MEASURING HETEROGENEITIES IN MULTIPLE-SCALE GEOLOGIC MEDIA

by

Susan Jennifer Colarullo

DISSERTATION

Presented to the Faculty of the Department of Earth and Environmental Science
in Partial Fulfillment of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

NEW MEXICO INSTITUTE OF MINING AND TECHNOLOGY
January 1998

To my daughter Rachel,
who gives new meaning to the phrase
'viscous flow'.

ABSTRACT

Characterization of aquifer heterogeneity in the presence of multiple-scale geologic structure is a fundamental problem facing hydrogeologists. In the presence of more than one scale of variation, randomness of hydrogeological properties such as hydraulic conductivity can be a function of observation scale, and spatial variation should be measured through instrument windows corresponding to scales that most dominate the particular flow or transport problem of interest.

A multiple-scale, multiple-realization design model was formulated to identify location and scale of hydraulic conductivity measurement that minimized concentration prediction uncertainty at a regulatory compliance point. The multiple-scale design was conditioned sequentially on log conductivity measurements collected at two scales of spatial variation. Five steady-state design scenarios were investigated. Preference for measurement at a given scale depended on how much log conductivity variance resided at that scale, source loading, size of source relative to scales of conductivity variation, distance to source, and principal directions of anisotropy relative to principal directions of transport.

Water content and tracer concentration observations collected during an infiltration/tracer test conducted adjacent to the Las Cruces trench were used as a basis for geostatistical validation of a two-scale saturated permeability model. Simulated ensemble variograms for a two-scale permeability model characterized by large-scale deterministic layering and small-scale intralayer random variation reproduced observed water content change variances better than a statistically-equivalent single-scale random permeability model at late experimental time. A future avenue of research is incorporation of large-scale MRF parameter uncertainty into the multiple-scale sequential design.

ACKNOWLEDGEMENTS

There are a number of people who should be recognized for contributing to this dissertation research. The greatest thanks is extended to all members of my dissertation committee. In particular, Dr. Allan Gutjahr's wisdom, guidance, technical expertise, and remarkable patience were singularly responsible for helping me bring this dissertation to completion. I also deeply appreciate the guidance and expertise of Drs. John Wilson and Fred Phillips during the course of this study. Thanks also to Dr. Bryan Travis for showing great enthusiasm when I presented some of my early ideas relating to scale issues, and for providing me with data from the Las Cruces trench infiltration experiment.

A special thanks to Edith Montoya, who offered me thoughtful words of encouragement on more than one occasion. Also thanks to my parents for their love, to Adeline, Baby Tom, and Big Tom DeForest, Linda Colarullo, Steve Snelgrove, and Kimberly Edlund for their friendship and support, and to Rachel Colarullo for tolerating countless late night visits to my office. I am grateful to Dr. Rick Aster, who showed unusual patience and consideration in allowing me to use his network of sun workstations for the design runs. Finally, I'd like to extend my appreciation to the user consultant named Kathy at the New Mexico Tech Computing Center, for helping to unravel the mysteries of an Interleaf problem that threatened completion of this dissertation.

Financial and technical support for this research was received from the Department of Energy Subsurface Science Program, Los Alamos and Sandia National Laboratories, and the University of Utah.

TABLE OF CONTENTS

Abstract iii
Acknowledgements iv

TABLE OF CONTENTS v
LIST OF TABLES x
LIST OF FIGURES xi

CHAPTER 1: INTRODUCTION: RANDOM GEOLOGIC STRUCTURE 1
1.1 STATEMENT OF THE PROBLEM 1
1.2 ORGANIZATION OF THE DISSERTATION 1
1.3 THE PROBLEM OF AQUIFER CHARACTERIZATION 3
1.3.1 Hydrogeologic Parameter Estimation 6
1.3.2 The Impact of Unresolved Heterogeneity on Predicted Mean Concentration 9
1.3.3 The Scale Effect 26
1.3.4 Multiple-Scale Characterization of Aquifer Heterogeneity 32
1.4 GEOSTATISTICAL CHARACTERIZATION OF DISTRIBUTED AQUIFER PROPERTIES 33
1.4.1 Single-Scale Random Processes 37
1.4.1.1 Second-Order Stationarity 38
1.4.1.2 The Ergodic Hypothesis 40
1.4.1.3 Intrinsic Random Processes 41
1.4.2 Multiple-Scale Random Processes 45
1.4.3 Randomness as a Function of Observation Scale 48
1.4.4 Multiple-Scale Variogram Behavior 54
1.5 SUMMARY 59
1.6 REFERENCES 60

CHAPTER 2: EFFECTIVE PROPERTIES AND MOVEMENT OF INFORMATION BETWEEN SPATIAL SCALES 67
2.1 THE REPRESENTATIVE ELEMENTARY VOLUME 69
2.2 CONSTITUTIVE VARIABLES AND THE 'IDEAL' INSTRUMENT 73
2.3 EQUIVALENT AND EFFECTIVE PROPERTIES 81

TABLE OF CONTENTS (Cont'd.)

2.4 MULTIPLE-SCALE GEOLOGIC STRUCTURE 87

2.5 EMBEDDING OF SCALES 94

2.5.1 Spatially-Continuous Versus Spatially-Discrete Geostatistical Models ... 97

2.5.2 Spatially-Discrete MRFs 98

2.6 EXAMPLES OF SINGLE-SCALE AND TWO-SCALE CONDUCTIVITY PROCESSES 99

2.7 DETERMINATION OF TWO-SCALE AND SINGLE-SCALE ENSEMBLE EFFECTIVE CONDUCTIVITY 112

2.8 DERIVATION OF ENSEMBLE MACRODISPERSIVITY TENSORS 128

2.9 TRANSPORT OF INFORMATION BETWEEN SCALES 138

2.10 SUMMARY 148

2.11 REFERENCES 149

CHAPTER 3: CHARACTERIZATION OF DISCRETE GEOLOGIC STRUCTURE USING MARKOV RANDOM FIELDS 154

3.1 LITHOLOGIC FACIES SEQUENCES AS ONE-DIMENSIONAL MARKOV CHAINS 155

3.2 MARKOV RANDOM FIELDS 159

3.2.1 Extension of the Markov Property to Higher-Dimensional Space 159

3.2.2 Two-Dimensional Gas Lattices and the Gibbs Distribution 162

3.2.3 The Joint Distribution 166

3.2.4 Conditional PDF as the Joint PDF Restricted to the Neighborhood 170

3.2.5 Local Interaction Energy in the Neighborhood 172

3.2.6 Markov Parameters 175

3.2.7 Importance Sampling Metropolis Algorithm 178

3.2.8 Maximum Likelihood Estimation of Markov Parameters 183

3.2.9 Goodness-of-Fit Hypothesis Testing 188

TABLE OF CONTENTS (Cont'd.)

3.3 APPLICATION OF MRFs TO CHARACTERIZATION OF DISCRETE GEOLOGIC STRUCTURE 190

3.3.1 Evolution of a 64-By-64 MRF Realization 190

3.3.2 Estimation of MRF Parameters for Alluvial-Fan Structure 192

3.3.3 χ^2 Goodness-of-Fit Testing 198

3.4 SUMMARY 204

3.5 REFERENCES 204

CHAPTER 4: OPTIMIZATION OF HYDRAULIC CONDUCTIVITY MEASUREMENT IN MULTIPLE-SCALE GEOLOGIC MEDIA .. 206

4.1 CONDUCTIVITY MEASUREMENT IN MULTIPLE-SCALE POROUS MEDIA 206

4.2 PREVIOUS RESEARCH RELATED TO PARAMETER MEASUREMENT NETWORK DESIGN 209

4.2.1 PARAMETER MEASUREMENT NETWORK DESIGN MODELS BASED ON KRIGING AND COKRIGING 210

4.2.2 DESIGN MODELS BASED ON PHYSICAL PROCESSES 213

4.3 FORMULATION OF A MULTIPLE-SCALE CONDUCTIVITY MEASUREMENT DESIGN MODEL 215

4.3.1 First-Order, Second-Moment Analysis 219

4.3.2 Single-Scale Conductivity Measurement Design 224

4.3.3 Multiple-Scale Conductivity Measurement Design 231

4.3.4 Multiple-Scale Conductivity Measurement Design Constrained by Multiple Conductivity Realizations 233

4.4 AVERAGING CONCENTRATION OVER DIFFERENT SCALES OF RESOLUTION 237

4.5 CALCULATION OF CONCENTRATION SENSITIVITIES USING ADJOINT TECHNIQUES 239

4.6 TRUE CONDUCTIVITY REALIZATIONS 255

4.7 OPTIMIZATION RESULTS 263

4.7.1 Optimal Conductivity Measurement Network Design 263

4.7.2 Design Problem 2: Reduced Y Variance 271

4.7.3 Problem 3: Continuous Source Loading 282

4.7.4 Problem 4: Reduced Source Extent 287

TABLE OF CONTENTS (Cont'd.)

4.7.5 Problem 5: Reduced Distance to Well 291

4.7.6 Rules of Thumb for Guiding Measurement of Conductivity in Multiple-Scale Geologic Media 295

4.7.7 Steady-State Vs. Transient Design 296

4.7.8 Conductivity Entropy Vs. Weighted, Squared Sensitivity 297

4.8 SUMMARY 299

4.9 REFERENCES 300

CHAPTER 5: GEOSTATISTICAL VALIDATION OF A MULTIPLE-SCALE AQUIFER CHARACTERIZATION MODEL 304

5.1 THE LAS CRUCES TRENCH INFILTRATION EXPERIMENTS 305

5.1.1 Characterization of Subsurface Saturated Permeability Structure 305

5.1.2 Experimental Design 316

5.1.2.1 One-Dimensional Infiltration/Tracer Experiment 316

5.1.2.2 Two-Dimensional Infiltration/Tracer Experiment 318

5.2 EXPERIMENTAL RESULTS AND PREVIOUS NUMERICAL TRANSPORT PREDICTIONS 321

5.2.1 One-Dimensional Infiltration/Tracer Experiment 321

5.2.2 Two-Dimensional Infiltration/Tracer Experiment 323

5.3 JUSTIFICATION FOR MULTIPLE-SCALE MODEL OF HETEROGENEITY IN LAS CRUCES TRENCH 328

5.3.1 Observed Variations in Spatial Structure Between Soil Horizons 328

5.3.2 Large Vertical Variation in Water Content 329

5.4 TWO-DIMENSIONAL VALIDATION OF STATE VARIABLE SPATIAL MOMENTS 330

5.4.1 TRACR3D Numerical Flow and Transport Simulations 330

5.4.2 Comparison of Observed and Simulated Ensemble Variogram Behavior 333

TABLE OF CONTENTS (Cont'd.)

5.4.2.1 Observed and Simulated Ensemble Water Content Change Variograms 333

5.4.2.2 Observed and Simulated Ensemble Bromide Concentration Variograms 338

5.4.3 Macrodispersive Flux 341

5.5 SUMMARY 344

5.6 REFERENCES 345

CHAPTER 6: CONCLUSIONS AND FUTURE RESEARCH DIRECTIONS 347

6.1 CONCLUSIONS 347

6.2 FUTURE RESEARCH DIRECTIONS 350

6.3 REFERENCES 358

APPENDIX A. MACRODISPERSIVITY VS. TIME PLOTS FOR INFORMATION TRANSFER EXPERIMENT A-1

APPENDIX B. DERIVATION OF LOG LIKELIHOOD DERIVATIVES B-1

APPENDIX C. DERIVATION OF ADJOINT EQUATION DERIVATIVE TERMS C-1

C.1 Derivation of the $\nabla_\sigma g(\bar{\sigma}, Y)$ Matrix C-2

C.2 Derivation of the $\nabla_Y g(\bar{\sigma}, Y)$ Matrix C-5

C.1 References C-7

APPENDIX D. SIMULATED WATER CONTENT CHANGE AND BROMIDE CONCENTRATION REALIZATIONS D-1

LIST OF TABLES

Table 2.1. Small-Scale Statistical Moments for Two-Scale Ln K Realizations. 102

Table 2.2 Node Equations in 5-By-5 Periodic Boundary Condition Example with Head Drop = Δh in the X-Direction and $\Delta x = \Delta y$. 118

Table 3.1 Results of Maximum Likelihood MRF Parameter Estimation and χ^2 Goodness-of-Fit Testing 201

Table 4.1 Statistical Parameters for Design Problem 1 257

Table 4.2 Results of Conductivity Measurement Design 265

Table 4.3 Specifications for Conductivity Measurement Network Design Problems 270

Table 4.4 Results of Conductivity Measurement Design for Problem 2 275

Table 4.5 Results of Conductivity Measurement Design for Problem 3 286

Table 4.6 Results of Conductivity Measurement Design for Problem 4 290

Table 4.7 Results of Conductivity Measurement Design for Problem 5 294

Table 5.1 Deterministic Parameters of TRACR3D Flow and Transport Simulations 308

LIST OF FIGURES

Figure 1.1 Construction of a Variogram.. 44

Figure 1.2 A Gedanken Experiment 49

Figure 1.3 Measurement Resolution and Parameter Uncertainty as Sources of Randomness 51

Figure 1.4 Measurement of Heterogeneities at Different Scales 53

Figure 1.5 A Two-Scale Process 55

Figure 1.6 Superimposed Variogram for Two-Scale Process in the Principal Longitudinal Direction of Anisotropy 57

Figure 2.1 REV of a Physical Property.. 70

Figure 2.2 Instrument Windows of a Discrete Geologic Hierarchy.. 75

Figure 2.3 Spectral Density of Ideal Instruments with Different Characteristic Frequencies.. 81

Figure 2.4 Sierra Ladrones Formation Deposits at the 1-Meter Observation Scale.. 88

Figure 2.5 Sierra Ladrones Formation Deposits at the 30-Meter Observation Scale... 89

Figure 2.6 Randomness and Structure as a Function of Observation Scale.. 92

Figure 2.7 Embedding Multiple Scales of Heterogeneity.. 95

Figure 2.8 Unconditioned Two-Scale and Statistically-Equivalent Single-Scale Conductivity Realizations.. 100

Figure 2.9 Rotation of Coordinate System.. 103

Figure 2.10 a. Two-Dimensional Ensemble Covariance Function for Two-Scale Ln K Process 110

b. Transects of Two-Dimensional Ensemble Covariance Function for Two-Scale Ln K Process in Longitudinal and Transverse Directions.. 111

Figure 2.11 Imposition of Periodic Boundary Conditions 115

Figure 2.12 Node Numbering for 5-By-5 Periodic Boundary Condition Example ... 119

Figure 2.13 Solution Matrix Structure for 5-By-5 Periodic Boundary Condition Problem 120

Figure 2.14 Four Adjacent Plots of Hydraulic Head Resulting from Application of Periodic Boundary Conditions 124

Figure 2.15 Traditional Boundary Conditions Used for Estimating Effective Conductivity 125

LIST OF FIGURES (Cont'd.)

Figure 2.16 Diagonalized Ensemble Log Effective Conductivity Tensors for Unconditioned Conductivity Processes (ln m/s) 127

Figure 2.17 Macrodispersivity vs. Time for 4 Two-Scale Conductivity Realizations 132

Figure 2.18 Macrodispersivity vs. Time for 4 Single-Scale Conductivity Realizations 134

Figure 2.19 Diagonalized Ensemble Macrodispersivity Tensors for Unconditioned Conductivity Processes (m) 135

Figure 2.20 Locations of Noisy Permeameter Data Clusters 139

Figure 2.21 Conditioned Two-Scale and Statistically-Equivalent Single-Scale Conductivity Realizations 141

Figure 2.22 Diagonalized Ensemble Log Effective Conductivity Tensors for Conductivity Processes Conditioned on Noisy Permeameter Measurements Using Point-to-Block Kriging for Large-Scale Conditioning (ln m/s) 142

Figure 2.23 Diagonalized Ensemble Macrodispersivity Tensors for Conductivity Processes Conditioned on Noisy Permeameter Measurements Using Point-to-Block Kriging for Large-Scale Conditioning (m) 143

Figure 2.24 Diagonalized Ensemble Log Effective Conductivity Tensors for Two-Scale Conductivity Process Unconditioned and Conditioned on Noisy Permeameter Measurements Using Geometric Mean for Large-Scale Conditioning (ln m/s) 145

Figure 2.25 Diagonalized Ensemble Macrodispersivity Tensors for Two-Scale Conductivity Process Unconditioned and Conditioned on Noisy Permeameter Measurements Using Geometric Mean for Large-Scale Conditioning (m) 146

Figure 3.1 A Vertical Geologic Sequence as a One-Dimensional Markov Chain.. . 156

Figure 3.2 Generation of a Markov Random Field Realization Via Conditioning with Respect to the Neighborhood.. 161

Figure 3.3 Homogeneous Gas Lattice.. 163

Figure 3.4 MRF States for M=9, L=4.. 170

Figure 3.5 Third-Order Neighborhood System for Calculating Pair-Potential Interaction Energy.. 176

Figure 3.6 The Metropolis Algorithm.. 179

Figure 3.7 Factorization of Joint Pdf Into Component Conditional Pdfs.. 182

Figure 3.8 Mutually Orthogonal Codings for a First-Order MRF.. 185

LIST OF FIGURES (Cont'd.)

Figure 3.9 Evolution of an MRF Realization from a Relaxation Chain.. 191

Figure 3.10 Percent Successful Exchanges vs. Attempted Number of Exchanges N for First-Order 64-By-64 MRF Realization.. 193

Figure 3.11 Sewer-Line Trench Excavation at Kirtland Air Force Base.. 195

Figure 3.12 Observed Lithofacies Type Distribution Along Main Stem of Sewer Trench.. 196

Figure 3.13 First-Order Log Likelihood Functions for Lithofacies Type Data Using Various Forms of Interaction Energy.. 199

Figure 3.14 Simulated Lithofacies Type Distribution Along Main Stem of Sewer Trench Using Estimated MRF Parameters.. 203

Figure 4.1 Multiple-Scale Conductivity Measurement 207

Figure 4.2 Examples of Predicted Plumes at Two Scales of Simulation 238

Figure 4.3 Adjoint Coefficient Matrix Structure (Solving from Left) 252

Figure 4.4 True Conductivity Realizations at Two Scales 256

Figure 4.5 True Plume Realizations at Two Scales of Simulation 261

Figure 4.6 Conductivity Measurement Design Results 264

Figure 4.7 Ensemble and Predicted Variance Reduction Vs. Iteration Number 269

Figure 4.8 True Conductivity Realizations at Two Scales for Reduced Variance Designs 272

Figure 4.9 True Plume Realizations at Two Scales of Simulation for Design Problem 2 273

Figure 4.10 Conductivity Measurement Design Results, Problem 2 274

Figure 4.11 Random Conductivity Measurement Design Results 278

Figure 4.12 Ensemble Variance Reduction vs. Iteration Number for Systematic and Random Designs 279

Figure 4.13 True Plume Realizations at Two Scales of Simulation for Design Problem 3 283

Figure 4.14 Conductivity Measurement Design Results, Problem 3 285

Figure 4.15 True Plume Realizations at Two Scales of Simulation for Design Problem 4 288

Figure 4.16 Conductivity Measurement Design Results, Problem 4 289

Figure 4.17 True Plume Realizations at Two Scales of Simulation for Design Problem 5 292

Figure 4.18 Conductivity Measurement Design Results, Problem 5 293

Figure 4.19 Final Distributions of Large- and Two-Scale Log K Entropy and Weighted, Squared Sensitivity 298

LIST OF FIGURES (Cont'd.)

Figure 5.1 Instrumentation for the First Las Cruces Trench Infiltration Experiment 306

Figure 5.2 Statistical Properties of Saturated Log Permeability (ln cm/day) 311

Figure 5.3 Large-Scale Deterministic and Two-Scale Random Permeability at the Las Cruces Site 313

Figure 5.4 Single-Scale Permeability Structure at the Las Cruces Site 314

Figure 5.5 Ensemble Variogram Estimates for Single-Scale and Two-Scale Saturated Permeability Processes 315

Figure 5.6 Instrumentation for the Second Las Cruces Trench Infiltration Experiment 320

Figure 5.7 Observed Water Content Change Distributions (y=2.0 m) 324

Figure 5.8 Observed Bromide and Tritium Concentration Distributions (y=0.5 m) 326

Figure 5.9 Vertical Spatial Discretization of Las Cruces Transport Domain for TRACR3D Simulation 334

Figure 5.10 Variograms for Observed and Simulated Water Content Change Perturbations About Simulated Ensemble Mean 335

Figure 5.11 Variograms for Observed and Simulated Bromide Concentration Perturbations About Simulated Ensemble Mean 339

Figure 5.12 Maximum Absolute Ensemble Macrodispersive Flux for Two-Scale Permeability Process Conditioned at Large Scale and for Unconditioned Single-Scale Permeability Process 343

Figure 6.1 Incorporation of MRF Parameter Uncertainty into Measurement Network Design 354

Figure A.1 Macrodispersivity vs. Time for 100 Single-Scale Realizations

Figure A.2 Ensemble Macrodispersivity vs. Time for Single-Scale Realizations

Figure A.3 Macrodispersivity vs. Time for 100 Two-Scale Realizations

Figure A.4 Ensemble Macrodispersivity vs. Time for 100 Two-Scale Realizations

Figure D.1 Simulated Water Content Change Realizations for Single-Scale Permeability Model

Figure D.2 Simulated Water Content Change Realizations for Two-Scale Permeability Model

Figure D.3 Simulated Bromide Concentration Realizations for Single-Scale Permeability Model

Figure D.4 Simulated Bromide Concentration Realizations for Two-Scale Permeability Model

Chapter 1. Introduction: Random Geologic Structure

1.1 STATEMENT OF THE PROBLEM

The physical properties of many geologic media depend on the scale of averaging. In such geologic media, predictions of state variables will likewise be scale-dependent. This dissertation research focuses on the problem of designing a hydraulic conductivity measurement network model that minimizes state-variable prediction uncertainty when conductivity varies as a function of scale of variation. Specifically, the research is directed towards answering the question: At what spatial scales should hydraulic conductivity be measured if we wish to minimize concentration prediction uncertainty at a regulatory compliance point? The question is an important one for hydrogeologists faced with the problem of predicting effectiveness of ground water remediation strategies or managing subsurface water supplies affected by the presence of dissolved contaminants.

1.2 ORGANIZATION OF THE DISSERTATION

The dissertation is divided into six chapters, including this introductory chapter dealing with random geologic structure. Chapter 1 is used to state the problem of subsurface characterization, with particular regard to the problem of predicting

contaminant transport in the presence of unresolved aquifer heterogeneities. The concept that randomness can be a function of observation scale, and that our ability to measure unresolved heterogeneities can depend on observation scale, is intended to motivate the multiple-scale conductivity measurement design algorithm presented in Chapter 4.

The framework for multiple-scale conductivity measurement design is established in Chapters 2 and 3. In Chapter 2, the conventional representative elementary volume (REV) construct is abandoned in favor of a more realistic, relativist approach to characterizing effective aquifer heterogeneities. Results of a numerical experiment demonstrate that hydraulic conductivity information can be changed when moved from small to large spatial scales using widely-accepted linear upscaling rules. On the basis of this observation, we conclude that aquifer properties should be measured independently at each of the natural scales of variation characteristic of the geologic medium. Together with the ideas presented in the first two chapters, the motivation for constructing a multiple-scale measurement design model is established.

Chapter 3 functions as a tutorial on Markov random fields, which are used to describe spatially-discrete structure typically encountered at the facies and stratigraphic scales. With the conclusion of Chapter 3, we have developed all tools necessary for building the multiple-scale measurement design model.

Chapter 4 presents a measurement design algorithm that identifies hydraulic conductivity measurement locations and observation scales that minimize concentration prediction uncertainty at a regulatory compliance point. The design

algorithm is demonstrated for a variety of physical problems and general rules of thumb regarding optimal measurement scales are developed for these differing scenarios.

In Chapter 5, the multiple-scale aquifer characterization model is geostatistically validated using realizations of water content and tracer concentration distribution observed during the Las Cruces trench infiltration and tracer experiments performed in 1991. Water content and concentration variograms generated on the basis of two-scale saturated permeability realizations are shown to differ from variograms based on single-scale permeability structure.

Chapter 6 serves to present overall conclusions of the dissertation research, along with future avenues of study that might provide additional insight into the problem of aquifer characterization in the presence of spatially-discrete, hierarchical geologic structure.

1.3 THE PROBLEM OF AQUIFER CHARACTERIZATION

Based on liberal assumptions of plume size and rate of subsurface migration, it is believed that as much as 2% of the total volume of useable ground water supplies in the U.S. has been contaminated by septic tanks, landfills, and industrial sources of ground water pollution in the United States (Lehr, 1985). Much of this contamination was generated over a time period spanning several decades, when pollutant contributions from natural sources, human activities, and industrialization were poorly understood and not subject to management or regulation. In many cases, parties responsible for past incidences of pollution have since been held accountable for

remediation of contaminated aquifers. Many of the controllable contaminant sources have been identified and removed. Non-point sources such as those associated with municipal landfills and agricultural practices are being managed in ways that, to varying degrees, control adverse impacts on remaining uncontaminated ground water supplies.

Hydrogeologists have been instrumental in helping to reduce the number of controllable pollution sources and minimizing the impact of ongoing natural and agricultural contamination, but most recognize the limitations of current state-of-the-art strategies for remediating contaminated supplies and are are more concerned with assessing the potential for future contamination of ground water resources, particularly in the context of landfill and nuclear waste repository siting, management, and containment. Central to the problem of predicting the fate of municipal and nuclear wastes in the subsurface is characterization of aquifer properties that influence migration of ground water and pollutants.

Since the distributions of aquifer properties are as varied as the natural geologic processes that give rise to them, aquifer characterization is by nature an extremely site-specific endeavor. Fortunately, when we deal with regional ground water flow, the details of site-specific geologic structure tend to be irrelevant — in lay terms, the effects of these details tend to average out at the regional scale where water is concerned. For example, local anomalies such as high-conductivity channel scours or fractures characterized by lengths on the order of tens of meters will typically not have significant impact when regional ground water flow over distances of tens of kilometers is

considered. Most aquifer characterization studies performed until the 1970's focused on exactly these types of regional flow problems.

However, the problem of predicting the fate of dissolved contaminants in the subsurface often requires that we pay close attention to these details. Contaminant transport predictions depend on whether a representative fluid element is either marked or unmarked with a contaminant. Prediction of how much contaminant moves across a regulatory boundary is determined by the movement of each marked fluid element, separate and distinct from the unmarked ones. This differs radically from the case of fluid flow, for which each fluid element is indistinguishable from every other fluid element, and the exact source and migration pathway of any given element does not affect how much fluid moves across a flow boundary. The differences in the movement of fluid and contaminant are mirrored by their associated partial differential equations — the parabolic nature of the flow equation dictates that hydraulic energy is dispersed or dissipated through the system much more rapidly than contaminant mass, whose movement is governed by a more hyperbolic-like partial differential equation characterized by slower rates of dissipation. As a result, prediction of the near-source potential impact of a municipal landfill or nuclear waste repository will typically demand that preferential migration pathways and other local aquifer anomalies that dictate minimum travel time to ground water receptor locations be delineated, as opposed to average global aquifer properties.

From these considerations, it is evident that gaining an understanding of local, site-specific aquifer structure can be critical to making reliable predictions of

contaminant migration in the subsurface. One way of accounting for the complexity of natural geologic media would be to build a deterministic transport model that incorporates all heterogeneities, including the small-scale details, in an explicit manner. In general, however, the costs of measuring all geologic properties that might influence the transport prediction, particularly the small-scale ones, are prohibitive. Moreover, the very act of measuring the aquifer property is typically so invasive that it can fundamentally and unpredictably alter the local flow geometry and change the nature of the contamination problem being investigated, with the unanticipated outcome that the measurement renders itself invalid. Even if one were to have access to property measurements at every point in space, the logistical difficulties of incorporating the data into a predictive model are generally too great to justify acquiring such a data set. When confronted with such problems, the hydrogeologist can benefit from geostatistical methods that do not require a value of the aquifer property be observed at every location. The inferential power of geostatistical methods can be exploited to characterize aquifer heterogeneities based on a limited number of aquifer property measurements.

1.3.1 Hydrogeologic Parameter Estimation

Numerical models of ground water flow and contaminant transport provide much of the basis for decisions about remediation of contaminated ground water supplies or siting of landfills and repositories, but the reliability of these models is only as good as the data used to parameterize them. A typical distributed-parameter

ground water flow model requires estimation of literally hundreds of parameters associated with aquifer properties, boundary geometries and conditions, and contaminant source location and magnitude. Most of these parameters are not directly measurable and can only be inferred from observation of hydraulic heads and contaminant concentrations using inverse models, which are notoriously ill-posed due to the nature of the partial differential equations describing fluid flow and, to a lesser degree, solute transport. In the case of fluid flow, ill-posedness produces extremely large errors in hydraulic parameters when small head observation errors are present. As a result, it becomes difficult to distinguish between alternative sets of parameters, because the sets produce similar distributions of head. A survey of parameter-estimation research is provided by Yeh (1986) and, more recently, Sun (1994).

Neuman (1973) attempted to overcome the problem of ill-posedness by minimizing an objective function defined as the sum of squared head prediction errors at observation points. Neuman and Yakowitz (1979) later introduced a Bayesian component by incorporating a plausibility constraint to penalize parameter estimates that differ greatly from available parameter measurements. Kalman filtering approaches to parameter identification have also been proposed (Wilson *et al.*, 1978). Recognizing that ill-posedness can at least partly be attributed to overparameterization, Emsellem and deMarsily (1971), Cooley (1977; 1979), Yakowitz and Duckstein (1980), Neuman (1980), and Yeh and Yoon (1981) reduced parameter dimension via regularization or zonation techniques. McLaughlin and Townley (1996)

recast the inverse problem using functional analysis, which can easily be used to address the issue of ill-posedness.

Compounding the problem of ill-posedness are difficulties of characterizing extreme geologic heterogeneities that occur naturally in the subsurface, where hydraulic conductivity can vary over many orders of magnitude within any given aquifer (Bakr *et al.*, 1978). Freeze and Cherry (1979) estimate that conductivity varies over 13 orders of magnitude when a wide range of geologic materials is considered. In the presence of such large variations in conductivity, characterization of local heterogeneities required for reliable contaminant transport prediction is problematic.

More recently, geostatistical approaches have been used to substantially reduce the number of parameters required for model identification (Neuman, 1982). While the number of parameters needed to explicitly identify conductivity at every location in a deterministic model comprised of N nodes is N, the number of parameters needed for a geostatistical model will depend only on the order of the probability law believed to govern the spatial distribution of conductivity. The geostatistical identification problem is solved by first determining geostatistical moments via maximum-likelihood or least-squares techniques. Following solution of the geostatistical direct problem, kriging or conditional gaussian simulation is performed to obtain minimum mean-square-error, unbiased parameter estimates everywhere in the domain (Clifton and Neuman, 1982; Kitanidis and Vomvoris, 1983; Kitandis, 1985b; Hoeksema and Kitanidis, 1984; Carrera and Neuman, 1986a,b,c).

Based on these considerations, it is evident that the problem of parameter estimation is far from trivial. The practicing hydrogeologist is confronted with the problem of constructing a reliable physical model based on incomplete knowledge of the very parameters that drive the model and in the presence of numerical difficulties associated with mathematical inversion. Not only must the hydrogeologist infer key aquifer parameters from an incomplete set of state-variable data, but he or she must be sufficiently confident in model predictions to inform decision makers about the efficacy of proposed remediation or siting strategies. While these problems stem from incomplete knowledge of the parameter set and can never be practically overcome, the field of geostatistics offers an ideal framework for systematically quantifying parameter estimation uncertainties, particularly with regard to how they impact our ability to make reliable predictions about the fate of contaminants in the subsurface. Specifically, geostatistics provides an operational framework within which one can control prediction uncertainty through rational design of parameter measurement networks.

1.3.2 The Impact of Unresolved Heterogeneity on Predicted Mean Concentration

If we cannot resolve all heterogeneity through parameter estimation, it might be instructive to develop an understanding of how these unresolved aquifer property variations affect our ability to make valid transport predictions. It is now widely accepted by the hydrogeologic community that migration of dissolved contaminants in the subsurface is driven primarily by the movement of ground water via advection.

Chapter 1. Introduction: Random Geologic Structure

Contaminant spreading previously attributed to large-scale values of field dispersivity is now recognized as principally due to differential advection caused by spatial variations in aquifer hydraulic properties. As discussed in section 1.1, an important factor affecting the reliability of contaminant migration prediction in the subsurface is understanding how these aquifer properties are distributed in space.

Although the advective migration of contaminants is driven by fluid flux, solute and fluid flux are characterized by different types of partial differential equations. The transient movement of ground water is governed by a second-order parabolic differential equation, while transient migration of dissolved contaminants in two-dimensional space is described by the second-order hyperbolic-like advection-dispersion equation (ADE) (Bear, 1979):

$$\frac{\partial}{\partial x_i}\left(D_{ij}\,\frac{\partial c}{\partial x_j}\right) - \frac{\partial}{\partial x_i}(v_i\,c) = \frac{\partial c}{\partial t} \qquad i, j = 1, 2 \qquad [1.1]$$

where c is spatially- and temporally-variable contaminant concentration [M/L^3], x_i, i= 1,2 are two-dimensional Cartesian coordinates [L], D_{ij} is the local, spatially-invariant dispersion tensor [L^2/T], v_i is distributed velocity in the ith-direction [L/T], and t is time [T]. Equation [1.1] is a mixed hyperbolic-parabolic equation, possessing both hyperbolic and parabolic features. It is the competition between these features that make it so difficult to solve mathematically.

Equation [1.1] is written in the standard Cartesian summation of Einstein. It is subject to initial and boundary conditions specific to the transport problem being investigated. For this study, it is assumed that flow is steady-state and velocity not a function of time. Note that to simplify the analysis, advection-dispersion equation [1.1] is presented for a two-dimensional transport problem.

The first term in advection-dispersion equation [1.1] represents dispersive transport of dissolved contaminants, and includes both molecular diffusion and mechanical mixing. Elements of the dispersion tensor are given by (Bear, 1979):

$$D_{ij} = D^* + \alpha_L \bar{v} \delta_{ij} + (\alpha_L - \alpha_T) \frac{\bar{v}_i \bar{v}_j}{\bar{v}} \qquad i, j = 1, 2 \qquad [1.2]$$

where D^* is the molecular diffusion coefficient for the particular solute being transported [L^2/T], modified by a tortuosity factor to account for the microscopic structure of the porous medium (Reddell and Sunada, 1970; Bear, 1979). α_L and α_T [L] represent longitudinal and transverse characteristic mixing lengths, or dispersivities. $\bar{v}_i$ and $\bar{v}_j$ are mean velocities in the ith- and jth-directions [L/T], $\bar{v}$ is mean resultant velocity [L/T], and δ_{ij} is the dimensionless Kronecker delta function.

In equations [1.1] and [1.2], it is often assumed that dispersivities α_L and α_T are spatially- and temporally-invariant and that dispersion can be treated as a diffusion, or Fickian, process that is dependent only on the gradient of concentration in space. One

characteristic of diffusive processes is that the spatial distribution of the diffusive variable assumes a normal, or gaussian, distribution in space about a centroid and that the rate of increase in spatial variance about the centroid with time be constant. Bear (1972) has demonstrated that dispersion in a homogeneous porous medium can be considered diffusive only when each tracer particle is displaced over many velocity correlation scales. His work was extended to a heterogeneous medium by Schwartz (1977), who used a particle tracker to demonstrate that advective migration of contaminants is of an essentially Fickian nature when the log conductivity distribution is gaussian-distributed and the scale of transport is sufficiently large relative to the scale of heterogeneities to produce many independent samples of velocity. He demonstrated that over sufficiently large displacement distance, local advection of contaminants produces concentration distributions that are virtually no different than those predicted on the basis of Fick's law of diffusion. He concluded that, under such Fickian conditions, dispersion can be treated as a global diffusive process.

This equivalence between dispersion and Fickian diffusion is significant because it allows prediction of transport in the presence of unresolved small-scale variations in velocity. However, in practice, numerical transport simulations conducted by Smith and Schwartz (1980) suggest that non-gaussian concentration distributions are the rule rather than the exception, and that spatially- and temporally-invariant dispersivities are unlikely to occur at field scales of observation. They attribute non-gaussian behavior to insufficient sampling of the velocity distribution by the plume. Under such conditions a contaminant particle does not

become fully informed of all possible velocity variations, total particle displacement is not the sum of statistically independent displacements, and the central limit theorem cannot be invoked (Scheidegger, 1964). Violation of the central limit theorem can occur even if the majority of particles encounter the full range of velocity variation, because the entire ensemble of particles must be fully informed on velocity variation if they are to collectively define a gaussian distribution in space.

Results of the numerical experiments of Smith and Schwartz (1980) suggest that, even in porous media characterized by a single scale of heterogeneity, a particle plume is rarely displaced a sufficient distance to sample all independent velocity variations. In real field situations, transport over tens of representative spatial scales may be required before differential advection attains its Fickian or scale-invariant value. In the majority of cases this limit is never approached. In some cases, it is approached in one direction, but not others. When displacement is sufficient and asymptotic values of dispersivity do exist, field-scale prediction of solute migration becomes a simple matter of replacing the small-scale dispersivities α_L and α_T with their effective field-scale Fickian counterparts, macrodispersivities A_L and A_T.

For a porous medium characterized by evolving scales of heterogeneity, sufficient displacement is even less likely to be realized because under such circumstances the plume continues to encounter different scales of velocity variation with increasing displacement. In such a case, mechanical mixing cannot be considered a 'scaled-up' version of Fickian diffusion. When more than one scale of variation is spanned by the problem, mechanical mixing must instead be explicitly considered as

an advection-based process, produced by spatial variations in velocity occurring at all scales. At the smallest scale, these variations manifest as tortuosity of flow paths through pore spaces. At intermediate and larger scales, velocity variations occur as fluid encounters boundaries between lithofacies and stratigraphic facies. These spatial variations in velocity serve to deflect water and dissolved contaminants from the average flow path defined by $\bar{v}_1$ and $\bar{v}_2$, causing dissolved contaminants to migrate through the subsurface in a non-Fickian manner. Under such circumstances, mechanical mixing should be treated as the non-Fickian process that it is, by removing it from consideration in the first, dispersive term of equation [1.1] and instead accounting for it in the second, advective term. This, however, requires that heterogeneities responsible for the non-Fickian behavior be fully resolved through knowledge of distributed velocities v_1 and v_2 in equation [1.1] rather than mean velocities $\bar{v}_1$ and $\bar{v}_2$ in equation [1.2].

When mechanical mixing is considered an advective rather than diffusive process and incorporated explicitly into the advective term of equation [1.1], D_{ij} reflects only microscopic diffusion independent of velocity or space and reduces to D^* for all i,j. Expanding the implied summation in equation [1.1], the transport equation for isotropic diffusion can be rewritten as:

$$D^* \frac{\partial^2 c}{\partial x_1^2} + D^* \frac{\partial^2 c}{\partial x_2^2} - \frac{\partial}{\partial x_1}(v_1 c) - \frac{\partial}{\partial x_2}(v_2 c) = \frac{\partial c}{\partial t} \qquad [1.3]$$

Chapter 1. Introduction: Random Geologic Structure

To demonstrate how unresolved velocity variations can influence predicted concentration, consider the transport equation given by [1.3] with both concentration and velocity expressed as first-order expansions about their means:

$$D^* \frac{\partial^2[\bar{c} + c']}{\partial x_1^2} + D^* \frac{\partial^2[\bar{c} + c']}{\partial x_2^2} - \frac{\partial[(\bar{v}_1 + v_1')(\bar{c} + c')]}{\partial x_1} - \frac{\partial[(\bar{v}_2 + v_2')(\bar{c} + c')]}{\partial x_2} = \frac{\partial[\bar{c} + c']}{\partial t} \qquad [1.4]$$

The expansions $c=\bar{c}+c'$ and $v_i=\bar{v}_i+v_i'$, $i=1,2$ are generally valid for small perturbations c' and v_i'. For simplicity of notation, the spatial and temporal dependence of mean concentration $\bar{c}$ and concentration perturbation c' as well as the spatial dependence of mean velocities $\bar{v}_1$ and $\bar{v}_2$ and velocity perturbations v_1' and v_2' are not indicated in the expansions. Substitution of first-order expansions about mean values in equations of flow and transport is often referred to as linearization, and represents a useful tool for assessing how small changes in variables or parameters affect the behavior of a system. The perturbations are assumed to be characterized by zero-valued means and well-defined spatial correlation structure. Note that in equation [1.4], D^* is treated as a deterministic quantity because the effects of variations in diffusion on predicted transport have been shown to be insignificant compared to variations in hydraulic conductivity (Gelhar *et al.*, 1979). Also note that velocity perturbations are assumed to occur in both x_1 and x_2 directions.

Chapter 1. Introduction: Random Geologic Structure

Expanding the terms of equation [1.4] yields:

$$D^* \frac{\partial^2 \bar{c}}{\partial x_1^2} + D^* \frac{\partial^2 c'}{\partial x_1^2} - \frac{\partial(\bar{v}_1 \bar{c})}{\partial x_1} - \frac{\partial(\bar{v}_1 c')}{\partial x_1} - \frac{\partial(v_1' \bar{c})}{\partial x_1} - \frac{\partial(v_1' c')}{\partial x_1} +$$

$$D^* \frac{\partial^2 \bar{c}}{\partial x_2^2} + D^* \frac{\partial^2 c'}{\partial x_2^2} - \frac{\partial(\bar{v}_2 \bar{c})}{\partial x_2} - \frac{\partial(\bar{v}_2 c')}{\partial x_2} - \frac{\partial(v_2' \bar{c})}{\partial x_2} - \frac{\partial(v_2' c')}{\partial x_2} = \frac{\partial \bar{c}}{\partial t} + \frac{\partial c'}{\partial t} \qquad [1.5]$$

Taking the expectation of equation [1.5] to obtain the mean transport equation, and assuming that the expected values of velocity and concentration perturbations are negligible, the equation for mean concentration is:

$$D^* \frac{\partial^2 \bar{c}}{\partial x_1^2} + D^* \frac{\partial^2 \bar{c}}{\partial x_2^2} - \frac{\partial(\bar{v}_1 \bar{c})}{\partial x_1} - \frac{\partial(\bar{v}_2 \bar{c})}{\partial x_2} - \left[\frac{\partial E(v_1' c')}{\partial x_1} + \frac{\partial E(v_2' c')}{\partial x_2} \right] = \frac{\partial \bar{c}}{\partial t} \qquad [1.6]$$

Aside from being expressed in terms of mean concentration $\bar{c}$ and mean velocities $\bar{v}_1$ and $\bar{v}_2$ rather than c, v_1, and v_2, the form of equation [1.6] differs from that associated with equation [1.3] in the inclusion of derivatives of unexplained flux terms $E(v_1'c')$ and $E(v_2'c')$ where $v_1'c'$ and $v_2'c'$ are characterized by the dimension of flux, $[M/L^2T]$.

Equation [1.6] can be further simplified by assuming x_1 and x_2 are principal transport directions, with average flow in the x_1 direction and $\bar{v}_2=0$. Rewriting

equation [1.6] for $\bar{v}_2=0$ and spatially-invariant $\bar{v}_1$,

$$D^* \frac{\partial^2 \bar{c}}{\partial x_1^2} + D^* \frac{\partial^2 \bar{c}}{\partial x_2^2} - \bar{v}_1 \frac{\partial \bar{c}}{\partial x_1} - \left[\frac{\partial E(v_1'c')}{\partial x_1} + \frac{\partial E(v_2'c')}{\partial x_2} \right] = \frac{\partial \bar{c}}{\partial t} \qquad [1.7]$$

If advection-driven dispersion acts globally as a diffusive process, the unexplained flux terms, referred to as macrodispersive fluxes, can be restated for zero-mean perturbations using Fick's law:

$$E(v_1'c') = Cov(v_1'c') = - D_L \frac{\partial \bar{c}}{\partial x_1}$$

$$E(v_2'c') = Cov(v_2'c') = - D_T \frac{\partial \bar{c}}{\partial x_2}$$

where D_L and D_T are large-scale effective dispersion coefficients. If it is further assumed that equation [1.2] holds for macrodispersivity with molecular diffusion D^* already accounted for in the first two terms of equation [1.7], then

$$E(v_1'c') = - A_L \bar{v}_1 \frac{\partial \bar{c}}{\partial x_1}$$

$$E(v_2'c') = - A_T \bar{v}_1 \frac{\partial \bar{c}}{\partial x_2}$$

where A_L and A_T are longitudinal and transverse macrodispersivities, $\bar{v}_2$ is equal to 0, and $\bar{v}_1$ is the magnitude of mean velocity in the longitudinal direction. Taking the spatial derivatives of $E(v_1'c')$ and $E(v_2'c')$ in the longitudinal (x_1) and transverse (x_2) directions yields:

$$\frac{\partial E(v_1'c')}{\partial x_1} = - A_L \bar{v}_1 \frac{\partial^2 \bar{c}}{\partial x_1^2}$$

$$\frac{\partial E(v_2'c')}{\partial x_2} = - A_T \bar{v}_1 \frac{\partial^2 \bar{c}}{\partial x_2^2} \qquad [1.8]$$

Equation [1.7] can be rewritten using equation [1.8] as follows:

$$\left[D^* + A_L \bar{v}_1\right] \frac{\partial^2 \bar{c}}{\partial x_1^2} + \left[D^* + A_T \bar{v}_1\right] \frac{\partial^2 \bar{c}}{\partial x_2^2} - \bar{v}_1 \frac{\partial \bar{c}}{\partial x_1} = \frac{\partial \bar{c}}{\partial t} \qquad [1.9]$$

Equation [1.9] suggests that, to first-order with respect to concentration and velocity, contaminant migration due to velocity variations about the mean in a system

characterized by Fickian behavior is the sum of local and global diffusive behavior. Since local diffusion occurs at a molecular scale, it is generally much smaller than global diffusion and equation [1.9] can be approximated as:

$$A_L \bar{v}_1 \frac{\partial^2 \bar{c}}{\partial x_1^2} + A_T \bar{v}_1 \frac{\partial^2 \bar{c}}{\partial x_2^2} - \bar{v}_1 \frac{\partial \bar{c}}{\partial x_1} = \frac{\partial \bar{c}}{\partial t} \qquad [1.10]$$

Unlike the case for equation [1.7], the influence of small-scale heterogeneities on transport has not been considered in equation [1.10]. Instead, this influence has been assumed to act the same as a diffusion process at the global scale. Equation [1.7] and [1.10] are of similar form except for inclusion of macrodispersive flux terms in equation [1.7] that explicitly account for heterogeneities left unresolved in equation [1.10]. In equation [1.7], macrodispersive fluxes $E(v_1'c')$ and $E(v_2'c')$ account for contaminant spreading caused by velocity variations about the mean occurring at a scale too large to be included in the diffusive terms $D^* \partial^2\bar{c}/\partial x_1^2$ and $D^* \partial^2\bar{c}/\partial x_2^2$. In equation [1.10], these same fluxes are considered to be produced by velocity variations that are too small-scaled to be resolved by the mean longitudinal velocity $\bar{v}_1$, and incorporated into the third advective term $\bar{v}_1 \partial\bar{c}/\partial x$. Instead, they are lumped into the first two diffusive terms.

Given this perspective, macrodispersive flux at the large scale can be viewed as unexplained flux due to random variation in velocity and concentration. This unexplained flux is equal to total mass flux minus flux that can be explained using

estimates of mean concentration and velocity. To demonstrate this, consider macrodispersive flux in the x_1-direction:

$$E(v_1'c') = E[(v_1 - \bar{v}_1)(c - \bar{c})] = E(v_1c) - \bar{v}_1E(c) - \bar{c}E(v_1) + E(\bar{v}_1)E(\bar{c})$$

$$= E(v_1c) - \bar{v}_1\bar{c} - \bar{c}\bar{v}_1 + \bar{v}_1\bar{c} = E(v_1c) - \bar{v}_1\bar{c} \qquad [1.11]$$

We wish to minimize macrodispersivity terms because they embody everything that we *don't* understand about the transport process. We can do this only by better resolving velocity perturbations v_1' and v_2' through local measurement of conductivity and incorporating that knowledge into our estimate of mean velocities $\bar{v}_1$ and $\bar{v}_2$ via conditioning. By so doing, v_1' and v_2' tend to vanish, $E(v_1'c')$ and $E(v_2'c')$ go to zero, A_L and A_T disappear, equation [1.7] reduces to deterministic equation [1.3], and linearized equation [1.10] can be used to make valid predictions of mean concentration. Alternatively, with full knowledge of the distribution of heterogeneities, a Lagrangian approach can be used to make solute flux predictions along streamlines originating from the source (Dagan *et al.*, 1992).

We have demonstrated that if unresolved velocity variations are present they contribute to unknown macrodispersivity flux terms, and that by resolving velocity perturbations v_1' and v_2', one is justified in using the deterministic form of the advection-dispersion equation given by equation [1.3]. From equation [1.7] it is evident that in the presence of unresolved velocity variations, predicted mean concentration $\bar{c}$

will depend on how strongly velocity variations are correlated to concentration perturbations. Since the degree of correlation will be a function of the flow geometry of the problem, it is application dependent and therefore not readily known.

We are primarily interested in how conductivity variations affect our ability to make transport predictions. It might be instructive to investigate how unresolved, small-scale conductivity heterogeneities contribute to distributed velocity using yet another first-order analysis, this time applying it to Darcy's law. Specifically, we want to determine how conductivity perturbations influence distributed velocity in equation [1.3]. Using Darcy's law, longitudinal velocity $v_1(\mathbf{x})$ is described as a function of hydraulic conductivity and hydraulic head gradient in the x_1-direction:

$$v_1(\mathbf{x}) = -\frac{K(\mathbf{x})}{n}\frac{\partial h(\mathbf{x})}{\partial x_1} \qquad [1.12]$$

where $v_1(\mathbf{x})$, $K(\mathbf{x})$, and $h(\mathbf{x})$ are spatially-distributed longitudinal velocity [L/T], isotropic hydraulic conductivity [L/T], and hydraulic head [L] at vector location $\mathbf{x}$, and n is spatially-invariant effective porosity [−]. The effects of distributed conductivity perturbations on velocity can be evaluated by expanding $v_1(\mathbf{x})$, $K(\mathbf{x})$, and $h(\mathbf{x})$ about their local means using the first-order Taylor series approximations:

$$\begin{aligned} v_1(\mathbf{x}) &= \bar{v}_1(\mathbf{x}) + v_1'(\mathbf{x}) \\ K(\mathbf{x}) &= \overline{K} + K'(\mathbf{x}) \\ h(\mathbf{x}) &= \bar{h}(\mathbf{x}) + h'(\mathbf{x}) \end{aligned} \qquad [1.13]$$

where $\bar{v}_1(\mathbf{x})$, $\overline{K}$, and $\bar{h}(\mathbf{x})$ are large-scale, slowly varying or constant deterministic functions representing mean longitudinal seepage velocity, conductivity, and head, while $v_1'(\mathbf{x})$, $K'(\mathbf{x})$, and $h'(\mathbf{x})$ are first-order, unresolved, random perturbations about the mean. Again, the perturbations are assumed to be characterized by zero-valued means and well-defined spatial correlation structure.

Substituting these linearizations into Darcy's law [1.12] yields:

$$v_1(\mathbf{x}) = \bar{v}_1(\mathbf{x}) + v_1'(\mathbf{x}) = -\left[\frac{\overline{K} + K'(\mathbf{x})}{n}\right]\left[\frac{\partial[\bar{h}(\mathbf{x}) + h'(\mathbf{x})]}{\partial x_1}\right] \qquad [1.14]$$

Assuming that the mean hydraulic gradient $\bar{J}=-\partial\bar{h}(\mathbf{x})/\partial x_1$ is spatially-invariant, we can expand the products as follows:

$$v_1(x) = \bar{v}_1(x) + v_1'(x)$$

$$= -\frac{1}{n}\left[\bar{K}\frac{\partial \bar{h}(x)}{\partial x_1} + \bar{K}\frac{\partial [h'(x)]}{\partial x_1} + K'(x)\frac{\partial \bar{h}(x)}{\partial x_1} + K'(x)\frac{\partial [h'(x)]}{\partial x_1}\right]$$

$$= -\frac{1}{n}\left[-\bar{K}\,\bar{J} + \bar{K}\frac{\partial [h'(x)]}{\partial x_1} - K'(x)\,\bar{J} + K'(x)\frac{\partial [h'(x)]}{\partial x_1}\right]$$

[1.15]

The fourth and final term of equation [1.15] involves second-order perturbation products that are generally considered to be negligible. While there exists some debate over the appropriateness of such an assumption, there appears to be a substantial body of evidence from Monte Carlo flow simulations, indicating that products of head and conductivity perturbations can generally be ignored even when variances of conductivity perturbations K'(x) are large, without seriously compromising linearized analytic solutions of the stochastic flow equation (Cushman, 1983; Gelhar and Axness, 1983a,b). However, the degree to which this is true for head gradient and conductivity perturbations is open to question.

For purposes of assessing the effect of unresolved conductivity heterogeneity on velocity in the general case, the fourth term of equation [1.15] is not considered negligible. Equation [1.15] suggests that, if we are interested in resolving distributed velocity v_1 as input to equation [1.3], we must have knowledge about not only mean

conductivity and mean hydraulic gradient but also about the distribution of conductivity and gradient perturbations in space. To the extent that gradient perturbations can be predicted from conductivity perturbations using the ground water flow equation, the problem of resolving distributed velocity appears to reduce to one of resolving the distribution of conductivity perturbations.

Note that, even if one were to substitute mean velocity $\bar{v}_1(\mathbf{x})$ for $v_1(\mathbf{x})$ in equation [1.3], information about the spatial distribution of conductivity and gradient perturbations is still required. To demonstrate this, the mean effect of unresolved small-scale conductivity variations on velocity is assessed by taking the expectation of equation [1.15]:

$$E[v_1(\mathbf{x})] = E[\bar{v}_1(\mathbf{x}) + v_1'(\mathbf{x})] = E[\bar{v}_1(\mathbf{x})] = \bar{v}_1(\mathbf{x})$$

$$= -\frac{1}{n} E\left[-\overline{K}\,\overline{J} + \overline{K}\frac{\partial[h'(\mathbf{x})]}{\partial x_1} - K'(\mathbf{x})\,\overline{J} + K'(\mathbf{x})\frac{\partial[h'(\mathbf{x})]}{\partial x_1} \right]$$

$$= \frac{1}{n}\left[\overline{K}\,\overline{J} - 0 + 0 - \mathrm{Cov}\left(K'(\mathbf{x}), \frac{\partial[h'(\mathbf{x})]}{\partial x_1}\right) \right]$$

$$= \frac{1}{n}\left[\overline{K}\,\overline{J} - \mathrm{Cov}\left(K'(\mathbf{x}), \frac{\partial[h'(\mathbf{x})]}{\partial x_1}\right) \right] \qquad [1.16]$$

where the expectation of the product of two zero-mean variables is equal to their covariance. Equation [1.16] demonstrates that, in the presence of small-scale unresolved variations in $K(\mathbf{x})$, mean velocity is not simply equal to the velocity

predicted from Darcy's law using the product of mean conductivity and mean gradient, $\overline{KJ}$. Rather, it also depends on how variations in conductivity about mean conductivity are correlated to local variations in the hydraulic head gradient. Like correlation between concentration and velocity variations, correlation between conductivity and hydraulic gradient perturbations will generally be a function of the unique geometry specific to the flow problem.

Equations [1.15] and [1.16] suggest that if we are interested in minimizing the macrodispersive flux terms in equation [1.7] by reducing $v_1'(\mathbf{x})$ and $v_2'(\mathbf{x})$ through better resolution of velocities $v_1(\mathbf{x})$ and $v_2(\mathbf{x})$, we need to be informed of spatial variations in conductivity at the smaller-scale, $K'(\mathbf{x})$, as well as the strength of correlation between $K'(\mathbf{x})$ and both $\partial h'(\mathbf{x})/\partial \mathbf{x_1}$ and $\partial h'(\mathbf{x})/\partial \mathbf{x_2}$. Thus, if we want the deterministic equation [1.3] to be a good predictor of the true mean concentration, we are required to resolve the spatial distribution of conductivity perturbations $K'(\mathbf{x})$, as well as the distribution of hydraulic gradient perturbations $\partial h'(\mathbf{x})/\partial \mathbf{x_1}$ and $\partial h'(\mathbf{x})/\partial \mathbf{x_2}$ at the smaller scale and understand how they are correlated to one another.

This differs dramatically from the case of steady-state fluid flow, where the two-dimensional flow equation for average flow in a heterogeneous, isotropic medium is:

$$\frac{\partial}{\partial x_1}\left[K(\mathbf{x})\frac{\partial h(\mathbf{x})}{\partial x_1}\right] + \frac{\partial}{\partial x_2}\left[K(\mathbf{x})\frac{\partial h(\mathbf{x})}{\partial x_2}\right] = 0 \qquad [1.17]$$

Equation [1.17] has the associated mean equation:

$$E\left[\frac{\partial}{\partial x_1}K(\mathbf{x})\frac{\partial h(\mathbf{x})}{\partial x_1}\right] + E\left[\frac{\partial}{\partial x_2}K(\mathbf{x})\frac{\partial h(\mathbf{x})}{\partial x_2}\right]$$

$$= E\left[\frac{\partial}{\partial x_1}\left[[\overline{K} + K']\frac{\partial[\overline{h} + h']}{\partial x_1}\right]\right] + E\left[\frac{\partial}{\partial x_2}\left[[\overline{K} + K']\frac{\partial[\overline{h} + h']}{\partial x_2}\right]\right]$$

$$= E\left[\frac{\partial}{\partial x_1}\left[\overline{K}\frac{\partial\overline{h}}{\partial x_1} + \overline{K}\frac{\partial h'}{\partial x_1} + K'\frac{\partial\overline{h}}{\partial x_1} + K'\frac{\partial h'}{\partial x_1}\right]\right] + E\left[\frac{\partial}{\partial x_2}\left[\overline{K}\frac{\partial\overline{h}}{\partial x_2} + \overline{K}\frac{\partial h'}{\partial x_2} + K'\frac{\partial\overline{h}}{\partial x_2} + K'\frac{\partial h'}{\partial x_2}\right]\right]$$

$$= \frac{\partial}{\partial x_1}\left[\overline{K}\frac{\partial\overline{h}}{\partial x_1} + 0 + 0 + \frac{\partial E(K'h')}{\partial x_1}\right] + \frac{\partial}{\partial x_2}\left[\overline{K}\frac{\partial\overline{h}}{\partial x_2} + 0 + 0 + \frac{\partial E(K'h')}{\partial x_2}\right]$$

$$= \frac{\partial}{\partial x_1}\left[\overline{K}\frac{\partial\overline{h}}{\partial x_1}\right] + \frac{\partial}{\partial x_2}\left[\overline{K}\frac{\partial\overline{h}}{\partial x_2}\right] + \frac{\partial^2 E(K'h')}{\partial x_1^2} + \frac{\partial^2 E(K'h')}{\partial x_2^2} = 0 \qquad [1.18]$$

1.3.3 The Scale Effect

As discussed in the previous section, central to use of deterministic transport equation [1.3] is that variations about mean velocity be resolved and incorporated into $\overline{v}_1$ and $\overline{v}_2$. When velocity variations occur at a scale too small to be fully resolved, these unresolved heterogeneities can be incorporated into the problem via use of macrodispersivities in equation [1.10]. As long as the plume is displaced a distance of many times the scale of velocity variation, local velocity variations v_1' and v_2' stop varying with plume movement, $E(v_1'c')$ and $E(v_2'c')$ become robust, scale-independent estimators, and both A_L and A_T are independent of the scale of the transport problem. At such a scale, the plume has effectively sampled all of the velocity variations and a

constant Fickian rate of spreading is approached. It should be noted that this limiting rate is generally not attained in practice along the transverse direction because sampling of independent variations in velocity perturbations occurs at a slower rate than in the longitudinal direction, typically resulting in quasi-Fickian transverse dispersion. Dagan (1990) has demonstrated that the transverse dimension of the solute source dictates the length of time required to approach true Fickian behavior.

Predictions made by Dagan (1984; 1988), Neuman and Zhang (1990), and Zhang and Neuman (1990) according to linear and quasi-linear theory suggest that, if mean velocity is uniform and flow is steady state, the rate of longitudinal transport will approach the Fickian limit after a mean displacement of only a few correlation scales even when log conductivity is not gaussian-distributed and the central limit theorem not valid except for very large mean displacement. However, this theoretically predicted constant Fickian rate of spreading is only rarely observed in the field.

Greenkorn and Kessler (1969) were among the first to note that field-scale dispersion tends to be much larger than dispersion observed under laboratory conditions. Gillham *et al.* (1984) noted that solute movement appears to be in a constant state of transition, never attaining asymptotic behavior but instead exhibiting intermittent Fickian behavior at successively large time/space scales. Rather than macrodispersivities approaching constant values with scale of displacement, longitudinal macrodispersivity may tend only intermittently to an asymptotic value, often appearing to never approach a constant value and increasing without bound with increasing scale of plume displacement. Using a particle tracker, Schwartz (1977)

demonstrated that a unique dispersivity value will not exist in a two-dimensional medium that is nonuniformly random or when conductivity is a function of space. Dagan (1991) accounted for dispersion caused by large-scale heterogeneity by considering the dispersion coefficient to be equal to the sum of an effective dispersion coefficient and dispersion caused by uncertainty in plume centroid location due to large-scale structure. Domenico and Robbins (1984) demonstrated that when an (n-1)-dimensional model is used to calibrate an n-dimensional system, dispersion coefficients must be scaled up in order to achieve a good fit to concentration observations, thus attributing the scale effect to a numerical artifact of the simulation procedure. The so-called scale effect has also been observed under field conditions by Fried (1975), Sudicky and Cherry (1979), and Sudicky *et al.* (1983).

What is the source of this never-ending scale dependency? Why do tracer tests performed at the field scale in aquifers with heterogeneities that appear uniformly random frequently show evidence of non-Fickian spread, rather than the constant Fickian rate predicted by equation [1.10]? Why do these tests suggest that the rate of tracer spread only rarely tends asymptotically to the Fickian limit? Cushman (1990) and Neuman (1990) attribute this failure to the presence of spatially discrete, nested structures characteristic of natural geologic media. Rather than possessing a single scale of variation characterized by one correlation scale, natural geologic media instead possess a *hierarchy* of scales of variation.

In more than 130 cases investigated by Neuman (1990), the rate of longitudinal plume spread in the subsurface appeared to increase without bound indefinitely as the

scale of the tracer test increased. He attributed the scale effect to evolving scales of conductivity heterogeneity that a plume continuously encounters as it migrates through the subsurface. In the presence of these evolving scales, variations in velocity also evolve in a scale-dependent manner. The scale-dependent variations in advection produced by scale-variant velocity render the rate of contaminant spread transient and the asymptotic Fickian limit unapproachable. As long as the greatest scale of heterogeneity in such a multiple-scale medium is large relative to the scale of plume displacement, velocity variations continue to enter the problem with increasing scale, and the differential advective spread of the plume increases without limit.

In a manner similar to that of Burrough (1983b), Neuman assumed that aquifers are comprised of discrete hierarchies of nested, mutually uncorrelated, uniformly random processes. In such a hierarchy, the plume intermittently approaches Fickian or quasi-Fickian behavior at each discrete scale of variation if it undergoes displacement by several correlation lengths at that scale before encountering the next larger scale. The variogram of such a hierarchical process can be constructed by superposing component variograms characteristic of each scale. When the discrete hierarchical medium is replaced by a continuous hierarchy in which there is no obvious decomposition of scales and it is assumed that this continuous hierarchy of scales is described by exponential component variograms for which variance increases with scale, the composite variogram reduces to a general power-law variogram of log conductivity characteristic of a random fractal process. This outcome demonstrates

that the observed scale effect is consistent with the assumption that log conductivity acts as a random fractal.

Neuman (1993) considered the resulting fractal scaling law to be universally applicable to a wide spectrum of geologic materials and over a multitude of scales, but emphasized that the rule does not necessarily reflect actual conditions at any given site. It instead represents transport behavior in a mean sense over a wide variety of geologic scales. He attributed large local deviations from the scaling rule to the prevalence of discrete scales of heterogeneity that universally prevail only over small ranges of scale intervals. In response to the work of Neuman (1990), Anderson (1991) suggested that due to the very different spatial structure of facies assemblages produced in different depositional systems, the proposed scaling rule is likely valid up to the scale of a sedimentary environment, with scaling behavior evident only at and below that scale.

In an effort to verify the scaling rule derived by Neuman (1990), Gelhar *et al.* (1992) reviewed a 59-point subset of the dispersivity data on which the scaling rule was based, with particular regard to data reliability. The dispersivity data values were obtained from tracer tests, contamination spills, and environmental events in a wide range of aquifer materials under a variety of flow conditions, and were estimated via calibration against numerical or analytic models, by breakthrough curve analysis, or by the method of moments. They noted that, while plots of estimated dispersivity versus scale of observation show that dispersivity does appear to increase with scale, there is a minimum of a two-order-of-magnitude range in dispersivity at any given scale. This very large range suggests that, if a systematic relationship between dispersivity and

scale does exist, it is obscured by error pervading the reported data. When Engesgaard *et al.* (1997) added a dispersivity value estimated for a sandy aquifer in Denmark at a plume displacement distance of 1000 m to the data set, no increase in dispersivity was found with increasing scale. While not precluding the possibility of scale-dependence of dispersivity, further study was suggested to improve reliability of dispersivity estimates, particularly at larger scales of observation using long-term, controlled tracer tests.

The implications of hierarchical geologic structure on transport prediction are significant. Macrodispersivities A_L and A_T are Fickian measures of differential advection in spatially-uniformly random geologic media. Their substitution for space-invariant values of α_L and α_T in equation [1.2] is not appropriate in hierarchical media, which are nonuniformly random in space and dominated by intermittent non-Fickian transport regimes. It is for this reason that both the deterministic advection-dispersion equation [1.3] and its associated mean concentration equation [1.10] have proven to be inadequate for describing the field-scale transport of solutes in hierarchical porous media, and have largely been abandoned in favor of simulation approaches that treat solute migration as the local advective process that it truly is, rather than as a global diffusive process. Principal among these methods is particle tracking, a technique that allows explicit recognition that local spatial variations in velocity and not global, Fickian diffusive processes are fundamentally responsible for driving the contaminant transport problem.

1.3.4 Multiple-Scale Characterization of Aquifer Heterogeneity

As discussed in the previous section, the observed rate of intermittent, field-scale plume spread in a wide variety of geologic settings is a strong indication that aquifers possess velocity variations at more than a single length scale. Journel and Huijbregts (1978), Burrough (1983a), Cushman (1990), Neuman (1990), and others have instead observed or inferred nested variability structures, for which the scale of variation increases discretely with the scale of observation. The prevalence of hierarchical structure can be traced to geologic processes of erosion, deposition, diagenesis, and fracturing responsible for imposing structure on natural geologic materials. Some of these processes act over many different spatial scales and others operate only at distinct scales or within ranges of scales (Burrough, 1983a). In the presence of geologic mechanisms that act over more than one spatial scale, geologic products exhibit scale-dependent structure that parallels the structure of the geologic process. When the effects of these processes are superimposed, structure at each scale will possess its own unique signature, characterized both by features that it shares with other scales and by those that are not shared.

Scale-dependent structure has, in fact, been directly observed at field and regional scales by researchers conducting aquifer characterization studies that seek to relate geologic processes to their products (Barton, 1994; Davis *et al.*, 1993; Davis, 1994). Recognition of the prevalence of scale-variant geologic structure has motivated much research into scaling of partial differential equations (Cushman, 1984; 1986;

1987; 1990) and treatment of aquifer heterogeneities as fractal structures (Wheatcraft and Tyler, 1988; Hewett and Behrens, 1988).

The importance of understanding multiple-scale aquifer heterogeneity to the problem of predicting field-scale contaminant transport has only been recognized during the past several decades. Within the past few years, significant efforts have been made to resolve geologic structure at both small and large scales via portable air-permeametry and lithofacies mapping. These studies generally focus on sedimentary aquifers, for which the erosional and depositional processes are fairly well-understood (Goggin, 1988; Davis, 1994; Barton, 1994). Since the intent of many such studies is to relate observed geologic structure to sedimentary processes, the resulting characterization tends to be somewhat unique to the specific depositional environment under investigation. To overcome the inherent site specificity of such causal studies, geostatistical models can be used to introduce a more descriptive, inferential component to the problem of quantifying multiple-scale aquifer heterogeneity.

1.4 GEOSTATISTICAL CHARACTERIZATION OF DISTRIBUTED AQUIFER PROPERTIES

Geostatistical methods rely on the assumption that general probabilistic laws can describe the spatial distribution of aquifer properties. While the actual distribution of an aquifer property such as hydraulic conductivity is fixed in space and thus deterministic, it is generally unknown to us except at a limited number of locations and is too costly to measure at every location in space. Compounding this information

uncertainty is intrinsic uncertainty associated with spatial variability of the property distribution. A highly heterogeneous aquifer, for example, will be characterized by a greater degree of intrinsic uncertainty than a spatially uniform medium. While information uncertainty is uncertainty associated with what we *know* about the distribution of the spatial property and can be reduced by acquiring more data, intrinsic uncertainty is a fundamental characteristic of the spatial property. The two sources of uncertainty are distinct, but under certain conditions within a probabilistic framework, intrinsic uncertainty can be used to assess information uncertainty, with the two types of uncertainty amounting to different philosophic ways of viewing aquifer property variability.

To compensate for our ignorance about the spatial distribution of conductivity, we can cast the characterization problem in this geostatistical framework, treating the conductivity distribution as a spatial stochastic process with the statistics of the random process estimated from the data set. This geostatistical framework allows us to infer information from the relatively limited amount of data typically available within the financial constraints of most aquifer characterization efforts, effectively replacing site-specific, causal, deterministic geologic laws with general, descriptive, probabilistic ones.

Such statistical parameterization almost always reduces the number of variables that need to be defined because it is based on inference. The inferential nature of the geostatistical model allows us to assume that, while the behavior of a random aquifer process at any given point in space cannot be exactly identified, it is sufficiently

uniformly random, or statistically homogeneous, to be able to do so *on average*. The data requirements for geostatistical characterization, then, will depend on the degree to which this is true. To illustrate how a geostatistical model can reduce the amount of data required to characterize aquifer heterogeneities, Farmer (1987) notes that in a spatially uncorrelated field a histogram constructed using on the order of 10 data values can supply as many as 10^5 samples of the uncorrelated process, effectively reducing the number of degrees of freedom by a factor of 10^4. Russo and Jury (1987) have reviewed data requirements for typical geostatistical analysis for correlated, uniformly homogeneous geologic processes. Although many more data points are required to characterize a correlated field than an uncorrelated one, it is evident that a geostatistical characterization model can significantly reduce data requirements compared to an equivalent deterministic model.

When data are supplemented by what is generally considered 'soft' or visually-observed geologic observations, data requirements can be even further reduced (Gelhar, 1986; Journel, 1986). Such soft data are generally available at the macroscopic scales, where structural information is accessible from visual geologic proxy observations. It should be emphasized that many practicing geologists take exception to the term 'soft' as it is applied by geostatisticians. They consider lithologic data obtained from direct observation of geologic material to be far less speculative than data obtained by assigning quantitative values to geologic material, because the latter requires numerous assumptions be made.

Chapter 1. Introduction: Random Geologic Structure

One of the earliest studies relating geologic products to geostatistical parameters was conducted by MacMillan and Gutjahr (1986), who attempted to relate dimensions of sandstone layers to porosity and permeability correlation scales. Operating on the assumption that conductivity is loosely related to macroscopically observable geologic features like color, texture, and grain size, Phillips and Wilson (1989) developed a semi-quantitative method of relating macroscopically observable, subjective geologic observations to correlation length using principles of threshold-crossing theory (vanMarcke, 1983). This theory allows average dimensions associated with regions of excursion above some threshold conductivity to define spatial statistics of the conductivity process. Brannan and Haselow (1993) have synthesized compound random fields using hierarchical indicator decision trees and level-crossing theory to incorporate soft geologic information. Other studies seeking to utilize geological observations for geostatistical characterization of heterogeneities include those of Goggin (1988), Barton (1994), and Davis (1994). While the scope of this study does not include development of a geologically-based geostatistical aquifer characterization model that incorporates well-understood geologic principles, it is hoped that this dissertation research will constitute a basis for further development of such a model. It is anticipated that merging of geostatistical and geologically-based approaches will significantly reduce data requirements relative to those of either strictly causal, deterministic, geologic process models or purely descriptive, geostatistical approaches.

1.4.1 Single-Scale Random Processes

In mathematical terms, a three-dimensional regionalized variable or random process is a function $Z(\mathbf{x})$ that assumes the value of a random variable at every vector location $\mathbf{x}$. This random process is characterized both by somewhat erratic local random behavior and a more global, systematic structure amenable to deterministic description; locally $Z(\mathbf{x})$ is a scalar random variable, but globally it has arbitrary orders of correlation described by a covariance tensor. The pointwise random property can be considered either to result from true unpredictability, or from incomplete information about the process.

In a manner analogous to a random variable being the outcome of a single probabilistic event, a realization of a random process is a single outcome of the collection of M random variables comprising the random process, $Z_0(\mathbf{x}_i)$, i=1, 2, 3, . . M. Similar to the probability density function (pdf) of a random variable, $Z(\mathbf{x})$ is characterized by joint pdfs of varying orders that describe probabilistic behavior among all M variables. To completely characterize a random process, the joint pdf between and among all M points in space is required. The joint pdfs describe how subsets of the M variables separated by certain distances are related to one another, and thus define the structural component of $Z(\mathbf{x})$. In general, it is not possible to infer this joint probability law from the single observed realization, because the realization represents only a single sample from the joint pdf.

$Z(\mathbf{x})$ is strictly stationary if the joint probability density function of $Z(\mathbf{x}_1)$, $Z(\mathbf{x}_2$ . . . , $Z(\mathbf{x}_M)$ is the same as the joint pdf for $Z(\mathbf{x}_1+\zeta)$, $Z(\mathbf{x}_2+\zeta)$, . . . , $Z(\mathbf{x}_M+\zeta)$, where

is an arbitrary translation vector. In other words, the joint pdf of a strictly stationary process is translationally space-invariant — the process is uniformly random up to arbitrary orders of joint behavior. Strict stationarity does not necessarily imply that the moments of the joint pdf exist (Myers, 1989). Since high-order joint moments are difficult to estimate and may not even exist, most practical geostatistics is based on a weaker form of stationarity known as second-order stationarity.

1.4.1.1 Second-Order Stationarity

Under the assumption of weak or second-order stationarity, the true joint probability laws governing the random process $Z(\mathbf{x})$ can be simplified to account for only the first- and second-order moments. $Z(\mathbf{x})$ is second-order stationary if it has constant mean and if the covariance $\text{Cov}[Z(\mathbf{x}+\zeta), Z(\mathbf{x})]$ exists and is translationally invariant. Weak stationarity implies that the first two moments of the distribution are sufficient to characterize the joint behavior of the process. This assumption is equivalent to assuming that the bivariate gaussian distribution is enough to define joint behavior. Although the gaussian distribution can possess nonzero high-order even joint moments, they are functions of the second-order moment. In practice, the assumption of gaussian behavior is almost always made because of the tremendous amount of data required to estimate third- and higher-order odd joint moments. For example, while second-order moments require knowledge of how pairs of $Z(\mathbf{x})$ values separated by a certain distance behave jointly probabilistically, estimation of the

third-order moment requires information about triplets of $Z(\mathbf{x})$ values associated with a large number of possible spatial configurations of data.

Freeze (1975) and Hoeksema and Kitanidis (1984) have demonstrated that gaussian probabilistic behavior describes most spatially-distributed hydrogeologic variables, usually after a logarithmic transformation. In the latter study, covariance matrix decomposition was used to decorrelate variables of random processes corresponding to aquifer properties from 31 regional studies, and univariate tests were applied to the decorrelated values in order to check for univariate gaussian behavior. Univariate mean, variance, skewness, and kurtosis were estimated and Kolmogorov-Smirnoff and χ^2 tests of normality performed on the uncorrelated residuals. In addition, maximum-likelihood techniques were used to estimate nugget variance, structural variance, and correlation scale for a number of transmissivity, conductivity, and storage coefficient data sets. Finally, the Akaike Information Criterion (AIC) was used to discriminate between untransformed and log-transformed models. Results of the statistical analyses suggested that the lognormal model describes the transmissivity, conductivity, and storativity data better than the normal model for 51 out of 52 successful analyses. Use of logarithmic values has the added benefit of reducing parameter variability. However, lognormality of storativity is in direct contradiction to the conclusion of Freeze (1975) that porosity appears normally-distributed. This might be attributed to the weak dependence of storativity on porosity, compared to the coefficient of compressibility.

The assumption that aquifer properties are described by univariate or multivariate gaussian probability distributions has been called into question by Journel and Alabert (1988; 1989) and Journel and Deutsch (1993). They argue that the tails of a gaussian distribution cannot account for extreme values observed in real geologic media. They also point out that the multivariate gaussian model maximizes spatial entropy and produces smoothly-varying structure that bears little resemblance to abrupt lithologic discontinuities typically encountered in the field.

1.4.1.2 The Ergodic Hypothesis

In practice, the true process $Z(\mathbf{x})$ and its underlying joint pdf will never be known to us because such knowledge requires that many samples, or realizations, of the process be drawn from a population or ensemble of the random process. We have access to only one realization of the geologic process and must rely on the ergodic hypothesis to estimate joint moments. The ergodic hypothesis, which can never be proved or disproved using anything less than a complete ensemble of realizations, states that the spatial variability observed in a realization can be used to obtain estimates of uncertainty at a point. In an ergodic or 'typical' process, the spatial average of the variance is equivalent to the ensemble variance and implicit uncertainty is a reflection of information uncertainty. When the process is ergodic, all possible states of the ensemble are represented in the realization (Dagan, 1986; 1990). The hypothesis allows us to make statements regarding information uncertainty using accessible spatial statistics. Note that stationarity is a sufficient condition for ergodicity. That is,

if a process is ergodic, it must also be stationary. However, the converse is not necessarily true.

While the ensemble interpretation permits us to evaluate uncertainty associated with prediction at a point, such an approach differs fundamentally from the spatial interpretation. From the perspective of spatial interpretation, the prediction represents a trend about which the property fluctuates. This is especially true in the context of solute transport, because the plume might sample only a small part of the conductivity realization relative to the scale of heterogeneity rather than the entire realization. As a consequence, the ergodicity hypothesis is less likely to be valid than for the case of fluid flow, where the flow domain is everywhere sampled by the fluid. Note that, since it is usually prediction error rather than spatial trends that interest us, the ensemble interpretation is generally favored.

1.4.1.3 Intrinsic Random Processes

Myers (1989) argues that, because geologic data only represent an incomplete, nonrandom sample from a single realization of the random geologic process, the presence of strict or weak stationarity can never be verified. He suggests that, because stationarity is a property of the process, and *not* the data, one can never actually test for stationarity using a sample from a single realization. Hence, as weak a form of it as possible should be invoked. A form of stationarity even weaker than second-order stationarity is the intrinsic hypothesis, which requires only second-order stationarity of the increments $[Z(\mathbf{x}+\xi)-Z(\mathbf{x})]$ (Delhomme, 1978).

$Z(\mathbf{x})$ can always be written as the sum of a slowly-varying deterministic mean $\overline{Z}(\mathbf{x})$ and a more erratic zero-mean, first-order perturbation $Z'(\mathbf{x})$:

$$Z(\mathbf{x}) = \overline{Z}(\mathbf{x}) + Z'(\mathbf{x}) \qquad [1.19]$$

where $Z'(\mathbf{x})$ is characterized by a specified covariance structure that probabilistically defines the average behavior of the fluctuations in space. For an intrinsic random function (IRF) or process of order zero, the mean of the increments must equal zero:

$$E[Z(\mathbf{x} + \xi) - Z(\mathbf{x})] = 0 \qquad [1.20]$$

which implies that $Z(\mathbf{x})$ has a spatially-constant mean. The zeroth-order increment $[Z(\mathbf{x}+\xi) - Z(x)]$ automatically filters out that spatially-constant mean. In certain cases, processes involving higher-order differences may be used to filter out a mean that behaves as a high-order polynomial over space. However, it is usually preferable to remove the mean and operate on the residual as if it were a zero-order intrinsic random process.

Random and systematic spatial characteristics of the IRF are embodied in the semi-variogram $\gamma(\xi)$, which is half the variance of the spatial increment of $Z(\mathbf{x})$ with respect to separation distance ξ. For mean-zero increments, the semi-variogram reduces to:

$$\gamma(\xi) = \frac{1}{2}\,\mathrm{var}[Z(\mathbf{x}+\xi) - Z(\mathbf{x})] = \frac{1}{2}\,E\{[Z(\mathbf{x}+\xi) - Z(\mathbf{x})]^2\} \qquad [1.21]$$

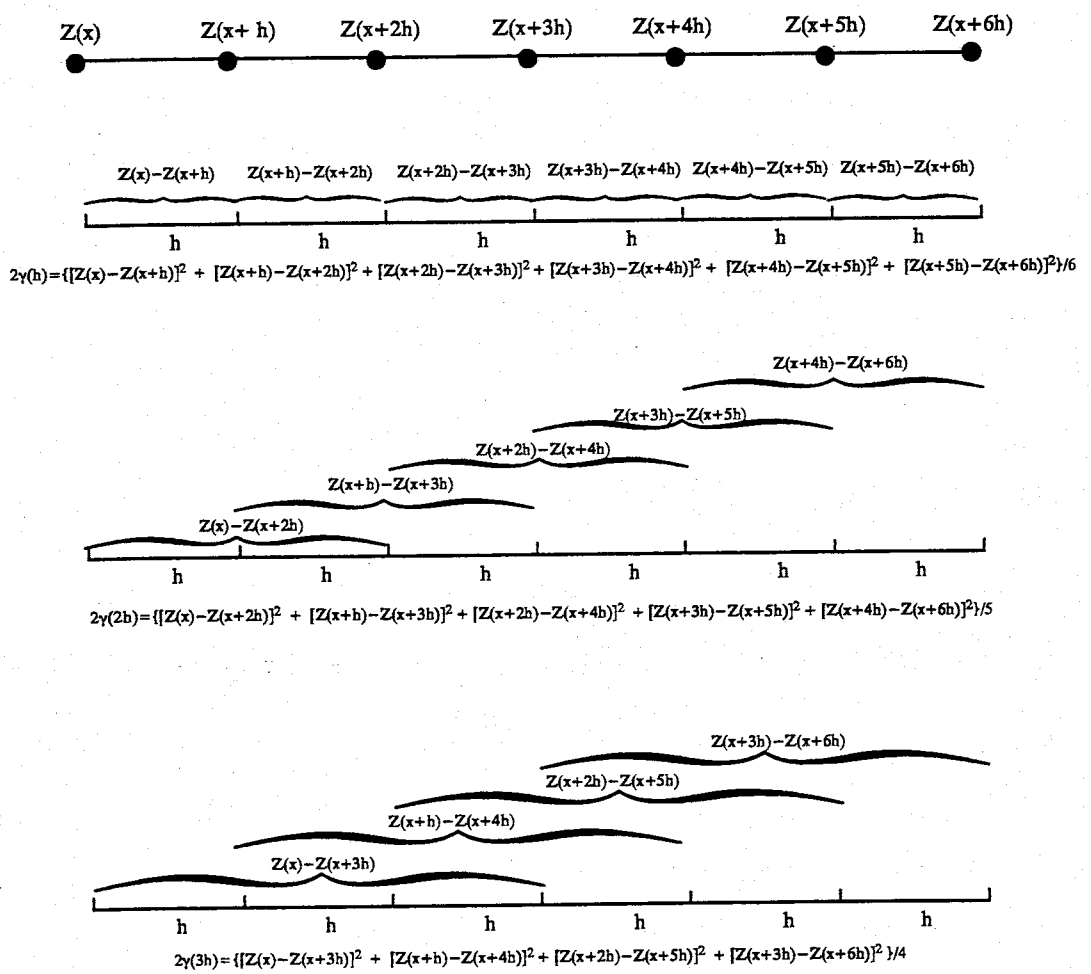
$\gamma(\xi)$ is a measure of relative variance with respect to separation distance, and can be multiplied by a scalar without altering its meaning. Hence, the variogram $2\gamma(\xi)$ is generally used to quantify the variance of an intrinsic random process. The variogram is constructed by estimating the variance of the increment associated with all pairs of data separated by distance vector ξ. Figure 1.1a illustrates how pairs of values separated by lag distances that are multiples of scalar separation distance h are used to estimate the experimental variogram in one-dimensional space. Nugget variance occurring at a separation distance of zero corresponds to unstructured variability, measurement error related to operational precision of the sampling device, or any variability that cannot be attributed to spatial separation of data points.

The variogram can be estimated from data using the nonparametric estimator (Russo and Jury, 1987):

$$2\hat{\gamma}(\xi) = \frac{1}{n(\xi)} \sum_{i=1}^{n(\xi)} \left[Z(\mathbf{x}_i) - Z(\mathbf{x}_i + \xi)\right]^2 \qquad [1.22]$$

where $n(\xi)$ is the number of pairs separated by lag ξ, or by using parametric maximum likelihood (ML) or least-squares techniques to fit theoretical variogram functions to the data (Kitanidis, 1985; Kitanidis and Lane, 1985). If the variogram increases to a

Chapter 1. Introduction: Random Geologic Structure



(a) Increments of Lag 1h, 2h, and 3h

(b) Second-Order Stationary Behavior

(c) Second-Order Intrinsic, Nonstationary Behavior

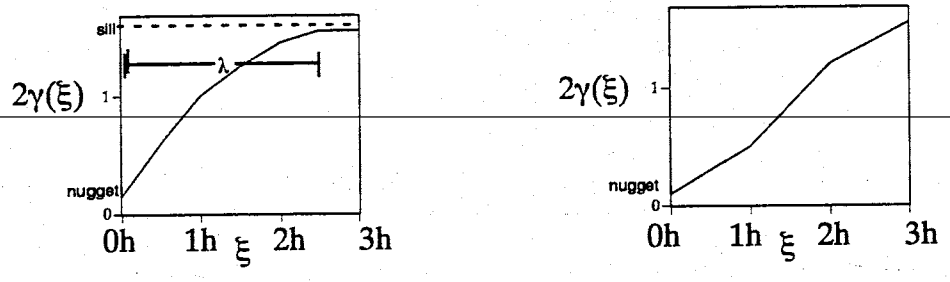


Figure 1.1. Construction of a Variogram.

constant variance of the increments, referred to as the sill of the variogram, then the random process conforms to the stricter assumption of second-order stationarity. In this case, the lag at which the sill is reached corresponds to the correlation or integral scale λ, equal to the average distance over which the process is correlated (see Figure 1.1b). Statistical methods of data interpolation, such as kriging and conditional gaussian simulation, tend to be very robust to the functional form of the variogram model and the estimated correlation scale (Kitanidis and Vomvoris, 1983).

In multiple-scale geologic media, the intrinsic conductivity process will not be second-order stationary except at scales much larger than the largest scale of variation. When this is the case and there are many discrete scales of spatial variation, the variance increases stepwise and apparently without bound as shown in Figure 1.1c, with progressively larger scales of observation recording changes in mean behavior that contribute ever larger amounts of variation. Random fractal structure, which follows a power-law variogram, is an ideal example of such structure when an infinite number of continuously evolving scales are present. Hydrologists have only recently begun to view most aquifer correlation structure, and thus many flow and transport problems, as the multiple-scaled, hierarchical phenomena that they truly are.

1.4.2 Multiple-Scale Random Processes

As mentioned above, fractals are ideally suited for reproducing continuously evolving scales of spatial variation. Fractals represent self-similar structures characterized by unbounded variance and infinite correlation scale (Wheatcraft and

Tyler, 1988; Hewett and Behrens, 1988). Due to the self-similarity of fractals, information moves in a straightforward manner, scaling according to some characteristic power of the scale of observation. While fractals provide a useful basis for characterizing scale-variant geologic structure, they describe an ideal type of structure rarely encountered in real field situations. Instead, Journel and Huijbregts (1978), Burrough (1983a), Cushman (1990), Neuman (1990) and others have observed or inferred intermittent scaling of information over discrete ranges or hierarchies of scales.

Recognizing that deterministic patterns are rarely observed in nature, Burrough (1983a) adopted a random fractal model known as fractional Brownian motion (fBm) process (Mandelbrot, 1975; 1977). He also recognized that abrupt changes in the mean at geologic boundaries and local second-order stationarity within certain scale intervals give rise to variograms that increase in a step-wise manner as larger scales of variation are sampled. Burrough has referred to this latter phenomenon as partial self-similarity, reflecting the fact that most geologic structure is not formed by a single process acting simultaneously at all scales and a single fractal dimension, but rather by many scale-dependent interacting geologic processes and with a fractal dimension that itself is a function of scale. Moreover, in a partially self-similar medium the relative dominance of a structure-forming process does not necessarily increase with scale, as a self-similar fractal implies. The assumptions of a single process acting over all scales characterized by increasing variability with scale, both which are central to the fBm model, must be reconsidered.

Chapter 1. Introduction: Random Geologic Structure

Burrough proposes that a non-Brownian model be developed to characterize structure caused by independent processes operating over distinct spatial scales. He argues that this model should accommodate local stationarity between geologic boundaries and account for discrete scaling behavior at geologic boundaries. To resolve the primary sources of geologic variation acting over discrete ranges of scales with differing degrees of dominance, the composite random process due to all structure-forming processes is described as a weighted sum of processes over all S scales:

$$Z(\mathbf{x}) = \beta_0 Z_0 + \beta_1 Z_1 + \ldots \beta_S Z_S \qquad [1.23]$$

where Z_s, s= 1, 2, . . . , S is the random geologic process acting over a range of scales r_s, and β_s is the relative weight of the sth geologic process. Equation [1.23] is simply a superposition of all random effects acting at all discrete scales, with variability of $Z(\mathbf{x})$ increasing with scale of observation s. Note that for additive processes, use of the superposition principle requires the assumption that independent processes prevail at each scale. That is, the process measured at each distinct scale is assumed to be independent of processes observed at all other scales.

One can minimize the square of the difference between the model variogram and the experimental variogram to obtain the relative weights β_s, s = 1, 2, . . . , S that best describe the multiple-scaled spatial variation and decompose the variogram into its component scale-dependent variograms. A larger magnitude associated with the sth weight might suggest that efforts be focused on measuring at the sth scale, because the

most information return would generally be associated with the scale that dominates the process.

Perhaps a more promising approach to characterizing geologic heterogeneity is offered by Painter (1996), who proposes that fractional Levy motion be used to describe permeability and porosity variations in sedimentary formations. The slowly-decaying tails associated with the Levy-stable family of distributions can reproduce spatially abrupt changes in geologic properties and scaling behavior associated with hierarchical media.

1.4.3 Randomness as a Function of Observation Scale

Burrough (1983a) made a distinction between systematic, or explained, variation that can be understood in the context of known physical variables such as parent material, topography, and climatic conditions and random, or unexplained, variation to which no cause can be assigned. He noted that unresolved behavior which might be expressed as infinite variance 'white noise' or uninterpretable signal at one scale may actually show finite-variance structure or interpretable signal when observed at another scale. Cushman (1987) has also noted that refining the scale of observation reveals progressively more detail in geologic structure.

To demonstrate this concept, a Gedanken experiment, in which a hypothetical zoom lens is used to capture a geologic image at progressively finer scales, is illustrated schematically in Figure 1.2. At the meter-scale of observation, centimeter-scale structure remains unresolved, or unmeasurable, because small-scale variability is

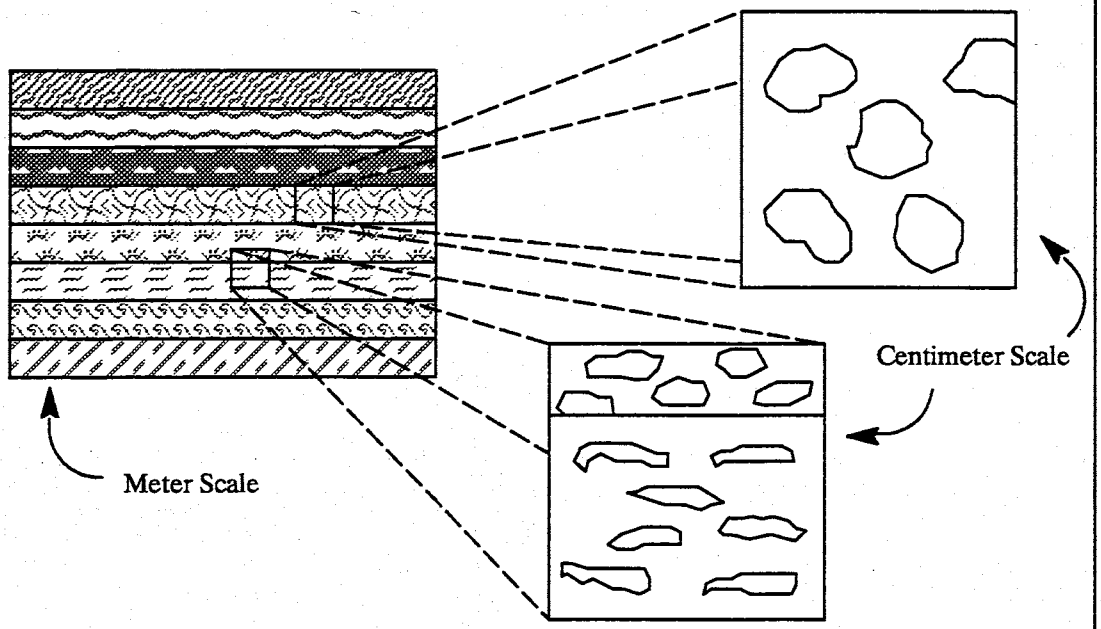


Figure 1.2. A Gedanken Experiment.

averaged by the meter-scale instrument — small-scale structure simply doesn't register on the larger-scale instrument. In this case, physical constraints on the resolution of the meter-scale measuring device are the source of the uncertainty. As the scale of observation is progressively refined, what appears as implicit unexplained behavior, or noise, at the coarser scales explicitly reveals itself as structure at finer scales until, at the finest scale, all variation is structure and fully explained but not completely measured over the entire process.

Conversely, systematic explainable structure such as layering present at the larger scale of meters may seem random and without explanation when observed at the smaller centimeter scale. In this instance, the centimeter-scale instrument won't see all systematic variation present at the larger scale unless measurements are made at every location in the image. Implied, unmeasurable heterogeneity is present because there are not a sufficient number of small-scale measurements to fully explain the large-scale behavior — the source of randomness, or lack of knowledge, is parameter uncertainty rather than poor instrument resolution.

In designing an instrument of fixed resolution such as a camera with an 'instrument window' containing a fixed number of pixels, there will always be a trade-off between randomness caused by small-scale parameter uncertainty and randomness caused by large-scale measurement resolution, as illustrated in Figure 1.3. The instrument resolution is generally obtained at the expense of increased parameter uncertainty but coarse resolution tends to produce bias in predictions of state variables (Vasco *et al.*, 1997). In both cases, uncertainty can be quantified in a geostatistical framework. Small-scale unseen variability within each of the large-scale instrument windows can be treated as a random process (Figure 1.3a), while the inferential nature of the geostatistical model can be used to 'fill in' the gray area associated with parameters that remain unmeasured by the small-scale instrument (Figure 1.3b).

The Gedanken experiment demonstrates that different scales of variation are measured depending on the scale of the measuring device. From these considerations, we can deduce that randomness itself can be a function of observation scale.

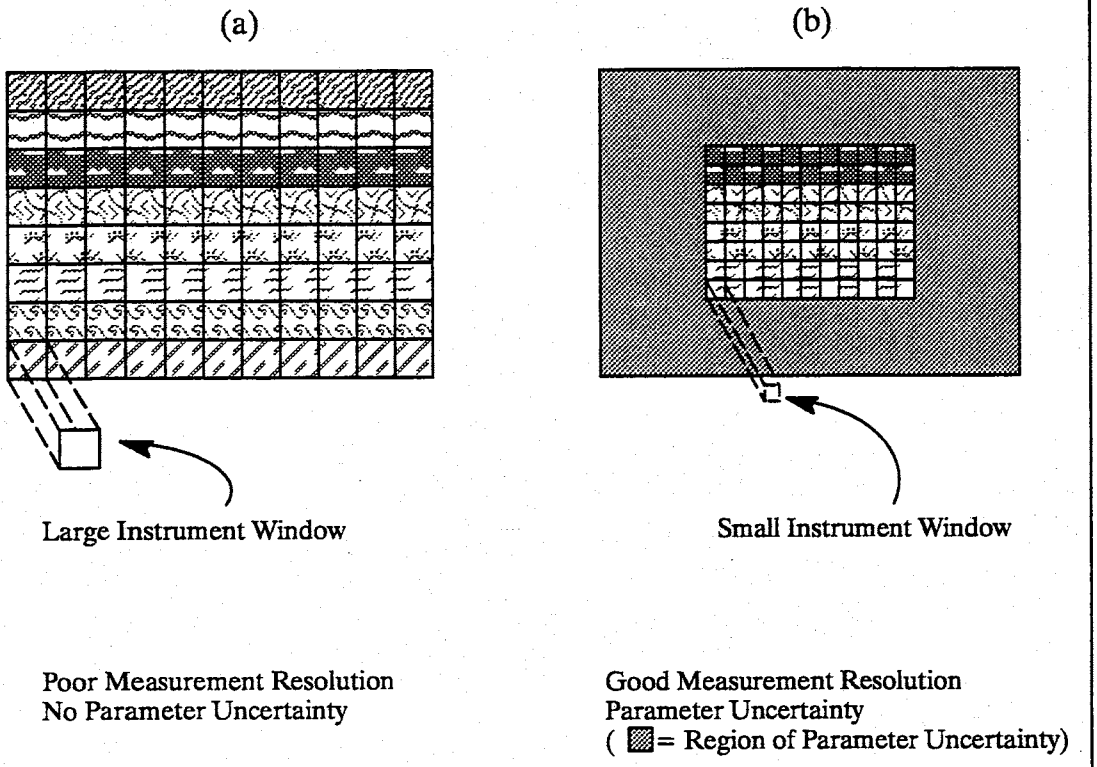


Figure 1.3. Measurement Resolution and Parameter Uncertainty as Sources of Randomness.

A common example of how different measurements might be obtained through different measurement windows is the monitoring of piezometric head. A small-diameter piezometer screened over a given interval can yield a different head measurement than a large-diameter well screened over the same interval, due to borehole storage effects. To illustrate why one might be interested in measuring

heterogeneities at different scales depending on the scale of a dynamic problem, three different hypothetical contamination events are presented in Figure 1.4. The first plume might have been generated by runoff of nitrates from an areally-extensive agricultural plot. In this situation, the plume has sampled so many small-scale heterogeneities that the effects of these heterogeneities have averaged out, and are not registered by the plume as distinct structure. In effect, the plume has already become informed of these small-scale heterogeneities, simply on the basis of its past history. Since the plume has not sampled many of the intermediate-scale heterogeneities and hasn't yet had the opportunity to sample even one of the large-scale heterogeneities, we can most effectively inform the model by measuring heterogeneities at those scales.

The second hypothetical case corresponds to a leachate plume from a landfill, which covers a smaller area than the non-point agricultural source. In this situation, the plume has obtained a representative sample of the small-scale heterogeneities, but is influenced by unmeasured, intermediate-scale structure. It is not at all affected by large-scale structure, which it hasn't yet sampled or 'seen'. We would be most interested in evaluating intermediate-scale heterogeneities, because most of the unresolved, unexplainable behavior affecting the plume is occurring at that scale. Large-scale structural features may not be relevant if we intend to remediate the aquifer before the plume begins to be impacted by them.

Finally, in the third case we are dealing with a point source. In this situation, the plume hasn't begun to fully sample small-scale variations in conductivity, and has even fewer samples of the intermediate structure. It is also not likely to span the larger

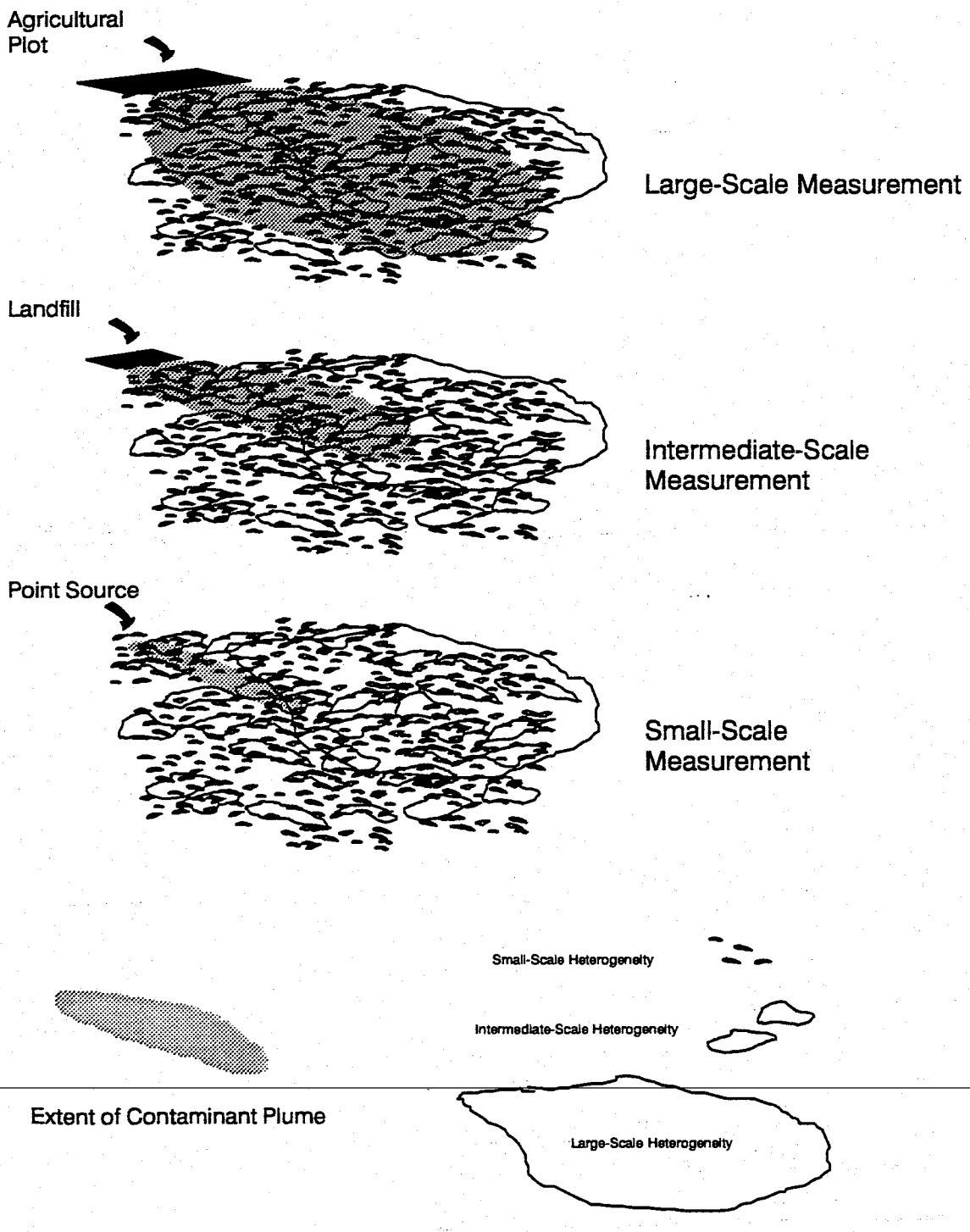


Figure 1.4. Measurement of Heterogeneities at Different Scales.

scale of variation in any reasonable amount of time. The most amount of information about future migration of the plume would be gained by concentrating data acquisition efforts on measuring aquifer properties at small and intermediate scales, because it is at those scales where unresolved plume behavior, or randomness, is most pronounced.

1.4.4 Multiple-Scale Variogram Behavior

The variograms presented in Figure 1.1 represent processes dictated by a single scale of variation (1.1b) and unbounded variance caused by discretely evolving structure (1.1c). In practice, most geologic structure is characterized by a hierarchy of scales of variation, manifested in the variogram as stepwise increases to multiple, transient sills. Figure 1.5 shows an example of a hypothetical two-scale logarithmic process. The correlation scales associated with small- and large-scale structures in the principal, longitudinal direction are roughly 20 and 90 units, respectively. Although the mean may differ between large-scale features, it is assumed for the sake of simplicity that small-scale variance is the same within each large-scale feature. It is further assumed that the small-scale instrument averages over a 5-unit length, while the large-scale instrument averages over a 30-unit length. We can view the variogram as a superposition of two distinct variograms observed through small- and large-scale instrument windows.

Small-scale variogram construction in the principal longitudinal direction of anisotropy proceeds beginning with a small separation distance of 5 units, and shows an increase in the variance as pairs are separated by a larger distance and become less

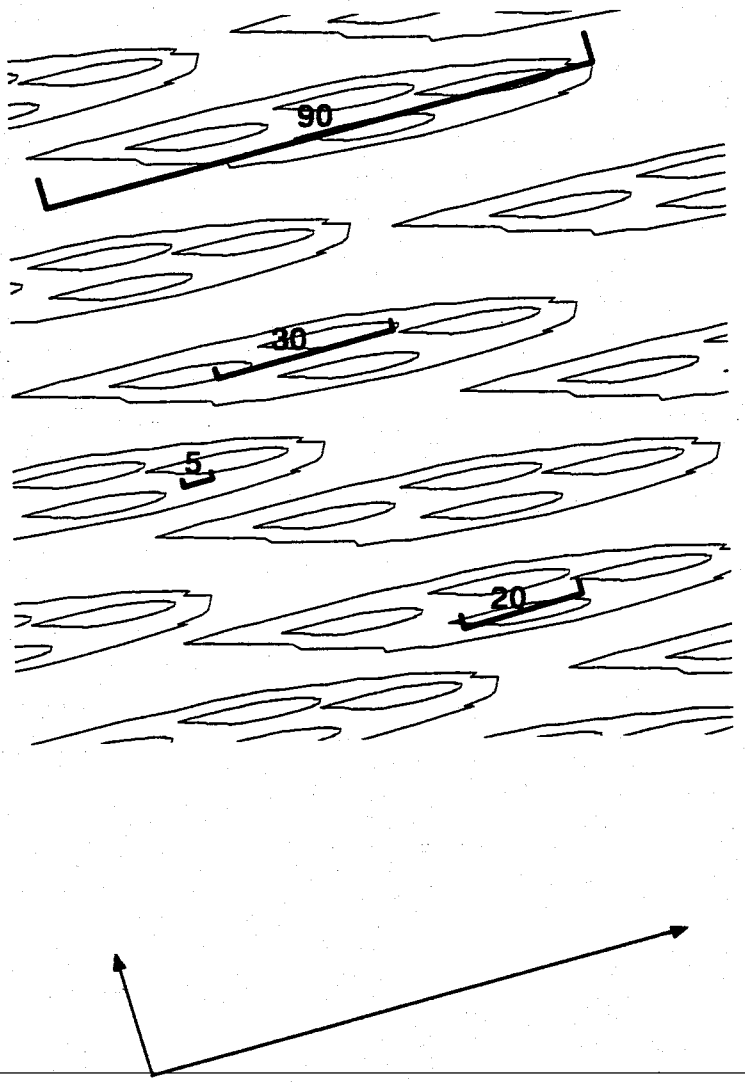
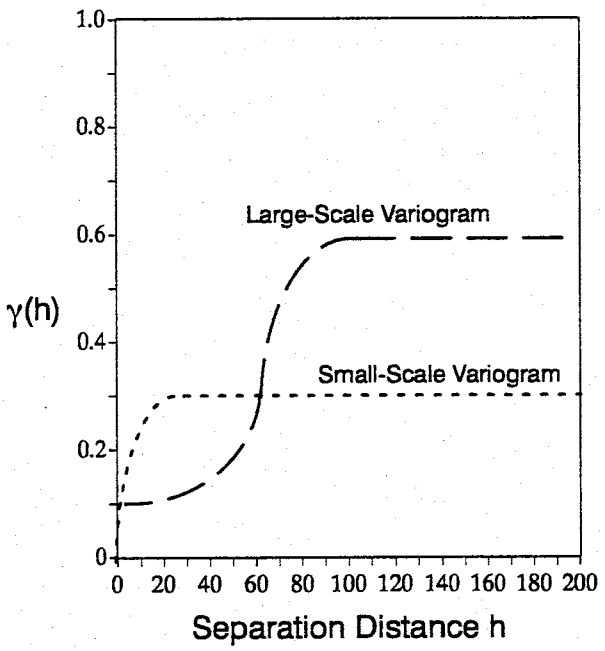


Figure 1.5. A Two-Scale Process.

correlated, as illustrated in Figure 1.6a. The small-scale instrument does not 'see' changes in the large-scale mean, so those changes must be removed from the process when constructing the small-scale variogram. They become known to us only through large-scale measurement. Because large-scale changes in the mean have been removed and the small-scale variance in the large-scale elements is the same, the small-scale process is stationary and the variogram approaches a sill at a 20-unit lag. At this separation distance all small-scale variance within each large-scale feature has been completely accounted for. At lags larger than 20 units, the small-scale measuring device samples from different large-scale structures, but because the large-scale mean has been removed from the two-scale process for purposes of constructing the small-scale variogram, the small-scale variance in each structure is the same, and the small-scale variogram continues to define a constant sill of 0.3 beyond a 20-unit lag.

Since the large-scale instrument can perceive changes in the large-scale mean, the mean is not removed prior to constructing the large-scale variogram. The large-scale variogram corresponding to the larger instrument window does not record structural variation at small lags because the small-scale correlation scale is on the order of 20 units. Thus, the averaging that occurs within the 30-unit measuring window obliterates all small-scale structure. Non-structural variance caused by the fact that the averaging volume is much larger than the size of small-scale heterogeneities would produce a large-scale nugget, perhaps on the order of 0.1. As lags increase from 30 to 90 units, increasing numbers of large-scale measurement pairs span more than one large-scale feature, introducing greater amounts of variance attributed to changes in

(a) Large- and Small-Scale Component Variograms



(b) Composite Two- Scale Variogram

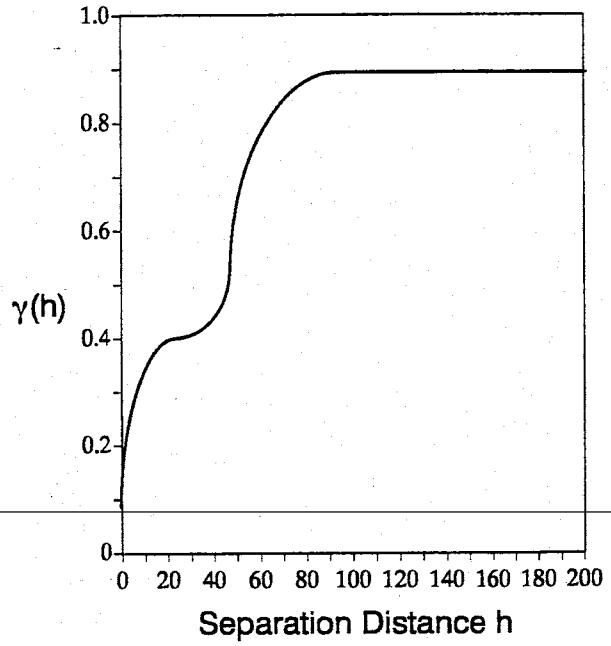


Figure 1.6. Superimposed Variogram for Two-Scale Process in the Principal Longitudinal Direction of Anisotropy.

the large-scale mean. When the separation distance increases above the 90-unit average length of the large-scale features, all of the pairs span all large-scale features. At this separation distance, the variance of the increments reflects all variance caused by changes in the mean between large-scale features. It is at this point that the large-scale variogram has reached its sill — one that corresponds to the variance contributed by changes in the mean among large-scale features. Small-scale variance is not included except as a nugget, because the large-scale device can't explicitly measure small-scale structure. And because changes in the mean typically dwarf small-scale variations, the large-scale variogram sill of 0.6 is larger than the smaller scale sill.

Note that, in order to highlight the concept of variable instrument windows, the supports for the two variograms are assumed different. Construction of the component variograms using the same support size would yield qualitatively similar results, but with different large-scale and small-scale variances and correlation scales than those in Figure 1.6. Small-scale structure would still be obliterated at the large scale because the small-scale process would not be correlated at large lags.

When superimposed, the small- and large-scale variograms completely describe the spatial properties of the two-scale random process, as shown in Figure 1.6b. The first inflection in the two-scale variogram corresponds to a single step up the hierarchy of the two-scale process from the small to the large scale. Stationary small-scale behavior is interrupted by an abrupt change in the mean behavior as large-scale variation begins to enter the variogram. Associated with large-scale variation is

another sill reflecting both large- and small-scale variance. Three-, four-, and nth-scaled processes would exhibit similar stepwise behavior, with multiple-scaled processes exhibiting unbounded variance and correlation scale. In the limit, as the number of scales of variation encountered increases continuously to ∞, the stepwise variogram approaches a power variogram characteristic of fractal structure (Neuman, 1990). In such a medium, there is no single characteristic scale of variation.

1.5 SUMMARY

Characterization of aquifer heterogeneity in the presence of multiple-scale geologic structure is a fundamental problem facing hydrogeologists. When unresolved aquifer heterogeneities occur, prediction of mean contaminant concentration using the deterministic advection-dispersion equation is valid only when the plume is displaced over many conductivity correlation scales and mechanical mixing approaches Fickian, or diffusive, behavior. However, in multiple-scale porous media no single representative scale of variation exists and a non-Fickian transport regime prevails. Under such conditions, unresolved conductivity variations must be resolved and incorporated into the advective term of the deterministic advection-dispersion equation.

Intermittent scaling of geologic structure has been observed or inferred by a number of researchers (Burrough, 1983a,b; Neuman, 1990). Such scaling results from the occurrence of discrete hierarchical, nested scales of variation. If hierarchical scales of conductivity variation are to be resolved and incorporated into the advective term

of the ADE, they must be measured at both small and large scales using instruments specifically designed to measure at those scales of averaging.

To reduce data required for resolving aquifer structure at all scales of variation, the inferential power of a geostatistical model can be exploited. Within this geostatistical framework, the nature of hierarchical spatial structure sampled by a plume spanning more than one scale of conductivity variation can be described using traditional geostatistical estimates such as the variogram. Geostatistical inference also lends a natural framework for dealing with poor instrument resolution at larger measurement scales, as well as with parameter uncertainty associated with a limited number of measurements observed through small 'instrument windows'.

1.6 REFERENCES

Anderson, M.P., Comment on 'Universal scaling of hydraulic conductivities and dispersivities in geologic media' by S.P. Neuman, *Water Resour. Res.*, 27(6), 1381-1382, 1991.

Bakr, A.A., L.W. Gelhar, A.L. Gutjahr, and J.R. Macmillan, Stochastic analysis of spatial variability in subsurface flows, 1. Comparison of one- and three-dimensional flows, *Water Resour. Res.*, 14(2), 263-271, 1978.

Barton, M.D., *Outcrop characterization of architecture and permeability structure in fluvial-deltaic sandstones, Cretaceous Ferron Sandstone, Utah,* Ph.D dissertation, 259 pp., The Univ. of Texas, Austin, 1994.

Bear, J., *Dynamics of Fluids in Porous Media,* Elsevier, New York, 764 pp., 1972.

Bear, J., *Hydraulics of Groundwater*, McGraw-Hill, New York, 1979.

Brannan, J.R., and J.S. Haselow, Compound random field models of multiple-scale hydraulic conductivity, *Water Resour. Res.,* 29(2), 365-372, 1993.

Burrough, P.A., Multiscale sources of spatial variation in soil, I. The application of fractal concepts to nested levels of soil variation, *J. Soil Science*, 34, 577-597, 1983a.

Burrough, P.A., Multiscale sources of spatial variation in soil, II. A non-Brownian fractal model and its application in soil survey, *J. Soil Science*, 34, 599-620, 1983b.

Carrera, J., and S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 1. Maximum likelihood method incorporating prior information, *Water Resour. Res.*, 22(2), 199-210, 1986a.

Carrera, J., and S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 2. Uniqueness, stability, and solution algorithms, *Water Resour. Res.*, 22(2), 211-227, 1986b.

Carrera, J., and S.P. Neuman, Estimation of aquifer parameters under transient and steady state conditions, 3. Application to synthetic and field data, *Water Resour. Res.*, 22(2), 228-242, 1986c.

Clifton, P.M., and S.P. Neuman, Effects of kriging and inverse modeling on conditional simulation of the Avra Valley aquifer in southern Arizona, *Water Resour. Res.*, 18(4), 1215-1234, 1982.

Cooley, R.L., A method for estimating parameters and assessing reliability for models of steady state groundwater flow, 1. Theory and numerical properties, *Water Resour. Res.*, 13(2), 318-324, 1977.

Cooley, R.L., A method for estimating parameters and assessing reliability for models of steady state groundwater flow, 2. Applications and statistical analysis, *Water Resour. Res.*, 15(2), 603-617, 1979.

Cushman, J.H., Comment on three-dimensional stochastic analysis of macrodispersion in aquifers by L.W. Gelhar and C.L. Axness, *Water Resources Research*, 19(6), 1641-1642, 1983.

Cushman, J.H., On unifying the concepts of scale, instrumentation, and stochastics in the development of multiphase transport theory, *Water Resour. Res.*, 20(11), 1668-1676, 1984.

Cushman, J.H., On measurement, scale, and scaling, *Water Resour. Res.*, 22(2), 129-134, 1986.

Cushman, J.H., More on stochastic models, *Water Resour. Res.*, 23(4), 750-752, 1987.

Cushman, J.H. (ed.), An introduction to hierarchical porous media, in *Fluids in Hierarchical Porous Media*, Academic Press, San Diego, Calif., 1-6, 1990.

Dagan, G., Solute transport in heterogeneous formations, *J. Fluid Mech.*, 145, 151-177, 1984.

Dagan, G., Statistical theory of groundwater flow and transport: Pore to laboratory, laboratory to formation, and formation to regional scale, *Water Resour. Res.*, 22(9), 120S-134S, 1986.

Dagan, G., Time dependent macrodispersion for solute transport in anisotropic heterogeneous aquifers, *Water Resour. Res.*, 24(9), 1491-1500, 1988.

Dagan, G., Transport in heterogeneous porous formations: Spatial moments, ergodicity, and effective dispersion, *Water Resour. Res.*, 26(6), 1281-1290, 1990.

Dagan, G., Dispersion of passive solute in non-ergodic transport by steady velocity fields in heterogeneous formations, *J. Fluid Mech.*, 223, 197-210, 1991.

Dagan, G., The significance of heterogeneity of evolving scales and of anomalous diffusion to transport in porous formations, *Water Resour. Res.*, 30(12), 3327-3336, 1994.

Dagan, G., V. Cvetkovic, and A. Shapiro, A solute flux approach to transport in heterogeneous formations, 1. The general framework, *Water Resour. Res.*, 28(5), 1369-1376, 1992.

Davis, J.M., *A conceptual sedimentological-geostatistical model of aquifer heterogeneity based on outcrop studies*, Ph.D. dissertation, Dept. of Geoscience, New Mexico Inst. of Mining and Tech., 234 pp., 1994.

Davis, J.M., R.C. Lohmann, F.M. Phillips, J.L. Wilson, and D.W. Love, Architecture of the Sierra Ladrones Formation, central New Mexico: Depositional controls on the permeability correlation structure, *Geol. Soc. Am. Bull.*, 105, 998-1007, 1993.

Delhomme, J.P., Kriging in the hydrosciences, *Advances in Water Resources*, 1, 251-266, 1978.

Domenico, P.A., and G.A. Robbins, A dispersion scale effect in model calibrations and field tracer experiments, *J. Hydrol.*, 70, 123-132, 1984.

Emsellem, Y., and G. deMarsily, An automatic solution for the inverse problem, *Water Resour. Res.*, 7(5), 1264-1283, 1971.

Engesgaard, P., K.H. Jensen, J. Molson, E.O. Frind, and H. Olsen, Large-scale dispersion in a sandy aquifer: Simulation of subsurface transport of environmental tritium, *Water Resour. Res.*, 32(11), 3253-3266, 1996.

Farmer, C.L., *The Generation of Stochastic Fields of Reservoir Parameters with Specified Geostatistical Distributions*, paper presented at the IMA Conference on The Mathematics of Oil Production, Robinson College, Cambridge Univ., July 6-7, 1987.

Freeze, R.A., A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media, *Water Resour. Res.*, 11(5), 725-741, 1975.

Freeze, R.A., and J.A. Cherry, *Groundwater*, Prentice-Hall, New Jersey, 1979.

Fried, J.J., Groundwater Pollution, Elsevier, New York, 1975.

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

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

Gelhar, L.W., and C.L. Axness, Reply to comment on three-dimensional stochastic analysis of macrodispersion in aquifers by L.W. Gelhar and C.L. Axness,*Water Resour. Res.*, 19(6), 1643-1644, 1983b.

Gelhar, L.W., A.L. Gutjahr, and R.L. Naff, Stochastic analysis of macrodispersion in a stratified aquifer, *Water Resour. Res.*, 15(6), 1387-1397, 1979.

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

Gillham, R.W., E.A. Sudicky, J.A. Cherry, and E.O. Frind, An advection-diffusion concept for solute transport in heterogeneous unconsolidated geological deposits, *Water Resour. Res.*, 20, 369-378, 1984.

Goggin, D.J., *Geologically-sensible modelling of the spatial distribution of permeability in eolian deposits: Page Sandstone (Jurassic), northern Arizona,* Ph.D dissertation, 417 pp., The Univ. of Texas at Austin, 1988.

Greenkorn, R.A., and D.P. Kessler, Dispersion in heterogeneous anisotropic porous media, *Ind. Eng. Chem.*, 61(9), 14-32, 1969.

Hewett, T.A., and R.A. Behrens, Conditional simulation of reservoir heterogeneity with fractals, Annual Conference and Exhibition, Soc. Petrol. Eng., *SPE 18326*, Houston, 1988.

Hoeksema, R.J., and P.K. Kitanidis, An application of the geostatistical approach to the inverse problem in two-dimensional groundwater modeling, *Water Resour. Res.*, 20(7), 1003-1020, 1984.

Journel, A.G., Nonparametric estimation and quantitative information — the soft kriging approach, *Math. Geol.*, 15, 269-286, 1986.

Journel, A.G., and F.G. Alabert, Focusing on spatial connectivity of extreme-value attributes: Stochastic indicator models of reservoir heterogeneities, Annual Technical Conference and Exhibition, *SPE Paper 18324*, Houston, 1988.

Journel, A.G., and F.G. Alabert, Non-gaussian data expansion in the earth sciences, *Terra Nova*, 1, 123-134, 1989.

Journel, A.G., and C.V. Deutsch, Entropy and spatial disorder, *Math. Geol.*, 25(3), 329-354, 1993.

Journel, A.G., and C.T. Huijbregts, *Mining Geostatistics*, Academic Press, New York, 1978.

Kitanidis, P.K., Minimum-variance unbiased quadratic estimation of covariances of regionalized variables, *Math. Geol.*, 17, 195-208, 1985.

Kitanidis, P.K., and R.W. Lane, Maximum likelihood parameter estimation of hydrologic spatial processes by the Gauss-Newton method, *J. Hydrol.*, 79, 53-71, 1985.

Kitanidis, P.K., and E.G. Vomvoris, A geostatistical approach to the inverse problem in groundwater modeling (steady state) in one-dimensional simulations, *Water Resour. Res.*, 19(3), 677-690, 1983.

Lehr, J.H., Calming the restless native: how ground water quality will ultimately answer the questions of ground water pollution, in *Ground Water Quality*, edited by C.H. Ward, W. Giger, and P.L. McCarty, p. 11-24, John Wiley and Sons, New York, 1985.

MacMillan, J., and A. Gutjahr, Geological controls and spatial variability for one-dimensional arrays of porosity and permeability normal to layering, in *Reservoir Characterization*, edited by L.W. Lake and H.B. Carroll, Jr., 265-292, 1986.

Mandelbrot, B.B., On the geometry of homogeneous turbulence, with stress on the fractal dimension of the iso-surfaces of scalars, *J. Fluid Mech.*, 72, 401-416, 1975.

Mandelbrot, B.B., *Fractals, Form, Chance and Dimension*, Freeman, London, 1977.

McLaughlin, D., and L. R. Townley, A reassessment of the groundwater inverse problem, *Water Resour. Res.*, 32(5), 1131-1161, 1996.

Myers, D.E., To be or not to be . . . stationary?, *Math. Geol.*, 21(3), 347-361, 1989.

Neuman, S.P., Calibration of distributed parameter groundwater flow models viewed as a multiple-objective decision process under uncertainty, *Water Resour. Res.*, 9(4), 1006-1021, 1973.

Neuman, S.P., A statistical approach to the inverse problem of aquifer hydrology, 3. Improved solution method and added perspectives, *Water Resour. Res.*, 16(2), 331-346, 1980.

Neuman, S.P., Statistical characterization of aquifer heterogeneities: An overview, *Geol. Soc. Amer. Spec. Pap. 189*, 81-102, 1982.

Neuman, S.P., Universal scaling of hydraulic conductivities and dispersivities in geologic media, *Water Resour. Res.*, 26(8), 1749-1758, 1990.

Neuman, S.P., Comment on 'A critical review of data on field-scale dispersion in aquifers' by L.W. Gelhar, C. Welty, and K.R. Rehfeldt, *Water Resour. Res.*, 29(6), 1863-1865, 1993.

Neuman, S.P., and S. Yakowitz, A statistical approach to the inverse problem of aquifer hydrology, 1. Theory, *Water Resour. Res.*, 15(4), 845-860, 1979.

Neuman, S.P., and Y.-K. Zhang, A quasi-linear theory of non-Fickian and Fickian dispersion, 1. Theoretical analysis with application to isotropic media, *Water Resour. Res.*, 26(5), 887-902, 1990.

Painter, S., Evidence for non-Gaussian scaling behavior in heterogeneous sedimentary formations, *Water Resour. Res.*, 32(5), 1183-1195, 1996.

Phillips, F.M., and J.L. Wilson, An approach to estimating hydraulic conductivity spatial correlation scales using geological characteristics, *Water Resour. Res.*, 25(1), 141-143, 1989.

Reddell, D.L., and D.K. Sunada, Numerical simulation of dispersion in groundwater aquifers, *Hydrol. Paper 41*, Colo. State Univ., Fort Collins, 79 pp., 1970.

Russo, D., and W.A. Jury, A theoretical study of the estimation of the correlation scale in spatially variable fields, 1. Stationary fields, *Water Resour. Res.*, 23(7), 1257-1268, 1987.

Scheidegger, A.E., Statistical hydrodynamics in porous media, *Advances in Hydrosciences*, 1, 161-181, 1964.

Schwartz, F.W., Macroscopic dispersion in porous media: The controlling factors, *Water Resour. Res.*, 13(4), 743-752, 1977.

Smith, L., and F.W. Schwartz, Mass transport, 1. A stochastic analysis of macroscopic dispersion, *Water Resour. Res.*, 16(2), 303-313, 1980.

Sudicky, E.A., and J.A. Cherry, Field observations of tracer dispersion under natural flow conditions in an unconfined sandy aquifer, *Water Pollut. Res. Can.*, 14, 1-17, 1979.

Sudicky, E.A., J.A. Cherry, and E.O. Frind, Migration of contaminants in groundwater at a landfill: A case study, 4. A natural gradient dispersion test, *J. Hydrol.*, 63, 81-108, 1983.

Sun, N.-Z., Theory and applications of transport in porous media, v. 6, in *Inverse Problems in Groundwater Modeling*, J. Bear (ed.), Kluwer Acad., Norwell, Mass., 1994.

vanMarcke, E., *Random Fields: Analysis and Synthesis*, MIT Press, Cambridge, 1993.

Vasco, D.W., Datta-Gupta, A.F., and J.C.S. Long, Resolution and uncertainty in hydrologic characterization, *Water Resour. Res.*, 33(3), 379-397, 1997.

Wheatcraft, S.W., and S.W. Tyler, An explanation of scale-dependent dispersivity in heterogeneous aquifers using concepts of fractal geometry, *Water Resour. Res.*, 24(4), 566-578, 1988.

Wilson, J.L., P.K. Kitanidis, and M. Dettinger, State and parameter estimation in groundwater models, in *Applications of Kalman Filtering to Hydrology,*

Hydraulics and Water Resources, edited by C.-L. Chiu, 657-679, Univ. of Pittsburgh, Pittsburgh, Penn., 1978.

Yakowitz, S., and L. Duckstein, Instability in aquifer identification: Theory and case studies, *Water Resour. Res.*, 16(6), 1045-1064, 1980.

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

Yeh, W.W.-G., and Y.S. Yoon, Aquifer parameter identification with optimum dimension in parameterization, *Water Resour. Res.*, 17(3), 664-672, 1981.

Zhang, Y.-K., and S.P. Neuman, A quasi-linear theory of non-Fickian and Fickian subsurface dispersion, 2. Application to anisotropic media and the Borden site, *Water Resour. Res.*, 26(5), 903-913, 1990.

Chapter 2.
Effective Properties and Movement of Information Between Spatial Scales

Scale-variant geologic structure has been the focus of much research during the past decade, initially motivated by efforts in the petroleum industry to maximize oil production in hydrocarbon reservoirs. Central to much of this research has been the recognition that traditional geostatistical characterization techniques are inadequate for describing natural porous media that are neither statistically homogeneous, nor characterized by structure possessing a single length scale. Most natural geologic variability is instead characterized either by multiple, discretely-nested scales of variation that produce parallel hierarchies of flow and transport processes, or by continuously evolving scales of variation manifested as fractal structure (Cushman, 1990). These discrete or continuous geologic hierarchies produce statistical nonstationarities that render the concept of the representative elementary volume (REV), and traditional continuum theories of transport based on the REV concept, effectively meaningless.

The work of Bhattacharya and Gupta (1983) was among the first to demonstrate mathematically how the mechanism of transport at each distinct scale can be physically different from that occurring at other scales. They showed that at each scale, the transport equations possess basically the same form, but require a different interpretation of the parameters. Cushman (1983) subsequently attempted to interpret

the multiple-scale transport problem using deforming volumes. Approaching the problem from a different perspective, Baveye and Sposito (1984) recognized that transport equations were valid for making predictions at the scale of the measuring device only when convolved with the weight function associated with the device. In response to that research, Cushman (1984) argued that noise is introduced into such convolved equations because the transport process measured at one scale cannot be fully explained by equations derived at another. He suggested that it is better to design instruments that measure at the appropriate length scale than convolve mathematical representations of existing devices with the transport equations.

Towards the end of the chapter, a numerical experiment is used to illustrate Cushman's thesis. The experiment involved introduction of noise to hydraulic conductivity realizations characterized both by single and multiple scales of spatial variation. Results of the experiment demonstrate that the practice of upscaling conductivity in an aquifer characterized by more than one scale of spatial variation can result in a significant change in predicted transport behavior. Results of the experiment suggest that, in porous media exhibiting multiple scales of variation, upscaling should be avoided whenever possible by acquiring conductivity measurements at each distinct scale of spatial variation that may affect the prediction.

Chapter 2. Effective Properties and Movement of Information Between Spatial Scales

2.1 THE REPRESENTATIVE ELEMENTARY VOLUME

If it were possible to know the exact geometry of pore spaces in a porous medium, the Navier-Stokes equation could be used to solve for fluid flow at the microscale, subject to boundary conditions of zero velocity at fluid-solid interfaces. However, the solution could not be extrapolated to other portions of the porous medium or to a different porous medium, and would therefore be of limited practical value. Instead, the flow problem must be addressed at a larger scale, by defining meaningful averages of microscopic properties. The question then becomes: what defines a meaningful averaging volume in the context of flow and transport through a hierarchical porous medium?

The representative elementary volume, or REV, has traditionally been used to define a meaningful averaging volume in the framework of macroscopic theories of fluid transport through porous media. These theories rest on the continuum hypothesis, which presumes that the spatial distribution of aquifer properties can be described by mathematical field variables assigned to centroids of REVs (Baveye and Sposito, 1984). Briefly stated, an REV represents a volume that is large relative to the size of pore spaces but small compared to the scale of heterogeneities (Bakr *et al.*, 1978). As long as the REV is small with respect to the scale of the heterogeneities, is indifferent to the actual macroscopic field variable of interest, and possesses geometrical characteristics that are temporally- and spatially-invariant, the REV average has meaning and the continuum description is valid (Bear, 1972). The REV

hypothesis then makes possible a continuum description of temporal and spatial dependence of field variables using a partial differential mass-balance equation.

Figure 2.1 illustrates the concept of the REV graphically via a plot of a physical property of the porous medium vs. length scale of the averaging volume. One can view the ordinate of the plot as a specific density-related hydrogeologic property such as porosity. At the scale of a point, either a solid phase or a void phase will be encountered, but not both. Hence, the REV is undefined at the point scale. As the size

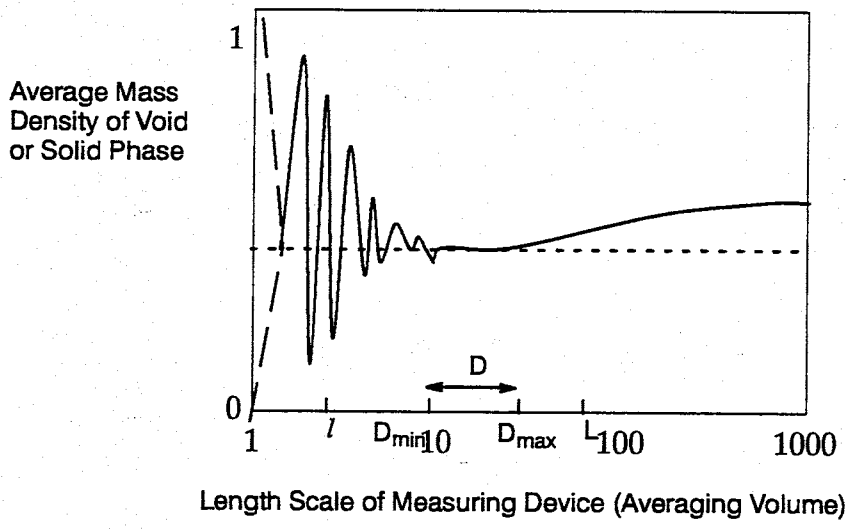


Figure 2.1. REV of a Physical Property (after Baveye and Sposito, 1984).

of the measuring device increases, so too does the scale of averaging. Erratic fluctuations will tend to occur as the fraction of void or solid space randomly dominates the measurement. Eventually, these fluctuations tend to zero, and the property remains fairly stable over some interval of averaging length scale D. Note that one could view the plot in Figure 2.1 as an indication of the uncertainty or measurement error associated with measuring volumes of increasing size, with the REV corresponding to the volume at which zero sampling error occurs, assuming infinite instrument precision.

Within interval D, the property measurement is independent of the averaging volume associated with the measurement — the REV is sufficiently large to sample many voids and solids, yet small enough to possess physical meaning in a macroscopic sense, in a mathematical neighborhood of its centroid (Sposito, 1978). Bear (1972) proposed that the most appropriate definition of the REV be the smallest averaging volume for which the measurement is independent of element size, or the lower limit on the interval D_{min}. If the medium is heterogeneous at the macroscopic scale, larger scale variations in the property begin to enter the average at larger length scales, producing low-frequency fluctuations, shown in Figure 2.1 at length scales larger than D_{max}. Thus, the length scale of the REV denoted by any value within range D must not only be much larger than the microscopic characteristic length l or mean pore diameter of the porous medium, but also much smaller than the macroscopic characteristic length L associated with gross inhomogeneities in the property (Baveye and Sposito, 1984):

$$l << D_{min} < D_{max} << L \qquad [2.1]$$

However, there do occur situations where the averaged property will not exhibit a range of length scales within which the property is independent of averaging volume. For example, in a layered medium, macroscopic length scale L varies from layer to layer and the interval D becomes ill-defined. Furthermore, for nonlinear systems, the point property of the system will be a function of system state. Under such conditions, an REV will exist only if there is a range of length scales common to the property over all physical states of the system for which equation [2.1] holds. In addition, the REV associated with one property may be different than that associated with another. In such cases, a range D common to all constitutive variables associated with the transport equation may not even exist. And finally, in a porous medium exhibiting multiple scales of variation, no single REV will exist that has meaning in the framework of continuum theory.

Even for situations in which a valid D does occur, its existence or lack of existence is inherently unverifiable because one can never prove or disprove either the invariance of the indifference property. Since the hypothesis on which the REV concept is built is inherently unverifiable, the very basis of the REV concept has been called into question by many researchers. Cushman (1983) and Baveye and Sposito (1984) were the first to propose that it be abandoned in favor of an approach that does not require unverifiable invariance and indifference hypotheses. They offer an alternative to the REV approach, replacing the absolute concept of an REV with one

in which the measured property is considered dependent on the method of measurement. That is, the existence of an absolute value of the property is denied — instead, the property is defined relative to the characteristics of the measuring device. Such a relativist concept, in which the measurement is 'relative' to the measurement operation, avoids adoption of unverifiable hypotheses inherent to application of the REV approach.

2.2 CONSTITUTIVE VARIABLES AND THE 'IDEAL' INSTRUMENT

The work of Cushman (1987a,b, 1990) focused on some of the more theoretical aspects of characterizing multiple-scale geologic processes and constitutive variables. Since constitutive variables are derived from state variables such as hydraulic head or concentration using inversion techniques, they possess meaning only in the context of the instruments used to measure the state variable. As one moves through the geologic hierarchy and averages over larger and larger length scales, new constitutive variables are created to act as place holders for the next level of averaging. Cushman argued that this process of upscaling results in an irreversible reduction in degrees of freedom, or loss of information, because each of these constitutive variables ignores useless information at larger or smaller scales. The information is useless in the sense that it is supplied at a scale too large or too small to be relevant at the particular scale at which averaging is actually performed.

For example, when averaging up from a smaller scale, conductivity details that are irrelevant to large-scale head predictions become lost. Conversely, when scaling

down through the hierarchy to smaller scales, larger-scale, lower-frequency conductivity trends become meaningless in the context of making small-scale predictions. At each scale, implicit, unmeasurable heterogeneities associated with other scales are present, but cannot be observed explicitly with the measuring device associated with that scale. However, these implicit heterogeneities have an impact on flow and transport predictions made at the scale of the averaging device. We can try to account for the implicit heterogeneities via geostatistical inference. Alternatively, implicit heterogeneities can be converted to explicit heterogeneities by altering the scale of the measuring device.

Cushman (1984) suggested that exactly this approach be used to conserve implicit information, by observing structure through a variety of instrument windows corresponding to distinct, natural scales of variation. In so doing, information observed at one scale is independent of information acquired at other scales, and the very averaging procedures responsible for loss of information are no longer required. The information observed at one scale can often be analyzed to determine whether it bears a systematic relation to information observed at another scale, and if so, the manner in which information moves between scales can be assessed.

As Figure 2.2 illustrates, multiple scales in natural geologic media are manifested as intermittently asymptotic REV limits or plateaus with increasing instrument window size. Variograms corresponding to such structure might exhibit stationary behavior at length scales where REV asymptotic limits are attained, with increasing variance inside transition zones between asymptotes. For example, the

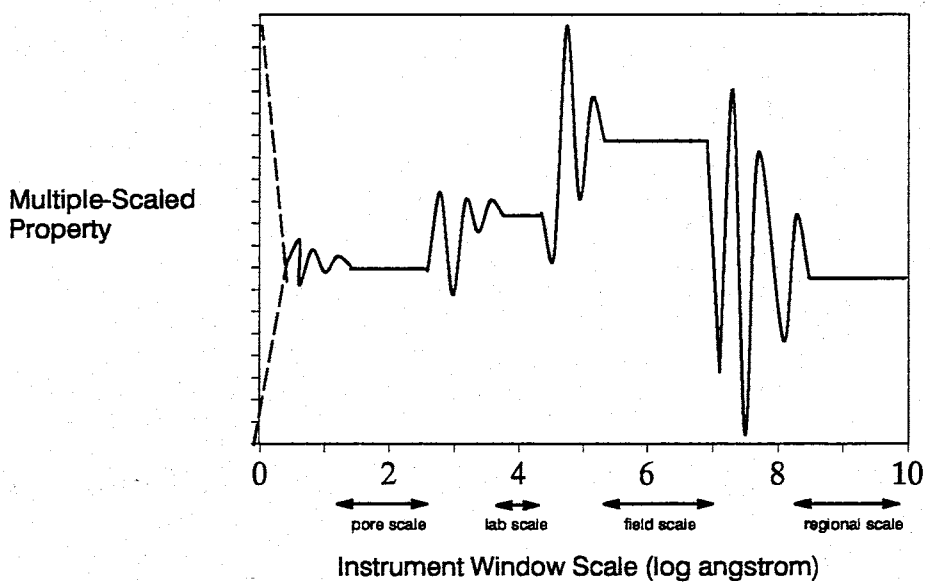


Figure 2.2. Instrument Windows of a Discrete Geologic Hierarchy (after Cushman, 1984).

variogram associated with each layer of Figure 1.3 might approach a sill distinct to that layer and corresponding to the variance of within-layer increments, but when separation distance exceeds layer thickness, variance will begin to increase until a sill reflecting average variability among all layers is attained. This situation differs from that associated with fractal structure, in which continuously evolving scales of heterogeneity contain information that can be scaled or compressed. Rather,

information in a discrete geologic hierarchy can only be decomposed at natural breaks in the REV plot denoted by the intermittent plateaus.

In order to perform such a decomposition, the hierarchical process must be observed at each distinct scale, through instrument windows designed specifically to measure variation at each scale. It is in this context that the relativist approach introduced earlier begins to take on meaning. While the REV approach requires that measurements collected over an REV are representative and theoretically valid, the relativist approach generalizes the concept of sampling domain so that it transcends the idea of a representative domain of measurement, with all measurements regardless of sampling domain size or geometry being equally valid.

Matheron (1965) was among the first to propose using a relativist framework to analyze ore deposit samples. Marle (1967) later mathematically formalized this relativist idea by defining a three-dimensional macroscopic field variable $Z_m(\bar{x})$ as the volumetric convolution of the point microscopic property $Z(\bar{x})$ with the weighting function $m(\bar{x})$ associated with both the property and the measuring device:

$$Z_m(\bar{x}) = \int_{V(\bar{x})} Z(\bar{x} + \bar{\xi})\, m(\bar{\xi})\, d^3\bar{\xi} \qquad [2.2]$$

The above convolution is also referred to as the regularization of $Z(\bar{x})$ in volume $V(\bar{x})$ by $m(\bar{x})$, with the weighting function $m(\bar{x})$ embodying the effect of the measuring device on the macroscopic property. The same type of filtering approach can be used to relate measurements collected on one support to those collected on another. Beckie and

Wang (1994) used such an approach to relate core-scale conductivity to conductivity estimated from slug testing. Similar filtering techniques have been used by Oliver (1990) and Desbarats (1994).

$Z_m(\bar{x})$ reflects not only characteristics of the porous medium, but also the unique properties of the measuring device. Note that this generalized definition of macroscopic property not only transcends the indifference and invariance properties required of an REV, but is also independent of the scale of the measuring device. The only constraint on the device is that it does not amplify the property, only that it average it:

$$\int_{V(\bar{x})} m(\bar{\xi})\, d^3\bar{\xi} = 1 \qquad [2.3]$$

Equation [2.2] allows for the unique influence of the measuring device to be incorporated into the macroscopic estimate. The 'size' of the instrument is irrelevant — it need not be much larger than *l* nor much smaller than L, because the unique way that the instrument records the macroscopic signature of the property is already fully accounted for in $m(\bar{x})$. Moreover, $m(\bar{x})$ also accounts for the specific geometry of the instrument window. The REV approach, on the other hand, does not explicitly incorporate the influence of the measuring device into the macroscopic average $Z_m(\bar{x})$:

$$Z_{REV}(\bar{x}) = \frac{1}{V_{REV}} \int_{V_{REV}} Z(\bar{\xi}) \; d^3\bar{\xi} \qquad [2.4]$$

Rather, it measures or 'weights' the microscopic property $Z(\bar{x})$ in the same way at all locations, in a manner inversely proportional to the size of the REV element, V_{REV}. It is only when all instruments record the microscopic property $Z(\bar{x})$ in the same way at all locations and scales that the absolutist REV and relativist macroscopic averages become indistinguishable.

In equation [2.2], $m(\bar{x})$ supplies a general description of an instrument that averages property information over space. In multiple-scale porous media, the concept of instrument $m(\bar{x})$ must be modified to account for different scales of measurement. Cushman extended the relativist framework embodied by the spatial convolution given by equation [2.2] to the frequency domain, deriving a generalized mathematical expression for an ideal instrument that measures a random process at a frequency characteristic of the instrument. This ideal instrument is mathematically equivalent to a measure that filters out all high-frequency information contained in the process and leaves all low-frequency information below the characteristic frequency of the instrument intact. Such an instrument removes small-scale, unstructured 'noisy' information while preserving larger-scale systematic information conveying more structured variation. Note that the terms 'structured, systematic' and 'unstructured,

noisy' are relative terms that refer to information impacting predictions made at a particular scale of measurement.

Cushman (1983) introduced the concept of a measure, $P(d\bar{s})$, as a mathematical analog to an instrument, where $d\bar{s}$ is a space increment. If $g(\bar{s})$ is the characteristic or weighting function associated with the device, then according to integration theory:

$$P(d\bar{s}) = g(\bar{s})\, d\bar{s} \qquad [2.5]$$

As in the space domain representation of a filter, an instrument is a filter if the measurement is not amplified during sampling:

$$\int_{R^3} P(d\bar{s}) = 1 \qquad [2.6]$$

where R is a characteristic length scale of the device. Equation [2.5] can be expressed in the spectral domain by performing a Fourier transform:

$$\hat{P}(\bar{\omega}) = \int_{R^3} e^{-i\bar{s}\bar{\omega}} P(d\bar{s}) \qquad [2.7]$$

Certain constraints must be imposed on the instrument to account for the fact that it cannot 'see' information at frequencies higher than the characteristic frequency of the

instrument. If $P\hat{}(\overline{\omega})$ goes to zero very rapidly for high frequencies $|\overline{\omega}| >> \omega_0$, the instrument filter is characterized by scale λ_0 and frequency $\omega_0 = \lambda_0^{-1}$.

One can't expect to be able to design a real, physical instrument for which high-frequency components go immediately to zero above some frequency cut-off. Instead, real constants ε and γ are defined such that the power of the instrument goes to zero within some finite frequency interval between ε and γ:

$$\begin{aligned} \hat{P}(\omega) &= 1 \qquad |\omega| < \varepsilon \leqq \lambda_0^{-1} \\ \hat{P}(\omega) &= 0 \qquad \lambda_0^{-1} < \gamma \leqq |\omega| \end{aligned} \qquad [2.8]$$

where ε and γ are small positive numbers. For discrete multiple-scale problems, we are interested in sequential averages that remove successively higher-frequency noise above distinct cut-offs, leaving the lower-frequency signal intact. Hypothetical examples of the spectral densities of such a succession of instruments are presented in Figure 2.3.

Using the mathematical representation of an ideal instrument and theorems that relate derivatives of averages to averages of derivatives, Cushman (1983) succeeded in building different scales of instruments into the transport equation. He did so by deriving transport equations at different scales of variability, convolving the transport equation with filters associated with various instrument frequencies, and using averaging theorems to complete the upscaling process. On the basis of this

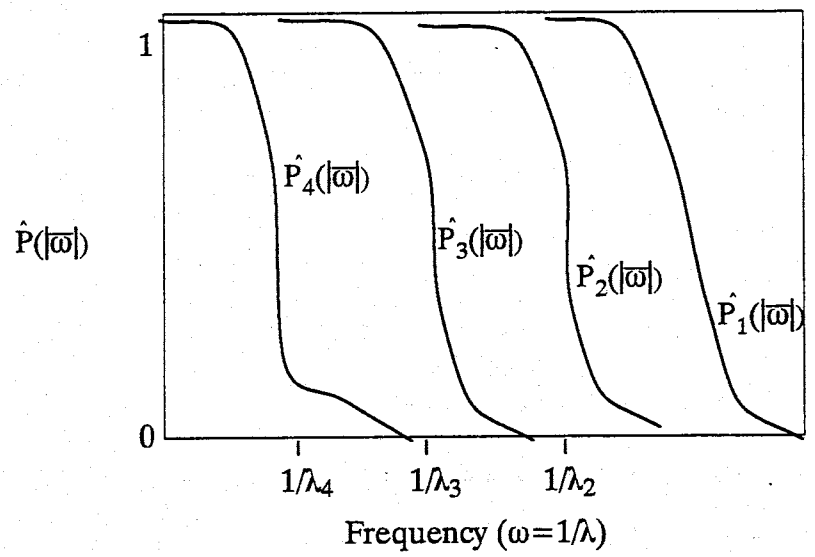


Figure 2.3 Spectral Density of Ideal Instruments with Different Characteristic Frequencies (after Cushman, 1984).

procedure, he was able to account for the net effects of unresolved heterogeneities not explicitly recorded by an instrument.

2.3 EQUIVALENT AND EFFECTIVE PROPERTIES

The issue of scale-dependent measurement relates closely to the problem of defining equivalent and effective properties. The problem of defining such properties is of great concern, because most hydrologic parameters such as hydraulic conductivity

are not typically measured at a scale on the same order as the size of finite-difference (FD) or finite-element (FE) blocks used to discretize the flow or transport domain. Due to constraints on computer CPU and disk space, most practicing hydrologists are interested in performing numerical simulations at scales much larger than the scale at which conductivity data are generally available, and are only concerned with small-scale variability to the extent that it influences the effective behavior of the large-scale block. This problem is referred to as upscaling because it involves defining the parameters over block sizes larger than the support size of their original measurement. Some type of upscaling rule must be used to translate small-scale conductivities to larger-scale values that can be assigned to FD or FE blocks. Without such systematic upscaling, many problems of practical interest would be hopelessly intractable.

Yet the prospect of losing valuable conductivity information during the upscaling procedure is almost as undesirable as construction of a numerical model containing thousands of FD or FE nodes. The upscaling rule must somehow preserve the effect of conductivity at small, unresolved scales while not requiring that the small-scale conductivity distribution be explicitly incorporated into the numerical model. It must do so in such a way that global aquifer behavior of interest is approximately reproduced, but not at the expense of losing a good approximation of local behavior. Finally, it should also handle data collected at a variety of instrument supports in a consistent manner so that information measured at one scale is not distorted or lost during the upscaling procedure at the expense of information observed

at another scale. Beckie (1996) refers to the problem of accounting for unresolved physics at subgrid scales of variation as the closure problem. For example, the closure model for contaminant transport is the ADE with macrodispersivities used as phenomenological coefficients to describe the effects of unresolved heterogeneities on the concentration prediction. Different scales of simulation will produce differing amounts of unresolved physics at the simulation scale and different coefficients. Thus, the upscaled value, and presumably the associated upscaling rule, can be expected to change with grid resolution. For unresolved physics at a scale much smaller than the grid scale, however, it can often be assumed that resolved and unresolved physics are independent and that grid-scale parameters are not strongly scale-dependent (Beckie, 1996).

It is the search for this elusive upscaling rule that has motivated research into equivalent and effective aquifer properties. Theories of upscaling in porous media have undergone rapid development during the past decade, inspired by research into homogenization theory that began during the late 1970's. Homogenization or diffusion theory involves replacement of complicated small-scale permeability structure by an asymptotically equivalent homogeneous structure possessing properties similar to the real medium. More specifically, the equivalent medium must have approximately the same diffusive properties as the real medium over long space and time scales. While homogenization theory is precise only in media exhibiting periodic structure, it can yield good approximations of effective properties in random porous geologic media when the size of the averaging domain is large relative to the characteristic length scale

of the heterogeneities (Durlofsky, 1991; Kitanidis, 1990) or when an ensemble approach is used to define effective behavior (Koebbe, 1995).

In the context of flow simulation, equivalent conductivity in an FD or FE block b is defined as the average flux through the block $<Q_b>$ predicted by the smaller-scale model, divided by the negative of the average uniform hydraulic gradient in the block, $<J_b>$:

$$K_{eq,\,b} = -\frac{< Q_b >}{< J_b >} \qquad [2.9]$$

Although equivalent conductivity is often used interchangeably with the concept of effective conductivity, equivalent and effective properties are in fact two distinct concepts. While equivalent conductivity varies spatially over a realization depending on changes in average block fluxes and hydraulic gradients and according to block geometry, effective conductivity represents mean behavior over an ensemble of realizations and is defined as:

$$K_{eff} = -\frac{< Q >}{< J >} \qquad [2.10]$$

where $<Q>$ and $<J>$ are spatial averages over a realization of an ergodic process. Implicit to the use of equation [2.10] is the assumption that the scale of averaging be large compared to the scale of heterogeneities, so that the averages are meaningful estimates of global behavior. Only then will flux and gradient processes be ergodic or

statistically representative of the true flow process and effective conductivity be an ensemble parameter rather than a spatial parameter, independent of flow conditions. Under conditions of ergodicity, as block size increases $K_{eq,b}$ becomes more of an ensemble statistic and eventually converges to K_{eff}. Because we are concerned with ensemble behaviors of single-scale and multiple-scale conductivity processes, our interest lies in effective properties.

Warren and Price (1961) used Monte Carlo simulation and Matheron (1967) relied on perturbation techniques to demonstrate that effective conductivity is equal to the geometric average of point conductivities when a mean, steady-state uniform gradient is imposed on a two-dimensional, isotropic random conductivity process. Gutjahr *et al.* (1978) verified this conclusion and developed first-order approximations to effective conductivity for isotropic structure. Gelhar and Axness (1983) further extended this work for the anisotropic case in stratified media. Dagan (1982) and Kitanidis (1990) extended the analysis to account for gradually varying mean flow. Desbarats (1987) numerically investigated effective conductivity in the presence of discontinuous shale barriers. Ababou and Wood (1990) and Desbarats (1992) employed power-law averaging techniques to estimate effective conductivity. Indelman *et al.* (1996) analytically derived an expression for equivalent conductivity under conditions of steady flow to a well. One of the more physically-rigorous approaches for defining equivalent properties was based on conserving the mean and covariance of the rate of energy dissipation in the block, so that global flow behavior is approximately reproduced (Indelman and Dagan, 1993a,b). Sanchez-Vila *et al.*

(1995) demonstrated that the approach of Rubin and Gomez-Hernandez (1990), which relied on a perturbation approach coupled with projection in the direction of mean flow, yields results equivalent to those of Indelman and Dagan (1993a,b). Their approach, which preserves average energy dissapation, was used by Bierkens (1996) to upscale conductivity of a confining layer. Flow and transport behavior predicted from the upscaled conductivity closely reproduced that of the original conductivity process.

Perturbation techniques have also been used to define effective field-scale dispersivity, or macrodispersivity (Gelhar and Axness, 1983; Dagan, 1984, 1988). Macrodispersivity represents asymptotically Fickian diffusion caused by field-scale heterogeneities. Expressions for ensemble macrodispersivity in statistically isotropic and stratified porous media derived by Gelhar and Axness (1983), Naff *et al.* (1988), and Dagan (1984, 1988) have agreed well with the Monte Carlo simulation results provided by Warren and Skiba (1964), Smith and Schwartz (1980), and Tompson and Gelhar (1990) and with estimates from field experiments (Sudicky *et al.*, 1983; Freyberg, 1986; Sudicky, 1986; Garabedian *et al.*, 1991).

We are most interested in how effective properties estimated on the basis of single-scale structure differ from those estimated when the discrete multiple-scale geologic structure is accounted for. More importantly, how does conductivity information scale-up differ in single-scale and multiple-scale geologic environments? The question is an important one because its answer lies in how information is changed, lost, or distorted as it is transported from one scale to another.

2.4 MULTIPLE-SCALE GEOLOGIC STRUCTURE

The existence of natural, embedded scale-variant structure has been observed in the Sierra Ladrones Formation (SLF) fluvial deposits in the Albuquerque Basin at several scales of observation (Davis *et al.*, 1990; Davis, 1994). An outcrop photograph of fluvial deposits observed in the SLF at a scale of observation on the order of 1 meter is presented in Figure 2.4. At this particular scale of observation, small-scale spatially continuous structure on the order of centimeters appears to be embedded within larger, submeter-scale discrete structure.

An interesting phenomenon occurs if we perform a Gedanken experiment similar to the one presented in Chapter 1, panning away from the outcrop to view the deposits from a larger, 30-meter scale of observation, as illustrated in Figure 2.5. The same sample of the outcrop as presented in Figure 2.4 is shown blocked off in Figure 2.5 to demonstrate that structure perceived as being discretely horizontally layered at the smaller, 1-meter scale of observation now appears more spatially continuous at the 30-meter observation scale. Moreover, the degree of spatial variability within the block appears to be smaller at the 30-meter scale of observation than at the 1-meter scale of observation.

Why is this happening? Why do we perceive the same sample of the outcrop differently at different scales of observation? Why do features that previously appeared spatially discrete at the 1-m scale of observation look more spatially continuous at the 30-m observation scale? And what is causing a reduction in the degree of spatial variability as we move to the larger scale of observation? All of these

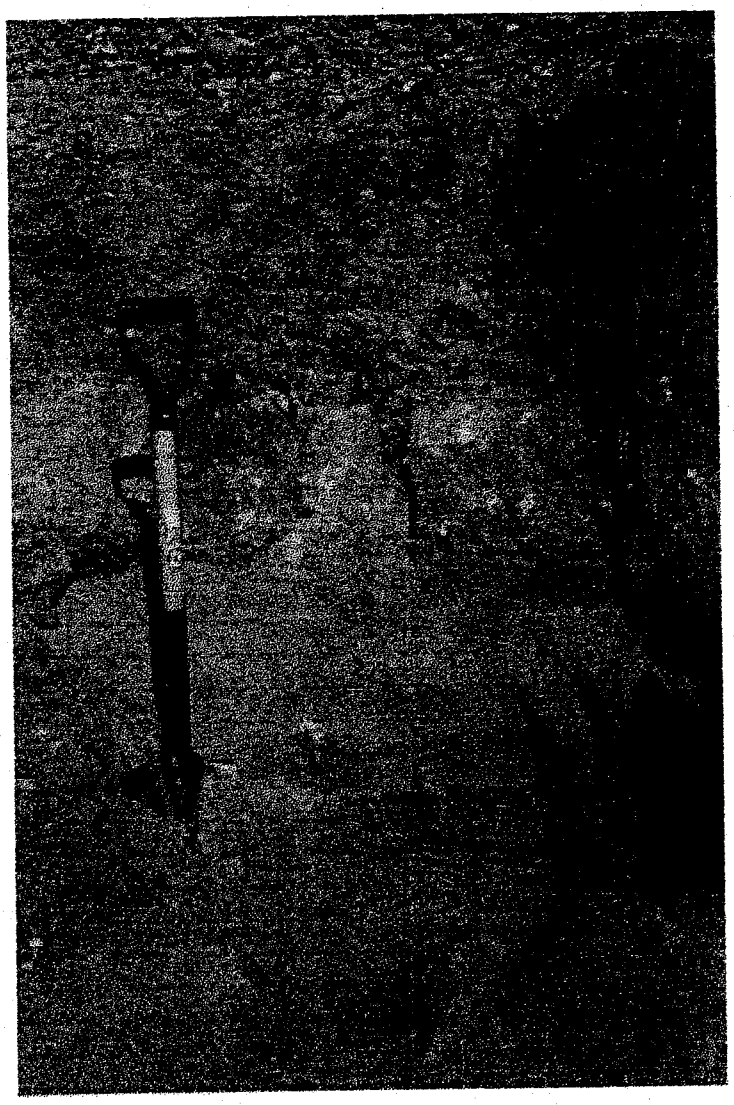


Figure 2.4. Sierra Ladrones Formation Deposits at the 1-Meter Observation Scale (courtesy J. Matthew Davis).

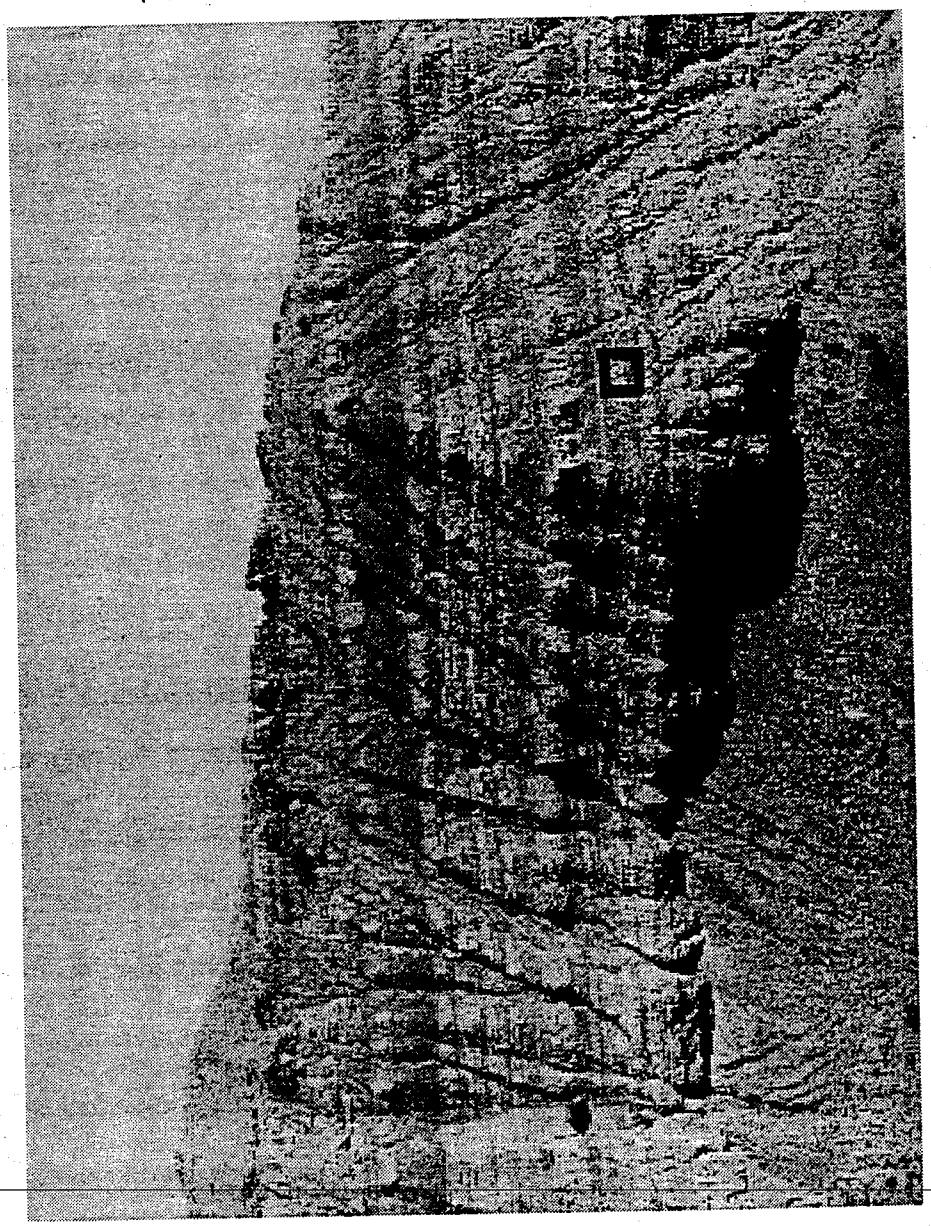


Figure 2.5. Sierra Ladrones Formation Deposits at the 30-Meter Observation Scale (courtesy J. Matthew Davis).

changes occur because the camera used to record the spatial variability is a fixed-resolution instrument, registering heterogeneity over a fixed number of pixels. Each pixel must average geologic information over a wider area as the scale of observation increases, resulting in more 'cloudy' vision as the distance between the point of focus and the camera increases. As we pan away from the outcrop and force the same number of pixels to record geologic information over a wider area, we perceive not only less spatial variability, but also experience changes in the nature of this variability as spatially discrete structure becomes obscured by the averaging process.

As before, the Gedanken experiment demonstrates that perceived randomness can be a function of scale. Randomness, rather than being a fixed property of a system, represents a mathematical abstraction used to deal with things that we can't measure, 'see', or explain. If all variables influencing our system of heterogeneities were fully known, or explained, no randomness exists in our system and the character of all heterogeneities in the outcrop could be exactly predicted. However, we rarely find ourselves in such a position. Instead, we must deal with the effects of unmeasurable or unknown influences that give rise to unpredictable or random behavior. As an example, consider the large-scale nonstationary spatial variation in one-dimensional conductivity shown in Figure 2.6. It can be considered deterministic because we can easily see it or infer it from large-scale geology, while smaller-scale variations superimposed on the larger-scale deterministic structure appear 'random' because they are unknowable or unmeasurable with the same instrumentation used to measure

the larger-scale variation. Yet when viewed from a closer perspective using an instrument that *is* capable of resolving smaller scales of variation as illustrated by the inset in the figure, the small-scale variation can be seen to exhibit deterministic structure as well. This process of zooming in on conductivity variation could continue *ad infinitum*, finding deterministic pattern in seemingly random variation. This is what is meant by the statement that randomness can be a function of observation scale.

With regard to prediction of fluid flow, there are strong implications associated with the concept that randomness depends on observation scale. If we have access to geologic information only at the 30-meter scale of observation, and can resolve heterogeneities only at scales of 1-meter or larger, we will not be aware of small, submeter-scale variability occurring in the outcrop. We know that submeter-scale variability is present either because we have photographically 'measured' it at the 1-meter scale, as recorded in Figure 2.4, or have some sense that it is there when we are unable to reproduce observed aquifer response using a numerical model with meter-scale discretization. However, in the absence of this information we will be somewhat oblivious to the presence of submeter-scale heterogeneities and inclined to assign a single value of conductivity to each 1-meter scale FD or FE blocks. In general, predicted head or concentration distributions will not match the observed distributions because the prediction model has not been informed on the submeter-scale heterogeneities, while the real aquifer has integrated those heterogeneities into its response. Note that, due to the different physics associated with the flow and transport

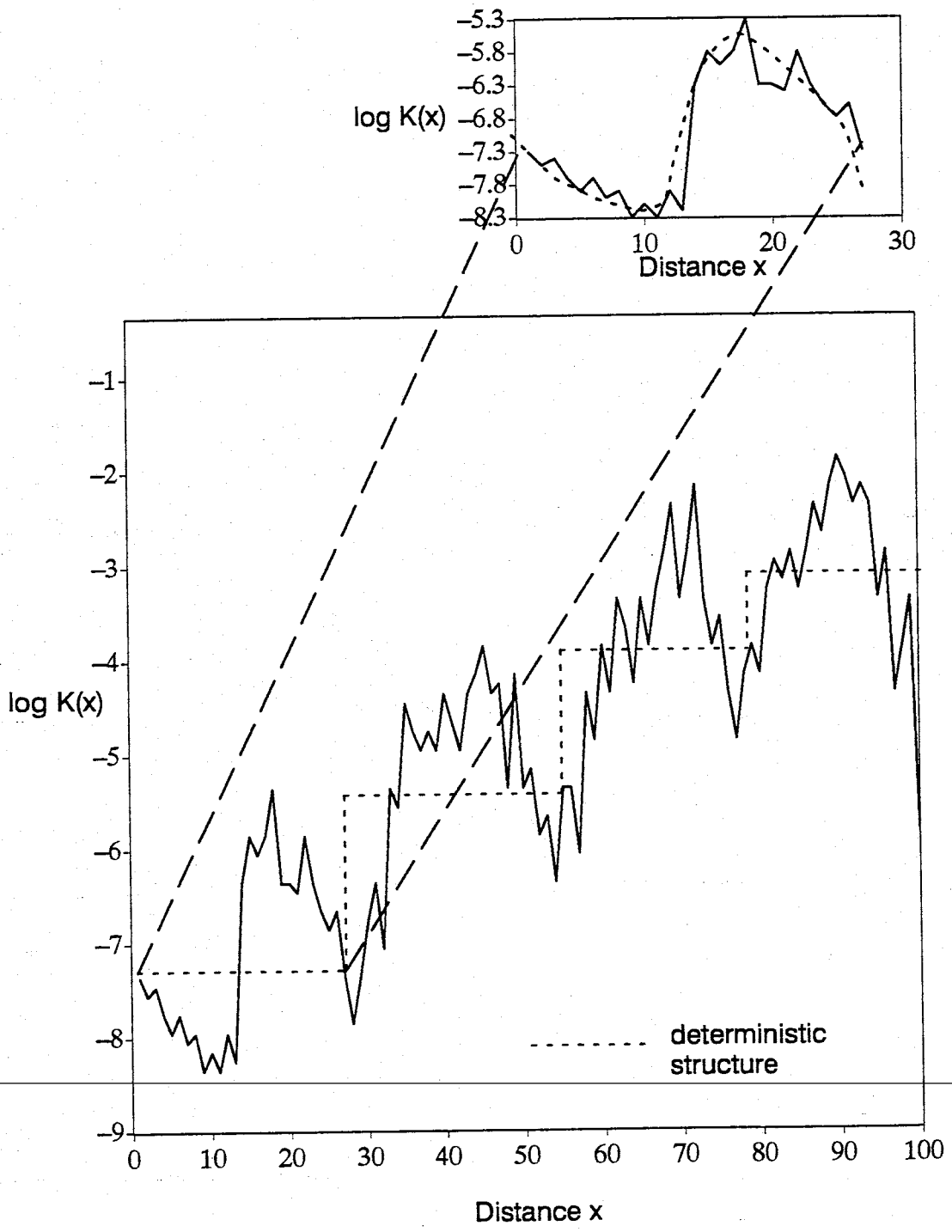


Figure 2.6. Randomness and Structure as a Function of Observation Scale.

equations, the degree of mismatch can be expected to be much more pronounced for the case of concentration prediction than for prediction of hydraulic head.

From these considerations, it is evident that measurement of a multiple-scale random geologic process through a single, arbitrary instrument window (the large-scale photograph) will not provide us with all the information contained in the process. Unlike the case for single-scale heterogeneity, for which a typical, or ergodic, sample of the process can be observed through all instrument windows larger than the scale of variation, there generally exists no such typical sample in a hierarchical geologic process except at scales greatly exceeding the scale of the largest structure. It would appear that successful characterization of multiple-scale aquifer variation depends on the extent to which the limitations of observing a multiple-scale process through a single instrument window can be overcome. The essence of Cushman's argument is that these limitations cannot be overcome, and that we need to choose a variety of instruments through which to observe the multiple-scaled geologic process.

Three key issues are of concern when characterizing multiple-scale variations. First, how does one go about reproducing randomness as a function of scale? For example, at the larger 30-meter scale shown in Figure 2.5, how can we account for geologic variation within a 1-m^2 sample of the outcrop that we cannot explicitly see at the larger scale, but that we know for certain to be there based on our observation in Figure 2.4? Secondly, how does conductivity information scale, or move between spatial scales? Clearly, if it changes when we try to move it to another scale, then the practice of using measurements collected at one scale to make predictions at another

scale must be called into question. Finally, how do we go about measuring hierarchical geologic heterogeneities if we want to be able to exert some degree of control over transport prediction uncertainties — specifically, if we want to reduce them?

2.5 EMBEDDING OF SCALES

As illustrated schematically in Figure 2.7, a multiple-scale random geologic process can be constructed in a purely descriptive fashion by embedding different scales of variation observed through progressively larger instrument windows, beginning with discrete spatial variation at coarser scales where the assumption of spatially continuous structure cannot typically be made. A spatially discrete model such as a Markov random field (MRF) model could be used to reproduce geologic structure. Such a model imposes correlation over discrete rather than continuous space. MRF models are discussed in greater detail in Chapter 3. Smaller-scale discrete structure continues to be progressively embedded as we move through the hierarchy until the scale of variation is very small relative to the scale of the problem, at which point discrete geologic structure will, for all intents and purposes, act spatially continuous. At this scale, spatially continuous structure such as that generated by conventional spatially-continuous gaussian models can be incorporated into the model. Note that spatially-continuous structure is considered desirable because analytical methods of synthesizing continuous structure are much simpler to implement than combinatorial methods of generating discrete MRF structure.

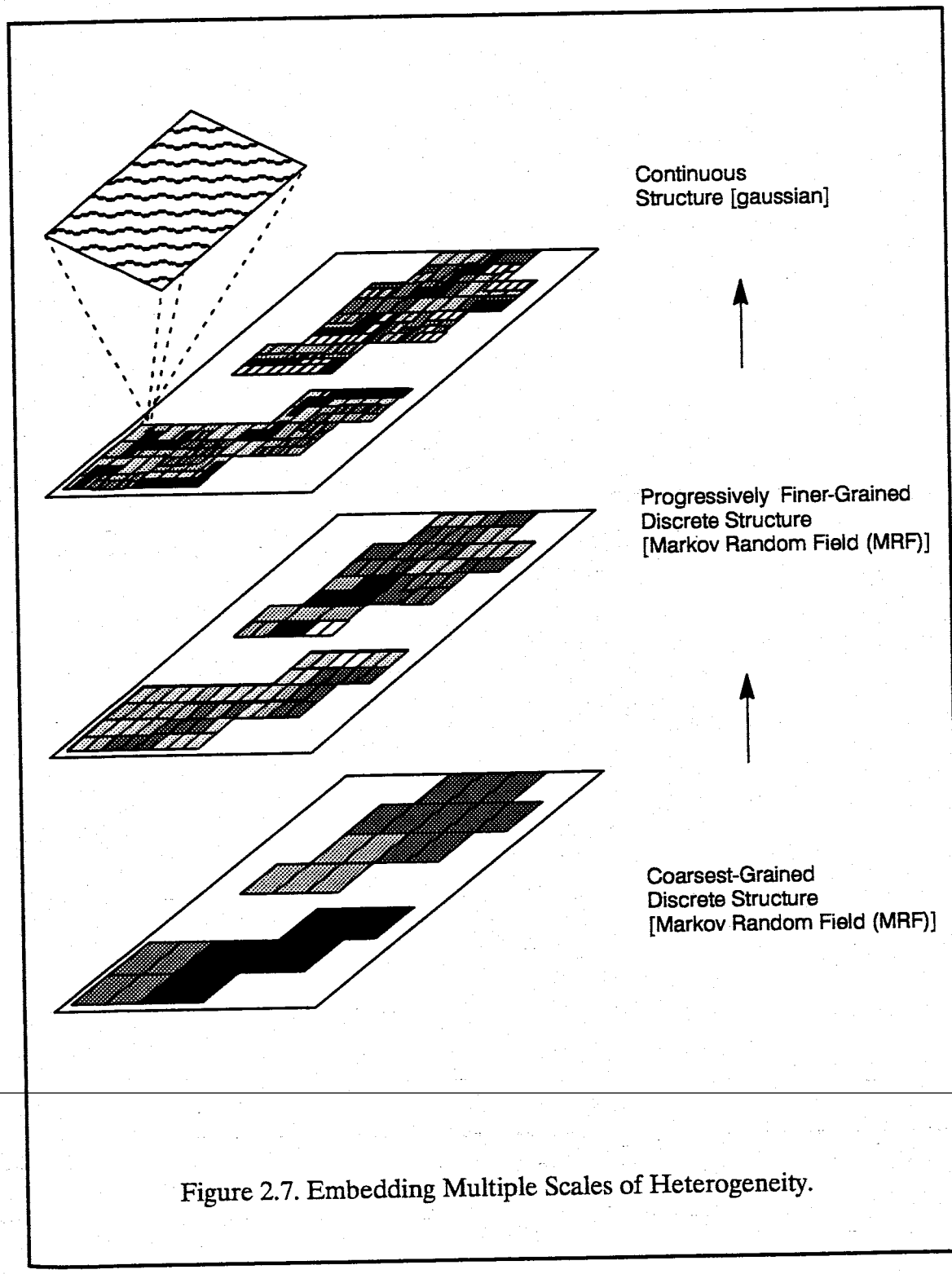


Figure 2.7. Embedding Multiple Scales of Heterogeneity.

One virtue of such an embedding approach stems from its nonanalytic nature, which allow dependencies between scales to be introduced arbitrarily. In addition, unlike analytically-based fractal models for which it is assumed that the same geologic process is acting simultaneously over many scales, an embedding approach can be used to characterize different processes, and different structure, at each distinct scale in the geologic hierarchy. For example, large-scale discrete MRF structure can describe the effects of abrupt, allocyclic events such as changes in climatic regime or tectonic uplift, while small-scale continuous structure within each feature of the large-scale process might help characterize more gradual autocyclic events such as stream avulsion. Autocyclic events such as debris flows, or any autocyclic event involving some type of nonconformal erosional or depositional process, would be best described using models based on spatially-discrete variation.

To demonstrate the impact of multiple scales of variation on prediction and prediction uncertainty without building an unduly complicated model of heterogeneity, this dissertation research has been restricted to two-scale random processes, comprised of smaller-scale continuous gaussian structure nested within a larger-scale, spatially discrete MRF process. Extension to three or more scales of spatial variation is fairly straightforward, but involves a geometric increase in computational burden to both generate hierarchical structure and to identify optimal scales of measurement.

2.5.1 Spatially-Continuous Versus Spatially-Discrete Geostatistical Models

Both spatially continuous and discrete models of random processes have been used to synthesize geologic phenomena. While continuous models can be quite useful in characterizing the structure of geologic phenomena at certain scales of variation, discrete models are uniquely suited to the problem of describing aquifer characteristics in geologic materials that have been deposited, eroded, or altered under abruptly changing geologic conditions. Only discrete models can reproduce the sharp lithologic interfaces typically observed in depositional environments at certain scales of observation.

The majority of geostatistical techniques developed during the past several decades rely on continuum-based gaussian models. Gaussian methods of random field generation include matrix decomposition techniques, spectral methods such as the fast Fourier transform algorithm, the turning bands method, and autoregressive techniques (Farmer, 1987). Common to all of these methods is use of the variogram $\gamma(\bar{\xi})$ or a similar bivariate statistic such as the covariance function $C(\bar{\xi})$. Either of these bivariate statistics are used to operate continuously on an arbitrary, spatially-continuous, uncorrelated process in either the real or frequency domain, to produce a spatially continuous process with the desired correlation scale λ. However, since the variogram and covariance functions are bivariate statistics, they can account only for second moment behavior and very restricted relationships between second- and high-order moments. Alternatively, methods of generating correlated and cross-correlated random fields by estimating the power spectrum directly from the data, without

intermediate estimation of the covariance function, have been developed by Gutjahr (1989) and Robin *et al.* (1993).

The failure of these traditional low-order linear or gaussian stochastic models has been documented extensively in the literature, most notably by Journel and Alabert (1988; 1989) and Deutsch and Journel (1993). Inadequacies of linear geostatistics can be at least partially attributed to the occurrence of significant third- and higher-order odd moments present in observed geologic structure. Significant higher-order moments in nongaussian processes can be manifested as interconnected high- or low-conductivity migration pathways. Since gaussian models can only describe even orders of correlation structure up to order two, they may need to be heavily conditioned on available data if the observed random process contains nongaussian features. Unfortunately, the amount of data required to fully incorporate third- and higher-order odd moment behavior is usually prohibitive.

2.5.2 Spatially-Discrete MRFs

Unlike gaussian methods of synthesis, a realization of an MRF process is generated by explicitly accounting for the joint probability density function (pdf) of the random process and not just its bivariate moments. Because the joint pdf reflects all moments of the process and not just the univariate and bivariate moments, realizations generated on the basis on the joint pdf are not limited to those with negligible higher-order odd moments. Methods of simulating realizations of random processes from high-dimensional joint pdfs proposed by Hastings (1970) include:

- Factorization of the joint pdf p(**Z**) into the product of one-dimensional conditional distributions
- Simulation from a joint pdf q(**Z**) similar to the true pdf p(**Z**) but easier to sample from, with moments computed using the ratio of true to approximate probabilities
- Simulation from the distribution of some transformed version of **Z** using the pdf of the transformed variable **Y** , h(**Y**), to obtain the sample realization **Z** as a function of **Y**.

Realizations of discrete Markov random field models are generated according to the first method, by factoring the joint pdf into a set of conditional distributions. These distributions are then used to locally sample conditional structure that ultimately evolves into a global joint structure characterized by the correct joint moments. The theory behind MRF models and the practical considerations of generating MRF realizations are discussed further in Chapter 3.

2.6 EXAMPLES OF SINGLE-SCALE AND TWO-SCALE CONDUCTIVITY PROCESSES

Figure 2.8 is an example of a two-dimensional, two-scale random conductivity realization, as seen in map view and loosely based on the Sierra Ladrones Formation study conducted by Davis (1994). At the larger scale, where structure is assumed dominated by spatially discrete channel scours that trend along the axis of deposition in a N45°W direction, a discrete second-order MRF model was used to generate the 16-by-16 large-scale component of a two-scale log conductivity process. The length and width of each large-scale block was equal to 16.0 m, producing a transport domain of size 256m-by-256m. Large-scale ln K random variables were allowed to assume only

Figure 2.8. Unconditioned Two-Scale and Statistically-Equivalent Single-Scale Conductivity Realizations.

five distinct values: -7.0, -6.0, -5.0, -4.0, and -3.0 in units of ln m/s, characteristic of conductivities ranging from silty sand to gravel (Freeze and Cherry, 1979). The initial random variables of the uncorrelated large-scale MRF base process were generated from a uniform, univariate probability distribution. Parameters for the second-order Markov process were 1.05 for the first-order neighbors in both the vertical and horizontal directions, and 0.20 and -0.90 for the second-order neighbors located in second and fourth quadrants and in first and third quadrants, respectively. A pair-potential interaction energy was assumed. 5000 potential exchanges per large-scale conductivity site were performed using the Metropolis algorithm. MRF parameters and neighborhood orders, as well as the Metropolis importance-sampling exchange algorithm used to generate MRF realizations, are discussed in greater detail in Chapter 3.

At the smaller scale, 16-by-16 gaussian realizations were embedded within each of the large-scale MRF blocks, with the length and width of each small-scale increment equal to 1.0 m. It was assumed that univariate and bivariate statistics of the small-scale gaussian structure, including mean, variance, and correlation scale, were dependent on the large-scale MRF random variable. In other words, the small-scale moments were conditioned on the large-scale random variables. These moments are listed in Table 2.1. It was also assumed that an exponential covariance structure best described small-scale spatial variation in ln K, following the work of Bakr (1976), Hoeksema and Kitanidis (1985), Sudicky (1986), and Rehfeldt and Gelhar (1992).

Table 2.1. Small-Scale Statistical Moments for Two-Scale Ln K Realizations.

Large-Scale Random Variable ($\ln m\ s^{-1}$)	Small-Scale Ln K Variance $(\ln m\ s^{-1})^2$	λ_x(m)	λ_y(m)
-7.0	2.0	8.0	2.6
-6.0	2.0	8.0	4.0
-5.0	2.0	6.0	4.2
-4.0	2.0	4.0	2.0
-3.0	2.0	4.0	4.0

In keeping with the oblique structure of the large-scale MRF conductivity process, small-scale gaussian features were generated oblique to the x- and y-axes. This required that the principal directions of statistical anisotropy be rotated to N45°W and N45°E when generating small-scale gaussian structure. This rotation was performed in the spectral domain as presented below (Gutjahr, 1995, personal communication).

The spectral density is defined as the Fourier transform of the covariance function:

$$S(\bar{k}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\bar{k}^T\bar{\xi}}\ C(\bar{\xi})\ d\bar{\xi} \qquad [2.11]$$

where $\bar{k}$ is the wave-number vector, $\bar{\xi}$ is the separation-distance vector, and the partial integration is implied over all spatial dimensions.

Let $C_R(\bar{\xi}')$ denote the covariance function in a new coordinate system $\bar{\xi}'$ obtained by rotating the old coordinates through arbitrary angle α. Then the spectral density given in rotated coordinates $\bar{\xi}'$ is:

$$S_R(\bar{k}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\bar{k}^T\bar{\xi}'} \, C_R(\bar{\xi}') \, d\bar{\xi}' \qquad [2.12]$$

Figure 2.9 illustrates how the unrotated and rotated coordinate systems are related for the two-dimensional problem. To express the spectral density as a function of the

$$\xi_1 = \xi_1' \cos\alpha - \xi_2' \sin\alpha$$
$$\xi_2 = \xi_1' \sin\alpha + \xi_2' \cos\alpha$$

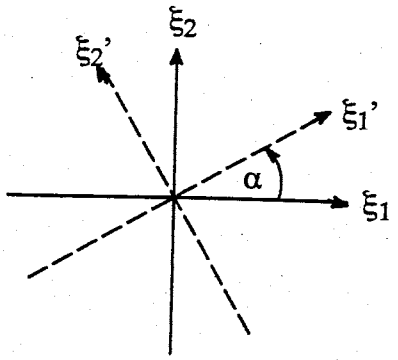


Figure 2.9. Rotation of Coordinate System.

non-rotated coordinates vector $\underline{\xi}$, $\underline{\xi}'$ and $\underline{\xi}$ are related by the coordinate transformation:

$$\underline{\xi} = \mathbf{P}\,\underline{\xi}' \qquad [2.13]$$

where **P** is the orthogonal rotation matrix

$$\mathbf{P} = \begin{bmatrix} \cos\alpha & -\sin\alpha \\ \sin\alpha & \cos\alpha \end{bmatrix}. \qquad [2.14]$$

Multiplying equation [2.13] by the inverse of **P**,

$$\underline{\xi}' = \mathbf{P}^{-1}\,\underline{\xi}. \qquad [2.15]$$

For orthogonal matrix **P**, $\mathbf{P}^{-1} = \mathbf{P}^{T}$, and this yields:

$$\underline{\xi}' = \mathbf{P}^{T}\,\underline{\xi} \qquad [2.16]$$

From equation [2.15], the rotated covariance function can thus be expressed as:

$$C_R(\underline{\xi}') = C(\underline{\xi}) = C(\mathbf{P}\underline{\xi}') \qquad [2.17]$$

Since the matrix **P** imposes only a rotation with no deformation, $|\mathbf{P}| = 1$ and $d\bar{\xi}' = d\bar{\xi}$. Substituting for the transformed vector $\bar{\xi}' = \mathbf{P}^T \bar{\xi}$, the rotated covariance $C_R(\bar{\xi}') = C(\mathbf{P}\bar{\xi}')$, and $d\bar{\xi}' = d\bar{\xi}$ in equation [2.12] yields:

$$S_R(\bar{k}) = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\,\bar{k}^T\bar{\xi}'}\; C_R(\bar{\xi}')\; d\bar{\xi}' = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\,\bar{k}^T\bar{\xi}'}\; C(\mathbf{P}\bar{\xi}')\; d\bar{\xi}'$$

$$= \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\,\bar{k}^T\mathbf{P}^T\bar{\xi}}\; C(\bar{\xi})\; d\bar{\xi} = \frac{1}{2\pi} \int_{-\infty}^{\infty} e^{-i\,(\mathbf{P}\bar{k})^T\bar{\xi}}\; C(\bar{\xi})\; d\bar{\xi} = S(\mathbf{P}\bar{k})$$

[2.18]

Thus, the 'rotated' spectrum, or the spectrum corresponding to a rotated covariance function, is simply the 'unrotated' spectrum evaluated at rotated wave number vector $\mathbf{P}\,\bar{k}$ (Mantaglou and Wilson, 1982). In other words, if $C_R(\bar{\xi}')$ is the covariance function rotated with respect to coordinate system $\bar{\xi}$, then $S(\mathbf{P}\bar{k})$ is the spectral density associated with the rotated covariance function.

Using the simple transform given by equation [2.18], oblique gaussian structure in each of the 16-by-16 large-scale MRF blocks was generated via the fast-Fourier transform (FFT) algorithm developed by Gutjahr (1989). For a second-order stationary, zero-mean process, the FFT algorithm correlates independent random

increments in Fourier space using the Fourier-Steiltjes representation:

$$f'(\mathbf{x}) = \int_{-\infty}^{\infty} e^{i\bar{k}^T\bar{x}}\, dZ_f(\bar{k}) \qquad [2.19]$$

where $f'(\mathbf{x})$ is the perturbation about the mean, $dZ_f(\bar{k})$ is the associated complex Fourier amplitude of fluctuations about zero mean as a function of wave number vector $\bar{k}=(k_1, k_2, k_3)$, and $\mathbf{x}$ is the vector of spatial coordinates. Following generation of each small-scale 16-by-16 unconditional gaussian realization, conditioning of small-scale realizations was performed along block interfaces whenever adjacent large-scale random variables were the same, allowing spatial continuity to be imposed within each channel scour. Since embedding of small-scale gaussian structure in the large-scale MRF realization proceeded from south to north along each column of the large-scale realization, conditioning on previously-generated small-scale structure was performed along the western and southern interfaces. Alternatively, the shape of large-scale channel scours could have been used to extract continuous gaussian realizations from realizations sized to the largest dimension of the channel scour. This, however, requires more disk space than interface-conditioning, an approach which essentially trades CPU time for computer storage requirements. The resulting 256-by-256 ln K embedded realization presented in Figure 2.8 is referred to as a two-scale realization because it possesses elements of two different scales. Since it was not conditioned on real data, it is also referred to as an unconditional realization.

Figure 2.8 also shows a 256-by-256 ln K gaussian realization that is statistically equivalent to the two-scale realization, where statistical equivalence was defined up to order two. The gaussian realization has, by definition, a single scale of variation given by its correlation scale. In order to synthesize equivalent single-scale gaussian structure that was statistically equivalent to the two-scale process, it was necessary to first determine principal directions of anisotropy and correlation scales associated with an ensemble of the two-scale ln K process. These principal directions and scales, along with means and variances, were then used to specify spatial variation in the ensemble of single-scale, gaussian, second-order equivalent realizations. A covariance model was fit to the two-scale ln K process by estimating covariances in the principal directions using the relation:

$$C(\bar{\xi}) = \frac{1}{N_\xi} \sum_{i=1}^{N_\xi} \left[Y(\mathbf{x}_i) - \mu_Y\right]\left[Y(\mathbf{x}_i + \bar{\xi}) - \mu_Y\right] \qquad [2.20]$$

where $Y(\mathbf{x}_i)$ is the log conductivity value at vector location $\mathbf{x}_i$, N_ξ is the number of pairs of Y values separated by distance vector $\bar{\xi}$ and μ_Y is the mean of $Y(\bar{x})$. Rewriting $C(\bar{\xi})$ in equation [2.20] as:

$$C(\bar{\xi}) = \frac{1}{N_\xi} \sum_{i=1}^{N_\xi} Y(\mathbf{x}_i)\, Y(\mathbf{x}_i + \bar{\xi}) - \mu_Y^2 \qquad [2.21]$$

$C(\bar{\xi})$ can be viewed as a spatial convolution of mean-removed values of $Y(\mathbf{x})$, and is equivalent to multiplying the Fourier transform of mean-removed $Y(\mathbf{x})$ and its complex conjugate in the spectral domain. Following subtraction of the μ_Y from the two-scale process, the Fourier transform of the resulting zero-mean $Y(\mathbf{x})$ process becomes:

$$F(\bar{\omega}) = \int_{-\infty}^{\infty} Y(\mathbf{x})\, e^{-i\bar{\omega}\mathbf{x}}\, d\mathbf{x} \qquad [2.22]$$

where $\bar{x}$ and $\bar{\omega}$ are space and frequency vectors, respectively, and integration is implied over two-dimensional space. Multiplication of $F(\bar{\omega})$ by its complex conjugate $F^*(\bar{\omega})$ yields the real-valued energy spectrum as a function of frequency:

$$S(\bar{\omega}) = F(\bar{\omega}) \cdot F^*(\bar{\omega}) \qquad [2.23]$$

Finally, the covariance is obtained by finding the inverse Fourier transform of $S(\bar{\omega})$:

$$C(\bar{\xi}) = \int_{-\infty}^{\infty} S(\bar{\omega})\, e^{i\bar{\omega}\bar{\xi}}\, d\bar{\omega} \qquad [2.24]$$

The ensemble covariance function for the two-scale ensemble, presented in Figure 2.10a, was estimated by summing $C(\bar{\xi})$ obtained in this manner over 100 two-scale realizations, and then dividing by 100.

Based on the covariance estimate, the variance of the two-scale ln K process was determined to be equal to 3.7 $(\ln m/s)^2$, including a nugget variance of roughly 0.3. It was also determined that the principal directions of statistical anisotropy were oriented along transects in the $N45^0W$ and $N45^0E$ directions, with principal correlation scales of 18.0 m and 10.0 m, respectively, estimated on the basis of an exponential covariance model from transects through the two-dimensional covariance function shown in Figure 2.10b. Statistical equivalence was imposed on the single-scale process by forcing each realization in the single-scale ensemble to reproduce the global mean of -5.0 ln m/s, as well as both structured and unstructured (nugget) variances and correlation scales of the two-scale ln K process. Nugget variance was incorporated into each single-scale gaussian realization following generation by the FFT algorithm by adding independent normally-distributed error characterized by a mean of 0.0 and a variance of 0.3 to each random variable in the realization.

Note that, while the two-scale realization is only locally stationary within a given channel scour, the equivalent single-scale gaussian realization is globally stationary. In other words, one could visit any neighborhood in the single-scale realization and determine that the local statistics are roughly the same and equal to the global statistics. However, local statistics in the two-scale realization depend on the channel scour and will be stationary only within the confines of each scour, a property perhaps better referred to as quasi-stationarity.

It should also be noted that, while quasi-stationarity of the two-scale process translates easily into the well-defined covariance function presented in Figure 2.10a,

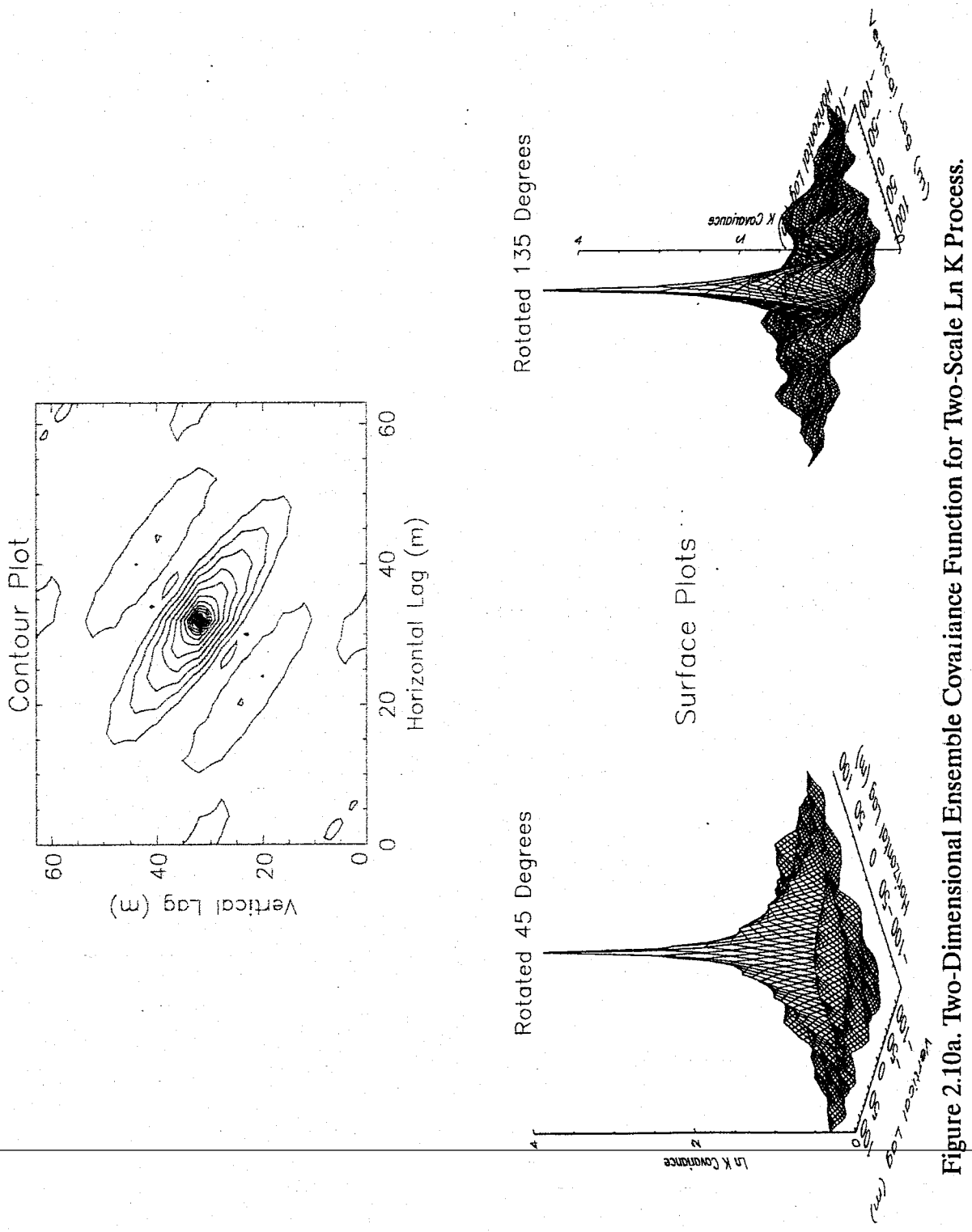


Figure 2.10a. Two-Dimensional Ensemble Covariance Function for Two-Scale Ln K Process.

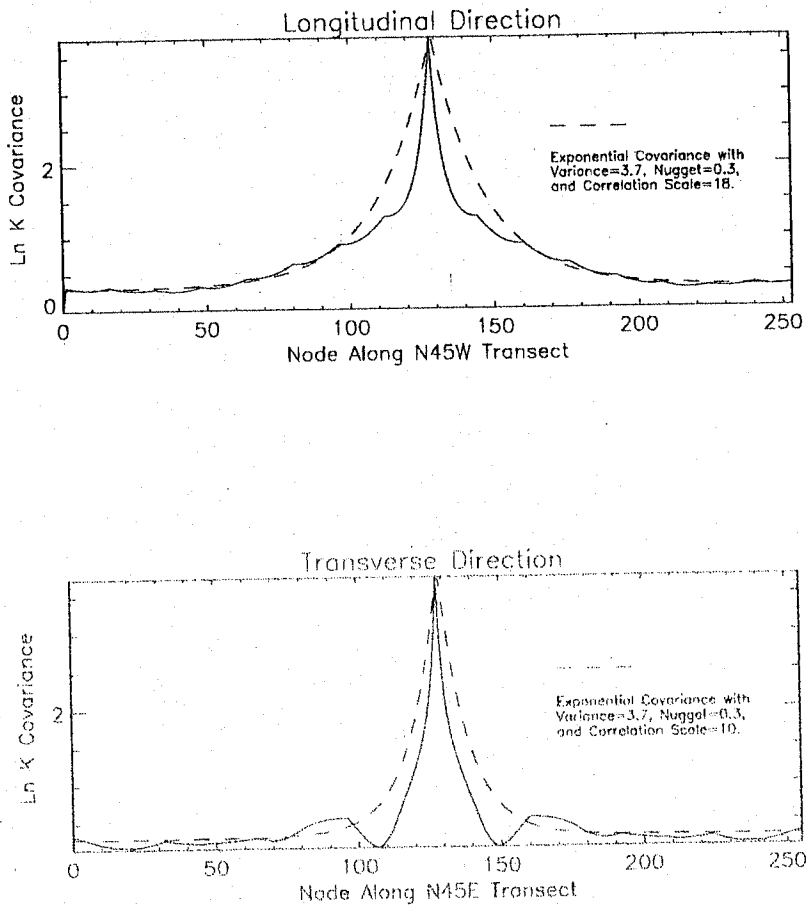


Figure 2.10b. Transects of Two-Dimensional Ensemble Covariance Function for Two-Scale Ln K Process in Longitudinal and Transverse Directions.

this covariance function can only describe the global average of bivariate behavior of the process. Local deviations from this behavior, as well as high-order moments not characterized by $C(\xi)$, can produce very different realizations from those associated with stationary single-scale gaussian behavior, as illustrated by the very disparate single-scale and two-scale realizations presented in Figure 2.8.

Of interest is the question of how ensembles of two-scale and equivalent single-scale realizations might respond differently to the same hydraulic and contaminant stresses. Specifically, how would estimates of effective conductivity and macrodispersivity differ between ensembles of both types of stochastic ln K processes?

2.7 DETERMINATION OF TWO-SCALE AND SINGLE-SCALE ENSEMBLE EFFECTIVE CONDUCTIVITY

To address this issue, 100 realizations each of both ln K processes were subjected to well-defined flow and transport problems. From inspection of equation [2.10], it is evident that the problem of defining effective conductivity reduces to prediction of average flux and gradient through an ensemble of conductivity realizations. A question then arises: what boundary conditions should be imposed along block or domain edges so that small-scale conductivity variations are appropriately sampled and upscaled? For the case of K_{eff}, the choice of boundary conditions is irrelevant in a statistically homogeneous ergodic realization, because all heterogeneities will be thoroughly sampled by the migrating fluid elements, independent of flow conditions. This is directly attributable to the fact that many independent measurements of stationary conductivity will be made in an ergodic stationary conductivity process, regardless of

the average flow trajectory. However, the two-scale ln K realizations are not statistically homogeneous and may not even be ergodic. As a consequence, K_{eff} will generally be dependent on the nature of boundary conditions used to impose flow through a two-scale realization.

Much research has been directed to overcoming this dependency of effective parameters on the choice of boundary conditions (Sanchez-Vila *et al.*, 1995). Freeze and Cherry (1979) derived equivalent conductivity for layered systems in which flow is either parallel or perpendicular to the layers. In such perfectly stratified systems, flow parallel to the layers gives rise to an effective conductivity equal to a weighted arithmetic mean, while flow perpendicular to layering produces a weighted harmonic mean. Warren and Price (1961) demonstrated that, when small-scale conductivity does not possess spatially anisotropic structure, effective conductivity is equal to the geometric mean regardless of imposed boundary conditions. Such generalization of boundary conditions is based on the unbiased nature of conductivity measurement that occurs when there is no distinct structure associated with conductivity and fluid elements are free to roam in a random manner, obtaining many independent measurements of conductivity without any particular geometry constraining the manner of sampling. White and Horne (1987) proposed a general model for equivalent conductivity by imposing a number of different boundary conditions on the block and averaging the equivalent conductivity values. While this general boundary condition model does not depend on any one set of boundary conditions and does not thus require an *a priori* assumption of the flow field, it produces physically meaningless

nonsymmetric tensors. Homogenization theory (Bourgeat, 1984) and Taylor dispersion theory (Kitanidis, 1990) have also been applied to the problem of developing a model that can account for general boundary conditions. Rubin and Gomez-Hernandez (1990) defined equivalent block conductivity to be the ratio of average local flow and average gradient, and Gomez-Hernandez (1991) addressed the problem of equivalent conductivity when average local flux is not parallel to the true head gradient. Neuman and Orr (1993) investigated the existence of effective conductivity by explicitly incorporating boundary conditions.

Saez *et al.* (1989) and Durlofsky (1991) proposed, in effect, an elimination of boundary effects by imposing periodic boundary conditions along block interfaces. The resulting equations are similar to those derived on the basis of homogenization or diffusion theory (Bourgeat, 1984; Kitanidis, 1990). Periodic boundary conditions serve to overcome the dependency of equivalent conductivity on flow conditions because they do not specify absolute head or flux along any of the boundaries. Rather, head and flux relations along opposing boundaries are stipulated. It is the very generality of such relative conditions that makes the approach so well-suited to defining equivalent parameters, because it presupposes no absolute flow geometry that can preclude unbiased sampling of conductivity.

Figure 2.11 illustrates how periodic boundary conditions are imposed on a flow domain. Two separate steady-state flow problems of the form $\nabla_{x,y}[K\nabla_{x,y}h]= 0$ are solved, where K is spatially-distributed conductivity. The first problem involves imposition of an arbitrary gradient in, say, the x direction:

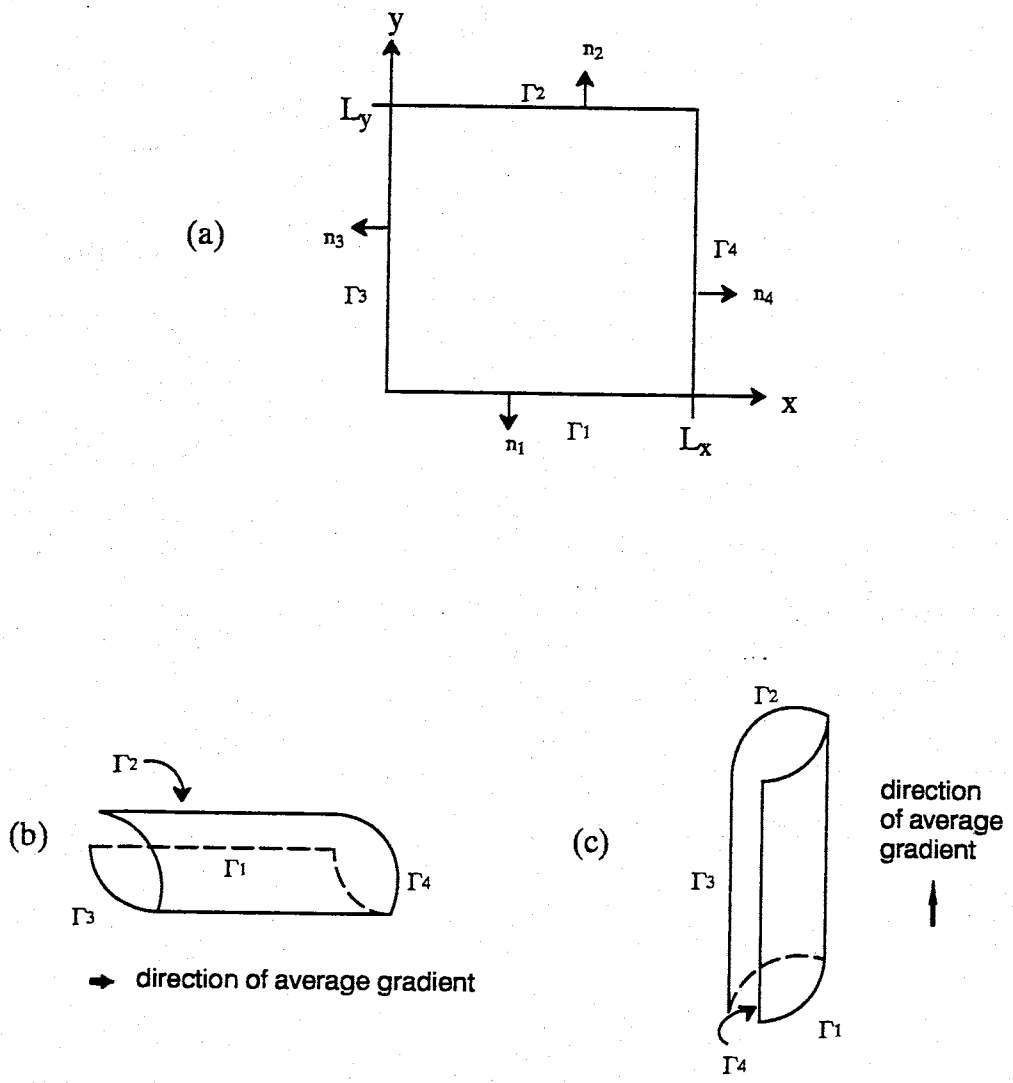


Figure 2.11. Imposition of Periodic Boundary Conditions.

$$h(x = 0, y) = h(x = L_x,\ y) - J_x * L_x \qquad \text{on } \Gamma_3 \text{ and } \Gamma_4 \qquad [2.25]$$

In addition to the gradient condition, fluxes along opposing boundaries Γ_3 and Γ_4 are required to be equal and opposite in sign:

$$q(x = 0,\ y) \cdot n_3 = -\, q(x = L_x,\ y) \cdot n_4 \qquad \text{on } \Gamma_3 \text{ and } \Gamma_4 \qquad [2.26]$$

where n_3 and n_4 are outward unit normal vectors along Γ_3 and Γ_4. Along boundaries Γ_1 and Γ_2, heads are forced to be equal according to the condition:

$$h(x,\ y = 0) = h(x,\ y = L_y) \qquad \text{on } \Gamma_1 \text{ and } \Gamma_2 \qquad [2.27]$$

and fluxes are again assumed equal and of opposite sign:

$$q(x,\ y = 0) \cdot n_1 = -\, q(x,\ y = L_y) \cdot n_2 \qquad \text{on } \Gamma_1 \text{ and } \Gamma_2 \qquad [2.28]$$

Note that, unlike equation [2.25], equation [2.27] imposes true periodicity because it involves no head jump across opposing boundaries.

Equations [2.25], [2.26], [2.27], and [2.28] define relationships between heads and fluxes on opposing boundaries — absolute heads and fluxes are never specified. Since the boundary-value problem $\nabla_{x,y}[K\nabla_{x,y}h]=0$ subject to equations [2.25] through [2.28] has no unique solution, an arbitrary head H is specified somewhere in the interior of the domain at $x=x_0$ and $y=y_0$:

$$h\ (x=x_0,\ y=y_0) = H \qquad [2.29]$$

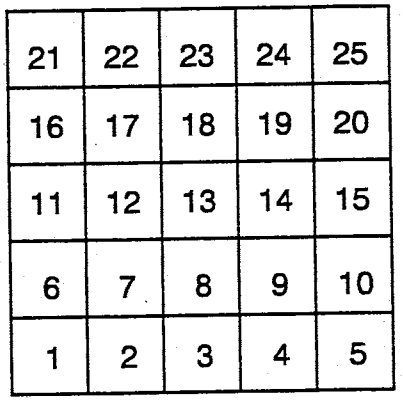
resulting in a well-posed flow problem.

Figure 2.11a illustrates how, in the x-direction parallel to average flow, the effect of the boundary conditions along Γ_1 and Γ_2 is the same as if the boundaries were not present at all. In effect, the two-dimensional domain is wrapped back onto itself along Γ_1 and Γ_2 to create a flow cylinder. In the absence of gradient J_x, the cylinder would also be wrapped back onto itself along Γ_3 and Γ_4 to create a torus topology.

Table 2.2 lists the equations that result at each of 25 nodes in the 5-by-5 example shown in Figure 2.12 when the periodic boundary conditions are imposed on the edges of a finite difference grid for average flow in the x-direction. The resulting matrix structure for the 5-by-5 example is presented in Figure 2.13. Due to the nonsymmetric nature of the matrix, a GMRES solver developed at Los Alamos National Laboratory and based on the algorithm of Saad and Schultz (1986) was used to solve the periodic flow problem.

Table 2.2. Node Equations in 5-By-5 Periodic Boundary Condition Example with Head Drop = Δh in the X-Direction and $\Delta x = \Delta y$. ($K_{i,j}$ = interface mean between blocks i and j)

Node	Equation
1	$h_1 = h_5 + \Delta h$
2	$h_2 = h_{22}$
3	$h_3 = h_{23}$
4	$h_4 = h_{24}$
5	$h_5 = h_{25}$
6	$h_6 = h_{10} + \Delta h$
7	$\overline{K}_{7,2}(h_7 - h_2) + \overline{K}_{7,6}(h_7 - h_6) + \overline{K}_{7,8}(h_7 - h_8) + \overline{K}_{7,12}(h_7 - h_{12}) = 0$
8	$\overline{K}_{8,3}(h_8 - h_3) + \overline{K}_{8,7}(h_8 - h_7) + \overline{K}_{8,9}(h_8 - h_9) + \overline{K}_{8,13}(h_8 - h_{13}) = 0$
9	$\overline{K}_{9,4}(h_9 - h_4) + \overline{K}_{9,8}(h_9 - h_8) + \overline{K}_{9,10}(h_9 - h_{10}) + \overline{K}_{9,14}(h_9 - h_{14}) = 0$
10	$\overline{K}_{9,10}(h_9 - h_{10}) = \overline{K}_{6,7}(h_6 - h_7)$
11	$h_{11} = h_{15} + \Delta h$
12	$\overline{K}_{12,7}(h_{12} - h_7) + \overline{K}_{12,11}(h_{12} - h_{11}) + \overline{K}_{12,13}(h_{12} - h_{13}) + \overline{K}_{12,17}(h_{12} - h_{17}) = 0$
13	$\overline{K}_{13,8}(h_{13} - h_8) + \overline{K}_{13,12}(h_{13} - h_{12}) + \overline{K}_{13,14}(h_{13} - h_{14}) + \overline{K}_{13,18}(h_{13} - h_{18}) = 0$
14	$\overline{K}_{14,9}(h_{14} - h_9) + \overline{K}_{14,13}(h_{14} - h_{13}) + \overline{K}_{14,15}(h_{14} - h_{15}) + \overline{K}_{14,19}(h_{14} - h_{19}) = 0$
15	$\overline{K}_{14,15}(h_{14} - h_{15}) = \overline{K}_{11,12}(h_{11} - h_{12})$
16	$h_{16} = h_{20} + \Delta h$
17	$\overline{K}_{17,12}(h_{17} - h_{12}) + \overline{K}_{17,16}(h_{17} - h_{16}) + \overline{K}_{17,18}(h_{17} - h_{18}) + \overline{K}_{17,22}(h_{17} - h_{22}) = 0$
18	$\overline{K}_{18,13}(h_{18} - h_{13}) + \overline{K}_{18,17}(h_{18} - h_{17}) + \overline{K}_{18,19}(h_{18} - h_{19}) + \overline{K}_{18,23}(h_{18} - h_{23}) = 0$
19	$\overline{K}_{19,14}(h_{19} - h_{14}) + \overline{K}_{19,18}(h_{19} - h_{18}) + \overline{K}_{19,20}(h_{19} - h_{20}) + \overline{K}_{19,24}(h_{19} - h_{24}) = 0$
20	$\overline{K}_{19,20}(h_{19} - h_{20}) = \overline{K}_{16,17}(h_{16} - h_{17})$
21	$h_{21} = h_{25} + \Delta h$
22	$\overline{K}_{17,22}(h_{17} - h_{22}) = \overline{K}_{2,7}(h_2 - h_7)$
23	$\overline{K}_{18,23}(h_{18} - h_{23}) = \overline{K}_{3,8}(h_3 - h_8)$
24	$\overline{K}_{19,24}(h_{19} - h_{24}) = \overline{K}_{4,9}(h_4 - h_9)$
25	$\overline{K}_{20,25}(h_{20} - h_{25}) = \overline{K}_{5,10}(h_5 - h_{10})$



21	22	23	24	25
16	17	18	19	20
11	12	13	14	15
6	7	8	9	10
1	2	3	4	5

Figure 2.12. Node Numbering for 5-By-5 Periodic Boundary Condition Example.

It should be noted that specification of equal fluxes along opposing boundaries using a finite-difference flow solver required that the fluxes actually be made equivalent just inside the opposing boundaries, because fluxes along the block-centered finite-difference interfaces are unknown. A finite-element formulation, which solves for head along all block interfaces, would be a means of avoiding this problem. However, if the size of the domain is large relative to the size of the finite-difference grid blocks, the error introduced by such an approximation to boundary fluxes can usually be considered negligible.

Following solution of the steady-state flow equation subject to relative boundary conditions [2.25] through [2.29], average fluxes along boundaries Γ_4 and Γ_1 (or, equivalently, along Γ_3 and Γ_2 due to the periodicity of the resulting head solution)

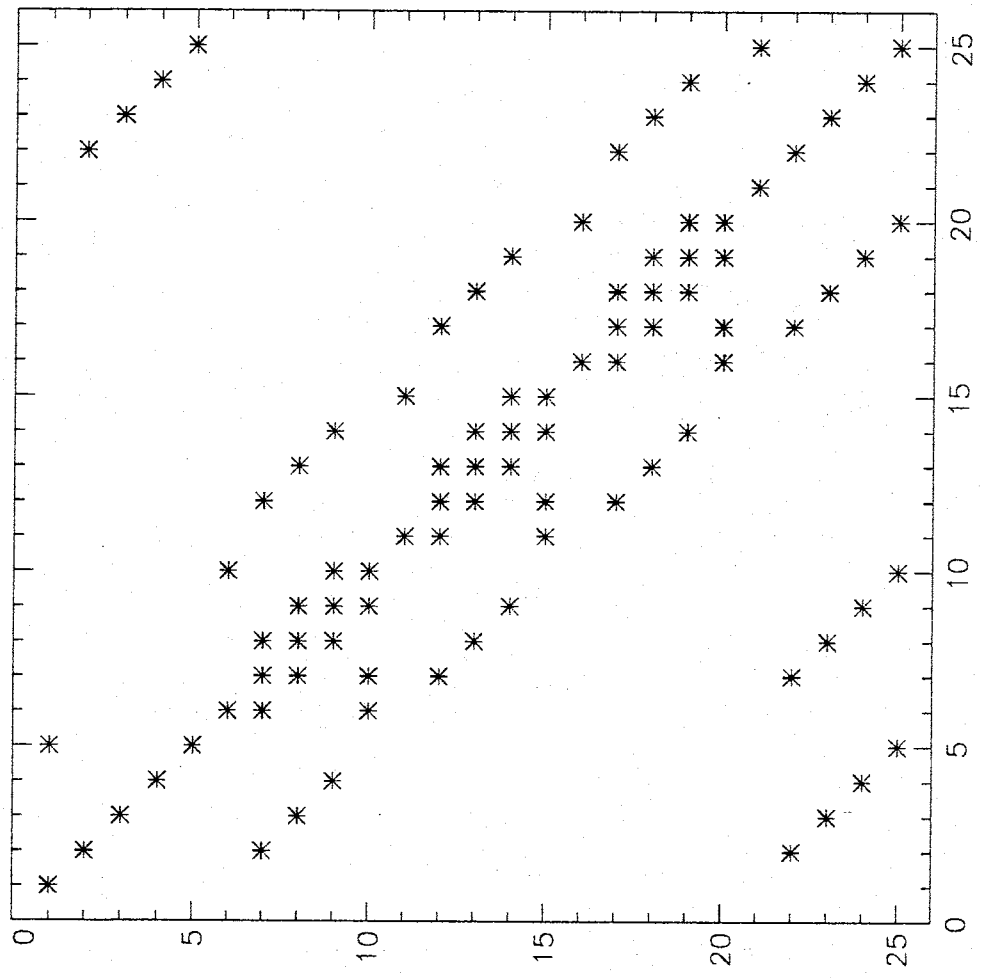


Figure 2.13. Solution Matrix Structure for 5-By-5 Periodic Boundary Condition Problem.

were calculated as $<Q_x>$ and $<Q_y>$. These values yielded two elements of the effective conductivity tensor:

$$K_{eff,xx} = \frac{-<Q_x>}{|J_x|}$$
$$K_{eff,yx} = \frac{-<Q_y>}{|J_x|} \qquad [2.30]$$

To determine the elements $K_{eff,yy}$ and $K_{eff,xy}$, the flow equation was solved again, subject to boundary conditions:

$$h(x,\ y = 0) = h(x\ ,\ y = L_y) - J_y * L_y \qquad \text{on } \Gamma_1 \text{ and } \Gamma_2 \qquad [2.31]$$

$$q(x\ ,\ y = 0\ ) \cdot n_1 = - q(x,\ y = L_y\ ) \cdot n_2 \qquad \text{on } \Gamma_1 \text{ and } \Gamma_2 \qquad [2.32]$$

$$h(x = 0,\ y) = h(x = L_x,\ y) \qquad \text{on } \Gamma_3 \text{ and } \Gamma_4 \qquad [2.33]$$

$$q(x = 0,\ y) \cdot n_3 = - q(x = L_y,\ y) \cdot n_4 \qquad \text{on } \Gamma_3 \text{ and } \Gamma_4 \qquad [2.34]$$

To determine the elements $K_{eff,yy}$ and $K_{eff,xy}$, the flow equation was solved again, subject to boundary conditions:

$$h(x,\ y = 0) = h(x\ ,\ y = L_y) - J_y * L_y \qquad \text{on } \Gamma_1 \text{ and } \Gamma_2 \qquad [2.35]$$

$$q(x\ ,\ y = 0\ ) \cdot n_1 \ = - q(x,\ y = L_y\ ) \cdot n_2 \qquad \text{on } \Gamma_1 \text{ and } \Gamma_2 \qquad [2.36]$$

$$h(x = 0,\ y) = h(x = L_x,\ y) \qquad \text{on } \Gamma_3 \text{ and } \Gamma_4 \qquad [2.37]$$

$$q(x\ = 0,\ y) \cdot n_3 \ = - q(x = L_y,\ y) \cdot n_4 \qquad \text{on } \Gamma_3 \text{ and } \Gamma_4 \qquad [2.38]$$

As shown in Figure 2.11(c), these boundary conditions create a cylinder similar to the one created in Figure 2.11(b), but with Γ_3 and Γ_4 defining the 'seam' of the cylinder. The same finite-difference algorithm used to construct the matrix for the x-direction flow problem was used to construct the matrix for the y-direction flow problem by simply taking the transpose of the conductivity realization. Again, averages fluxes $<Q_x>$ and $<Q_y>$ were determined along any two orthogonal boundaries and the remaining elements of the effective conductivity tensor determined as follows:

$$K_{eff,\,yy} = \frac{-<Q_y>}{|J_y|} \qquad [2.39]$$

$$K_{eff,\,xy} = \frac{-<Q_x>}{|J_y|}$$

Non-zero values for $K_{eff,xy}$ and $K_{eff,yx}$ are responsible for lack of colinearity in streamline and gradient directions except along principal directions of anisotropy. $K_{eff,xy}$ and $K_{eff,yx}$ should be equal to one another to within machine accuracy, producing a symmetric effective conductivity tensor and real eigenvalues. The tensor is also positive definite, denoting flow in a downgradient direction. Both properties are guaranteed by the solution of a well-posed boundary-value problem and the consistency of the boundary conditions (Durlofsky, 1991).

The steady-state head solution for a gradient imposed in the x-direction for a 128-by-128 problem with mean ln K of -5.0 ln m/s, ln K variance of 3.7 $(\ln \text{m/s})^2$, principal directions of statistical anisotropy equal to N45^{0}W and N45^{0}E, and longitudinal and transverse correlation scales of 18.0 m and 10.0 m is presented in Figure 2.14 for an unconditional, two-scale conductivity realization. Based on four adjacent plots of the head solution, one can easily see that the periodic boundary conditions give rise to a single-period head solution in the flow domain, and by extension, a similarly periodic flux solution.

Note that the traditional method of estimating the effective conductivity tensor $\mathbf{K}_{\mathbf{eff}}$ by imposing a gradient in one direction and no-flux conditions in the other direction, as illustrated in Figure 2.15, does not allow determination of off-diagonal tensor elements. No-flux boundary conditions, by definition, prevent the movement of water across the boundaries. For the case of flow imposed in the x-direction, such an assumption is tantamount to assuming that a gradient in the x-direction does not deflect fluid in the y-direction across the Γ_1 and Γ_2 boundaries. In general, this will be true only

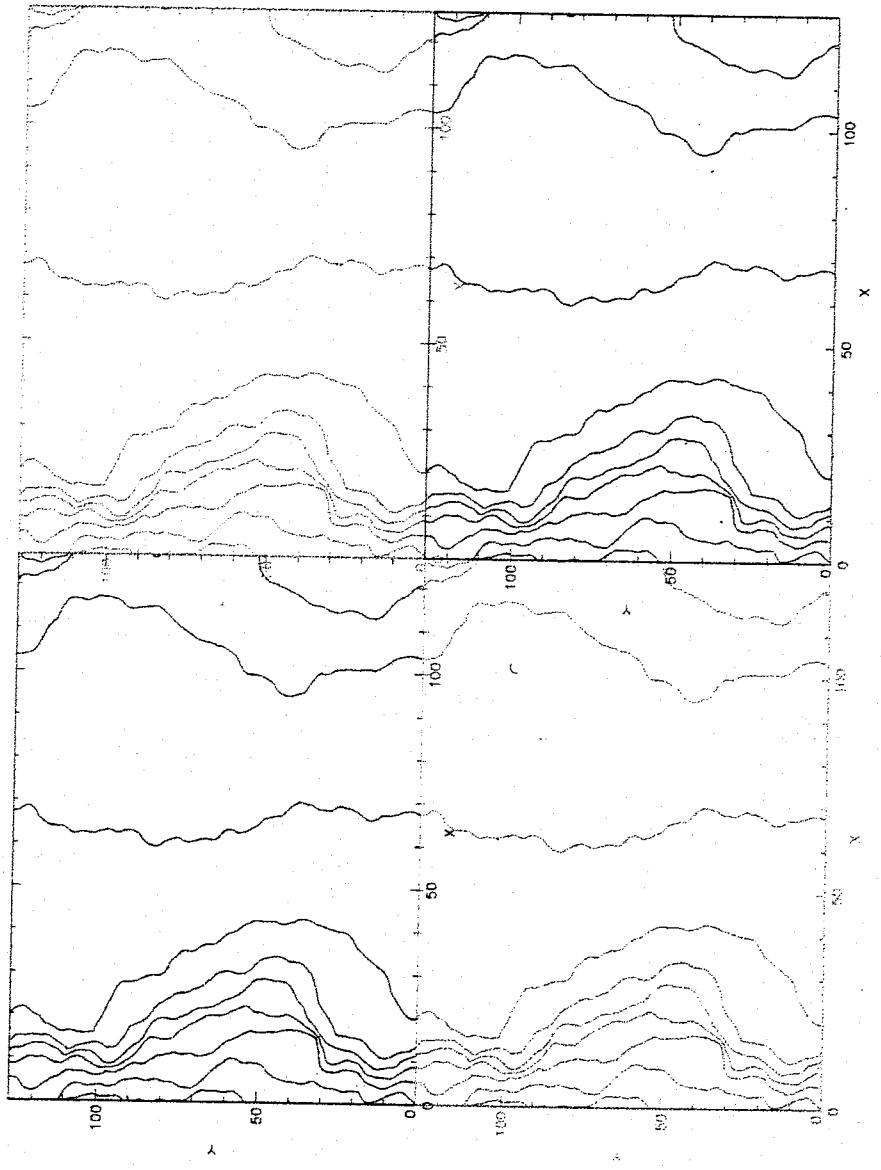


Figure 2.14 Four Adjacent Plots of Hydraulic Head Resulting from Application of Periodic Boundary Conditions.

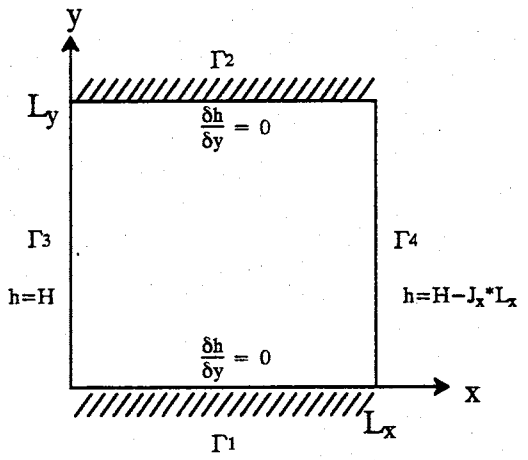


Figure 2.15. Traditional Boundary Conditions Used for Estimating Effective Conductivity.

for a medium which is either physically homogeneous or in which the principal directions are, in fact, perfectly oriented along the x- and y-axes. For the vast majority of heterogeneous media, the off diagonals of $\mathbf{K_{eff}}$ will be non-zero valued and the flow geometry illustrated in Figure 2.15 cannot be expected to yield correct values of effective conductivity. It should also be noted that the danger of using the traditional approach to defining $\mathbf{K_{eff}}$ is not just that the off-diagonal terms will be calculated incorrectly, but that the diagonal terms will also generally be wrong. For example, if the gradient is in the x-direction, imposition of no-flux conditions along Γ_1 and Γ_2 will funnel fluid through the domain from Γ_3 to Γ_4, producing an overestimate of flux in the x-direction and too large a value of K_{xx}. Use of periodic boundary conditions

eliminates this restriction, because fluid is free to migrate across Γ_1 and Γ_2 when flow is deflected by heterogeneities not aligned in the x-direction, and across Γ_3 and Γ_4 when flow is deflected by heterogeneities not aligned in the y-direction. However, the periodic boundary condition method may not perform well when within-block head variations are highly nonlinear (Durlofsky, 1991).

It should also be emphasized that neither the traditional nor the periodic boundary condition approach will yield correct elements of the effective conductivity tensor when the conductivity realization is not ergodic, or statistically typical, of the true conductivity process. When preferential pathways occur or the size of the dominant scale is on the same order or smaller than the averaging domain, averaging will not produce a valid effective conductivity. The averaging volume or area should always be sufficiently large to include a representative sample of the conductivity process at the particular scale of interest.

Diagonalized ensemble effective conductivity tensors, obtained as a result of applying periodic boundary conditions to all 100 realizations comprising both the single-scale and two-scale ln K processes, are presented in Figure 2.16. It is not surprising that ln $\mathbf{K_{eff}}$ estimates for the small-scale and two-scale cases were similar, considering that we required the global properties of the two ln K processes to be the same. Note that the standard errors of estimate were somewhat larger in the two-scale case, where small changes in large-scale structure could produce very different ln K realizations and where flow estimates were subject to greater error as a result of more pronounced nonperiodicities in ln K. Effective longitudinal and transverse

Single-Scale

$$\begin{bmatrix} -2.8 \pm 0.11 & 0.0 \\ 0.0 & -4.2 \pm 0.10 \end{bmatrix}$$

Two-Scale

$$\begin{bmatrix} -2.7 \pm 0.12 & 0.0 \\ 0.0 & -3.7 \pm 0.12 \end{bmatrix}$$

Figure 2.16. Diagonalized Ensemble Log Effective Conductivity Tensors for Unconditioned Conductivity Processes (ln m/s) (Including Standard Errors of Estimate).

conductivities for both cases ranged from one to two orders of magnitude larger than the theoretical approximation derived by Gelhar and Axness (1983, eq. 52) for two-dimensional flow in a perfectly-stratified medium parallel to bedding:

$$\ln K_{eff,l} = \ln\left[K_G \left[\left(1 + \frac{\sigma^2_{\ln K}}{2} \right) - \sigma^2_{\ln K}\left(\frac{1}{1 + \lambda_l/\lambda_t} \right) \right] \right] = \ln\left[e^{-5} \left[\left(1 + \frac{3.7}{2} \right) - 3.7\left(\frac{1}{1 + 1.8} \right) \right] \right] = -4.6 \text{ ln m/s}$$

$$\ln K_{eff,t} = \ln\left[K_G \left[\left(1 + \frac{\sigma^2_{\ln K}}{2} \right) - \sigma^2_{\ln K}\left(\frac{\lambda_l/\lambda_t}{1 + \lambda_l/\lambda_t} \right) \right] \right] = \ln\left[e^{-5} \left[\left(1 + \frac{3.7}{2} \right) - 3.7\left(\frac{1.8}{1 + 1.8} \right) \right] \right] = -5.8 \text{ ln m/s}$$

Overestimation of $K_{eff,l}$ and $K_{eff,t}$ was attributed to the tendency for effective conductivity in an imperfectly stratified medium both perpendicular and parallel to the direction of bedding to be larger than for a perfectly stratified medium because fluid in an imperfectly stratified medium is freer to move around low conductivity barriers in either direction.

2.8 DERIVATION OF ENSEMBLE MACRODISPERSIVITY TENSORS

To simulate transport of contaminant through each realization and estimate macrodispersivity, steady-state fluid flow was imposed from northwest to southeast by specifying constant fluxes of $2.5 \text{ X}10^{-5}$ m^3/s along the top and left boundaries and fluxes of $-2.5 \text{ X}10^{-5}$ m^3/s along the bottom and right boundaries of the 256m-by-256m domain. These flow conditions produced a uniform mean hydraulic gradient, with no globally converging or diverging flow to distort subsequent effective dispersivity calculations. Seepage velocities were estimated on the basis of a porosity of 0.2, and N=1024 particles were distributed uniformly over a 64-by-64 m^2 square contaminant source area located in the far northwest corner of the domain ($0.0m \leq x \leq 64.0m$, $192.0m \leq y \leq 256.0m$). A particle tracker developed by Schafer-Perini and Wilson (1991) was then used to simulate transient contaminant migration through all 200 two-scale and equivalent single-scale realizations using 1000 time steps with a constant duration of 500 seconds to facilitate subsequent calculation of time-dependent covariance derivatives. A principal advantage of using a particle tracker to predict transport is that it did not introduce the artifice of dispersivity into the problem, instead

permitting evaluation of how different scales of heterogeneity might advectively influence transport. Moreover, particle tracking is not subject to the numerical problems associated with methods based on the advection-dispersion equation with $D^* \simeq 0$, and it could be assumed that numerical errors introduced to the single-scale and two-scale Monte Carlo transport simulations were negligible compared to ensemble uncertainty. Movement of contaminant particles was simulated for up to 1000 time steps until one or more particles exited the transport domain.

Macrodispersivities were estimated using the method of moments according to the Fickian relation:

$$A = \frac{1}{2\bar{v}} \frac{d\sigma^2}{dt} = \frac{1}{2\bar{v}} \frac{d}{dt} \begin{bmatrix} \sigma^2_{xx} & \sigma^2_{xy} \\ \sigma^2_{yx} & \sigma^2_{yy} \end{bmatrix} \qquad [2.40]$$

where $\bar{v}$ is the average linear flow velocity, σ^2_{xx} and σ^2_{yy} are the ensemble mean particle location variances with respect to the plume centroid in the x- and y- directions and σ^2_{xy} and σ^2_{yx} are ensemble mean location cross-covariances in directions oriented at angles of 45^0 and 135^0 with respect to the x-axis. Particle location covariances and cross-covariances were defined with respect to the plume centroid. When divided by solute mass, these represent moments of inertia of the plume.

Dagan (1984) showed that, if the spatial distribution of ensemble mean concentration is gaussian, it represents a solution to the ADE when macrodispersivity given by equation [2.40] is used as the effective dispersion coefficient. Equation [2.40] is valid for a single realization only when the coefficient of variation for the plume

centroid location and moments of inertia are small because these conditions are sufficient to satisfy the ergodicity requirements for a gaussian concentration distribution (Dagan, 1990). When ergodicity requirements are not met, macrodispersivity for each realization will generally depend on travel time and initial and source conditions, and will not be a valid effective property. Given that ergodicity of the two-scale realizations was not anticipated, macrodispersive transport was not expected to behave as a diffusive, Fickian process. Therefore, the method of moments given by equation [2.40] was considered to be a more reliable means of determining macrodispersivity than fitting the simulated plume to advection-dispersion equation predictions.

In order to evaluate equation [2.40], particle location variances were determined by finding the centroid, or center of mass $(\bar{x}, \bar{y})$, of the plume for each time step prior to the exit of a single particle. x and y locations of all N=1024 particles were then subtracted from the centroids, and the variances of the differences were determined as follows:

$$\sigma_x^2 = \frac{1}{N-1}\sum_{i=1}^{N} d_{x_i}^2 = \frac{1}{N-1}\sum_{i=1}^{N} |x_i - \bar{x}|^2$$

$$\sigma_y^2 = \frac{1}{N-1}\sum_{i=1}^{N} d_{y_i}^2 = \frac{1}{N-1}\sum_{i=1}^{N} |y_i - \bar{y}|^2 \qquad [2.41]$$

Covariances were estimated in a similar manner, except that rather than summing the square of the deviations separately in each direction, the cross-products

were summed:

$$\sigma_{xy}^2 = \sigma_{yx}^2 = \frac{1}{N-1} \sum_{i=1}^{N} d_{xy_i}^2 = \frac{1}{N-1} \sum_{i=1}^{N} (x_i - \bar{x})(y_i - \bar{y}) \qquad [2.42]$$

Finally, the averages were differenced over time and divided by 2 times the average seepage velocity $\bar{v}$, which was calculated as the resultant of the average velocity vectors in both the x- and y- directions:

$$\bar{v} = \sqrt{v_x^2 + v_y^2} \qquad [2.43]$$

Since each of the 100 two-scale realizations had a unique local arrangement of large-scale heterogeneities and corresponding unique macrodispersivity versus time estimates, ensemble A_L and A_T vs. time plots were not meaningful for two-scale realizations. Instead, four A_L and A_T vs. time realizations are presented in Figure 2.17 (the conditional plots are discussed later), with the plots for the entire ensemble provided in Appendix A. It appears that a single asymptotic value of dispersivity often, but not always, occurs at late times and large plume displacement distance for the two-scale ln K process. Note that transverse macrodispersivity for both processes was approximately equal to 0.0 at all times due to the absence of diffusion. From inspection of the macrodispersivity plots, there is some evidence of multiple-scale transport behavior paralleling the two-scale conductivity process, as exhibited by multiple

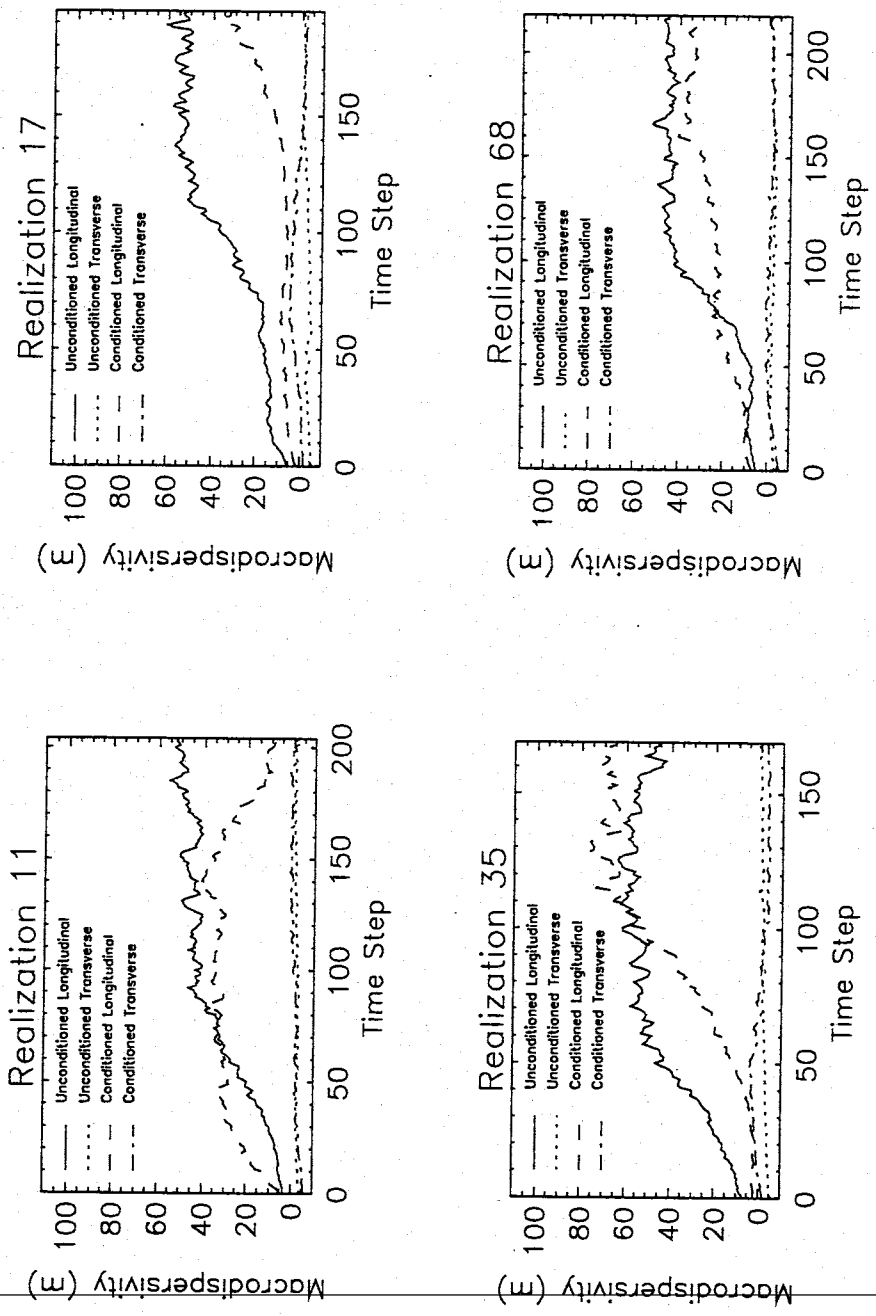


Figure 2.17. Macrodispersivity vs. Time for 4 Two-Scale Conductivity Realizations.

asymptotes. However, the vast majority of macrodispersivity vs. time plots exhibit no evidence of an asymptote corresponding to the length scale associated with the smaller, gaussian conductivity structure. This can be attributed to the smoothing effect of the transport equation, which tends to dampen variations in concentration via spatial integration.

Similar plots for four single-scale realizations are presented in Figure 2.18, with the plots for the ensemble again presented in Appendix A. Fickian behavior appears to prevail at late times. Appendix A also includes ensemble plots for the single-scale case as well as for the two-scale case. They both exhibit classic Fickian behavior associated with a single asymptote, illustrating that Fickian behavior occurs only in the ensemble. This suggests that neither set of single-scale and two-scale transport processes fulfilled ergodicity requirements that source area scale and travel distance scale be much larger than the scale of heterogeneity. Nonergodicity could be attributed to uncertainty in plume centroid location across both two-scale and single-scale ensembles (Dagan, 1990). These results highlight the need to condition on site-specific concentration measurements (Andricevic, 1996).

Macrodispersivity tensor elements were estimated for all conductivity realizations on the basis of the final asymptote in each of the macrodispersivity vs. time plots presented in Appendix A, as determined by the criteria that final macrodispersivity changed by less than 20 m over at least 50 consecutive time steps. The ensemble macrodispersivity tensor estimated on the basis of equation [2.40] was

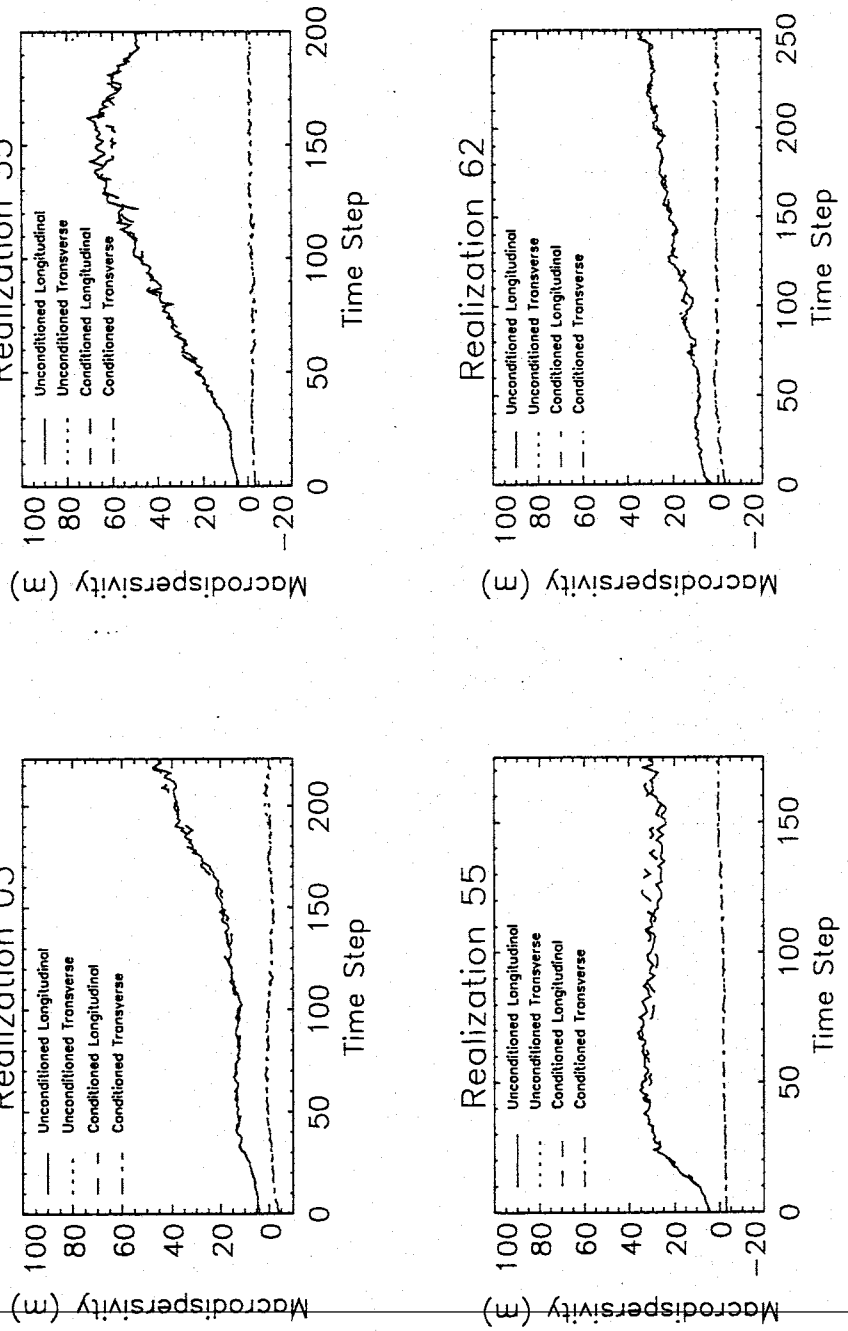


Figure 2.18 Macrodispersivity vs. Time for 4 Single-Scale Conductivity Realizations.

hen diagonalized in order to obtain principal macrodispersivities. These tensors are ɔresented in Figure 2.19.

Single-Scale	Two-Scale
$\begin{bmatrix} 33 \pm 1.7 & 0.0 \\ 0.0 & -0.8 \pm 2.0 \end{bmatrix}$	$\begin{bmatrix} 42 \pm 2.2 & 0.0 \\ 0.0 & -1.1 \pm 2.2 \end{bmatrix}$

Figure 2.19. Diagonalized Ensemble Macrodispersivity Tensors for Unconditioned Conductivity Processes (m) (Including Standard Errors of Estimate).

Unlike the elements of $\ln \mathbf{K}_{\mathbf{eff}}$, the elements of **A** for the single-scale and two-scale processes showed some discrepancy, despite the fact that the single-scale and two-scale processes were forced to have the same global properties. Again, this could be related to the greater spatial smoothing associated with second-order integrations performed during flow simulations. The standard errors of estimate for the two-scale ensemble were somewhat larger than for the single-scale case because pronounced variability in the large-scale component of any given two-scale conductivity realization

introduced widespread spatial variability to the realization, which under the ergodic hypothesis translated to larger ensemble uncertainty. The larger standard error can also be viewed as the result of greater error associated with sampling a conductivity realization off the joint pdf for the two-scale case, which has a larger variance than the single-scale joint pdf as a result of the large variance associated with the large-scale conductivity component.

Estimated longitudinal macrodispersivity for both two-scale and two-scale cases agreed well with the theoretical estimate provided by Gelhar and Axness (1983, eq. 46) for a three-dimensional, statistically anisotropic medium with mean flow parallel to bedding, exponential covariance, and zero-valued local dispersivity:

$$A_L \approx \sigma^2_{\ln K} \lambda_l /(1+\sigma^2_{\ln K}/6)^2 = 3.7(18.0)/(1.+3.7/6)^2 = 25.5 \text{ m}$$

This theoretical relation corresponds to the limit as longitudinal and transverse horizontal correlation scales become much larger than the vertical correlation scale, and perfect stratification occurs. A_L also agreed well with the depth-averaged two-dimensional theoretical longitudinal macrodispersivity value given by equation 71 of Gelhar and Axness (1983) for $\phi=0^0$:

$$\gamma = \frac{\exp\left[\sigma^2_{\ln K}\left(\frac{1}{2}-\frac{\lambda_l/\lambda_t}{1+\lambda_l/\lambda_t}\right)\right]}{\left[\sin^2\phi + \cos^2\phi \exp\left[\sigma^2_{\ln K}\left(\frac{1}{1+\lambda_l/\lambda_t}-\frac{\lambda_l/\lambda_t}{1+\lambda_l/\lambda_t}\right)\right]\right]} = \frac{\exp\left[3.7\left(\frac{1}{2}-\frac{1.8}{1+1.8}\right)\right]}{\exp\left[3.7\left(\frac{1}{1+1.8}-\frac{1.8}{1+1.8}\right)\right]} = 1.68$$

$$A_L = \frac{\sigma^2_{\ln K}\lambda_l\lambda_t}{\gamma^2(\lambda_l^2\sin^2\phi + \lambda^2_t\cos^2\phi)^{1/2}} = \frac{3.7(18.0)(10.0)}{1.68^2(0.0 + 10.0^2)^{1/2}} = 23.6 \text{ m}$$

$$A_T \equiv 0.0 \text{ m}$$

Discrepancies were assumed related to violation of first-order assumptions by the fairly large ln K variance of 3.7 (ln m/s)^2, as well as the absence of perfect stratification. The latter factor would cause longitudinal Monte Carlo ensemble macrodispersivity estimates to be significantly larger than the theoretical value because lenses of heterogeneity produce greater advective mixing in the longitudinal direction than perfectly stratified heterogeneities. Gelhar and Axness (1983) point out that the linearization required to obtain theoretical longitudinal macrodispersivity values may not be valid for ln K perturbation variances in excess of 1. In any event, published analytical results of depth-averaged parameters in three-dimensional conductivity fields are not strictly comparable to parameters obtained for two-dimensional fields.

Note that transverse macrodispersivity for the single scale case was not statistically different from zero. Zero-valued A_T is, in fact, predicted theoretically by Gelhar and Axness (1983) for cases in which local dispersivity D^* is equal to zero

because transverse dispersivity is diffusion-controlled in a stratified medium with flow parallel to bedding.

2.9 TRANSPORT OF INFORMATION BETWEEN SCALES

An important concern is the issue of how conductivity information scales, or how it moves between different scales of variation. To address this issue, both two-scale and single-scale ln K processes were corrupted by a set of 196 synthetic permeameter measurement errors distributed in 4 clusters of 49 evenly-distributed data points each, as shown in Figure 2.20. Of interest was how unstructured information such as error changes effective properties as it is moved from one scale to another via conventional upscaling techniques. Errors were assumed to be independent, gaussian-distributed, zero-mean random variables with a log variance equal to 0.037 $(\ln m/s)^2$, amounting to roughly 1% of the ensemble unconditional ln K variance of 3.7 $(\ln m/s)^2$. Corrupted ln K data for all two-scale and single-scale realizations were obtained by adding permeameter error to random variables extracted from each of the 100 unconditioned single-scale realizations at the location where permeameter error was observed. The resulting values represented error-free conductivities perturbed by uncorrelated error. The implied assumption was that permeameter errors were independent of the unconditional measurement.

The point permeameter errors were moved to the large scale of variation in each two-scale realization by conditioning twice — once at the large scale using point-to-block kriging and then again at the small-scale using point-to-point kriging.

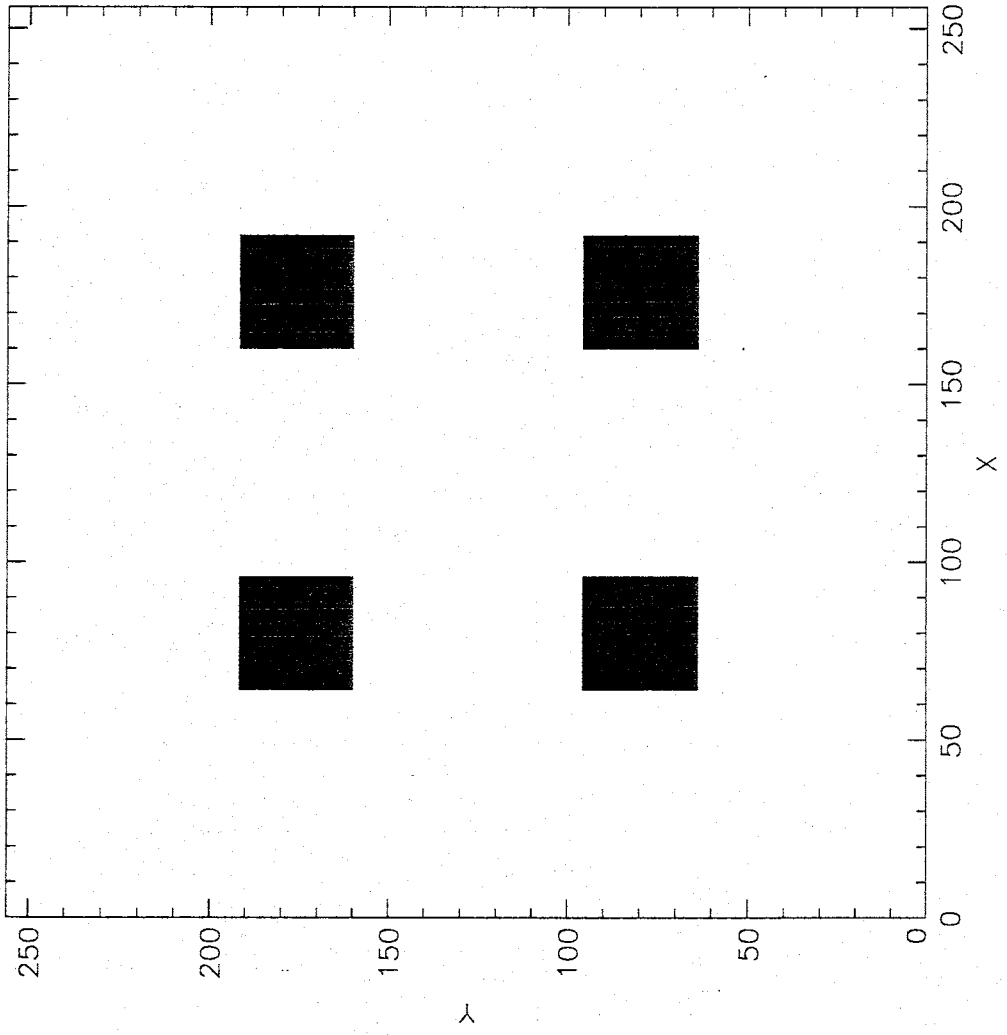


Figure 2.20. Locations of Noisy Permeameter Data Clusters.

Point-to-block kriging was performed in each large-scale block by comparing the point-to-block kriging result to all possible large-scale states (in this case, −7., −6., −5., −4., −3. ln m/s), and assigning the state closest to the kriged value to the block. This large-scale conditioned log permeability was held constant during execution of the Metropolis exchange algorithm. Point-to-block kriging is a linear operation and can be considered a form of linear upscaling. Finally, the small-scale gaussian realization embedded in the large-scale block was conditioned on the same error in the block using point-to-point kriging of log permeability. This conditioning was performed in order to justify later comparisons with results from the conditioned two-scale process, and was not related to any upscaling of permeameter information between spatial scales. The conditioned realizations corresponding to the realizations shown in Figure 2.8 are presented in Figure 2.21.

Following conditioning of all 200 realizations, the same periodic flow and transport problems as specified for the unconditional realizations were executed, with effective conductivity and macrodispersivity tensors again estimated for both processes. Macrodispersivity vs. time plots for both ensembles are presented in Appendix A. Note that conditioning on measurement error had a much less pronounced effect on the macrodispersivity vs. time behavior in all of the single-scale realizations than for the majority of two-scale realizations. The resulting diagonal tensors $\mathbf{ln\ K_{eff}}$ and **A** are presented in Figure 2.22 and Figure 2.23.

The $\mathbf{ln\ K_{eff}}$ tensors changed very little in response to conditioning on noisy permeameter data, presumably due to the efficient dissapation of energy for the case of

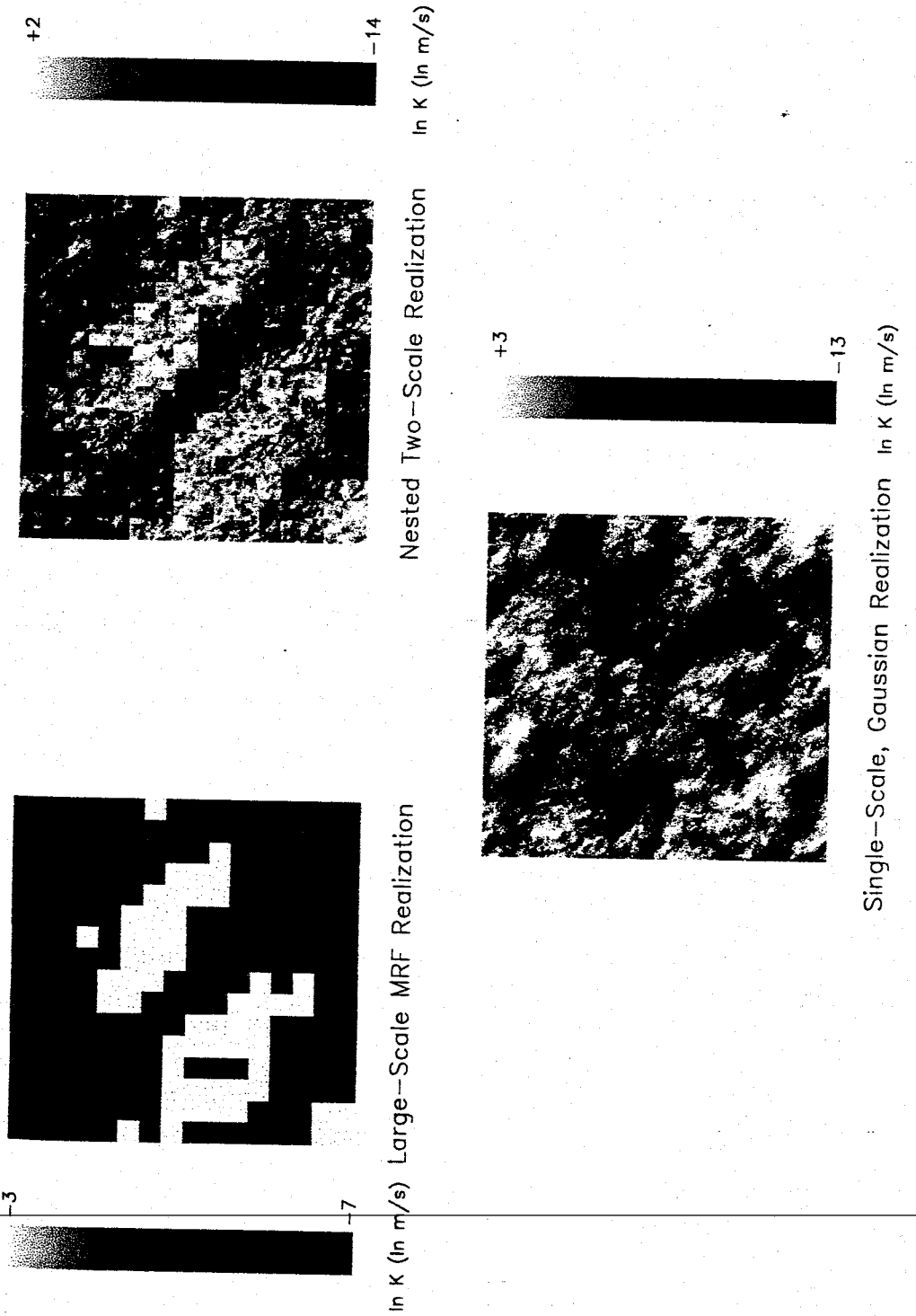


Figure 2.21. Conditioned Two-Scale and Statistically-Equivalent Single-Scale Conductivity Realizations.

Single-Scale

$$\begin{bmatrix} -2.8 \pm 0.11 & 0.0 \\ 0.0 & -4.2 \pm 0.10 \end{bmatrix}$$

Two-Scale

$$\begin{bmatrix} -2.6 \pm 0.11 & 0.0 \\ 0.0 & -3.7 \pm 0.12 \end{bmatrix}$$

Figure 2.22. Diagonalized Ensemble Log Effective Conductivity Tensors for Conductivity Processes Conditioned on Noisy Permeameter Measurements Using Point-to-Block Kriging for Large-Scale Conditioning (ln m/s).

the parabolic PDE governing subsurface flow (see Figure 2.16). However, from inspection of Figure 2.23 and Figure 2.19, it is evident that changes in the two-scale ensemble macrodispersivity tensor in response to conditioning were larger than changes in the equivalent single-scale tensor. This is not surprising, considering that conditioning of the two-scale realizations was performed not only at the small scale, but also at the large scale where conditioning had a spatially widespread impact on conductivity realizations. These results suggest that measurement error, or unexplained behavior, becomes amplified as we try to move it from the smaller scale to the larger scale. Note that such magnification might prove to be much smaller for the fully three-dimensional

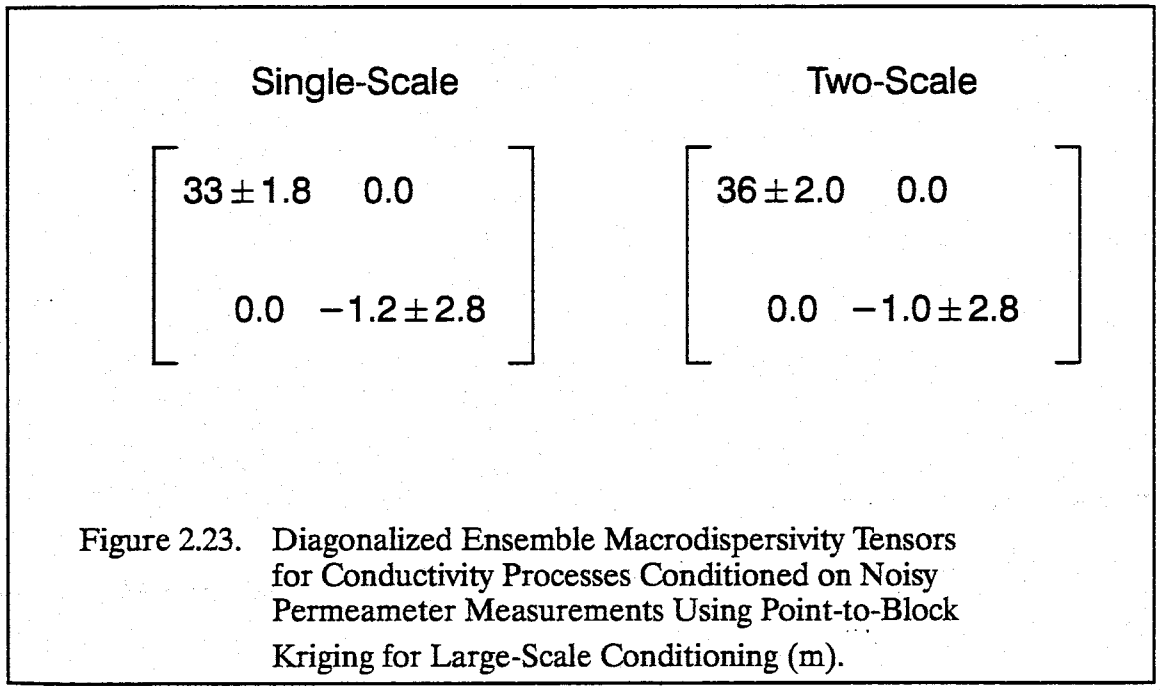


Figure 2.23. Diagonalized Ensemble Macrodispersivity Tensors for Conductivity Processes Conditioned on Noisy Permeameter Measurements Using Point-to-Block Kriging for Large-Scale Conditioning (m).

case, in which conductivity perturbations would likely have substantially less impact on effective properties as a result of the larger number of degrees of freedom associated with higher-dimensional flow and transport. It is possible that use of process models such as SEDSIM (Tetzlaff and Harbaugh, 1989) or geometric models such as the Braided Channel Simulator developed by Webb and Anderson (1996) to generate large-scale structure may also yield substantially different results following upscaling of error.

Permeameter measurement error was chosen to condition both ensembles of ln K processes because it was considered to be physically meaningful way to perturb the

system. However, even in the absence of error, the problem of moving information from one scale to another would be expected to persist. In fact, it would be anticipated that movement of structured information would likely produce even more pronounced changes in effective properties. Thus, perturbation using uncorrelated error provides a conservative estimate of change.

As demonstrated in Chapter 5, the very act of conditioning large-scale structure on data reduces uncertainty in transport prediction and results in smaller macrodispersive flux and a decrease in macrodispersivity. Therefore, the more pronounced change in macrodispersivities for the two-scale case is not particularly surprising. Of greater interest is how the reduction in **A** elements for the two-scale case depends on the type of linear upscaling rule used to condition the large-scale conductivity process. Figure 2.24 and Figure 2.25 show $\mathbf{ln\ K_{eff}}$ and **A** estimated for unconditioned and conditioned two-scale flow and transport processes when the geometric mean of noisy permeameter measurements, rather than point-to-block kriging, is used to condition large-scale structure. Since the geometric average translates to a linear average in ln K, it can be consider another form of linear upscaling. For both tensors, point-to-point kriging was retained for small-scale conditioning. Note that unconditioned $\mathbf{ln\ K_{eff}}$ and **A** are different from the previous example because some unconditioned realizations used in that example resulted in one or more particles exiting the domain prior to the time at which criteria used to estimate macrodispersivity were satisfied, and had to be replaced with other realizations that did satisfy the criteria.

Unconditioned	Conditioned
$\begin{bmatrix} -3.5 \pm 0.09 & 0.0 \\ 0.0 & -4.6 \pm 0.10 \end{bmatrix}$	$\begin{bmatrix} -3.5 \pm 0.08 & 0.0 \\ 0.0 & -4.5 \pm 0.10 \end{bmatrix}$

Figure 2.24. Diagonalized Ensemble Log Effective Conductivity Tensors for Two-Scale Conductivity Process Unconditioned and Conditioned on Noisy Permeameter Measurements Using Geometric Mean for Large-Scale Conditioning (ln m/s).

Ln K_{eff} for both processes showed little evidence of change in response to large-scale conditioning on the geometric mean of noisy permeameter values within each block. While no changes in A_L and A_T were observed for the single-scale case, A_L and A_T for the two-scale case changed by roughly 7 m and 0.2 m, compared to 6 m and 0.1 m in the previous example. Since both kriging and geometric averaging can be considered different forms of linear upscaling, these results demonstrate that movement of information between spatial scales is a function of the type of rule used to perform the linear upscaling. Via the process of linear upscaling from the point to the block scale using both point-to-block kriging and geometric averaging, we obtained

Unconditioned	Conditioned
$\begin{bmatrix} 42 \pm 2.2 & 0.0 \\ 0.0 & -1.1 \pm 2.1 \end{bmatrix}$	$\begin{bmatrix} 35 \pm 1.9 & 0.0 \\ 0.0 & -0.9 \pm 1.6 \end{bmatrix}$

Figure 2.25. Diagonalized Ensemble Macrodispersivity Tensors for Two-Scale Conductivity Process Unconditioned and Conditioned on Noisy Permeameter Measurements Using Geometric Mean for Large-Scale Conditioning (m).

measurably different estimates of the effective characteristics of our system for a perturbation amounting to only 1% of the unconditional variance. Substantially larger differences could be expected for more typical ranges of measurement error encountered under normal field conditions, as well as for the case of spatially correlated error. The outcome of the numerical experiment is fundamentally at odds with the whole premise of upscaling, which is to *preserve* the same characteristics of the system as reflected by the point properties.

Most likely, information scales as some highly nonlinear spatial function that depends on the particular physics of the transport problem. Depending on the nature

of the nonlinearity, changes in $\ln K_{eff}$ or A may be smaller or larger than those predicted using a linear upscaling rule. In any case, we will not generally have a very good understanding of what the scaling rule might be. The exception to this is an ideal fractal process, for which information systematically scales as a function of observation scale (Wheatcraft and Tyler, 1988).

Results of the conductivity-information scaling experiment suggest that, in the absence of knowledge about the upscaling rule, and with evidence that even simple linear upscaling of uncorrelated error can alter the properties of a hierarchically-structured system, we must find a way to avoid the problem of measuring at one scale and using these measurements to make predictions at other scales. The only means of doing so is by freeing ourselves from the constraint of viewing hydraulic conductivity through some arbitrary instrument window — in this case, the permeameter measurement window. As proposed by Cushman (1984), the scales must be decoupled by acquiring information separately at each scale relevant to the problem using a succession of instruments, each designed specifically to measure the process at a distinct scale of natural variation. Tidwell and Wilson (1997) used just this approach to estimate statistics of conductivity as a function of variable support size, by measuring permeability of Berea sandstone via a gas permeameter with variable tip seal size. They discovered distinct changes in mean, variance, and semi-variogram with changing support size, but whether these changes are evidence of upscaling behavior remains to be seen. Chapter 4 expands further on this idea of using more than one instrument to characterize scale-dependent variations in random hierarchical geologic processes.

2.10 SUMMARY

A relativist approach to measuring heterogeneities generalizes the concept of measurement domain so that all measurements in a multiple-scale porous medium are equally valid, regardless of instrument support. In the presence of more than one scale of variation, randomness can be a function of observation scale and spatial variation should be measured through instrument windows corresponding to scales that most dominate the particular flow or transport problem of interest.

Effective conductivity and macrodispersivity for ensembles of nested, two-scale ln K realizations and statistically-equivalent single-scale gaussian ln K realizations were estimated. When conditioned on 4 clusters of noisy permeameter errors characterized by 1% of the ln K variance, longitudinal macrodispersivity for the two-scale conductivity process was more dependent on the type of linear upscaling rule used than longitudinal macrodispersivity for the equivalent single-scale process. The fact that macrodispersivity of the system changed with the upscaling rule is in violation of the whole premise underlying upscaling theory. Effective conductivity was significantly less affected by conditioning on noisy data than macrodispersivity due to energy dampening tendencies inherent to the flow equation. To overcome the problem of using ln K measurements collected at one scale to make transport predictions at another scale, an attempt should be made to decouple dominant scales of variation by acquiring conductivity information independently at each scale relevant to the flow or transport problem.

2.11 REFERENCES

Ababou, R., and E.F. Wood, Comments on 'Effective groundwater model parameter values: Influence of spatial variability of hydraulic conductivity, leakance, and recharge" by J.J. Gomez-Hernandez and S.M. Gorelick, *Water Resour. Res.*, 26(8), 1843-1846, 1990.

Andrecevic, R., Evaluation of sampling in the subsurface, *Water Resour. Res.*, 32(4), 863-874, 1996.

Bakr, A.A., *Stochastic analysis of the effect of spatial variations in hydraulic conductivity on groundwater flow*, Ph.D dissertation, New Mexico Inst. Mining and Tech., 1976.

Bakr, A.A., L.W. Gelhar, A.L. Gutjahr, and J.R. MacMillan, Stochastic analysis of spatial variability in subsurface flows, 1. Comparison of one- and three-dimensional flows, *Water Resour. Res.*, 14(2), 263-271, 1978.

Baveye, P. and G. Sposito, The operational significance of the continuum hypothesis in the theory of water movement through soils and aquifers, *Water Resour. Res.*, 20(5), 521-530, 1984.

Bear, J. *Dynamics of Fluids in Porous Media,* American Elsevier, New York, 1972.

Beckie, R., Measurement scale, network sampling scale, and groundwater model parameters, *Water Resour. Res.*, 32(1), 65-76, 1996.

Beckie, R., and B. Wang, A numercial method to characterize the averaging process invoked by a slug test, in *Computational Methods in Water Resources,* A. Peters (ed.), Kluwer Acad., Norwell, Mass., 1994.

Bhattacharya, R.N., and V.K. Gupta, A theoretical explanation of solute dispersion in saturated porous media at the Darcy scale, *Water Resour. Res.*, 19(4), 938-944, 1983.

Bierkins, M.F.P., Modeling hydraulic conductivity of a complex confining layer at various spatial scales, *Water Resour. Res.*, 32(8), 2369-2382, 1996.

Bourgeat, A., Homogenized behavior of two-phase flows in naturally fractured reservoirs with uniform fracture distribution, *Comput. Meth. Appl. Mech. Eng.*, 47, 205-216, 1984.

Cushman, J.H., Multiphase transport equations: 1. General equation for macroscopic statistical, local, space-time homogeneity, *Transp. Theor. Stat. Phys.*, 12, 35-71, 1983.

Cushman, J.H., On unifying the concepts of scale, instrumentation, and stochastics in the development of multiphase transport theory, *Water Resour. Res.*, 20(11), 1668-1676, 1984.

Cushman, J.H., More on stochastic models, *Water Resour. Res.*, 23(4), 750-52, 1987a.

Cushman, J.H., Development of stochastic partial differential equations for subsurface hydrology, *Stochastic Hydrol. Hydraul.*, 1(4), 241-262, 1987b.

Cushman, J.H. (ed.), An introduction to hierarchical porous media, in *Fluids in Hierarchical Porous Media*, Academic Press, San Diego, Calif., pp. 1-6, 1990.

Dagan, G., Stochastic modeling of groundwater flow by unconditional and conditional probabilities, 1. Conditional simulation and the direct problem, *Water Resour. Res.*, 18(4), 813-833, 1982.

Dagan, G., Solute transport in heterogeneous porous formations, *J. Fluid Mech.*, 145, 151-177, 1984.

Dagan, G., Time-dependent macrodispersion for solute transport in anisotropic heterogeneous aquifers, *Water Resour. Res.*, 24(9), 1491-1500, 1988.

Dagan, G., Transport in heterogeneous porous formations: Spatial moments, ergodicity, and effective dispersion, *Water Resour. Res.*, 26(6), 1281-1290, 1990.

Davis, J.M., *A conceptual sedimentological-geostatistical model of aquifer heterogeneity based on outcrop studies,* Ph.D dissertation, 234 pp., New Mexico Institute of Mining and Technology, 1994.

Davis, J.M., R.C. Lohmann, S.J. Colarullo, and F.M. Phillips, Alluvial aquifer heterogeneities in the Rio Grande Valley: Implications for ground water contamination, *New Mexico Resources Research Institute Technical Completion Report*, 1990.

Desbarats, A.J., Numerical estimation of effective permeability in sand-shale formations, *Water Resour. Res.*, 23(2), 273-286, 1987.

Desbarats, A.J., Spatial averaging of hydraulic conductivity in three-dimensional heterogeneous porous media, *Math. Geol.*, 24(3), 249-267, 1992.

Desbarats, A.J., Spatial averaging of hydraulic conductivity under radial flow conditions, *Math. Geol.*, 26(1), 1-21, 1994.

Deutsch, C., and A. Journel, Entropy and spatial disorder, *Math. Geol.*, 25(3), 329-355, 1993.

Durlofsky, L.J., Numerical calculation of equivalent grid block permeability tensors for heterogeneous porous media, *Water Resour. Res.*, 27(5), 699-708, 1991.

Farmer, C.L., *The Generation of Stochastic Fields of Reservoir Parameters with Specified Geostatistical Distributions*, paper presented at the IMA Conference on The Mathematics of Oil Production, Robinson College, Cambridge Univ., July 6-7, 1987.

Freeze, R.A., and J.A. Cherry, *Groundwater*, Prentice-Hall, New Jersey, 1979.

Freyberg, D.L., A natural gradient experiment on solute transport in a sand aquifer, 2. Spatial moments and the advection and dispersion of nonreactive tracers, *Water Resour. Res.*, 22(13), 2031-2046, 1986.

Garabedian, S.P., D.R. LeBlanc, L.W. Gelhar, and M.A. Celia, Large-scale natural-gradient tracer test in sand and gravel, Cape Cod, Massachusetts, 2. Analysis of tracer moments for a nonreactive tracer, *Water Resour. Res.*, 27(5), 911-924, 1991.

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

Gomez-Hernandez, J.J., *A stochastic approach to the simulation of conductivity fields conditioned upon data measured at a smaller scale,* Ph.D. dissertation, Dept. of Applied Sci., Stanford Univ., Calif., 1991.

Gutjahr, A.L., *Fast Fourier Transforms for Random Field Generation,* project report for Los Alamos grant, contract 4-R58-2690R, N.M. Institute of Mining and Technology, Socorro, 1989.

Gutjahr, A.L., L.W. Gelhar, A.A. Bakr, and J.R. MacMillan, Stochastic analysis of spatial variability in subsurface flows, 2. Evaluation and application, *Water Resour. Res.*, 14(5), 953-959, 1978.

Hastings, W.K., Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* 57, 97-109, 1970.

Hoeksema, R.J., and P.K. Kitanidis, Analysis of the spatial structure of properties of selected aquifers, *Water Resour. Res.*, 21(4), 563-572, 1985.

Indelman,P., and G. Dagan, Upscaling of permeability of anisotropic heterogeneous formations, 1. The general framework, *Water Resour. Res.,* 29(4), 917-923, 1993a.

Indelman,P., and G. Dagan, Upscaling of permeability of anisotropic heterogeneous formations, 2. General structure and small perturbation analysis, *Water Resour. Res.,* 29(4), 925-933, 1993b.

Indelman, P., A. Fiori, and G. Dagan, Steady flow toward wells in heterogeneous formations: Mean head and equivalent conductivity, *Water Resour. Res.,* 32(7), 1975-1983, 1996.

Journel, A.G., and F.G. Alabert, Focusing on spatial connectivity of extreme-value attributes: Stochastic indicator models of reservoir heterogeneities, Annual Technical Conference and Exhibition, *SPE Paper 18324*, Houston, 1988.

Journel, A.G., and F.G. Alabert, Non-gaussian data expansion in the earth sciences, *Terra Nova*, 1, 123-134, 1989.

Kitanidis, P.K., Effective hydraulic conductivity for gradually varying flows, *Water Resour. Res.*, 26(6), 1197-1208, 1990.

Koebbe, J.V., *Documentation and Installation Guide for HomCode: A Code for Scaling Up Permeabilities Using Homogenization*, Dept. of Mathematics and Statistics Report No. 83, Utah State University, 1995.

Mantaglou, A., and J.L. Wilson, The turning bands method for simulation of random fields using line generation by a spectral method, *Water Resour. Res.*, 18(5), 1379-1394, 1982.

Marle, C.-M., Ecoulements monophasiques en milieu poreux, *Rev. Inst. Fr. Petrol.*, 22, 1471-1509, 1967.

Matheron, G., *Les Variables Regionalisees et Leur Estimation*, Massson, Paris, 1967.

Naff, R.L., T.-C.J. Yeh, and M.W. Kemblowski, A note on the recent natural gradient test at the Borden site, *Water Resour. Res.*, 24(12), 2099-2013, 1988.

Neuman, S.P., and S. Orr, Prediction of steady state flow in nonuniform geologic media by conditional moments: Exact nonlocal formalism, effective conductivities and weak approximation, *Water Resour. Res.*, 29(2), 341-364, 1993.

Oliver, D.S., The averaging process in permeability estimation from well-test data, *SPE Form. Eval.*, 5, 319-324, 1990.

Rehfeldt, K. R., and L.W. Gelhar, Stochastic analysis of dispersion in unsteady flow in heterogeneous aquifers, *Water Resour. Res.*, 28(8), 2085-2099, 1992.

Robin, M.J.L., A.L. Gutjahr, E.A. Sudicky, and J.A. Wilson, Cross-correlated random field generation with the direct Fourier transform method, *Water Resour. Res.*, 29(7), 2385-2397, 1993.

Rubin, Y., and J.J. Gomez-Hernandez, A stochastic approach to the problem of upscaling of conductivity in disordered media: Theory and unconditional numerical simulations, *Water Resour. Res.*, 26(4), 691-701, 1990.

Saad, Y., and M.H. Schultz, GMRES: A generalized minimum residual algorithm for solving non-symmetric linear systems, *SIAM Jour. Sci. Stat. Comput.*, 7(3), 856-869, 1986.

Saez, A.E., C.J. Otero, and I. Rusinek, The effective homogeneous behavior of heterogeneous porous media, *Transp. Porous Media*, 4, 213-238, 1989.

Sanchez-Vila, X., J.P. Girardi, and J. Carrera, A synthesis of approaches to upscaling of hydraulic conductivities, *Water Resour. Res.*, 31(4), 867-882, 1995.

Schafer-Perini, A.L., and J.L. Wilson, Efficient and accurate front tracking for two-dimensional groundwater flow models, *Water Resour. Res.*, 27(7), 1471-1485, 1991.

Smith, L., and F.W. Schwartz, Mass transport, 1. A stochastic analysis of macroscopic dispersion, *Water Resour. Res.*, 16(2), 303-313, 1980.

Sposito, G., The statistical mechanical theory of water transport through unsaturated soil, 1. The conservation laws, *Water. Resour. Res.*, 14, 474-478, 1978.

Sudicky, E.A., A natural gradient experiment on solute transport in a sand aquifer, 4. Spatial variability of hydraulic conductivity and its role in the dispersion process, *Water Resour. Res.*, 22(13), 2069-2082, 1986.

Sudicky, E.A., J.A. Cherry, and E.O. Frind, Hydrogeological studies of a sandy aquifer at an abandonned landfill, 4. A natural gradient dispersion test, *J. Hydrol.*, 63, 81-108, 1983.

Tidwell, V.C., and J.L. Wilson, Laboratory method for investigating permeability upscaling, *Water Resour. Res.*, 33(7), 1607-1616, 1997.

Tompson, A.F.B., and L. Gelhar, Numercial simulation of solute transport in three-dimensional, randomly heterogeneous porous media, *Water Resour. Res.*, 26(10), 2541-2562, 1990.

Warren, J.E., and H.S. Price, Flow in heterogeneous porous media, *Soc. Petrol. Eng. J.*, 1, 153-169, 1961.

Warren, J. E., and F.F. Skiba,, Macroscopic dispersion, *Trans. Amer. Inst. Min. Metal. Petrol. Eng.*, 231, 215-230, 1964.

Webb, E.K., and M.P. Anderson, Simulation of preferential flow in three-dimensional, heterogeneous conductivity fields with realistic internal architecture, *Water Resour. Res.*, 32(3), 533-545, 1996.

Wheatcraft, S.W., and S.W. Tyler, An explanation of scale-dependent dispersivity in heterogeneous aquifers using concepts of fractal geometry, *Water Resour. Res.*, 24(4), 566-578, 1988.

White, C.D., and R.N. Horne, *Computing absolute transmissibility in the presence of field-scale heterogeneity,* paper presented at the SPE Symposium on Reservoir Simulation, Soc. Pet. Eng., San Antonio, Texas, 1987.

Chapter 3. Characterization of Discrete Geologic Structure Using Markov Random Fields

The information-transfer experiment conducted in Chapter 2 relied on the use of a Markov random field (MRF) model to generate large-scale spatially discrete geologic structure. Unlike traditional stochastic methods of gaussian random process generation, MRF models allow spatial correlation to be imposed over discrete rather than continuous space, offering a means of reproducing abrupt changes in geologic structure encountered at many scales of variation. Before proceeding with the study of aquifer characterization in the presence of multiple-scale heterogeneities, it might be instructive to discuss MRFs in greater detail.

The use of Markov random fields was first introduced by Ising (1925) to study ferromagnetism. Although MRFs were ultimately found to be inadequate for describing magnetism, they represent good approximations of phase-separated alloys, ideal gases, and crystals. During the past decade, Markov random fields have gained widespread acceptance as tools for the restoration of degraded images (Cross and Jain, 1983). In the application of MRFs to image restoration, a prior probability density function (pdf) is conditioned on nondegraded portions of the image to obtain a maximum likelihood *a posteriori* distribution. The problem of characterizing discrete aquifer structure differs from the problem of image restoration, in which prior information extracted from the degraded image is used to reproduce the texture of an

ideal image. Rather, aquifer characterization requires a feature-based approach, directed toward finding a stochastic geologic process that can reproduce the statistical characteristics of the field without regard to a single, ideal image. In other words, we are more interested in finding a *possible* MRF realization than in finding the *most likely* one.

3.1 LITHOLOGIC FACIES SEQUENCES AS ONE-DIMENSIONAL MARKOV CHAINS

MRFs are a mathematical extension of one-dimensional Markov processes to multidimensional space. As illustrated in Figure 3.1, one-dimensional Markov chain models have traditionally been used to statistically characterize discrete geologic structure. The probability of the lithology assuming a state S_j at time n, conditional upon all states preceding it at time n-1, n-2, . . . , 1, is given by:

$$p(x_n = S_j \mid x_{n-1} = S_{i_{n-1}},\ x_{n-2} = S_{i_{n-2}},\ \ldots,\ x_1 = S_{i_1}) \qquad [3.1]$$

where i_{n-k} is the lithologic state index at time n-k, k=1, 2, . . . , n-1. The sequence x_1, $x_2, \ldots, x_n$ is a Markov chain if the distribution of x_n is independent of all previous states except for the one immediately prior to it (Hammersley and Handscomb, 1964):

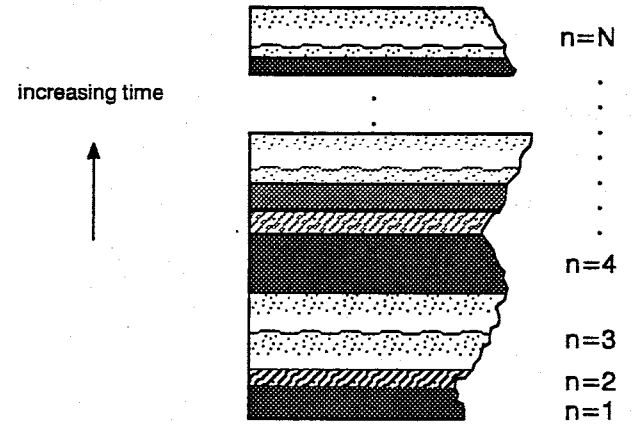


Figure 3.1. A Vertical Geologic Sequence as a One-Dimensional Markov Chain.

$$p(x_n = S_j \mid x_{n-1} = S_{i_{n-1}},\ x_{n-2} = S_{i_{n-2}},\ \ldots,\ x_1 = S_{i_1}) = p(x_n = S_j \mid x_{n-1} = S_{i_{n-1}}) \qquad [3.2]$$

If it can be further assumed that the Markov process is stationary, transition probabilities $p_{ij} = p(x_n = S_j \mid x_{n-1} = S_i)$ do not depend on n and represent elements in a transition probability matrix **P**. Elements of this transition probability matrix describe the likelihood of making a transition from lithology i to any other lithology j, where the probability of a certain lithology is assumed to depend, or be conditioned, only on the lithology preceding it in depositional time, independent of lithologic states in the

distant past. **P** embodies statistical information contained in the geologic record and can be determined through inspection of long vertical sequences of lithologic facies.

The Markov chain is updated by conditioning the current estimate of the probabilistic distribution of lithologies $\bar{\pi}_n$ on both the prior estimate $\bar{\pi}_{n-1}$ and **P**:

$$\bar{\pi}_n = \bar{\pi}_{n-1}\,\mathbf{P} \qquad [3.3]$$

One can view equation [3.3] from a Bayesian perspective, with $\bar{\pi}_{n-1}$ representing the prior estimate of the state probability vector and $\bar{\pi}_n$ representing the posterior estimate, conditioned on transition probabilities. To extend the conditioning through an arbitrary time period, the state probability vector at time n can be determined on the basis of the previous state probability vector at any time m by simply raising **P** to the n-m power. Note that for a continuous-state situation in which the number of possible lithologic states is infinite, it becomes more meaningful to view **P** as a density function rather than as a probability transition matrix.

Equation [3.3] is considered the master equation of statistical mechanics because it dictates the stochastic evolution in time of a one-dimensional Markov process. It can be shown that equation [3.3] is nothing more than a probability-mass conservation equation that balances probability fluxes between states in the state space, with zero net flux at equilibrium (Whittle, 1986):

$$\pi_j - \sum_j \pi_i p_{ij} = 0 \qquad [3.4]$$

If the Markov sequence is sufficiently long and if certain other conditions are met, the distribution of lithologies in the sequence will converge to an equilibrium state probability vector $\bar{\pi}$ that describes the probability of any given lithology occurring, independent of initial lithology:

$$\lim_{n\to\infty} \bar{\pi}_n = \bar{\pi} \qquad [3.5]$$

Central to attainment of the equilibrium state probability vector is the requirement that the chain be both irreducible and aperiodic. In an irreducible chain, there is an absence of absorbing or self-avoiding states, and every state is accessible from every other state. Irreducibility essentially amounts to a condition of probabilistic connectedness. Aperiodicity refers to the property whereby the current state can return to a given state without first passing through a fixed number of states, and is not subject to oscillation between different sets of states in a regular periodic manner. When the Markov chain is irreducible and aperiodic, it is free to visit all lithologic states and complete information about the probability distribution $\bar{\pi}$ can be obtained by simply allowing the chain to approach an infinite length.

Note that an element of $\bar{\pi}$ represents the probability of a certain state when all other states are considered simultaneously, or jointly. That is, the equilibrium state

probability vector $\bar{\pi}$ summarizes the joint probabilistic behavior of the geologic sequence. Equation [3.5] states that, through repeated application of equation [3.3], we are able to build joint structure based on more accessible conditional behavior, as summarized by transition probability matrix **P**. The equilibrium state eventually converges to a sample off the joint distribution for large n.

3.2 MARKOV RANDOM FIELDS

3.2.1 Extension of the Markov Property to Higher-Dimensional Space

In order to extend the Markov property to two or three dimensions in space, the concept of unilateral time must be extended to multilateral space. In the one-dimensional Markov chain shown schematically in Figure 3.1, the vertical axis can be considered analogous to time. A point along that axis effectively partitions time into past and future, with the future conditioned on the past. However, if we want to extend the Markov property to two- or three-dimensional space, there exists no such natural partitioning because there is no concept of past and future and no preferred direction of conditioning. Therefore, if we want to generate a multidimensional Markov random *process*, we must replace the idea of past and future with the concept of a neighborhood. The probability that a random variable assumes a certain value at a site will then depend on the variables at locations contained in some specified neighborhood of the site rather than on a single variable observed at some prior time. Geman and Geman

(1984) credit Dobrushin (1968) with developing the concept of a Markov random field as an extension of the Markov property. A principal advantage of Markov random field models over traditional spatially-continuous models is that the limits of the neighborhood can be spatially finite, allowing correlation to be imposed over discrete rather than continuous space.

Figure 3.2 illustrates how conditioning with respect to the neighborhood produces a sequence of lattice states along a relaxation chain, analogous to a one-dimensional Markov chain. The chain is referred to as a relaxation chain because an operator is applied iteratively to relax it to final equilibrium state. Starting with an arbitrary uncorrelated process, random variables at two randomly chosen lattice sites at each chain level are locally conditioned on values within their immediate neighborhood in the preceding lattice state. A sequence of lattice states is generated by locally conditioning with respect to the prior lattice state in the neighborhoods of the two selected sites rather than with respect to the entire vector state preceding it, using a conditional operator to update within the neighborhood. Like the case for the one-dimensional Markov chain, the relaxation chain eventually converges to an equilibrium state that represents a sample from the joint pdf.

For an MRF, the conditional operator **P** is a simulation analogy to equation [3.3]. The problem of synthesizing a realization of a Markov random field reduces to one of finding the correct conditional operator and generating a sequence of states in two- or three-dimensional space that will eventually converge to a realization with the

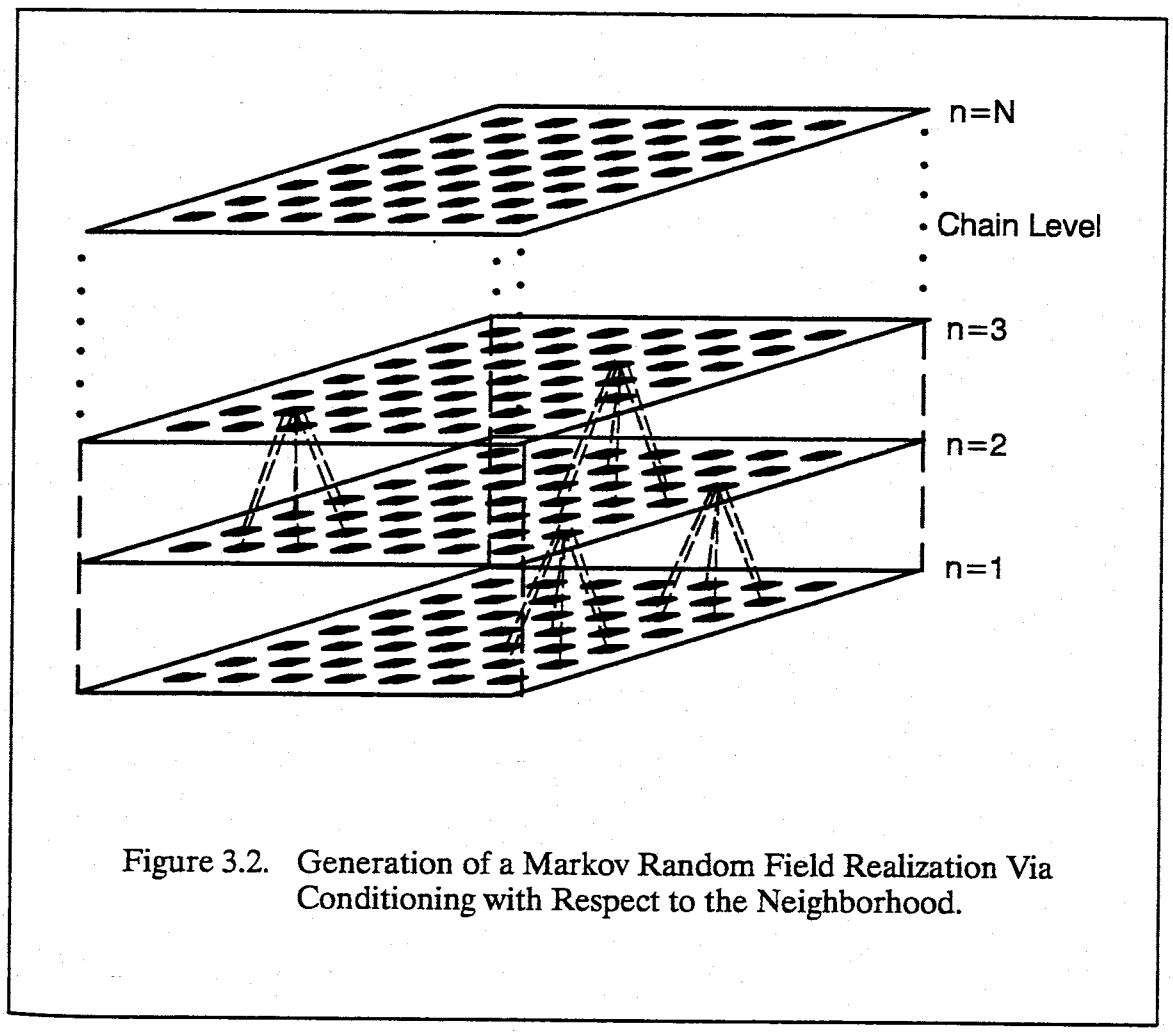


Figure 3.2. Generation of a Markov Random Field Realization Via Conditioning with Respect to the Neighborhood.

desired joint probabilistic structure, just as the vector $\bar{\pi}_n$ converges to the joint probability vector $\bar{\pi}$ in equation [3.5].

This conditional approach to generating MRF structure allows us to break the joint problem into smaller, more manageable conditional subproblems (Rothman, 1985). We simply make local updates in each subproblem while maintaining the correct

relationship between subproblems by allowing them to interact through their overlapping neighborhoods. Since all sites are neighbors of one another through overlapping neighborhoods, the graph of the MRF is completely connected, and local conditional characteristics are ultimately propagated globally throughout the domain to produce the desired joint structure.

3.2.2 Two-Dimensional Gas Lattices and the Gibbs Distribution

Along the relaxation chain shown in Figure 3.2, the conditional probability associated with making a local state transition at a lattice site depends on the current level of spatial interaction within the neighborhood of the site and on whether there is another, more likely level of interaction in the neighborhood. The conditional operator must be defined in a manner that accounts for the likelihood of certain types and degrees of physical interaction in the neighborhood.

One of the simplest and most powerful models for describing spatial interaction in the field of statistical mechanics is the homogeneous gas lattice. Consider a closed system of gas molecules undergoing Brownian motion. If the gas system is closed, then the average energy E(U) of each gas molecule is constant and equal to KT, where T is absolute temperature [^{o}K], K is Boltzman's constant [$ML^2/^{o}KT^2$], and U represents the equilibrium kinetic energy of a single gas molecule. Figure 3.3 schematically illustrates a homogeneous gas lattice of M=25 such molecules, where each molecule is at one of L distinct energy states indicated by its shading. The gas molecules move independently of one another, but do transfer kinetic energy among themselves. This

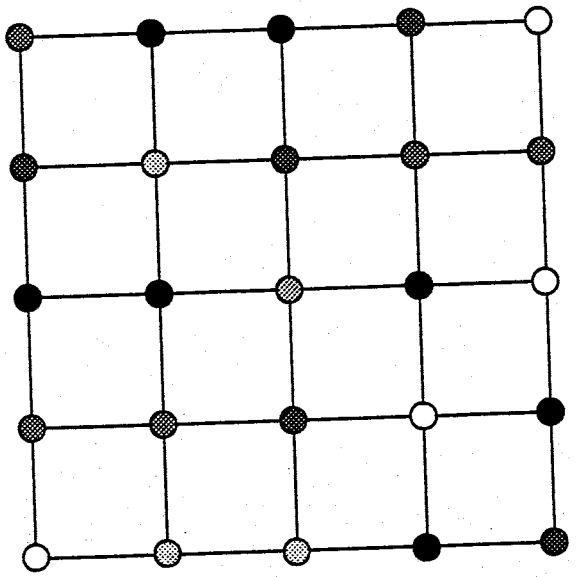


Figure 3.3. Homogeneous Gas Lattice.

situation is analogous to a random process, which behaves as a random variable at a point, but exhibits some degree of structure over space. In much the same way as there is an exchange of energy between colliding gas molecules, there is exchange of information between adjacent conductivity sites, with both processes representing manifestations of the same thing — spatial interaction. In both cases, the problem reduces to one of specifying the nature of the spatial interaction appropriately.

How do we go about defining a probability law p(U) for describing equilibrium kinetic energy of a molecule? p(U) should be least biased toward any particular value

of U, with the minimum number of 'spikes' centered about particular energies. In other words, we want p(U) to be as impartial as possible. We can do this by maximizing entropy:

$$H(U) = - \sum_{i=1}^{L} p(U_i) \ \ln \ p(U_i) \qquad [3.6]$$

where L is the number of possible outcomes, or energy states. The maximization must, however, be made subject to certain constraints. The first of these constraints:

$$\sum_{i=1}^{L} p(U_i) = 1 \qquad [3.6a]$$

requires that the sum of probabilities normalize to 1. The maximization must also be constrained by available data. For the homogeneous system of gas molecules, the average energy is fixed and equal to KT:

$$\sum_{i=1}^{L} U_i \ p(U_i) = KT \qquad [3.6b]$$

Thus, we are interested in maximizing the constrained objective function:

$$- \sum_{i=1}^{L} p(U_i) \ln p(U_i) + \eta \left[\sum_{i=1}^{L} p(U_i) - 1 \right] + \lambda \left[\sum_{i=1}^{L} U_i \, p(U_i) - KT \right] \qquad [3.7]$$

where η and λ are Lagrange multipliers for the normalization and mean energy constraints. Recasting the constrained entropy objective function as function $H_c(U)$ in continuous space yields:

$$H_c(U) = - p(U) \ln p(U) + \eta \, p(U) + \lambda \, U \, p(U) \qquad [3.8]$$

and setting the derivative of $H_c(U)$ with respect to $p(U)$ equal to 0,

$$\frac{\partial H_c(U)}{\partial p(U)} = \frac{\partial \left[- p(U) \ln p(U) + \eta \, p(U) + \lambda \, U \, p(U) \right]}{\partial p(U)}$$

$$= \frac{- p(U)}{p(U)} - \ln p(U) + \eta + \lambda \, U$$

$$= - 1 - \ln p(U) + \eta + \lambda \, U = 0 \qquad [3.9]$$

yields:

$$-1 + \eta + \lambda U = \ln p(U)$$

$$p(U) = e^{(-1+\eta+\lambda U)} \qquad [3.10]$$

Multipliers η and λ are determined by substituting $p(U)=e^{(-1+\eta+\lambda U)}$ into the two constraints given by equations [3.6a] and [3.6b], resulting in the Boltzmann, or Gibbs, distribution:

$$p(U) = \frac{e^{-U/KT}}{KT} \qquad [3.11]$$

Thus, the Gibbs distribution is the maximum prior ignorance, or maximum entropy, estimate of the probability law. It incorporates only available information about mean energy via the energy constraint into the estimate and does not assume any additional information than is contained in that physical constraint.

3.2.3 The Joint Distribution

If we wish to extend the univariate Gibbs distribution given by equation [3.11] to describe the joint, multivariate behavior of the entire collection of M gas molecules rather than a single molecule, energy levels at each of the M molecules must be

considered simultaneously, by defining a collective energy $U(\mathbf{v})$ where $\mathbf{v}$ represents the M-dimensional state vector $[v_0, v_1, v_2, \ldots, v_{M-1}]$. In addition, the Gibbs distribution must be normalized by some measure of the total energy associated with the set of all L^M possible states of the collective system, where L is the number of discrete energy states that can be assumed at any one of the M sites:

$$p(\mathbf{v}) = \frac{e^{-U(\mathbf{v})}}{\int_{\mathbf{v_p}} e^{-U(\mathbf{v})} d\mathbf{v}} = \frac{e^{-U(\mathbf{v})}}{Z} \qquad [3.12]$$

$\mathbf{v_p}$, $p=1, 2, \ldots, L^M$ is the set of all possible state vectors and $U(\mathbf{v})$ now describes the joint energy of the lattice rather than the energy at a single lattice site. Z represents a scalar normalization factor, referred to as the partition function.

The normalized multivariate Gibbs distribution [3.12] is equivalent to the joint pdf of the system energy state because it describes the probability of the entire gas lattice assuming a given lattice state $\mathbf{v}$. It has been shown that, if the joint probability law of a random process is given by equation [3.12], the associated random process is an MRF (Spitzer, 1971; Isham, 1981).

Since the converse of this is also true, in theory a realization of an MRF can be obtained by sampling a lattice state from the normalized multivariate Gibbs distribution given by equation [3.12]. Formally, an MRF is a joint pdf on the set of all possible lattice states, subject to the conditions of (1) positivity:

$$p(\mathbf{v}) > 0 \qquad \text{for all } \mathbf{v}$$

(2) Markovianity:

$$p[v_{i,j} | v_{l,m},\ (l,m) \neq (i,j)] = p[v_{i,j} | v_{l,m},\ (l,m) \in \mathcal{N}_{i,j}]$$

where $\mathcal{N}_{i,j}$ is a neighborhood system of the MRF containing all sites in the neighborhood of the lattice location (i,j) but not including site (i,j), and (3) homogeneity, where

$$p[v_{i,j} | v_{l,m},\ (l,m) \in \mathcal{N}_{i,j}]$$

is translationally invariant (Cross and Jain, 1983).

Note that, by constraining entropy maximization only on mean energy and not on second- or higher-order moments, the multivariate Gibbs distribution is free to describe any arbitrary higher-order moment behavior. Thus, unlike more traditional methods of synthesis that rely on weak stationarity, equation [3.12] defines the full joint pdf with arbitrary orders of joint behavior. This suggests that, besides permitting synthesis of spatially discrete processes, MRFs can be used to characterize natural geologic structure that exhibits significant high-order odd moments. We need not

quantify these moments explicitly — they are implicit to the Markov parameters used to define interaction energy U(**v**).

The multivariate Gibbs distribution appears to be a powerful tool for simulating discrete geologic structure. Unfortunately, it is not a simple matter to generate a sample lattice state realization from the joint distribution. Since the total mass of the joint pdf must be equal to one, the value of Z must be known if a valid probability distribution is to be defined. However, calculation of Z is generally prohibitively cumbersome for most practical problems due to the large number of possible states L^M that can be assumed by the lattice, particularly if the number of possible site states L is large. Approximation of the integral in equation [3.12] using a large subset of $\mathbf{v_p}$ is problematic, because the approximation will tend to be biased toward low energy, high-frequency states, where most of the density of the exponential Gibbs distribution is concentrated.

Figure 3.4 illustrates that even for problems characterized by small lattice size M, calculation of Z is computationally prohibitive. The process is comprised of M=9 lattice sites, with the random variable assuming only one of L=4 different states denoted by varying shades of grey. For this simple uncorrelated case, the number of possible distinct states that the entire system can assume is 4^9, or well over a quarter of a million. When spatial interaction is considered, many states become excluded because they either lack spatial correlation structure or possess the wrong order of correlation, but the total number of possible states generally remains too high to permit

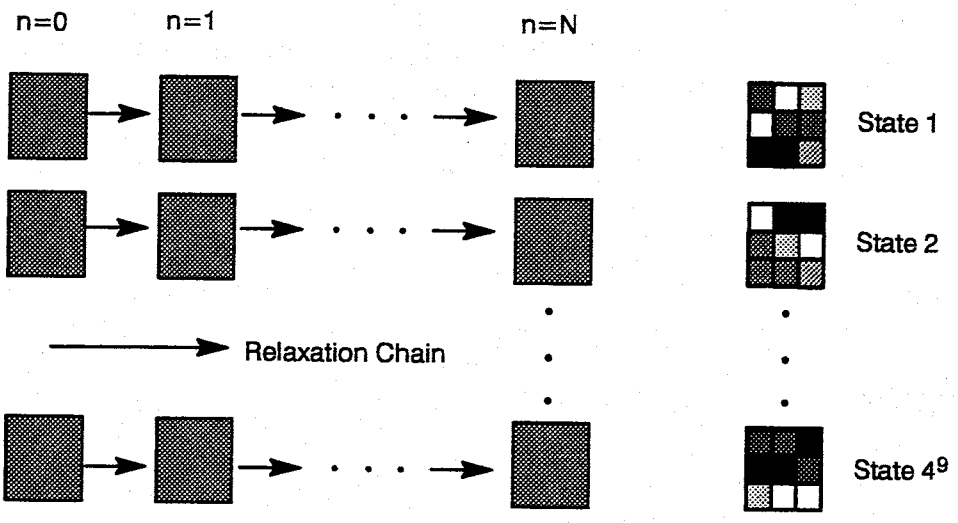


Figure 3.4. MRF States for M=9, L=4.

evaluation of the integral Z. Calculation of Z is clearly not feasible for most systems of practical interest.

3.2.4 Conditional PDF as the Joint PDF Restricted to the Neighborhood

The fact that Z cannot generally be calculated and a joint pdf defined effectively precludes development of an analytic method to generate MRF realizations from equation [3.12]. To overcome the problem of estimating Z, the normalizing factor is restricted to account only for all possible states within a small region, or neighborhood,

of a point on the lattice. This restricted form of the joint pdf is the conditional pdf:

$$p(v|.) = p[v(\mathbf{x}_0) \mid v(\mathbf{x}) \text{ at } \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)] = \frac{e^{-U'[v(\mathbf{x}_0) \mid v(\mathbf{x}) \text{ at } \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)]}}{\int_{v_p(\mathbf{x}),\ \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)} e^{-U'[v(\mathbf{x}_0) \mid v(\mathbf{x}) \text{ at } \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)]}\, dv} \qquad [3.13]$$

where $v(\mathbf{x})$ is the state at site location $\mathbf{x}$, $v_p(\mathbf{x})$, $p=1, 2, \ldots, L$ is the set of possible site states at $\mathbf{x}$, and $U'[v(\mathbf{x}_0) \mid v(\mathbf{x}) \text{ at } \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)]$ is local neighborhood interaction energy within the neighborhood. The denominator in equation [3.13] is the cumbersome normalizing factor Z restricted to the neighborhood of $\mathbf{x}_0$, $\mathcal{N}(\mathbf{x}_0)$. As discussed in the previous section, although Z has been restricted to the neighborhood it may still be intractable even for small, simple problems. This is especially true if the random variable is continuous-state, because in such a case there are an infinite number of possible neighborhood states.

The denominator of equation [3.13] is not accessible analytically because it is now restricted artificially to the neighborhood. To recast the problem in a tractable form, the ratio of conditional probabilities at two sites is used to make an informed decision about the relative strength of the conditional probabilities at both sites. This is made possible by the fact that, for a strictly stationary process, the denominator in equation [3.13] is the same in all neighborhoods, and drops from consideration when

the ratios of conditional probabilities are used to impose joint structure. The ratio is the basis for an importance-sampling method known as the Metropolis algorithm (Metropolis *et al.*, 1953). Using the Metropolis algorithm, realizations that honor the joint pdf are generated by locally and repeatedly enforcing the conditional probability law at all points in the field. If a large number of sites are visited, the chain eventually relaxes to a realization from the joint pdf. Geman and Geman (1984) refer to equation [3.13] as the Gibbs sampler because it is used to locally sample off the joint distribution.

3.2.5 Local Interaction Energy in the Neighborhood

In order to implement the importance-sampling algorithm, local interaction energy $U'[v(\mathbf{x}_0) \mid v(\mathbf{x}) \text{ at } \mathbf{x} \in \mathcal{N}(\mathbf{x}_0)]$ within a neighborhood of each of two sampled sites must be estimated. A neighborhood for a two-dimensional process is defined as follows: a site (l, m) is a neighbor of site (i, j) in a Markov sense if the conditional probability of the random variable $V_{i,j}$ depends on the random variable $V_{l,m}$:

$$p(v|.) = p[V_{i,j} = v_{i,j} \mid V_{l,m} = v_{l,m},\ (l,\ m) \neq (i,\ j)] \qquad [3.14]$$

The neighborhood sites (l, m) are, in general, in close physical proximity to site (i, j), but this need not necessarily be the case.

We can establish the MRF neighborhood system as follows. If $X_M = \{(i, j) : 1 \leq i, j \leq M\}$ represents a two-dimensional M X M lattice system and $V = \{V_{i,j},\ (i,j \in X_M\}$ is the lattice process, then the neighborhood of lattice sites (i, j) is defined as a subset

of lattice sites in X_M, $\mathcal{N}_{i,j}$, $(i,j) \in X_M$. Since the random variable at site (i, j) depends only on the random variable at points in its neighborhood, the conditional probability statement for the Markov random field becomes (Geman and Geman, 1984):

$$p(v|.) = p[V_{i,j} = v_{i,j} \mid V_{l,m} = v_{l,m}, \ (l, m) \neq (i, j)]$$

$$= p[V_{i,j} = v_{i,j} \mid V_{l,m} = v_{l,m}, \ (l, m) \in \mathcal{N}_{i,j}] \qquad [3.15]$$

The probability p(v|.) must be assessed according to the Gibbs distribution specified in equation [3.13], but first we must assess local neighborhood energy potentials U'(v|.). They can be defined as the sum of potentials over the neighborhood system $\mathcal{N}_{i,j}(v)$, with each potential weighted by a unique parameter:

$$U'_{i,j}(v|.) = U'[v_{i,j}|v_{l,m} \text{ at } (l,m) \in \mathcal{N}_{i,j}] = \sum_{N_{i,j} \in \mathcal{N}_{i,j}} w_{N_{i,j}}(v)\, u_{N_{i,j}}(v) \qquad [3.16]$$

where $u_{Ni,j(v)}$ is the portion of local interaction energy associated with neighborhood subset $N_{i,j}$, $w_{Ni,j(v)}$ is the weight corresponding to energy contribution $u_{Ni,j(v)}$, and $N_{i,j}$ is a subset of all neighborhood system $\mathcal{N}_{i,j}(v)$. Weights are considered spatially uniform in a stationary MRF, and are referred to as Markov parameters. Under the conditional form of the problem given by equations [3.13] and [3.16], energy perturbations at sites located outside of the neighborhood in no way alter the distribution of energy within

the neighborhood. As specified by the neighborhood system $\mathcal{N}_{i,j}$ and the Markov parameters $w_{Ni,j}(v)$, $U'(v|.)$ uniquely defines the conditional operator [3.13] and the Markov random process.

Note that the local nature of the conditional energy $U'(v|.)$ presents a fundamental problem with applying MRF theory to practical problems: one does not usually know *a priori* how local energies will become manifested globally in the realization. The evolution of global structure is not generally predictable from the form of the local neighborhood energy function or from the neighborhood configuration. Hence, one often has no means of predicting and thus controlling the ultimate global, joint structure. This is the price paid in going from a joint to a conditional formulation — a more tractable problem is derived, but it occurs at the expense of not knowing how locally-specified parameters will become manifested globally as the energy propagates through the realization during relaxation. The inaccessible nature of the relationship between local and global parameters is the inherent drawback associated with the MRF approach to random field generation. However, the problem of reproducing joint behavior through locally specified parameters usually presents few problems of a practical nature because data are naturally sparse and joint behavior not generally observable anyway. That is to say, even if one could sample directly off the joint pdf, the moments of the pdf cannot be determined from the limited data typically available in most practical field problems.

3.2.6 Markov Parameters

Positivity requirements for p(v|.) will automatically be fulfilled in equation [3.13] as long as U'(v|.) is real. Therefore, U'(v|.) can be defined using arbitrary neighborhoods and parameters. Figure 3.5 shows an example of a neighborhood system for a two-dimensional, third-order field. A typical third-order pair-potential interaction energy might be of the form:

$$U'[v_{i,j} \mid v_{l,m},\ (l,m) \in \mathcal{N}(i,j)\ ] = \alpha\ [\ |v_{i+1,j} - v_{i,j}\ | + |v_{i-1,j} - v_{i,j}\ |\ ]$$
$$+ \beta\ [\ |v_{i,j+1} - v_{i,j}\ | + |v_{i,j-1} - v_{i,j}\ |\ ] + \gamma\ [\ |v_{i+1,j-1} - v_{i,j}| + |v_{i-1,j+1} - v_{i,j}|\ ]$$
$$+ \eta\ [\ |v_{i+1,j+1} - v_{i,j}\ | + |v_{i-1,j-1} - v_{i,j}\ |\ ] + \mu\ [\ |v_{i+2,j} - v_{i,j}| + |v_{i-2,j} - v_{i,j}|\ ]$$
$$+ \xi\ [\ |v_{i,j+2} - v_{i,j}\ | + |v_{i,j-2} - v_{i,j}\ |\ ] \qquad [3.17]$$

where the energy of interaction is proportional to the sum of energy flows between the site and each site in its neighborhood, and α, β, γ, η, μ, and ς are Markov parameters.

Markov parameters reflect the degree of local energy interaction, and are essentially equivalent to local expressions of Lagrange multipliers in the maximum entropy formulation used to derive the Gibbs distribution. While positive Markov parameters encourage growth of correlation structure through attraction, negative parameters inhibit it through repulsion; positive parameters become manifested as clustering of like values and negative parameters produce a checkerboard effect caused by repulsion of like values. Control of the sizes of geologic features is made possible

through the use of combined repulsive-attractive effects. This control is effected by specifying low-order parameters to be positive in order to enhance growth of the features, with higher-order parameters made negative in order to inhibit further growth of the features. A zero-valued parameter implies lack of either positive or negative correlation. This can be seen by substituting $\alpha=\beta=\gamma=\eta=\mu=\xi=0$ into the pair-potential function equation [3.17] and noting that the conditional pdf equation [3.13] predicts equal likelihood of all random variables at a point regardless of the values at the surrounding neighbors, giving rise to a field of uncorrelated structure. Anisotropy can be incorporated into the field by specifying different Markov parameters associated with different directions of energy exchange.

For purposes of calculating local potential energy, boundary sites are treated somewhat differently than interior sites because their natural neighborhoods are

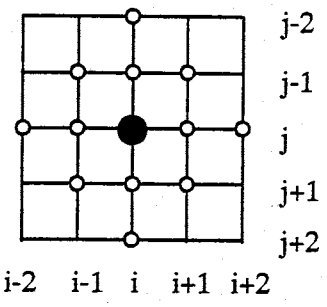


Figure 3.5. Third-Order Neighborhood System for Calculating Pair-Potential Interaction Energy.

truncated at the boundary. In statistical mechanics applications, edge effects are traditionally handled by repeating the field at its boundaries to produce a toroidal structure and duplicate the effects of an infinite field, but this procedure introduces a periodicity to the field that may not generate realistic geologic structure. Geman and Geman (1984) suggest simple alteration of the neighborhood at the field edges to induce the free-boundary effect used for this study. For the free boundary, it is assumed that there is zero energy interaction across the boundary by allowing the boundary site to depend only on values located at interior points and not on any conditions on the exterior side of the boundary. A boundary point is thus considered like any point, except that it has fewer neighbors than interior points.

One can even introduce nonlinear terms into the potential energy function equation [3.17] or use energies involving interactions among cliques of three or more sites, where a clique is a subset of neighborhood sites for which every pair of sites are neighbors. For example, when pair-potentials are squared, weighted, and summed, second-order interaction is imposed and a gaussian-distributed MRF produced. In the third-order neighborhood shown in Figure 3.5, a clique may involve site (i,j) and the three adjacent sites (i+1,j), (i,j−1), and (i+1,j−1) which are themselves adjacent to one another in a Markovian sense. U'(v|.) would then be defined as the weighted sum of potentials among all sites in the clique. By increasing the order of the neighborhood system and the complexity of interaction among neighborhood sites, the model can be made more complex.

3.2.7 Importance Sampling Metropolis Algorithm

As mentioned in section 3.2.4, analytic synthesis of MRF realizations is not possible because of the local, artificially restricted nature of the conditional operator. Instead, a combinatorial simulation approach must be adopted to generate MRF structure. Figure 3.6 illustrates how the Metropolis importance sampling algorithm is used to impose spatially-discrete correlation. Farmer (1987) likens the Metropolis site-replacement algorithm to random walk sampling of a multivariate pdf over the space of the random variable vector $\mathbf{v}(\mathbf{x})$.

Like more traditional gaussian random process generators, the Metropolis algorithm constructs a new correlated process from an old uncorrelated one. The method involves making local changes in the process through exhaustive sampling and site exchange until the correct global properties evolve and an MRF realization generated, driven by a conditional operator that represents the primary rule of re-ordering. We must sample and exchange many times in order to overcome the local implications of the conditional pdf, for which the set of random variables contained in a neighborhood constitutes only a partial realization of the joint structure. The decision to exchange is made on the basis of the ratio of conditional probabilities estimated at two sites. Since the final steady-state joint pdf will have no memory of the initial process, any base process can be used to initialize the chain. However, the rate of convergence will depend on the initial configuration (Cross and Jain, 1983).

The specifics of the importance-sampling algorithm are as follows. A pair of sites in the old state are randomly visited at lattice locations $\mathbf{x}_1$ and $\mathbf{x}_2$ and conditional

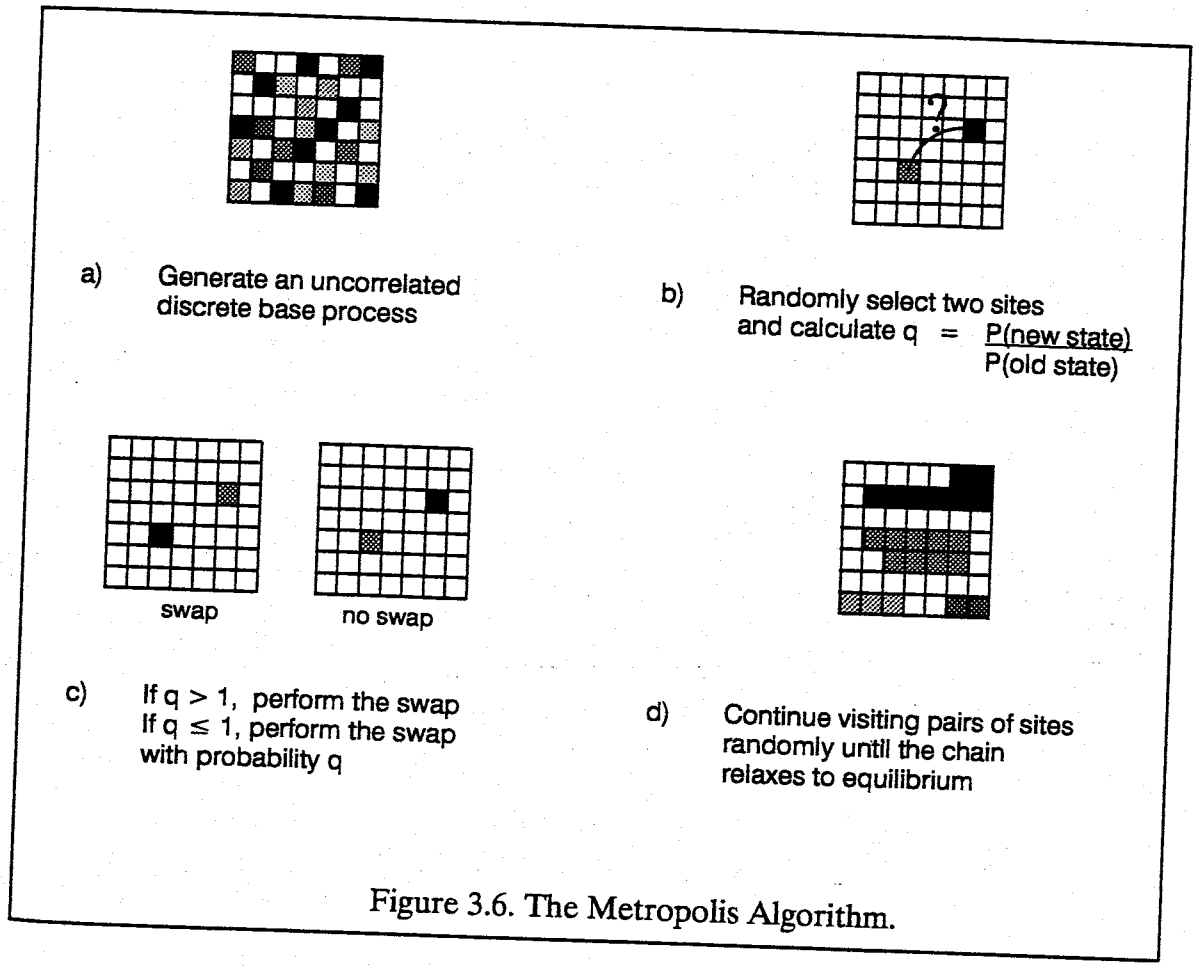


Figure 3.6. The Metropolis Algorithm.

probabilities estimated using the neighborhoods of each pair between which a potential exchange will occur. The ratio q of the conditional probability associated with the new state to that of the old state is determined as:

$$q = \frac{p[\mathbf{v}_{new}]}{p[\mathbf{v}_{old}]} = \frac{p[v_{new}(\mathbf{x}_1) \mid .]\; p[v_{new}(\mathbf{x}_2) \mid .]}{p[v_{old}(\mathbf{x}_1) \mid \cdot]\; p[v_{old}(\mathbf{x}_2) \mid \cdot]} = \frac{e^{-U'[v_{new}(\mathbf{x}_1)|.]-U'[v_{new}(\mathbf{x}_2)|.]}}{e^{-U'[v_{old}(\mathbf{x}_1)|.]-U'[v_{old}(\mathbf{x}_2)|.]}} = e^{-\Delta U'} \qquad [3.18]$$

where q is the ratio of the new-state probability to that of the old state. $\Delta U'$ is the difference in interaction energy between new and old states and is a measure of the degree to which the new state is more likely than the old. When either of the selected sites is a data conditioning point, the pair is simply ignored and sampling continued. Note that q is the ratio of conditional probabilities given by equation [3.13]. For a stationary MRF process, the unknown integral in the denominator drops completely from the ratio.

If q is greater than 1, $\Delta U'[v(x_0)| .]$ is less than zero, the new state has lower energy and more organization than the old state, and a swap is made because such an exchange results in a more likely lower-energy state. If q is less or equal to 1, then the swap is performed with probability q to allow occasional escape from a local energy minimum caused by spurious structure in the initial process. Inclusion of this option for escape from local energy minima prevents the MRF realization from being ordered just because local order happened to be artificially present in the base process. From a physical perspective, this procedure is analogous to adding erratic energy to the field in order to enter into a higher-energy state, making more subsets of lower-energy states accessible from the current state and honoring chain ergodicity requirements. Note

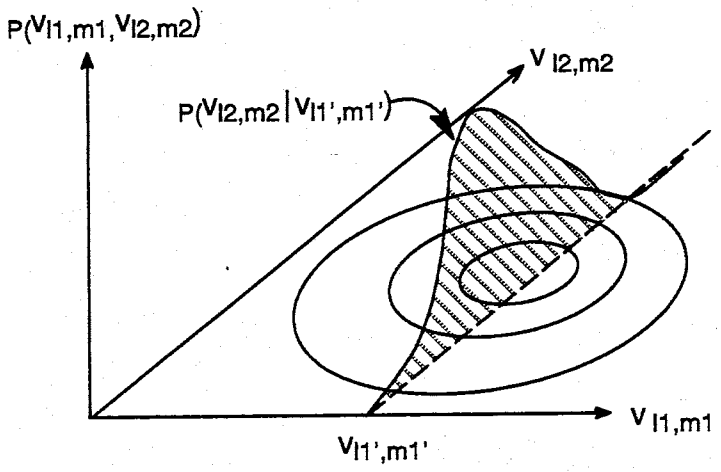
that, without the addition of this energy, lower energy states are always selectively and exponentially favored. Typically, importance sampling continues until the number of successful replacements drops below some percent of total attempts.

The Metropolis algorithm is similar to simulated annealing, except that we do not necessarily desire the most likely structure — rather, we seek only a *possible* structure. It allows us to avoid the problem of simulating states uniformly off the joint Gibbs pdf which, depending on the form of the local energy potential U'(v|.), may almost completely exclude unlikely, high-energy states as discussed in section 3.2.3. Rather than sampling unequally-probable states uniformly off the joint pdf, the Metropolis algorithm samples them nonuniformly from the state space. As a result, even the most improbable realization has a chance of being represented in the ensemble.

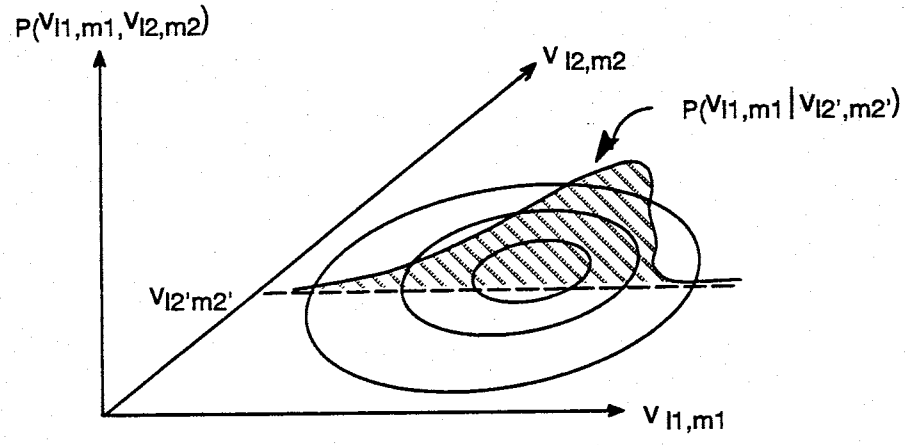
To obtain an understanding of how repeated application of the conditional operator produces a sample from a joint pdf, consider the joint pdf as a product of sequentially conditional probabilities over the entire process at lattice sites (i,j):

$$p(v) \equiv p[v_{i,j} |\ v_{l,m},\ (l,m) \in \mathcal{N}_{i,j}]\ p[v_{l,m},\ (l,m) \notin (i,j)] \qquad [3.19]$$

The equivalence is approximate because there are dependencies not accounted for in the above relation. Figure 3.7 illustrates conceptually how a two-dimensional joint pdf can be factored into the product of two conditionals if the conditional distributions are independent of one another. The approximation is exact if the contours of the joint pdf



a) Conditional Probability of $v_{l2,m2}$ Given $V_{l1,m1}$



b) Conditional Probability of $v_{l1,m1}$ Given $V_{l2,m2}$

Figure 3.7. Factorization of Joint Pdf Into Component Conditional Pdfs.

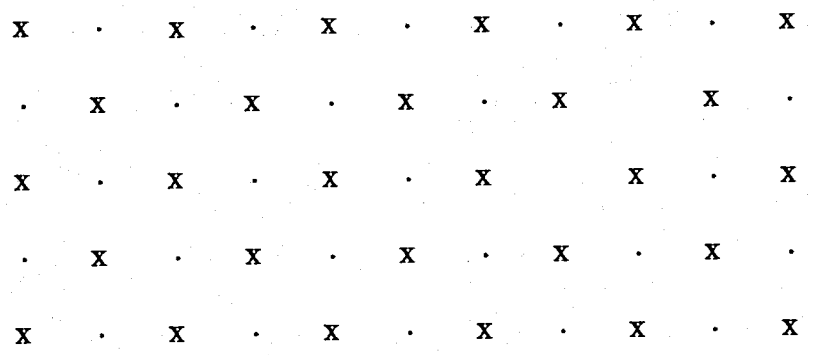
are concentric circles, which is indicative of statistical independence. When the contours are ellipses oriented obliquely to the random variable coordinate axes, the two variables possess some dependence and the degree of approximation associated with equation [3.19] will degrade with increasing eccentricity of the ellipses. With each iteration of the importance sampling algorithm, yet another term is incorporated to the factorization. Such factorization makes it possible to reduce what is essentially an intractable global, absolute problem into a series of manageable local, relative ones.

3.2.8 Maximum Likelihood Estimation of Markov Parameters

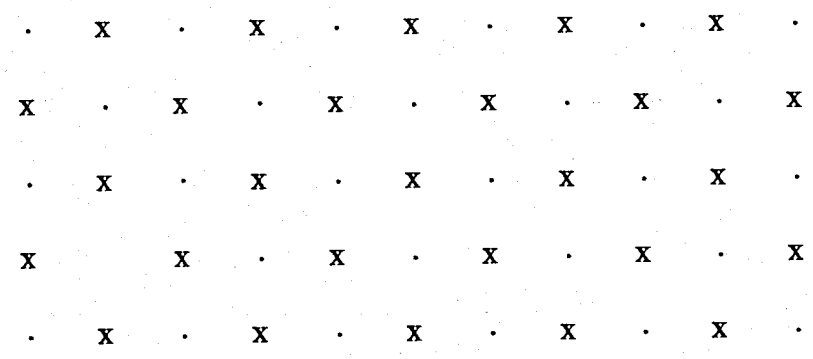
Our interest lies in generating discrete MRF realizations that behave, in a statistical sense, like observed geologic structure. Two final questions remain: how are Markov parameters estimated from geologic data? and does an MRF model with the optimal parameters adequately describe the observed geologic structure? Due to the combinatorial nature of the method used to generate MRF realizations, the issue of Markovianity and the estimation of parameters can only be approached by again resorting to simulation techniques. Since simulation of an MRF requires as input the Markov parameters, we must first estimate these parameters as if the field were an MRF by assuming a specific functional form of $U'(\mathbf{v}|.)$. Methods of hypothesis testing, in which independent subsets of the observed and theoretically predicted processes are compared, can then be used to assess the likelihood that the observed process is not only an MRF of the assumed order and local energy function, but also characterized by the estimated parameters.

Since the optimal Markov parameters are those most likely to describe the observed MRF, maximum likelihood (ML) techniques can be used to estimate the true parameter set. The Markov parameters are defined in the context of conditional probabilities and must be estimated from observed data that has been divided, or coded, into mutually orthogonal sets of lattice sites. With locations in a given coding independent of one another in a Markovian sense, the conditional probabilities can be used to form a product that approximates the joint density of the process as given by equation [3.19]. This approximation is then maximized over the Markov parameter space for all coded samples to yield the most likely set of parameters to have produced the observed data. Besag (1972) was the first to use coding and conditional maximum likelihood techniques as a means of MRF parameter estimation, pointing out their obvious intuitive appeal.

Figure 3.8 illustrates how one would go about coding a first-order realization of an MRF, where each value depends on the values at the four surrounding sites immediately to the north, east, south, and west. If the observed data are considered to be from a first-order MRF realization, the two coding schemes shown in Figure 3.8 represent independent locations where the sites in any given coding are not within the neighborhoods of one another in a Markovian sense. In a similar manner, one could estimate parameters for a second-order MRF by coding the field into four disjoint sets of lattice sites. For a lattice of M sites with k codings, there are M/k independent samples of the conditional probability (Cross and Jain, 1983). Following coding into independent sets, each set is analyzed separately for maximum-likelihood parameters.



First Coding



Second Coding

Figure 3.8. Mutually Orthogonal Codings for a First-Order MRF.

The codings commonly produce very different and potentially correlated estimates of the Markov parameters. This correlation stems from the fact that any single piece of data is used more than once. For example, in a first-order process, each data value is used four times as a neighbor and once as the conditioning value. Nevertheless, these sets of estimates are typically averaged to define representative parameters.

The maximum likelihood procedure amounts to unconstrained maximization of a nonlinear likelihood function over the Markov parameter space. The set of parameters which define the optimal likelihood are those that are most likely to occur, given their implicit manifestation through the behavior of the data. From a conditional probability standpoint the likelihood product, defined as $\ell'(k)$, is approximately equal to the product of all conditional probabilities in a given coding k:

$$\ell'(k) \approx \prod_{\mathbf{X}(k)} p[\, v(\mathbf{x}) \mid . \,] \qquad [3.20]$$

where the product is taken over all locations $\mathbf{X}(k)$ in the kth lattice coding. Note that the denominator of conditional probability $p[v(\mathbf{x})|.]$ is Z restricted to the neighborhood and, for a stationary process, is the same for each term in the likelihood product. As a result, the denominators in the product do not affect the maximization.

To facilitate analysis, the logarithm of the conditional likelihood $\ell'(k)$ is determined:

$$\mathcal{L}(k) = \ln \mathcal{L}'(k) \approx \sum_{\mathbf{X}(k)} P[v(\mathbf{x}) \mid . \,] \qquad [3.21]$$

Since $\mathcal{L}(k)$ is a monotonically increasing function of $\mathcal{L}'(k)$, maximization of $\mathcal{L}(k)$ is equivalent to maximization of $\mathcal{L}'(k)$. Furthermore, objective function $\mathcal{L}(k)$ is in the form of a sum, which facilitates derivation of gradient and Hessian functions as input to a modified-Newton nonlinear gradient search algorithm. And finally, since $\mathcal{L}'(k)$ is always between 0 and 1, the logarithmic transformation serves to stretch the objective function along the likelihood axis and makes it easier to find the optimum. After substituting the form of the conditional probabilities given by equation [3.13] and the values of the process at the coded lattice locations, $\mathcal{L}(k)$ becomes a function of the neighboring values of $v(\mathbf{x})$ and the Markov parameters. The Markov parameters are determined by maximizing log likelihood over the parameter space.

There are several drawbacks to the ML estimation of MRF parameters. Unfortunately, due to the local nature of the parameters and their estimation, the parameter estimation methods are unavoidably sensitive to even the smallest of nonstationarities. This might explain why it is notoriously difficult to obtain reliable parameter estimates in a small domain, where conditional probability outliers may dominate the analysis. Besag (1974) also notes that the coding techniques are not fully efficient because only a portion of the data is used. For example, in first-order codings only 50% of the data is used for each test. However, he believes that this lack of full efficiency is more than offset by the simplicity and flexibility of the coding techniques

and maximum-likelihood parameter estimation. Finally, Besag (1974) suggests that one maximize the pseudo-likelihood, equal to the product of conditional probabilities over the entire field without regard to coding, because this method produces parameter estimates that are consistent for increasing field size.

3.2.9 Goodness-of-Fit Hypothesis Testing

It is critical to determine whether maximum likelihood parameters do, in fact, characterize the overall structure of the observed MRF. Goodness-of-fit testing represents a means of objectively determining how well the MRF model reproduces the overall character of the observed process. Since one must simultaneously determine the Markov parameters and neighborhood order when performing parameter estimation, a poor fit between MRF predicted structure and observed structure may indicate the presence of nonstationarities in the form of conditional outliers, an incorrectly guessed neighborhood order, violation of the assumption of Markovianity, or some combination of these problems. Under conditions of nonstationarity, Markov parameters estimated via a maximum likelihood approach may, in a global sense, be the most likely but may not represent the parameters adequately in a local sense. If the field is stationary and thus described by a single unique set of parameters, but no reasonable neighborhood order results in acceptance of the hypothesis that the field is an MRF, then one may be forced to conclude that the data is not a sample from an MRF but from some other stationary process.

χ^2 goodness-of-fit testing can be used to determine how well the MRF parameters actually describe the observed realization, as well as whether the observed realization is of the assumed MRF order. The hypothesis tested is that coded samples are drawn from the conditional pdf given by equation [3.13], with the same neighborhood and interaction energy function as was assumed for ML parameter estimation. The null hypothesis is that the samples are not drawn from the conditional pdf with the assumed neighborhood and interaction energy. Given that χ^2 testing involves comparison of observed and theoretical independent conditional frequency distributions to assess goodness-of-fit, we must again encode the lattice into mutually orthogonal subsets. χ^2 testing on conditional frequencies proceeds as follows:

- In each coding, every point and its uncoded neighbors represent an independent sample from the conditional probability $p[v(x)|\ .]$. Observed conditional frequencies are tabulated for all possible combinations associated with the site value and its neighboring values, referred to as conditional frequency classes.

- The expected conditional frequencies F_i , $i=1, 2, \ldots, N_F$ are determined, where N_F is the number of conditional frequency classes, using Markov parameters estimated from the maximum likelihood analysis.

- The expected conditional frequency distribution is compared to the observed conditional frequency distribution F_{oi}, $i=1, 2, \ldots, N_F$. If the χ^2 statistic for the appropriate degree of freedom indicates acceptance of the hypothesis that the two distributions are statistically equivalent, then the observed realization is accepted as a sample from an MRF characterized by the estimated parameters and the assumed neighborhood order. Otherwise, the process is not Markovian, not of the assumed order, or simply not statistically homogeneous.

Cross and Jain (1983) recommend that since test results from the codings are not independent, the significance level be assumed equal to $k\phi$ rather than α^k, where α is the significance level applied simultaneously for all of the k codings and ϕ represents the level at which the greatest significance occurs in the k tests. The former statistic is more likely lead to rejection when this is done, and therefore results in a more conservative acceptance rule. Note that large χ^2 test statistics may result if the sample is small because small samples can produce very low conditional frequencies in some of the classes. In the extreme case, if zero conditional frequencies occur, the positivity condition is violated. Such a situation would be especially likely for high-order fields for which the size of the domain is small.

3.3 APPLICATION OF MRFs TO CHARACTERIZATION OF DISCRETE GEOLOGIC STRUCTURE

3.3.1 Evolution of a 64-By-64 MRF Realization

Figure 3.9 is an example of how a 64-by-64 MRF realization evolves along a relaxation chain. Parameters of the second-order MRF process were assumed to be α=1.0, β=0.5, γ=1.0, η=0.5, μ=0.0, ξ=0.0 and an interaction energy of the form provided by equation [3.17] specified. Lattice site states were permitted to assume only the values −3.0, −4.0, −5.0, −6.0, and −7.0. From the figure, it is clear that relaxation is a highly nonlinear process — much more reordering occurs early in the chain because little structure exists and marginal information gained from additional importance

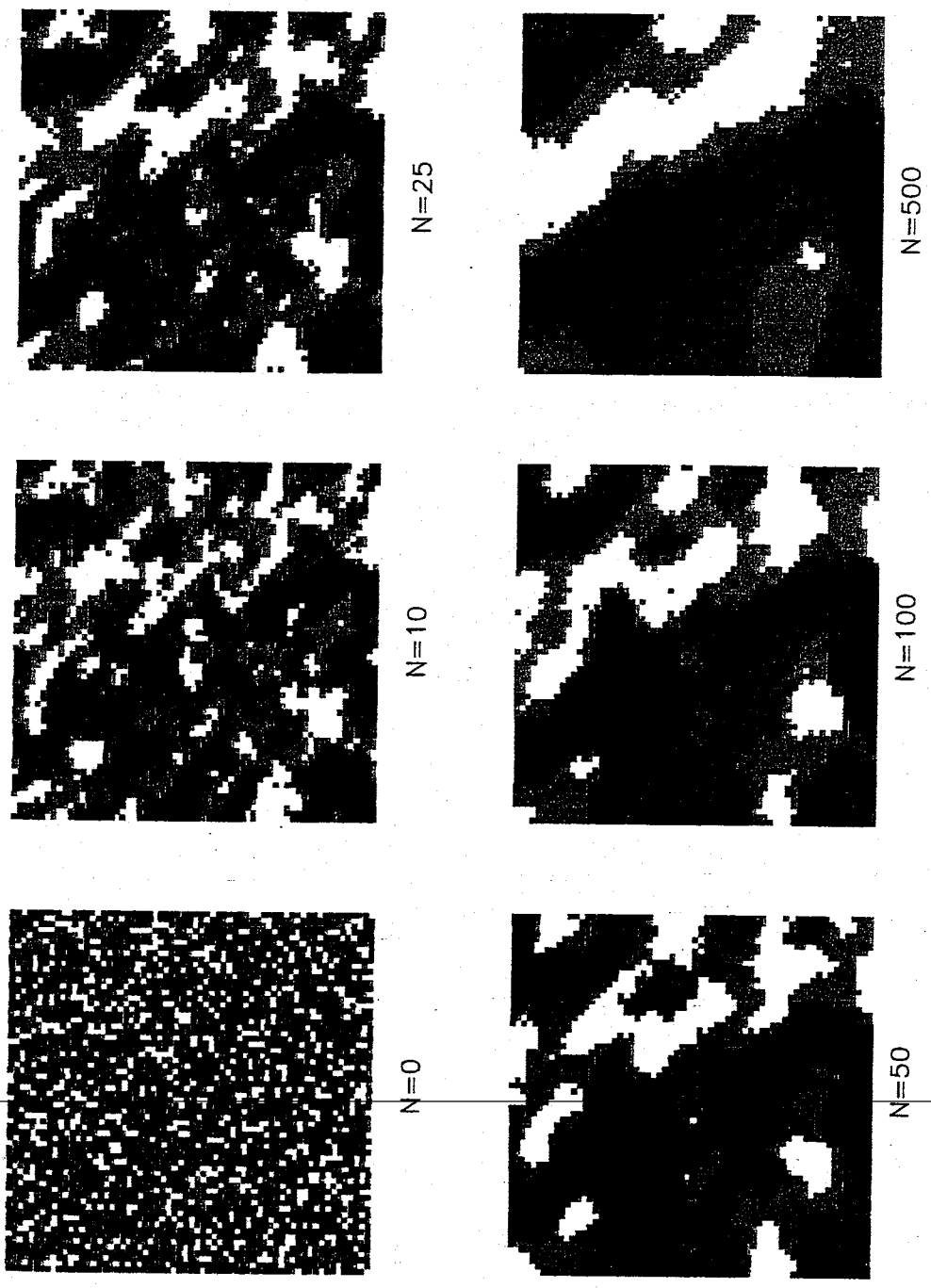


Figure 3.9. Evolution of an MRF Realization from a Relaxation Chain.

sampling is very valuable, introducing significant added structure to the process. As the chain progresses and energy becomes more organized, acquired information becomes more redundant and less useful in terms of increasing MRF structure. Farmer (1989) reports that use of the Metropolis method for simulated annealing requires on the order of 100 attempted or 10 successful replacements per site for convergence to the most likely state. Since the theoretical rate of convergence for large N is of order $\sqrt{N}$, this slowdown in convergence rate is to be expected (Geman and Geman, 1984). Shown in Figure 3.10 is percent successful exchanges vs. attempted exchanges N for the first-order 64-by-64 MRF relaxation chain. It is evident from the plot that the rate of successful exchange approaches $1/\sqrt{N}$ for large N. Likewise, the error of estimate for moments of the joint pdf is $O(N^{1/2})$ (Hammerseley and Handscomb, 1964).

3.3.2 Estimation of MRF Parameters for Alluvial-Fan Structure

To illustrate how useful MRF models can be for describing realistic geologic structure, a discrete MRF model was used to characterize two-dimensional alluvial-fan structure underlying Kirtland Air Force Base in the Albuquerque Basin, central New Mexico. Lithofacies-type data became available when a six-mile long sewer line trench was excavated into the fan deposits in 1991 (see Figure 3.11). Six distinct lithofacies types (L=6) were defined along the walls of the excavation (Gaither *et al.*, 1993; Crowson *et al.*, 1993). Figure 3.12 shows the distribution of lithofacies type along roughly 3.5 miles in the main stem of the trench, assumed to be a linear feature for purposes of data analysis and presentation. It is evident that the top of the trench tends

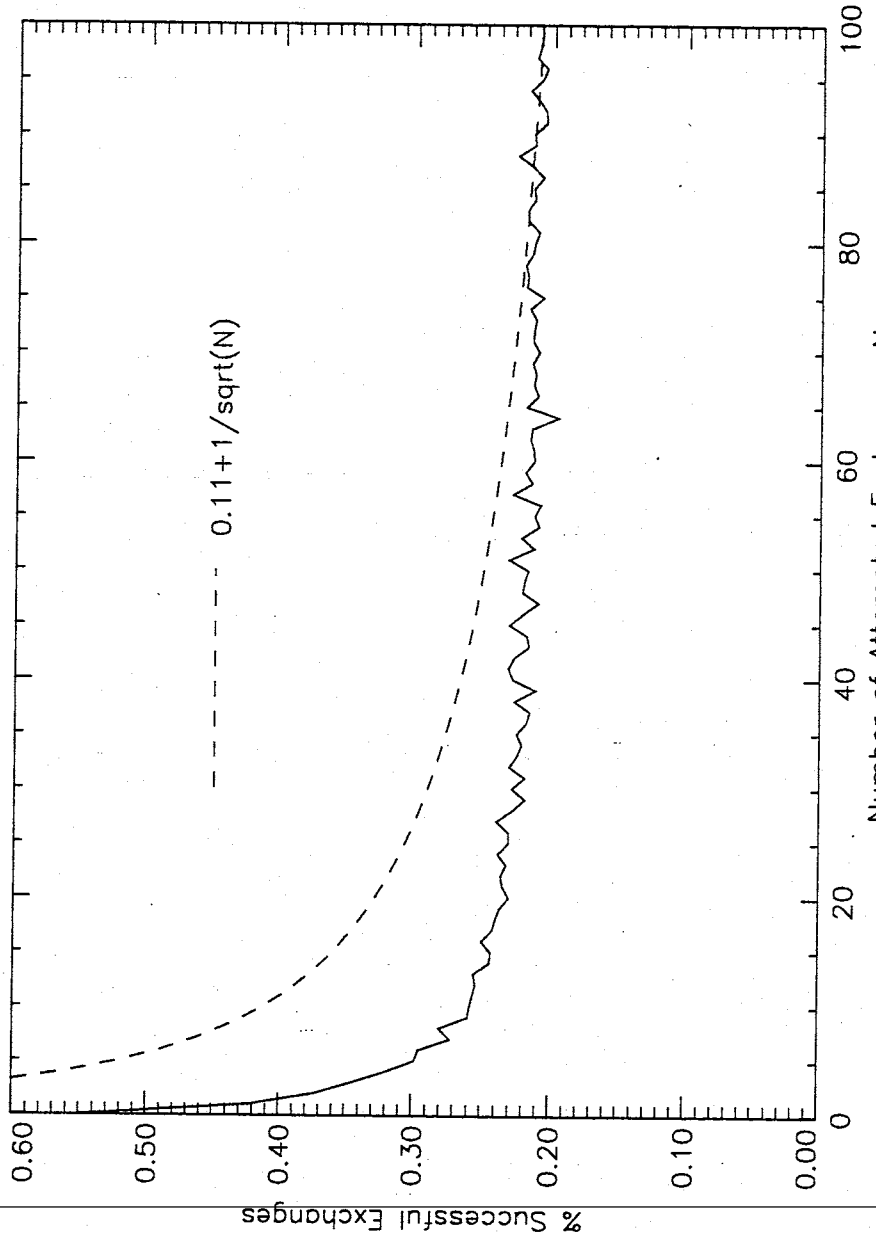


Figure 3.10. Percent Successful Exchanges vs. Attempted Number of Exchanges N for First-Order 64-By-64 MRF Realization.

to show a spatially-uniform lithology indicative of surficial weathering, while the middle and bottom portions exhibit more heterogeneous behavior.

To estimate MRF parameters for the fan deposits, a maximum-likelihood approach was used. The ML algorithm relies on mutually-orthogonal codings of the lattice sites, which depend on the order of the neighborhood assumed. It was initially assumed that the observed lithofacies distribution was a realization from a second-order MRF, with pair potential interaction energy given by equation [3.17] for $\mu=\zeta=0$. However, when second-order ML parameters corresponding to oblique structure γ and η were found to be consistently close to 0.0 compared to the first-order parameters α and β, a first-order MRF was instead adopted and the number of sites in a neighborhood, including the coded site itself, reduced from 9 to 5. This assumption considerably decreased the computational effort involved in subsequent χ^2 goodness-of-fit testing, reducing the number of conditional frequency classes from approximately $L^M=6^9=10{,}077{,}696$ to roughly $6^5=7776$.

The likelihood function was defined according to equation [3.20] using a discrete summation approximation of conditional probability given by equation [3.13]:

$$\mathcal{L}'(k) \approx \prod_{i=1}^{M(k)} p(v_i \mid .) = \prod_{i=1}^{M(k)} \frac{e^{U'(i)}}{\sum_{l=1}^{L} e^{U'(i,\,l)}} \qquad [3.22]$$

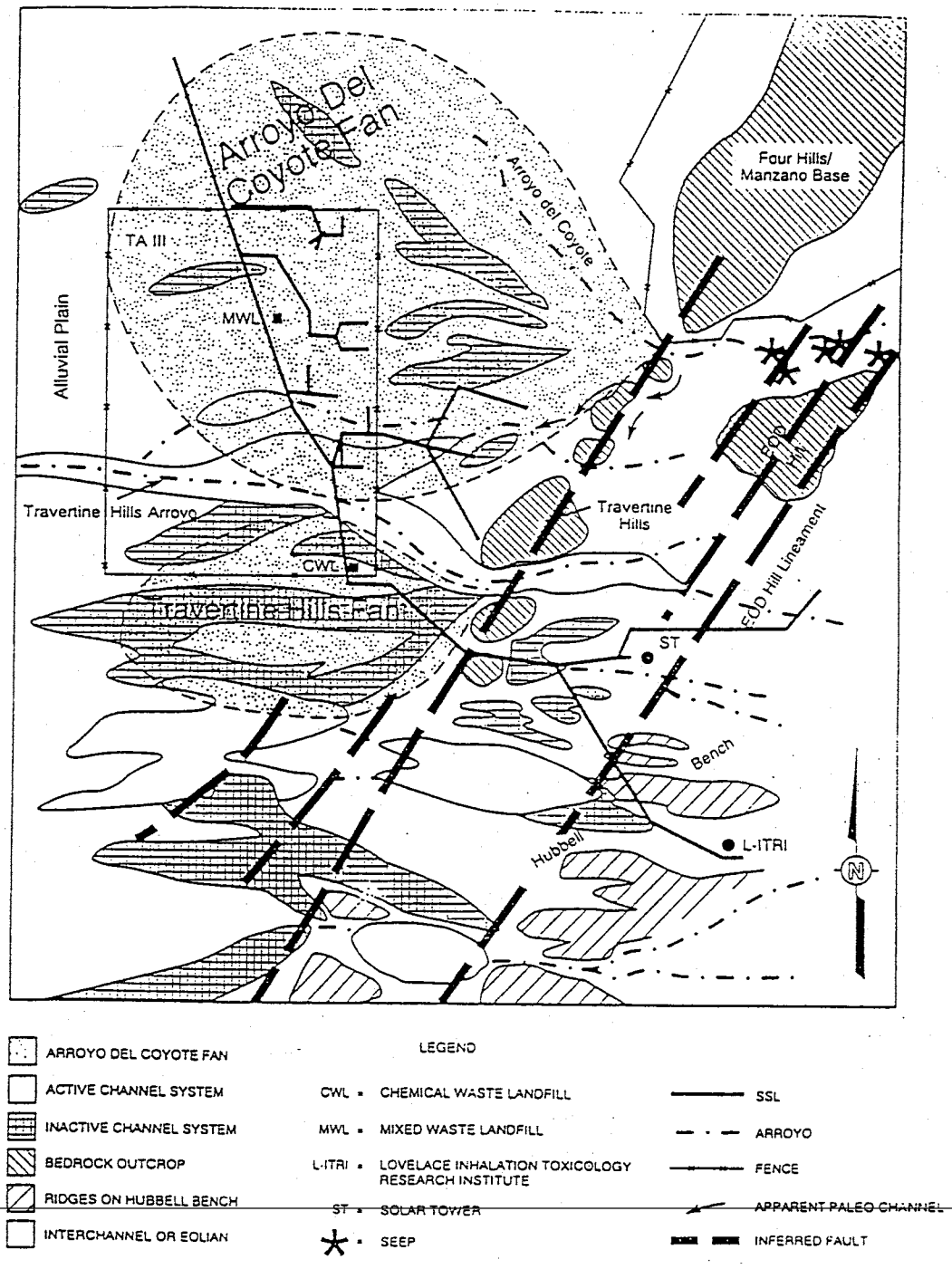


Figure 3.11. Sewer-Line Trench Excavation at Kirtland Air Force Base (courtesy of Sandia National Laboratories).

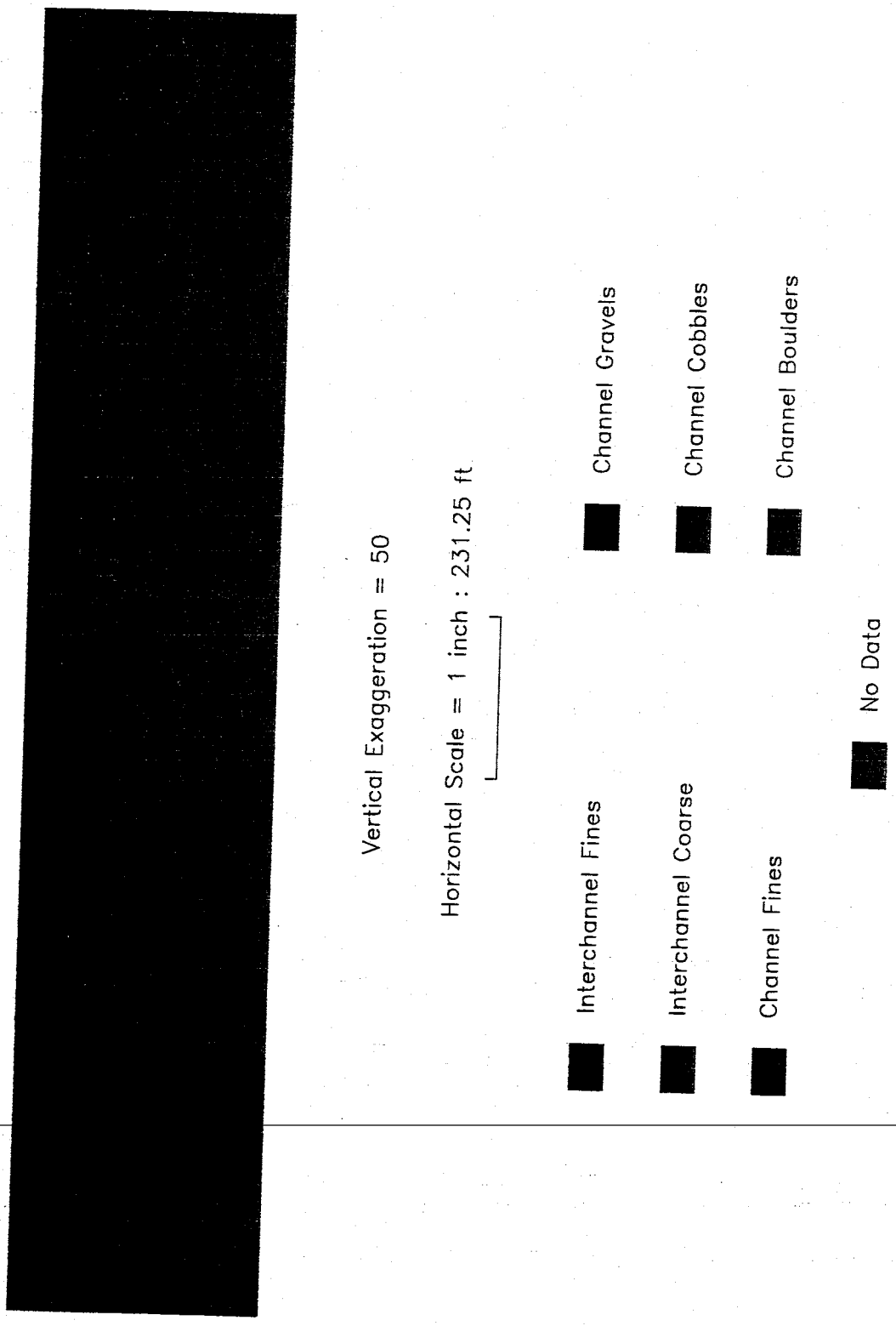


Figure 3.12. Observed Lithofacies Type Distribution Along Main Stem of Sewer Trench.

where U'(i) is the interaction energy based on the current state at the ith coded site, U'(i,*l*) is the interaction energy at the ith coded site if state level *l* were substituted at the site, L is the total number of possible state levels at each site, and M(k) is the total number of coded sites for the kth coding. Taking the logarithm of $\mathcal{L}(k)$ yields:

$$\mathcal{L}(k) = \ln \mathcal{L}'(k) \approx \ln \left[\prod_{i=1}^{M(k)} \frac{e^{U'(i)}}{\sum_{l=1}^{L} e^{U'(i,l)}} \right]$$

$$= \sum_{i=1}^{M(k)} \ln \frac{e^{U'(i)}}{\sum_{l=1}^{L} e^{U'(i,l)}} = \sum_{i=1}^{M(k)} \left[U'(i) - \ln \sum_{l=1}^{L} e^{U'(i,l)} \right]$$

[3.23]

Since $\mathcal{L}(k)$ is a nonlinear objective function, both first- and second-order derivatives with respect to the parameters α and β were required to perform a modified-Newton gradient search. These are shown derived in Appendix B.

Equations [B.1], [B.2], and [B.3] presented in Appendix B form the basis for separate modified-Newton gradient searches conducted over the log of the likelihood function for each independent coding of the assumed first-order MRF. Three different forms of pair-potential interaction energy were tested. One was based on absolute values of differences between lithology indices as given by equation [3.17], another

involved squaring the absolute differences, and the third was defined using the square root of the absolute differences.

Shown in Figure 3.13 are the log-likelihood functions for each tested form of the interaction energy and for each of two independent, first-order codings. From inspection of the log-likelihood plots, it is evident that parameters estimated for the same energy function but different codings agreed closely. Although optimal α and β were similar between two of the three models, the log-likelihood function tended to be more pronouncedly peaked for the square-root interaction energy model. Note that, for each form of the pair-potential interaction energy, the maximum-likelihood parameter estimation procedure identified the most likely parameters given the assumptions that the MRF is order 1, stationary, and with interaction energy of the assumed form, but does not provide any indication of how appropriate those assumptions are.

3.3.3 χ^2 Goodness-of-Fit Testing

To address this issue, χ^2 goodness-of-fit tests were conducted to determine how closely the observed realization of trench lithologies conformed to a first-order MRF model for each form of the interaction energy. In each case, the null hypothesis was that the observed lithofacies spatial distribution was drawn from a first-order MRF with the assumed form of the interaction energy. To this end, conditional frequencies associated with the observed facies-type realization presented in Figure 3.12 were compared to expected conditional frequencies. Expected conditional frequencies were

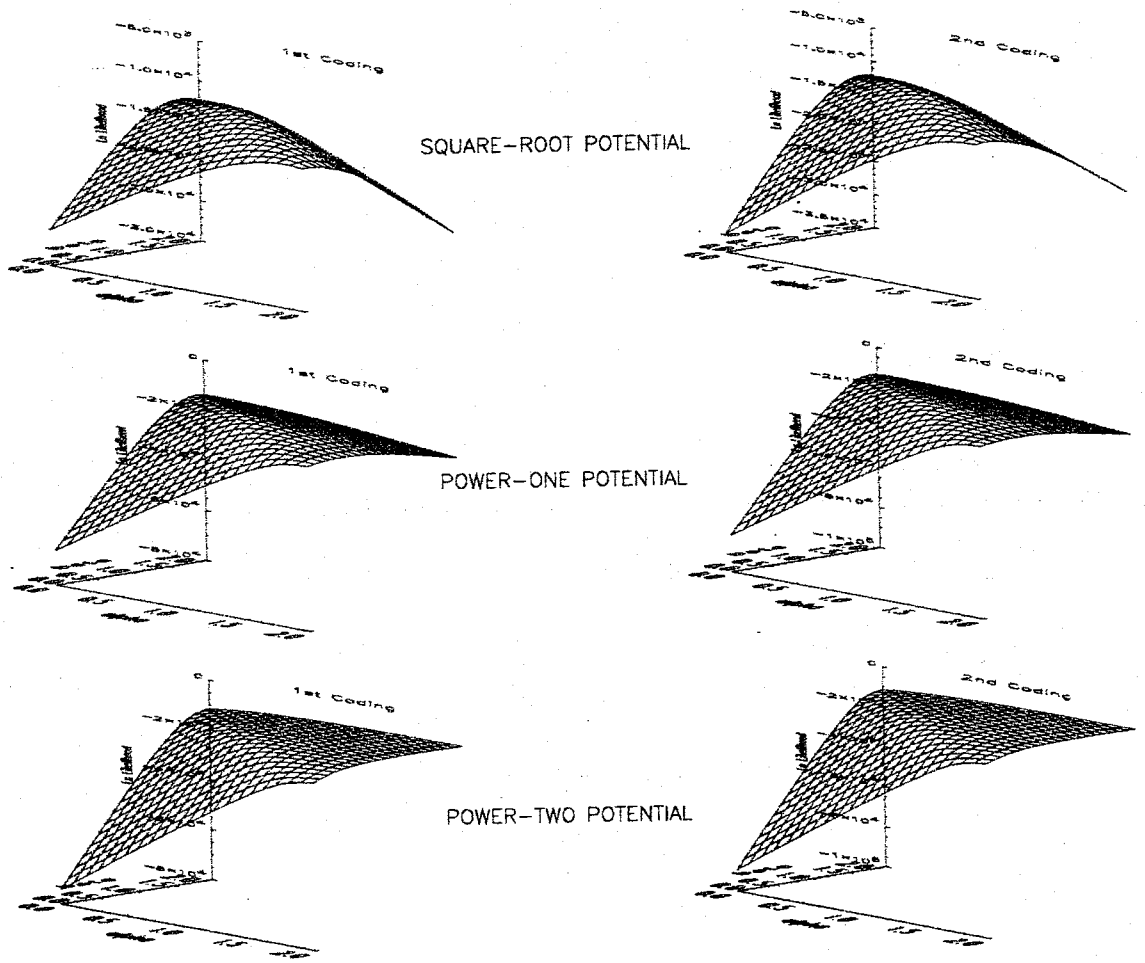


Figure 3.13. First-Order Log Likelihood Functions for Lithofacies Type Data Using Various Forms of Interaction Energy.

obtained by multiplying conditional probabilities given by equation [3.13] by the total expected frequency of each distinct neighborhood configuration $N_x N_z/6^4$, where N_x=740 is the number of horizontal sites, N_z=15 equals the number of sites in the vertical direction, and 6^4 neighborhood configurations are possible. The summation in the denominator of equation [3.22] was made tractable by this fairly small number of possible neighborhood configurations.

The number of degrees of freedom associated with all local configurations, including both site and neighbors, was $(6\text{-}1)6^4$. This was reduced by 2 to reflect the number of MRF parameters (α and β) estimated previously from the same data via maximum likelihood, yielding a value of 6478. The χ^2 statistic corresponding to a significance level of 0.05 and 6478 degrees of freedom was estimated from the approximation $df=1/2[(2(df)\text{-}1)^{1/2}-1.64]^2$ for large df to be equal to 6292 (Benjamin and Cornell, 1970). Note that, for this large value of df, the χ^2 statistic is sensitive to the data and may not be a good estimate for the data as a whole.

Results of the χ^2 tests are presented in Table 3.1. The chi-squared statistic used to accept or reject the assumed MRF model was twice the statistic for the given degrees of freedom because the statistics for both codings were dependent, even though the codings themselves were mutually orthogonal. Results suggest that the hypothesis cannot be accepted at a significance level of 5% for any pair-potential interaction-energy model nor for either independent coding. Based on these results, there is greater than a 5% chance of accepting the first-order, pair-potential energy MRF model when it is incorrect, regardless of assumed energy interaction power. It

Table 3.1. Results of Maximum Likelihood MRF Parameter Estimation and χ^2 Goodness-of-Fit Testing.

Power of Pair-Potential Interaction Energy	α	β	$\chi^2 = \sum_{i=1}^{N_F} \frac{(F_{oi} - F_i)^2}{F_i}$		$2\chi^2_{0.05,\,6478}$ = 2 (6292)
			Coding 1	Coding 2	
0.5	1.06	1.13	107641	17892	>12584
1.0	0.81	0.78	219653	26578	>12584
2.0	0.24	0.13	690130	75503	>12584

does appear, however, that the degree of peakedness of the log likelihood function in Figure 3.13 is an excellent indicator of relative goodness-of-fit, with the smallest estimated χ^2 value associated with the square-root model.

Figure 3.14 presents an unconditional MRF realization of lithofacies type synthesized with the Metropolis algorithm using the ML optimal parameters of the first-order, square-root model equal to α=1.06 and β=1.13, obtained after 5000 attempted exchanges per site. The tendency toward spatially-uniform properties at the top of the trench is not captured because such uniformity is essentially a nonstationary influence that the strictly stationary MRF model cannot possibly reproduce without recourse to local conditioning on observed lithofacies type. The failure of the χ^2 test and the poor reproduction of observed fan deposit structure were both attributed primarily to nonstationarities pervading the observed lithofacies type data. While the multivariate Gibbs pdf can account for arbitrary joint orders of correlation, it cannot incorporate randomness that is spatially invariant without benefit of extensive data conditioning.

It should be pointed out that, had permeabilities been assigned to each facies type and those permeabilities used to generate a permeability realization, the realization and MRF parameters would have been somewhat different. This is due to dependency of interaction energy on differences between random variables within the neighborhood.

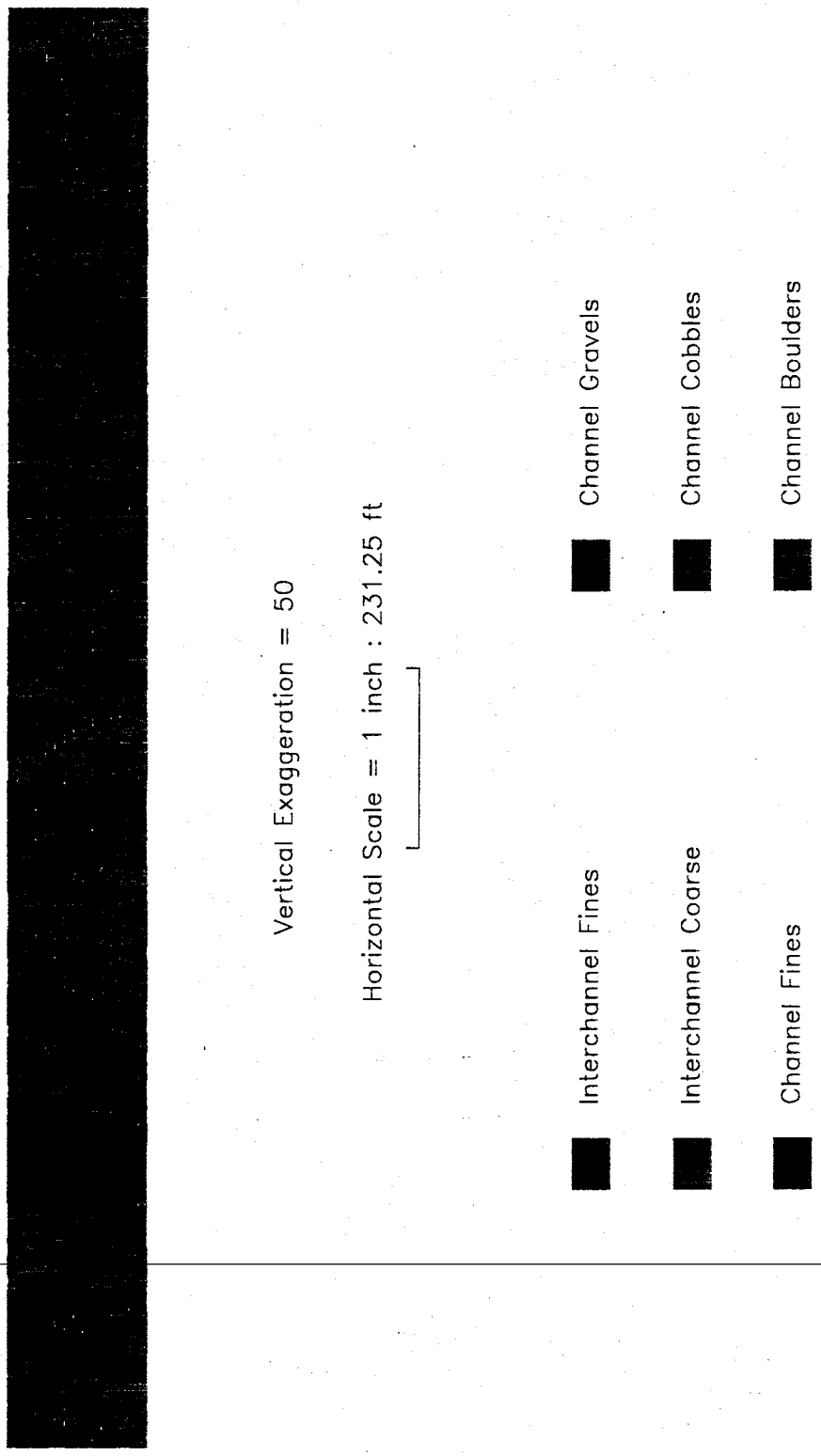


Figure 3.14. Simulated Lithofacies Type Distribution Along Main Stem of Sewer Trench Using Estimated MRF Parameters.

3.4 SUMMARY

Markov random fields (MRFs) impose spatial correlation over discrete space and are ideally suited for generation of geologic structure at larger scales of variation, where abrupt lithologic and stratigraphic boundaries typically prevail. They represent the extension of one-dimensional Markov chains to multivariate space. MRF realizations are synthesized using an importance sampling algorithm and a conditional probability operator analogous to the transition probability matrix used to generate one-dimensional Markov chains. Conditioning is performed at each level of the chain with respect to neighborhoods associated with pairs of randomly visited sites. The rate of chain convergence to a *bona fide* MRF realization is of order $\sqrt{N}$, where N is the number of exchanges per site.

MRF parameters for alluvial fan structure in the Albuquerque Basin were estimated on the basis of mutually independent subsets of observed lithofacies type data using maximum likelihood techniques. The observed lithofacies type realization could not be accepted as a sample from a first-order MRF characterized by any one of three pair-potential interaction energies, presumably due to strong nonstationarities evident in the observed lithofacies type distribution.

3.5 REFERENCES

Benjamin, J.R., and C. Allin Cornell, *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill, New York, 1970.

Besag, J., Nearest-neighbor systems and the auto-logistic model for binary data, *J. Royal. Stat. Soc., series B*, 34, 75-83, 1972.

Besag, J., Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. Royal. Stat. Soc., series B*, 36, 192-226, 1974.

Cross, G.R., and A.K. Jain, Markov random field texture models, *IEEE Trans. on Pattern Recog. and Mach. Intel.*, PAMI-5(1), 1983.

Crowson, D., J.T. McCord, R. Holt, J. Piper, and E.K. Webb, Detailed mapping and preliminary geostatistical analysis of alluvial fan deposits exposed in six miles of trench, *EOS, Transactions, American Geophysical Union, 1993 Spring Meeting,* 74(16), 143, 1993.

Dobrushin, R.L., The description of a random field by means of conditional probabilities and conditions of its regularity, *Theory Prob. Appl.*, 13, 197-224, 1968.

Farmer, C.L., *The Generation of Stochastic Fields of Reservoir Parameters with Specified Geostatistical Distributions*, paper presented at the IMA Conference on The Mathemetics of Oil Production, Robinson College, Cambridge Univ., July 6-7, 1987.

Farmer, C.L., *Numerical Rocks, The Mathematical Generation of Reservoir Geology,* paper presented at the joint IMA/SPE Conference on The Mathemetics of Oil Recovery, Robinson College, Cambridge Univ., July 25-27, 1989.

Gaither, K., J. Piper, and R. Holt, Geologic mapping of the sanitary sewer line excavation, southern Kirtland Air Force Base, Albuquerque, New Mexico, *Sandia Report SAND 92-2517*, 1993.

Geman, S., and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. on Patt. Recog. and Mach. Intel.*, PAMI-6(6), 1984.

Hammersley, J.M., and D.C. Handscomb, *Monte Carlo Methods*, Methuen, London, 1964.

Isham, V., An introduction to spatial point processes and Markov random fields, *Int. Stat. Rev.,* 49, 21-43, 1981.

Ising, E., Zeitschrift Physik, 31, 253, 1925.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087-1092, 1953.

Rothman, D.H., Nonlinear inversion, statistical mechanics, and residual statics estimation, *Geophysics*, 50(12), 2784-2796, 1985.

Spitzer, F., Markov random fields and Gibbs ensembles, *Amer. Math. Mon.,* 78, 142-154, 1971.

Whittle, P., *Systems in Stochastic Equilibrium*, J. Wiley and Sons, Chichester, 1986.

Chapter 4. Optimization of Hydraulic Conductivity Measurement in Multiple-Scale Geologic Media

4.1 CONDUCTIVITY MEASUREMENT IN MULTIPLE-SCALE POROUS MEDIA

In Chapter 2, it was demonstrated empirically using a numerical experiment that effective properties can change significantly when a multiple-scale conductivity process is conditioned on point noisy permeameter data. Results of that information transfer experiment suggest that, in geologic media characterized by more than one scale of spatial variation, use of conductivity measurements collected at the small scale to make predictions at a larger scale should be avoided whenever possible. Instead, conductivity should be measured independently at each scale relevant to the particular flow or transport problem of interest, ideally using instrument windows capable of measuring at characteristic frequencies that most dominate the problem.

This issue of scale-dependent randomness and measurement naturally leads to the question: At what scale should we measure random multiple-scale hydraulic conductivity if we want to minimize some measure of transport prediction uncertainty?. For example, if we are interested in minimizing concentration prediction uncertainty at a regulatory compliance point such as a monitoring well, how should we go about choosing the scale at which to measure conductivity? Clearly, this issue of

measurement scale has meaning only when posed in the context of a particular contaminant transport problem.

Depending on the nature of the transport problem and on the character of the heterogeneities, more marginal information about the prediction might be gained by favoring certain scales of data collection over another. For example, if we have reason to believe that most of the impact on contaminant migration is occurring at the smaller scale, we would stand to gain the most information about well concentration from *in situ* permeameter measurements and lab testing of core samples (see Figure 4.1). If,

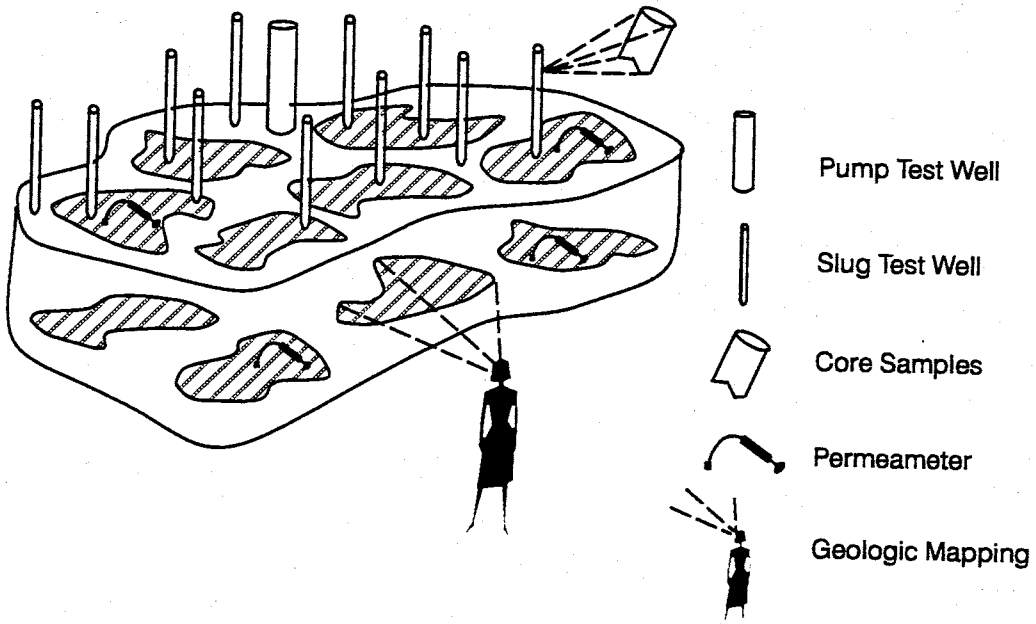


Figure 4.1. Multiple-Scale Conductivity Measurement.

on the other hand, there is evidence that the plume is impacted by intermediate-scale structure, we would want to shift the emphasis of measurement strategy toward slug testing and short-term pump testing to reduce prediction uncertainty. And finally, if the plume is of regional extent, we might best enhance our ability to predict well concentration by conducting long-term pump tests and mapping the geology. Our central consideration should be to measure conductivity at the spatial scale or scales that most drive the particular transport problem of interest. In Chapter 5, for example, it is demonstrated that concentration prediction uncertainty at the Las Cruces site is reduced as a result of conditioning on large-scale horizontally layered permeability structure. However, one can easily imagine a situation involving a plume that spans only the small, intralayer scale of variability, for which small-scale permeability measurement might provide greater reduction in concentration prediction uncertainty.

To systematize the search for optimal conductivity measurement scale, the multiple-scale measurement design algorithm must operate at all scales relevant to the problem, balancing data requirements among these scales. Ideally, the measurement design algorithm should also parallel the observed natural geologic hierarchy and focus on only natural scales of variation typically encountered in the field.

4.2 PREVIOUS RESEARCH RELATED TO PARAMETER MEASUREMENT NETWORK DESIGN

Networks can be designed to minimize either uncertainty in parameter measurement or uncertainty in state variables such as head or concentration. Design of state variable sampling networks has been investigated by a number of researchers, including Carrera *et al.* (1984), Meyer and Brill (1988), Hsu and Yeh (1989), Graham and McLaughlin (1989b, 1991), Loaiciga (1989), Hudak and Loaiciga (1992), Knopman and Voss (1989), Knopman *et al.* (1991), Andricevic (1990), Andricevic and Foufoula-Georgiou (1991), Lee and Kitanidis (1991), Tucciarelli and Pinder (1991), McKinney and Loucks (1992), James and Gorelick (1994), and Meyer *et al.* (1994). Loaiciga *et al.* (1992) provide an extensive review of research related to ground water quality monitoring network design. In general, objectives of water quality monitoring network design include maximizing probability of contaminant detection, minimizing the number of monitoring wells, or minimizing concentration variance.

Parameter measurement network design problems focus on parameter identification and can be classified into two categories. In the first category are those that identify a configuration and scheduling of state-variable sampling that minimizes some measure of aquifer parameter estimation variance. Those that seek to identify parameter measurement locations, where a measurement will most reduce state-variable prediction uncertainty, comprise the second category. The first class of problems can be considered a variant of the parameter identification or inverse problem, which attempts to minimize differences between predicted and observed state variables and parameters. The latter group is more in the realm of the classic

forward, predictive problem. Both classes of parameter measurement design problem rely on maximizing a scalar measure of the information matrix. However, information maximization translates into very different objectives depending on the class of design.

For the first class of design problems, information maximization requires that a scalar measure of the parameter covariance matrix be minimized (Yeh, 1986; Sun and Yeh, 1990). For the second class of design problem, information maximization translates into maximizing a measure of the state-variable covariance matrix. While this dissertation focuses on the problem of parameter measurement network design in the presence of hierarchical spatial variation, the parameter measurement network design algorithm can easily be extended to the problem of designing state-variable sampling networks in the presence of multiple scales of random spatial variation.

Both classes of design can further be divided into problems which explicitly account for physical relationships among and between state variables and parameters, and those that rely on kriging, cokriging, or other interpolation methods to relate them (McKinney and Loucks, 1992).

4.2.1 PARAMETER MEASUREMENT NETWORK DESIGN MODELS BASED ON KRIGING AND COKRIGING

Kriging is a statistical interpolation technique that linearly operates on available data to predict values at locations where data are not available, using estimated statistical properties to constrain the interpolation. In addition to providing

linearly interpolated values, the kriging algorithm also permits calculation of estimation variance at locations where values have been interpolated. However, Philip and Watson (1986) and Srivastava (1986) suggest that kriging estimation variance is more of an index of data configuration than it is an estimate of prediction variance. Although kriging estimation variances are not true ensemble variances, they can be used as a guide to choosing sampling or measurement locations by simply finding the locations associated with the largest variance. Since sampling or measuring at any location reduces variance to zero in the absence of measurement error, doing so at the location of maximum kriging variance produces the largest possible variance reduction.

Journel and Huijbregts (1978) used maximum conductivity kriging variance to determine optimal conductivity measurement networks via trial-and-error, and Delhomme (1978) determined locations where measurement of transmissivity reduced kriging estimation variance by a prescribed amount. Hughes and Lettenmaier (1981) recognized that, because kriging estimation variance at a potential measurement location depends on other measurement locations and not on the actual measured values, it is ideal for network design. Other researchers, including Bardossy and Bogardi (1983), Bogardi *et al.* (1985), Carrera *et al.* (1984), and Carrera and Szidarovzsky (1985), exploited this feature of kriging to formalize the variance reduction problem by incorporating kriging variance into the objective function and minimizing it over a set of integer decision variables representing potential measurement locations. The kriging equations were included as constraints in the design. However, a kriging variance objective function generally possesses many local

minima rather than one global minimum, violating the convexity requirement upon which linear and nonlinear programming algorithms are based.

Rouhani and Hall (1988) designed leakance measurement networks by finding locations where the risk of high leakance was large and where total kriging estimation variance reduction across the entire domain was at a maximum. Total variance reduction was not just variance reduction at a measurement point, but also included reduction at all other locations due to information obtained at the measurement location. This was considered a more meaningful measure of information than point estimation variance. It was not clear whether total variance reduction gave rise to a convex objective function.

While each of these studies seek to minimize parameter estimation error, a more meaningful measure of model reliability can be found in the predicted state of the system, rather than in the model parameters. After all, it is not uncertainty in conductivity *per se* that concerns us, but instead how that uncertainty affects our ability to make reliable head or concentration predictions. Concentration prediction error, for example, most compromises our ability to fulfill regulatory constraints on contaminant concentration at a point of compliance.

The conductivity measurement location that most reduces concentration or head variance can also be determined using cokriging estimation variance. Cokriging is similar to kriging, but utilizes information about how two different random processes are related to one another over space. Although cokriging does not explicitly account for the physical relationship between conductivity and concentration, it does so

implicitly based on how observations of conductivity and concentration relate to one another in space. McLaughlin and Graham (1986) estimated cross-covariance functions between heads, concentrations, and log conductivity and, based on cokriging estimation error, identified conductivity measurement networks that most reduce cokriging variance. Maximum likelihood geostatistical parameter estimation was used by Loaiciga (1989), in conjunction with mixed-integer programming, to identify an optimal concentration monitoring network design. This method involved geostatistical estimation of concentration moments based on the spatial structure of aquifer parameters, and can be viewed as a variant of statistical interpolation procedures such as cokriging.

A method that predicts concentration variance using physical principles is preferable to cokriging or maximum likelihood approaches because it explicitly accounts for the unique dynamics of the particular problem of interest. To overcome the problem of incorrect predictions of conditional variance, Dagan *et al.* (1996) used a Lagrangian approach to develop inversion equations for transport. A number of other approaches have been used to incorporate our understanding of dynamic physical processes into network design.

4.2.2 DESIGN MODELS BASED ON PHYSICAL PROCESSES

Rather than use cokriging of conductivity and concentration observations to identify the location where conductivity measurement will most reduce concentration variance, the actual dynamics of the transport system should be considered in the

design of a conductivity measurement network. Magnuson and Wilson (1987) were among the first researchers to account for predicted head uncertainty when identifying optimal transmissivity network design. Using a first-order, second-moment uncertainty analysis of the flow equation, they chose values of log transmissivity at specified measurement locations that minimized the trace of the head covariance matrix. Massman and Freeze (1987a,b) developed a network design using Bayesian updating of parameter statistics to minimize travel time uncertainty and concentration uncertainty at a regulatory point, relying on Monte Carlo simulation to perform flow and transport simulations. Multiple deterministic transport simulations were used by Tucciarelli and Pinder (1991) to identify a minimum cost of remediation design that also minimized costs of measuring transmissivity. Reduction in concentration variance produced by transmissivity measurement was incorporated into the design. Graham and McLaughlin (1989b) relied on extended Kalman filtering techniques to condition concentration and velocity predicted by the unconditional model of Graham and McLaughlin (1989a) on conductivity, head, and concentration observations.

The degree to which any of these methods of accounting for physical processes compare with cokriging will depend on the presence of deterministic boundary conditions and variables, as well as on whether the random variables are related through high-order terms not considered by the gaussian cokriging operator. While all of the approaches cited above account in different ways for the true physics impacting prediction uncertainty, the first-order second-moment method of defining prediction

uncertainty was adopted in this dissertation to estimate prediction uncertainty, following the work of Magnuson and Wilson (1987) and McKinney and Loucks (1992).

4.3 FORMULATION OF A MULTIPLE-SCALE CONDUCTIVITY MEASUREMENT DESIGN MODEL

Our goal is to choose conductivity measurement scale and location so that concentration predictions at a compliance point have minimum prediction uncertainty. We wish to avoid use of trial-and-error techniques for choosing conductivity measurement locations that most reduce prediction uncertainty, because they are time-consuming and inefficient. Moreover, one can never hope to fully enumerate all possible network configurations using guesswork. The time spent conducting a trial-and-error search is better utilized by building a model that systematically identifies the effects of conductivity uncertainty on prediction error. While Monte Carlo and extended Kalman filtering approaches account for the true physics impacting prediction uncertainty, a first-order second-moment method of defining prediction uncertainty was considered the least computationally intensive.

In order to build a model that systematically identifies the conductivity measurement scale and location that most improves concentration prediction at the monitoring well, some means of physically relating conductivity to well concentration must be established. Such a relation provides us with a way to quantify concentration uncertainty in the presence of a given amount of conductivity uncertainty. Rather than use perturbation analysis to perturb the flow and transport equations and solve them stochastically for concentration moments, we instead assume a linear relationship

between spatially-distributed log conductivity $\mathbf{Y}=\ln \mathbf{K}$ and concentration $\mathbf{C}$ using the first-order Taylor series expansion:

$$\mathbf{C}(\mathbf{Y}) \stackrel{1}{=} \mathbf{C}(\hat{\mathbf{Y}}) + \frac{\partial \mathbf{C}(\hat{\mathbf{Y}})}{\partial \mathbf{Y}}(\mathbf{Y} - \hat{\mathbf{Y}}) = \mathbf{C}(\hat{\mathbf{Y}}) + \mathbf{J}\,(\mathbf{Y} - \hat{\mathbf{Y}}) \qquad [4.1]$$

where

$\mathbf{Y}$ = the vector of ln K values at M potential measurement sites
$\hat{\mathbf{Y}}$ = the M-dimensional vector of mean ln K values at potential measurement sites
$\mathbf{C}(\mathbf{Y})$ = the N-dimensional vector of concentration at regulatory points
$\mathbf{C}(\hat{\mathbf{Y}})$ = the N-dimensional vector of concentration at regulatory points predicted on the basis of mean $\mathbf{Y}$, $\hat{\mathbf{Y}}$

and where $\stackrel{1}{=}$ denotes equal to first order.

Unknown concentration moments can be derived as a function of known log conductivity moments on the basis of the above relation. Note that the equivalence in equation [4.1] is to first order and omits terms of second order and higher. Equation [4.1] essentially expands solution of the transport equation about the ensemble mean deterministic solution, adding deviations to this deterministic solution that are attributable to fluctuations of $\mathbf{Y}$ about its mean. This first-order expansion yields the same moment information as a first-order perturbation analysis of the governing partial differential equations (Dettinger and Wilson, 1981).

Use of equation [4.1] requires that we assume variation in concentration **C** is only affected by variations in **Y**. Since the effects of porosity and molecular diffusion variations on transport have been determined to be insignificant compared to the effects of variations in conductivity (Naff, 1978; Gelhar *et al.*, 1979), this assumption does not generally pose a problem. While uncertainties with respect to random measurement error, discretization error, and systematic model error stemming from improper mathematical specification of the system can also contribute to concentration uncertainty, these other sources of uncertainty are likewise considered to be small compared to uncertainty in **Y**, and are consequently ignored. Some of these uncertainties could easily be included in the expansion as additional terms associated with derivatives of concentration with respect to additional sources of error. Since there is generally no way to distinguish between measurement and model errors or to systematize them, they are typically lumped into a single nonsystematic error term.

In addition to being invalidated by the existence of unaccounted sources of **Y** variation, the validity of expansion [4.1] can also be compromised when variations in **Y=ln K** are large relative to the mean of **Y** (i.e. when **Y** is not in the neighborhood of its mean) or if the relation between **C** and **Y** is not approximately linear in the vicinity of mean Y. Dettinger and Wilson (1981) suggest that the coefficient of variation for the independent variable **Y** be much smaller than 1, so that second-order moments of **Y** be small. This will generally not impose a significant constraint on using equation [4.1] because variations in **ln K** are typically small even when **K** variations are large. A high degree of nonlinearity in the system may also contribute to truncation error in

the expansion, because nonlinearity can produce large high-order derivatives of **C(Y)** with respect to **Y**. If high-order derivatives are large, uncertainty in **Y** can influence concentration predictions in a highly nonlinear manner, even if variations in **Y** produce small high-order moments. For example, changes in **Y** that contribute to large changes in flow geometry can result in pronounced nonlinearities in the **C-Y** relationship. Such nonlinearities will be a function of boundary conditions and magnitude of localized recharge/discharge. However, reduction of truncation error through consideration of second- and higher-order derivatives can generally be bought only at the expense of forfeiting the analytic convenience of equation [4.1].

Notwithstanding these considerations, the sensitivities $\partial C_i/\partial Y_j$, i=1, 2, 3, . . . , N, j=1, 2, 3, . . . , M, define an N-by-M Jacobian matrix of concentration sensitivities:

$$\mathbf{J} = \begin{bmatrix} \frac{\partial C_1}{\partial Y_1} & \frac{\partial C_1}{\partial Y_2} & \frac{\partial C_1}{\partial Y_3} & \cdots & \frac{\partial C_1}{\partial Y_M} \\ \frac{\partial C_2}{\partial Y_1} & \frac{\partial C_2}{\partial Y_2} & \frac{\partial C_2}{\partial Y_3} & \cdots & \frac{\partial C_2}{\partial Y_M} \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \cdot & \cdot & \cdot & \cdots & \cdot \\ \frac{\partial C_N}{\partial Y_1} & \frac{\partial C_N}{\partial Y_2} & \frac{\partial C_N}{\partial Y_3} & \cdots & \frac{\partial C_N}{\partial Y_M} \end{bmatrix} \qquad [4.2]$$

where N is the number of regulatory points at which concentration is to be monitored and M is the number of potential **Y** measurement sites, not including previously measured or inaccessible sites. **C** and **J** reduce to a scalar and an M-dimensional row vector, respectively, for the case of a single well with concentration observed at a single time. The sensitivities in equation [4.2] are evaluated at mean **Y**, the current estimate of the log conductivity process, and are updated during the course of the design as described later. They embody the unique physics of the transport problem, and are a crucial part of the measurement design algorithm, supplanting the need to solve the flow and transport equations for concentration as an explicit function of **Y**. If systematic model error is present, it might be incorporated into the elements of **J**.

In this study, adjoint techniques are used to facilitate estimation of monitoring-well concentration sensitivities with respect to **Y** at M conductivity measurement locations distributed uniformly over a discretized flow and transport domain.

4.3.1 First-Order, Second-Moment Analysis

Based on the first-order expansion given by equation [4.1], we can perform a second-moment analysis to obtain explicit expressions for unknown monitoring well concentration moments as functions of known moments of **Y**. First- and second-moment analysis of the first-order expansion replaces the more traditional Monte Carlo method which, while offering a full-distribution technique for estimating concentration moments, cannot be posed in the compact analytical form provided by

equation [4.1] and is therefore not as easily incorporated into the measurement design algorithm.

Taking the expectation of equation [4.1],

$$E[C(\mathbf{Y})] \stackrel{1}{=} E\left[C(\hat{\mathbf{Y}}) + \frac{\partial C(\hat{\mathbf{Y}})}{\partial \mathbf{Y}}(\mathbf{Y} - \hat{\mathbf{Y}})\right]$$

$$= E[C(\hat{\mathbf{Y}})] + E\left[\frac{\partial C(\hat{\mathbf{Y}})}{\partial \mathbf{Y}}(\mathbf{Y} - \hat{\mathbf{Y}})\right] = C(\hat{\mathbf{Y}}) = \hat{C} \qquad [4.3]$$

Equation [4.3] states that to first order, the expectation of monitoring well concentration C is simply the well concentration predicted for an aquifer in which **Y** is everywhere equal to the spatially-distributed ensemble mean of **Y**. Since the expected concentration is dependent only on the mean of **Y**, the first-order model for mean concentration is a deterministic one. Dettinger and Wilson (1981) show that a second-order expansion of C(**Y**) about mean **Y** yields an expected concentration that is second-order conditional upon both the mean and covariance of **Y**.

Equation [4.3] provides an expression for mean well concentration as a function of mean **Y**. However, it is not so much the expectation of C(**Y**) that is of interest to us, but rather some measure of the uncertainty in C(**Y**) associated with uncertainty in **Y**. Such an estimate of concentration uncertainty as a function of **Y** uncertainty is obtained as follows (Dettinger and Wilson, 1981):

$$\mathrm{Cov}[\mathbf{C(Y)}] \overset{1}{=} \mathrm{E}\left[\left[\mathbf{C(Y)} - \hat{\mathbf{C}}\right]\left[\mathbf{C(Y)} - \hat{\mathbf{C}}\right]^{\mathrm{T}}\right] = \mathrm{E}\left[\left[\mathbf{J}\ (\mathbf{Y} - \hat{\mathbf{Y}})\right]\left[\mathbf{J}\ (\mathbf{Y} - \hat{\mathbf{Y}})\right]^{\mathrm{T}}\right]$$

$$= \mathrm{E}\left[\ \mathbf{J}\ (\mathbf{Y} - \hat{\mathbf{Y}})\ (\mathbf{Y} - \hat{\mathbf{Y}})^{\mathrm{T}}\ \mathbf{J}^{\mathrm{T}}\right] = \left(\frac{\partial \mathbf{C}}{\partial \mathbf{Y}}\right) \mathrm{E}\left[(\mathbf{Y} - \hat{\mathbf{Y}})\ (\mathbf{Y} - \hat{\mathbf{Y}})^{\mathrm{T}}\right]\left(\frac{\partial \mathbf{C}}{\partial \mathbf{Y}}\right)^{\mathrm{T}}$$

$$= \left(\frac{\partial \mathbf{C}}{\partial \mathbf{Y}}\right) \mathrm{Cov}(\mathbf{Y}) \left(\frac{\partial \mathbf{C}}{\partial \mathbf{Y}}\right)^{\mathrm{T}} = \mathbf{J}\mathrm{Cov}(\mathbf{Y})\mathbf{J}^{\mathbf{T}} \qquad [4.4]$$

The concentration covariance given by equation [4.4] is what we're truly interested in, because it summarizes the uncertainty in predicted concentration C caused by uncertainty in the **Y** process. Equation [4.4] states that to first order, uncertainty in C will depend not just on uncertainty in **Y**, but also on how changes in **Y** translate into changes in C as expressed by sensitivities $\partial C/\partial \mathbf{Y}$. Equation [4.4] suggests that the variance of well concentration depends not only on the covariance of **Y**, but also on how sensitive C is to **Y**, evaluated at the current estimate of **Y**. Note that, since Cov(**Y**) is positive semi-definite for real matrix $\partial C/\partial \mathbf{Y}$, Cov(C) will always meet the covariance requirements of positive semi-definiteness (Dettinger and Wilson, 1981). McKinney and Loucks (1992) designed a conductivity measurement network on the basis of equation [4.4], using diagonal terms of the **Y** kriging covariance matrix to approximate Cov(**Y**).

Dettinger and Wilson (1981) also note that, since $(\partial C/\partial \mathbf{Y})(\partial C/\partial \mathbf{Y})^{\mathrm{T}}$ is generally not a diagonal matrix, Cov(**C**) will not typically be diagonal, even if Cov(**Y**) possesses diagonal structure. Consequently, correlation of **C** can occur in the absence of correlation in **Y**, demonstrating how the flow and transport equations introduce

correlation structure to state variables via the process of spatial integration. Moreover, the distance over which concentrations are correlated will depend not only on the distance over which **Y** is correlated, but also the distance over which concentration is sensitive to **Y**. It is this latter distance that generally produces longer correlation distances in **C** than are present in **Y**.

For a single concentration monitoring well and time, **J** is an M-dimensional row vector and Cov[C(**Y**)] reduces to a scalar concentration variance σ^2_C:

$$\mathbf{J}\mathrm{Cov}(\mathbf{Y})\mathbf{J}^T = \sigma^2_C \qquad [4.5]$$

Our goal is to define a single location and scale where **Y** contributes the most uncertainty, or variance, to predicted well concentration because knowledge gained by measuring that particular element of **Y** will most close the gap between the current state of knowledge regarding C and its true, or ensemble mean, value. That is, we wish for the weighted log conductivity covariance matrix $\mathbf{J}\mathrm{Cov}(\mathbf{Y})\mathbf{J}^T$ to be in some sense large, because this will be equivalent to maximizing the reduction in concentration variance σ^2_C. This differs fundamentally from the concentration sampling design problem investigated by Cleveland and Yeh (1990), for which the goal is to maximize some measure of the information matrix $\mathbf{J}\mathrm{Cov}^{-1}(\mathbf{C})\mathbf{J}^T$, equivalent to minimizing a measure of the parameter covariance matrix.

McKinney and Loucks (1992) used equation [4.5] to develop a single-scale conductivity measurement network. They demonstrated that the location of maximum

change in concentration variance with respect to change in **Y** variance is equivalent to the location where concentration sensitivity to **Y** is at a maximum. From equation [4.5],

$$\sigma_C^2 = \left(\frac{\partial C}{\partial \mathbf{Y}}\right) \mathrm{Cov}(\mathbf{Y}) \left(\frac{\partial C}{\partial \mathbf{Y}}\right)^T$$

Taking the derivative of σ^2_C with respect to Cov(**Y**) yields,

$$\frac{\partial \sigma_C^2}{\partial \mathrm{Cov}(Y)} = \left(\frac{\partial C}{\partial \mathbf{Y}}\right) \left(\frac{\partial C}{\partial \mathbf{Y}}\right)^T$$

For a single **Y** measurement location Y_j, $\mathrm{Cov}(Y_j)=\sigma^2_{Yj}$ and the above equation reduces to:

$$\frac{\partial \sigma_C^2}{\partial \sigma_{Y_j}^2} = \left(\frac{\partial C}{\partial Y_j}\right)^2 \qquad [4.6]$$

From equation [4.6], it can be seen that maximizing $(\partial C/\partial Y_j)^2\, \sigma_{Yj}^2$ and $(\partial C/\partial Y_j)^2/\, \sigma^2_C$, or more generally, maximization of $\mathbf{J}\mathrm{Cov}(\mathbf{Y})\mathbf{J}^T$ and $\mathbf{J}\mathrm{Cov}^{-1}(\mathbf{C})\mathbf{J}^T$, are basically an expression of the same problem — find the location where concentration prediction C is most sensitive to changes in **Y**, with the squared sensitivity weighted by the appropriate covariance or its inverse, depending on whether one is interested in

obtaining concentration variance reduction (measurement design), or conductivity variance reduction (sampling design).

4.3.2 Single-Scale Conductivity Measurement Design

A single-scale measurement network design algorithm can easily be developed using standard integer-programming techniques that 'turn on' measurement at any one of M potential measurement locations to yield the most additional information about well concentration C, providing us with a more reliable predictive model. Expanding the terms of equation [4.1] for scalar concentration and weighting each term by a binary decision variable vector **u** yields:

$$C(\mathbf{Y}) = C(\hat{\mathbf{Y}}) + \sum_{j=1}^{M} u_j \left[\frac{\partial C(\hat{Y}_j)}{\partial Y_j} \right] (Y_j - \hat{Y}_j)$$

where the elements of **u** are either 1 to denote measurement or 0 otherwise. Again, we are interested in maximizing concentration variance with respect to **Y** at location index j:

$$\max_j I = \max_j \left[\text{var}[C(\mathbf{Y})]\right] = \text{var}\left[C(\hat{\mathbf{Y}}) + \sum_{j=1}^{M} u_j \left[\frac{\partial C(\hat{Y}_j)}{\partial Y_j}\right](Y_j - \hat{Y}_j)\right]$$

$$= \sum_{j=1}^{M} u_j^2 \left[\frac{\partial C(\hat{Y}_j)}{\partial Y_j}\right]^2 \text{var}\,(Y_j - \hat{Y}_j) + \sum_{j \neq k}^{M} u_j u_k \left[\frac{\partial C(\hat{Y}_j)}{\partial Y_j}\right]\left[\frac{\partial C(\hat{Y}_k)}{\partial Y_k}\right]\text{Cov}(Y_j, Y_k) \qquad [4.7]$$

Subject to: $u_j = 0$ or 1; $j = 1, 2, \ldots, M$ [4.7a]

Since sensitivities in equation [4.7] are evaluated at the prior estimate of **Y**, an outlier **Y** realization has the potential to produce distorted sensitivity estimates and yield an extremely suboptimal design, particularly during early stages of site characterization when little information about **Y** is available. In an attempt to guard against that possibility, sequential design is introduced via the additional constraint:

$$\sum_{j=1}^{M} u_j = 1 \qquad [4.7b]$$

Constraint [4.7b] protects the design against prior uncertainty in **Y** by restricting the design to measure at only a single location. With the maximization subject to constraints [4.7a] and [4.7b], $u_j u_k = 0$ for $j \neq k$ and the second summation in equation

[4.7] equals zero. Therefore, as a natural consequence of the sequential design constraint [4.7b], the off-diagonal terms of the Cov(**Y**) matrix automatically drop from consideration. These off-diagonal terms account for conductivity information shared by two **Y** measurements. When only a single **Y** measurement location is considered, there is no such redundancy present. Ignoring the off-diagonal terms of Cov(**Y**) is equivalent to assuming no correlation between sensitivities (Knopman and Voss, 1987; 1989)

The sequential measurement design given by constraint [4.7b] is analogous to identification of A-optimal, or incremental, sampling design networks that expand on existing networks by minimizing the weighted trace of the parameter covariance matrix (Steinberg and Hunter, 1984). D-optimal, or total designs, involve design of entire networks by maximizing total variance, and account for off-diagonal terms that reflect shared information at more than one measurement location. D-optimality is obtained by minimizing the determinant of the weighted parameter covariance matrix, which is directly related to the size of the parameter estimate confidence ellipsoid. In general, there is insufficient prior information to obtain a D-optimal design in most natural geologic settings. This is because uncertainty in **Y** tends to identify measurement locations that are only locally optimal, or optimal in the neighborhood of the current information point defined by the most recent guess of **Y**. Convergence properties of

sequential, A-optimal designs are discussed by Hsu and Yeh (1989) and Nishikawa and Yeh (1989).

Objective function [4.7] is further simplified by noting that $u_j^2=u_j$ for all j. Together with the simplification provided by dropping off-diagonal terms, the objective function reduces to:

$$\max_j I = \max_j \left[\text{var}[C(\mathbf{Y})]\right] = \sum_{j=1}^{M} u_j \left[\frac{\partial C(\hat{Y}_j)}{\partial Y_j}\right]^2 \sigma^2_{Y_j} \qquad [4.8]$$

The information objective function can be viewed as a weighted sum of squared sensitivities, with the weights provided by the decision variable **u** and the ensemble variance of the **Y** process, both evaluated at the jth measurement location. A budgetary constraint may also be imposed on the design to limit costs of measurement.

Objective function [4.8], subject to constraints [4.7a] and [4.7b], is a mixed-integer programming problem that maximizes gain in concentration information in order to best inform our prediction. As might be anticipated solely on the basis of intuition, information gain will be greatest at the measurement location j where monitoring well concentration is the most sensitive to **Y** and where uncertainty in **Y** is the largest, because it is at these locations where measurement of **Y** will most improve C prediction reliability. The sequential design constraint forces the design model to choose the single location where **Y** uncertainty contributes the most uncertainty to predicted C, with the basis for the selection predicated on the current estimate of **Y**.

The dominance of any one term in objective function [4.8] will depend on both concentration sensitivity and variance in **Y**. If variance in **Y** at the jth location is large but sensitivity $\partial C/\partial Y_j$ is extremely small, the jth term may be small relative to other terms in the sum despite the high degree of parameter uncertainty at that location. On the other hand, Y_j might be known with a great deal of certainty, but the jth measurement location may have a very large sensitivity associated with it, in which case the contribution from the jth term may be large despite the high degree of certainty. This interplay between variance and sensitivity is complicated by the fact that they are often inversely related, with smaller **Y** variance limiting the number of possible pathways for a contaminant to migrate, producing large well concentration sensitivities. This inverse relationship is, in fact, suggested by equation [4.4] for constant concentration covariance.

The implications of constraint [4.7b] are significant and warrant further attention. While we can never expect to find a truly optimal design in the absence of perfect knowledge regarding **Y**, constraint [4.7b] constitutes an effort to minimize suboptimality. In addition to restricting measurement to a single location, the incremental approach to measurement design allows us to collect a measurement of **Y** at the single location specified by the model, and condition our current guess of realization **Y** on the newly-acquired value, thereby reducing the adverse effects of conductivity uncertainty on the design. By repeating the cycle of optimization and conditioning, we can reduce the potential for an extremely suboptimal design. Sequential conditioning typically continues until either the variance in concentration

or the change in the variance decrease below some value acceptable to the regulatory agency (McKinney and Loucks, 1992). Conditioning can also extend to Bayesian updating of the statistics of the **Y** process following each sequential measurement. In this particular study only the **Y** process itself, and not the statistics, was conditioned on the new conductivity measurement.

Note that, even with the benefit of sequential measurement and conditioning, we must still operate without complete knowledge of the true **Y** process. As a result, a truly optimal design is unlikely to be identified. The potential for suboptimality in the presence of uncertain **Y** will depend on how effectively the system disperses mass. If contaminant mass disperses easily through the system, then there will be a greater number of measurement locations where mass resides and where some information is available than if the system is not dispersive. To understand this, consider a system characterized by a low degree of dispersion. In such a system, the advection-dispersion equation acts hyperbolically, mass has finite velocity, there are many locations where mass rarely appears, and concentration sensitivities at these locations will tend to be small. Conversely, when dispersion is very large and mass (information) moves instantaneously through the system, the advection-dispersion equation acts as a parabolic partial differential equation, with mass being transported to many locations in the domain. In highly dispersive systems, the optimization will tend to correct itself for poor prior information more readily than in less dispersive systems. Under such circumstances, the design will be less likely to get 'caught' in a suboptimal, yet robust, design when poor prior information is used. In highly dispersive systems, the sequential

conditioning constraint [4.7b] may be relaxed by allowing more than one measurement at a time to be identified.

Conditioning on concentration and head observations, as well as on conductivity observations, can serve to further reduce uncertainty in predicted monitoring well concentration and alleviate the impact of poor prior Y information. This is especially true during later phases of site characterization, when measurements of state variables typically become available. Conditioning information gain I on concentration and head measurements was considered beyond the scope of this particular study.

In addition to the problem of uncertain prior information and suboptimality, the validity of measurement design depends in part on the degree to which the first-order approximation given by equation [4.1] is valid. This can be assessed by estimating concentration variance reduction as a result of the most recent conductivity measurement at the optimal measurement location using equation [4.8], with $u_j=1$ for the location where conductivity was last measured and 0 at all unmeasured locations and locations where measurement occurred during previous iterations of the design. If the estimated variance reduction during the preceding iteration varies significantly from well concentration variance reduction estimated via Monte Carlo simulation, the poor agreement can be related to either nonlinearities not accounted for in the linear C-Y approximation, nonergodic behavior of the transport process, or uncertainty in the Monte Carlo concentration prediction.

4.3.3 Multiple-Scale Conductivity Measurement Design

The single-scale design developed in the previous section was extended to account for more than one scale of variation in **Y**. Measurement of **Y** at the smallest possible scale will resolve heterogeneity but will not generally minimize prediction uncertainty because of the problem of parameter uncertainty. Conversely, measurement at larger scales does not resolve spatial variability, but will tend to reduce parameter uncertainty. Recall from Chapter 1 that both poor measurement resolution and high parameter uncertainty are sources of prediction error. We wish to find a trade-off between the two sources in the context of a particular transport problem using the multiple-scale measurement design.

Our objective now is to maximize the amount of marginal information per unit cost (IPC) at the concentration monitoring well over both scale s and location j_s, where j_s is itself now a function of scale:

$$\max_{j_s,\, s} \text{IPC} = \sum_{s=0}^{S} f(\mathbf{Y}^{(s)})/r^{(s)} \qquad [4.9]$$

again subject to the sequential design constraint:

$$f(\mathbf{Y}^{(s)}) = \sum_{j_s=1}^{M_s} u_{j_s}^{(s)} \left[\frac{\partial C^{(s)}}{\partial Y_{j_s}^{(s)}}\right]^2 \sigma^2_{Y_{j_s}^{(s)}} \qquad s = 0, 1, 2, \ldots\ldots, S \qquad [4.9a]$$

$$u_{j_s}^{(s)} = 0 \text{ or } 1 \qquad s = 0, 1, 2, \ldots\ldots, S \quad j_s = 1, 2, \ldots, M_s \qquad [4.9b]$$

$$\sum_{s=0}^{S} \mathbf{u}^{(s)^T} \mathbf{u}^{(s)} = 1 \qquad [4.9c]$$

where S is the total number of distinct scales of natural variation, and $r^{(s)}$ the cost of measuring conductivity at the sth scale. Because the conductivity scales are considered to be independent of one another (this being, in fact, the reason that a viable upscaling rule frequently does not exist), it is assumed that no interaction between uncertainty contributions from different scales occurs and that total information is simply the sum of separate scale-dependent contributions. The $f(\mathbf{Y}^{(s)})$ functions given by constraint [4.9a] represent the weighted sums of squared sensitivities, where conductivity is now a function of scale. $C^{(s)}$ represents the concentration at the regulatory compliance point averaged over the sth scale, discussed further in section 4.4. The information

objective [4.9] is to maximize the amount of information per unit cost at the monitoring well over all locations and over all scales of measurement.

Note that the multiple-scale design algorithm parallels the single-scale algorithm, with the exception that **Y** and C are indexed with a scale parameter s and the integer decision variable vectors $\mathbf{u}^{(s)}$ are different for each measurement scale. With the scale-dependent sequential design constraint [4.9c], we instruct the model to find the single location at the single scale where $\mathbf{Y}^{(s)}$ uncertainty most contributes to concentration prediction uncertainty. As before, the sequential design constraint allows us to condition $\mathbf{Y}^{(s)}$ on newly acquired measurements and avoid extremely suboptimal designs.

4.3.4 Multiple-Scale Conductivity Measurement Design Constrained by Multiple Conductivity Realizations

As a further hedge against outlier realizations of $\mathbf{Y}^{(s)}$ and suboptimal design, we can incorporate NR different conditional realizations of the **Y** process into the multiple-scale measurement design, where NR is the number of realizations used to constrain each iteration of the sequential design. Wagner and Gorelick (1989) used such a multiple conductivity realization approach to constrain a stochastic ground water quality management model. NR scale-dependent $\mathbf{Y}^{(s)}$ realizations are incorporated into the design as follows:

$$\max_{j_s,\, s} \text{IPC} = \sum_{s=0}^{S} f(\mathbf{Y}^{(s)})/r^{(s)} \qquad [4.10]$$

Subject to :

$$f(\mathbf{Y}^{(s)}) = \sum_{j_s = 1}^{M_s} RML_{j_s}^{(s)} \, u_{j_s}^{(s)} \qquad [4.10a]$$

$$u_{j_s}^{(s)} = 0 \text{ or } 1 \qquad s = 0, 1, 2, \ldots\ldots, S \qquad j_s = 1, 2, \ldots\ldots, M_s \qquad [4.10b]$$

$$\sum_{s = 0}^{S} \mathbf{u}^{(s)^T} \mathbf{u}^{(s)} = 1 \qquad [4.10c]$$

where $RML^{(s)}_{js}$ is the maximum likelihood rank of variance-weighted, squared sensitivity at location j_s and scale s over an ensemble of NR realizations. Again, the $\mathbf{u}^{(s)}$ integer decision variable vectors are not indexed to the realization index k and are the same for every realization.

Modified maximum likelihood rank $RML^{(s)}_{js}$ determines which term in the objective function sum [4.10] is dominant. It is based on the raw rank $R_{Y(s),js,k}$ of weighted squared sensitivity at each location and scale and within each realization:

$$R_{Y^{(s)}_{j_s,k}} = R\left[\left[\frac{\partial C^{(s)}}{\partial Y^{(s)}_{j_s,k}}\right]^2 \sigma^2_{Y^{(s)}_{j_s,k}}\right] \qquad \begin{array}{l} j_s = 1, 2, \ldots, M_s \\ s = 0, 1, 2, \ldots\ldots, S \\ k = 1, 2, \ldots\ldots, NR \end{array}$$

[4.10d]

where R is a numerical ranking, dependent on variance-weighted squared sensitivity, and operating in descending order from M_s to 1. $R_{Y(s),js,k}$ is used to build histograms $p^{(s)}_{js}$ across the ensemble for each scale s and at each location j_s. Maximum likelihood rank at the sth scale and jth location, $RML^{(s)}_{js}$, is then defined as the rank associated with the greatest ensemble probability:

$$RML^{(s)}_{j_s} = \max_{l} p^{(s)}_{j_s}(l) \qquad l = 1, 2, \ldots, M_s$$

[4.10e]

where $p^{(s)}_{js}(l)$ is the ensemble probability of rank l occurring at the j_sth location and the sth scale. $RML^{(s)}_{js}$ is modified to account for no single preferred rank (all ranks having exactly the same probability) by setting $RML^{(s)}_{js}$ equal to 0 to denote a 'do nothing' policy, as well as for rank ties by multiplying $RML^{(s)}_{js}$ by ensemble mean weighted squared sensitivity $E[(\partial C^{(s)}/\partial Y^{(s)}_{js})^2 \sigma_{Y(s)js}{}^2]$ at the j_sth location and sth scale. For $E[(\partial C^{(s)}/\partial Y^{(s)}_{js})^2 \sigma_{Y(s)js}{}^2] << RML^{(s)}_{js}$ the ensemble mean acts to break the tie

between locations and scales holding equivalent rankings, but permits the value of $RML^{(s)}_{js}$ to dominate each term in the objective function sum and remain fairly unaffected by the presence of outlier **Y** realizations. The equations associated with these latter two constraints are algorithmic and will be referred to as constraints [4.10f]. Note that, by using maximum likelihood rank in place of ensemble mean rank, the design is protected against extreme rank outliers in addition to extreme **Y** outliers.

The multiple-scale design algorithm given by objective function [4.10] subject to constraints [4.10a] through [4.10f] finds the location and scale where conductivity measurement $Y^{(s)}_{js}$ is most likely to provide the greatest reduction in monitoring well uncertainty. As NR increases, measurement network design is more likely to account for all uncertainty in **Y** and to represent the true optimal design.

As a final consideration, an average variance reduction can be obtained following each sequential measurement iteration by evaluating equation [4.8] for each realization and finding the mean reduction over the ensemble due to the new measurement. Ensemble mean concentration variance reduction affected by measurement of $Y^{(sopt)}_{jopts}$ can be expressed as:

$$\Delta I = \Delta \sigma_C^2 = E\left[\left[\frac{\partial C^{(sopt)}\left(Y_{jopt_s}^{(sopt)}\right)}{\partial Y_{jopt_s}^{(sopt)}}\right]^2 \sigma^2_{Y^{(sopt)}_{jopt_s}}\right] \qquad [4.11]$$

where sopt and $jopt_s$ are the optimal scale and location of the most recent measurement and E denotes the expectation operator over all NR realizations. Equation [4.11] can

be used after each iteration of the sequential multiple-scale, multiple-realization design as a check against ensemble C variance reduction estimated from the NR simulations of well concentration to determine the validity of the linear C-**Y** approximation or the presence of uncertainty in simulated C.

4.4 AVERAGING CONCENTRATION OVER DIFFERENT SCALES OF RESOLUTION

The concept of averaging concentration over the sth scale as mentioned in section 4.3.3 requires further clarification. Figure 4.2 shows examples of concentration solutions $C^{(s)}=C(\mathbf{Y}^{(s)})$ to the transport equation at two different scales of **Y** resolution obtained using a particle tracker. In effect, the transport problem was treated within its functionally hierarchical framework by conducting independent transport simulations separately at each scale. Therefore, it was the transport equations, and not post-simulation concentrations, that were averaged to obtain the solutions. In other words, large- and small-scale transport solutions were treated as realizations of two entirely different random processes.

The first transport solution corresponds to a 16-by-16 large-scale MRF log conductivity realization with state levels −7.0, −6.0, −5.0, −4.0, −3.0 ln m/s, and the second to a 256-by-256 two-scale log conductivity realization obtained by embedding 16-by-16 zero-mean log conductivity gaussian structure of variance 2.0 $(\ln \text{m/s})^2$ in each block of the MRF realization. In both cases, an instantaneous source of 1024 particles was uniformly distributed over a 64-by-64 square located in the NW corner of the

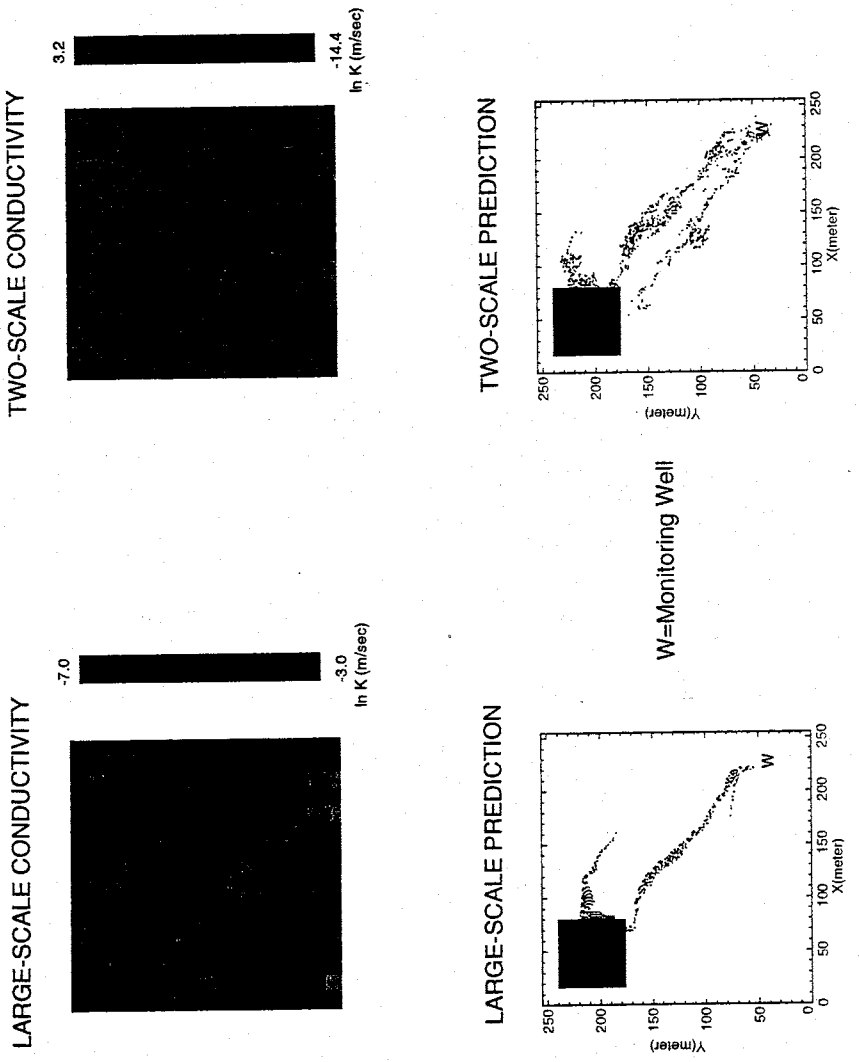


Figure 4.2. Examples of Predicted Plumes at Two Scales of Simulation.

transport domain, with average linear steady-state velocity of $1.0X10^{-6}$ m/s from NW to SE. Particle tracking was performed for a simulation period of 5.8 years.

It is evident that, while the plumes in Figure 4.2 have the same overall shape, they might predict a different concentration at a given point. For example, if the concentration monitoring well is located as indicated by W in the figure, one can see that very different concentration predictions would be made depending on the scale of simulation. In fact, since the degree of spatial variability differs between the two solutions, one would expect that under the ergodic hypothesis, concentration prediction uncertainty at the monitoring well would also be different for the two scales of simulation. The more dispersed nature of the two-scale solution can be attributed to the larger number of independent velocities sampled at the small scale relative to the number sampled at the large scale. This larger number of independent velocities causes a greater tendency for Fickian spreading of solute in the two-scale simulation compared to that associated with the larger scale of resolution. The fact that the two-scale concentration solution behaves more diffusively results in small-scale sensitivities that are significantly smaller than large-scale sensitivities.

4.5 CALCULATION OF CONCENTRATION SENSITIVITIES USING ADJOINT TECHNIQUES

For a given scale of variation, each realization **Y** produces a realization of monitoring well concentration sensitivity $\partial C/\partial \mathbf{Y}$. Central to generation of a sensitivity realization is estimation of $\partial C/\partial \mathbf{Y}$ via well-understood deterministic physical laws. These laws can be imposed either via forward or adjoint techniques.

The influence coefficient method for determining sensitivities is a forward technique, requiring that each element of **Y** be perturbed one at a time and well concentrations $C(Y_1, Y_2, \ldots, Y_i+\Delta Y_i, \ldots, Y_M)$ and $C(Y_1, Y_2, \ldots, Y_i, \ldots, Y_M)$ obtained from solution of the flow and transport equations. Based on those concentrations, a first-order estimate of sensitivity can be expressed as:

$$\frac{\partial C}{\partial Y_i} \overset{1}{=} \frac{C(Y_1, Y_2, \ldots, Y_i + \Delta Y_i, \ldots Y_M) - C(Y_1, Y_2, \ldots, Y_i, \ldots Y_M)}{\Delta Y_i}$$

$$i = 1, 2, \ldots, M \qquad [4.12]$$

Equation [4.12] is a discrete version of the first-order Taylor series expansion given by equation [4.1] and must be solved M times to define every element in the Jacobian vector. Solution of the sensitivity equation, obtained by directly differentiating the partial differential equations governing flow and transport with respect to Y_i, offers no real advantage over the influence coefficient method because it requires solution of as many equations as the influence coefficient method (Sykes *at al.*, 1985). However, such a procedure does yield exact sensitivities, while the accuracy of sensitivities provided by equation [4.12] depends on round-off and truncation errors associated with the chosen value of ΔY_i.

Adjoint-state theory makes it possible to determine sensitivities of all state variables with respect to all model parameters or inputs with minimal computational effort. The theory permits determination of concentration sensitivity to perturbations

in conductivity without the burden of having to solve the entire flow and transport problem for each conductivity value that we wish to perturb. Almost as important, solution of the adjoint equations is exact — it is not subject to either round-off errors associated with too small an imposed perturbation or to violation of the linearity assumption attendant to the Taylor-series expansion when the perturbation ΔY_i is too large. Ahlfeld *et al.* (1988) have developed coupled flow and transport adjoint equations to obtain sensitivities of concentrations with respect to remediation well injection and pumping rates, and used them to define the shape of nonlinear constraints for identification of minimum cost remediation well operation. The concentration constraints were nonlinear with respect to well flow rate because concentration predicted via the advection-dispersion equation is a nonlinear function of flow rate. Sensitivities were also used to define gradients for an objective function, also a nonlinear function of flow rate, that minimized total contaminant mass. The methodology of Ahlfeld *et al.* (1988) was adopted for this study, with some modification to reflect the finite-difference nature of the flow and transport models used to generate forward solutions as input to the adjoint problem.

Consider a vector of simulation functions implicitly relating M parameters to n state variables:

$$\mathbf{g}(\bar{\sigma}, \mathbf{Y}) = \mathbf{0} \qquad [4.13]$$

where

$\mathbf{g}(\bar{\sigma}, \mathbf{Y})$ = an n-by-1 vector of discretized equations describing steady-state fluid flow, steady-state velocity, and transient contaminant transport

$$= \begin{bmatrix} \mathbf{g}_{\mathbf{flow}}(\bar{\sigma}, \mathbf{Y}) \\ \mathbf{g}_{\mathbf{darcy}}(\bar{\sigma}, \mathbf{Y}) \\ \mathbf{g}_{\mathbf{tran0}}(\bar{\sigma}, \mathbf{Y}) \\ \mathbf{g}_{\mathbf{tran}}(\bar{\sigma}, \mathbf{Y}) \end{bmatrix}$$

$\bar{\sigma}$ = n-by-1 vector of state variables = $[\bar{h}, \bar{v}, \bar{c}_0, \bar{c}]$

$\bar{h}$ = vector of steady-state heads [L]

$\bar{v}$ = vector of steady-state velocities [L/T]

$\bar{c}_0$ = vector of previous concentrations [M/L^3]

$\bar{c}$ = vector of current concentrations [M/L^3]

$\mathbf{Y}$ = M-by-1 vector of parameters = $\mathbf{ln\ K}$

The functions $\mathbf{g}(\bar{\sigma},\mathbf{Y})$ are equations describing fluid flow $\mathbf{g_{flow}}(\bar{\sigma},\mathbf{Y})$, darcy velocity $\mathbf{g_{darcy}}(\bar{\sigma},\mathbf{Y})$, and prior and current concentrations $\mathbf{g_{tran0}}(\bar{\sigma},\mathbf{Y})$ and $\mathbf{g_{tran}}(\bar{\sigma},\mathbf{Y})$ as functions of parameter $\mathbf{Y}$, with velocity vector $\bar{v}$ dependent on head through $\mathbf{g_{darcy}}(\bar{\sigma},\mathbf{Y})$ and $\bar{c}_0$ and $\bar{c}$ dependent on $\mathbf{Y}$ through velocity vector $\bar{v}$ via $\mathbf{g_{tran0}}(\bar{\sigma}, \mathbf{Y})$ and $\mathbf{g_{tran}}(\bar{\sigma}, \mathbf{Y})$. Although prior $\mathbf{Y}$ measurement and inaccessibility of certain measurement sites will reduce the number of potential measurement locations to less than M, the nature of

the adjoint problem solution is such that adjoint states must be determined throughout the transport domain. **Y** measurement at locations of previous measurement or inaccessibility are withdrawn from consideration in objective function [4.10] by setting the associated variance to zero.

It should be pointed out that, while the advection-dispersion equation was used to define primary transport problems $\mathbf{g_{tran0}}(\bar{\sigma}, \mathbf{Y})$ and $\mathbf{g_{tran}}(\bar{\sigma}, \mathbf{Y})$, functions were actually obtained using a particle tracker that, by definition, does not consider dispersion of the contaminant. Use of these concentrations and the values of $D_L=D_T=0$ implied by them did not appear to create numerical difficulties when solving the adjoint problem. Also note that concentration associated with more than two time steps must be incorporated into the problem if transient design is desired. When the transit time of the plume is large relative to the design period, the design can be treated as steady-state for all practical purposes.

In addition to the simulation functions, consider a p-dimensional vector of performance measures:

$$\bar{\theta}(\bar{\sigma}, \mathbf{Y}) = \int_A f(\bar{\sigma}, \mathbf{Y})\, dA \qquad [4.14]$$

where $f(\bar{\sigma},\mathbf{Y})$ is a function of system state vector $\bar{\sigma}$, as well as parameter vector **Y**, integrated spatially over the domain A (Sykes *et al.*, 1985). The performance function sensitivity is derived as a total derivative of $\bar{\theta}(\bar{\sigma},\mathbf{Y})$ with respect to parameter vector **Y**:

$$\frac{d\bar{\theta}(\bar{\sigma}, \mathbf{Y})}{d\mathbf{Y}} = \frac{\partial\bar{\theta}(\bar{\sigma}, \mathbf{Y})}{\partial\mathbf{Y}} + \frac{\partial\bar{\theta}(\bar{\sigma}, \mathbf{Y})}{\partial\bar{\sigma}}\frac{\partial\bar{\sigma}}{\partial\mathbf{Y}}$$

$$= \nabla_{\mathbf{Y}}\bar{\theta}(\bar{\sigma}, \mathbf{Y}) + \nabla_{\sigma}\bar{\theta}(\bar{\sigma}, \mathbf{Y})\,\nabla_{\mathbf{Y}}\bar{\sigma} \qquad [4.15]$$

where $d\bar{\theta}(\bar{\sigma},\mathbf{Y})/d\mathbf{Y}$, $\nabla_{\mathbf{Y}}\bar{\theta}(\bar{\sigma},\mathbf{Y})$, $\nabla_{\sigma}\bar{\theta}(\bar{\sigma},\mathbf{Y})$, and $\nabla_{\mathbf{Y}}\bar{\sigma}$ are derivative matrices of dimension p-by-M, p-by-M, p-by-n, and n-by-M, respectively. $\nabla_{\mathbf{Y}}\bar{\sigma}$ is referred to as the state sensitivity. The first term of the total derivative, $\nabla_{\mathbf{Y}}\bar{\theta}(\bar{\sigma}, \mathbf{Y})$, describes all changes to $\bar{\theta}(\bar{\sigma}, \mathbf{Y})$ caused directly, or explicitly, by perturbations in parameters $\mathbf{Y}$, while the second term includes all indirect or implicit contributions to changes in $\bar{\theta}(\bar{\sigma}, \mathbf{Y})$ attributable to changes in state variables $\bar{\sigma}$ produced by perturbations in $\mathbf{Y}$. Since $\bar{\theta}(\bar{\sigma}, \mathbf{Y})$ is a known function of $\mathbf{Y}$ and $\bar{\sigma}$, $\nabla_{\mathbf{Y}}\bar{\theta}(\bar{\sigma}, \mathbf{Y})$ and $\nabla_{\sigma}\bar{\theta}(\bar{\sigma}, \mathbf{Y})$ are easily determined from the current estimate of $\mathbf{Y}$ and its associated state-variable forward solution. However, the state sensitivity matrix $\nabla_{\mathbf{Y}}\bar{\sigma}$ is unknown and must be determined from the simulation functions [4.13] if we are to use equation [4.15] to calculate performance measure sensitivity. In fact, for this study where we are interested in estimating $\partial C/\partial\mathbf{Y}$ representing a single element of $\nabla_{\mathbf{Y}}\bar{\sigma}$, solution for $\nabla_{\mathbf{Y}}\bar{\sigma}$ is sufficient for constructing objective function [4.10]. However, as mentioned previously, the computational burden of determining $\nabla_{\mathbf{Y}}\bar{\sigma}$ by perturbing $\mathbf{Y}$ at each location and solving the primary simulation equations for $\bar{\sigma}$ following each perturbation can be quite prohibitive. Instead, an adjoint approach was used to indirectly solve for concentration sensitivities via equation [4.15].

There are two ways to obtain the performance measure sensitivity matrix $\nabla_{\mathbf{Y}}(\bar{\sigma}, \mathbf{Y})$. The total derivative of $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$ with respect to $\mathbf{Y}$ can be written as:

$$\frac{d\mathbf{g}(\bar{\sigma}, \mathbf{Y})}{d\mathbf{Y}} = \nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) + \nabla_{\sigma} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) \nabla_{\mathbf{Y}} \bar{\sigma} \qquad [4.16]$$

Since $\mathbf{g}(\bar{\sigma}, \mathbf{Y}) \equiv \mathbf{0}$, then $d\mathbf{g}(\bar{\sigma},\mathbf{Y})/d\mathbf{Y} = \mathbf{0}$ and equation [4.16] can be restated as:

$$\nabla_{\sigma} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) \nabla_{\mathbf{Y}} \bar{\sigma} = -\nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) \qquad [4.17]$$

Equation [4.17] can then be solved for $\nabla_{\mathbf{Y}}\bar{\sigma}$, and equation [4.15] used to obtain performance measure sensitivity for known $\nabla_{\mathbf{Y}}\bar{\theta}(\bar{\sigma},\mathbf{Y})$, and $\nabla_{\sigma}\bar{\theta}(\bar{\sigma},\mathbf{Y})$.

Alternatively, one can solve for the adjoint state by adding an additional term to $d\bar{\theta}(\bar{\sigma}, \mathbf{Y})/d\mathbf{Y}$ in equation [4.15] as follows:

$$\frac{d\bar{\theta}(\bar{\sigma}, \mathbf{Y})}{d\mathbf{Y}} = \nabla_{\mathbf{Y}}\bar{\theta}(\bar{\sigma}, \mathbf{Y}) + \nabla_{\sigma} \bar{\theta}(\bar{\sigma}, \mathbf{Y}) \nabla_{\mathbf{Y}}\bar{\sigma} + [\, \bar{\psi}^{*} \nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) + \bar{\psi}^{*} \nabla_{\sigma} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) \nabla_{\mathbf{Y}} \bar{\sigma} \,] \qquad [4.18]$$

where $\bar{\psi}^{*}$ is an arbitrary matrix called the adjoint state or importance function equal to $\nabla_{\mathbf{g}}\bar{\theta}(\bar{\sigma}, \mathbf{Y})$. Again, because $\mathbf{g}(\bar{\sigma}, \mathbf{Y}) \equiv 0$, the term in brackets is equal to 0. Since $\bar{\psi}^{*}$ is an arbitrary function, the unknown $\nabla_{\mathbf{Y}} \bar{\sigma}$ in equation [4.18] can be removed from consideration by setting the sum of terms containing $\nabla_{\mathbf{Y}} \bar{\sigma}$ to zero (Sykes *et al.*, 1985):

$$\nabla_\sigma \bar{\theta}(\bar{\sigma}, \mathbf{Y})\, \nabla_\mathbf{Y} \bar{\sigma} + \bar{\psi}^* \nabla_\sigma \mathbf{g}(\bar{\sigma}, \mathbf{Y})\, \nabla_\mathbf{Y} \bar{\sigma} = 0$$

or

$$\bar{\psi}^* \nabla_\sigma \mathbf{g}(\bar{\sigma}, \mathbf{Y}) = -\nabla_\sigma \bar{\theta}(\bar{\sigma}, \mathbf{Y})$$

which can be restated in a form that allows for matrix solution of $(\bar{\psi}^*)^T$:

$$\nabla_\sigma \mathbf{g}(\bar{\sigma}, \mathbf{Y})^T\, (\bar{\psi}^*)^T = -\nabla_\sigma \bar{\theta}(\bar{\sigma}, \mathbf{Y})^T \qquad [4.19]$$

For Dirichlet or Neuman boundary conditions associated with the primary problem, boundary values of $(\bar{\psi}^*)^T$ are set equal to zero to reflect lack of change in the performance measure $\bar{\theta}(\bar{\sigma}, \mathbf{Y})$ caused by changes in fluid or solute flux $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$. Solution of equation [4.19] is referred to as solving the adjoint problem from the left.

In order to understand the physical meaning of the adjoint state $\bar{\psi}^* = \nabla_\mathbf{g} \bar{\theta}(\bar{\sigma}, \mathbf{Y})$, one can view the simulation equations $\mathbf{g}(\bar{\sigma}, \mathbf{Y}) = \mathbf{0}$ as mass balance residuals. A change in $\mathbf{g_{flow}}(\bar{h}, \mathbf{Y})$, for example, can be viewed as a perturbation in fluid flux, while a change in $\mathbf{g_{tran}}(\bar{v})$ is equivalent to perturbation of mass flux during the current time step. $\bar{\psi}^* = \nabla_\mathbf{g} \bar{\theta}(\bar{\sigma}, \mathbf{Y})$ then represents the change in performance measure produced by unit fluxes of fluid, velocity, and contaminant mass everywhere in the physical domain. When the performance measure is the state variable vector, $\bar{\theta}(\bar{\sigma}, \mathbf{Y}) = \bar{\sigma}$, and $\bar{\psi}^*$ reduces to Green's functions for the flow, velocity, and transport equations. The utility of the adjoint state $\bar{\psi}^*$ is that it integrates all effects of system perturbation, whether they be

perturbations in state variables or parameters, into a single mass-balance residual $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$. Changes in the performance measure due to any system parameter or variable can easily be determined because the residuals $\mathbf{g}(\bar{\sigma},\mathbf{Y})$ are known functions of the parameters and variables. It should be noted, however, that $\bar{\psi}^*$ only locally represents the adjoint state for the assumed parameter vector **Y**. That is, a change in **Y** will yield a different adjoint state $\bar{\psi}^*$.

Following solution of equation [4.19] for the adjoint state nodal values, $\bar{\psi}^*$ is substituted into equation [4.18] to obtain performance measure sensitivity:

$$\frac{d\bar{\theta}(\bar{\sigma}, \mathbf{Y})}{d\mathbf{Y}} = \nabla_{\mathbf{Y}} \bar{\theta}(\bar{\sigma}, \mathbf{Y}) + \bar{\psi}^* \nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) \qquad [4.20]$$

where $\nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y})$ is evaluated at the state vector $\bar{\sigma}$, obtained from solution of the primary problem $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$ at the current estimate of **Y** and, as before, all terms involving $\nabla_{\mathbf{Y}}\bar{\sigma}$ sum to zero. Thus, the burden of calculating changes in $\bar{\sigma}$ with respect to each of the M values of Y is never a consideration.

Note that $d\bar{\theta}(\bar{\sigma},\mathbf{Y})/d\mathbf{Y}$ can either be determined from solution of equations [4.17] and [4.15], or from solution of equations [4.19] and [4.20]. Because $\nabla_{\mathbf{Y}} \bar{\sigma}$ and $(\bar{\psi}^*)^T$ are matrices, and not vectors, equations [4.17] and [4.19] can both be viewed as a series of matrix equations of the form $\mathbf{Ax=b}$, where **A** is a matrix and **x** and **b** are column vectors. Whether one solves equation [4.17] or [4.19] thus depends on how many columns occur in $\nabla_{\mathbf{Y}}\bar{\sigma}$ and $(\bar{\psi}^*)^T$. If the number of parameters in vector **Y** is

greater than the number of performance measures in vector $\bar{\theta}(\bar{\sigma}, \mathbf{Y})$, then the matrix $(\bar{\psi}^*)^T = [\nabla_g \bar{\theta}(\bar{\sigma}, \mathbf{Y})]^T$ has fewer columns than matrix $\nabla_Y \bar{\sigma}$ and equation [4.19] requires less computational effort. When the dimension of **Y** is less than the dimension of $\bar{\theta}(\bar{\sigma},\mathbf{Y})$, $\nabla_Y \bar{\sigma}$ has fewer columns than $(\bar{\psi}^*)^T$ and equation [4.17] is more efficiently solved. For this particular study, $\bar{\theta}(\bar{\sigma},\mathbf{Y})=C$ is a scalar of dimension 1, and always smaller than the dimension of **Y** (M). Equation [4.19] was therefore solved for $(\bar{\psi}^*)^T$, followed by substitution into equation [4.20]. Note that the number of additions and multiplications involved in equations [4.15] and [4.20] are the same and equal to n and n*M, respectively. Post-solution calculations thus have no bearing on whether one should solve for $\nabla_Y \bar{\sigma}$ or $(\bar{\psi}^*)^T$.

To solve matrix equation [4.19] for $(\bar{\psi}^*)^T$, the right-hand-side and coefficient matrices must be specified. For the general case of $\bar{\theta}(\bar{\sigma},\mathbf{Y})=\bar{\sigma}$, the right-hand-side reduces to an identity load matrix given by, $\nabla_\sigma \bar{\theta}(\bar{\sigma}, \mathbf{Y})=\nabla_\sigma \bar{\sigma}$, which can be represented as:

$$\nabla_\sigma \bar{\sigma} = \begin{bmatrix} \frac{\partial \bar{h}}{\partial \bar{h}} & \frac{\partial \bar{h}}{\partial \bar{v}_x} & \frac{\partial \bar{h}}{\partial \bar{v}_y} & \frac{\partial \bar{h}}{\partial \bar{c}_0} & \frac{\partial \bar{h}}{\partial \bar{c}} \\ \frac{\partial \bar{v}_x}{\partial \bar{h}} & \frac{\partial \bar{v}_x}{\partial \bar{v}_x} & \frac{\partial \bar{v}_x}{\partial \bar{v}_y} & \frac{\partial \bar{v}_x}{\partial \bar{c}_0} & \frac{\partial \bar{v}_x}{\partial \bar{c}} \\ \frac{\partial \bar{v}_y}{\partial \bar{h}} & \frac{\partial \bar{v}_y}{\partial \bar{v}_x} & \frac{\partial \bar{v}_y}{\partial \bar{v}_y} & \frac{\partial \bar{v}_y}{\partial \bar{c}_0} & \frac{\partial \bar{v}_y}{\partial \bar{c}} \\ \frac{\partial \bar{c}_0}{\partial \bar{h}} & \frac{\partial \bar{c}_0}{\partial \bar{v}_x} & \frac{\partial \bar{c}_0}{\partial \bar{v}_y} & \frac{\partial \bar{c}_0}{\partial \bar{c}_0} & \frac{\partial \bar{c}_0}{\partial \bar{c}} \\ \frac{\partial \bar{c}}{\partial \bar{h}} & \frac{\partial \bar{c}}{\partial \bar{v}_x} & \frac{\partial \bar{c}}{\partial \bar{v}_y} & \frac{\partial \bar{c}}{\partial \bar{c}_0} & \frac{\partial \bar{c}}{\partial \bar{c}} \end{bmatrix} = \begin{bmatrix} \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} & \mathbf{0} \\ \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{0} & \mathbf{I} \end{bmatrix} = \mathbf{I}_{\mathbf{total}} \qquad [4.21]$$

where **I** is an identity matrix of size n/5-by-n/5, **0** is of dimension n/5-by-n/5, $\mathbf{I_{total}}$ is a n-by-n identity matrix, and the velocity vector has been resolved into its component vectors $\bar{v}_x$ and $\bar{v}_y$. Note that all off-diagonal submatrices in matrix [4.21] are identically equal to zero because state variables are related to other state variables only through the simulation equations $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$. For example, while changes in head $\bar{h}$ will affect x-direction velocity $\bar{v}_x$, they will not do so directly, but rather through the flow simulation equation $\mathbf{g_{darcy}}(\bar{\sigma}, \mathbf{Y})$.

Finally, solution of equation [4.19] requires that the derivative matrix $\nabla_\sigma \mathbf{g}(\bar{\sigma}, \mathbf{Y})^T$ be determined. The simulation functions $\mathbf{g}(\bar{\sigma}, \mathbf{Y})=\mathbf{0}$ for steady-state fluid flow, darcy velocity, and concentration take the following forms:

$$
\begin{aligned}
\mathbf{g_{flow}}(\bar{h}, \mathbf{Y}) &= \mathbf{A_{flow}}(\mathbf{Y})\ \bar{h} - \bar{b}_{flow}(\mathbf{Y}) = \mathbf{0} \\
\mathbf{g_{darcy}}(\bar{h}, \mathbf{Y}) &= \mathbf{I}\ \bar{v} - \mathbf{A_{darcy}}(\mathbf{Y})\ \bar{h} = \mathbf{0} \\
\mathbf{g_{tran0}}(\bar{v}, \bar{c}_0, \mathbf{Y}) &= \mathbf{A_{trans}}(\bar{v})\ \bar{c}_0 + \mathbf{A_{trant}}\ \bar{c}_{-1} - \bar{b}_{to}(\mathbf{Y}) = \mathbf{0} \\
\mathbf{g_{tran}}(\bar{v}, \bar{c}_0, \bar{c}, \mathbf{Y}) &= \mathbf{A_{trans}}(\bar{v})\ \bar{c} + \mathbf{A_{trant}}\ \bar{c}_0 - \bar{b}_t(\mathbf{Y}) = \mathbf{0}
\end{aligned}
\qquad [4.22]
$$

where:

$\mathbf{A_{flow}(Y)}$ = coefficient matrix for finite-difference steady-state flow equation [L^2/T]

$\mathbf{A_{darcy}(Y)}$ = coefficient matrix resulting from finite-difference discretization of Darcy's law in x- and y-directions [1/T]

$\mathbf{A_{trans}}(\bar{v})$ = coefficient matrix corresponding to finite-difference spatial derivatives of the advection-dispersion equation [1/T]

$\mathbf{A_{trant}}$ = coefficient matrix corresponding to finite-difference temporal derivatives of the advection-dispersion equation, where each coefficient is equal to $1/\Delta t$ [1/T]

$\bar{b}_{flow}(\mathbf{Y})$ = right-hand-side vector of known values in flow equation [L^3/T]

$\bar{b}_{to}(\mathbf{Y})$ = right-hand-side vector of known values in advection-dispersion equation for prior concentrations [M/TL^3]

$\bar{b}_{t}(\mathbf{Y})$ = right-hand-side vector of known values in advection-dispersion equation for current concentrations [M/TL^3]

$\bar{c}_{-1}$ = vector of concentrations at two time steps prior to current time [M/L^3]

$\bar{c}_0$ = vector of concentrations at one time step prior to current time [M/L^3]

$\mathbf{I}$ = identity matrix having the same dimension as $\mathbf{A_{darcy}(Y)}$

and where other state variables are defined previously.

Derivatives of these functions with respect to σ are presented in Appendix C. Some of these derivatives depend on state variables predicted from the primary simulation equations $\mathbf{g}(\bar{\sigma}, \mathbf{Y})=\mathbf{0}$ using the current estimate of **Y**. Concentrations predicted from particle tracking were used to assess the derivatives despite the fact that the forms of $\mathbf{g_{tran0}}(\bar{\sigma}, \mathbf{Y})$ and $\mathbf{g_{tran}}(\bar{\sigma}, \mathbf{Y})$ contain dispersion terms. This approach

appeared to present no numerical difficulties. Note that, since the multigrid flow solver and the particle-tracking algorithm used to predict the state variables were based on finite-difference discretization techniques, the $\mathbf{g}(\bar{\sigma},\mathbf{Y})$ functions were likewise defined using a finite-difference approach.

Substituting $\nabla_{\sigma}\bar{\theta}(\bar{\sigma},\mathbf{Y})^{T}=-\mathbf{I}_{\mathbf{total}}$ for the right-hand-side of equation [4.19], and expanding the terms of both the matrices $\nabla_{\sigma}\mathbf{g}(\bar{\sigma},\mathbf{Y})^{T}$ and $(\bar{\psi}^{*})^{T}$, equation [4.19] can be written as:

$$\begin{bmatrix} \frac{\partial g_{flow}}{\partial \bar{h}} & \frac{\partial g_{darcy,x}}{\partial \bar{h}} & \frac{\partial g_{darcy,y}}{\partial \bar{h}} & \frac{\partial g_{tran0}}{\partial \bar{h}} & \frac{\partial g_{tran}}{\partial \bar{h}} \\ \frac{\partial g_{flow}}{\partial \bar{v}_x} & \frac{\partial g_{darcy,x}}{\partial \bar{v}_x} & \frac{\partial g_{darcy,y}}{\partial \bar{v}_x} & \frac{\partial g_{tran0}}{\partial \bar{v}_x} & \frac{\partial g_{tran}}{\partial \bar{v}_x} \\ \frac{\partial g_{flow}}{\partial \bar{v}_y} & \frac{\partial g_{darcy,x}}{\partial \bar{v}_y} & \frac{\partial g_{darcy,y}}{\partial \bar{v}_y} & \frac{\partial g_{tran0}}{\partial \bar{v}_y} & \frac{\partial g_{tran}}{\partial \bar{v}_y} \\ \frac{\partial g_{flow}}{\partial \bar{c}_0} & \frac{\partial g_{darcy,x}}{\partial \bar{c}_0} & \frac{\partial g_{darcy,y}}{\partial \bar{c}_0} & \frac{\partial g_{tran0}}{\partial \bar{c}_0} & \frac{\partial g_{tran}}{\partial \bar{c}_0} \\ \frac{\partial g_{flow}}{\partial \bar{c}} & \frac{\partial g_{darcy,x}}{\partial \bar{c}} & \frac{\partial g_{darcy,y}}{\partial \bar{c}} & \frac{\partial g_{tran0}}{\partial \bar{c}} & \frac{\partial g_{tran}}{\partial \bar{c}} \end{bmatrix} \begin{bmatrix} \frac{\partial \bar{h}}{\partial g_{flow}} & \frac{\partial \bar{v}_x}{\partial g_{flow}} & \frac{\partial \bar{v}_y}{\partial g_{flow}} & \frac{\partial \bar{c}_0}{\partial g_{flow}} & \frac{\partial \bar{c}}{\partial g_{flow}} \\ \frac{\partial \bar{h}}{\partial g_{darcy,x}} & \frac{\partial \bar{v}_x}{\partial g_{darcy,x}} & \frac{\partial \bar{v}_y}{\partial g_{darcy,x}} & \frac{\partial \bar{c}_0}{\partial g_{darcy,x}} & \frac{\partial \bar{c}}{\partial g_{darcy,x}} \\ \frac{\partial \bar{h}}{\partial g_{darcy,y}} & \frac{\partial \bar{v}_x}{\partial g_{darcy,y}} & \frac{\partial \bar{v}_y}{\partial g_{darcy,y}} & \frac{\partial \bar{c}_0}{\partial g_{darcy,y}} & \frac{\partial \bar{c}}{\partial g_{darcy,y}} \\ \frac{\partial \bar{h}}{\partial g_{tran0}} & \frac{\partial \bar{v}_x}{\partial g_{tran0}} & \frac{\partial \bar{v}_y}{\partial g_{tran0}} & \frac{\partial \bar{c}_0}{\partial g_{tran0}} & \frac{\partial \bar{c}}{\partial g_{tran0}} \\ \frac{\partial \bar{h}}{\partial g_{tran}} & \frac{\partial \bar{v}_x}{\partial g_{tran}} & \frac{\partial \bar{v}_y}{\partial g_{tran}} & \frac{\partial \bar{c}_0}{\partial g_{tran}} & \frac{\partial \bar{c}}{\partial g_{tran}} \end{bmatrix} = -\mathbf{I}_{\mathbf{total}} \qquad [4.23]$$

where expansion of the terms on the left side yields the desired identity.

Figure 4.3 shows the form of the adjoint state coefficient matrix, $\nabla_{\sigma}\, g(\bar{\sigma},\, \mathbf{Y})^{T}$, when solving equation [4.19] for the adjoint state. The matrix corresponds to a 5-by-5 finite-difference problem with nonzero contaminant concentration at the monitoring well during current and previous times. Because less than 1% of the adjoint matrix elements are nonzero, solution is best performed using sparse-matrix storage techniques. However, since the adjoint matrix is not symmetric, an efficient iterative

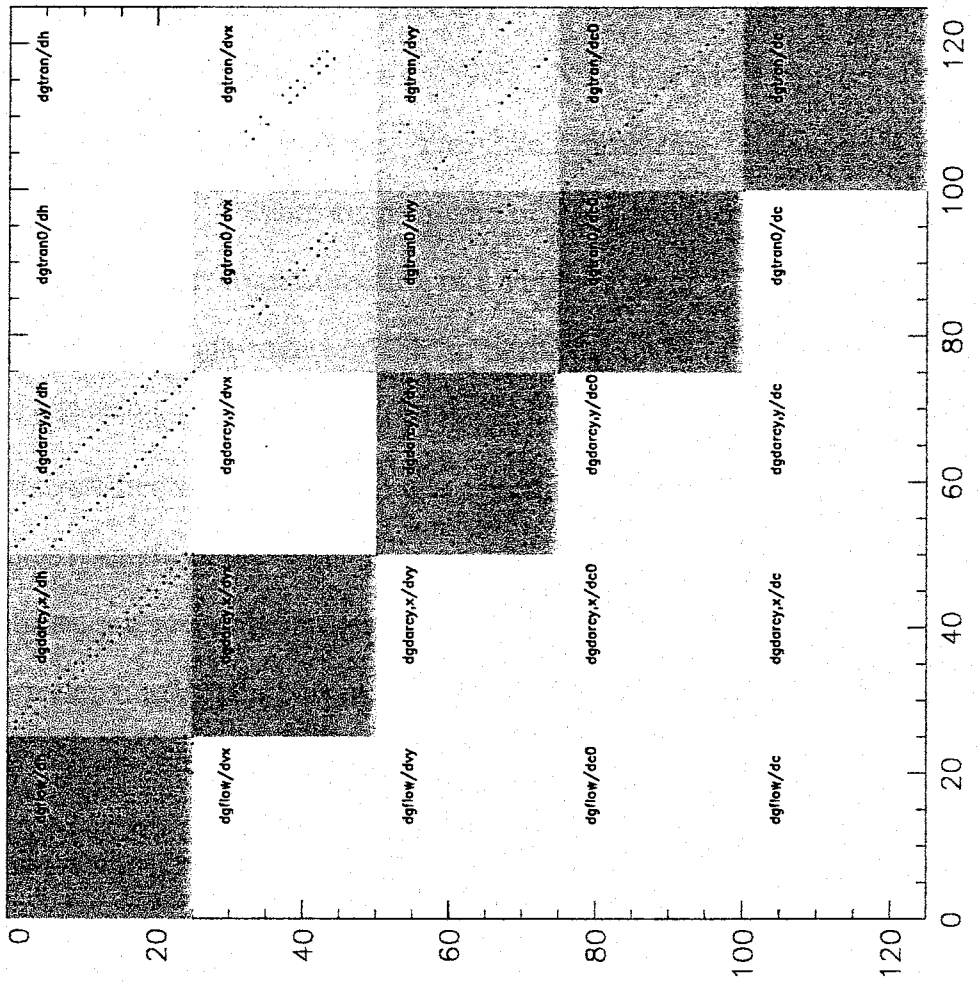


Figure 4.3. Adjoint Coefficient Matrix Structure (Solving from Left).

solver such as the conjugate-gradient (CG) algorithm cannot be applied to the problem of solving equation [4.23]. The generalized minimum residual equation solver (GMRES), which does not require a symmetric matrix, was instead used to solve the adjoint problem. The solver was developed at Los Alamos National Laboratory using the methodology of Saad and Schultz (1986). Note that while equation [4.17] can be decomposed into three separate, uncoupled problems, equation [4.19] cannot. Solution of equation [4.17] for each column of $\nabla_{\mathbf{Y}} \bar{\sigma}$ produces sensitivities of all state variables with respect to perturbation of a single value of Y. When one is interested in sensitivity of a single state variable to perturbation in all values of Y, solution of equation [4.17] must be performed M times, a huge computational effort that cannot usually be justified unless disk space is at a premium. Instead, equation [4.19] is solved, but without benefit of decoupling.

Since we are only interested in obtaining sensitivities of compliance point concentration C to **Y** at the current time, equation [4.23] is solved for the column vector of $(\bar{\psi}^*)^T$ corresponding to $\partial C/\partial \mathbf{g}$:

$$\begin{bmatrix} \frac{\partial g_{flow}}{\partial \bar{h}} & \frac{\partial g_{darcy,x}}{\partial \bar{h}} & \frac{\partial g_{darcy,y}}{\partial \bar{h}} & \frac{\partial g_{tran0}}{\partial \bar{h}} & \frac{\partial g_{tran}}{\partial \bar{h}} \\ \frac{\partial g_{flow}}{\partial \bar{v}_x} & \frac{\partial g_{darcy,x}}{\partial \bar{v}_x} & \frac{\partial g_{darcy,y}}{\partial \bar{v}_x} & \frac{\partial g_{tran0}}{\partial \bar{v}_x} & \frac{\partial g_{tran}}{\partial \bar{v}_x} \\ \frac{\partial g_{flow}}{\partial \bar{v}_y} & \frac{\partial g_{darcy,x}}{\partial \bar{v}_y} & \frac{\partial g_{darcy,y}}{\partial \bar{v}_y} & \frac{\partial g_{tran0}}{\partial \bar{v}_y} & \frac{\partial g_{tran}}{\partial \bar{v}_y} \\ \frac{\partial g_{flow}}{\partial \bar{c}_0} & \frac{\partial g_{darcy,x}}{\partial \bar{c}_0} & \frac{\partial g_{darcy,y}}{\partial \bar{c}_0} & \frac{\partial g_{tran0}}{\partial \bar{c}_0} & \frac{\partial g_{tran}}{\partial \bar{c}_0} \\ \frac{\partial g_{flow}}{\partial \bar{c}} & \frac{\partial g_{darcy,x}}{\partial \bar{c}} & \frac{\partial g_{darcy,y}}{\partial \bar{c}} & \frac{\partial g_{tran0}}{\partial \bar{c}} & \frac{\partial g_{tran}}{\partial \bar{c}} \end{bmatrix} \begin{bmatrix} \frac{\partial C}{\partial g_{flow}} \\ \frac{\partial C}{\partial g_{darcy,x}} \\ \frac{\partial C}{\partial g_{darcy,y}} \\ \frac{\partial C}{\partial g_{tran0}} \\ \frac{\partial C}{\partial g_{tran}} \end{bmatrix} = \begin{bmatrix} \frac{\partial C}{\partial \bar{h}} \\ \frac{\partial C}{\partial \bar{v}_x} \\ \frac{\partial C}{\partial \bar{v}_y} \\ \frac{\partial C}{\partial \bar{c}_0} \\ \frac{\partial C}{\partial \bar{c}} \end{bmatrix} = (\psi_C{}^*)^T = -\mathbf{i} \qquad [4.24]$$

where $\mathbf{i}$ is a column vector comprised entirely of zeros, except at the location corresponding to the monitoring well concentration at the current time, where the element is equal to one. The column of $(\bar{\psi}^*)^T$ corresponding to monitoring well concentration at the current time is $(\bar{\psi}_C{}^*)^T$. Once $(\bar{\psi}_C{}^*)^T$ is obtained from equation [4.24], equation [4.20] is used to estimate the monitoring well concentration sensitivities, where $\bar{\theta} \equiv C$:

$$\frac{\partial C}{\partial \mathbf{Y}} = \bar{\psi}_C{}^* \nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y}) \qquad [4.25]$$

and where $\nabla_{\mathbf{Y}} \bar{\theta}(\bar{\sigma}, \mathbf{Y}) = \mathbf{0}$ when there is no explicit, direct impact on C as a result of perturbing $\mathbf{Y}$. Since the simulation equations $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$ are known functions of $\mathbf{Y}$, the derivative matrix $\nabla_{\mathbf{Y}} \mathbf{g}(\bar{\sigma}, \mathbf{Y})$ in the above equation was easily derived (see Appendix C).

It should be noted that, despite the fact that off-diagonal terms of the $\mathbf{Y_c}$ covariance matrix are ignored when translating log conductivity spatial variability to concentration prediction uncertainty, covariance structure is implicitly present in the forward solutions used to solve for adjoint states. Thus, the sensitivities automatically account for joint structure in the manner that they direct the design strategy.

4.6 TRUE CONDUCTIVITY REALIZATIONS

Using the multiple-scale, multiple-realization design given by objective function [4.10] subject to constraints [4.10a] through [4.10f], optimal conductivity measurement design was identified sequentially on the basis of hypothetical true realizations for two-scale log conductivity presented in Figure 4.4. The true large-scale realization was generated as a sample of a 14-by-14 second-order MRF process with state levels of $-7.0, -6.0, -5.0, -4.0, -3.0$ ln m/s, pair-potential neighborhood interaction energy, and MRF parameters 8.0, 8.0, 0.5, and -8.0, using the Metropolis algorithm with 5,000 attempted exchanges per site. The true two-scale realization was obtained by embedding oblique small-scale 10-by-10 gaussian structure with space increments dx=dy=1 m in each large-scale MRF block using the approach outlined in Chapter 2. This involved adding the zero-mean gaussian random variable to the large-scale value of **Y** associated with the MRF block in which the small-scale realization was located. **Y** at the MRF block scale was then equal to the spatial arithmetic average of small-scale **Y** over the block. Note that arithmetic averaging of **Y** translates into geometric averaging of **K**.

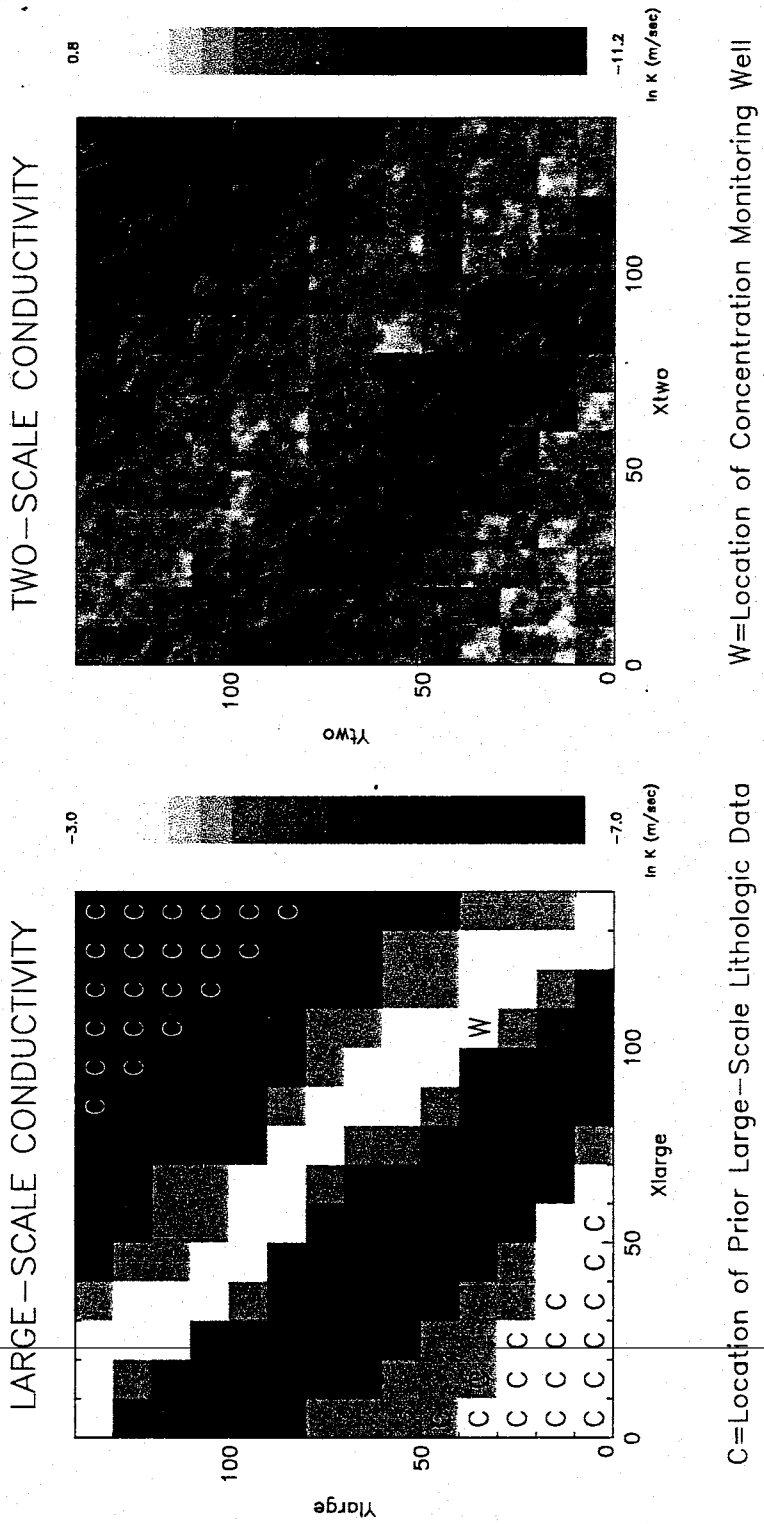


Figure 4.4 True Conductivity Realizations at Two Scales.

In constructing both the true and the conditional realizations used to constrain the design, small-scale gaussian structural parameters were assumed dependent on the large-scale MRF random variable, with **Y** variance and principal correlation scales in the N45^{0}W (λ_L) and N45^{0}E (λ_T) directions listed in Table 4.1. In real field situations for which no such dependency of small-scale structure on the large-scale random variable can necessarily be expected to occur, this assumption introduces artificial scale-dependency that is gradually removed as measurements are collected independently at each scale during the sequential design.

The large-scale true realization was conditioned on prior information made available from a hypothetical preliminary data acquisition effort, using the technique discussed in Chapter 3 to condition the MRF realization on large-scale data. If small-scale prior information is available, kriging methods outlined in Chapter 2 can be used to condition both large-scale and small-scale components of the true two-scale

Table 4.1. Statistical Parameters for Design Problem 1.

Large-Scale MRF Y State (ln m/s)	σ^2_Y (ln m/s)^2	λ_L (m)	λ_T (m)
−7.0	2.0	5.	1.
−6.0	2.0	5.	2.
−5.0	2.0	3.	1.
−4.0	2.0	3.	3.
−3.0	2.0	5.	5.

Y realization on small-scale prior information. Prior data can be viewed as defining an information point about which the measurement design is incrementally expanded, a perspective that highlights the marginal nature of the information that is gained from any measurement. For this study, the large C's shown in Figure 4.4 represent large-scale conditioning locations at which soft lithologic information is available from a preliminary mapping effort. If the NW-SE trending diagonal through the large-scale realization corresponds to the axis of a depositional basin, for example, the C's might represent locations of outcrop data collected along the margins of the basin from geologic mapping. Prior conductivity information can also be introduced to the optimization by conditioning on available state-variable measurements such as head or concentration using cokriging or extended Kalman filtering techniques (Hoeksema and Kitanidis, 1985; Graham and McLaughlin, 1989b).

There may also be access to information that allows us to determine the form of the covariance function for small-scale gaussian structure corresponding to at least one of the large-scale MRF site states. This follows the work of Rajaram and McLaughlin (1990), who extrapolated small-scale structure observed over a small intensively-sampled area to a larger area. The covariance structure can perhaps be generalized to other states by scaling the variance and correlation scale according to the mean conductivities of the MRF states. Alternatively, if no small-scale prior information is available, gaussian statistics can be borrowed from other aquifers believed to have formed under similar geologic conditions. As a final recourse, one can always guess at the small-scale statistics and update them as new data are acquired.

In this hypothetical study, when a piece of conductivity data is 'measured' off one of the two true, scale-dependent realizations, it conforms exactly to the assumed small-scale gaussian covariance structure or large-scale MRF parameter set used to generate the realization. In reality, the structure introduced by a newly-acquired data value may conflict with the prior structural parameters and will not necessarily represent a sample from the assumed joint distribution of two-scale log conductivity. How does conductivity data that contradicts the assumed structural parameters affect the optimal design if structural parameters are not re-estimated using this new data? Rouhani (1983) addressed this question by defining three distinct spaces: the data or input space, the structural parameter space, and the action or design space based on model predictions. By perturbing data values with noise, he demonstrated that the structural parameter space exhibited a great deal of instability with respect to the noisy input space. His findings are substantiated by those of Kitanidis (1983). However, the action or design space was very stable with respect to the parameter space, suggesting that even if the structure imposed by a new data value is not particularly representative of a sample from the joint pdf implied by the structural parameters, it will generally not significantly distort designs based on flow or transport equation predictions. For purposes of this study, this was assumed to be the case.

The 14-by-14 large-scale and 140-by-140 two-scale realizations shown in Figure 4.4 represent scale-variant samples of the two-scale conductivity process. The $Y_{js,k}^{(s)}$, k=1,2,. . . , NR=10 realization constraints were sequentially conditioned at locations j_s=1, 2, . . . , M_s , and scales s=1 (large-scale) and 2 (two-scale), where

M_1=14*14 and M_2=140*140, against these true realizations during each iteration of the measurement design. Of a total true **Y** variance amounting to 3.5 (ln m/s)^2, 1.9 (ln m/s)^2 resided at the large scale and 1.6 (ln m/s)^2 at the small-scale. The same techniques used to condition on prior data were employed to condition on newly-acquired conductivity data collected from Figure 4.4 during the course of the design. However, in order to prevent the occurrence of an ill-conditioned kriging matrix, a constraint was imposed to prohibit small-scale measurement at any location immediately adjacent to a previous small-scale conditioning location. It should be noted that random measurement error in **Y** was not incorporated into the sequential design, although it could have easily been included at the small scale by introducing a nugget to the small-scale gaussian covariance functions. As a result, $\sigma^2_{Y(2)}$ was reduced to zero rather than to instrument measurement error following small-scale measurement.

As a final consideration, incorporation of conditioning data from Figure 4.4 at both scales produced spatially-distributed variance $\sigma_{Y(s)}{}^2$ with respect to location index j_s, and required that ensemble variances be calculated at each location and at each scale across the NR-realization ensemble at the end of each design iteration. Ensemble variances were considered superior to kriging variances, which are spatial rather than true ensemble statistics, particularly given that ergodicity of the two-scale log conductivity process was not likely.

Figure 4.5 presents two scales of transport solutions corresponding to the true **Y** realizations with an instantaneous source of contamination. Prior to transport,

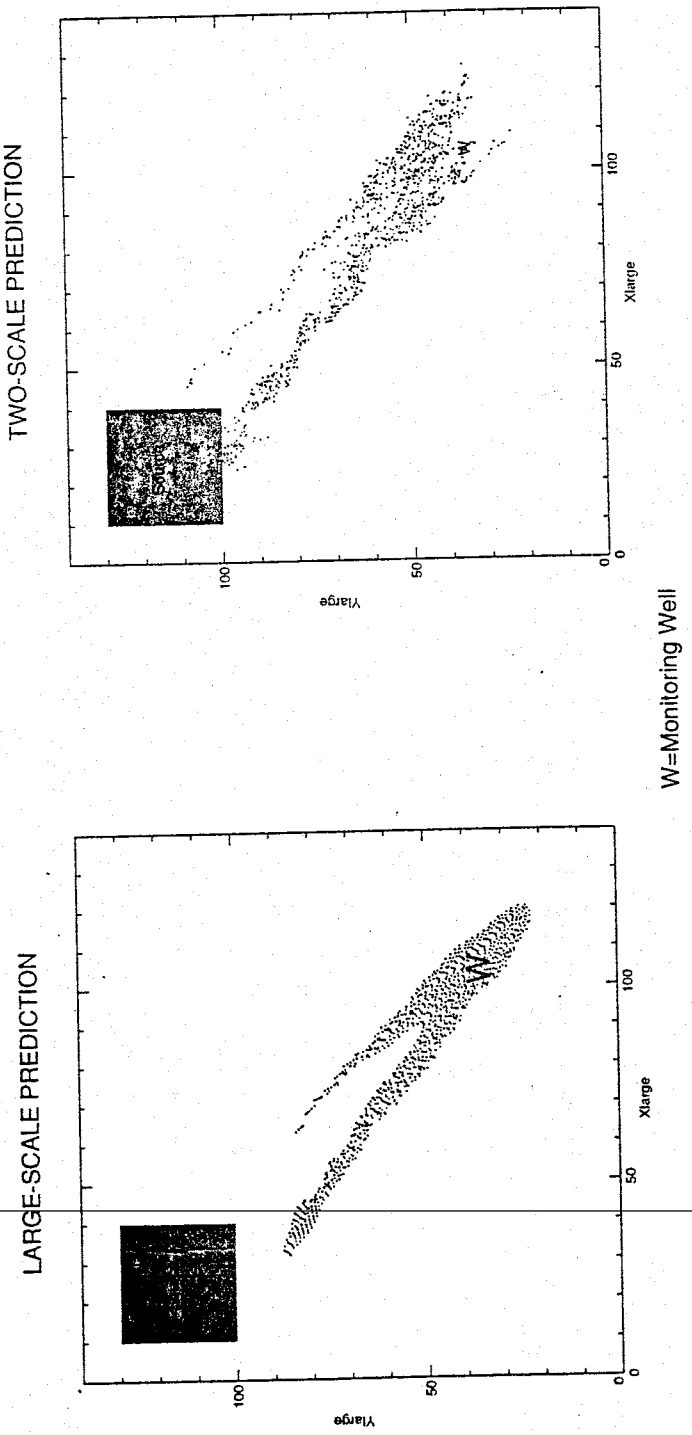


Figure 4.5 True Plume Realizations at Two Scales of Simulation.

steady-state velocities were predicted at each scale of resolution using a multigrid flow solver developed by Stephen Shaffer at New Mexico Institute of Mining and Technology that is able to account for extreme heterogeneities in conductivity with little loss of accuracy. An average, uniform steady-state darcy flux from the northwest to southeast equal to 1.0 X 10^{-6} m/s was imposed on both large- and two-scale true realizations for flow simulation. Following prediction of the steady-state head distribution in both realizations, a particle tracker developed by Schafer-Perini and Wilson (1991) was used to simulate transient migration of a conservative contaminant. A total of 1024 particles representing a concentration of 1.0 mg/*l* were released instantaneously over the 30m-by-30m source area shown in Figure 4.5, and advective transport through both realizations predicted at the end of 230 days. For the design problem, it was assumed that source strength and time since release were known with certainty. However, as this will rarely be the case, the release history can be estimated from concentration observations using Bayesian or other geostatistical techniques (Snodgrass and Kitanidis, 1997). Skaggs and Kabala (1994) provide a comprehensive review of methodology used to estimate source release history.

As indicated in Figure 4.5, the monitoring well at which concentration prediction uncertainty was minimized was located at x=105 m and y=35 m within large-scale MRF block (11,4). Costs of measuring at the larger MRF 10-m scale and the smaller 1-m gaussian scale were assumed to be $10,000/measurement and $100/measurement, respectively. Steady-state flow and transient advective transport simulations similar to those conducted for the true Y realization were performed for

each of the NR=10 **Y** realizations during each iteration of the design. The resulting head and concentration solutions were subsequently used to define coefficients for the adjoint problem given by equation [4.24], and the resulting vector $\overline{\psi}_C^*$ used to determine sensitivities via equation [4.25].

In order to find the single optimal measurement location and scale most common to all NR realizations, the probability mass function $p(R_{Y(s),js,k})$ of weighted squared sensitivity rankings was determined by constructing a histogram across the ensemble of realizations for each scale and for each location. Following determination of $p(R_{Y(s),js,k})$, modified maximum-likelihood ranking $RML^{(s)}_{js}$ was assigned to each location and scale and used to define the $f(\mathbf{Y}^{(s)})$ weighting functions given by equation [4.10a].

4.7 OPTIMIZATION RESULTS

4.7.1 Optimal Conductivity Measurement Network Design

Results of the first conductivity measurement design are presented in Figure 4.6 and Table 4.2. Initially, optimal measurement occurred at the small scale immediately to the east and west of the monitoring well. After two iterations of small-scale measurement, the algorithm reverted to large-scale measurement somewhat north and east of the well block. The two early iterations of small-scale measurement within the well block were apparently sufficient to reduce large-scale

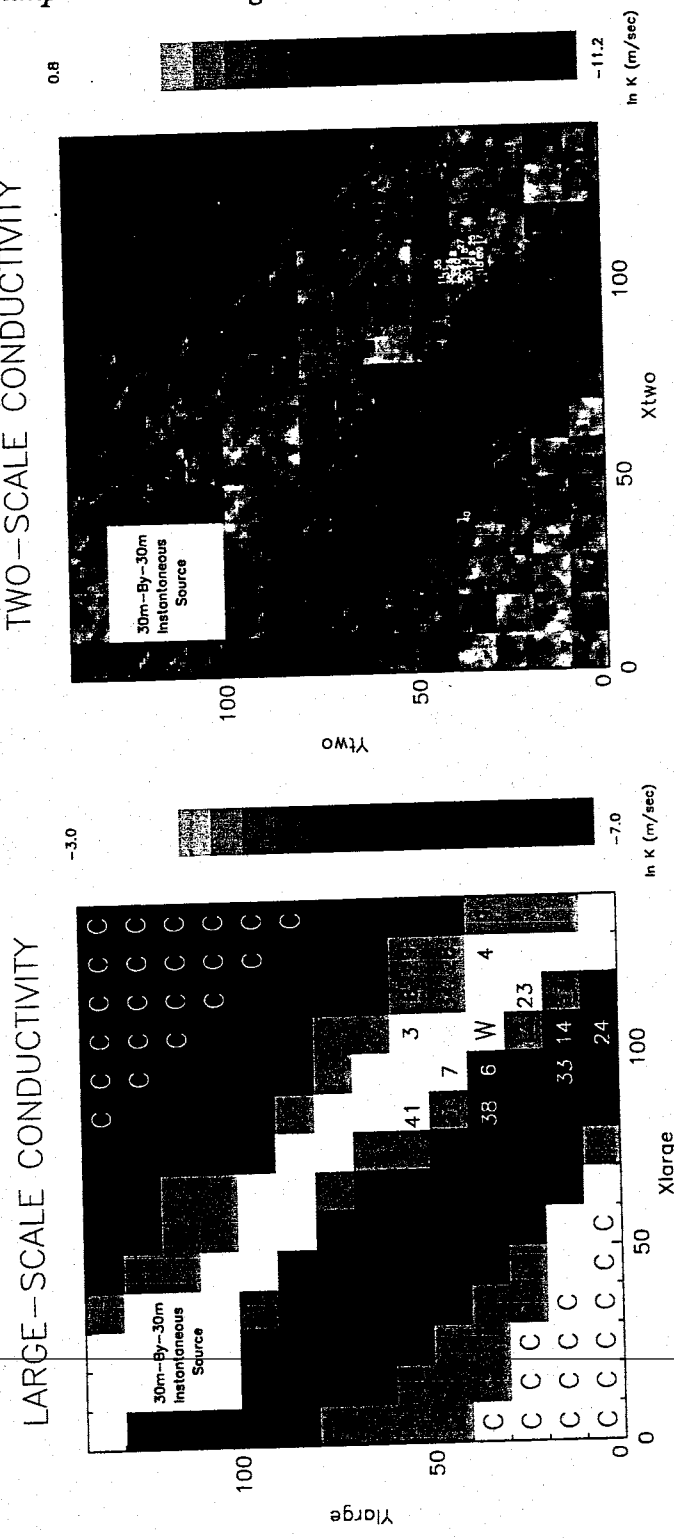


Figure 4.6. Conductivity Measurement Design Results.

Table 4.2. Results of Conductivity Measurement Design.

Sequential Design Iteration	Measurement Scale	Measurement Location Index (i_s, j_s)
1	small	106,35
2	small	104,35
3	large	11,6
4	large	13,4
5	small	108,35
6	large	10,4
7	large	10,5
8	large	12,6
9	large	14,4
10	small	105,37
11	small	101,41
12	small	105,33
13	large	11,3
14	large	11,2
15	small	104,31
16	small	106,31
17	small	111,31
18	small	107,38
19	small	107,33
20	small	102,34
21	large	9,5
22	large	11,7
23	large	12,3
24	large	11,1
25	small	111,33
26	small	105,39
27	small	110,36
28	large	13,5
29	large	13,6
30	small	101,36
31	small	103,40
32	small	103,37
33	large	10,2
34	large	12,5
35	small	105,42
36	small	101,39
37	large	12,2
38	large	9,4
39	small	108,31
40	small	40,36

well block variance below that which would justify subsequent large-scale measurement in that block.

The initial preference for small-scale measurement near the well prior to collection of more informative large-scale measurement was presumably due to the lower cost associated with small-scale measurement. Moreover, small-scale measurement helped to define how many particles were deflected into the point well, once they entered the well block. Small-scale measurement served a dual function in establishing crucial small-scale conductivity near the well and providing a rough estimate of conductivity for the large-scale well block, with both types of information collected at low cost. Once the design algorithm gained information about small-scale transport behavior near the well and provided an estimate of large-scale conductivity near the well, the remaining iterations were about evenly distributed between the two scales of measurement. This tendency for both scales to be equally favored was related to the equal proportions of variance residing at both scales. Larger concentration sensitivities associated with the large scale were apparently offset by larger costs of measurement incurred at the large scale.

As expected, the specific patterns of measurement at both scales were dictated by the geometry of the heterogeneities. After the first two large-scale measurements were collected along the principal longitudinal direction of large-scale anisotropy, large-scale measurement began to work its way out perpendicular to this direction. Such a pattern of measurement can be explained by noting the design algorithm will tend to measure **Y** along the centerline of the plume, where contaminant mass is at a

maximum and sampling of Y by the plume most intense. For this particular problem, the plume centerline happened to coincide with a principal direction of large-scale anisotropy, where correlation was at its highest. By measuring along the plume centerline, the design was also measuring along a direction where Y correlation was at its highest and the decrease in marginal information was most rapid, encountering information redundancy at the large scale. In attempting to avoid this large-scale redundancy, the design sought to measure in a direction perpendicular to the principal longitudinal direction of anisotropy and transport, eventually forfeiting preference for large-scale measurement as conditioning caused large-scale variance to decrease throughout the measurement domain.

A pattern similar to that observed at the large scale emerged during small-scale measurement in the well block, with early small-scale measurement occurring along the principal longitudinal direction of heterogeneity and transport and subsequent measurement tending to occur transverse to that. In other words, the large-scale pattern of measurement was imitated at the small scale, because the principal direction of anisotropy at the small-scale was the same as that occurring at the large scale.

Note that the first two small-scale measurements bound the well located at x_{two}=105 m, y_{two}=35 m along the x-direction, with subsequent small-scale measurements during iterations 10 and 12 occurring immediately to the north and south of the well location. Small-scale measurement at the well itself did not occur because, after the first small-scale measurement the restriction against conditioning at an adjacent point prohibited measurement at the well. At iteration 40, the design

algorithm suddenly chose to measure **Y** at the small-scale location far to the east of the well because a number of the 10 plume realizations constraining the design bifurcated around the low-conductivity area between the source and the well. The occurrence of low conductivities associated with true **Y** realization in this area, coupled with flowlines becoming abruptly re-oriented from northwest-southeast to a more westerly-easterly direction and providing a means for the subsidiary plume to enter the vicinity of the well, made small-scale measurement in the western area important to the problem of quantifying how much contaminant had the potential to enter the monitoring well. In effect, the design attempted to investigate the impact of plume bifurcation on prediction via this sequence of inexpensive small-scale measurements.

Variance for the two-scale **Y** process was equal to 3.5 $(\ln m/s)^2$, somewhat larger than that typically encountered in field situations (Freeze, 1975; Bakr, 1976). Since a large variance can result in violation of the assumptions implicit to the first-order approximation given by equation [4.1], the accuracy of the approximation was investigated. As illustrated in the plot of log variance reduction versus iteration number shown in Figure 4.7, variance reduction estimated via equation [4.11] was consistently many orders of magnitude smaller than the ensemble variance reduction obtained from the 10 conditional monitoring well concentrations predicted during each iteration of the design, suggesting that the linear approximation given by equation [4.1] might not be valid. To prevent the occurrence of suboptimal design due to nonlinearities in the C-**Y** relation, **Y** variance was reduced during subsequent designs. Additional examples of measurement design, discussed in the sections to follow, are

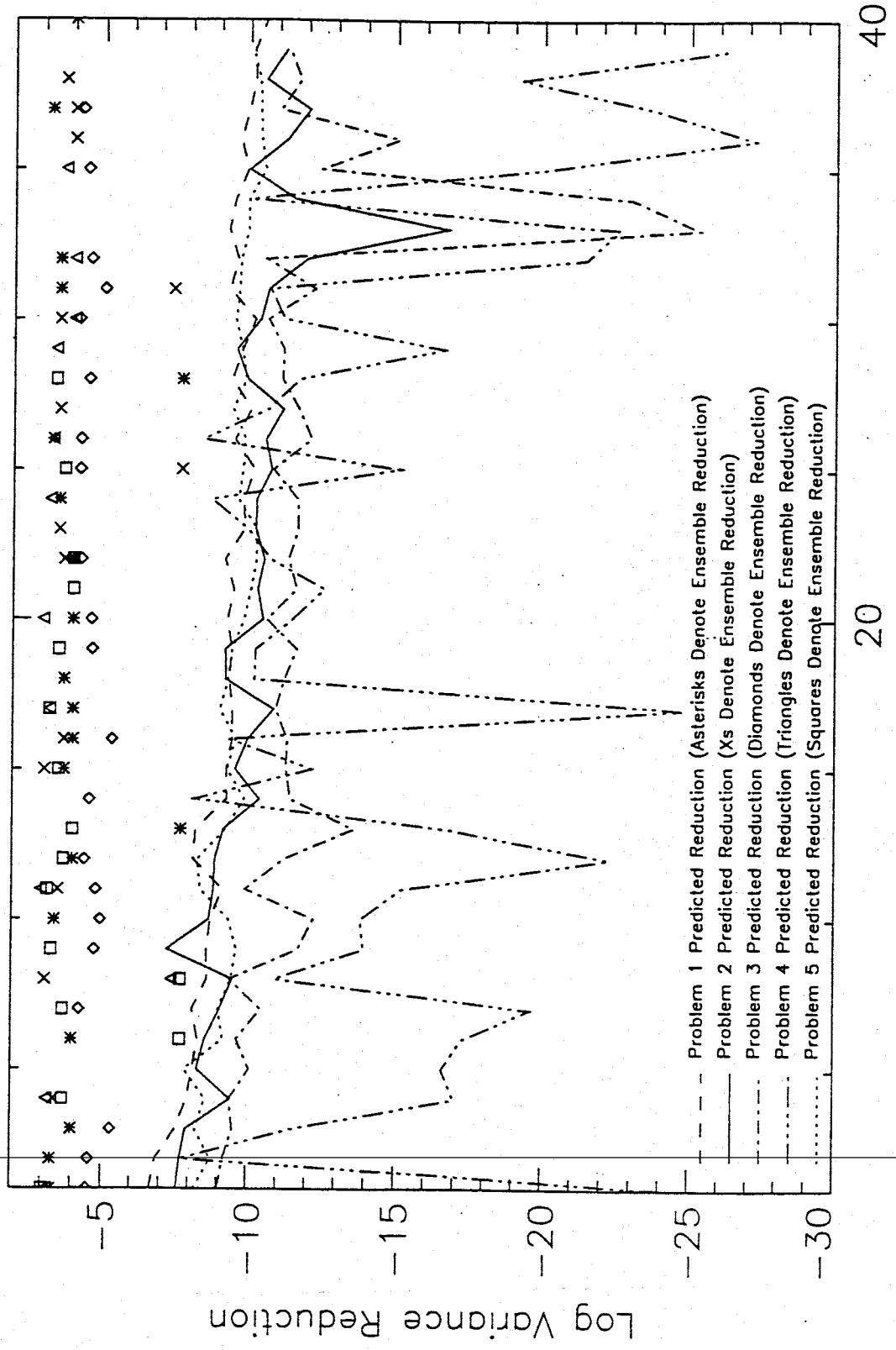


Figure 4.7. Ensemble and Predicted Variance Reduction Vs. Iteration Number.

summarized in Table 4.3. 1,024 particles were used to predict concentrations for all problems.

Table 4.3. Specifications for Conductivity Measurement Network Design Problems.

Design Problem Number	Source Size (m)	Source Type (I=instantaneous) (C=continuous)	Distance(m) Between SE Corner of Source and Well	True Large-Scale **Y** Variance $(\ln m/s)^2$	Nominal Small-Scale **Y** Variance $(\ln m/s)^2$	Two-Scale **Y** Variance $(\ln m/s)^2$
1	30-by-30	I	85	1.9	2.0	3.5
2	30-by-30	I	85	0.5	1.0	1.2
3	30-by-30	C	85	0.5	1.0	1.2
4	10-by-10	I	85	0.5	1.0	1.2
5	30-by-30	I	43	0.5	1.0	1.2

For the case of design problem 1, the moderately large variance of 3.5 $(\ln m/s)^2$ may have contributed to a small predicted concentration variance reduction compared to the true ensemble variance reduction estimated from the 10 conditional realizations, potentially rendering the design invalid. In design problem 2, problem 1 specifications were retained but nominal small-scale variance was reduced to 1.0 $(\ln m/s)^2$ when generating the true realizations. To further reduce total **Y** variance and lend greater validity to linear approximation [4.1], large-scale MRF state levels were changed from $-7.0, -6.0, -5.0, -4.0, -3.0$ ln m/s to $-6.0, -5.5, -5.0, -4.5$, and -4.0 ln m/s. This

reduced total two-scale **Y** variance from 3.5 to 1.2 (ln m/s)^2, of which 0.5 (ln m/s)^2 or 40% of the total variance resided at the large scale. The change in state levels slightly altered the structure of the large-scale MRF realization, producing somewhat less organization due to reduction in strength of neighborhood pair potential energies.

4.7.2 Design Problem 2: Reduced Y Variance

As a consequence of decreased **Y** variance, each contaminant particle had fewer options for getting from the source to the well, and the true plume did not reach the monitoring well at either scale of simulation after 230 days. In practice, a well would not be considered a candidate for monitoring ground water remediation effectiveness unless *some* contaminant were observed in it at a certain point in time. To make the design meaningful, the duration of plume displacement was increased to 345 days, and the monitoring well location was reassigned to x_{two}=95 m, y_{two}=35 m in MRF block (9,4). Note that the true plume simulated at the large scale of resolution did not quite reach the monitoring well. All transport and flow parameters other than well location were the same as those specified in design problem 1.

True log conductivity realizations and contaminant plumes for the reduced variance case are shown in Figure 4.8 and Figure 4.9. Figure 4.10 and Table 4.4 present the optimal design for problem 2. Despite the reduction in **Y** variance, Figure 4.7 shows evidence that a fairly large discrepancy between the linearized prediction of variance reduction and the true ensemble variance reduction persisted, and even worsened, suggesting that high **Y** variance was not the cause of inaccurate prediction

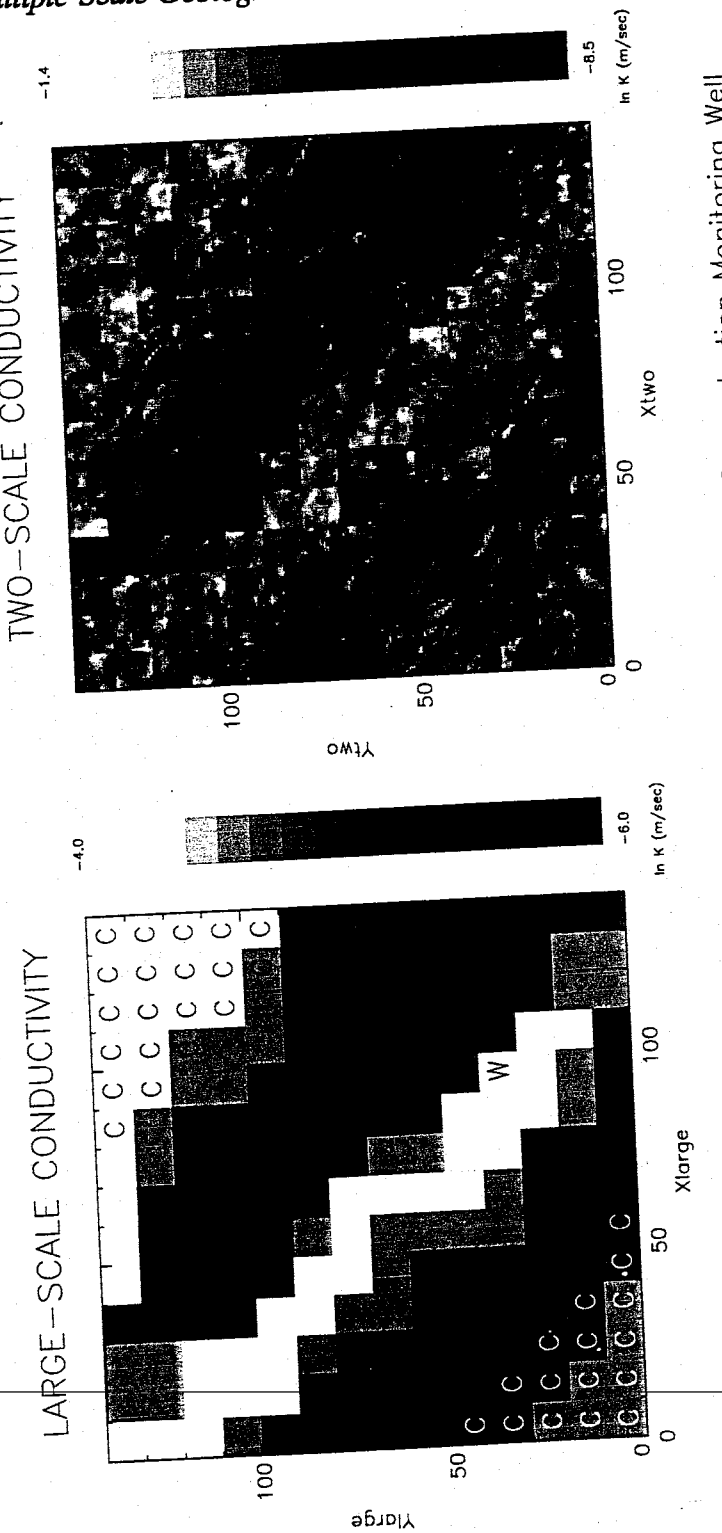


Figure 4.8. True Conductivity Realizations at Two Scales for Reduced Variance Designs.

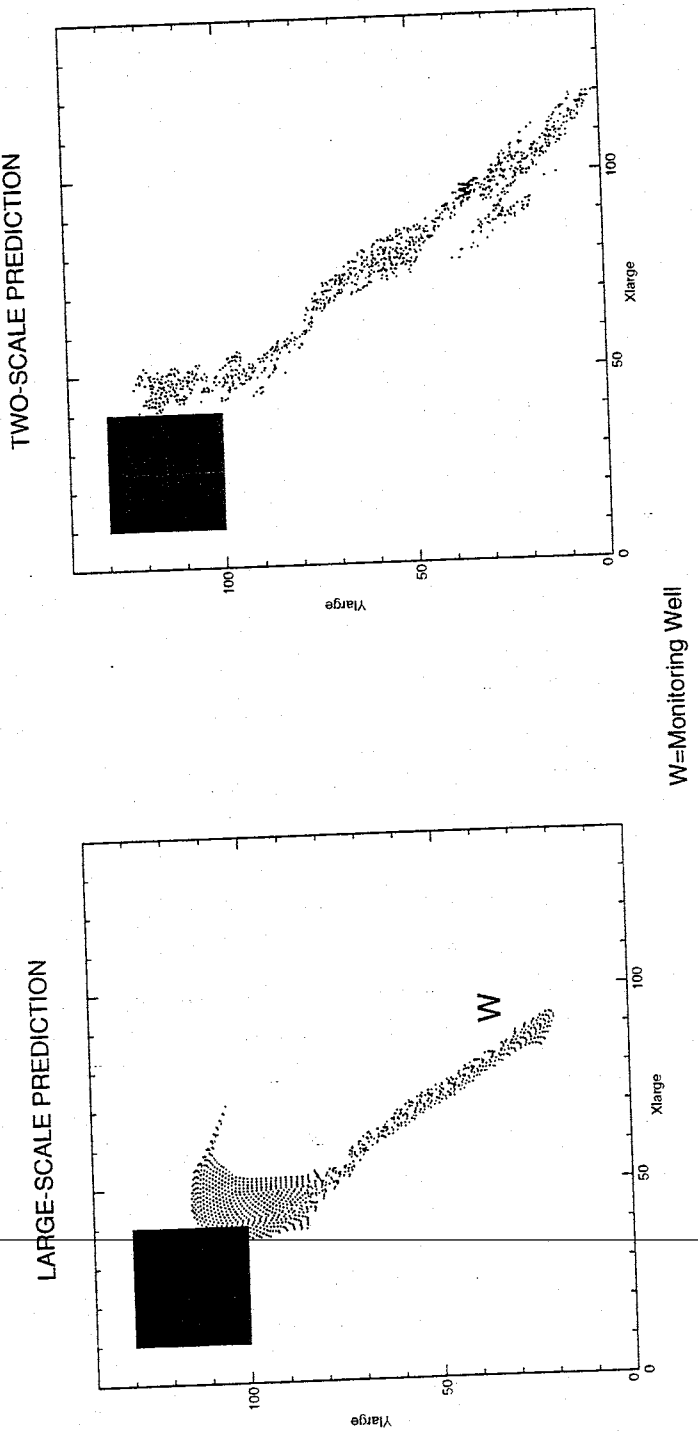


Figure 4.9. True Plume Realizations at Two Scales of Simulation for Design Problem 2.

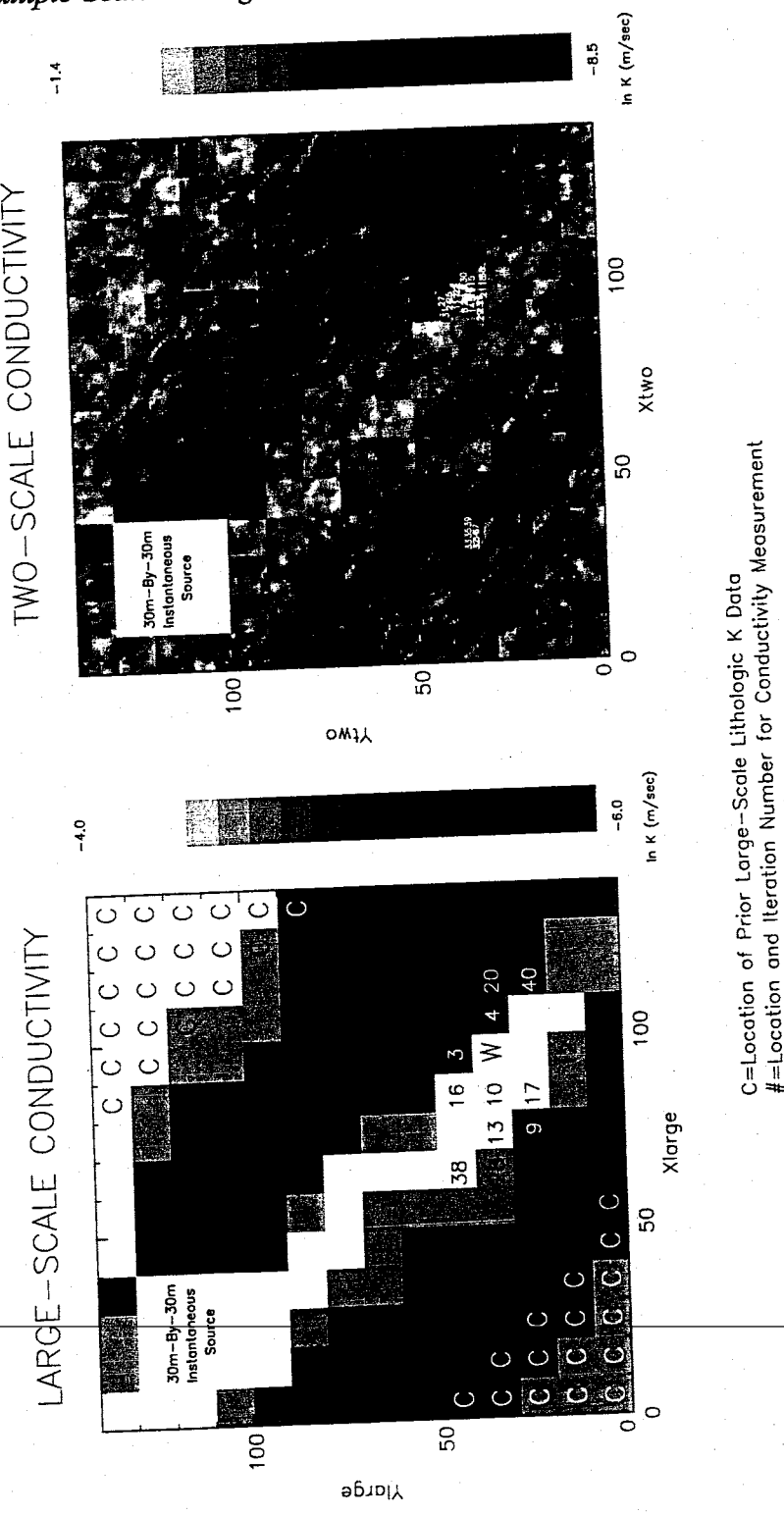


Figure 4.10. Conductivity Measurement Design Results, Problem 2.

Table 4.4. Results of Conductivity Measurement Design for Problem 2.

Sequential Design Iteration	Measurement Scale	Measurement Location Index (i_s, j_s)
1	small	96,35
2	small	94,35
3	large	10,5
4	large	11,4
5	small	95,33
6	small	98,35
7	small	96,37
8	small	92,35
9	large	8,3
10	large	9,4
11	small	94,37
12	large	10,2
13	large	8,4
14	small	93,33
15	small	100,33
16	large	9,5
17	large	9,3
18	small	100,30
19	small	97,33
20	large	12,4
21	small	97,30
22	small	98,37
23	small	93,31
24	small	94,39
25	small	96,39
26	small	95,31
27	small	95,41
28	large	7,4
29	small	91,31
30	small	101,35
31	small	92,40
32	small	31,34
33	small	31,36
34	small	33,34
35	small	34,36
36	small	102,30
37	small	35,34
38	large	7,5
39	small	37,36
40	large	12,3

of variance reduction. In addition, ensemble concentration variance reduction at the monitoring well did not decrease with iteration number as one would expect, despite the fact that the design made physical sense. In some cases, ensemble variance reduction was either 0.0 or negative (denoting a variance increase), and were not plotted. These observations suggest that the linearized prediction of variance reduction, which is a spatial measure of variance reduction based on concentration sensitivities $\partial C/\partial \mathbf{Y}$, may not be a good predictor of ensemble variance reduction. Errors in estimating the concentration distribution when a finite number of particles are used to perform tracking simulations may be introducing an additional source of uncertainty into the design. These estimation errors become amplified when the concentration estimates are used to determine sensitivities via the adjoint algorithm, because sensitivities involve calculation of concentration derivatives.

Nonergodicity of the concentration process may also be a factor influencing the accuracy of the linearized variance reduction prediction. Lack of ergodicity can occur when any predicted concentration realization does not contain all possible concentration states of the ensemble, and in the context of transport is more the rule than the exception. The problem of nonergodic behavior with respect to transport has been discussed in some detail by Dagan (1986; 1990). Unfortunately, the problem of nonergodicity in the concentration process is rooted in the very nature of transport and is related to the degree to which the transport system disperses mass, as discussed earlier. To understand how the problem of nonergodicity is especially relevant to transport prediction, consider that the true plume and 10 plumes generated during each

iteration of the design will depend on the unique spatial arrangement of ln K encountered as each particle migrates through the realization. This situation differs substantially from the case of fluid flow, in which the hydraulic head process is not determined by local variations in ln K and is less inclined to show nonergodic behavior.

A numerical experiment was conducted to determine if the reduced-variance design based on a spatial measure of variance change was robust and not strongly affected either by uncertainty in estimated concentration or nonergodic influences. The question of robustness was assessed for the case of a random design implemented by sampling measurement locations and scales from two uniform pdf's, bypassing the adjoint sensitivity calculations during each iteration. To maintain an equivalent total cost of measurement, the random design was forced to collect the same number of small- and large-scale measurements.

Figure 4.11 shows the results of the random measurement design. Ensemble variance reduction vs. iteration number for the random design is shown in Figure 4.12 along with systematic ensemble variance reduction for design problem 2. While overall concentration variance reduction during the course of the design for problem 2 was equal to 0.0002 $(mg/l)^2$, variance increased by 0.0002 $(mg/l)^2$ for the random design. These results suggest that the reduced-variance design algorithm does maximize ensemble variance reduction compared to random measurement.

Physical transport specifications for design problems 1 and 2 were equivalent, differing only in the duration of plume displacement. Therefore, differences between the design obtained for problem 1 and the reduced-variance design in problem 2 were

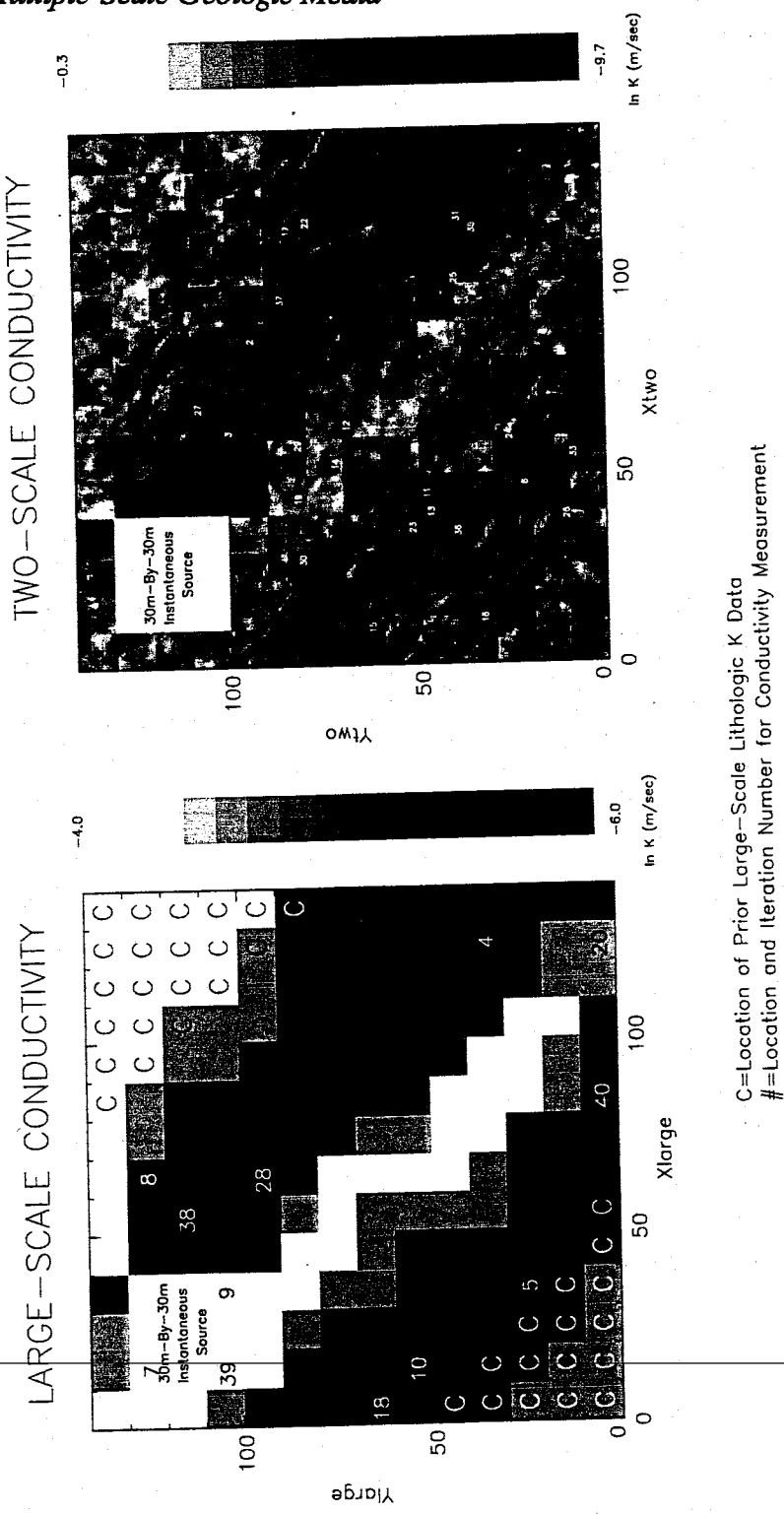


Figure 4.11. Random Conductivity Measurement Design Results.

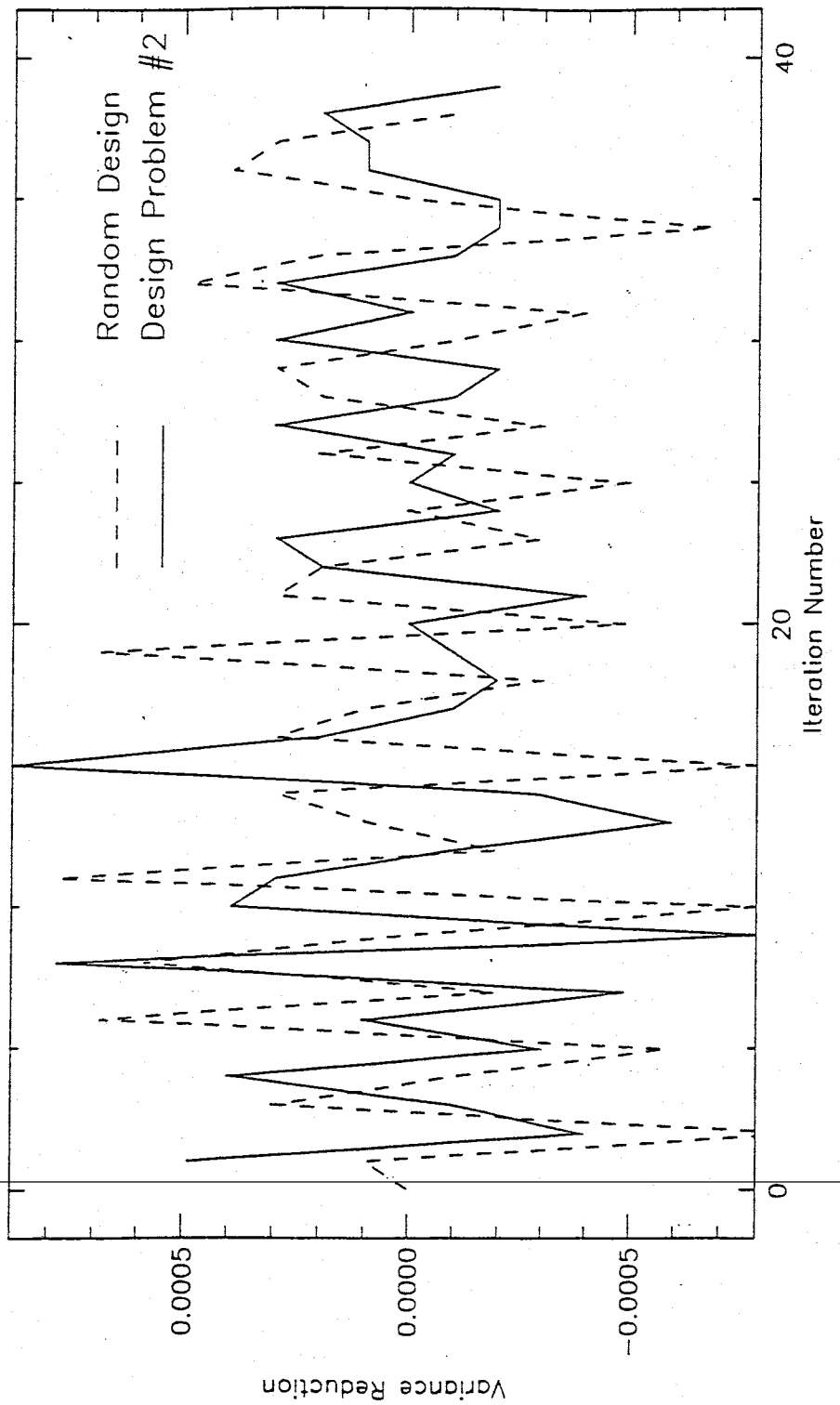


Figure 4.12. Ensemble Variance Reduction vs. Iteration Number for Systematic and Random Designs.

assumed solely related to differences in **Y** variance. Slight changes in degree of MRF correlation caused by changing the MRF state vector from [−7.0, −6.0, −5.0, −4.0, −3.0] to [−6.0, −5.5, −5.0, −4.5, −4.0] ln m/s were assumed to contribute insignificantly to changes in design. The diminished preference for large-scale measurement in design problem 2, for example, was believed related only to the smaller proportion of variance now residing in the large-scale component of **Y** [0.5 (ln m/s)^2, or 42%, instead of 1.9 (ln m/s)^2, or 54%, in design 1]. The change in proportion of measurements at any given scale changed linearly with change in percent variance at that scale. For design 1, 43% of measurements occurred at the large scale, where 54% of the Y variance resided. When large-scale percent of total variance was reduced to 42%, the proportion of large-scale measurements equalled 29%, or roughly 0.43X(0.42/0.54)=33%. These results suggest that squared sensitivity remains approximately constant, despite the change in **Y** variance. It appears that sensitivities are not strongly dependent on **Y** variance for moderate changes in σ_Y^2 when the geometry of the transport problem remains unchanged.

Like the design for problem 1, small-scale measurement immediately east and west of the well was initially preferred by design 2, with two large-scale measurements following along the principal axis of large-scale anisotropy. The design again worked its way out perpendicular to the principal axis of large-scale anisotropy. The pattern of large-scale measurement for design 2 was less dispersed than in design 1 because the portion of the plume transported between source and well was somewhat thinner. Like the case for design 1, small-scale measurement remained confined to the well block or

just outside of it. As mentioned earlier, larger **Y** variance associated with problem 1 produced true plumes at the two different scales of simulation that looked much more similar than the two plumes generated for problem 2. As a result, design 2 seemed to be characterized by a somewhat greater tendency for scale preference in measurement than design 1.

Given that the hydrogeologist will not be able to control the degree of **Y** variability in field settings, that the variance of 1.2 (ln m/s)^2 associated with problem 2 can be considered fairly representative of the magnitude of variance typically encountered in the field, and that problem 2 appeared sufficiently robust with respect to inaccuracies in the concentration estimate and to possible nonergodicities to produce a physically meaningful design, design problem 2 was assumed to constitute a realistic base case design against which results of all subsequent designs were compared.

Patterns of measurement dictated by design 2, as in design 1, suggest that small-scale measurement should be confined to the well block, that small-scale well block measurement may obviate the need to measure at the large scale in the well block, and that measurement at both scales when principal directions of heterogeneity and transport coincide should occur first along the longitudinal direction and then along the transverse direction. For geometrically similar transport problems, sensitivities will remain roughly constant with moderate changes in **Y** variance and, as a result, the proportion of measurements required at a given scale will change linearly with increasing variance at that scale. Finally, inaccuracies in the linearized variance

reduction given by equation [4.11] are likely related to errors in the concentration estimate or to nonergodicities of the concentration process, are not caused by high variance in **Y**, and may be an unavoidable aspect of dealing with transport in a stochastic framework.

4.7.3 Problem 3: Continuous Source Loading

For problem 3, the design algorithm was conditioned against the reduced-variance true **Y** realizations shown in Figure 4.8, but this time for a temporally-continuous contaminant source with all flow and transport parameters equivalent to those specified in design 2. Figure 4.13 shows true plumes produced by continuous source loading at both scales of simulation. Modification of source type from instantaneous to continuous loading altered the geometry of the transport problem without changing the geostatistical properties of the aquifer or the transport parameters. The plume for the continuous loading was generated over 60 time steps by repeatedly introducing 1024 particles over regular time intervals, rather than by shifting the instantaneous 1024-particle plume at regular displacement intervals to duplicate the effects of continuous loading via convolution of the instantaneous solution. This required that a larger time step be used to keep the total number of particles manageable. As a result, the continuous-source plume was characterized by a slightly different downgradient extent than the instantaneous-source plume despite equivalent displacement time.

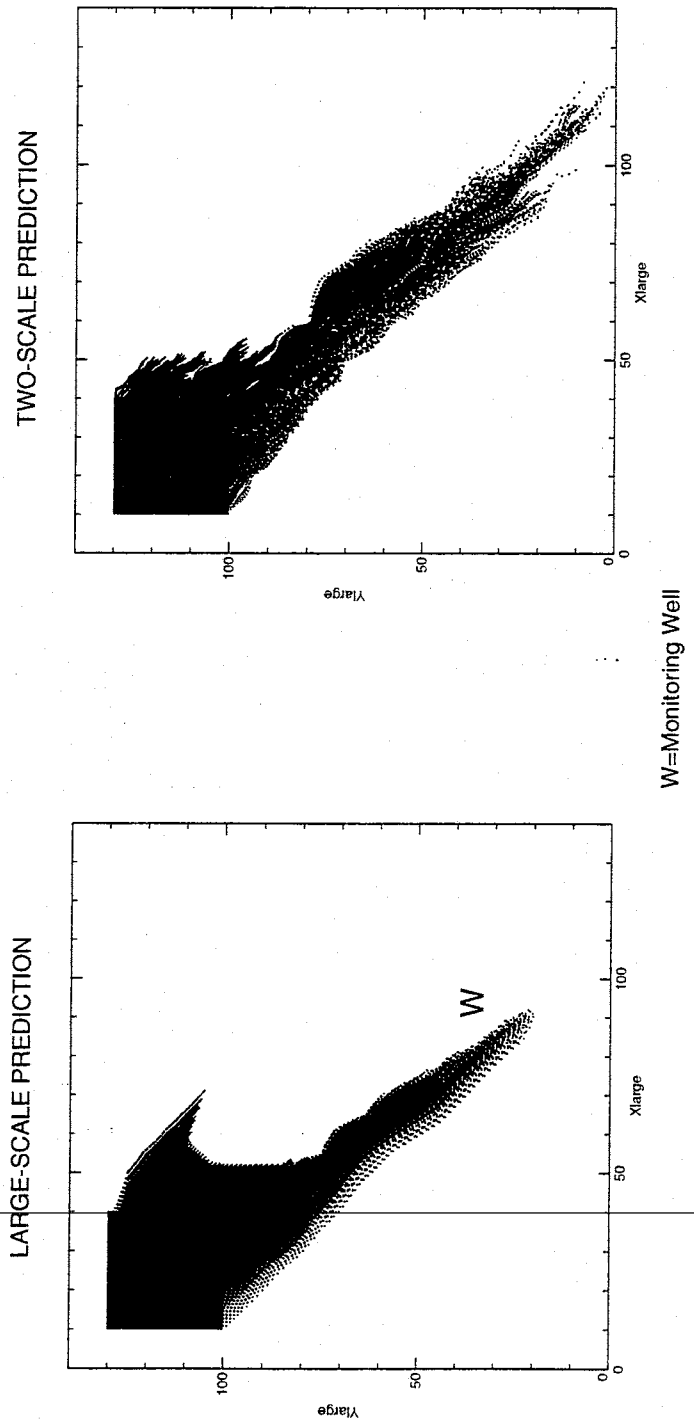


Figure 4.13. True Plume Realizations at Two Scales of Simulation for Design Problem 3.

To an even greater degree than the first, low-variance case given by design 1, the two scale-variant plumes predicted for the true **Y** realizations in the continuous-source loading case were characterized by similar shape and extent. This particular feature of the problem was related to the large extent of the plume compared to the size of both large- and small-scale heterogeneities. The larger plume sampled more large-scale heterogeneities than the smaller instantaneous plume used in design 2, and was considered more likely to reflect an ergodic concentration process. However, design 3 shows evidence that the predicted variance reduction deviated somewhat more from the simulated reduction than was observed in design 2, suggesting that errors in the concentration estimate may be affecting validity of equation [4.11].

Results of the design, presented in Figure 4.14 and Table 4.5, suggest that the design was less inclined to measure at the large scale than design 2. The more pronounced advective spreading of particles associated with the continuous-source case produced a more dispersed problem, reducing sensitivities. For an equal proportion of sensitivity reduction at the small and large scales, the squared sensitivity for the greater large-scale sensitivity will be reduced more than squared sensitivity at the smaller scale. As a result of this effect, an overall reduction in sensitivity proved to have a pronounced effect on reducing the number of large-scale measurements. Scale invariance of the plumes might also have contributed to the fewer number of large-scale measurements. For such a scale-invariant transport problem, the design would be expected to forfeit expensive large-scale measurement in favor of less expensive small-scale measurement.

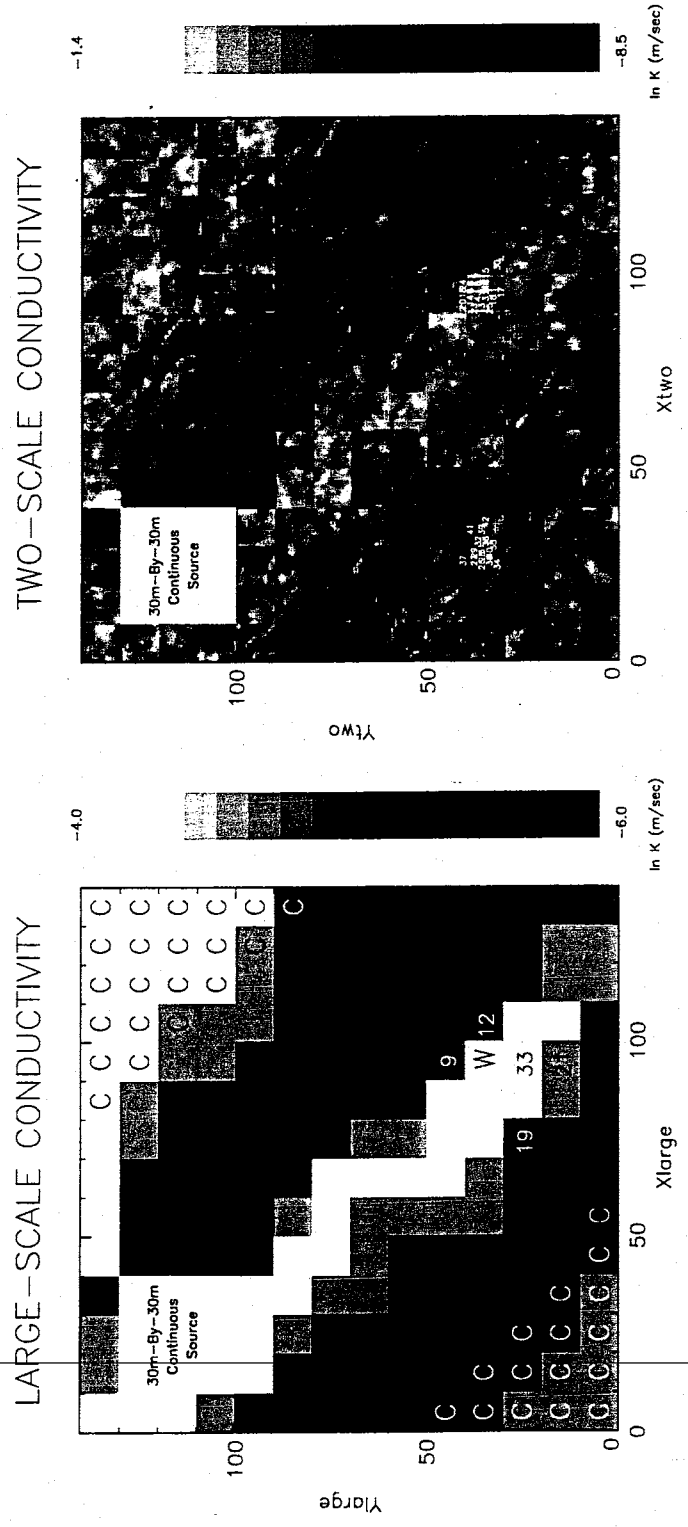


Figure 4.14. Conductivity Measurement Design Results, Problem 3.

Table 4.5. Results of Conductivity Measurement Design for Problem 3.

Sequential Design Iteration	Measurement Scale	Measurement Location Index (i_s, j_s)
1	small	95,36
2	small	95,34
3	small	93,34
4	small	93,36
5	small	97,36
6	small	95,38
7	small	99,36
8	small	97,38
9	large	10,5
10	small	97,34
11	small	95,32
12	large	11,4
13	small	91,35
14	small	93,38
15	small	100,34
16	small	93,32
17	small	95,30
18	small	95,40
19	large	8,3
20	small	92,40
21	large	10,2
22	small	90,40
23	small	91,37
24	small	97,40
25	small	91,33
26	small	25,35
27	small	26,37
28	small	27,35
29	small	28,37
30	small	102,31
31	small	29,35
32	small	31,36
33	large	10,3
34	small	25,31
35	small	30,32
36	small	26,33
37	small	26,40
38	small	31,34
39	small	34,35
40	small	28,33

4.7.4 Problem 4: Reduced Source Extent

Design 4 involved a reduction in instantaneous source size from 30m-by-30m to 10m-by-10m, with the downgradient southeast corner of the source area remaining at x_{two}=40m, y_{two}=100m. The true plumes are shown in Figure 4.15, with the optimal design presented in both Figure 4.16 and Table 4.6. Note that the decrease in source extent has produced a true large-scale plume that is no longer displaced into the well block.

Concentration of plume particles over a small source area and reduced plume spread resulted in preferentially greater squared sensitivity at the large scale and a less dispersive transport problem at the large scale. However, the reduced size of the source may have produced nonergodic behavior in transport, especially at the larger scale of simulation where the size of the source was equal to that of the MRF grid block. Figure 4.7, in which neither 0.0 or negative reductions and were plotted, illustrates that the spatial measure of variance reduction given by equation [4.11] was a poor estimate of ensemble uncertainty reduction and that the design was extremely erratic in predicting variance reduction.

The pattern of measurement was changed relative to that of design 2. Large-scale measurement for design 4 was displaced substantially downgradient, and was also extended downgradient and transverse to the plume centerline so that the large-scale measurement pattern actually covered a wider area than design 2, especially at later iterations, despite the smaller source size. Unlike every other design, design 4 exhibited no evidence of even a slight overall reduction in ensemble prediction

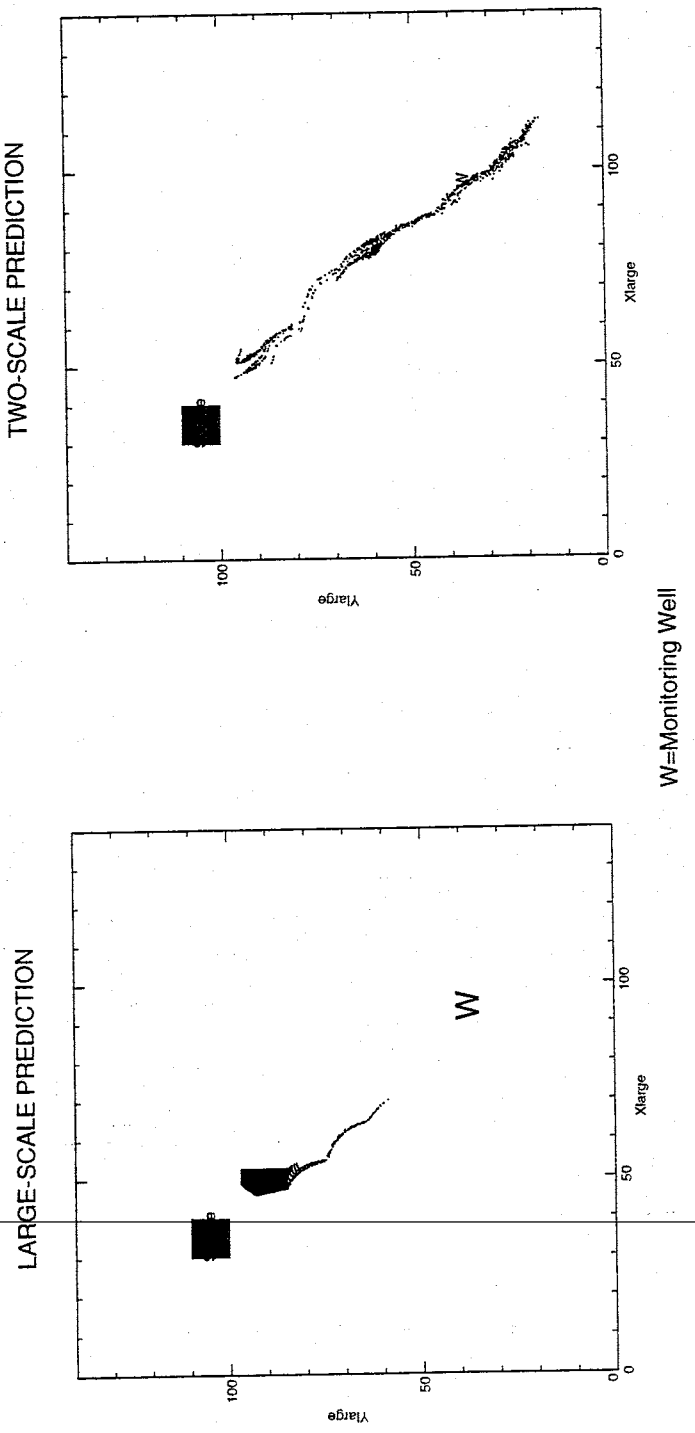


Figure 4.15. True Plume Realizations at Two Scales of Simulation for Design Problem 4.

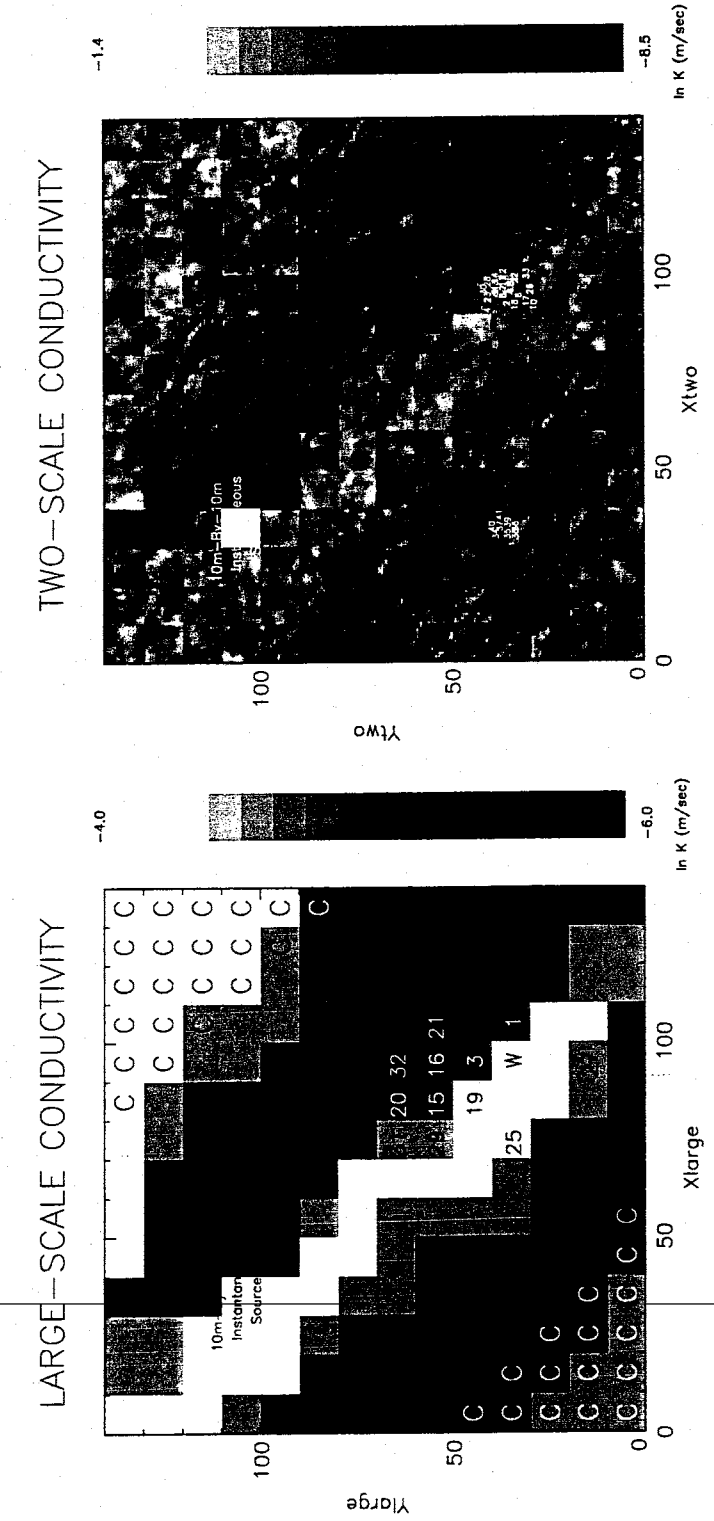


Figure 4.16. Conductivity Measurement Design Results, Problem 4.

Table 4.6. Results of Conductivity Measurement Design for Problem 4.

Sequential Design Iteration	Measurement Scale	Measurement Location Index (i_s, j_s)
1	large	11,4
2	small	92,35
3	large	10,5
4	small	95,34
5	small	97,34
6	small	94,32
7	small	96,36
8	small	94,36
9	small	95,38
10	small	92,28
11	small	98,36
12	small	100,36
13	small	97,38
14	small	99,38
15	large	9,6
16	large	10,6
17	small	93,30
18	small	92,33
19	large	9,5
20	large	9,7
21	large	11,6
22	small	99,33
23	small	94,40
24	small	90,40
25	large	8,4
26	small	96,29
27	large	10,2
28	small	98,40
29	large	8,6
30	small	96,41
31	large	8,7
32	large	10,7
33	small	100,30
34	small	33,38
35	small	33,35
36	small	35,33
37	small	35,37
38	small	33,33
39	small	36,35
40	small	35,39

uncertainty over the course of the design, and appeared to be wasting limited financial resources on suboptimal measurement.

Interestingly, large-scale measurement remained the same as in design 2, despite the disproportionate increase in squared sensitivity. A possible explanation for this might be that a nonergodic transport problem occurs at the large scale, where a small source can produce very different plumes for the 10 different Y realizations used to constrain the design. In the presence of such nonergodic influences, the design mistakes the large ensemble uncertainty in concentration prediction for a large spatial variance at the large scale, requiring more information at the large scale. Consequently, in an effort to better delineate a large-scale plume characterized by erratic shape and displacement distance, large-scale measurement is preferred.

4.7.5 Problem 5: Reduced Distance to Well

The fifth design problem involved moving the monitoring well to location x=65m, y=65m in order to determine the influence of near-source effects on the design. This new placement moved the monitoring well 42 m closer to the source than in design 2. Figure 4.17 shows well placement relative to the plume and source area, with results of design 5 presented in Figure 4.18 and Table 4.7.

By far, the greatest total number of large-scale measurements occurred in this design. Relative to design 2, which required intensive large-scale measurement during early and intermediate stages of the design, design 5 required large-scale conductivity information much more uniformly over much of the course of the design and seemed

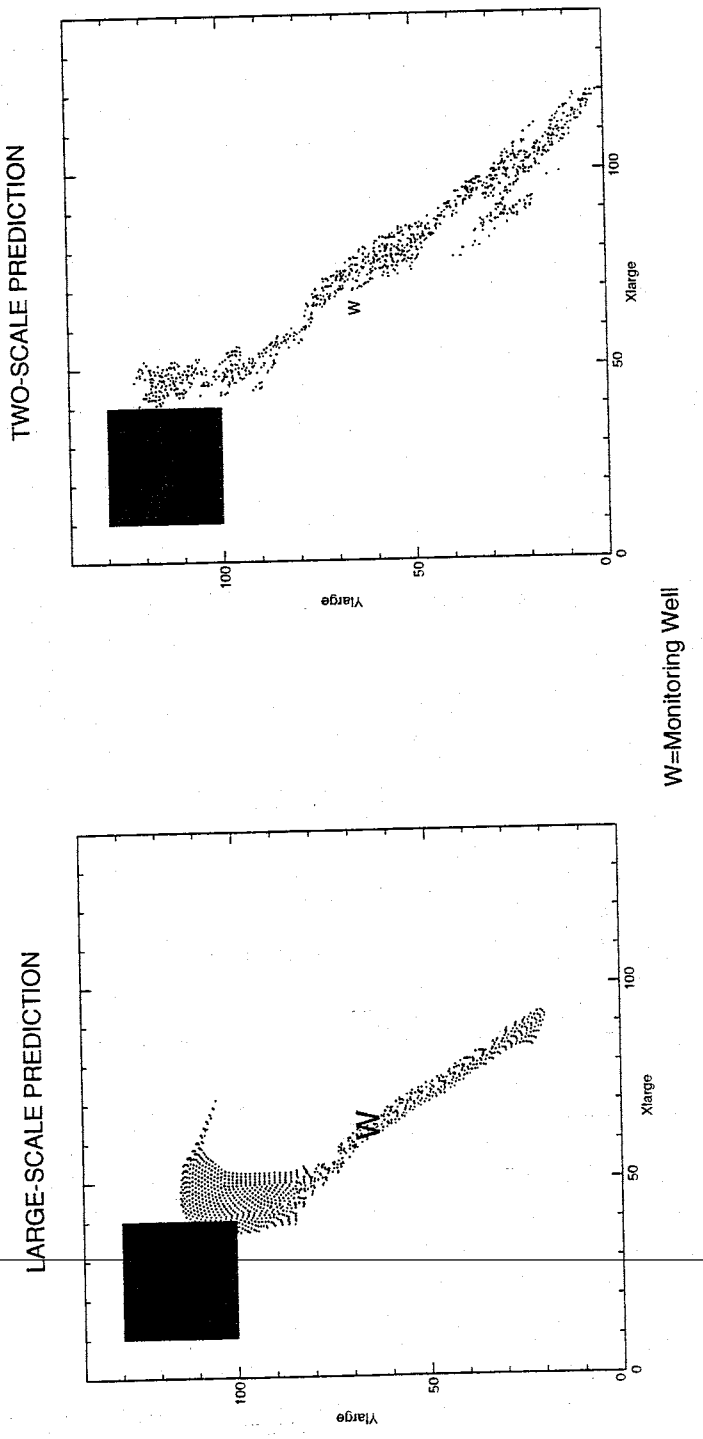


Figure 4.17. True Plume Realizations at Two Scales of Simulation for Design Problem 5.

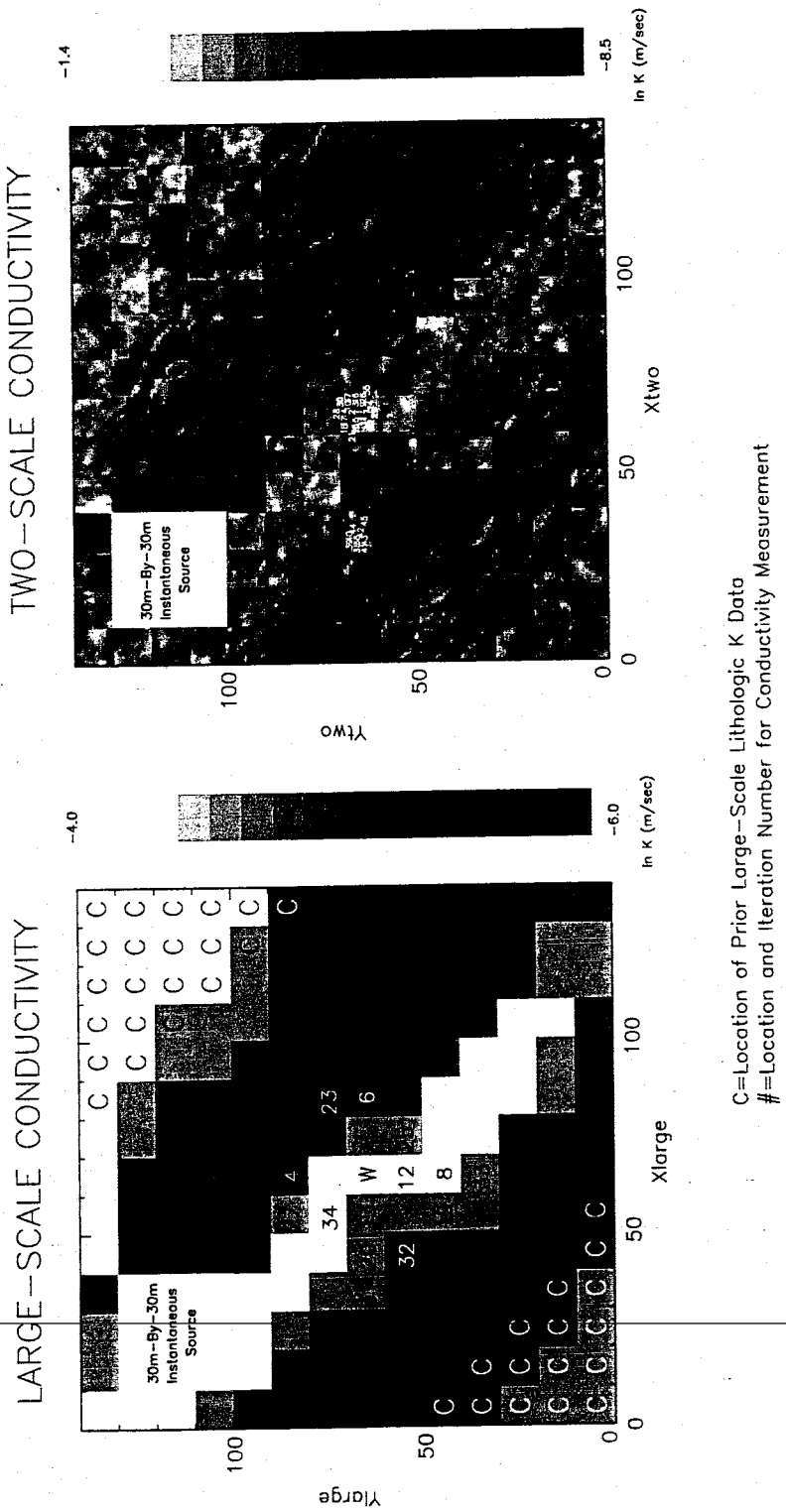


Figure 4.18. Conductivity Measurement Design Results, Problem 5.

Table 4.7. Results of Conductivity Measurement Design for Problem 5.

Sequential Design Iteration	Measurement Scale	Measurement Location Index (i_s, j_s)
1	small	65,64
2	small	65,66
3	small	67,65
4	large	7,9
5	small	63,65
6	large	9,7
7	small	65,62
8	large	7,5
9	small	67,63
10	small	67,67
11	small	63,63
12	large	7,6
13	large	6,7
14	small	65,68
15	small	61,65
16	small	69,65
17	small	63,68
18	small	61,68
19	large	6,5
20	large	5,7
21	small	59,66
22	small	62,61
23	large	9,8
24	small	67,61
25	large	8,6
26	small	69,63
27	large	6,9
28	small	65,70
29	small	65,60
30	small	68,69
31	large	4,7
32	large	5,6
33	large	7,4
34	large	6,8
35	small	71,62
36	small	30,65
37	small	69,67
38	small	32,65
39	small	31,67
40	small	33,67

to require a great deal of large-scale information during late stage iterations. This tendency to back load large-scale measurements may be an artifact of the wider extent of the plume close to the source. Once a sufficient number of large-scale measurements succeeded in determining how the northern leg of the true bifurcated plume influenced concentration in the monitoring well, it became most informative to collect measurements in the vicinity of the southern leg of the plume.

It should be pointed out that the variance reduction plots in Figure 4.7 show a consistent reduction in weighted, squared concentration sensitivity with iteration number for all designs, demonstrating that incremental improvement of a measurement network reduces with the addition of each measurement (Strecker *et al.*, 1985).

4.7.6 Rules of Thumb for Guiding Measurement of Conductivity in Multiple-Scale Geologic Media

Based on the five designs presented, a number of general conclusions can be made regarding measurement of hydraulic conductivity in geologic media characterized by more than one scale of spatial variation.

- Low-cost small-scale measurements should initially be collected near the well. Depending on how closely the small-scale measurements approximate the large-scale mean, they will tend to make large-scale measurement near the well redundant and unnecessary.
- For the case where the principal directions of anisotropy and transport coincide, measurements should be collected first along the principal directions and then along transverse directions. When the principal directions do not coincide, one might expect that measurements should be collected along directions

intermediate to the principal directions where information redundancy is minimized and plume mass maximized simultaneously.

- A change in proportion of **Y** variance residing at a given scale will change the fraction of optimal measurements at that scale as long as the change in **Y** variance is moderate and the same transport geometry is being investigated.
- A transport problem involving a continuous source will produce less scale-variant behavior than the equivalent instantaneous source problem. Since the design will tend to gain equivalent information at each scale, less expensive small-scale measurements will be favored.
- The greater the advective spreading of contaminant particles, the smaller the concentration sensitivities. Conversely, smaller spreading extent produces larger sensitivities. For an equal percent reduction or increase in large-scale and small-scale sensitivities, the absolute magnitude of large-scale squared sensitivity will be significantly reduced or increased disproportionately, and large-scale measurement will be more affected than small-scale measurement.
- Inaccuracies in estimates of concentration related to use of a particle tracker and nonergodic behavior of the transport process may compromise the validity of equation [4.11] and produce suboptimal design. Inaccuracies in the concentration estimate are of particular concern because concentrations are differenced in the adjoint equations to obtain sensitivities, resulting in magnification of this error.
- Large-scale measurement tends to be more favored overall when the well is closer to the source and the plume bifurcates upgradient of the well.

4.7.7 Steady-State Vs. Transient Design

It should be emphasized that the designs discussed earlier are based on a small design period during which a 'snapshot' of the plume at a certain time dictates measurement strategy. In many cases, characteristics of the plume will change

significantly during the time required to collect conductivity measurements, and a transient design should be performed. Since this involves incorporating a number of time steps into the adjoint problem, transient design can add substantially to the computational burden. For purposes of this study, it was assumed that conductivity data were acquired within a small time frame relative to the 345-day design period.

4.7.8 Conductivity Entropy Vs. Weighted, Squared Sensitivity

To demonstrate that the distribution of conductivity uncertainty can differ from the distribution of concentration uncertainty, conductivity entropy after 40 iterations of design problem 2 at the sth scale, H_s, was calculated according to the relation:

$$H_s = - \sum_{i_s = 1}^{M_s} p_{s,i_s} \ln p_{s,i_s} \qquad [4.26]$$

where p_{s,j_s} represents that ensemble probability of log conductivity at location i_s and M_s is the total number of measurement locations at the sth scale. Entropy represents conductivity information gained by measuring conductivity, while weighted squared concentration sensitivity is a measure of concentration information gained by measuring conductivity. Figure 4.19 presents initial and final ensemble large-scale and two-scale conductivity entropy and weighted squared sensitivity for problem 2. Ten intervals were used to tabulate frequency over the 10-realization ensemble. From inspection of the plots, it appears that the two measures of uncertainty are distributed

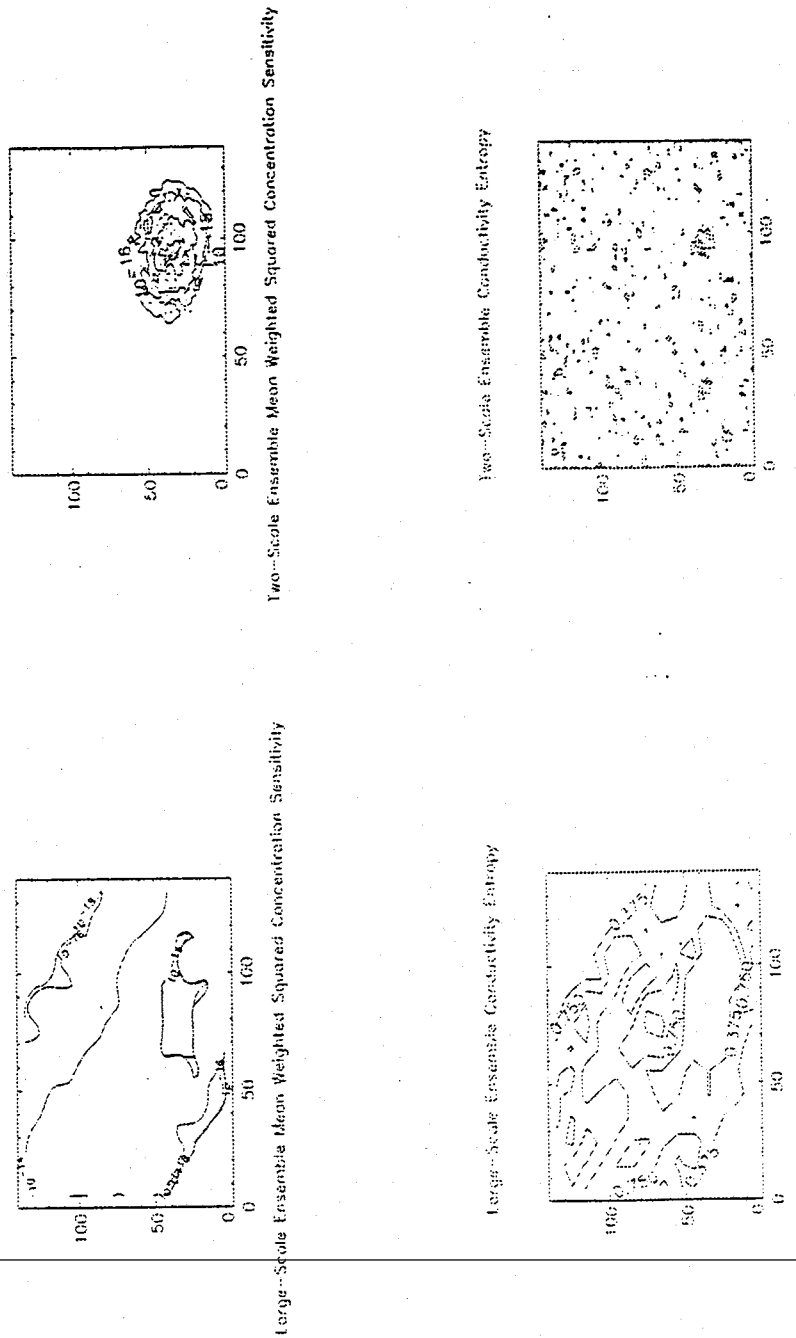


Figure 4.19. Final Distributions of Large- and Two-Scale Log K Entropy and Weighted, Squared Sensitivity.

somewhat differently, particularly at the large scale of variation, illustrating the importance of evaluating the worth of Y data through its ability to generate reliable well concentration predictions.

4.8 SUMMARY

A multiple-scale, multiple-realization design model was formulated to identify location and scale of hydraulic conductivity measurement that minimized concentration prediction uncertainty at a regulatory compliance point. First-order, second-moment analysis was used to relate log conductivity covariance to concentration prediction variance, with sensitivities determined via adjoint techniques used to incorporate physical meaning to the linear relationship. The multiple-scale design was conditioned sequentially on log conductivity measurements collected at two scales of spatial variation, and the validity of the first-order approximation assessed *a posteriori* to the design.

Five steady-state design scenarios were investigated. Preference for measurement at a given scale depended on how much log conductivity variance resided at that scale, source loading, size of source relative to scales of conductivity variation, distance to source, and principal directions of anisotropy relative to principal directions of transport.

4.9 REFERENCES

Ahlfeld, D.P., J.M. Mulvey, G.F. Pinder, and E.F. Wood, Contaminated groundwater remediation design using simulation, optimization, and sensitivity theory, 1. Model development, *Water Resour. Res.*, 24(3), 431-441, 1988.

Andricevic, R., Cost-effective network design for groundwater flow monitoring, *Stochastic Hydrol. Hydraul.*, 4(1), 27-41, 1990.

Andricevic, R., and E. Foufoula-Georgiou, A transfer function approach to sampling network design for groundwater contamination, *Water Resour. Res.*, 27(10), 2759-2769, 1991.

Bakr, A.A., *Stochastic analysis of the effects of spatial variations in hydraulic conductivity on groundwater flow*, Ph.D. dissertation, New Mexico Institute of Mining and Technology, Socorro, 1976.

Bardossy, A. and I. Bogardi, Network design for the spatial estimation of environmental variables, *Appl. Math. Comput.*, 12, 339-365, 1983.

Bogardi, I., A. Bardossy, and L. Duckstein, Multicriterion network design using geostatistics, *Water Resour. Res.*, 21(2), 199-208, 1985.

Carrera, J., and F. Szidarovzsky, Numerical comparison of network design algorithms for regionalized variables, *Appl. Math. Comput.*, 16, 189-202, 1985.

Carrera, J., E. Usunoff, and F. Szidarovszky, A method for optimal observation network design for groundwater management, *J. Hydrol.*, 73, 147-163, 1984.

Cleveland, T.G. and W.W.-G. Yeh, Sampling network design for transport parameter identification, *J. Water Resour. Plng. and Mgmt., ASCE*, 116(6), 764-783, 1990.

Dagan, G., Statistical theory of groundwater flow and transport: Pore to laboratory, laboratory to formation, and formation to regional scale, *Water Resour. Res.*, 22(9), 120S-134S, 1986.

Dagan, G., Transport in heterogeneous porous formations: Spatial moments, ergodicity, and effective dispersion, *Water Resour. Res.*, 26(6), 1281-1290, 1990.

Dagan, G., Butera, I., and E. Grella, Impact of concentration measurements upon estimation of flow and transport parameters: The Lagrangian approach, *Water Resour. Res.*, 32(2), 297-306, 1996.

Delhomme, J.P., Kriging in the hydrosciences, *Adv. Water Resour.*, 1(5), 251-266, 1978.

Dettinger, M.D., and J.L. Wilson, First-order analysis of uncertainty in numerical models of groundwater flow, 1. Mathematical development, *Water Resour. Res.*, 17, 149-161, 1981.

Freeze, R.A., A stochastic-conceptual analysis of one-dimensional groundwater flow in nonuniform homogeneous media, *Water Resour. Res.*, 11(5), 725-741, 1975.

Gelhar, L.W., A.L. Gutjahr, and R.L. Naff, Stochastic analysis of macrodispersion in a stratified aquifer,*Water Resour. Res.*, 15(6), 1387-1397, 1979.

Graham, W., and D. McLaughlin, Stochastic analysis of nonstationary subsurface solute transport, 1. Unconditional moments, *Water Resour. Res.*, 25(2), 215-232, 1989a.

Graham, W., and D. McLaughlin, Stochastic analysis of nonstationary subsurface solute transport, 2. Conditional moments, *Water Resour. Res.*, 25(11), 2331-2355, 1989b.

Graham, W., and D. McLaughlin, A stochastic model of solute transport in groundwater: Application to the Borden, Ontario, tracer test, *Water Resour. Res.*, 27(6), 1345-1359, 1991.

Hoeksema, R.J., and P.K. Kitanidis, Comparison of gaussian conditional mean and kriging in the geostatistical solution to the inverse problem, *Water Resour. Res.*, 21(6), 825-836, 1985.

Hsu, N.S., and W.W.-G. Yeh, Optimum experimental design for parameter identification in groundwater hydrology, *Water Resour. Res.*, 25(5), 1025-1040, 1989.

Hudak, P. F., and H.A. Loaiciga, A location modeling approach for groundwater monitoring network augmentation, *Water Resour. Res.*, 28(3), 643-649, 1992.

Hughes, J.P., and D.P. Lettenmaier, Data requirements for kriging: Estimation and network design, *Water Resour. Res.*, 17(6), 1641-1650, 1981.

James, B.R., and S.M. Gorelick, When enough is enough: The worth of monitoring data in aquifer remediaton design,*Water Resour. Res.*, 30(12), 3499-3513, 1994.

Journel, A. and C. Huijbregts, *Mining Geostatistics*, Academic, San Diego, Calif., 1978.

Kitanidis, P.K., Statistical estimation of polynomial generalized covariance functions and hydrologic applications, *Water Resour. Res.*, 19(4), 909-921, 1983.

Knopman, D.S., and C.I. Voss, Behavior of sensitivities in the one-dimensional advection-dispersion equation: Implications for parameter estimation and optimal design, *Water Resour. Res.*, 23(2), 253-272, 1987.

Knopman, D.S., and C.I. Voss, Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport, *Water Resour. Res.*, 25(10), 2245-2258, 1989.

Knopman, D.S., C.I. Voss, and S.P. Garabedian, Sampling design for groundwater solute transport: Tests of methods and analysis of Cape Cod tracer test data, *Water Resour. Res.*, 27(5), 925-949, 1991.

Lee, S.-I., and P.K. Kitanidis, Optimal estimation and scheduling in aquifer remediations with incomplete information, *Water Resour. Res.*, 27(9), 2203-2217, 1991.

Loaiciga, H.A., An optimization approach for groundwater quality monitoring network design, *Water Resour. Res.*, 25(8), 1771-1782, 1989.

Loaiciga, H.A., R.J. Charbeneau, L.G. Everett, G.E. Fogg,, B.F. Hobbs, and S. Rouhani, Review of ground-water quality monitoring network design, *J. Hydraul. Eng.*, 118(1), 11-37, 1992.

Magnuson, S., and J.L. Wilson, Optimization of transmissivity measurement networks, *EOS Trans. Amer. Geophys. Union,* 68(44), p. 1266, 1987.

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

Massmann, J., and R.A. Freeze, Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1. Results, *Water Resour. Res.*, 23(2), 368-380, 1987b.

McKinney, D.C., and D.P. Loucks, Network design for predicting groundwater contamination, *Water Resour. Res.*, 28(1), 133-147, 1992.

McLaughlin, D.B., and W.D. Graham, Design of cost-effective programs for monitoring groundwater contaminantion, *IAHS Publ. 158*, 231-245, 1986.

Meyer, P.D., and E.D. Brill, A method for locating wells in a groundwater monitoring network under uncertainty, *Water Resour. Res.*, 24(8), 1277-1282, 1988.

Meyer, P.D., A.J. Valocchi, and J.W. Eheart, Monitoring network design to provide initial detection of groundwater contamination, *Water Resour. Res.,* 30(9), 2647-2659, 1994.

Naff, R.L., *A continuum approach to the study and determination of field longitudinal dispersion coefficients*, Ph.D. dissertation, New Mexico Institute of Mining and Technology, Socorro, 1978.

Nishikawa, T., and W.W.-G. Yeh, Optimal pumping test design for the parameter estimation of groundwater systems, *Water Resour. Res.*, 25(7), 1737-1747, 1989.

Philip, G.M., and D.F. Watson, Matheronian geostatistics, *Math. Geol.*, 18(1), 93-117, 1986.

Rajaram, H., and D.B. McLaughlin, Identification of large-scale spatial trends in hydrologic data, *Water Resour. Res.*, 26(10), 2411-2424, 1990.

Rouhani, S., *Optimal Data Collection in Random Fields*, Ph.D. dissertation, Div. of Applied Science, Harvard University, Cambridge, Mass., 1983.

Rouhani, S., and T.J. Hall, Geostatistical schemes for groundwater sampling, *J. Hydrol.*, 103, 85-102, 1988.

Saad, Y., and M.H. Schultz, GMRES: A generalized minimum residual algorithm for solving non-symmetric linear systems, *SIAM Jour. Sci. Stat. Comput.*, 7(3), 856-869, 1986.

Schafer-Perini, A.L., and J.L. Wilson, Efficient and accurate front tracking for two-dimensional groundwater flow models, *Water Resour. Res.*, 27(7), 1471-1485, 1991.

Skaggs, T.H., and Z.J. Kabala, Recovering the release history of a ground water contaminant, *Water Resour. Res.*, 30(1), 71-79, 1994.

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

Srivastava, R.M., Philip and Watson — Quo Vadunt?, *Math. Geol.*, 18(1), 141-146, 1986.

Steinberg, D.M., and W.G. Hunter, Experimental design: Review and comment, *Technometrics*, 26(2), 71-97, 1984.

Strecker, E.W., and W.-S. Chu, Parameter identification of groundwater contaminant transport model, *Groundwater*, 24(1), 56−62, 1986.

Sun, N.-Z., and W. W.-G. Yeh, Coupled inverse problems in groundwater modeling, 2., Identifiability and experimental design, *Water Resour. Res.*, 26(10), 2527-2540, 1990.

Sykes, J.F., J.L. Wilson, and R.W. Andrews, Sensitivity analysis for steady-state ground water flow using adjoint operators, *Water Resour. Res.*, 21(3), 359-371, 1985.

Tucciarelli, T., and G. Pinder, Optimal data acquisition strategy for the development of a transport model for groundwater remediation, *Water Resour. Res.*, 27(4), 577-588, 1991.

Wagner, B.J., and S.M. Gorelick, Reliable aquifer remediation in the presence of spatially variable hydraulic conductivity: From data to design, *Water Resour. Res.*, 25(10), 2211-2225, 1989.

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

Chapter 5.
Geostatistical Validation of a Multiple-Scale Aquifer Characterization Model

Chapter 4 dealt with the issue of quantifying the relative worth of data collected at different scales. In this chapter, the question of whether a two-scale approach to aquifer characterization is better able to reproduce the spatial moments of observed state variables than an equivalent single-scale model is addressed. In the very loosest sense of the term, this process of reproducing the spatial moments of a single realization of observed state variables will be referred to as geostatistical validation. Such validation should be performed by matching moments of the observed and predicted ensemble mean state variables, but the observed ensemble mean is not accessible when dealing with geologic variables. Instead, we have access only to a single realization of the observed state variable process.

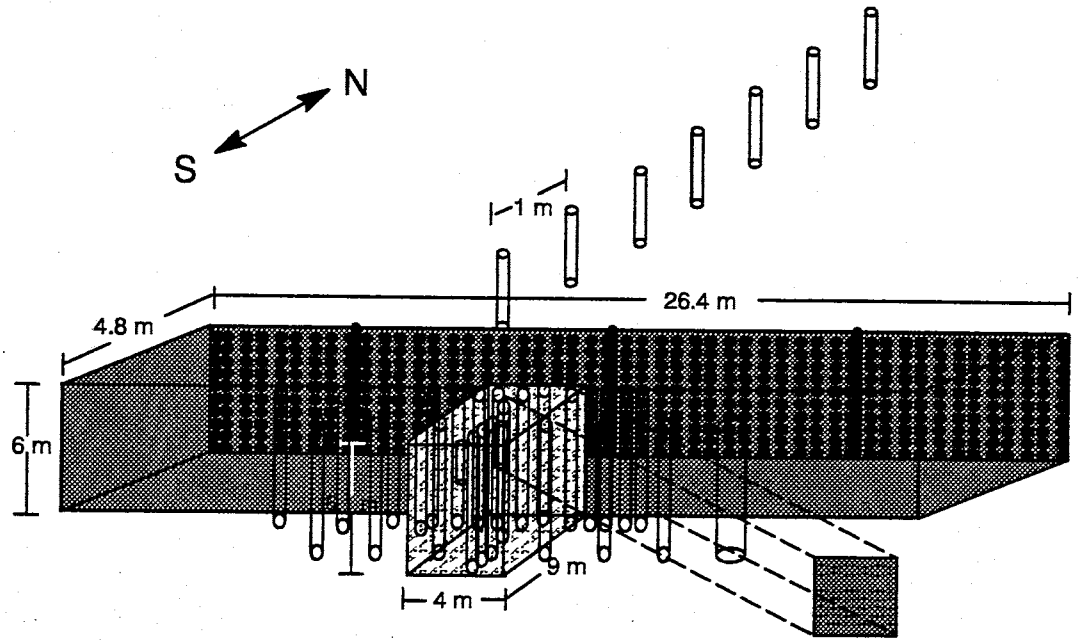
The validation focused on duplicating spatial moments of water content and tracer concentration distributions observed during a tracer test conducted at the Las Cruces site in southern New Mexico. Designed as a collaborated effort among New Mexico State University, University of Arizona, Massachusetts Institute of Technology, and Battelle Pacific Northwest Laboratories, the Las Cruces trench experiment represents one of the first attempts to stochastically characterize natural spatial variations in a subsurface medium. The stochastic study involved measurement of the spatial distribution of permeability, as well as other distributed properties of the soils,

in a trench excavation. Following subsurface characterization along the walls of the trench, the rate of fluid and tracer movement was observed in desiccated soils underlying an experimental plot adjacent to the trench during the course of two infiltration/tracer experiments. Objectives of the Las Cruces study were to build a stochastic model of flow and transport in the partially saturated soil underlying the site and compare numerical model predictions to observed water content and tracer distributions.

5.1 THE LAS CRUCES TRENCH INFILTRATION EXPERIMENTS

5.1.1 Characterization of Subsurface Saturated Permeability Structure

The Las Cruces Trench site is located on a basin slope of Mount Summerford in the Dona Ana Mountains, roughly 40 km northeast of Las Cruces, New Mexico, in a semi-arid environment. Mid-piedmont young soils of the Dona Ana and Onite series and lower-piedmont older soils dominated by the Berino series are predominant (Gee and Nicholson, 1987). A trench 26.4 m long by 4.8 m wide by 6 m deep was excavated under the site and used as a basis for characterizing subsurface heterogeneities (see Figure 5.1). Using observations of soil horizon morphology and measured hydraulic properties along trench walls, nine distinct soil layers were identified beneath the site. Each layer is comprised of morphological horizons possessing roughly equivalent hydraulic properties, as determined from soil samples extracted from trench walls. These soils include sands, sandy loams, loamy sands, and sandy clay loams (Wierenga



+ Horizontal Tensiometer or Soil Solution Sampler

Neutron Access Tube (1.5m and 6.1m)

Trench

Transect of Disturbed and Undisturbed Core Soil Samples and Guelph Permeameter Measurements (Coarsely Spaced)

Irrigation Area

Transect of Disturbed and Undisturbed Core Soil Samples (Finely Spaced)

Figure 5.1. Instrumentation for the First Las Cruces Trench Infiltration Experiment.

et al., 1991). While gravelly sandy loams comprising buried arroyos exhibited a lack of structure, other soils exhibited distinct, visually-observable structure.

The trench afforded access to the subsurface beneath test irrigation plots, and also permitted collection of soil samples for characterizing large-scale deterministic layered and small-scale random subsurface heterogeneities. 594 disturbed and 594 undisturbed core samples of soil, distributed at fixed intervals of 0.5 m in the horizontal direction within each of 9 distinct soil layers and at 0.13-m intervals along 3 vertical transects, were collected from the north wall of the trench and subjected to laboratory testing to measure bulk densities, saturated permeabilities, and soil moisture retention curve parameters. Water retention model parameters α and n were estimated from core samples via a least-squares analysis using the van Genuchten (1983) relation for effective saturation S_e:

$$S_e = \frac{\theta - \theta_r}{\theta_s - \theta_r} = \frac{1}{[1 + (\alpha\psi)^n]^m}\ ,\ m = 1 - \frac{1}{n} \qquad [5.1]$$

where θ, θ_r, θ_s, represent volumetric water content, residual water content, and saturated water content (or porosity) and ψ is equal to tension head. Table 5.1 lists estimates of these parameters for each soil layer. Note that parameters for only the top 7 layers are listed because the bottom 2 layers were not sampled by the tracer and water during the course of the experiment. Unsaturated permeability in the core

samples was defined as a function of water content and saturated permeability k_s using a simplified form of the Maulem model (van Genucthen, 1983):

$$k = k_s S_e^{1/2}[1 - (1 - S_e^{1/m})^m]^2 \qquad [5.2]$$

In addition to conductivities obtained via laboratory testing, *in situ* saturated conductivities were estimated in 30-cm diameter/15-cm deep holes located near

Table 5.1. Deterministic Parameters of TRACR3D Flow and Transport Simulations (after Wierenga *et al.*, 1991).

Layer	Porosity (θ_s)	θ_r	n	α (cm^{-1})
3	0.3104	0.0726	1.4177	0.02719
4	0.2942	0.0896	1.7117	0.07029
5	0.3021	0.0716	1.5496	0.04039
6	0.3129	0.0714	1.5373	0.06772
7	0.3359	0.0849	1.5742	0.05960
8	0.3434	0.0914	1.5278	0.06237
9	0.3483	0.0949	1.9026	0.04194

bulk matrix density = 1.54 for all layers

undisturbed core samples using a Guelph permeameter (Wierenga *et al.*, 1991). Although there was good qualitative agreement between lab- and field-based saturated conductivity measurements, the quantitative match was poor. In general, field-based saturated conductivity measurements exhibited a larger range than lab-based measurements, and maximum field conductivities were significantly larger than maximum lab conductivities (Wierenga *et al.*, 1991). This was at least partly attributed to the larger uncertainties inherent to field measurement.

Using a nonparametric Mann-Whitney test to determine whether mean saturated conductivities in each soil layer were indistinguishable from one another, it was found that the means of laboratory-measured conductivities in layer 1, layers 2 through 5, and layers 6 through 9 were not significantly different at the 10% significance level, while field-measured conductivities for layers 1 through 3, layers 4 through 8, and layer 9 were determined to be not statistically different at the same level of significance (Wierenga *et al*, 1991). However, the test was based on the assumption that mean saturated permeabilities associated with each layer are statistically independent of those measured in other layers, and does not test for similarity or dissimilarity of small-scale intralayer variance, which tended to vary considerably between layers.

Estimated statistical parameters for log conductivity in cm/day, including mean, variance, and horizontal and vertical correlation scales, are presented in Figure 5.2 for each of the 9 layers. When all *in situ* permeability data were grouped together, a variogram analysis of logarithmic saturated permeabilities indicated that the horizontal and vertical correlation lengths were equal to 2.5 m and 0.5 m (Jacobson,

1990). The most pronounced horizontal correlation occurred in layers 4 through 7. In fact, when the data were grouped into layers 1−3, 4−5, and 6−7, only layers 4-7 showed significant horizontal correlation, with a scale of 2.0 m. When all core and soil sample data collected along three vertical transects was used, the vertical correlation scale was 0.15 m, far less than the 0.5-m vertical scale estimated on the basis of all data (Wierenga *et al.*, 1991).

Only saturated permeability was treated as a random process, with all other soil properties assumed to be known deterministically. This assumption is justified by Warrick and Nielsen (1980) and Rao *et al.* (1983), who noted that water content and suction head exhibit much less spatial variability than saturated permeability. Note that, since saturated permeability k_s is treated as a random process, unsaturated permeability k also varied randomly through the functional dependence given by equation [5.2].

Due to the thinness of some layers with respect to the vertical correlation scales listed in Figure 5.2, nonergodic influences were likely present in the small-scale permeability process within many of the layers. That is, some layers were so thin that the statistical behavior of the small-scale processes was not fully expressed in the vertical direction, and estimates of structural parameters such as vertical correlation scale may not have been representative, or typical, of the layer.

Due to the well-defined horizontal layering that prevailed at the site, it was assumed that the large-scale saturated permeability process was a spatially-discrete deterministic structure. Geostatistical properties unique to each soil layer and listed

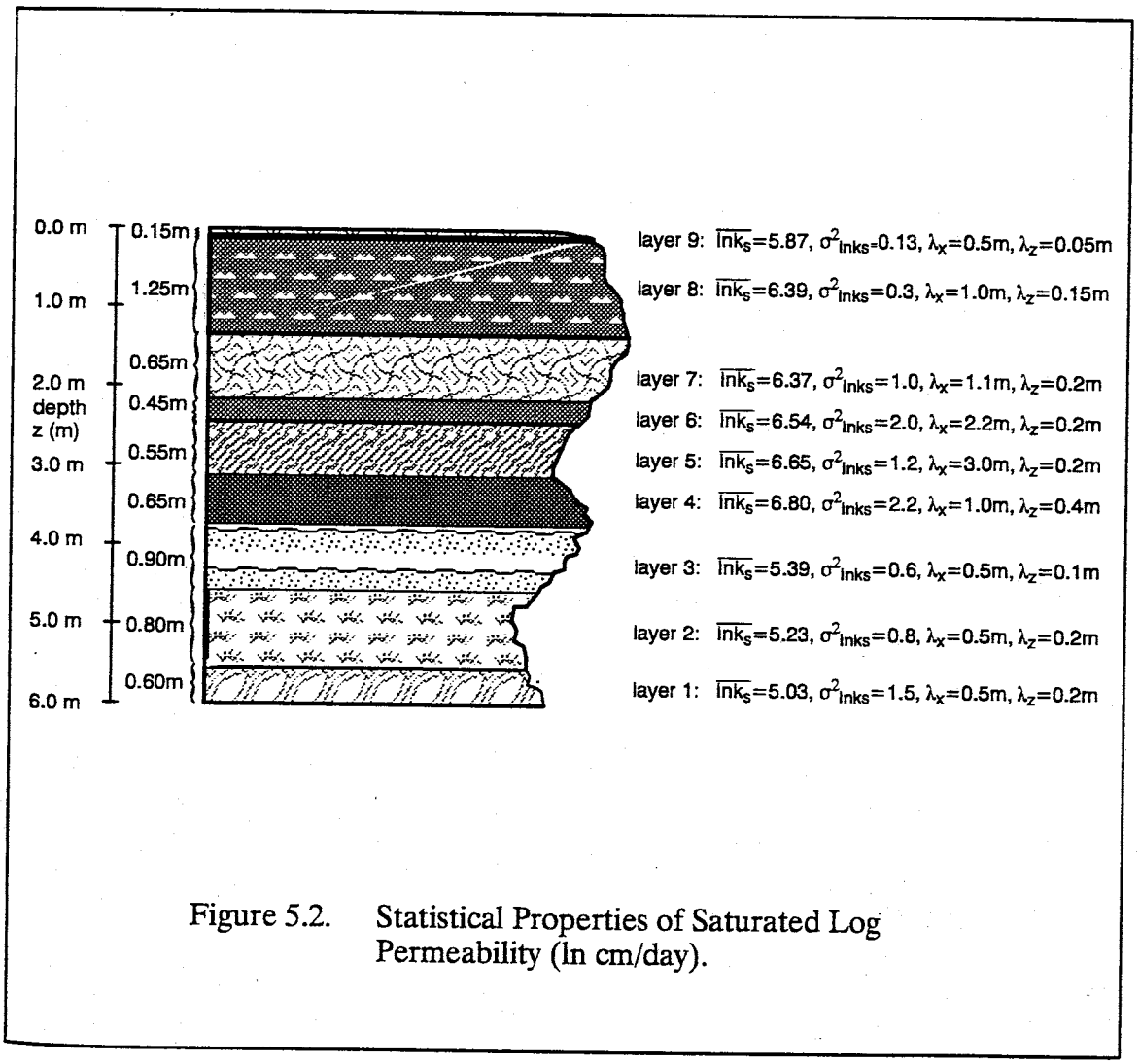


Figure 5.2. Statistical Properties of Saturated Log Permeability (ln cm/day).

in Figure 5.2 were then used to embed small-scale spatially continuous gaussian random structure with exponential covariance. Again, only the top 7 of the 9 layers were incorporated into the model because the tracers and infiltrating water didn't sample the bottom two layers during the course of the infiltration experiment.

Figure 5.3 shows an example of large-scale and nested, two-scale saturated permeability realizations generated on the basis of the deterministic structure and small-scale intralayer statistics. Since the large-scale structure was deterministic, it remained the same for all realizations. Figure 5.4 shows a statistically-equivalent single-scale saturated permeability realization.

Figure 5.5 presents an ensemble variogram estimate for log saturated permeability obtained using 20 realizations of the two-scale random saturated permeability process, including the realization presented in Figure 5.3. Since horizontal permeability is not a nested structure, only one sill was evident in the horizontal variogram. The vertical permeability variogram exhibited two distinct sills characteristic of a multiple-scale permeability process. The first of these sills was fairly transient and corresponded to a range of about 0.2-0.3 m, which represents the average vertical correlation scale for small-scale, intralayer permeability variation. The second sill, occurring at a range of roughly 0.65 m, corresponded to the average thickness of the deterministic layers. Also shown in Figure 5.5 are fitted exponential vertical and horizontal variogram estimates exhibiting single sills at different ranges corresponding to vertical and horizontal correlation scales of the equivalent single-scale gaussian process, equal to 0.18 m and 0.9 m, respectively. The variance of the fitted exponential variograms was equal to 1.2 (ln cm/day)^2. The fitted exponential variograms were used to generate 20 single-scale saturated permeability realizations.

It was believed that a poor match between observations and model predictions obtained at late times by Hills *et al.* (1991) using both uniform and layered soil

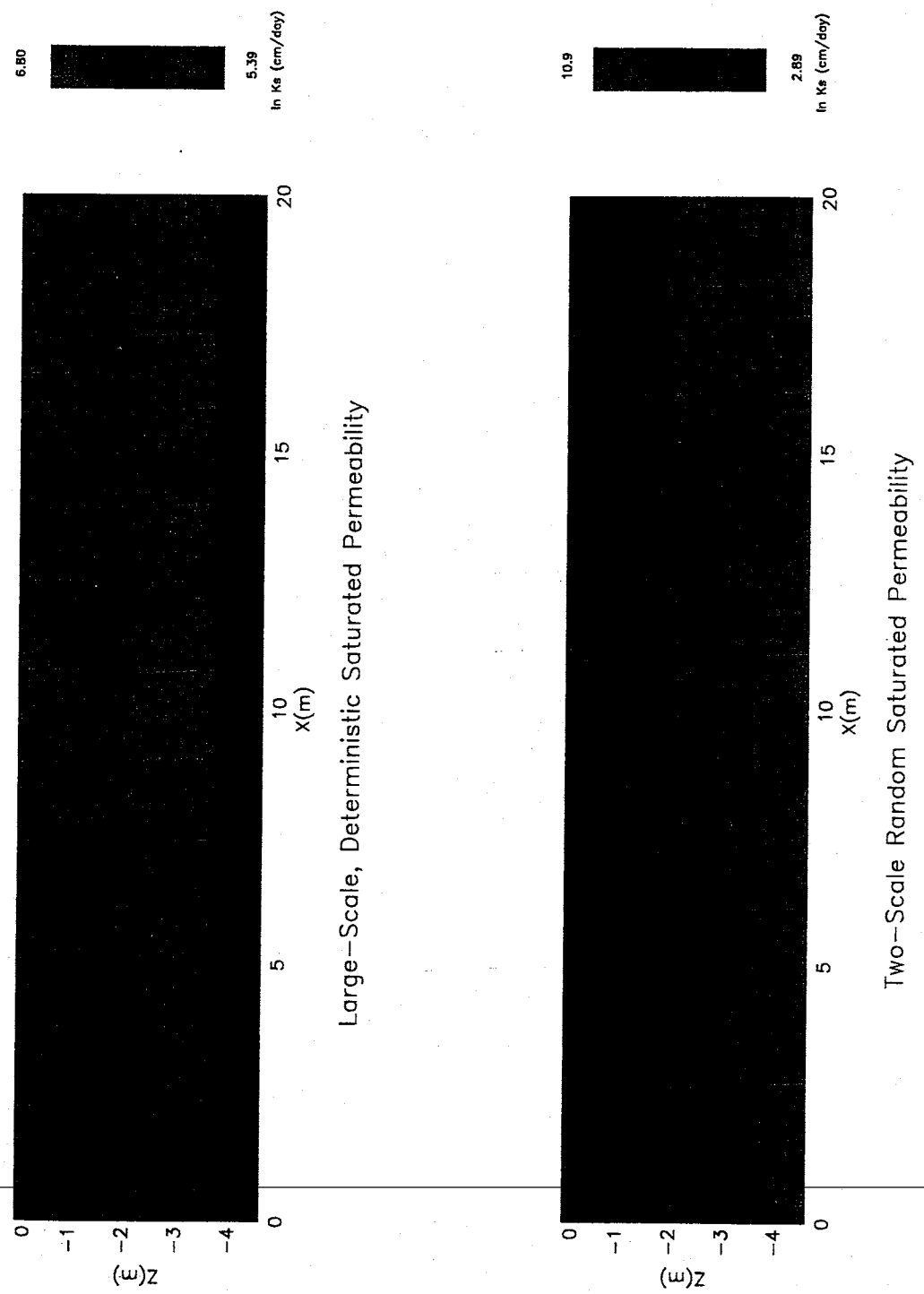


Figure 5.3. Large-Scale Deterministic and Two-Scale Random Permeability at the Las Cruces Site.

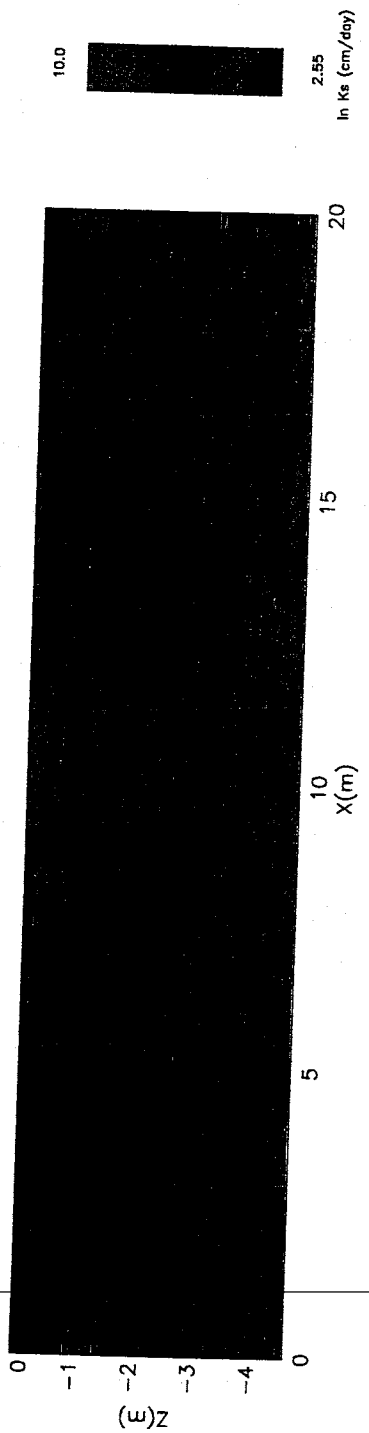


Figure 5.4. Single-Scale Permeability Structure at the Las Cruces Site.

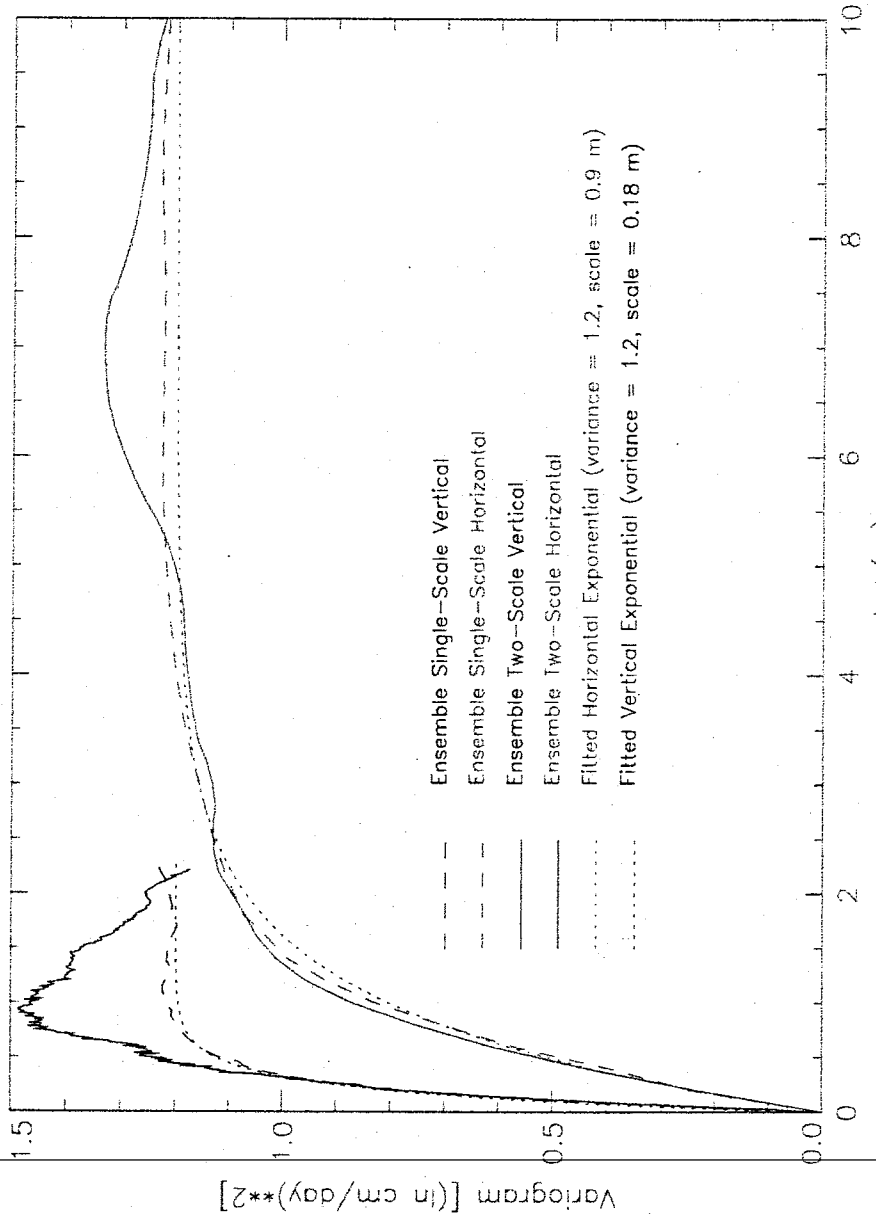


Figure 5.5. Ensemble Variogram Estimates for Single-Scale and Two-Scale Saturated Permeability Processes.

properties could be at least partly attributed to dissolved tracers encountering different scales of heterogeneity as they migrated downward from one soil layer to the next. The ensembles of two-scale and statistically equivalent single-scale k_s realizations provided a basis for assessing whether model predictions could be improved using a multiple-scale approach to permeability characterization.

5.1.2 Experimental Design

5.1.2.1 One-Dimensional Infiltration/Tracer Experiment

Following trench construction and soil sampling, an area encompassing 4 m by 9 m near the south side of the trench was instrumented with tensiometers and neutron probe access tubes to monitor water movement and with soil suction samplers to measure the distribution of bromide and tritium tracers in the subsurface following controlled application of tracer water (see Figure 5.1). Prior to trench construction, 18 6.0-m and 5 1.5-m access tubes were installed over and adjacent to the irrigation plot and calibrated onsite via gravimetric techniques. After trench excavation, seven additional access tubes were installed in a transect perpendicular to the north trench wall on the north side of the trench, and soil samples extracted at nine distinct depths used to psychometrically determine field tension heads. Finally, tensiometers were installed subhorizontally 50 cm into a portion of the south trench wall near the plot at various depths and tension monitored daily following passing of the wetting front, when tensions would not be too high to be reliably measured. Suction lysimeter solution

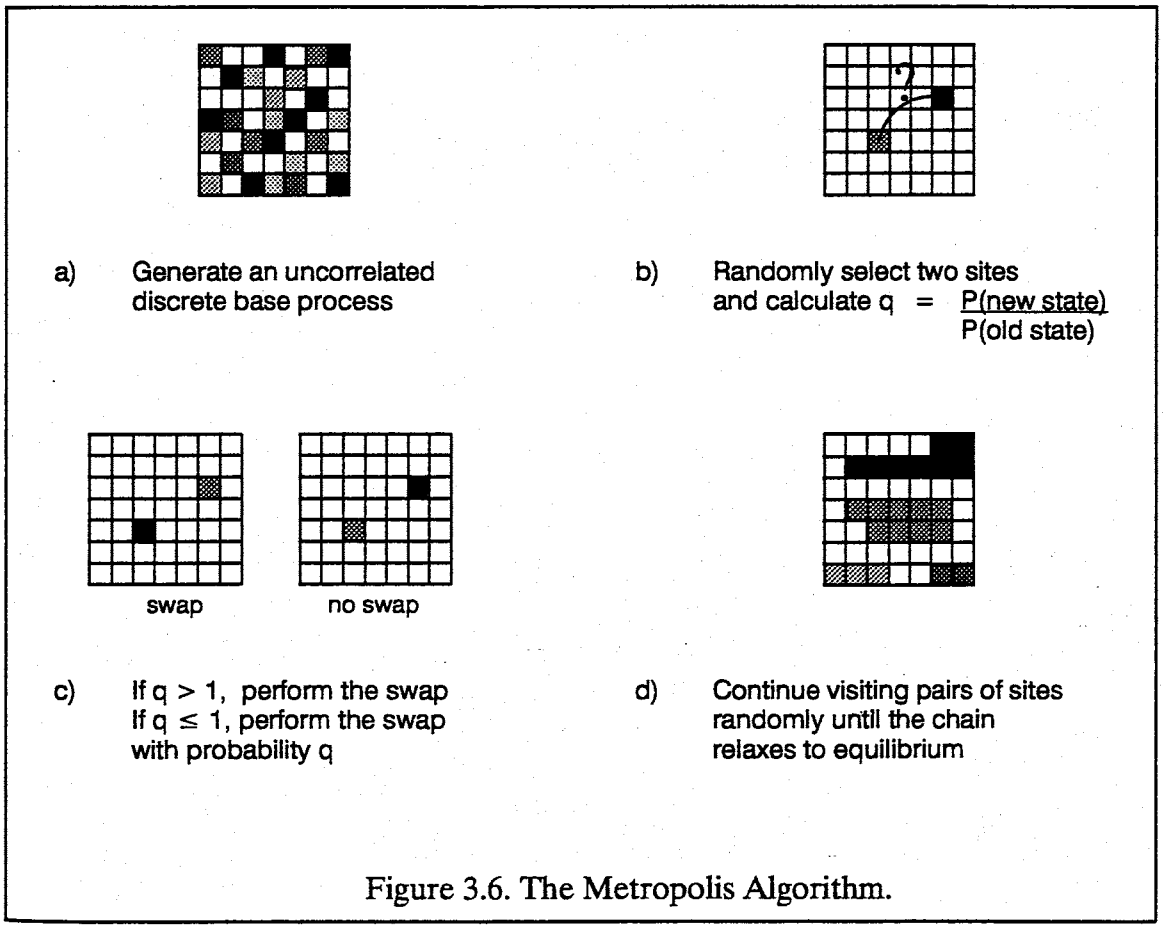


Figure 3.6. The Metropolis Algorithm.

probabilities estimated using the neighborhoods of each pair between which a potential exchange will occur. The ratio q of the conditional probability associated with the new state to that of the old state is determined as:

$$q = \frac{p[\mathbf{v}_{new}]}{p[\mathbf{v}_{old}]} = \frac{p[v_{new}(\mathbf{x}_1) \mid .]\; p[v_{new}(\mathbf{x}_2) \mid .]}{p[v_{old}(\mathbf{x}_1) \mid .]\; p[v_{old}(\mathbf{x}_2) \mid .]} = \frac{e^{-U'[v_{new}(\mathbf{x}_1)|.]-U'[v_{new}(\mathbf{x}_2)|.]}}{e^{-U'[v_{old}(\mathbf{x}_1)|.]-U'[v_{old}(\mathbf{x}_2)|.]}} = e^{-\Delta U'} \qquad [3.18]$$

where q is the ratio of the new-state probability to that of the old state. $\Delta U'$ is the difference in interaction energy between new and old states and is a measure of the degree to which the new state is more likely than the old. When either of the selected sites is a data conditioning point, the pair is simply ignored and sampling continued. Note that q is the ratio of conditional probabilities given by equation [3.13]. For a stationary MRF process, the unknown integral in the denominator drops completely from the ratio.

If q is greater than 1, $\Delta U'[v(x_0) \mid .]$ is less than zero, the new state has lower energy and more organization than the old state, and a swap is made because such an exchange results in a more likely lower-energy state. If q is less or equal to 1, then the swap is performed with probability q to allow occasional escape from a local energy minimum caused by spurious structure in the initial process. Inclusion of this option for escape from local energy minima prevents the MRF realization from being ordered just because local order happened to be artificially present in the base process. From a physical perspective, this procedure is analogous to adding erratic energy to the field in order to enter into a higher-energy state, making more subsets of lower-energy states accessible from the current state and honoring chain ergodicity requirements. Note

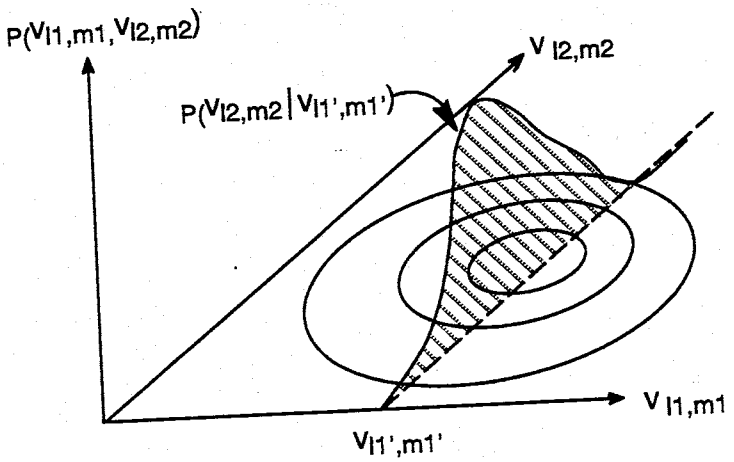
that, without the addition of this energy, lower energy states are always selectively and exponentially favored. Typically, importance sampling continues until the number of successful replacements drops below some percent of total attempts.

The Metropolis algorithm is similar to simulated annealing, except that we do not necessarily desire the most likely structure — rather, we seek only a *possible* structure. It allows us to avoid the problem of simulating states uniformly off the joint Gibbs pdf which, depending on the form of the local energy potential U'(v|.), may almost completely exclude unlikely, high-energy states as discussed in section 3.2.3. Rather than sampling unequally-probable states uniformly off the joint pdf, the Metropolis algorithm samples them nonuniformly from the state space. As a result, even the most improbable realization has a chance of being represented in the ensemble.

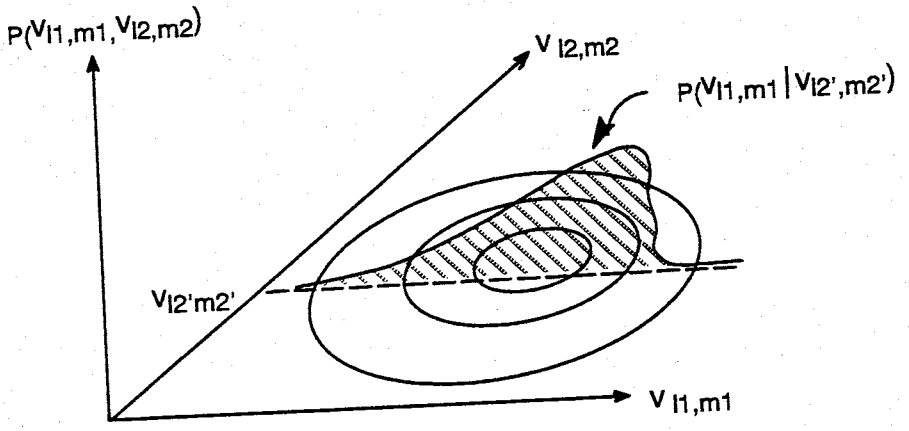
To obtain an understanding of how repeated application of the conditional operator produces a sample from a joint pdf, consider the joint pdf as a product of sequentially conditional probabilities over the entire process at lattice sites (i,j):

$$p(\mathbf{v}) \equiv p[v_{i,j} |\ v_{l,m},\ (l,m) \in \mathcal{N}_{i,j}]\ p[v_{l,m},\ (l,m) \notin (i,j)] \qquad [3.19]$$

The equivalence is approximate because there are dependencies not accounted for in the above relation. Figure 3.7 illustrates conceptually how a two-dimensional joint pdf can be factored into the product of two conditionals if the conditional distributions are independent of one another. The approximation is exact if the contours of the joint pdf



a) Conditional Probability of $v_{l2,m2}$ Given $v_{l1,m1}$



b) Conditional Probability of $v_{l1,m1}$ Given $v_{l2,m2}$

Figure 3.7. Factorization of Joint Pdf Into Component Conditional Pdfs.

are concentric circles, which is indicative of statistical independence. When the contours are ellipses oriented obliquely to the random variable coordinate axes, the two variables possess some dependence and the degree of approximation associated with equation [3.19] will degrade with increasing eccentricity of the ellipses. With each iteration of the importance sampling algorithm, yet another term is incorporated to the factorization. Such factorization makes it possible to reduce what is essentially an intractable global, absolute problem into a series of manageable local, relative ones.

3.2.8 Maximum Likelihood Estimation of Markov Parameters

Our interest lies in generating discrete MRF realizations that behave, in a statistical sense, like observed geologic structure. Two final questions remain: how are Markov parameters estimated from geologic data? and does an MRF model with the optimal parameters adequately describe the observed geologic structure? Due to the combinatorial nature of the method used to generate MRF realizations, the issue of Markovianity and the estimation of parameters can only be approached by again resorting to simulation techniques. Since simulation of an MRF requires as input the Markov parameters, we must first estimate these parameters as if the field were an MRF by assuming a specific functional form of $U'(\mathbf{v}|.)$. Methods of hypothesis testing, in which independent subsets of the observed and theoretically predicted processes are compared, can then be used to assess the likelihood that the observed process is not only an MRF of the assumed order and local energy function, but also characterized by the estimated parameters.

Since the optimal Markov parameters are those most likely to describe the observed MRF, maximum likelihood (ML) techniques can be used to estimate the true parameter set. The Markov parameters are defined in the context of conditional probabilities and must be estimated from observed data that has been divided, or coded, into mutually orthogonal sets of lattice sites. With locations in a given coding independent of one another in a Markovian sense, the conditional probabilities can be used to form a product that approximates the joint density of the process as given by equation [3.19]. This approximation is then maximized over the Markov parameter space for all coded samples to yield the most likely set of parameters to have produced the observed data. Besag (1972) was the first to use coding and conditional maximum likelihood techniques as a means of MRF parameter estimation, pointing out their obvious intuitive appeal.

Figure 3.8 illustrates how one would go about coding a first-order realization of an MRF, where each value depends on the values at the four surrounding sites immediately to the north, east, south, and west. If the observed data are considered to be from a first-order MRF realization, the two coding schemes shown in Figure 3.8 represent independent locations where the sites in any given coding are not within the neighborhoods of one another in a Markovian sense. In a similar manner, one could estimate parameters for a second-order MRF by coding the field into four disjoint sets of lattice sites. For a lattice of M sites with k codings, there are M/k independent samples of the conditional probability (Cross and Jain, 1983). Following coding into independent sets, each set is analyzed separately for maximum-likelihood parameters.

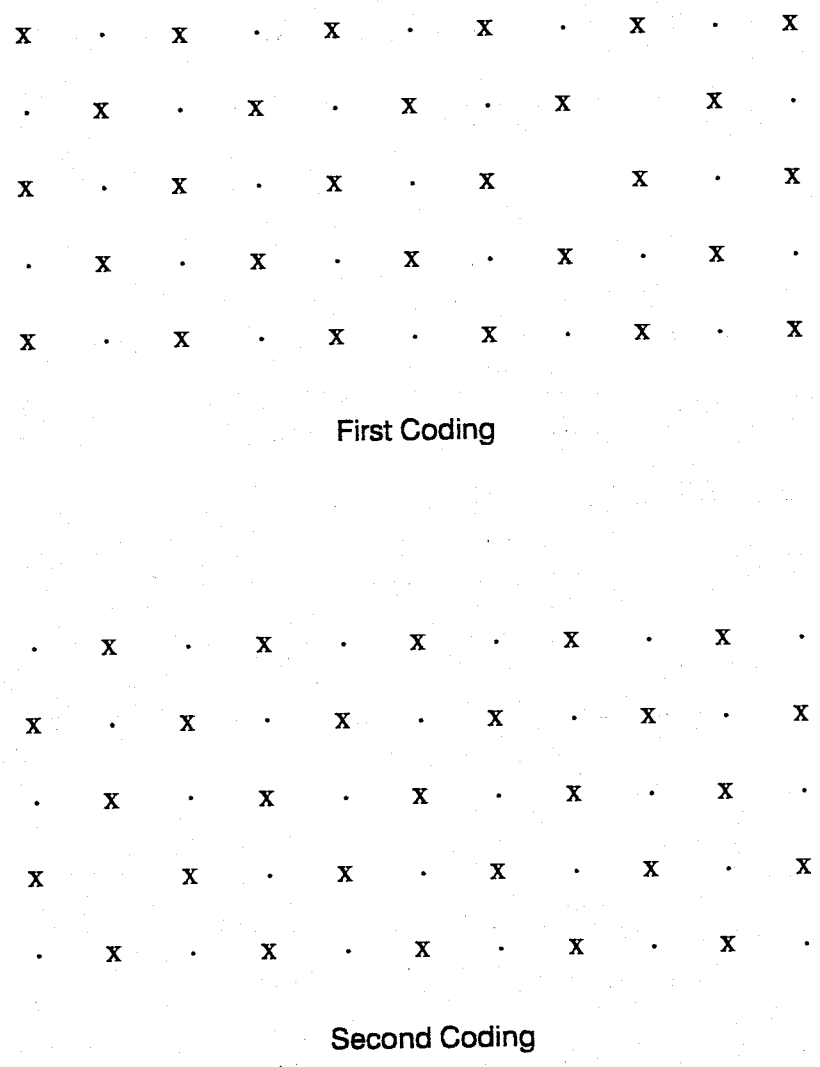


Figure 3.8. Mutually Orthogonal Codings for a First-Order MRF.

The codings commonly produce very different and potentially correlated estimates of the Markov parameters. This correlation stems from the fact that any single piece of data is used more than once. For example, in a first-order process, each data value is used four times as a neighbor and once as the conditioning value. Nevertheless, these sets of estimates are typically averaged to define representative parameters.

The maximum likelihood procedure amounts to unconstrained maximization of a nonlinear likelihood function over the Markov parameter space. The set of parameters which define the optimal likelihood are those that are most likely to occur, given their implicit manifestation through the behavior of the data. From a conditional probability standpoint the likelihood product, defined as $\mathcal{L}'(k)$, is approximately equal to the product of all conditional probabilities in a given coding k:

$$\mathcal{L}'(k) \approx \prod_{\mathbf{X}(k)} p[\, v(\mathbf{x}) \mid . \,] \qquad [3.20]$$

where the product is taken over all locations $\mathbf{X}(k)$ in the kth lattice coding. Note that the denominator of conditional probability $p[v(\mathbf{x})|.]$ is Z restricted to the neighborhood and, for a stationary process, is the same for each term in the likelihood product. As a result, the denominators in the product do not affect the maximization.

To facilitate analysis, the logarithm of the conditional likelihood $\mathcal{L}'(k)$ is determined:

$$\mathcal{L}(k) = \ln \mathcal{L}'(k) \approx \sum_{\mathbf{X}(k)} P[v(\mathbf{x}) \mid . \,] \qquad [3.21]$$

Since $\mathcal{L}(k)$ is a monotonically increasing function of $\mathcal{L}'(k)$, maximization of $\mathcal{L}(k)$ is equivalent to maximization of $\mathcal{L}'(k)$. Furthermore, objective function $\mathcal{L}(k)$ is in the form of a sum, which facilitates derivation of gradient and Hessian functions as input to a modified-Newton nonlinear gradient search algorithm. And finally, since $\mathcal{L}'(k)$ is always between 0 and 1, the logarithmic transformation serves to stretch the objective function along the likelihood axis and makes it easier to find the optimum. After substituting the form of the conditional probabilities given by equation [3.13] and the values of the process at the coded lattice locations, $\mathcal{L}(k)$ becomes a function of the neighboring values of $v(\mathbf{x})$ and the Markov parameters. The Markov parameters are determined by maximizing log likelihood over the parameter space.

There are several drawbacks to the ML estimation of MRF parameters. Unfortunately, due to the local nature of the parameters and their estimation, the parameter estimation methods are unavoidably sensitive to even the smallest of nonstationarities. This might explain why it is notoriously difficult to obtain reliable parameter estimates in a small domain, where conditional probability outliers may dominate the analysis. Besag (1974) also notes that the coding techniques are not fully efficient because only a portion of the data is used. For example, in first-order codings only 50% of the data is used for each test. However, he believes that this lack of full efficiency is more than offset by the simplicity and flexibility of the coding techniques

and maximum-likelihood parameter estimation. Finally, Besag (1974) suggests that one maximize the pseudo-likelihood, equal to the product of conditional probabilities over the entire field without regard to coding, because this method produces parameter estimates that are consistent for increasing field size.

3.2.9 Goodness-of-Fit Hypothesis Testing

It is critical to determine whether maximum likelihood parameters do, in fact, characterize the overall structure of the observed MRF. Goodness-of-fit testing represents a means of objectively determining how well the MRF model reproduces the overall character of the observed process. Since one must simultaneously determine the Markov parameters and neighborhood order when performing parameter estimation, a poor fit between MRF predicted structure and observed structure may indicate the presence of nonstationarities in the form of conditional outliers, an incorrectly guessed neighborhood order, violation of the assumption of Markovianity, or some combination of these problems. Under conditions of nonstationarity, Markov parameters estimated via a maximum likelihood approach may, in a global sense, be the most likely but may not represent the parameters adequately in a local sense. If the field is stationary and thus described by a single unique set of parameters, but no reasonable neighborhood order results in acceptance of the hypothesis that the field is an MRF, then one may be forced to conclude that the data is not a sample from an MRF but from some other stationary process.

χ^2 goodness-of-fit testing can be used to determine how well the MRF parameters actually describe the observed realization, as well as whether the observed realization is of the assumed MRF order. The hypothesis tested is that coded samples are drawn from the conditional pdf given by equation [3.13], with the same neighborhood and interaction energy function as was assumed for ML parameter estimation. The null hypothesis is that the samples are not drawn from the conditional pdf with the assumed neighborhood and interaction energy. Given that χ^2 testing involves comparison of observed and theoretical independent conditional frequency distributions to assess goodness-of-fit, we must again encode the lattice into mutually orthogonal subsets. χ^2 testing on conditional frequencies proceeds as follows:

- In each coding, every point and its uncoded neighbors represent an independent sample from the conditional probability $p[v(x)|\ .]$. Observed conditional frequencies are tabulated for all possible combinations associated with the site value and its neighboring values, referred to as conditional frequency classes.

- The expected conditional frequencies F_i , $i=1, 2, \ldots, N_F$ are determined, where N_F is the number of conditional frequency classes, using Markov parameters estimated from the maximum likelihood analysis.

- The expected conditional frequency distribution is compared to the observed conditional frequency distribution F_{oi}, $i=1, 2, \ldots, N_F$ If the χ^2 statistic for the appropriate degree of freedom indicates acceptance of the hypothesis that the two distributions are statistically equivalent, then the observed realization is accepted as a sample from an MRF characterized by the estimated parameters and the assumed neighborhood order. Otherwise, the process is not Markovian, not of the assumed order, or simply not statistically homogeneous.

Cross and Jain (1983) recommend that since test results from the codings are not independent, the significance level be assumed equal to $k\phi$ rather than α^k, where α is the significance level applied simultaneously for all of the k codings and ϕ represents the level at which the greatest significance occurs in the k tests. The former statistic is more likely lead to rejection when this is done, and therefore results in a more conservative acceptance rule. Note that large χ^2 test statistics may result if the sample is small because small samples can produce very low conditional frequencies in some of the classes. In the extreme case, if zero conditional frequencies occur, the positivity condition is violated. Such a situation would be especially likely for high-order fields for which the size of the domain is small.

3.3 APPLICATION OF MRFs TO CHARACTERIZATION OF DISCRETE GEOLOGIC STRUCTURE

3.3.1 Evolution of a 64-By-64 MRF Realization

Figure 3.9 is an example of how a 64-by-64 MRF realization evolves along a relaxation chain. Parameters of the second-order MRF process were assumed to be α=1.0, β=0.5, γ=1.0, η=0.5, μ=0.0, ξ=0.0 and an interaction energy of the form provided by equation [3.17] specified. Lattice site states were permitted to assume only the values −3.0, −4.0, −5.0, −6.0, and −7.0. From the figure, it is clear that relaxation is a highly nonlinear process — much more reordering occurs early in the chain because little structure exists and marginal information gained from additional importance

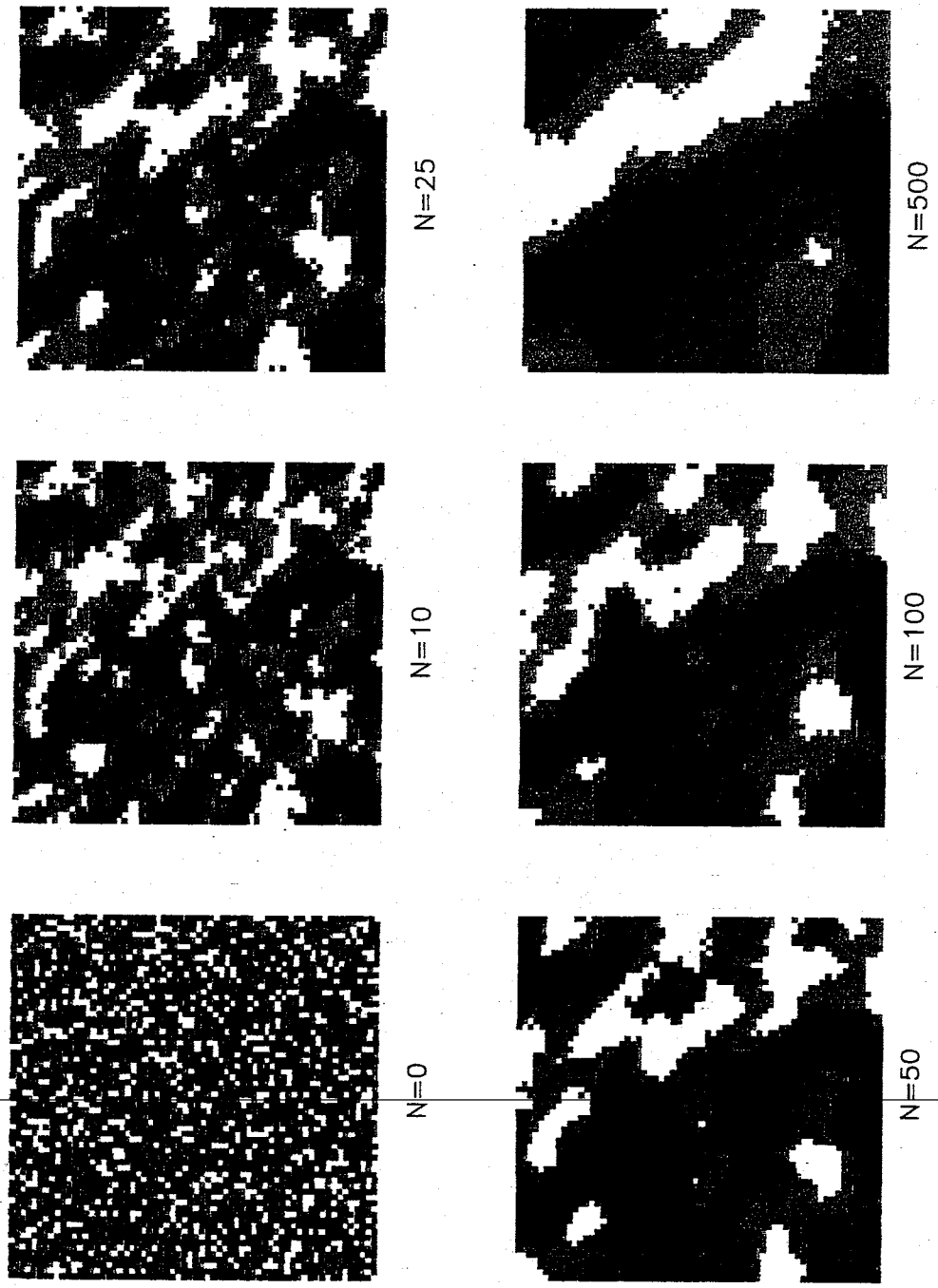


Figure 3.9. Evolution of an MRF Realization from a Relaxation Chain.

sampling is very valuable, introducing significant added structure to the process. As the chain progresses and energy becomes more organized, acquired information becomes more redundant and less useful in terms of increasing MRF structure. Farmer (1989) reports that use of the Metropolis method for simulated annealing requires on the order of 100 attempted or 10 successful replacements per site for convergence to the most likely state. Since the theoretical rate of convergence for large N is of order $\sqrt{N}$, this slowdown in convergence rate is to be expected (Geman and Geman, 1984). Shown in Figure 3.10 is percent successful exchanges vs. attempted exchanges N for the first-order 64-by-64 MRF relaxation chain. It is evident from the plot that the rate of successful exchange approaches $1/\sqrt{N}$ for large N. Likewise, the error of estimate for moments of the joint pdf is $O(N^{1/2})$ (Hammerseley and Handscomb, 1964).

3.3.2 Estimation of MRF Parameters for Alluvial-Fan Structure

To illustrate how useful MRF models can be for describing realistic geologic structure, a discrete MRF model was used to characterize two-dimensional alluvial-fan structure underlying Kirtland Air Force Base in the Albuquerque Basin, central New Mexico. Lithofacies-type data became available when a six-mile long sewer line trench was excavated into the fan deposits in 1991 (see Figure 3.11). Six distinct lithofacies types (L=6) were defined along the walls of the excavation (Gaither *et al.*, 1993; Crowson *et al.*, 1993). Figure 3.12 shows the distribution of lithofacies type along roughly 3.5 miles in the main stem of the trench, assumed to be a linear feature for purposes of data analysis and presentation. It is evident that the top of the trench tends

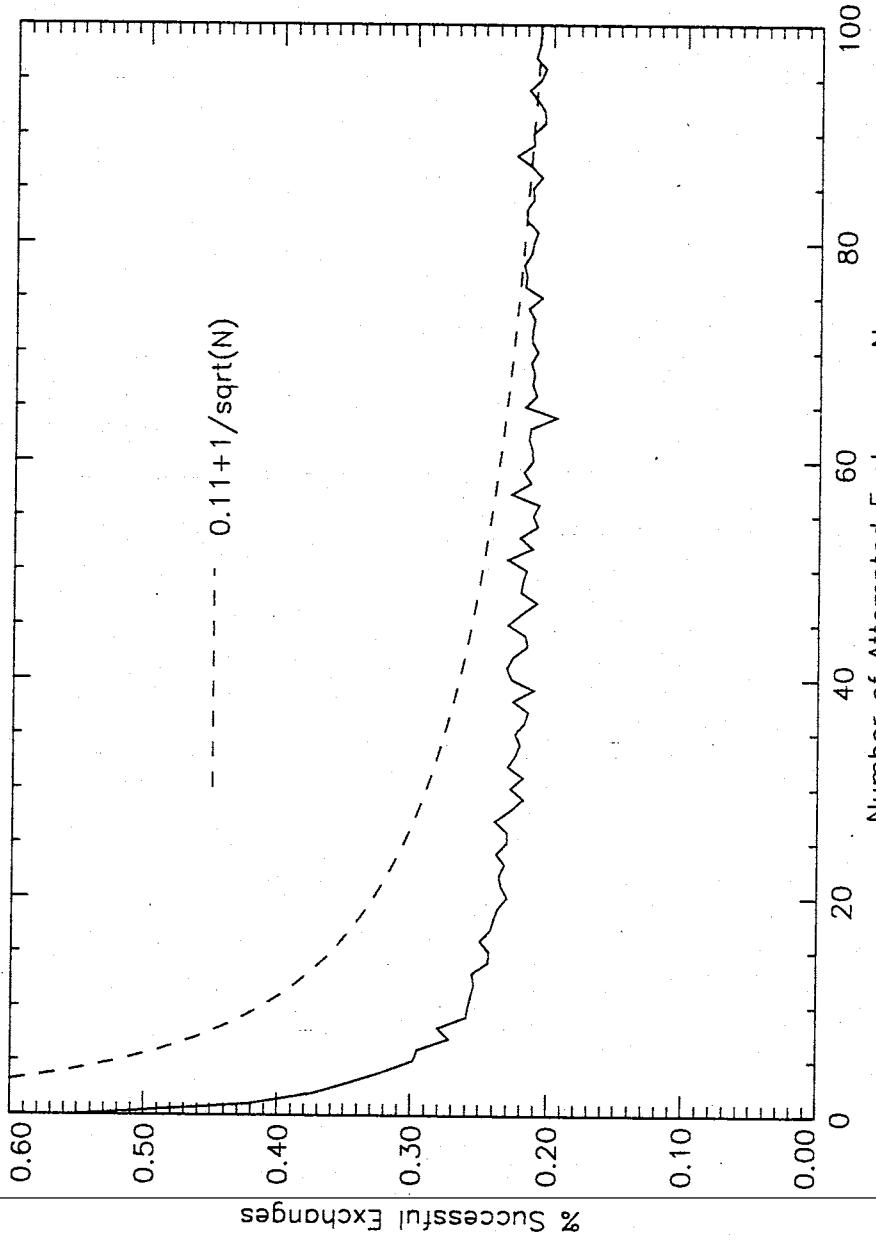


Figure 3.10. Percent Successful Exchanges vs. Attempted Number of Exchanges N for First-Order 64-By-64 MRF Realization.

to show a spatially-uniform lithology indicative of surficial weathering, while the middle and bottom portions exhibit more heterogeneous behavior.

To estimate MRF parameters for the fan deposits, a maximum-likelihood approach was used. The ML algorithm relies on mutually-orthogonal codings of the lattice sites, which depend on the order of the neighborhood assumed. It was initially assumed that the observed lithofacies distribution was a realization from a second-order MRF, with pair potential interaction energy given by equation [3.17] for $\mu=\zeta=0$. However, when second-order ML parameters corresponding to oblique structure γ and η were found to be consistently close to 0.0 compared to the first-order parameters α and β, a first-order MRF was instead adopted and the number of sites in a neighborhood, including the coded site itself, reduced from 9 to 5. This assumption considerably decreased the computational effort involved in subsequent χ^2 goodness-of-fit testing, reducing the number of conditional frequency classes from approximately $L^M=6^9=10{,}077{,}696$ to roughly $6^5=7776$.

The likelihood function was defined according to equation [3.20] using a discrete summation approximation of conditional probability given by equation [3.13]:

$$\mathcal{L}'(k) \approx \prod_{i=1}^{M(k)} p(v_i \mid .) = \prod_{i=1}^{M(k)} \frac{e^{U'(i)}}{\sum_{l=1}^{L} e^{U'(i,\,l)}} \qquad [3.22]$$

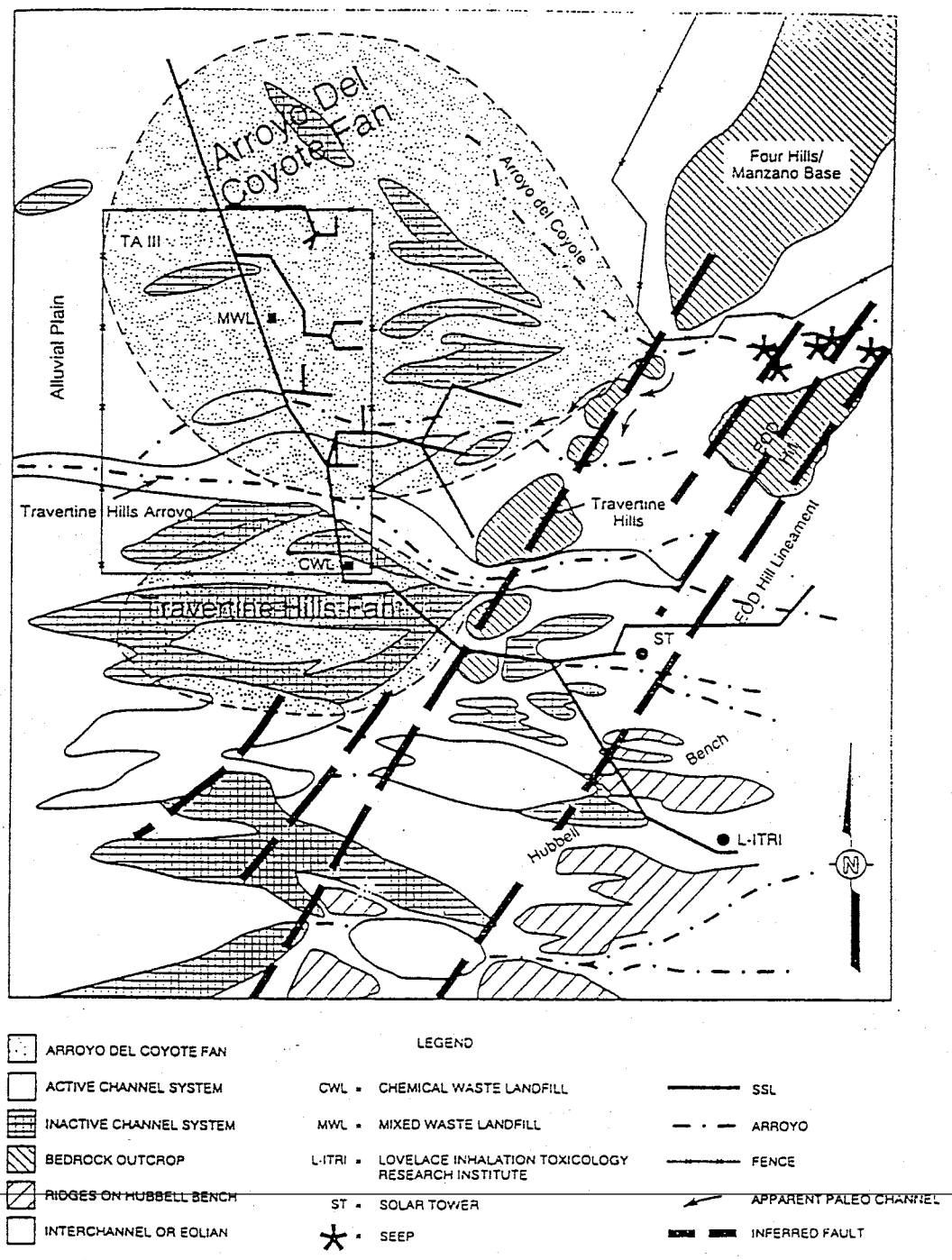


Figure 3.11. Sewer-Line Trench Excavation at Kirtland Air Force Base (courtesy of Sandia National Laboratories).

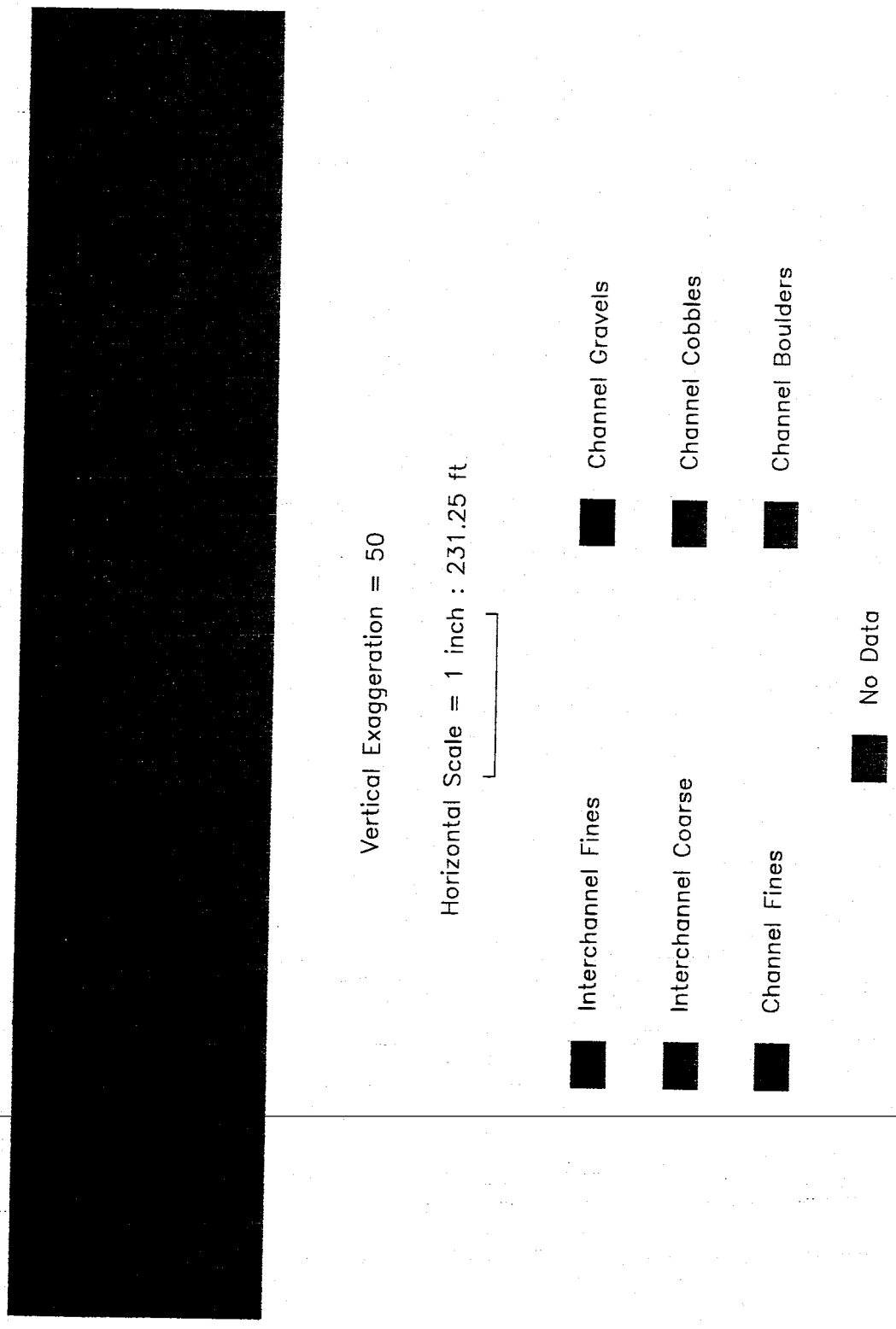


Figure 3.12. Observed Lithofacies Type Distribution Along Main Stem of Sewer Trench.

where U'(i) is the interaction energy based on the current state at the ith coded site, U'(i,*l*) is the interaction energy at the ith coded site if state level *l* were substituted at the site, L is the total number of possible state levels at each site, and M(k) is the total number of coded sites for the kth coding. Taking the logarithm of $\mathcal{L}(k)$ yields:

$$\mathcal{L}(k) = \ln \mathcal{L}'(k) \approx \ln \left[\prod_{i=1}^{M(k)} \frac{e^{U'(i)}}{\sum_{l=1}^{L} e^{U'(i,l)}} \right]$$

$$= \sum_{i=1}^{M(k)} \ln \frac{e^{U'(i)}}{\sum_{l=1}^{L} e^{U'(i,l)}} = \sum_{i=1}^{M(k)} \left[U'(i) - \ln \sum_{l=1}^{L} e^{U'(i,l)} \right]$$

[3.23]

Since $\mathcal{L}(k)$ is a nonlinear objective function, both first- and second-order derivatives with respect to the parameters α and β were required to perform a modified-Newton gradient search. These are shown derived in Appendix B.

Equations [B.1], [B.2], and [B.3] presented in Appendix B form the basis for separate modified-Newton gradient searches conducted over the log of the likelihood function for each independent coding of the assumed first-order MRF. Three different forms of pair-potential interaction energy were tested. One was based on absolute values of differences between lithology indices as given by equation [3.17], another

involved squaring the absolute differences, and the third was defined using the square root of the absolute differences.

Shown in Figure 3.13 are the log-likelihood functions for each tested form of the interaction energy and for each of two independent, first-order codings. From inspection of the log-likelihood plots, it is evident that parameters estimated for the same energy function but different codings agreed closely. Although optimal α and β were similar between two of the three models, the log-likelihood function tended to be more pronouncedly peaked for the square-root interaction energy model. Note that, for each form of the pair-potential interaction energy, the maximum-likelihood parameter estimation procedure identified the most likely parameters given the assumptions that the MRF is order 1, stationary, and with interaction energy of the assumed form, but does not provide any indication of how appropriate those assumptions are.

3.3.3 χ^2 Goodness-of-Fit Testing

To address this issue, χ^2 goodness-of-fit tests were conducted to determine how closely the observed realization of trench lithologies conformed to a first-order MRF model for each form of the interaction energy. In each case, the null hypothesis was that the observed lithofacies spatial distribution was drawn from a first-order MRF with the assumed form of the interaction energy. To this end, conditional frequencies associated with the observed facies-type realization presented in Figure 3.12 were compared to expected conditional frequencies. Expected conditional frequencies were

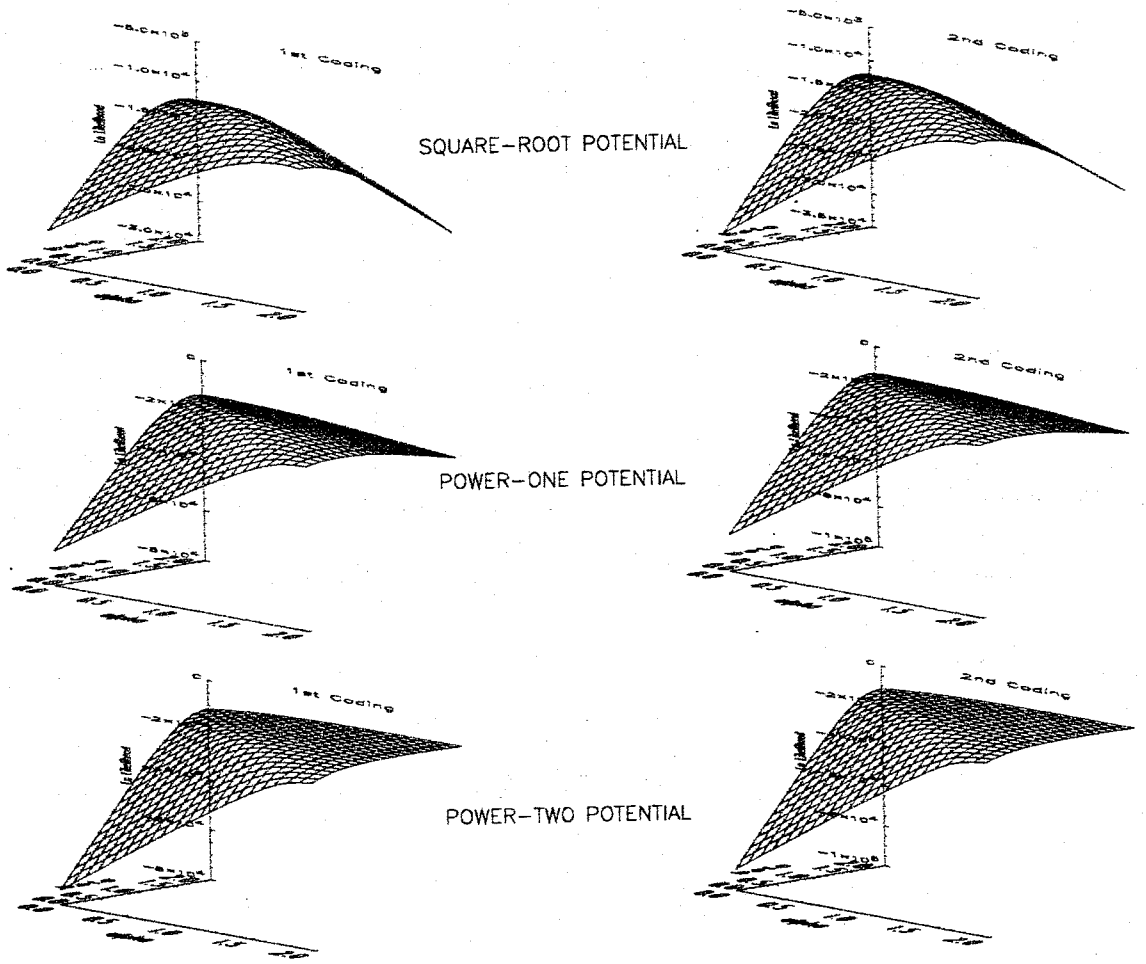


Figure 3.13. First-Order Log Likelihood Functions for Lithofacies Type Data Using Various Forms of Interaction Energy.

obtained by multiplying conditional probabilities given by equation [3.13] by the total expected frequency of each distinct neighborhood configuration $N_x N_z/6^4$, where N_x=740 is the number of horizontal sites, N_z=15 equals the number of sites in the vertical direction, and 6^4 neighborhood configurations are possible. The summation in the denominator of equation [3.22] was made tractable by this fairly small number of possible neighborhood configurations.

The number of degrees of freedom associated with all local configurations, including both site and neighbors, was $(6\text{-}1)6^4$. This was reduced by 2 to reflect the number of MRF parameters (α and β) estimated previously from the same data via maximum likelihood, yielding a value of 6478. The χ^2 statistic corresponding to a significance level of 0.05 and 6478 degrees of freedom was estimated from the approximation $df=1/2[(2(df)\text{-}1)^{1/2}-1.64]^2$ for large *df* to be equal to 6292 (Benjamin and Cornell, 1970). Note that, for this large value of *df*, the χ^2 statistic is sensitive to the data and may not be a good estimate for the data as a whole.

Results of the χ^2 tests are presented in Table 3.1. The chi-squared statistic used to accept or reject the assumed MRF model was twice the statistic for the given degrees of freedom because the statistics for both codings were dependent, even though the codings themselves were mutually orthogonal. Results suggest that the hypothesis cannot be accepted at a significance level of 5% for any pair-potential interaction-energy model nor for either independent coding. Based on these results, there is greater than a 5% chance of accepting the first-order, pair-potential energy MRF model when it is incorrect, regardless of assumed energy interaction power. It

Table 3.1. Results of Maximum Likelihood MRF Parameter Estimation and χ^2 Goodness-of-Fit Testing.

Power of Pair-Potential Interaction Energy	α	β	$\chi^2 = \sum_{i=1}^{N_F} \frac{(F_{oi} - F_i)^2}{F_i}$		$2\chi^2_{0.05,\,6478} = 2\,(6292)$
			Coding 1	Coding 2	
0.5	1.06	1.13	107641	17892	> 12584
1.0	0.81	0.78	219653	26578	> 12584
2.0	0.24	0.13	690130	75503	> 12584

does appear, however, that the degree of peakedness of the log likelihood function in Figure 3.13 is an excellent indicator of relative goodness-of-fit, with the smallest estimated χ^2 value associated with the square-root model.

Figure 3.14 presents an unconditional MRF realization of lithofacies type synthesized with the Metropolis algorithm using the ML optimal parameters of the first-order, square-root model equal to α=1.06 and β=1.13, obtained after 5000 attempted exchanges per site. The tendency toward spatially-uniform properties at the top of the trench is not captured because such uniformity is essentially a nonstationary influence that the strictly stationary MRF model cannot possibly reproduce without recourse to local conditioning on observed lithofacies type. The failure of the χ^2 test and the poor reproduction of observed fan deposit structure were both attributed primarily to nonstationarities pervading the observed lithofacies type data. While the multivariate Gibbs pdf can account for arbitrary joint orders of correlation, it cannot incorporate randomness that is spatially invariant without benefit of extensive data conditioning.

It should be pointed out that, had permeabilities been assigned to each facies type and those permeabilities used to generate a permeability realization, the realization and MRF parameters would have been somewhat different. This is due to dependency of interaction energy on differences between random variables within the neighborhood.

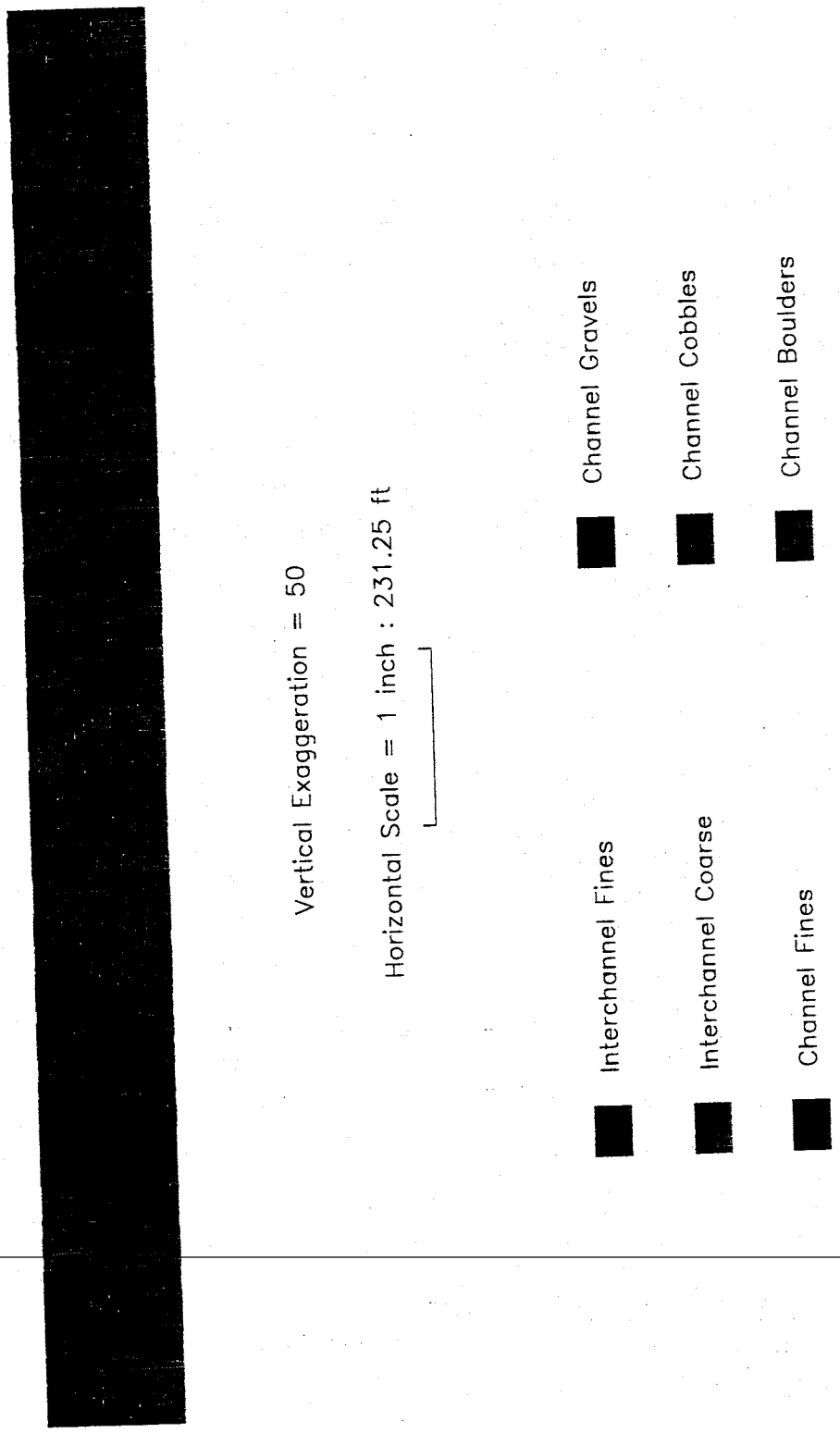


Figure 3.14. Simulated Lithofacies Type Distribution Along Main Stem of Sewer Trench Using Estimated MRF Parameters.

3.4 SUMMARY

Markov random fields (MRFs) impose spatial correlation over discrete space and are ideally suited for generation of geologic structure at larger scales of variation, where abrupt lithologic and stratigraphic boundaries typically prevail. They represent the extension of one-dimensional Markov chains to multivariate space. MRF realizations are synthesized using an importance sampling algorithm and a conditional probability operator analogous to the transition probability matrix used to generate one-dimensional Markov chains. Conditioning is performed at each level of the chain with respect to neighborhoods associated with pairs of randomly visited sites. The rate of chain convergence to a *bona fide* MRF realization is of order $\sqrt{N}$, where N is the number of exchanges per site.

MRF parameters for alluvial fan structure in the Albuquerque Basin were estimated on the basis of mutually independent subsets of observed lithofacies type data using maximum likelihood techniques. The observed lithofacies type realization could not be accepted as a sample from a first-order MRF characterized by any one of three pair-potential interaction energies, presumably due to strong nonstationarities evident in the observed lithofacies type distribution.

3.5 REFERENCES

Benjamin, J.R., and C. Allin Cornell, *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill, New York, 1970.

Besag, J., Nearest-neighbor systems and the auto-logistic model for binary data, *J. Royal. Stat. Soc., series B*, 34, 75-83, 1972.

Besag, J., Spatial interaction and the statistical analysis of lattice systems (with discussion), *J. Royal. Stat. Soc., series B*, 36, 192-226, 1974.

Cross, G.R., and A.K. Jain, Markov random field texture models, *IEEE Trans. on Pattern Recog. and Mach. Intel.*, PAMI-5(1), 1983.

Crowson, D., J.T. McCord, R. Holt, J. Piper, and E.K. Webb, Detailed mapping and preliminary geostatistical analysis of alluvial fan deposits exposed in six miles of trench, *EOS, Transactions, American Geophysical Union, 1993 Spring Meeting*, 74(16), 143, 1993.

Dobrushin, R.L., The description of a random field by means of conditional probabilities and conditions of its regularity, *Theory Prob. Appl.*, 13, 197-224, 1968.

Farmer, C.L., *The Generation of Stochastic Fields of Reservoir Parameters with Specified Geostatistical Distributions*, paper presented at the IMA Conference on The Mathemetics of Oil Production, Robinson College, Cambridge Univ., July 6-7, 1987.

Farmer, C.L., *Numerical Rocks, The Mathematical Generation of Reservoir Geology*, paper presented at the joint IMA/SPE Conference on The Mathemetics of Oil Recovery, Robinson College, Cambridge Univ., July 25-27, 1989.

Gaither, K., J. Piper, and R. Holt, Geologic mapping of the sanitary sewer line excavation, southern Kirtland Air Force Base, Albuquerque, New Mexico, *Sandia Report SAND 92-2517*, 1993.

Geman, S., and D. Geman, Stochastic relaxation, Gibbs distributions, and the Bayesian restoration of images, *IEEE Trans. on Patt. Recog. and Mach. Intel.*, PAMI-6(6), 1984.

Hammersley, J.M., and D.C. Handscomb, *Monte Carlo Methods*, Methuen, London, 1964.

Isham, V., An introduction to spatial point processes and Markov random fields, *Int. Stat. Rev.*, 49, 21-43, 1981.

Ising, E., Zeitschrift Physik, 31, 253, 1925.

Metropolis, N., A.W. Rosenbluth, M.N. Rosenbluth, A.H. Teller, and E. Teller, Equations of state calculations by fast computing machines, *J. Chem. Phys.*, 21, 1087-1092, 1953.

Rothman, D.H., Nonlinear inversion, statistical mechanics, and residual statics estimation, *Geophysics*, 50(12), 2784-2796, 1985.

Spitzer, F., Markov random fields and Gibbs ensembles, *Amer. Math. Mon.*, 78, 142-154, 1971.

Whittle, P., *Systems in Stochastic Equilibrium*, J. Wiley and Sons, Chichester, 1986.

Chapter 4. Optimization of Hydraulic Conductivity Measurement in Multiple-Scale Geologic Media

4.1 CONDUCTIVITY MEASUREMENT IN MULTIPLE-SCALE POROUS MEDIA

In Chapter 2, it was demonstrated empirically using a numerical experiment that effective properties can change significantly when a multiple-scale conductivity process is conditioned on point noisy permeameter data. Results of that information transfer experiment suggest that, in geologic media characterized by more than one scale of spatial variation, use of conductivity measurements collected at the small scale to make predictions at a larger scale should be avoided whenever possible. Instead, conductivity should be measured independently at each scale relevant to the particular flow or transport problem of interest, ideally using instrument windows capable of measuring at characteristic frequencies that most dominate the problem.

This issue of scale-dependent randomness and measurement naturally leads to the question: At what scale should we measure random multiple-scale hydraulic conductivity if we want to minimize some measure of transport prediction uncertainty?. For example, if we are interested in minimizing concentration prediction uncertainty at a regulatory compliance point such as a monitoring well, how should we go about choosing the scale at which to measure conductivity? Clearly, this issue of

measurement scale has meaning only when posed in the context of a particular contaminant transport problem.

Depending on the nature of the transport problem and on the character of the heterogeneities, more marginal information about the prediction might be gained by favoring certain scales of data collection over another. For example, if we have reason to believe that most of the impact on contaminant migration is occurring at the smaller scale, we would stand to gain the most information about well concentration from *in situ* permeameter measurements and lab testing of core samples (see Figure 4.1). If,

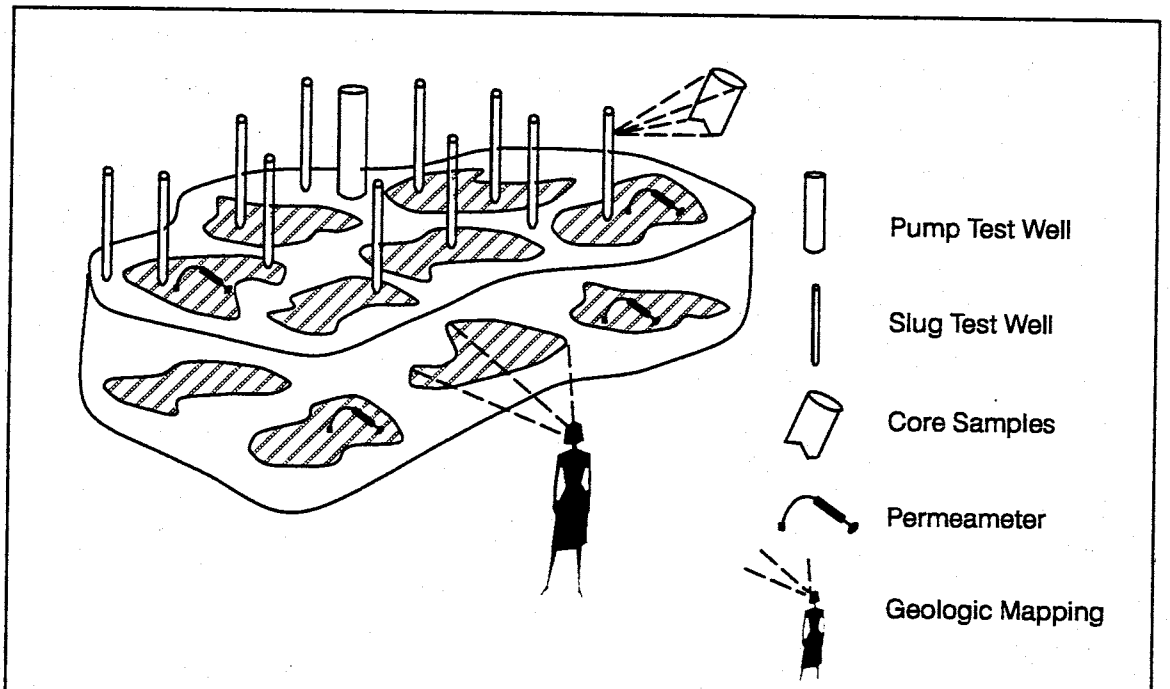


Figure 4.1. Multiple-Scale Conductivity Measurement.

samplers were also installed through the south trench wall at varying depths and were used to obtain solute samples following passing of the wetting front (Wierenga *et al.*, 1991). After the instrumentation was complete, a string grid was installed over the instrumented portion of the south wall so that tensiometer and neutron probe access tube readings of the wetting front advance could be verified through visual observation of moisture movement.

The one-dimensional infiltration experiment involved application of water containing dissolved, conservative tracers uniformly over the 4 m by 9 m infiltration area using a drip irrigation system. Tritium was dissolved in the infiltrating water only during the first 10 days of the 86-day experiment at a concentration of 0.01 mCi/m*l*. The average flux of the infiltrating water, applied intermittently for 17 minutes four times per day to avoid ponding, was 1.82 cm/day, resulting in a total mass of introduced tritium equal to 62.6 mCi in 6260 *l* water. A pond liner was used to protect the infiltration plot and trench areas from both evaporation and rainfall.

Following application of the water, the wetting front was visually observed along the trench face and measured via neutron probes and tensiometers every few hours during infiltration and every few days during redistribution (Wierenga *et al.*, 1991). The wetting front at 0.5 m from the wall was defined as the locus of all points at which neutron-probe moisture content was halfway between initial and saturated content. Since the neutron probe data were not very effective in delineating sharp wetting fronts, near-front data were used to estimate when tensions were sufficiently low to be measured via tensiometer. Tensions observed at a distance of 0.5 m from the

trench wall were used to more quantitatively determine the advance of the front, which was arbitrarily defined as the location where tension head decreased and remained below 100 cm. When tension heads were greater than 150 cm and thus too high to be measured using the tensiometers, they were inferred from neutron-probe water content measurements using the characteristic curve of water content (θ) vs. suction head. Visual observations and measurements were assumed to comprise statistically independent data sets because they were obtained using different methods.

The wetting front, as visually observed along the trench wall, exhibited a slight asymmetry that could have been related to accidental introduction of rainfall. The tension values showed evidence of a front in advance of the one defined through visual inspection, but appeared to reaffirm some asymmetry in the advance of the front. The slower rate of front movement observed visually at the wall, compared to 0.5 meter from the wall, may have been related to greater soil desiccation occurring along the trench wall. These dryer conditions were caused by evaporation of moisture from the trench following its construction. The front measured perpendicular to the wall showed further evidence of slower movement near the wall, and also suggested that there was a small degree of lateral spreading of water as one moved away from the wall.

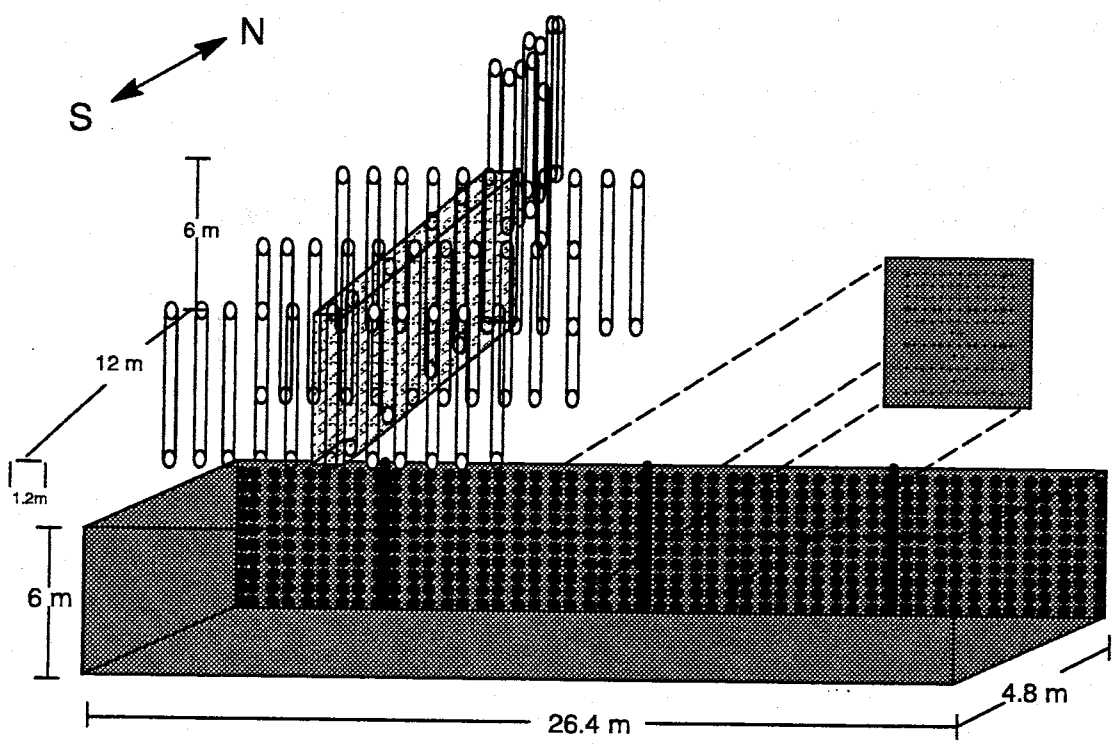
5.1.2.2 Two-Dimensional Infiltration/Tracer Experiment

Due to the lack of lateral spreading of water and tracer observed during the one-dimensional experiment, a two-dimensional experiment was designed using an less areally extensive, smaller magnitude irrigation source than was used for the first

experiment. By applying water and tracer over an irrigation plot smaller than the horizontal correlation length, it was anticipated that lateral spreading would be much more pronounced than for the first experiment. Such lateral spreading would allow the two-dimensional subsurface correlation structure to be sampled by the infiltrating water and migrating tracer.

As shown in Figure 5.6, an area extending over 1.2 m by 12 m adjacent to the north side of trench was instrumented with 43 neutron access tubes, each calibrated through gravimetric sampling to a depth of 6 m. Water content was sampled using neutron probes at 0.25 m intervals every 4-5 days until irrigation was completed, after which sampling frequency was reduced. As for the one-dimensional experiment, soil solution samplers were installed and tensiometers inserted 50 cm subhorizontally into the wall of the trench and both soil suction and tension measurements collected every 2-3 days until completion of irrigation, again with reduced frequency following irrigation. These tensiometers were used as soil solution samplers after passage of the wetting front in order to obtain additional tracer solution measurements. A number of tensiometers were also installed vertically within the irrigation plot. Neutron probe, tensiometer, and soil sampler instrumentation provided information over several vertical two-dimensional grids that permitted global tracking of moisture content over three dimensions as well as measurement of vertical water and tracer redistribution locally over two dimensions.

The two-dimensional infiltration/tracer experiment was conducted by imposing an average surface flux of 0.43 cm/d over the 1.2 m by 12 m irrigation plot for 75.5 days.



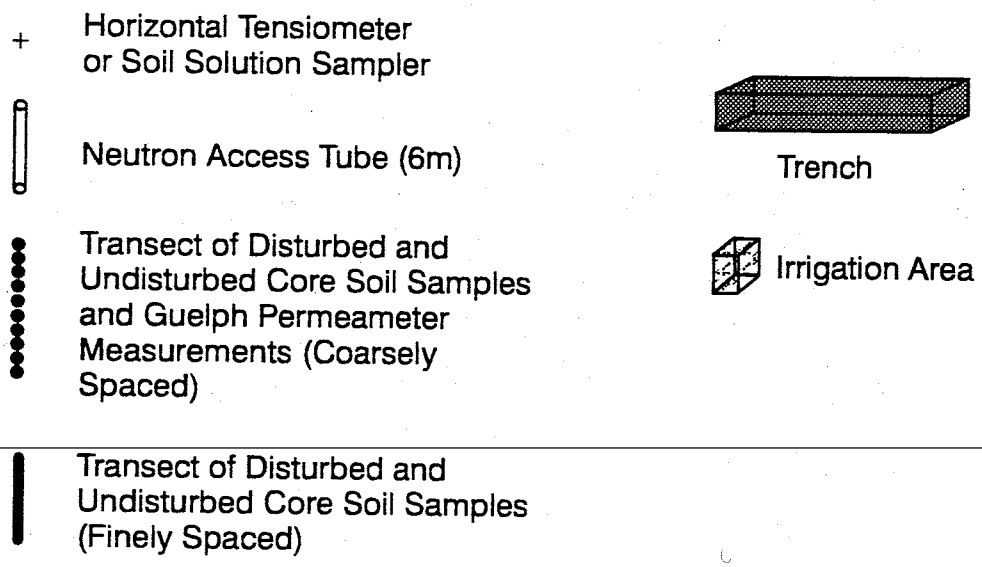


Figure 5.6. Instrumentation for the Second Las Cruces Trench Infiltration Experiment.

Irrigation water contained concentrations of 0.1 mCi/*l* and 939 mg/*l* of tritium and bromide, respectively, for the first 11.5 days of the experiment.

5.2 EXPERIMENTAL RESULTS AND PREVIOUS NUMERICAL TRANSPORT PREDICTIONS

5.2.1 One-Dimensional Infiltration/Tracer Experiment

A one-dimensional numerical infiltration model based on Richards' equation was designed to predict the distribution of water during the first infiltration experiment. The independent variable was chosen to be water content because, for the extremely desiccated soils found under the site, it resulted in a more accurate mass balance than head-based methods. However, moisture content is not generally continuous across soil layers, where abrupt changes in soil properties typically occur. Mass balance was imposed by requiring both suction head and fluid flux to be equivalent on either side of each water-content discontinuity. Hills *et al.* (1989 a,b) found that, under desiccated conditions when tensions exceed 50 bars, a water-content based model was more computationally efficient than the more traditional tension-based model.

When tensions were not too large to be measured, water contents used as initial conditions for transient numerical flow simulations were determined from the van Genuchten water retention model on the basis of tensions measured in soil samples extracted at 9 different depths from 7 neutron access tubes located on the north side of the trench. Otherwise, soil sample measurements of tension were obtained

psychometrically. Linear interpolation was used to obtain initial tensions between measurement depths. It was assumed that the moisture capacity was roughly constant within the range of moisture contents encountered during the experiment, even though this required that initial moisture prior to simulation be taken as somewhat high.

A flux boundary condition was imposed at the surface to simulate flow of 1.82 cm/d into the subsurface from the drip irrigation system, while a constant water-content boundary condition equal to the initial water content was specified at a depth of 5.74 m, considered to be sufficiently large to be unaffected by the advancing wetting front. Using a uniform grid spacing of 2 cm and a time step of 0.05 days, transient infiltration and redistribution were simulated for 35 days of the 86-day experiment assuming both uniform homogeneous and layered heterogeneous soil models (Wierenga *et al.*, 1991).

Results of the one-dimensional simulations demonstrated that the layered soil model was a better predictor of initial water content than the uniform soil model, but there was no clear advantage to either model after 19 and 35 days of simulation, respectively (Wierenga *et al.*, 1991). At all times, the uniform soil model predicted a slower advance of the front than the neutron access tube data, but a more rapid advance than visual inspection or tensiometer data suggested. The fact that the wetting front predictions for the uniform model were bounded by field observations was encouraging, especially considering that physical properties input to the numerical model were calculated on the basis of measurements collected from the opposite north side of the trench, that there was a significant variation in saturated conductivity over the trench site, and that a theoretical moisture retention model was used to help define

initial water content and unsaturated permeability via equations [5.1] and [5.2]. Close agreement between uniform and layered model predictions suggested that, while there may have been local variations in moisture content and saturated permeability, the global movement of water tended to average these variations out. On that basis, Wierenga *et al.* (1991) argued that at the scale of this particular flow problem, it made little sense to attempt to quantify spatial variations in permeability. This conclusion was reinforced by the fact that differences between water-content predictions based on uniform and layered permeability were smaller than differences among water content observations made along the centerline of the irrigation plot, suggesting that greater refinement in delineation of soil layers desiccation would not appreciably improve the accuracy of one-dimensional model predictions. However, it was speculated that within-layer lateral variations in permeability might be important when predicting response in two dimensions. This could be an issue when surface ponding occurs because it tends to induce significant lateral spreading. Under such circumstances, information related to small-scale horizontal variability could vastly improve our ability to predict the redistribution of water in the subsurface.

5.2.2 Two-Dimensional Infiltration/Tracer Experiment

Figure 5.7 shows water content change distributions observed at various times during the second, two-dimensional infiltration experiment at a distance 2.0 m north of the northern trench wall, in a plane parallel to that wall. Note that, at late times the observed water content exhibited a distinct asymmetry, suggesting that there was some

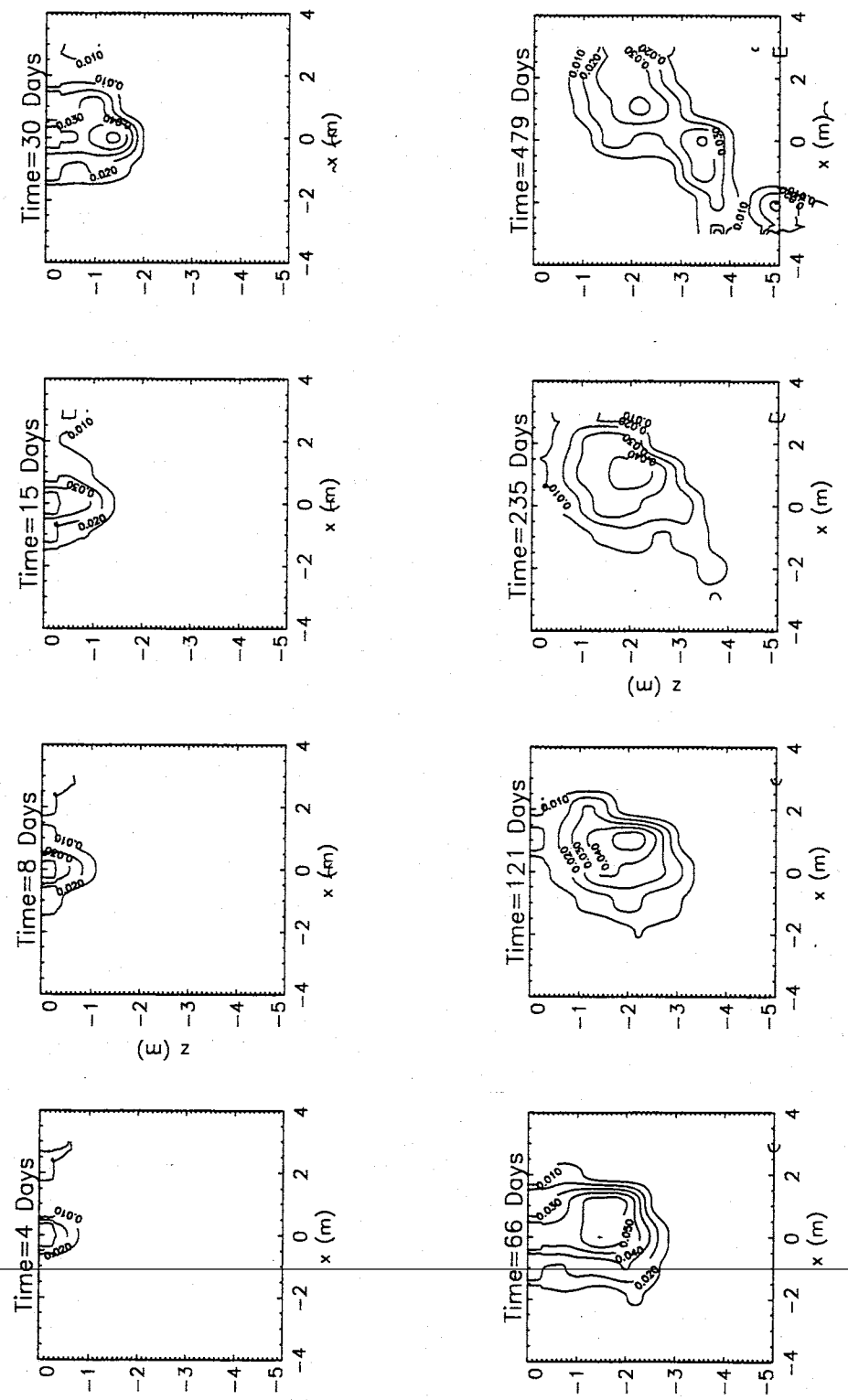


Figure 5.7. Observed Water Content Change Distributions (y=2.0 m).

type of subsurface variability affecting subsurface redistribution of water. Bromide and tritium distributions observed in a plane 0.5 m north of the wall and presented in Figure 5.8 shows somewhat less evidence of asymmetry. The researchers attributed asymmetric moisture and tracer distributions to the existence of lateral heterogeneities that were not accounted for by the spatially uniform model. Bromide appears to be moving a little faster than tritium, presumably due to anion exclusion effects. It should be emphasized that the distributions presented in the figures are only single realizations of the observed state variables. The moments of these realizations are later compared to the ensemble moments of predicted distributions simulated on the basis of single-scale and two-scale saturated permeability realizations using numerical flow and transport models.

A two-dimensional uniform model based on Richards' equation was used by Hills *et al.* (1991) to predict water movement during and after application of tracer and water on the north side of the trench. Again, water content was selected as the independent variable. Constant mixed boundary conditions, equal to 0.43 cm/d during the 75.5-day infiltration period and 0.0 cm/d following this period, were imposed at the surface boundary. A constant flux of 0.0 cm/d was specified at a depth sufficiently large to be unaffected by the advancing wetting front for times less than 300 days. In addition, zero-dispersive-flux conditions were imposed at horizontal boundaries located at sufficiently large distances from the experiment to be unaffected by lateral spreading of the water front before 300 days. Like the one-dimensional experiment, initial water contents were estimated on the basis of tension data using the soil retention model and

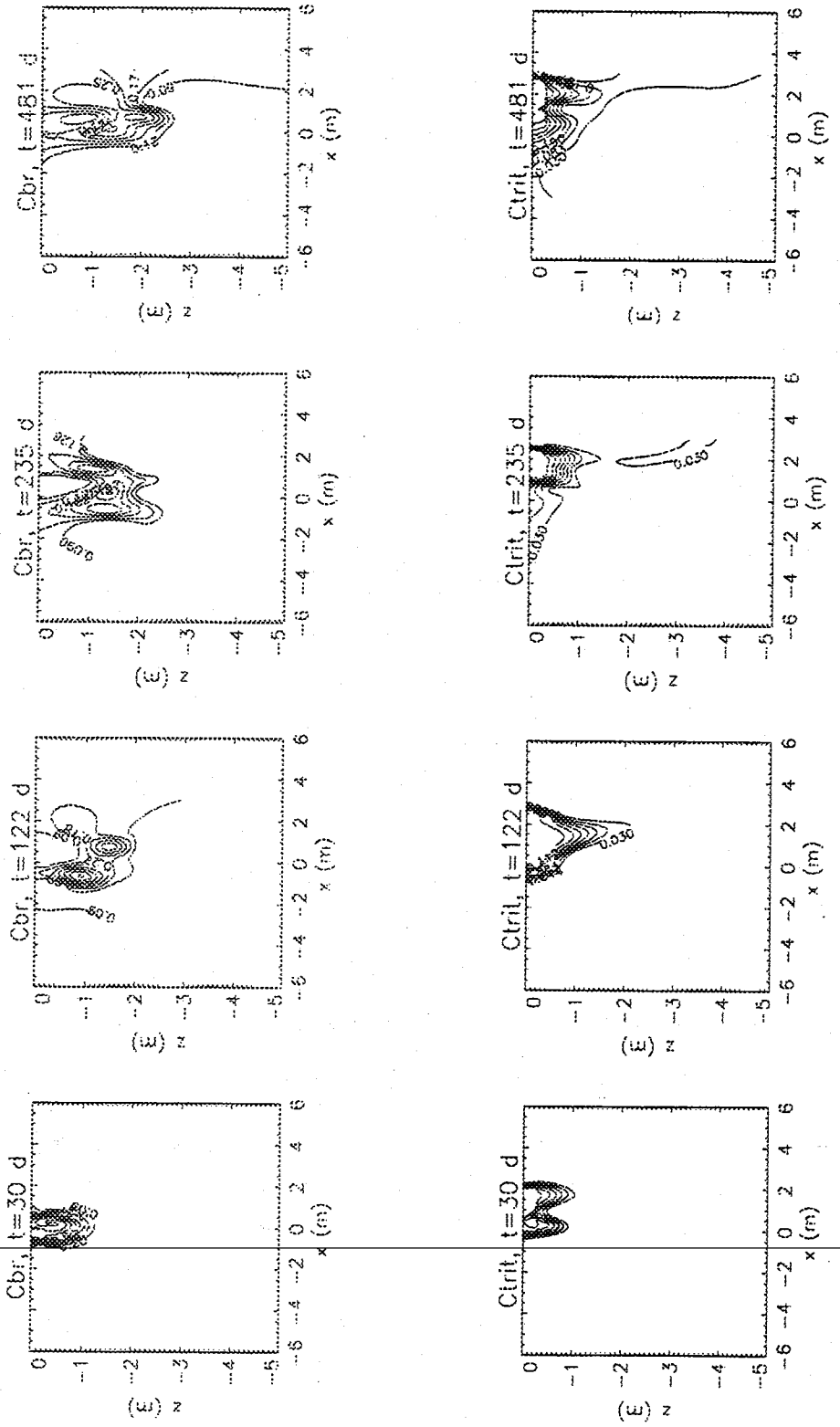


Figure 5.8. Observed Bromide and Tritium Concentration Distributions (y=0.5 m).

linear interpolation between tensiometer measurement depths. When tensions were too high to be measured, soil sample psychrometer data from access tube samples were used instead. Averages of these depth-dependent water contents were obtained using available tensiometer data collected from the same depth, resulting in a one-dimensional profile of initial water content. The two-dimensional uniform model predictions reproduced overall characteristics of water content change for up to 276 days following the onset of infiltration. However, the model failed to reproduce the observed late-time asymmetric moisture distributions shown in Figure 5.7.

Two-dimensional transport was also simulated for the two tracers, which were both assumed to be subject to adsorption. Initial concentrations of both tracers were specified as equal to 0.0, with constant fluxes of unit relative concentrations along the surface boundary and no-flux conditions on lateral and lower boundaries located sufficiently far from the tracer plumes to be unaffected by them. A small nodal spacing was selected so that sharp tracer fronts would not produce numerical instability problems. Simulation of transient tracer concentrations at each time step involved solving the Richards' equation for transient fluid flux, then solving the transient advective-dispersive equation for tracer concentrations. Predictions of tritium and bromide concentrations were in reasonably good agreement with field observations during the 75-day infiltration experiment and during a subsequent 200-day redistribution period.

5.3 JUSTIFICATION FOR MULTIPLE-SCALE MODEL OF HETEROGENEITY IN LAS CRUCES TRENCH

It is clear that, while use of one-dimensional and two-dimensional spatially uniform models by previous researchers reproduced overall characteristics of observed water content and tracer distributions, the models were not able to duplicate asymmetries observed during late stages of redistribution. The one-dimensional layered model, while offering the potential for incorporating large-scale deterministic structure, appeared to produce no significant improvement in predictive capabilities (Wierenga *et al.,* 1991). However, changes in spatial variability of observed state variables with respect to time suggest that a time dependent, or equivalently, a displacement-distance dependent source of variability, might be present in the subsurface.

5.3.1 Observed Variations in Spatial Structure Between Soil Horizons

Based on permeability testing of soil samples, as well as *in situ* permeameter tests, soils underlying the Las Cruces site appear to be characterized by structure that can be described stochastically as a weakly-stationary random process. Jacobson (1990) found that variograms constructed using all Las Cruces permeability data showed stationary behavior with correlation scales of 2.5 m and 0.5 m in the horizontal and vertical directions, respectively. While these scales may be good estimates of the global structure of heterogeneity underlying the site, visual observation of the soils suggest that the structure of heterogeneity may vary considerably between soil layers. A two-scale approach that accounts for deterministic layered structure and within-layer

random variation may offer advantages over the more traditional single-scale approach, for which the process is treated as globally uniform in its statistical properties. In addition, conditioning on large-scale deterministic structure based on easily-observable 'soft' lithologic information may help to reduce state-variable prediction uncertainty.

5.3.2 Large Vertical Variation in Water Content

Comparison of one-dimensional water content predictions made using both uniform and layered permeability suggest that there is smaller variation between model predictions than there is between measured water contents made in the vertical direction along the irrigation plot centerline (Wierenga *et al.*, 1991). Thus, the observed amount of vertical variation in water content cannot be fully be accounted for by introducing vertically layered structure. Another source of variation must be present. Assuming that spatial variations in saturated permeability dominate variations in all other parameters impacting subsurface water redistribution, the additional source of water content spatial variability might reside in smaller-scale, intralayer permeability variations.

5.4 TWO-DIMENSIONAL VALIDATION OF STATE VARIABLE SPATIAL MOMENTS

The degree to which validation of spatial moments based on a single realization of an observed state variable can be justified hinges on the ergodicity of the state

variable process. If the observed process is ergodic, or statistically typical, then geostatistical validation based on the single observed realization can be considered more justified than if it were not. Note that, since ergodicity is a hypothesis that can neither be proved nor disproved, the validity of comparing observed spatial moments estimated from a single realization to ensemble moments can never be fully assessed.

5.4.1 TRACR3D Numerical Flow and Transport Simulations

The integrated finite-difference flow and transport algorithm TRACR3D (Travis and Birdsell, 1991) was used to numerically simulate two-dimensional partially saturated fluid flow and tracer transport through two-scale realizations of saturated permeability similar to the realizations presented in Figure 5.3, as well as through equivalent single-scale realizations. Each layer in the two-scale process was assigned uniform deterministic values of porosity θ_s, irreducible water saturation θ_r, bulk density, and soil moisture parameters n and α as determined from laboratory testing and presented in Table 5.1. Only saturated permeability was allowed to vary in each layer according to the statistics provided in Figure 5.2. Since dispersivity of the conservative tracers was considered to be dominated by an advection-based mechanism of transport, a small local dispersion coefficient D* of $1.16X10^{-5}$ cm^2/s was specified to account for molecular diffusion of both tritium and bromide. This small value helped circumvent numerical problems typically associated with D*=0. Since the half-life associated with tritium was 12.5 years, radioactive decay was considered negligible over the period of the test. Equilibrium sorption coefficients (K_d) of 0.0 and

-0.023 m*l*/mg were assumed for tritium and bromide, respectively. The negatively-valued sorption coefficient for bromide reflects anionic exclusion effects.

Input suction heads were obtained by interpolating heads from tensiometer measurements collected prior to the onset of the experiment and presented in Hills *et al.* (1991). Initial bromide and tritium concentrations were specified as 0.0 mg/*l* and 0.0 mCi/*l* everywhere in the domain. Finally, no-flow boundary conditions were assigned to the bottom, left, and right boundaries with respect to both water and tracers. These conditions required that the boundaries be located sufficiently far from the migrating water and solute to not significantly influence their movement. Lateral boundaries were located 10 m from the irrigation plot centerline and the bottom boundary was located 4.6 m from the surface, for a total flow domain of 20 m in the horizontal and 4.6 m in the vertical. No-flow conditions were also imposed along the top boundary, except at node locations corresponding to the irrigation plot. At these nodes, a water flux of 0.43 cm/day was specified for 75.5 days and tracer fluxes of 0.1 mCi/*l* and 0.939 mg/cm^3 of tritium and bromide were introduced during the first 11.5 days of the simulation. Finally, it was assumed that the air pressure everywhere in the subsurface profile was equal to atmospheric pressure, with transient water flow governed by the Richards' equation.

The covariance structure associated with water content change and concentration state variables depended on the covariance structure of input saturated permeability. Hence, it was critical that the saturated permeability structure be preserved by discretizing the transport domain in such a way that there were as many

finite-difference cells as possible contained within each permeability correlation length. It was assumed that a minimum of 5 cells per correlation length was required to adequately preserve correlation structure, especially in the vertical dimension where most of the differences in model response were anticipated when comparing two-scale and single-scale models.

Using spectral techniques, the average ensemble correlation permeability scales for the two-scale permeability process were estimated as 0.9 m and 0.18 m in the horizontal and vertical directions, respectively. Equivalent single-scale simulations were performed using these average correlation scales, along with averages of the deterministic properties listed in Table 5.1 over all 7 layers sampled by the water and tracers. For the single-scale case, the finite-difference transport domain was discretized using dx=0.2 m and a constant dz of 0.046 m. These space increments produced grid spacings per correlation scale of 4.5 and 3.9 in the horizontal and vertical dimensions, respectively. While larger values would have been more desirable, particularly in the vertical direction, limitations of machine memory available at the time of simulation made finer discretization impractical.

The two-scale model was discretized somewhat differently than the equivalent single-scale model. Since the horizontal discretization had to be valid for all layers, dx was chosen to be 0.2 m (20 cm). This value was sufficiently small to allow at least 2.5 and up to 15 grid spacings per horizontal correlation scale and also had the added benefit of representing as many as 6 or 7 correlation lengths over the 20 m horizontal length of the flow and transport domain. Discretization in the vertical dimension varied

along the profile to honor the rule of 5 cells per correlation length. Figure 5.9 shows how the 20m-by-4.6m domain was discretized in the vertical dimension for the two-scale model.

Both single-scale and two-scale models were allowed to run without infiltration for a simulation time of $1.0X10^{+7}$ seconds (115.7 days) in order to obtain a realistic initial head distribution for the transient flow prediction. Following attainment of steady state, the simulations were initiated by infiltrating water for a period of 75.5 days, with tracers dissolved in the infiltrating water only during the first 11.5 days. To maintain accurate mass balance, the time step was allowed to vary from 0.0 to $1.0 X10^{7}$ s. An accurate mass balance was considered to be one that maintained a flux residual of $1.0X10^{-8}$ cm^3/s or smaller.

5.4.2 Comparison of Observed and Simulated Ensemble Variogram Behavior

5.4.2.1 Observed and Simulated Ensemble Water Content Change Variograms

Simulated water content change distributions for all 20 two-scale and single-scale statistically equivalent saturated permeability processes are presented in Appendix D. Variograms for observed water content change perturbations about the spatially-variable ensemble mean water content in both horizontal and vertical directions were estimated using equation [1.22] and are presented in Figure 5.10. Ensemble mean water content change at each observation point was determined from the two-scale ensemble of 20 TRACR3D water content simulations. No sill was evident

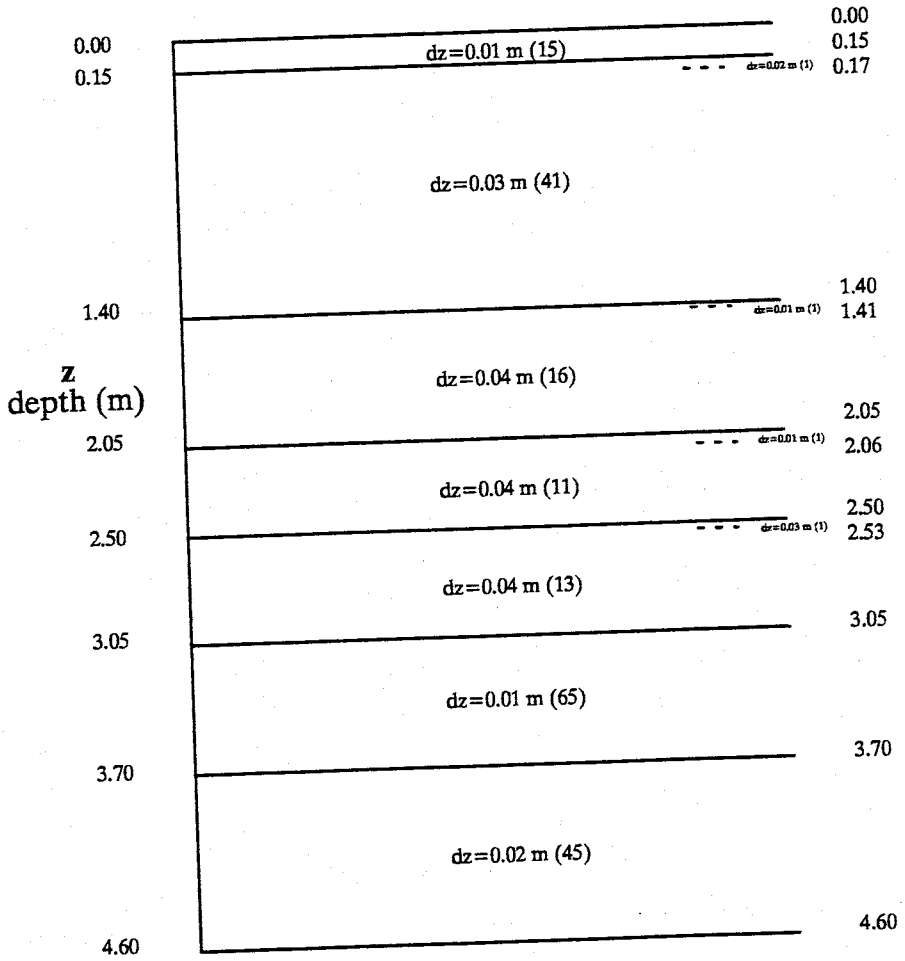


Figure 5.9. Vertical Spatial Discretization of Las Cruces Transport Domain for TRACR3D Simulation (number in parenthesis is number of cells).

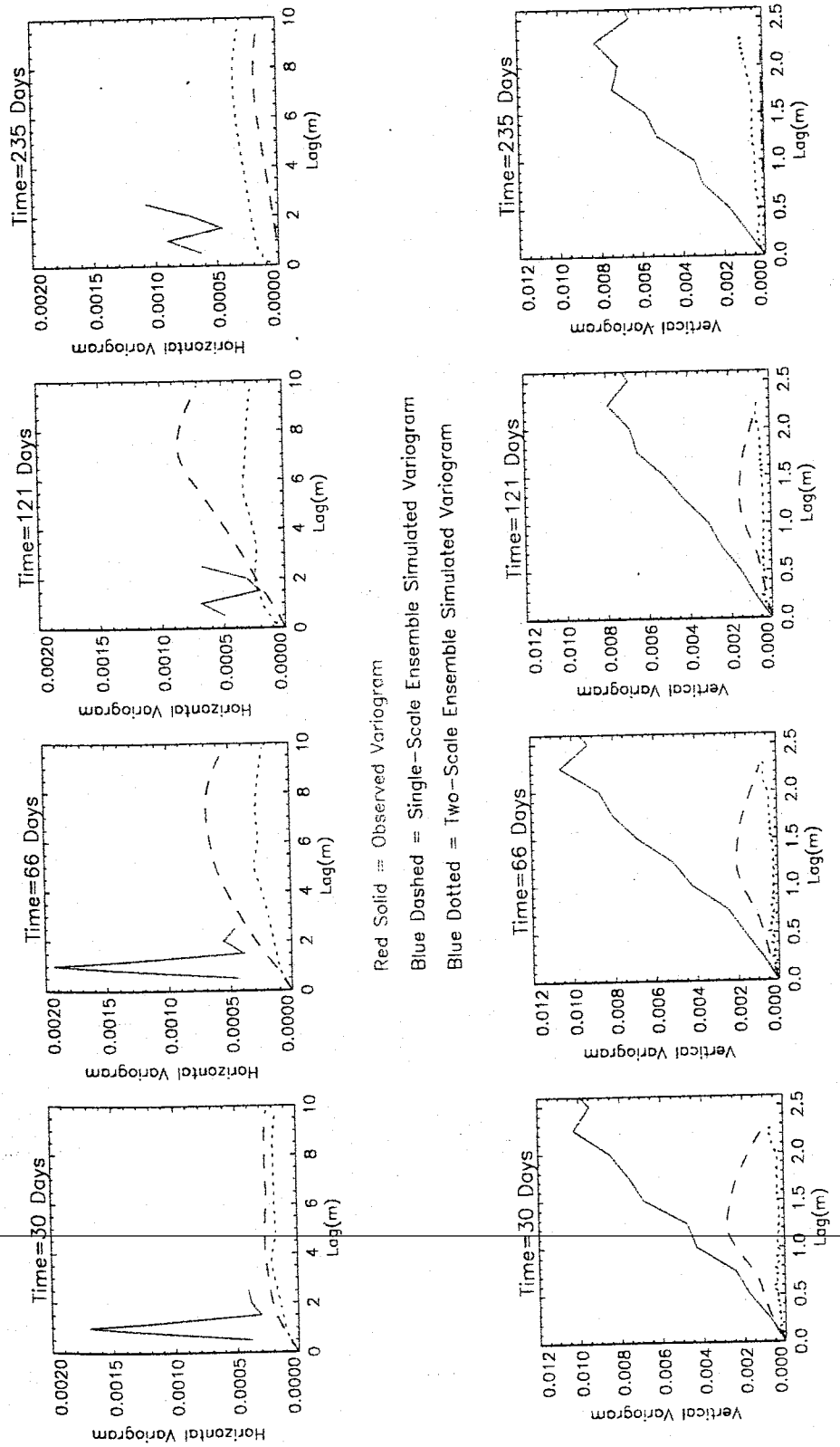


Figure 5.10. Variograms for Observed and Simulated Water Content Change Perturbations About Simulated Ensemble Mean.

in the observed horizontal water content change variogram. The observed vertical variogram showed evidence of a sill corresponding to roughly 2.0 m, larger than the sill of 0.65 m associated with layering. Lack of small-scale, intralayer structure in observed water content was likely caused by the dampening effect of the flow equation, which acts as a low-pass filter to screen out high-frequency small-scale variability. This buffering effect might be especially pronounced in the extremely desiccated soils underlying the site.

Figure 5.10 also shows ensemble variograms of simulated water content change perturbations about the ensemble mean from 20 TRACR3D simulations of water content for both statistically-equivalent single-scale and two-scale saturated permeability realizations. In both cases, the ranges of the sills tended to increase with time along both directions, as would be anticipated for state-variable processes affecting larger portions of the transport domain with increasing time. In the two-scale case, a small-scale intralayer sill and a sill corresponding to the layered structure are evident in the vertical variograms, with the ranges increasing slightly over time. These ranges tended to be larger than the ranges corresponding to intralayer and layer variability for the two-scale saturated permeability process. Such behavior appears to be caused by the smoothing effect of the Richards' equation, which tended to screen out high-frequency components of variation.

This damping effect is further substantiated by the fact that simulated ensemble vertical variograms are more bell-shaped than the observed vertical variograms near lag 0, having continuously differentiable behavior at the origin suggestive of a spatially

smooth process. The observed water-content variograms, along with the log conductivity variograms shown in Figure 5.5, exhibit a more exponential type covariance behavior typical of less connected, more anatomized random structure.

It is evident that, while the single-scale ensemble variograms more closely reproduced the variance of the observed process at early experimental time than two-scale ensemble variograms in both the horizontal and vertical directions, the converse was true at late experimental time. This outcome is related to the fact that layered structure is not completely sampled until later stages of infiltration. Therefore, the full variance of subsurface permeability structure was not brought into the problem until late time, when water began to encounter layers with very different statistical properties.

From inspection of the water content realizations, it appears that layer 6, located at a depth of approximately 2 m and characterized by a larger mean ln k_s as well a significantly larger ln k_s variance than any of the overlying three soil layers, began to introduce a large variance to the simulated two-scale process after roughly 30 days. Prior to that time, the two-scale process was largely influenced by the top three layers, which are all characterized by variances smaller than the equivalent variance associated with the single-scale ln k_s process, as well as small, similar means. As a result, the simulated two-scale water content process shows limited variance compared to the single-scale process, which is influenced by a larger uniform mean and variance. After a time of 30 days, the abrupt change in mean ln k_s and the larger ln k_s variance encountered by water at the top of layer 6 produces a slow increase in two-scale

variance. After approximately 180 days, the two-scale variance begins to exceed the single-scale variance as the two-scale model starts to reflect changes in the large-scale mean induced by layering. By a time of 235 days, the additional variability caused by changes in the mean as water encounters different layers manifests as a much greater total variance in water content.

Note that, while the two-scale model describes more of the total variance of the observed saturation process, neither model explains all of it. Clearly, other sources of variation remain unaccounted for. These sources might include random variations in other aquifer properties, in recharge or boundary conditions, or in other model specifications. Permeability variations may also be larger than those estimated on the basis of available trench wall permeability measurements.

5.4.2.2 Observed and Simulated Ensemble Bromide Concentration Variograms

Simulated bromide concentration distributions for all 20 single-scale and 20 two-scale saturated permeability realizations are presented in Appendix D. Observed and simulated ensemble variograms for bromide concentration perturbations about ensemble mean predicted concentration for the single-scale and two-scale cases are shown plotted in Figure 5.11. The erratic nature of the observed variograms along both directions is attributed to the poor resolution of the bromide data, which was collected at 0.25-m intervals in the vertical direction and 0.5-m intervals in the horizontal direction, compared to water content observations made at 0.05-m intervals. Like the water content variograms, there was more pronounced bell-shaped behavior

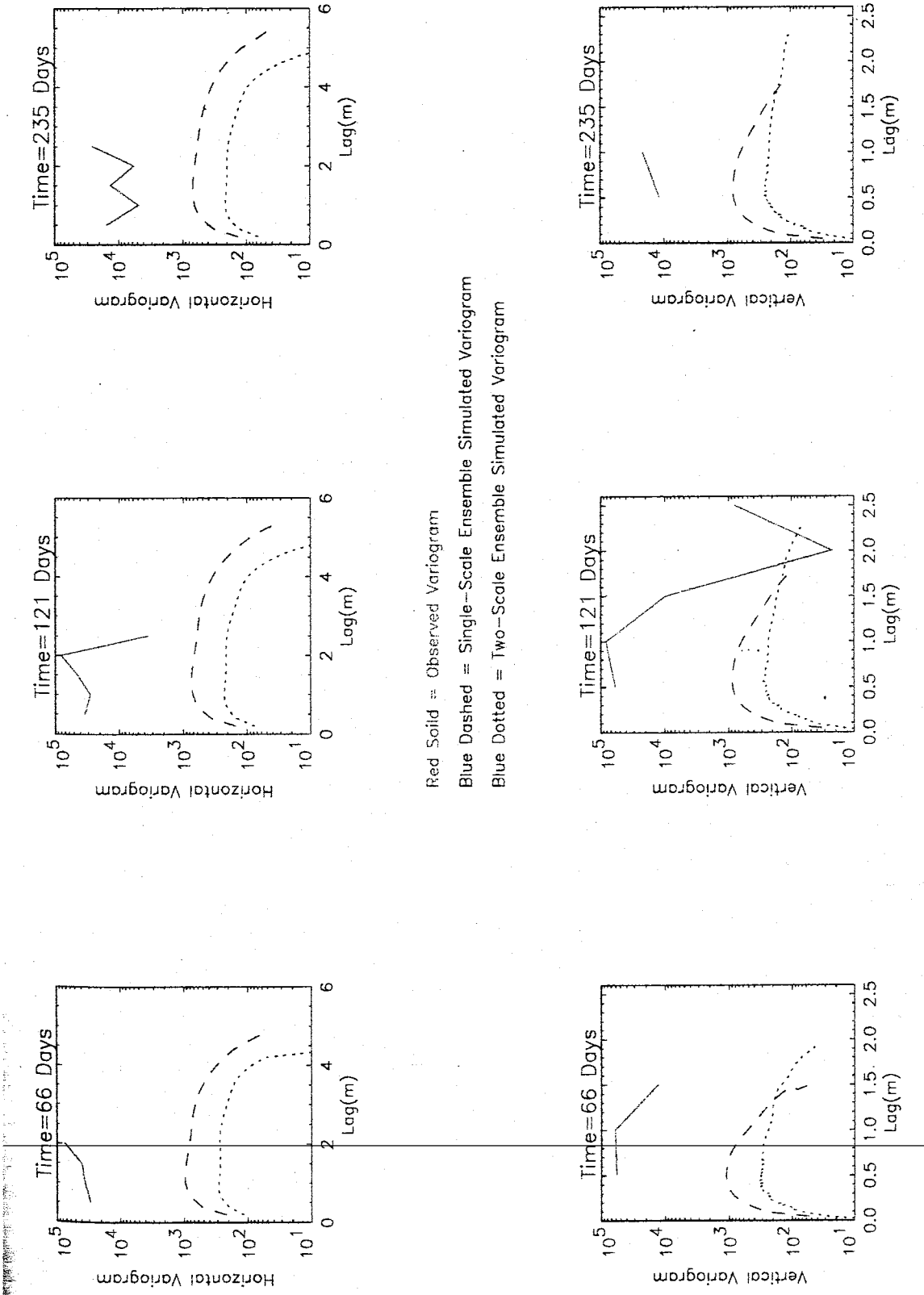


Figure 5.11. Variograms for Observed and Simulated Bromide Concentration Perturbations About Simulated Ensemble Mean.

associated with the simulated variograms. For the case of bromide concentration, however, the variance of the single-scale process more closely reproduced the observed variance than the two-scale process at all times. This outcome was related to the limited degree of layer sampling by tracer plumes prior to the time of final concentration measurement. From inspection of the observed plumes in Figure 5.8, it appears that by the end of the experiment, neither plume had even begun to encounter layer 6, where much larger variability in saturated permeability resided compared to variability in overlying layers. Again, neither model fully accounted for all observed spatial variation in the process. Similar results were demonstrated for the case of tritium.

While observed water content and tracer concentration realizations showed evidence of asymmetry at late time, there was no such asymmetry in predicted state-variable realizations. Lack of asymmetry in the simulated water content and tracer concentration distributions at large time for the two-scale case was attributed to the fact that displacement distances for both processes greatly exceeded the intralayer scales of permeability variation in both x- and z-directions. Under the central limit theorem, the large number of independent permeability and velocity samples gave rise to a Fickian, or diffusive, process that was characterized by spatial symmetry and gaussian behavior.

Results of the geostatistical validation suggest that the two-scale model is better able to reproduce state-variable variances as long as the state variables are observed at times when they have sampled the full range of two-scale variability. Prior to that

time, the single-scale model may actually provide a better prediction of state variable variability.

The total variance associated with the two-scale permeability model can be viewed as a sum of two variances — one, a small-scale variance caused by local, random fluctuations of the process Z about a constant mean within each layer (m_l), and the second caused by fluctuations of the layer mean (m_l) about the global mean of the process (m):

$$\sigma_{TS}^2 = \text{var}\,(m_l - m) + \text{var}\,(Z - m_l) \qquad [5.3]$$

For the 20 two-scale saturated permeability realizations, the ensemble average of var($m_l - m$) was equal to 1171 darcy^2, or 52% of the total variance of 2260 darcy^2, suggesting that if all large-scale changes in the mean are accounted for, unexplained variance in saturated permeability can be halved.

5.4.3 Macrodispersive Flux

Of interest is the question of how inclusion of deterministic layered structure at the larger scale impacts our ability to predict concentration at the Las Cruces site. Ensemble macrodispersive flux represents spreading that occurs at a scale too large to be truly Fickian but too small to be explained advectively using large-scale, mean velocity. As demonstrated by equation [1.11], macrodispersive flux is equal to total mass flux minus advective flux associated with mean flow and concentration and cannot

be assigned either a diffusive or an advective cause. As a result, it becomes manifested as an extra term in the transport equation, separate and distinct from the first-order advection term and the second-order diffusive term.

From equation [1.7], it can be seen that macrodispersive flux is equivalent to the velocity-concentration cross covariance, with larger macrodispersive fluxes corresponding to higher degrees of correlation between velocity and concentration perturbations. For large variations in velocity about mean velocity, larger fluctuations in concentrations about mean concentration will occur, accompanied by greater transport uncertainty. Consequently, larger fluxes are indicative of larger ensemble prediction uncertainty than smaller fluxes. It should be emphasized that macrodispersive flux is an ensemble, and not a spatial, property.

Using 20 realizations of transient bromide concentration and flow velocity for both the single-scale and two-scale Las Cruces permeability models, time-dependent spatially-distributed macrodispersive fluxes were calculated in both longitudinal (vertical) and transverse (horizontal) directions. Flux was determined at each location and in each realization by finding the velocity-concentration cross-covariance centered about the ensemble means and finding the expectation at each point in space and time across the ensemble. As shown in Figure 5.12, ensemble maximum absolute macrodispersive flux for the two-scale permeability model was smaller than flux predicted using a single-scale approach to characterizing permeability, except at very early time, before concentration had been significantly affected by the onset of layered structure at z=0.15 m. These results demonstrate that large-scale conditioning on

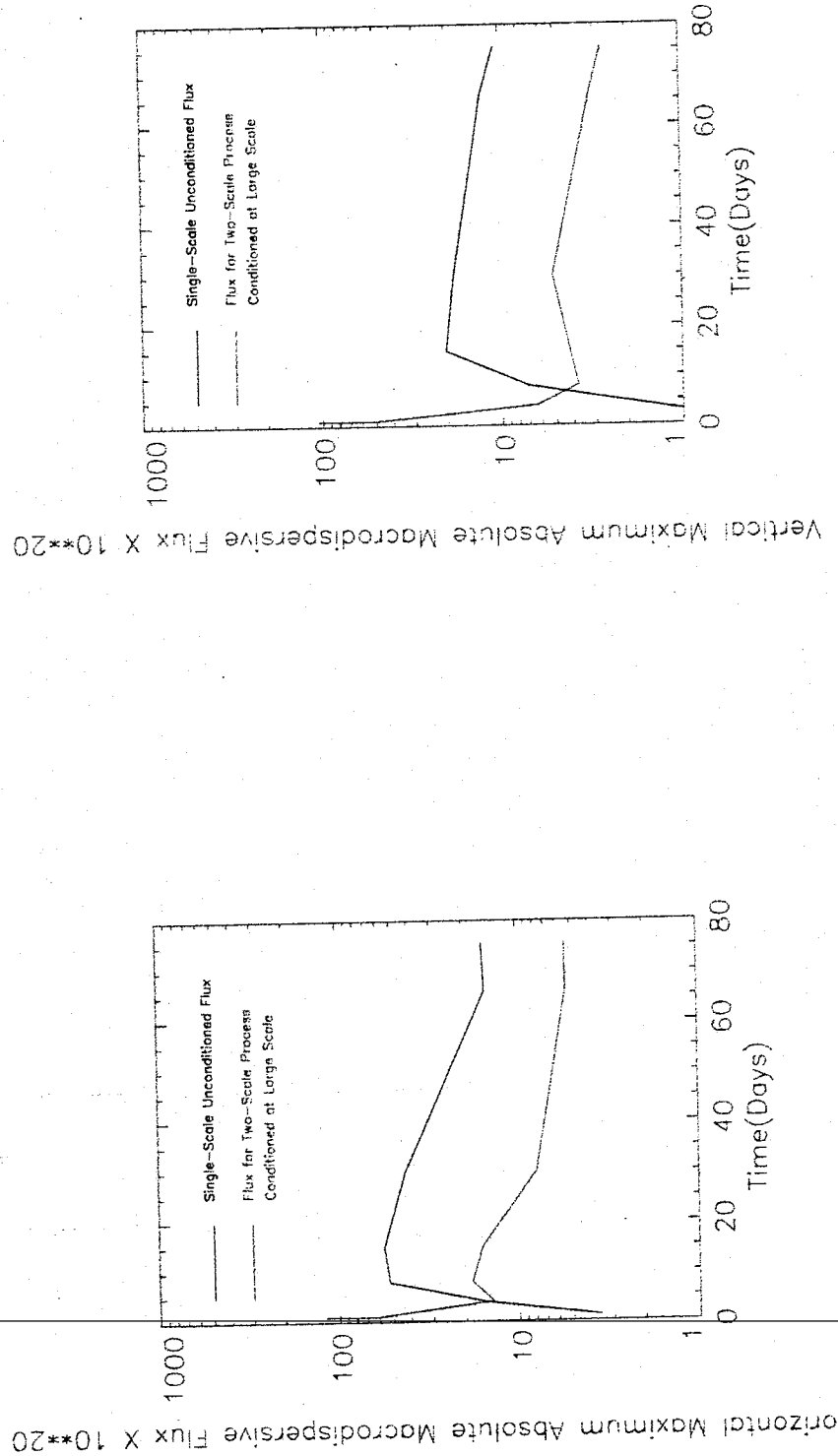


Figure 5.12. Maximum Absolute Ensemble Macrodispersive Flux for Two-Scale Permeability Process Conditioned at Large Scale and for Unconditioned Single-Scale Permeability Process.

deterministic information generally reduces the uncertainty of predicting tracer transport because it explains more of the spatial variation in velocity and allows us to better predict advective field-scale tracer movement. Therefore, access to easily observed deterministic 'soft' permeability information at the large scale can significantly improve our ability to predict contaminant migration when the dominant mode of transport is non-Fickian.

5.5 SUMMARY

Water content and tracer concentration observations collected during an infiltration/tracer test conducted adjacent to the Las Cruces trench were used as a basis for geostatistical validation of a two-scale saturated permeability model. Previous research suggested fact that differences between predicted one-dimensional uniform and layered permeability water contents were smaller than the variability of water content observations, indicating that there might be a source of permeability uncertainty in addition to layering. This source of uncertainty was assumed to be smaller-scale intralayer variations in permeability.

Simulated ensemble variograms for a two-scale permeability model characterized by large-scale layering and small-scale intralayer variability reproduced observed water content change variances better at late experimental time than a statistically-equivalent single-scale permeability model. Results of the geostatistical validation suggest that the two-scale model is better able to reproduce state-variable variances as long as the state variables are observed at times when they have sampled

the full range of two-scale variability. Prior to that time, the single-scale model may actually provide a better prediction of state variable variability.

When large-scale layered permeability structure in the two-scale permeability model was conditioned on easily-observed 'soft' information, the two-scale model produced smaller macrodispersive fluxes than the single-scale model, demonstrating that conditioning of large-scale components of permeability on observed geologic structure can significantly reduce concentration prediction uncertainty.

5.6 REFERENCES

Gee, G., and T.J. Nicholson, INTRAVAL Project Status Report, U.S. Nuclear Regulatory Commission, Office of Nuclear Regulatory Research, 1987.

Hills, R.G., I. Porro, D.B. Hudson, and P.J. Wierenga, Modeling one-dimensional infiltration into very dry soils, 1. Model development and evaluation, *Water Resour. Res.*, 25(6), 1259-1269, 1989a.

Hills, R.G., D.B. Hudson, I. Porro, and P.J. Wierenga, Modeling one-dimensional infiltration into very dry soils, 2. Estimation of the soil water parameters and model predictions, *Water Resour. Res.*, 25(6), 1271-1282, 1989b.

Hills, R.G., P.J. Wierenga, D.B. Hudson, and M.R. Kirkland, The second Las Cruces trench experiment: Experimental results and two-dimensional flow predictions, *Water Resour. Res.*, 27(10), 2707-2718, 1991.

Jacobson, E.A., *Investigation of the spatial correlation structure of saturated hydraulic conductivities from a vertical wall of a trench*, paper presented at the Canadian/American Conference on Hydrogeology — Parameter Identification and Estimation for aquifer and Reservoir Characterization, Alta. Res. Council, Ass. Groundwater Sci. and Eng., Calgary. Alta, Canada, Sept. 18-20, 1990.

Rao, P.S.C., R.E. Jessup, A.C. Hornsby, D.K. Cassel, and W.A. Pollans, Scaling soil microhydrologic properties of Lakeland and Konowa soils using similar media concepts, *Agric. Water Manage.*, 6, 277-290, 1983.

Travis, B.J., and K.H. Birdsell, *TRACR3D: A Model of Flow and Transport in Porous Media: Model Description and User's Manual*, Los Alamos Manual LA-11798-M, 1991.

van Genucten, M.T., An Hermitian finite element solution of the two-dimensional saturated-unsaturated flow equation, *Adv. Water Resour.*, 6, 106-111, 1983.

Warrick, A.W., and D.R. Nielsen, Spatial variability of soil physical properties in the field, in *Applications of Soil Physics*, D. Hillel (ed.), 319-344, Academic, Orlando, 1980.

Wierenga, P.J., R.G. Hills, and D.B. Hudson, The Las Cruces Trench Site: Characterization, experimental results, and one-dimensional flow predictions, *Water Resour. Res.*, 27(10), 2695-2705, 1991.

Chapter 6. Conclusions and Future Research Directions

6.1 CONCLUSIONS

Movement of ground water is not strongly determined by the presence of local variations in hydraulic conductivity. Transport of dissolved contaminants in subsurface geologic media, however, is sensitive to these local variations. Consequently, reliable prediction of contaminant movement in the subsurface hinges on how well we can characterize spatial variations in conductivity. At large plume displacement distance, the spreading of contaminant asymptotically approaches Fickian behavior if a large number of independent velocities are sampled by the plume. Under such conditions, the central limit theorem applies and the plume assumes a gaussian shape characteristic of a diffusive process. Much research has been invested in defining the effective, macroscopic diffusion coefficients associated with such a global Fickian process using Lagrangian (Dagan, 1984; 1987; 1988) and Eulerian (Gelhar and Axness, 1983) transport theories. These theories have application when conductivity is statistically uniform over space and characterized by finite variance and integral scale. When this is the case, effective diffusion coefficients known as macrodispersivities can be estimated using summary measures of log conductivity statistical behavior.

However, in the presence of nested or evolving scales of spatial variation, nonstationarities are present in the conductivity random process. The effect of these nonstationarities is that velocities are sampled independently from a probability distribution that varies with displacement. The resulting rate of contaminant spreading will, in general, not be characterized by Fickian behavior and classic diffusive transport theories will no longer apply. In such hierarchical geologic media, the statistical properties associated with log conductivity change over space, global macroscopic transport behavior does not mimic local microscopic diffusion, and an effective Fickian diffusion coefficient will not exist. Rather, a non-local non-Fickian transport regime will prevail.

The non-local nature of transport in the presence of nested or evolving scales of variation renders the traditional absolutist REV measurement approach essentially meaningless. It requires that we view the spatial distribution of aquifer properties through scale-variant instrument windows paralleling natural scales of variation that dominate the transport regime. Such a relativist measurement philosophy acknowledges the contribution of instrument support to measurement outcome. It also avoids the problem of upscaling, which can alter effective properties in a multiple-scale geologic medium depending on the upscaling method used. The relativist measurement approach establishes a formal basis for incorporating scale-variant randomness into a multiple-scale conductivity measurement network design model.

Chapter 6. Conclusions and Future Research Directions

The multiple-scale conductivity measurement design model built during the course of this dissertation research treats the transport prediction problem as separate and distinct problems at each scale of hydraulic conductivity variation. This implicit functional hierarchy parallels the explicit hierarchy associated with the natural geologic structure. Scale-variant transport predictions were used to generate scale-variant adjoint solutions relating spatial variations in conductivity to concentration uncertainty at two scales of resolution. These scale-dependent adjoint solutions systematized a sequential search for conductivity measurement location and scale that most reduced contaminant prediction uncertainty at a regulatory compliance point. When the multiple-scale design was conditioned sequentially on conductivity measurements and constrained by a number of scale-variant conductivity realizations, adverse effects of conductivity uncertainty on identification of an optimal measurement design strategy were reduced.

Results of sequential two-scale conductivity measurement design for a number of different scenarios demonstrated that preference for measurement at a given scale depended on how much log conductivity variance resided at that scale, source loading, size of source relative to scales of conductivity variation, distance to source, and principal directions of anisotropy relative to principal directions of transport — in short, on the geometry of the transport problem.

The validity of a two-scale approach to characterizing aquifer properties was investigated using Las Cruces trench experiment data. Saturated permeability was characterized as both a two-scale random process with large-scale horizontally layered

deterministic structure and small-scale embedded gaussian random structure, as well as a second-order statistically equivalent single-scale gaussian process. Vertical-profile water content change distributions simulated on the basis of the two-scale saturated permeability process better reproduced the variance of distributions observed during late stages of the experiment. This was attributed to the fact that the two-scale conductivity process accounted for variance associated with abrupt changes in the mean of saturated permeability due to horizontal layering, in addition to variance related to small-scale intralayer structure.

When large-scale horizontal layering was treated as deterministic structure and two-scale saturated permeability conditioned on this deterministic component, macrodispersive flux was reduced relative to that associated with transport through the equivalent single-scale saturated conductivity process. These results demonstrated that removal of saturated permeability uncertainty at the large scale extracted low-frequency, predictable trends in velocity, explaining more of the transport mechanism as a permeability-based advection process. Residual uncertainty following large-scale conditioning occurred as a result of velocity variations present at scales too small to be explained by large-scale horizontal layering, but too large to be attributed to small-scale intralayer gaussian variability and explained as a local Fickian process.

6.2 FUTURE RESEARCH DIRECTIONS

Principal avenues of future study related to this dissertation research involve incorporating log conductivity parameter uncertainty at all scales of spatial variation

into the measurement design, formal investigation of the degree of design robustness with respect to number of conditional log conductivity constraints, and consideration of model error.

It is not difficult to conceive of situations for which uncertainty in geostatistical parameters of **Y** contributes more to concentration prediction uncertainty than does uncertainty in **Y** itself. Under such circumstances, a Y measurement may contribute more information to the design in its capacity to reduce parameter uncertainty than with respect to its ability to reduce uncertainty in the conditional process, $\mathbf{Y_c}$. When this is the case, neglect of geostatistical parameter uncertainty in the design can produce conductivity measurement strategies that are grossly suboptimal.

Uncertainty in the Markov parameters can be critical when the plume exists at a spatial scale of the same order or larger than the scale of large-scale heterogeneities. For the designs listed in Table 4.3, the same Markov parameters used to synthesize the true realizations were used to generate the 10 large-scale **Y** realizations for the sequential design. In effect, this assumes perfect knowledge of the true MRF parameters, which will rarely be the case. Instead, the MRF parameters should themselves be treated as random variables, and Bayesian techniques used to condition the MRF statistics on **Y** observations.

A substantial amount of research has been conducted to minimize statistical parameter uncertainty using all available conductivity data. Kitanidis and Vomvoris (1983) and Wagner and Gorelick (1989) utilized a maximum likelihood approach to fit geostatistical parameters to the data, assuming that all data are jointly gaussian

distributed. For the design presented in this dissertation research, the ensemble of large-scale **Y** realizations used to constrain the design was conditioned on both current large-scale measurements and on the MRF parameter vector α.

The ensemble expectation of **Y** realizations can be written as $E(\mathbf{Y}|\mathbf{Y_c}, \alpha)$. We are interested in accounting for MRF parameter vector uncertainty and removing α from the ensemble mean estimate to obtain $E(\mathbf{Y}|\mathbf{Y_c})$. The two expectations are related as follows:

$$E(\mathbf{Y}|\mathbf{Y_c}) = \underset{a}{E}\{E[\mathbf{Y}|\alpha,\mathbf{Y_c}]\} = \int_{a} E[\mathbf{Y}|\alpha,\mathbf{Y_c}]p''(\alpha)d\alpha \qquad [6.1]$$

where $p''(\alpha)$ is the current, or posterior, estimate of the Markov parameter vector pdf. Equation [6.1] explicitly recognizes that there is parameter uncertainty and averages this uncertainty over the set of all possible parameter vectors α by taking the expectation over the entire parameter space. It accounts for randomness not only inherent to **Y**, but also to parameter vector α. Since the integration requires knowledge of the prior pdf of α, $p'(\alpha)$, the goal is to find a good prior estimate, based on currently available **Y** information.

To obtain an updated, or posterior, parameter pdf $p''(\alpha)$, the prior pdf is conditioned on the currently available large-scale **Y** data using the Bayesian relation:

$$p''(\alpha) = \frac{p(Y_c|\alpha)\ p'(\alpha)}{p(Y_c)} \qquad [6.2]$$

where $p(Y_c) = \int_{\alpha} p(Y_c|\alpha)\ p'(\alpha)\ d\alpha$

and $p'(\alpha)$ denotes the prior parameter probability density function (pdf), estimated before conditioning is performed on the next value of $\mathbf{Y_c}$, and $p''(\alpha)$ is the posterior pdf, or the distribution most recently informed on the data. $p(\mathbf{Y_c}|\alpha)$ is the probability of the current **Y** observations given the current MRF parameter vector and is used to condition the prior distribution $p'(\alpha)$ on currently observed large-scale **Y** data. The resulting posterior pdf then becomes the most current estimate of the probability of any MRF parameter vector. Since an initial estimate of $p'(\alpha)$ is generally not available, it can be assumed to take on the characteristics of a diffuse, or uniform, pdf.

To work the above integration into a practical design, it would be necessary to limit the analysis to a finite number of possible parameter vectors α, generate a finite set of α vectors directly off the posterior pdf, and synthesize one realization for each vector to produce a stack of realizations as shown in Figure 6.1. By weighting the state-variable response of each realization in the stack by the normalized posterior probability $p''(\alpha)$ of its associated parameter vector, a new realization is defined as a weighted sum of realizations in the stack. That realization is then used to condition the

posterior pdf of α, without recourse to stacks of subrealizations. In this case, weighting and normalization are not performed because only one sample off the joint distribution is used to construct each large-scale MRF realization.

Figure 6.1 demonstrates that, in effect, incorporation of MRF parameter uncertainty into the design adds another level of uncertainty to the design problem.

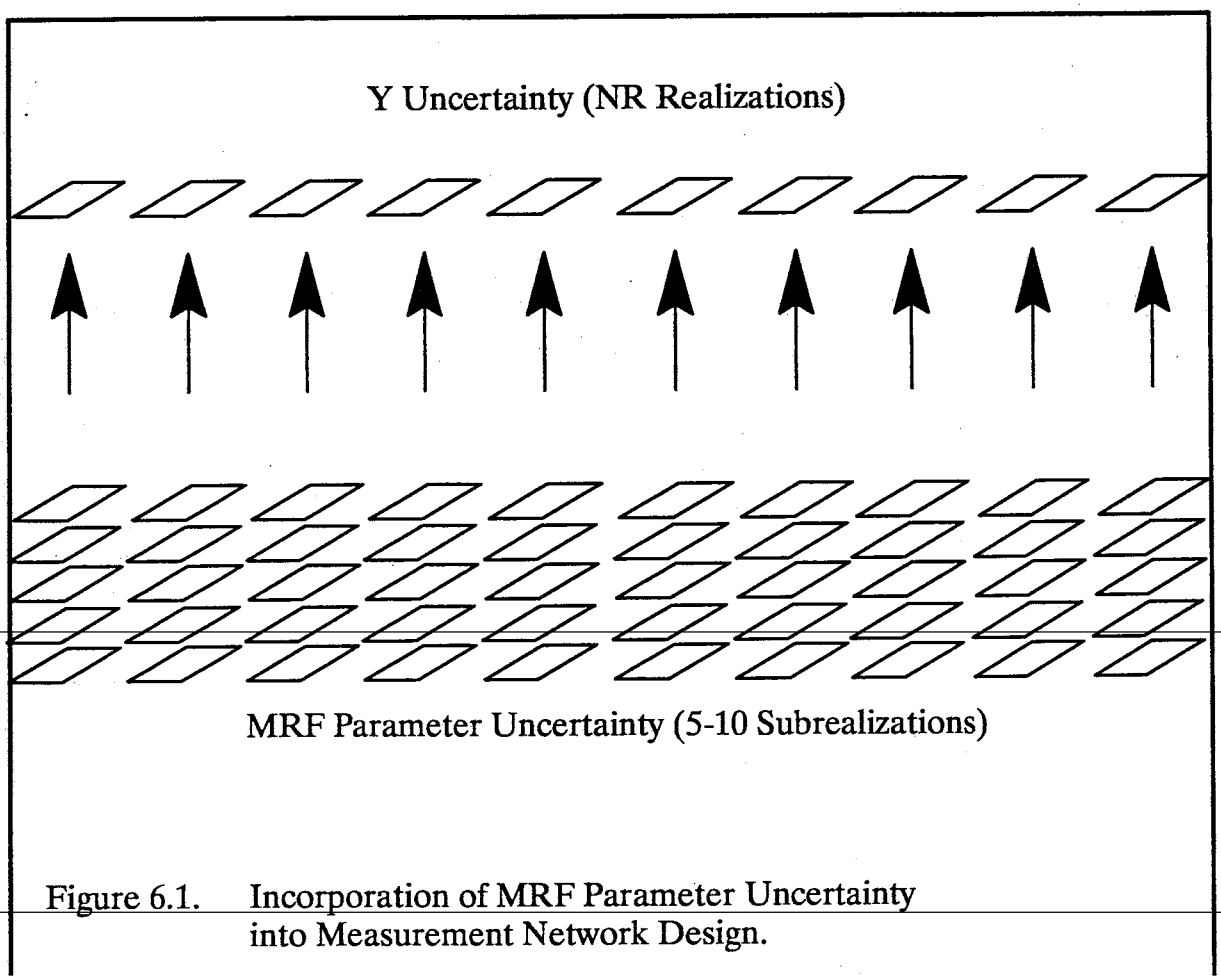


Figure 6.1. Incorporation of MRF Parameter Uncertainty into Measurement Network Design.

The first level is related to uncertainty in **Y**, while the second level is associated with uncertainty in the large-scale MRF stochastic parameters. When this level of uncertainty is included in the design and problem 2 re-solved, it is expected that measurement will, in general, linger longer at the larger-scale to account for uncertainty in the large-scale parameters as well as uncertainty in **Y** itself.

The issue of the robustness, or stability, of the measurement design should also be investigated. A robust design will tend to be insensitive to the addition or subtraction of any of the NR realizations used to constrain the design. Statistical sampling error in **Y** associated with an inadequate number of conditional realizations is an important issue, because it can yield extremely suboptimal measurement design.

The optimal value of NR can, in general, only be determined *a posteriori* by assessing whether there is a significant change in prediction variance with the addition of another **Y** realization. Previous studies involving stacking of multiple realizations (Wagner and Gorelick, 1989) have depended on Monte Carlo simulations to verify the reliability of the design *a posteriori* to the optimization. Robustness of a variance reduction design by Cleveland and Yeh (1990) was evaluated using Monte Carlo methods by varying parameters about their optimal values until the probability of the design varied significantly from 1.

Sampling error is particularly a problem when the covariance of the joint $\mathbf{Y_c}$ distribution is large. For high-variance cases, a larger number of realizations will be required to achieve a certain degree of design reliability. Under such circumstances, use of 10 or even hundreds of realizations to constrain each iteration of the sequential

measurement design might not be expected to account for all spatial variability at both scales of variation and produce an optimal network maximizing concentration information at the regulatory compliance point. In high-variance cases, one may need recourse to computationally parallel machines to perform the large number of flow, transport, and adjoint simulations required to produce a reliable design algorithm. Clearly, the impact of $\mathbf{Y_c}$ uncertainty on design robustness warrants further investigation within a formal analytic framework.

Clearly, a larger value of NR allows more uncertainty in $\mathbf{Y_c}$ to be sampled, guaranteeing a more robust design. The question is : what value of NR is sufficient to guarantee robustness yet not so large as to require an excessive computational burden? Too few realizations will not sufficiently sample reality, but too many realizations would contain information redundancy with each additional realization producing smaller information returns. Chan (1993) attempted to develop rules by which the *a priori* selection of NR realizations could be determined based on the desired level of design reliability. He came to the rather surprising conclusion that the level of reliability appears to depend primarily on NR, and does not seem sensitive to the specific physical or statistical properties of the problem. This suggests that, despite the fact that damping of variance imposed via the simulation functions $\mathbf{g}(\bar{\sigma},\mathbf{Y_c})$ is problem dependent, fairly consistent rules appear to exist for defining a minimum number of realizations NR required to attain a certain level of design accuracy. The issue of design robustness as a function of NR is an important one that warrants further investigation.

As a final consideration, the presence of model error can lead to seriously suboptimal measurement design. Designs presented in Chapter 4 for a hypothetical aquifer are assumed unaffected by model error. In real applications of multiple-scale design, however, errors in specifying appropriate boundary conditions, initial conditions, source location and magnitude, nature of chemical and physical processes, and geostatistical parameters relating to spatial variability of parameters all conspire to yield measurement designs that, while near-optimal for the assumed model, can be grossly suboptimal in the context of the true transport model. Knopman and Voss (1988; 1989) have investigated the problem of discriminating among many possible flow and transport models in the context of sampling design. Optimal measurement locations will tend to be those for which model predictions at selected points vary the most from model to model, because it is at these locations where the most information about the model resides and discrimination power is optimized. Design for purposes of model discrimination may conflict with measurement designs based on minimization of parameter estimation variance, particularly in the presence of parameter-robust designs that perform well over a range of parameters but may not be model sensitive and able to discriminate among possible physical models. Incorporation of a model discrimination component to the multiple-scale measurement network design would be a fairly straightforward extension of this dissertation research.

6.3 REFERENCES

Chan, N., Robustness of the multiple realiozation method for stochastic hydraulic aquifer management,*Water Resour. Res.*, 29(9), 3159-3167, 1993.

Cleveland, T.G. and W.W.-G. Yeh, Sampling network design for transport parameter identification, *J.Water Resour. Plng. and Mgmt., ASCE,*116(6), 764-783, 1990.

Dagan, G., Solute transport in heterogeneous porous formations, *J. Fluid Mech.*, 145, 151-177, 1984.

Dagan, G., Theory of solute transport by groundwater, *Ann. Rev. Fluid Mech.*, 19, 183-215, 1987.

Dagan, G., Time-dependent macrodispersion for solute transport in anisotropic heterogeneous aquifers, *Water Resour. Res.*, 24(9), 1491-1500, 1988.

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

Kitanidis, P.K., and E.G. Vomvoris, A geostatistical approach to the inverse problem in groundwater modeling (steady state) and one-dimensional simulations, *Water Resour. Res.*, 19(3), 677-690, 1983.

Knopman, D.S., and C.I. Voss, Discrimination among one-dimensional models of solute transport in porous media: Implications for sampling design, *Water Resour. Res.*, 24(11), 1859-1876, 1988.

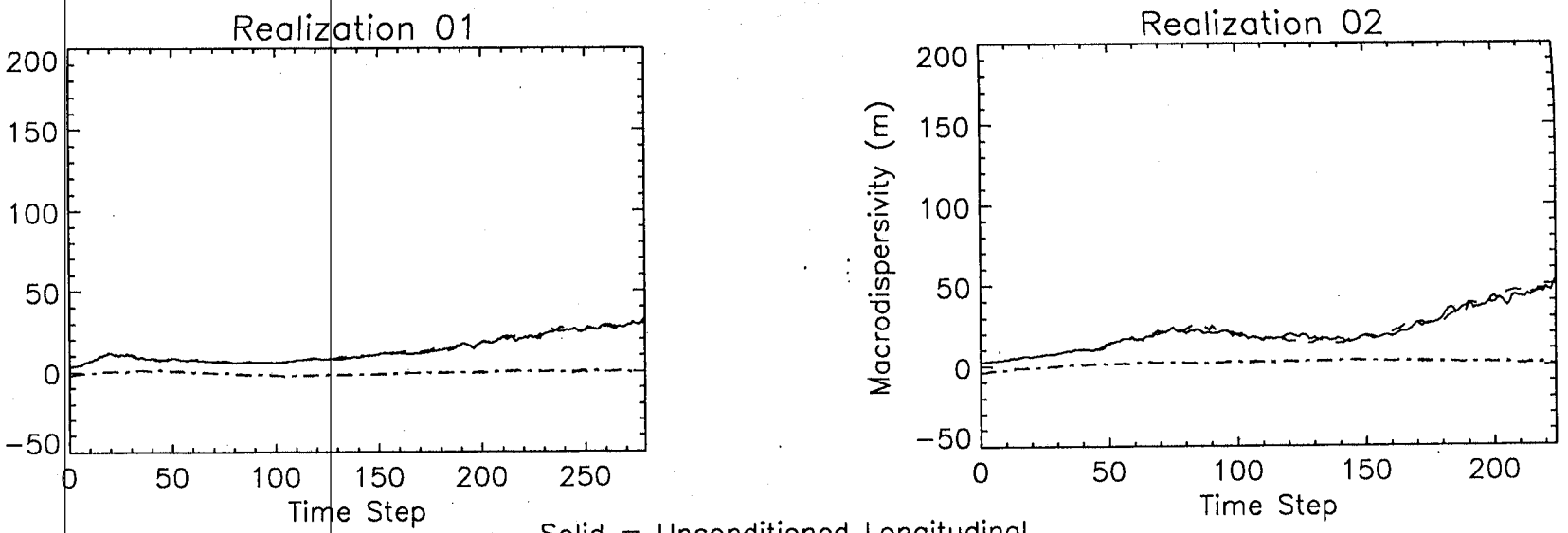
Knopman, D.S., and C.I. Voss, Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport, *Water Resour. Res.*, 25(10), 2245-2258, 1989.

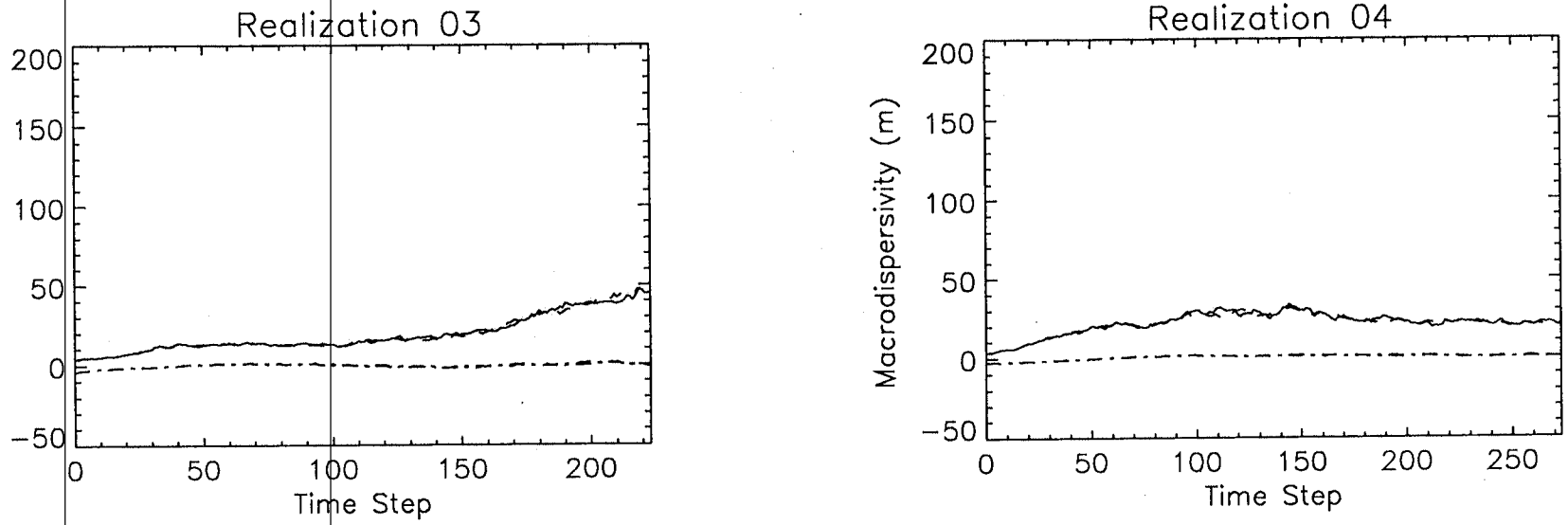
Wagner, B.J., and S.M. Gorelick, Reliable aquifer remediation in the presence of spatially variable hydraulic conductivity: From data to design, *Water Resour. Res.*, 25(10), 2211-2225, 1989.

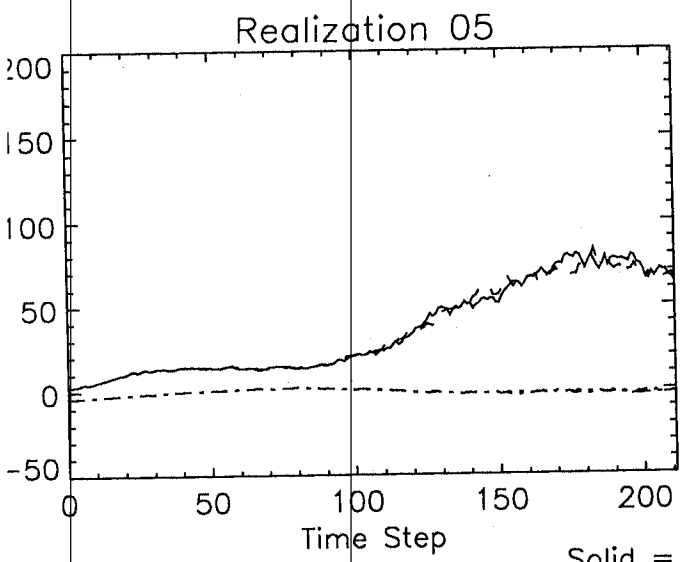
Appendix A.
Macrodispersivity vs. Time Plots for Information Transfer Experiment

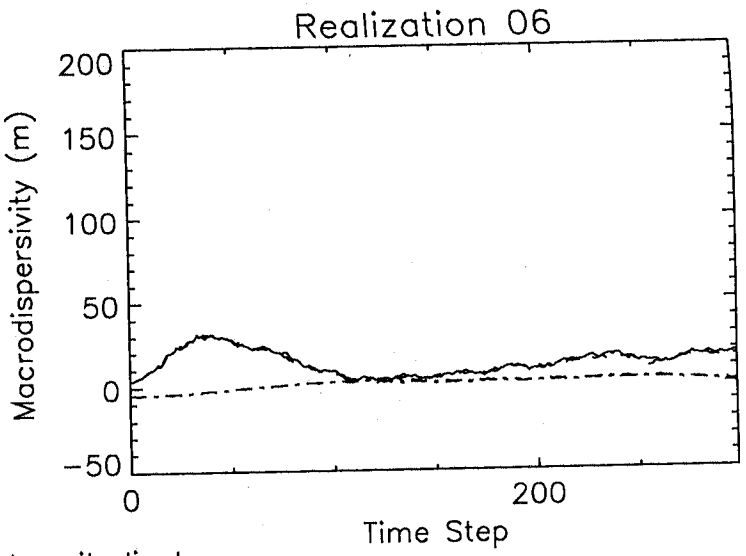
Appendix A: Macrodisersivity vs. Time Plots for Information Transfer Experiment

Figure A.1. Macrodispersivity vs. Time for 100 Single-Scale Realizations.

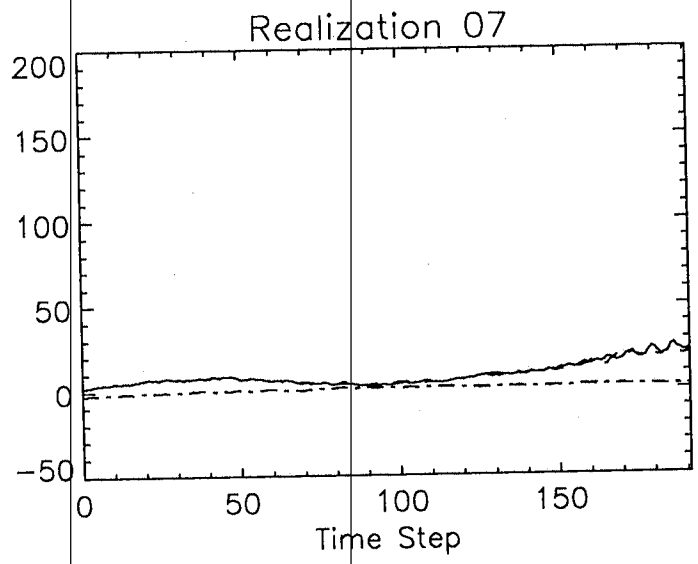

Realization 01
Realization 02
Macrodispersivity (m)
Time Step
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

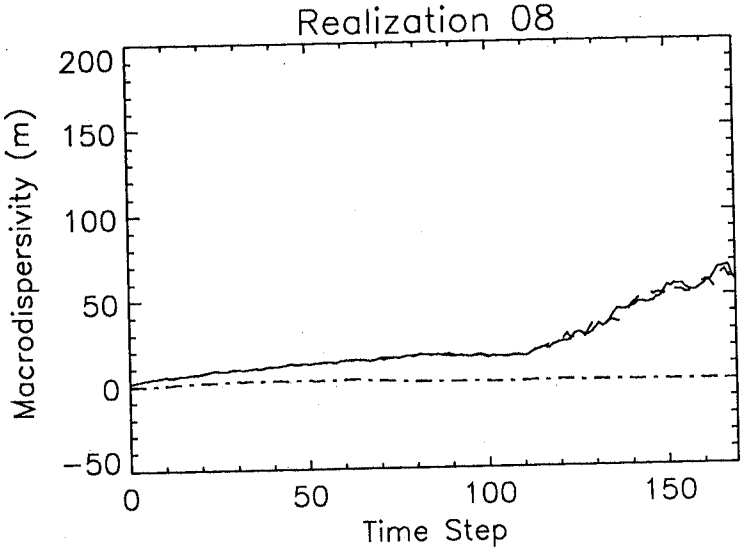

Realization 03
Realization 04
Macrodispersivity (m)
Time Step

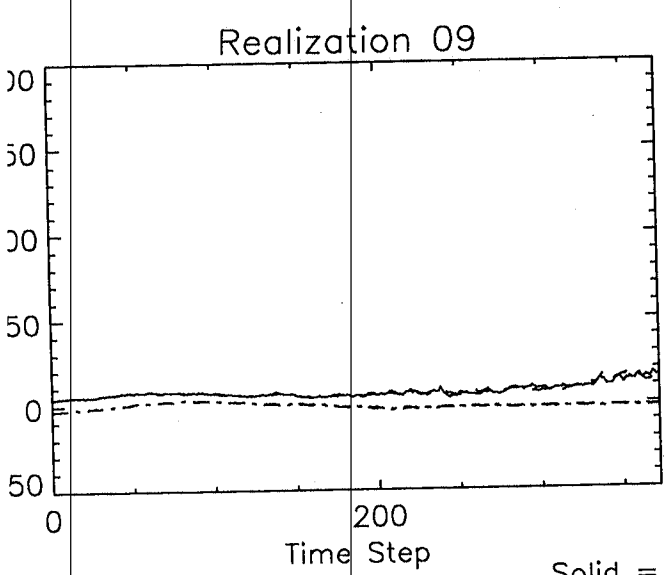


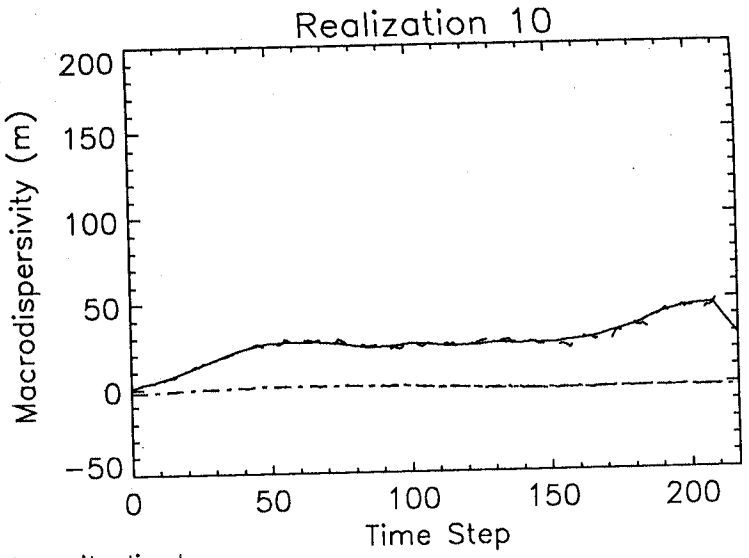


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

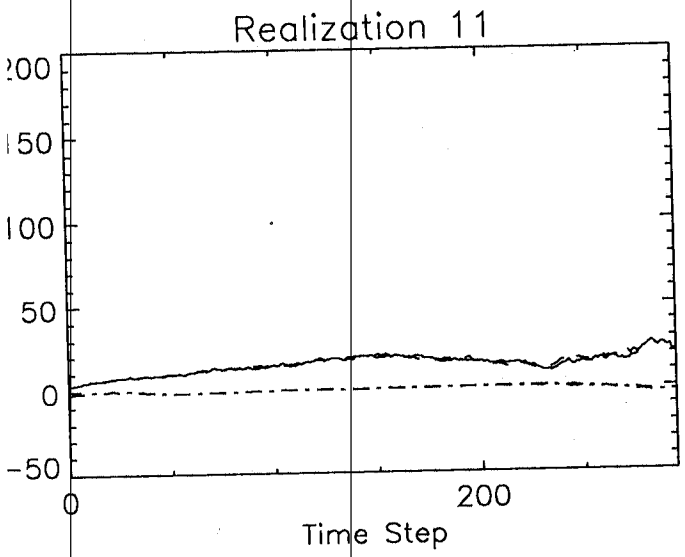


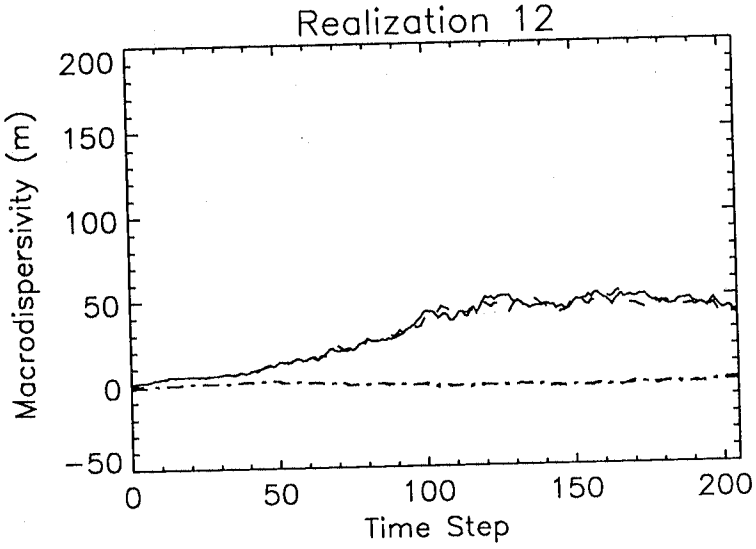


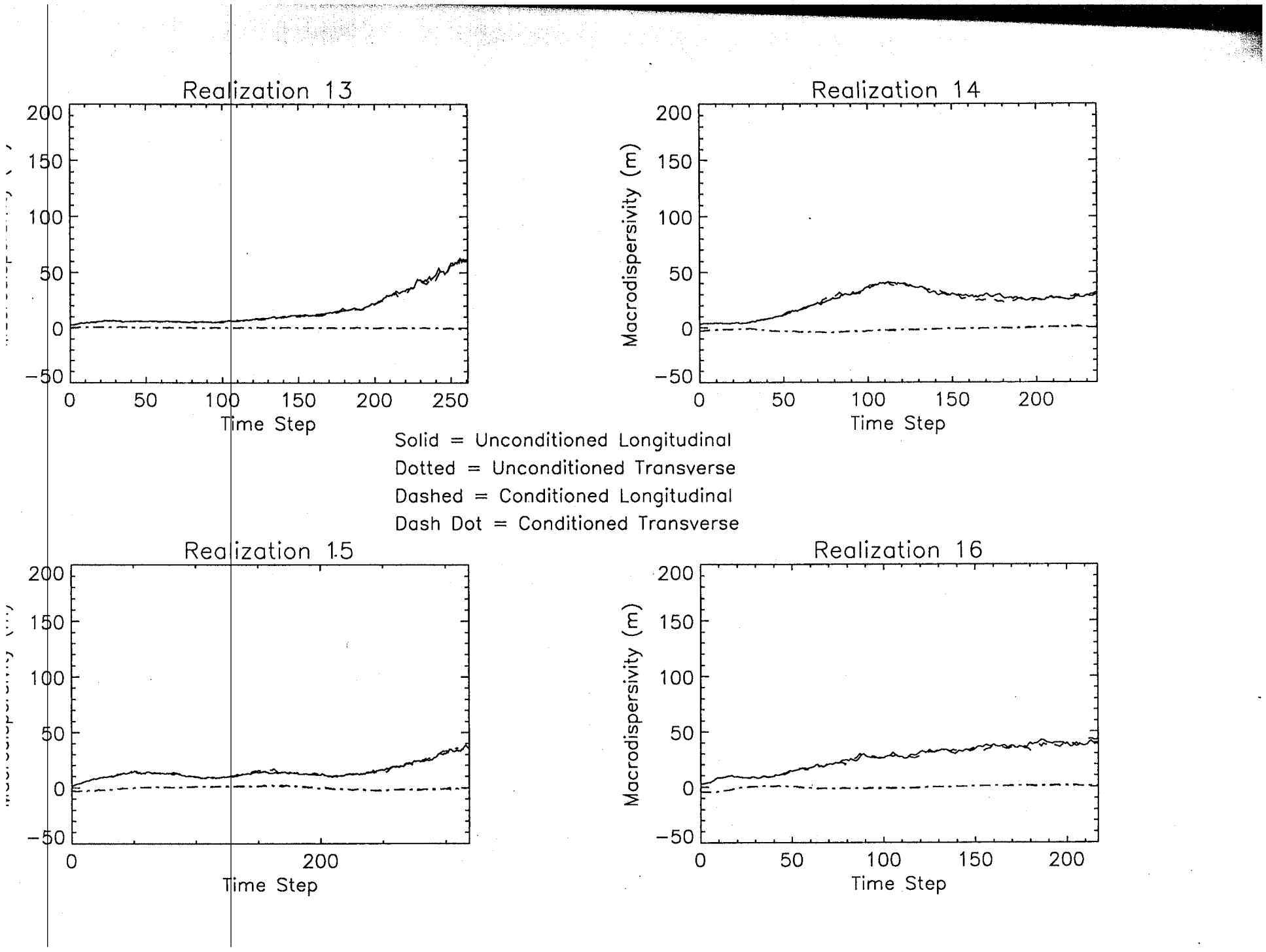


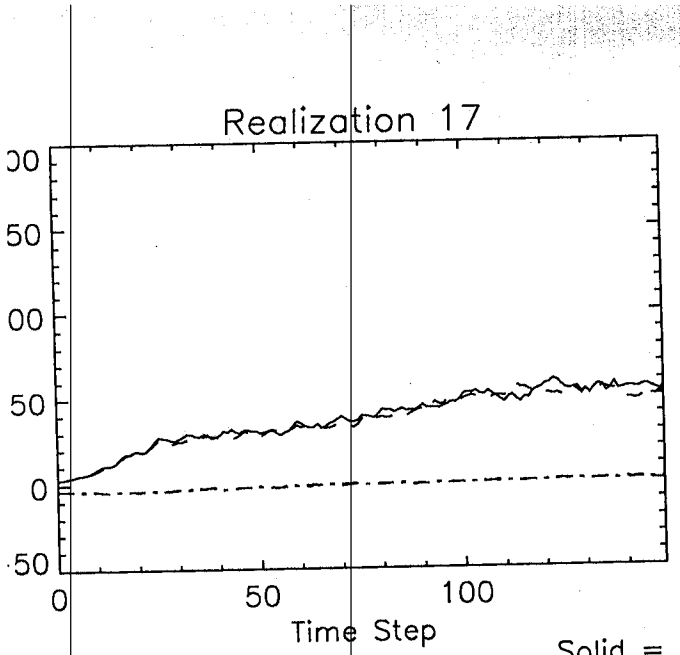


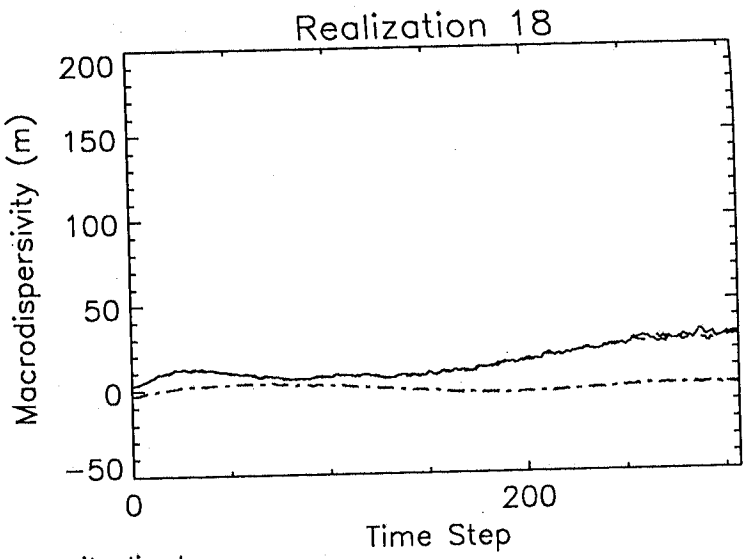
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse



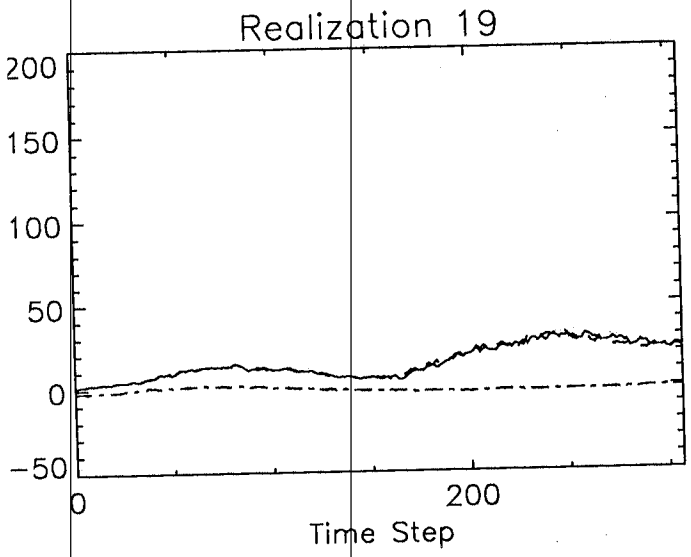


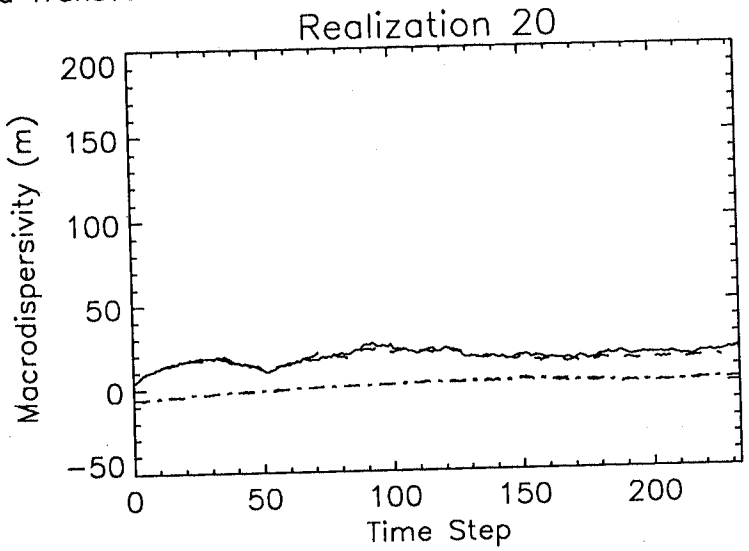

Realization 13
Realization 14
Realization 15
Realization 16
Macrodispersivity (m)
Time Step
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

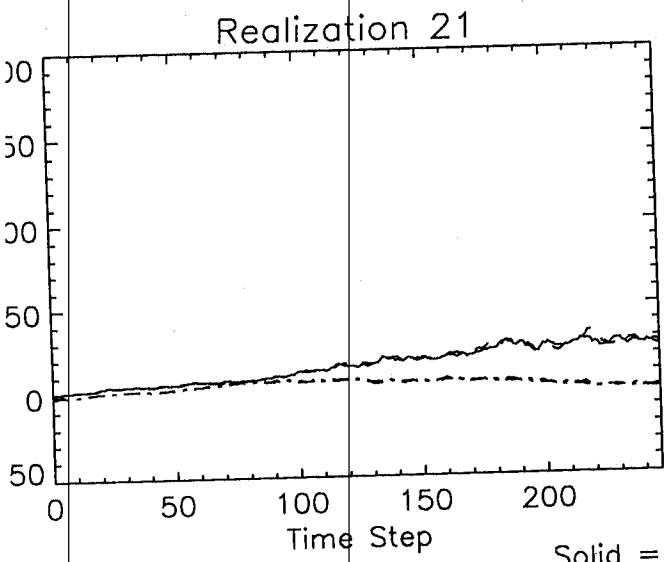


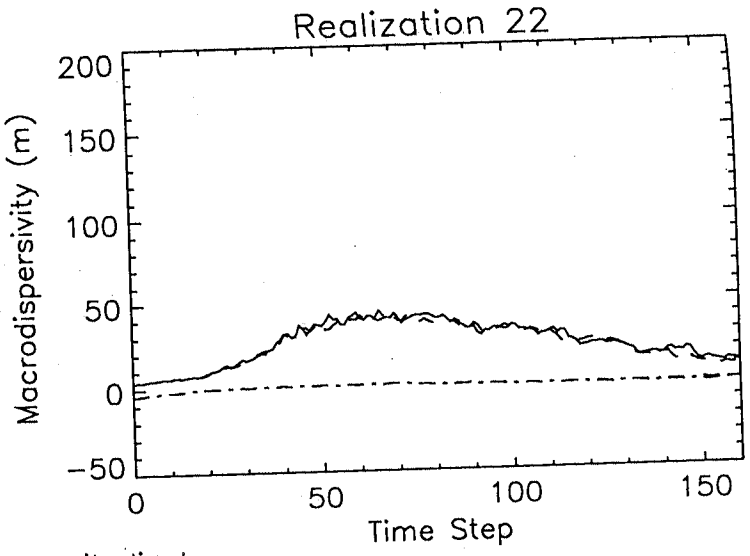


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

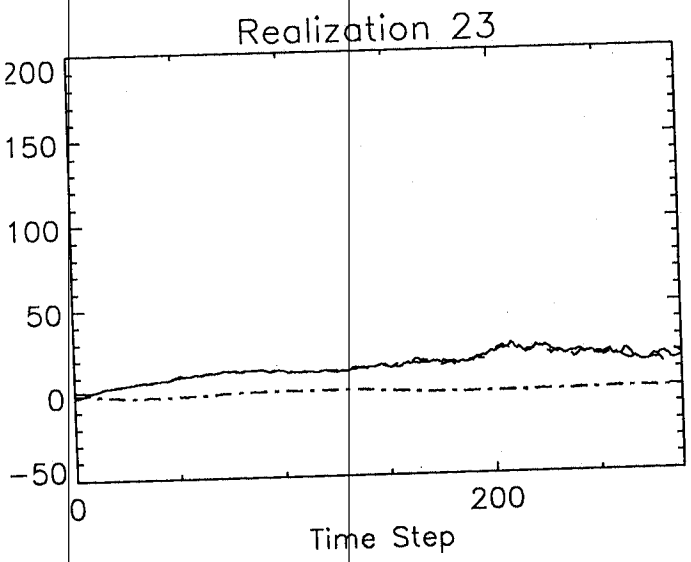


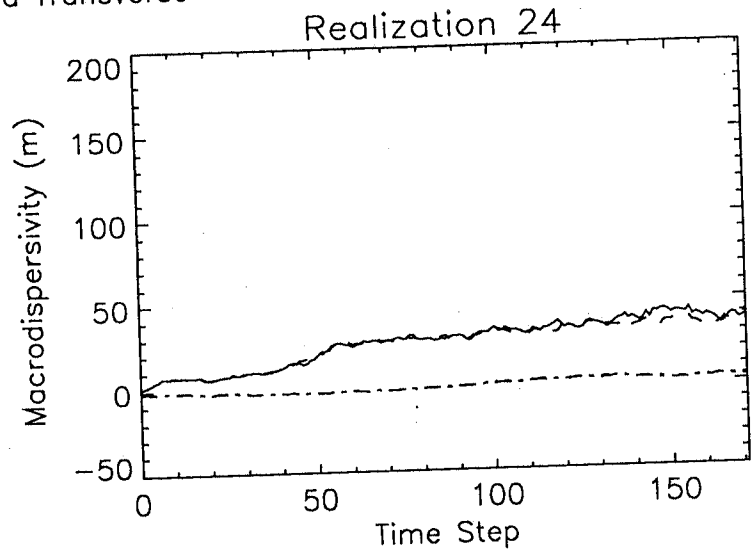


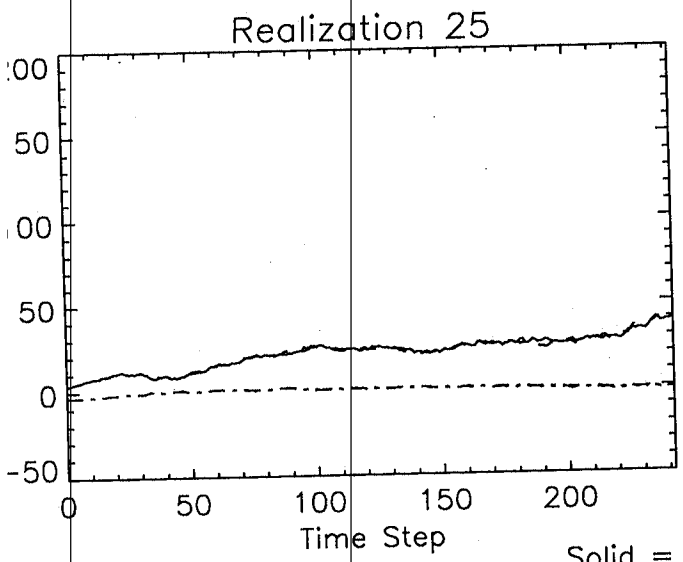


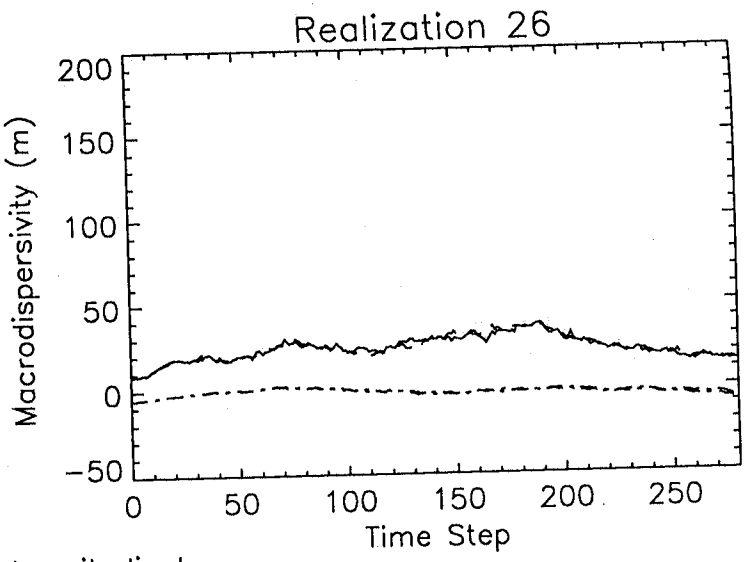


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

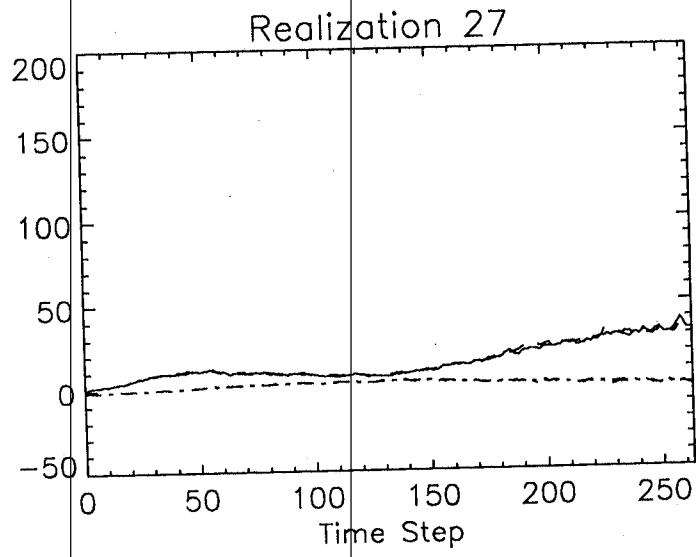


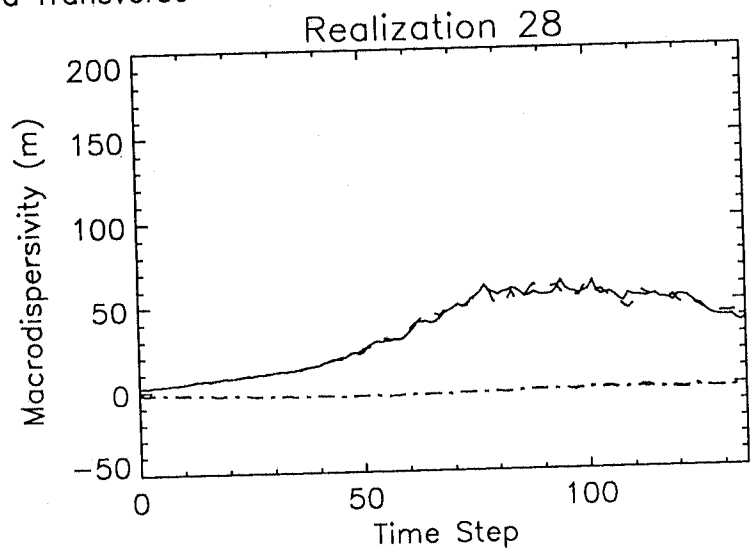


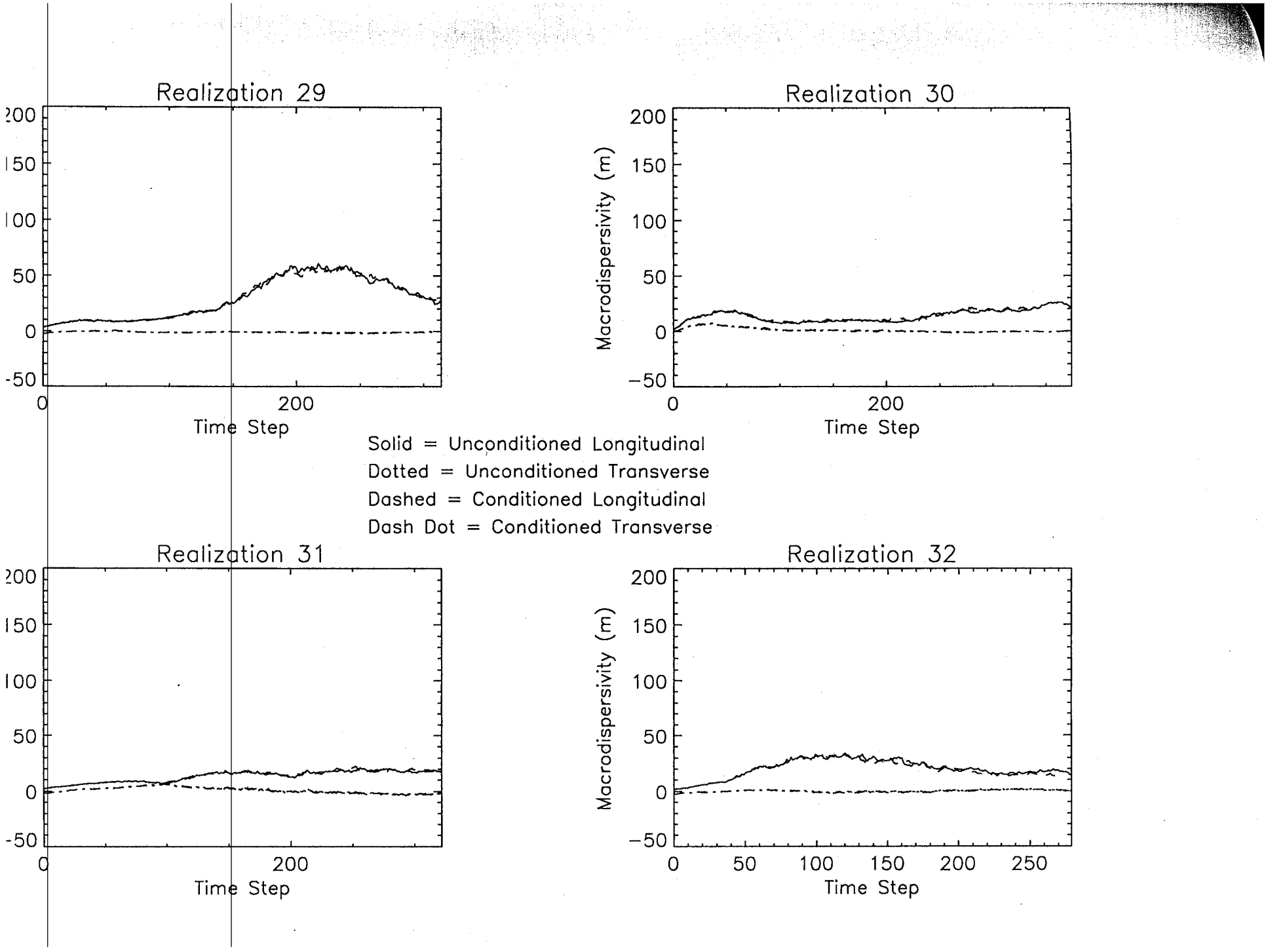


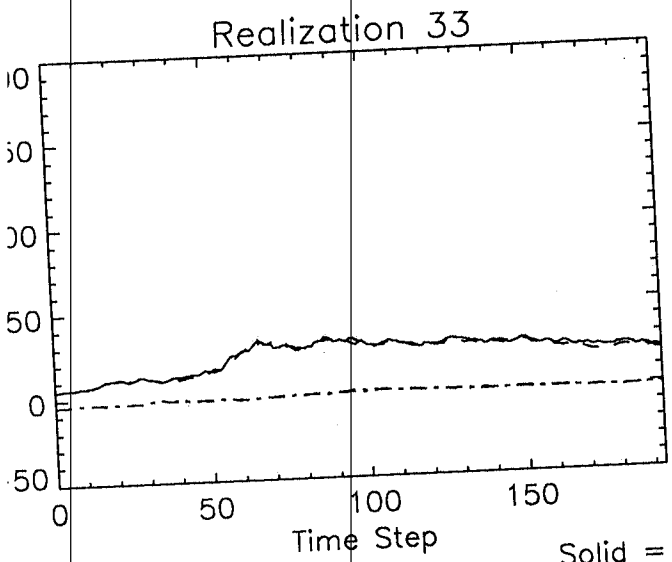


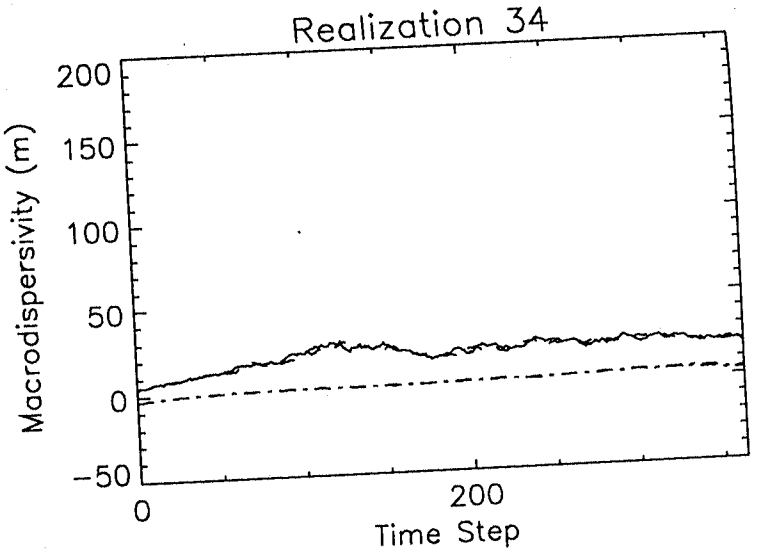
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse



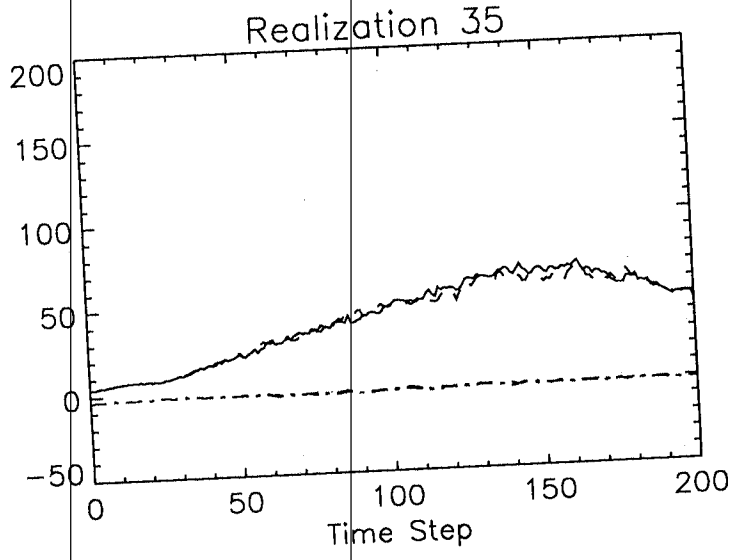


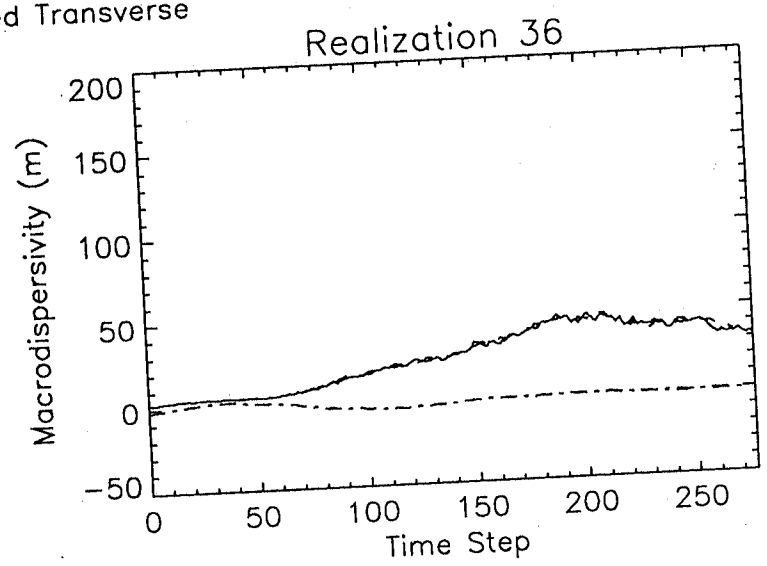

Realization 29
Realization 30
Realization 31
Realization 32
Macrodispersivity (m)
Time Step
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

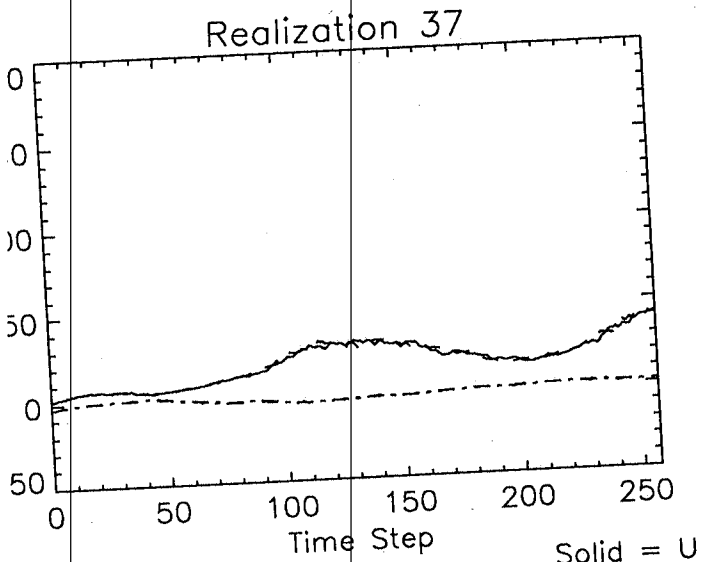


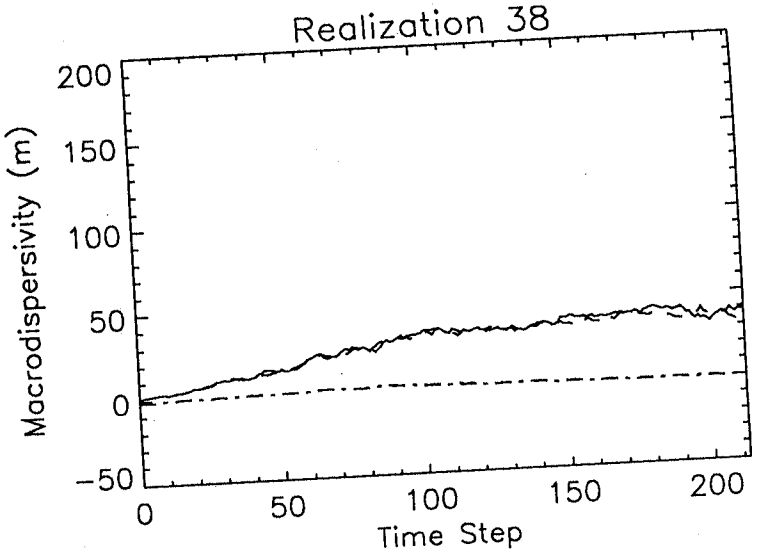


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

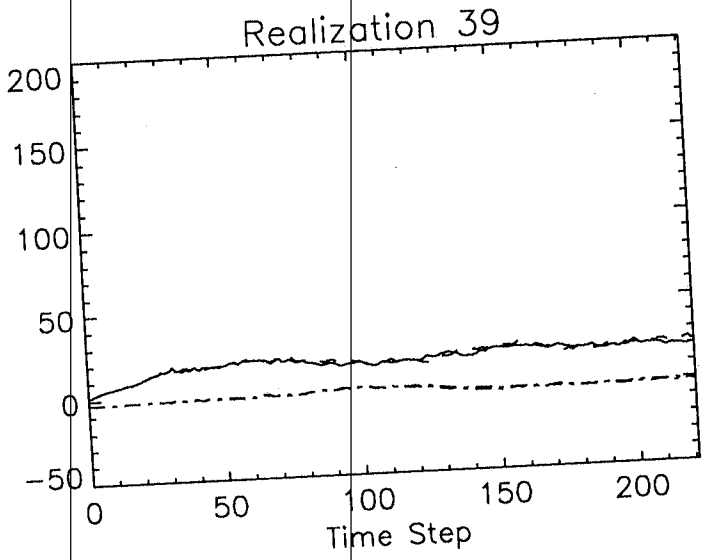


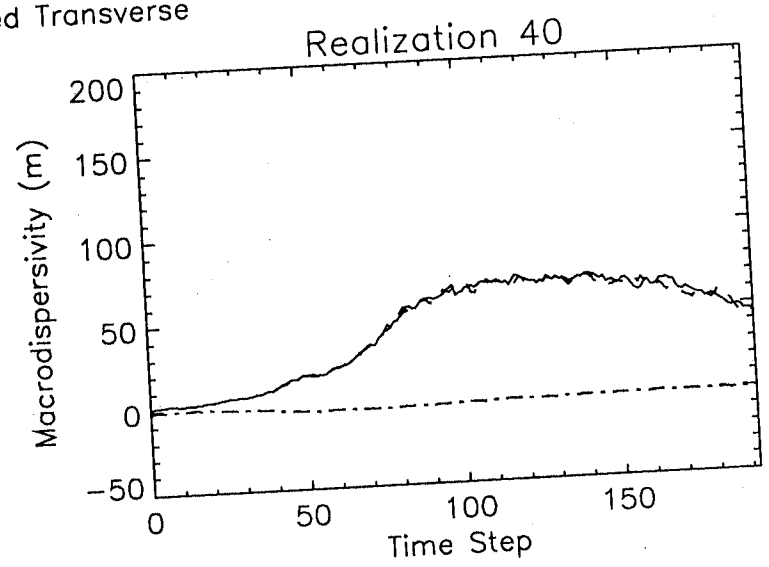


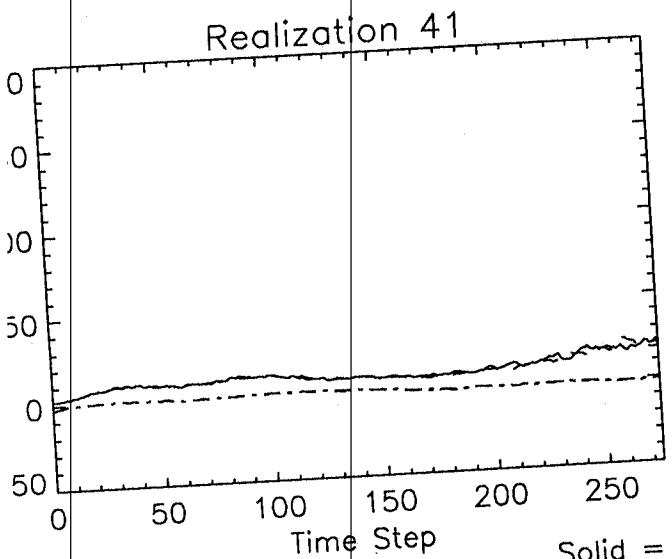


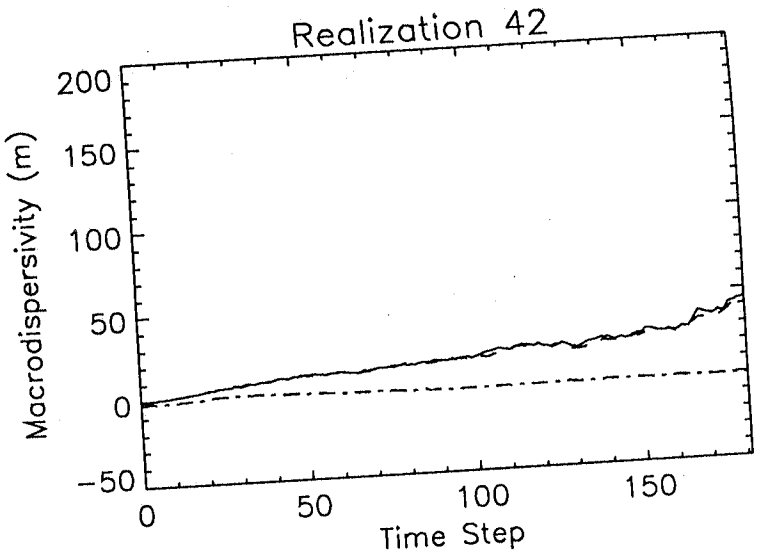


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

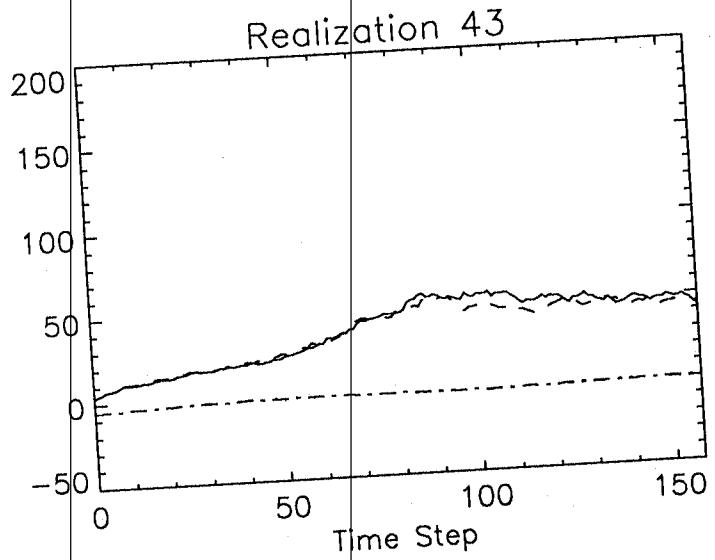


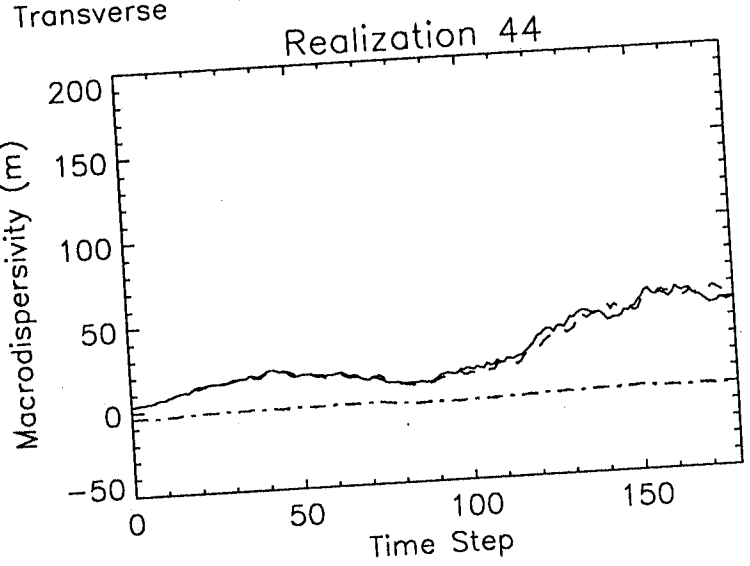


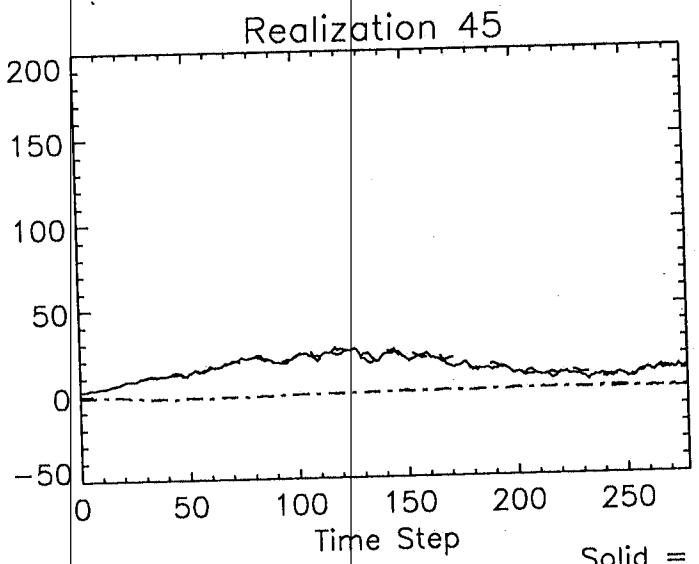


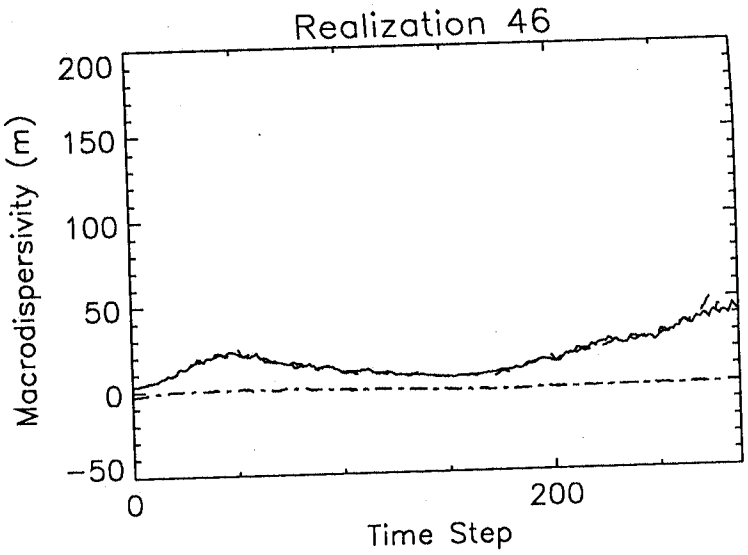


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

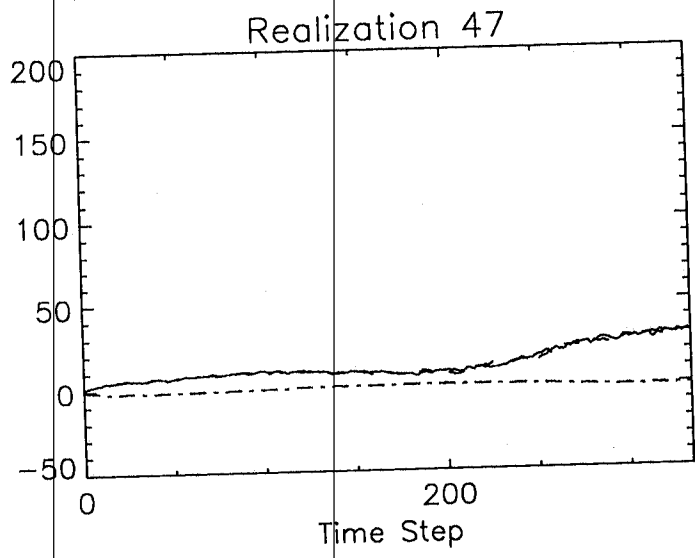


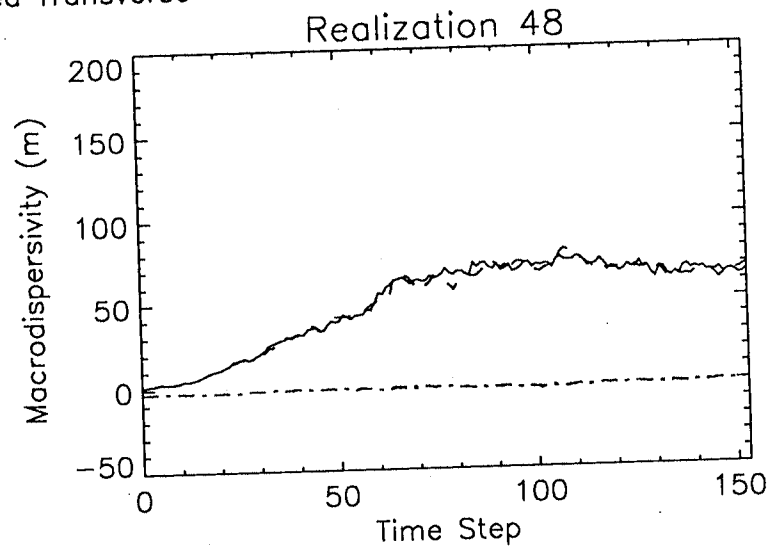


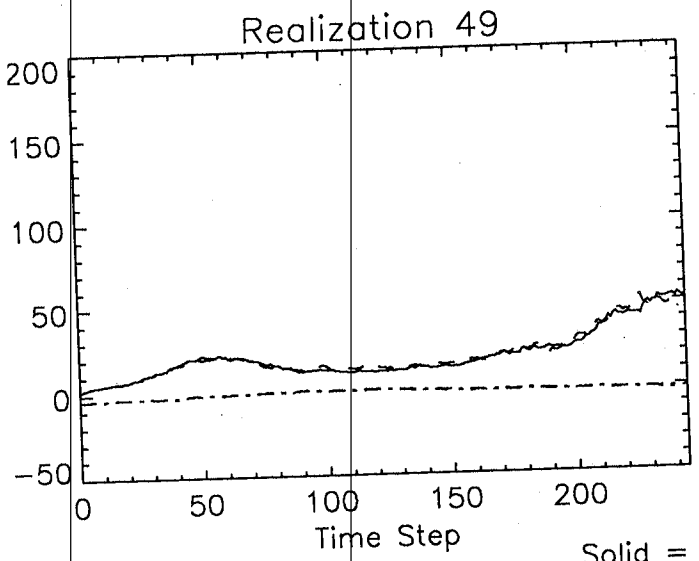


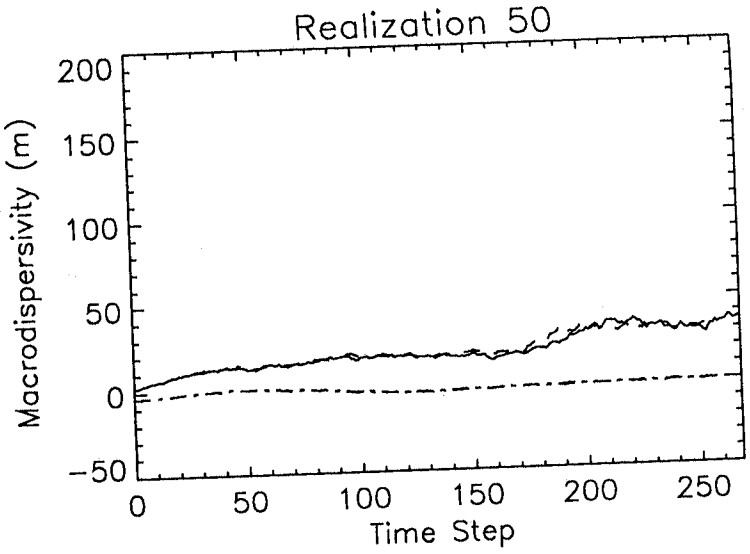


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

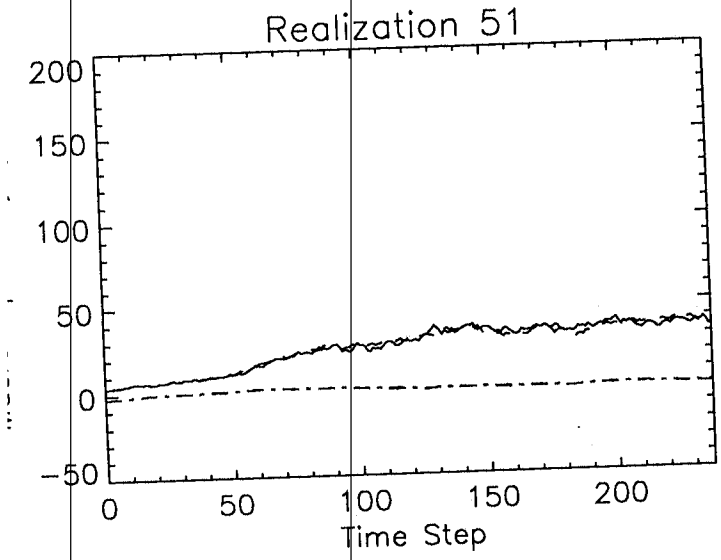


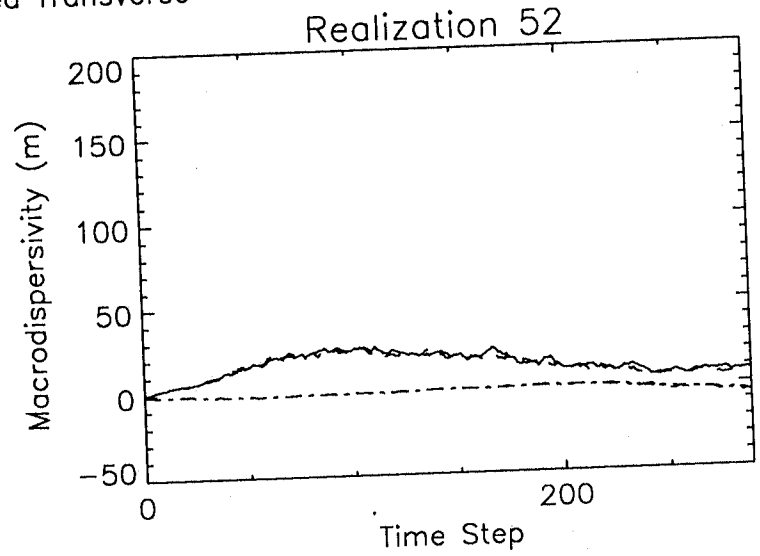


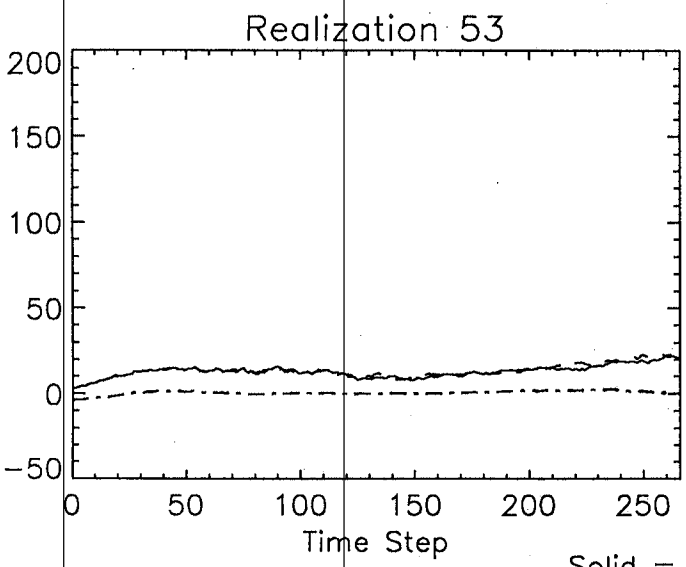


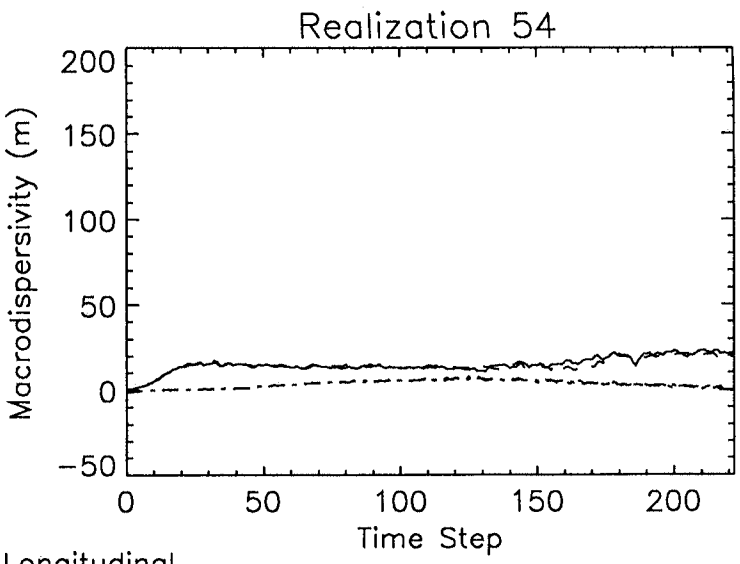


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

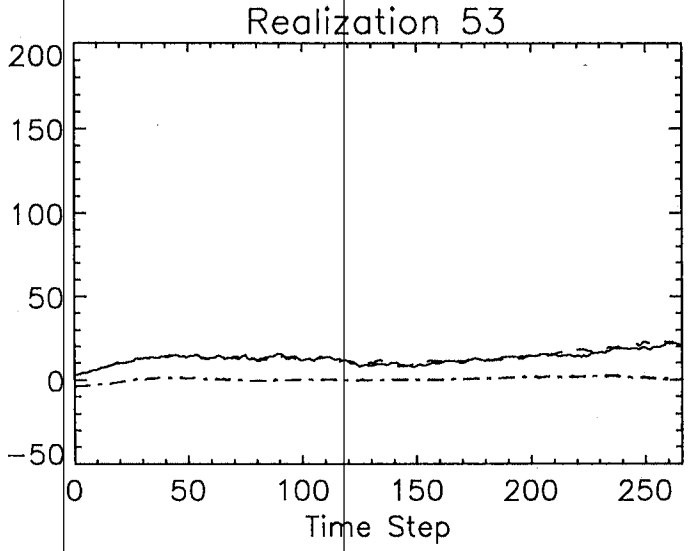


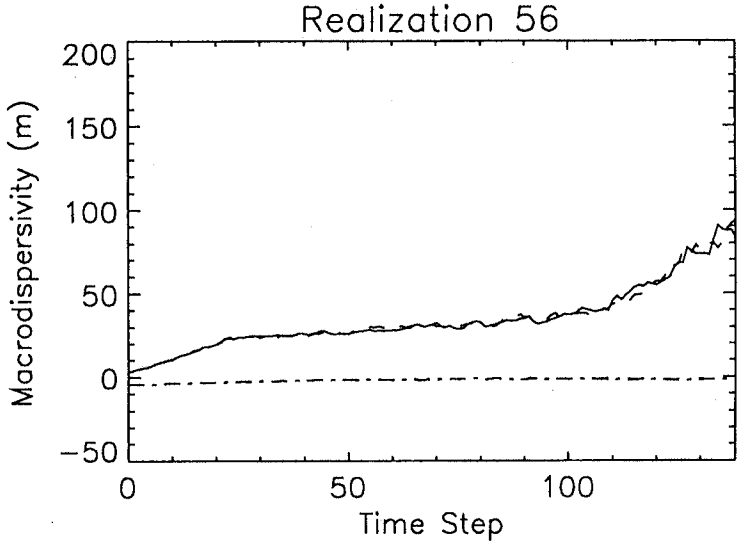


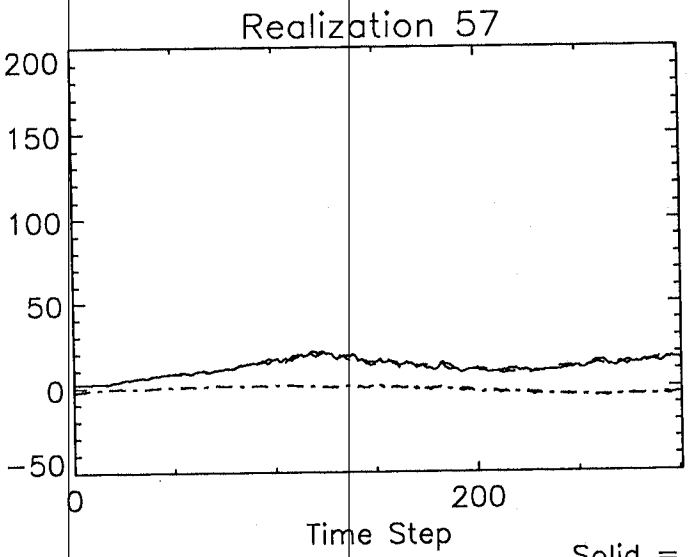


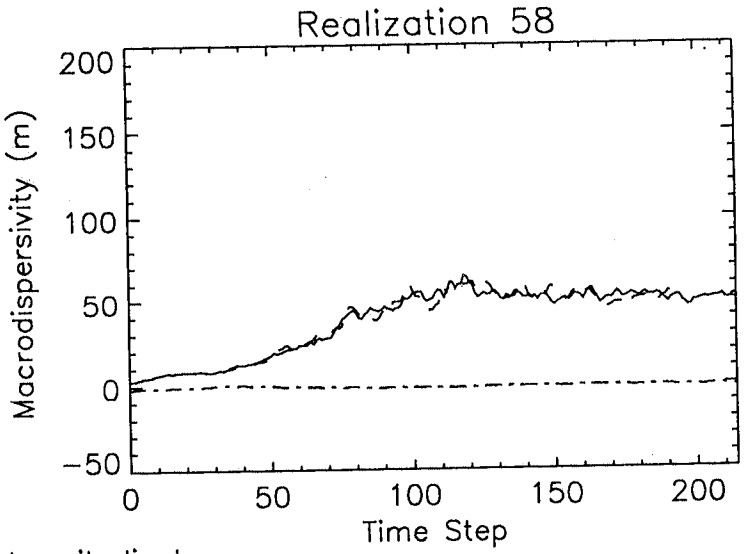


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

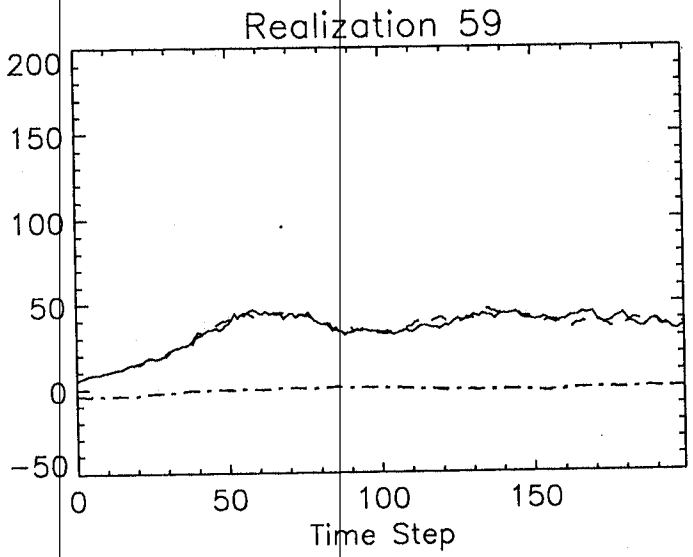


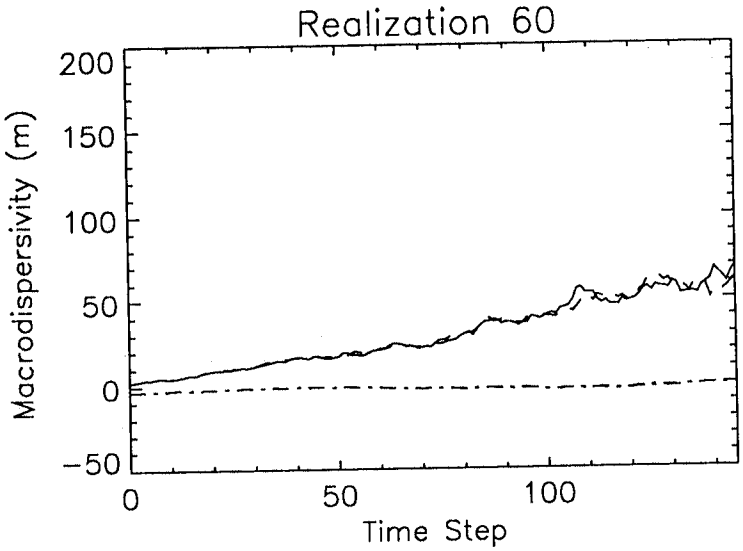


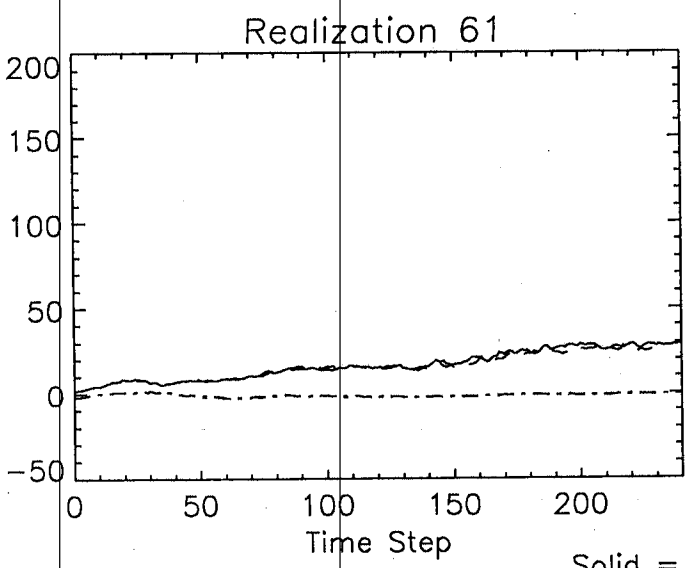


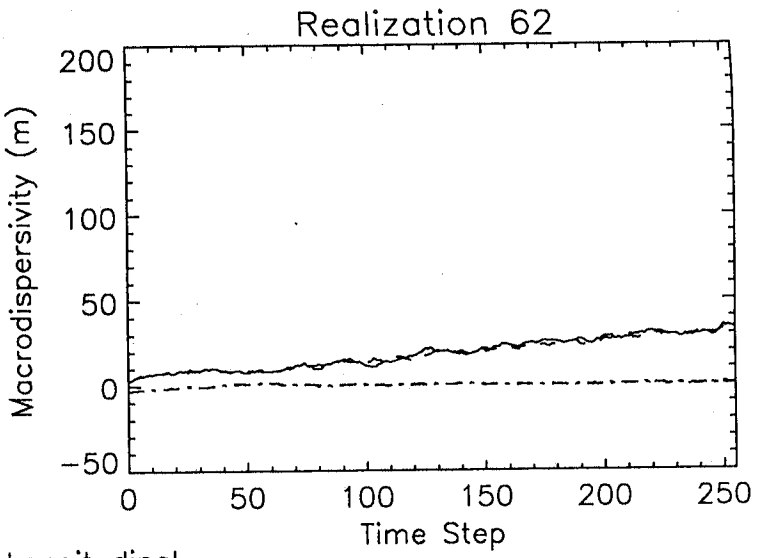


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

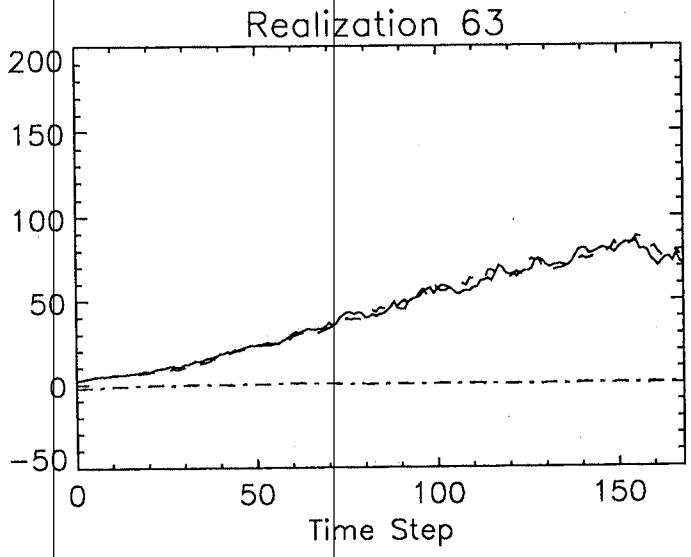


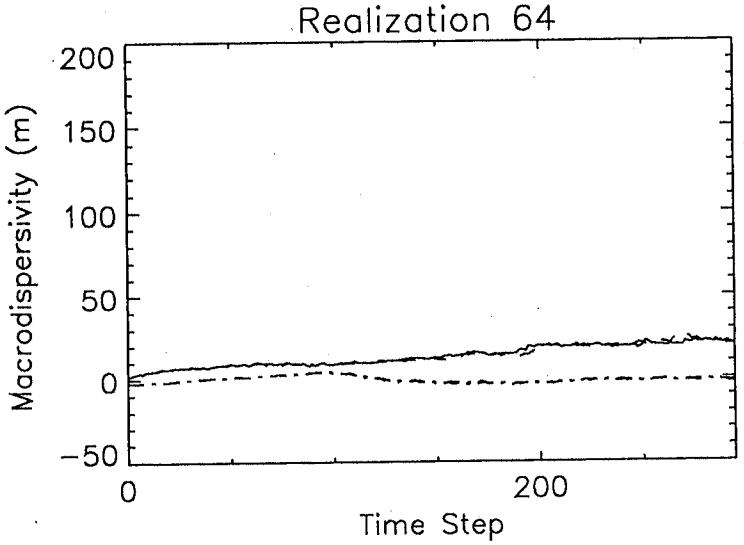


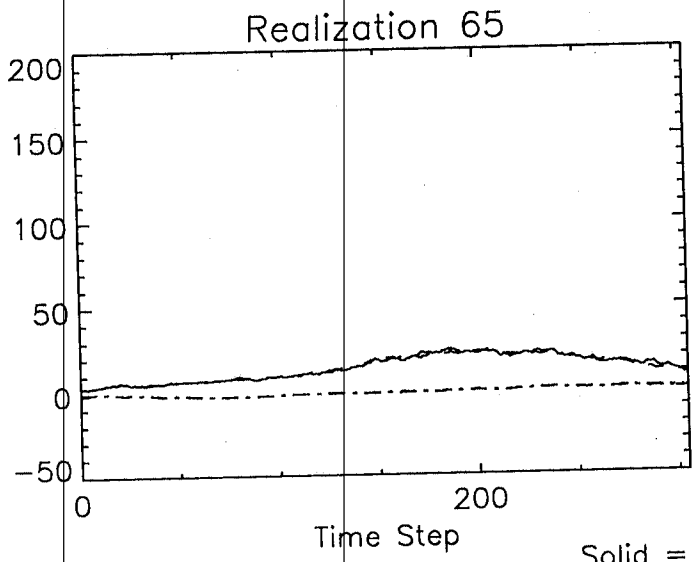


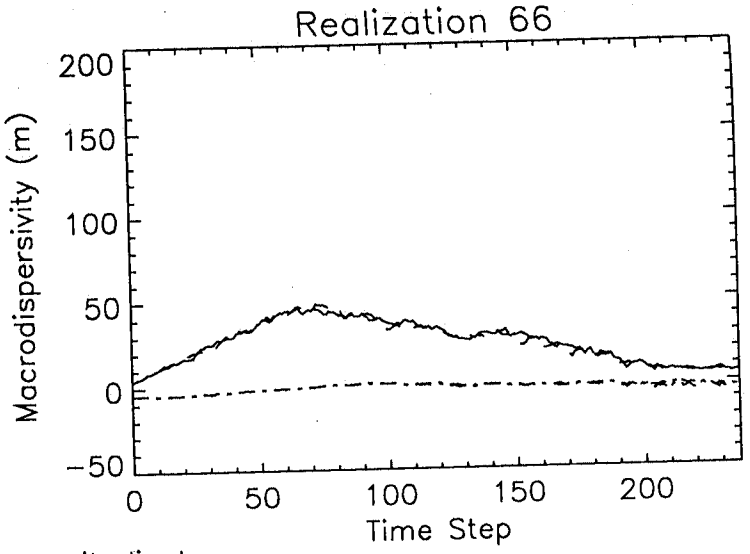


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

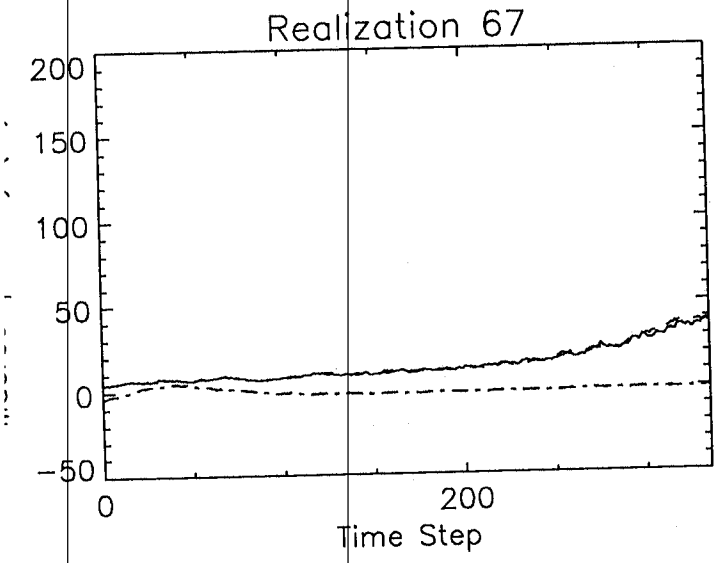


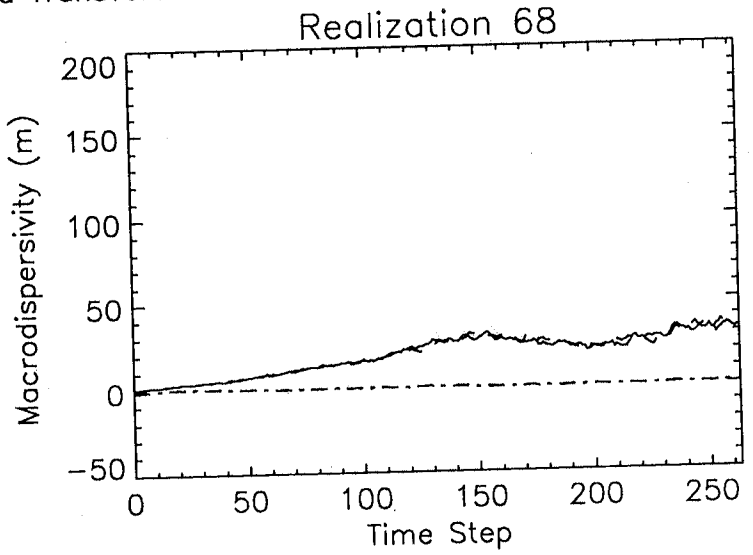


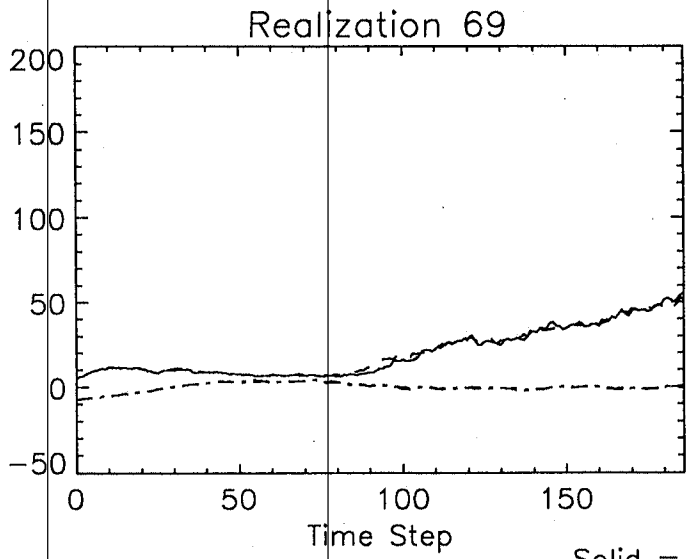


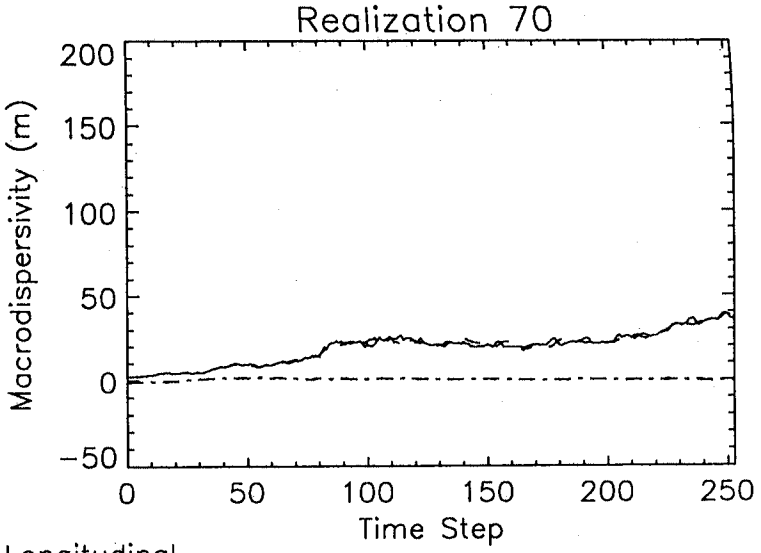


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

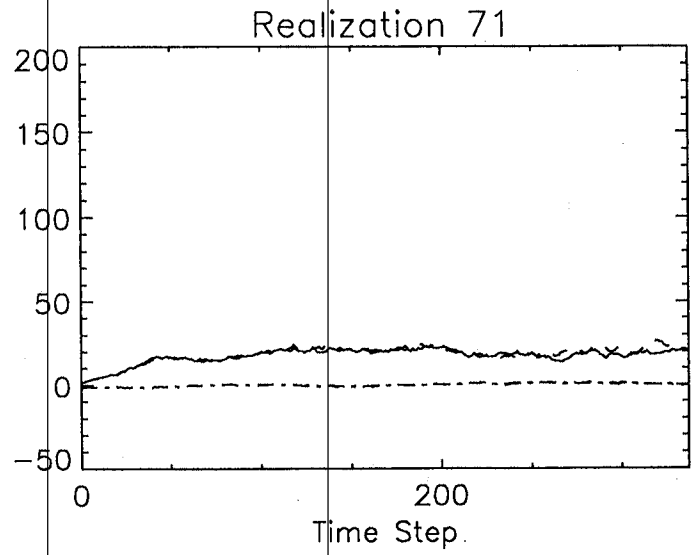


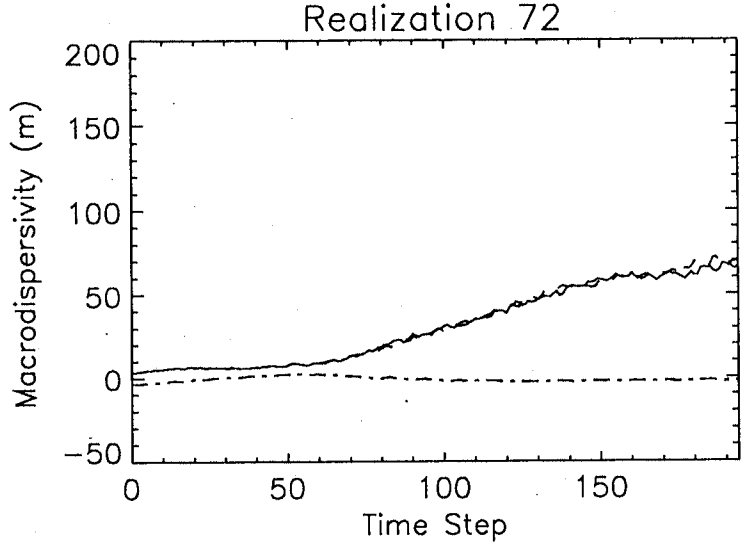


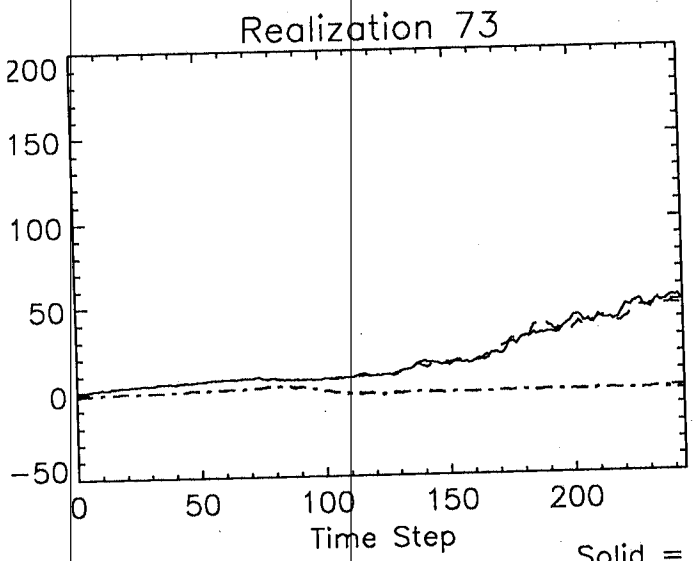


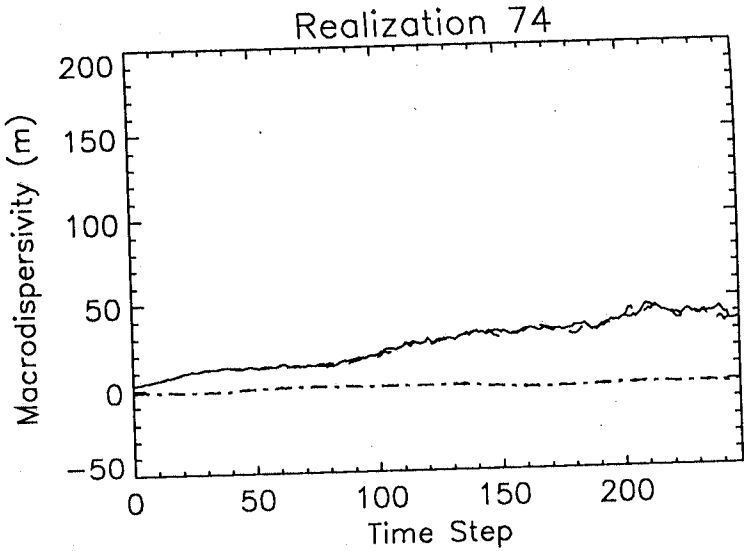


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

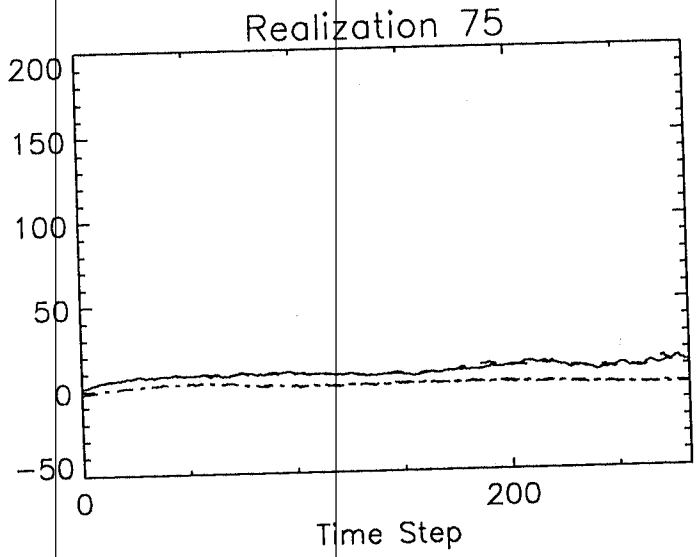


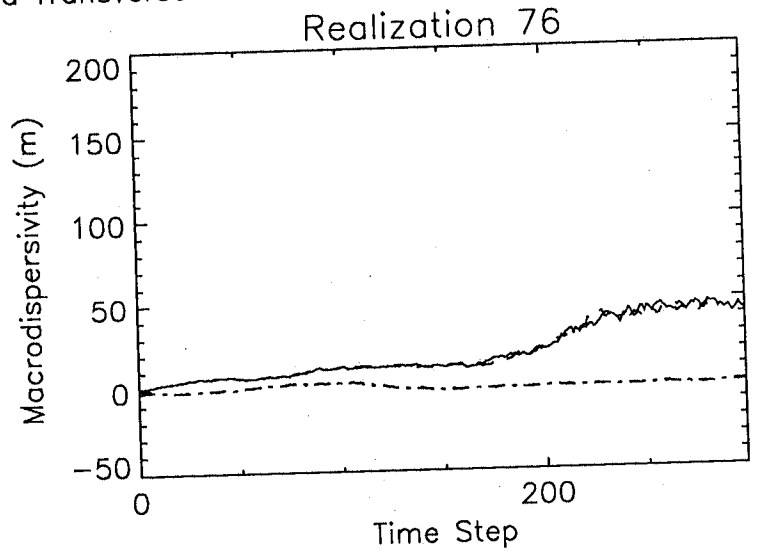


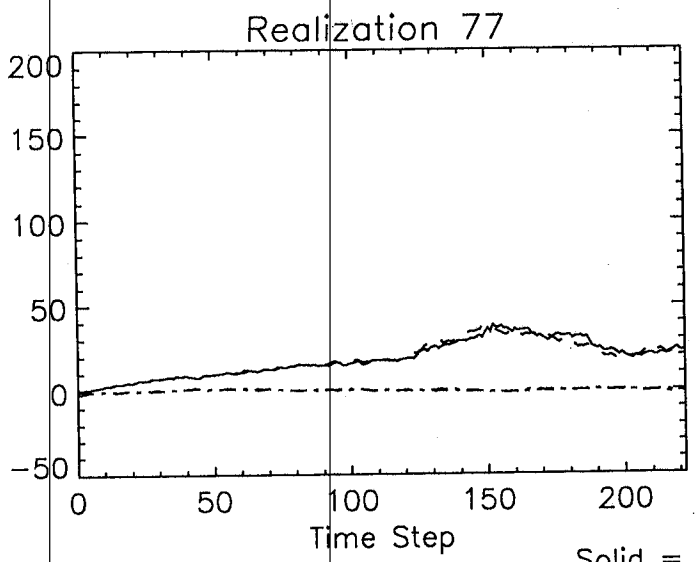


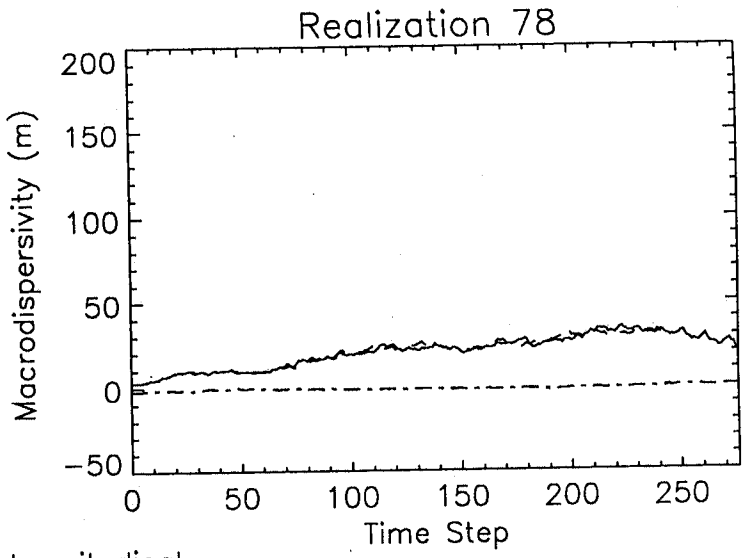


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

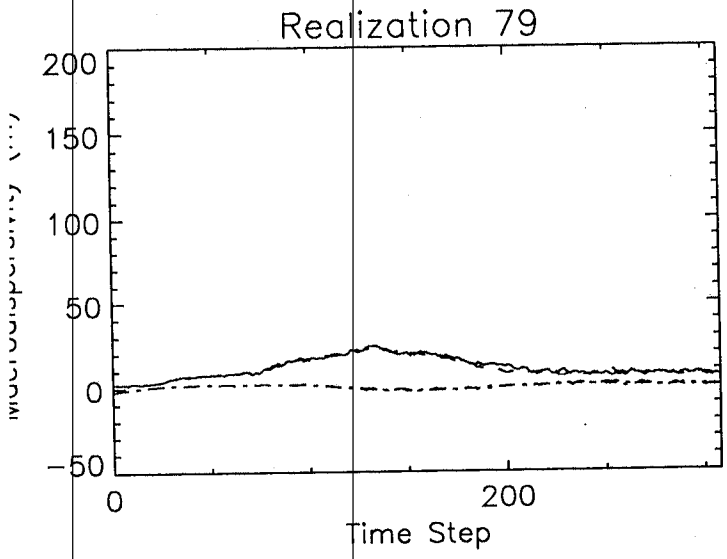


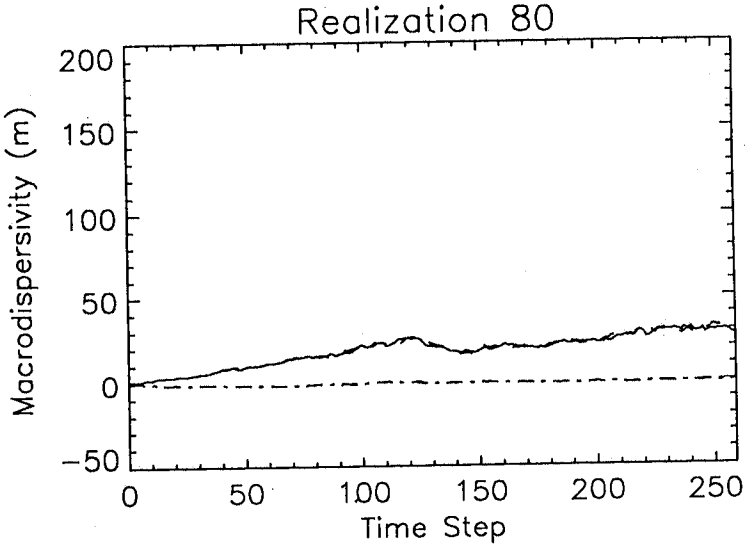


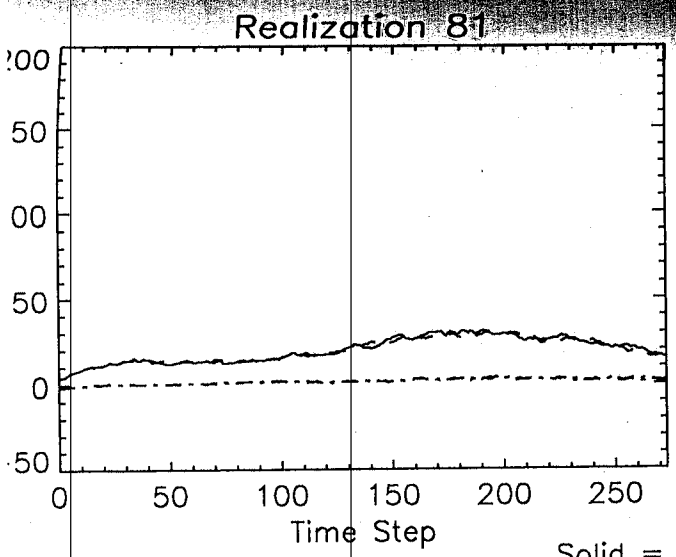


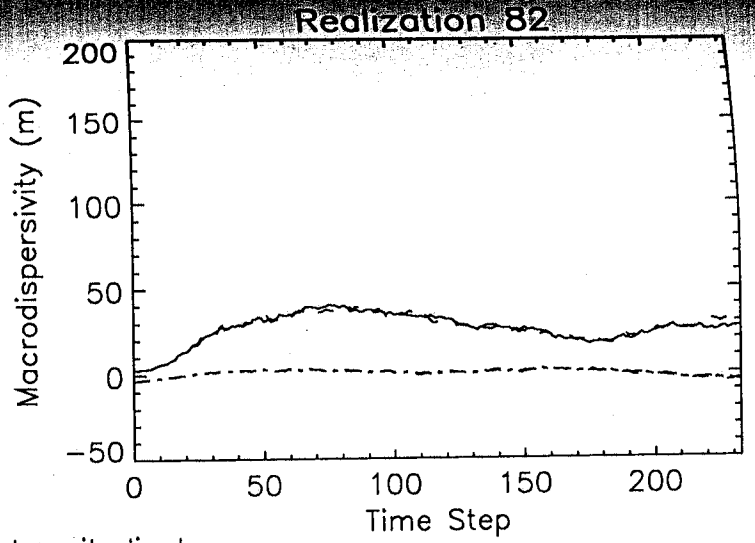


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

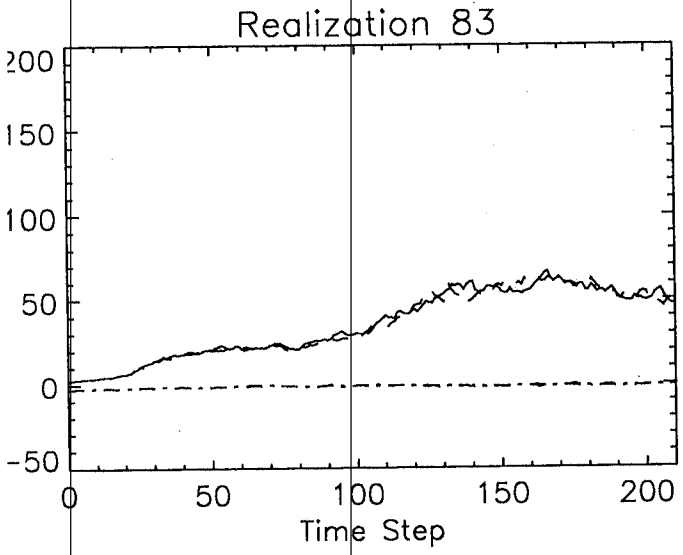


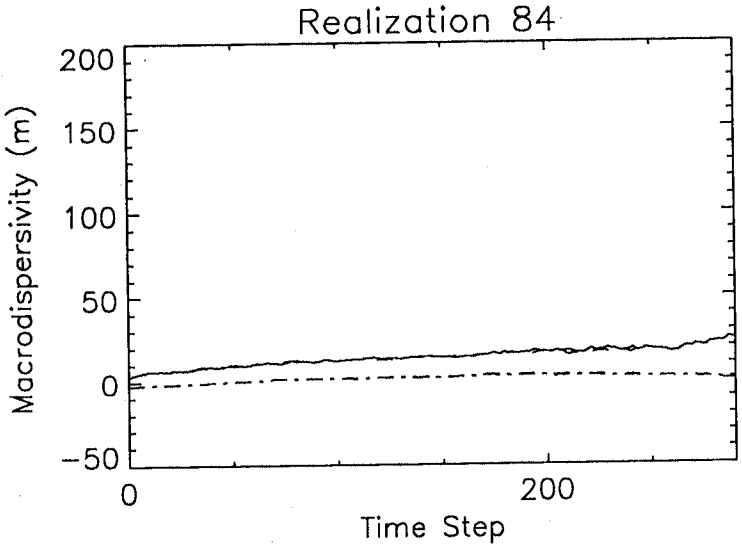


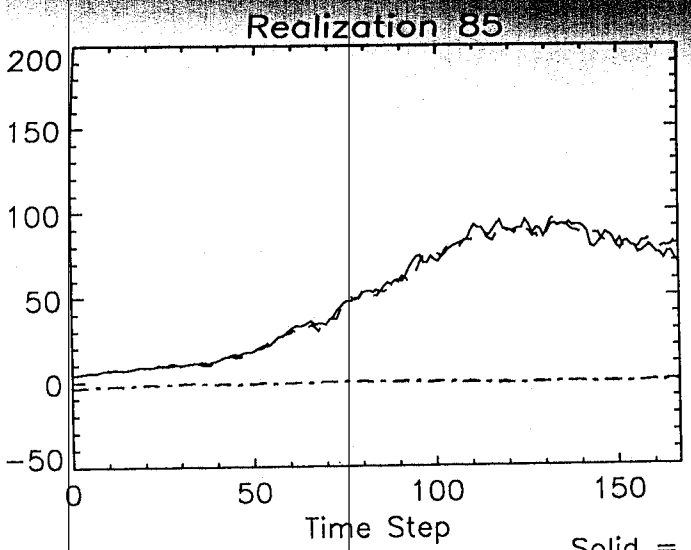


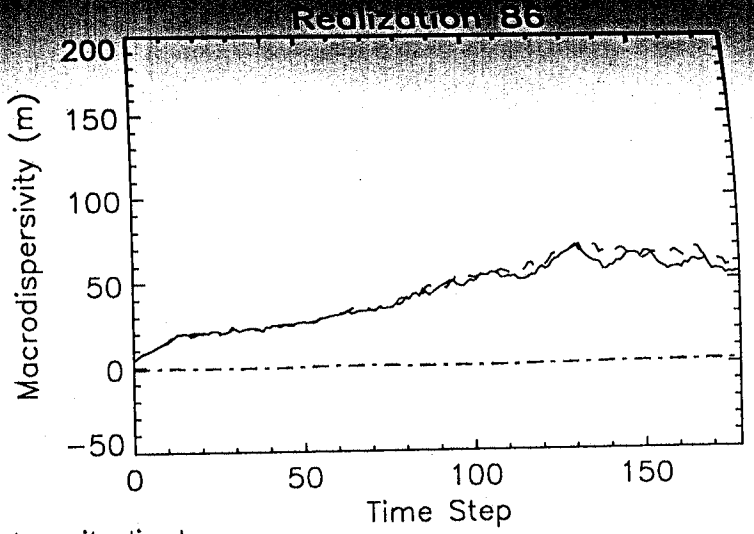


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

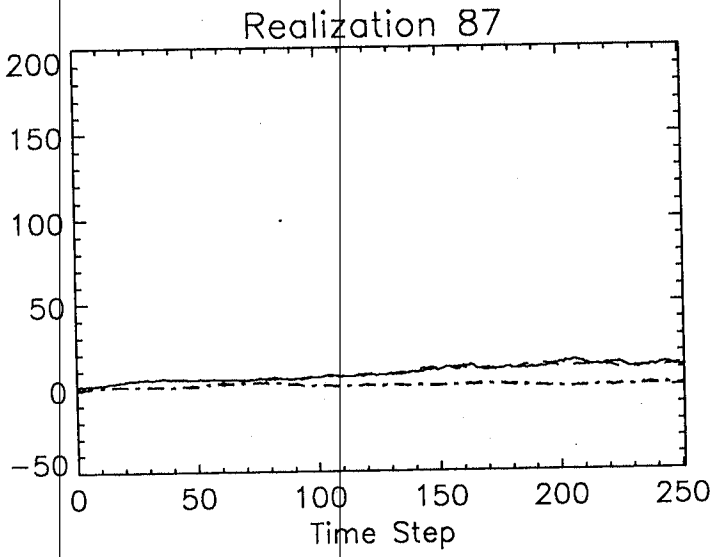


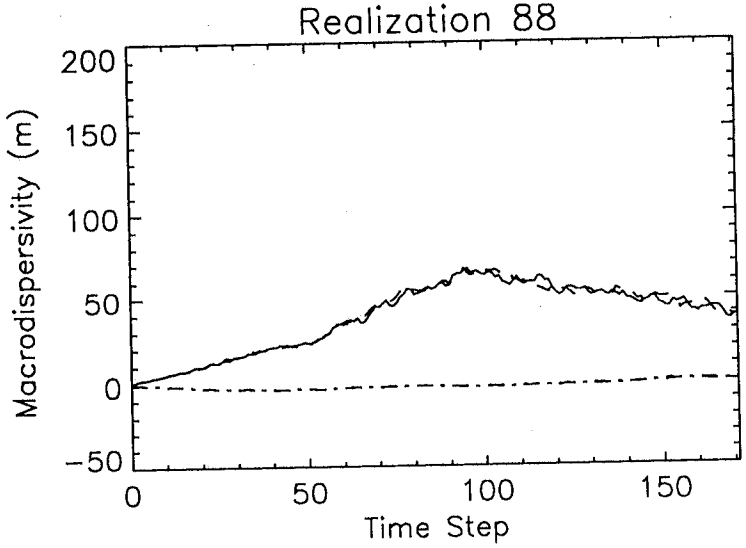


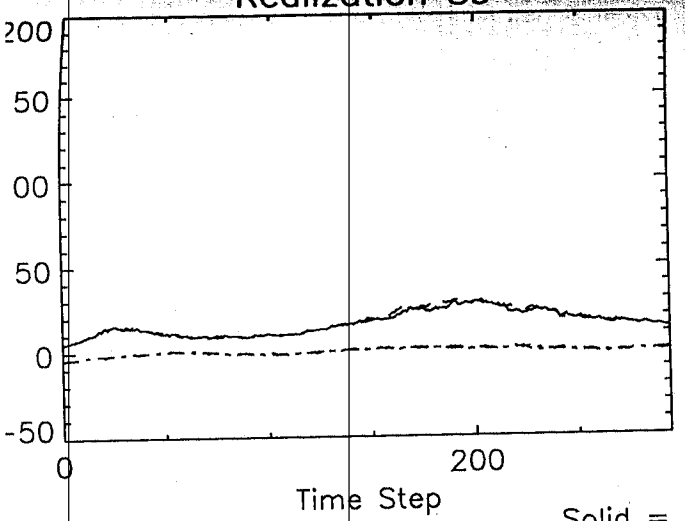


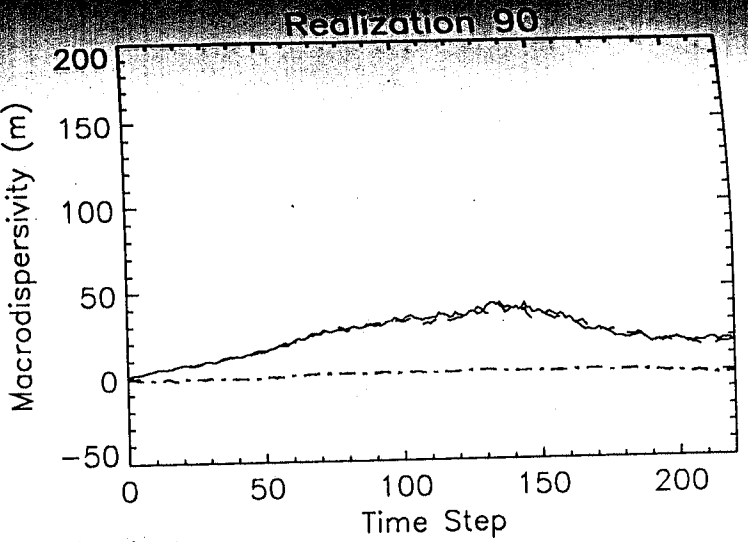


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

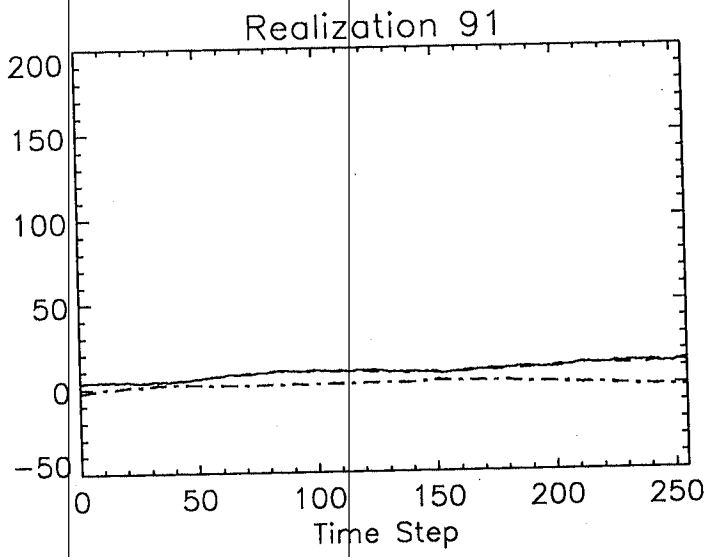


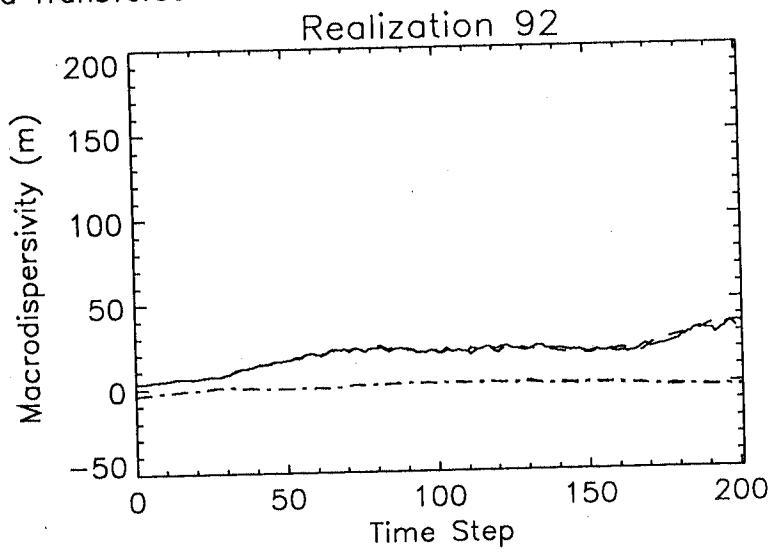


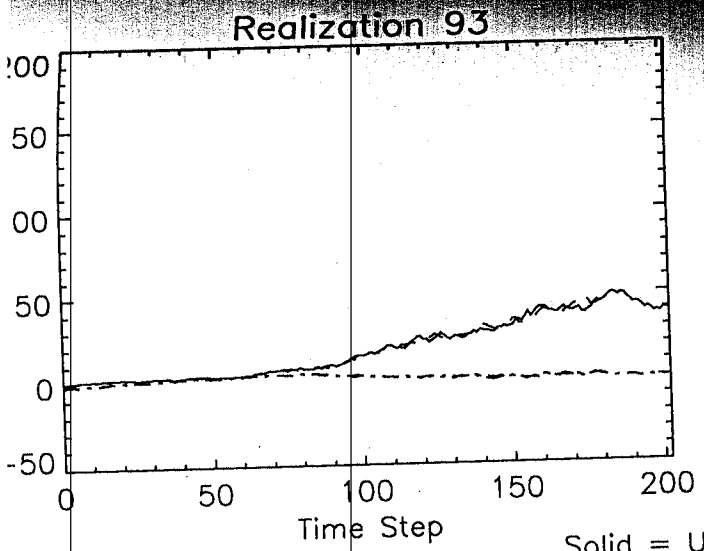


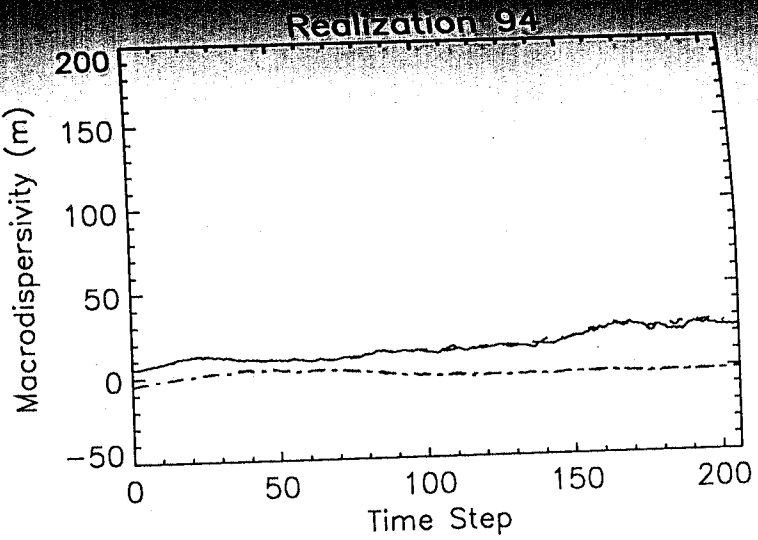


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

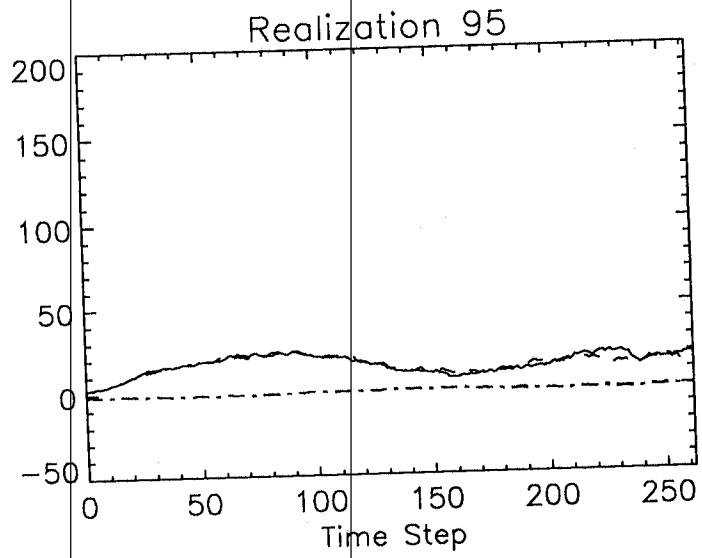


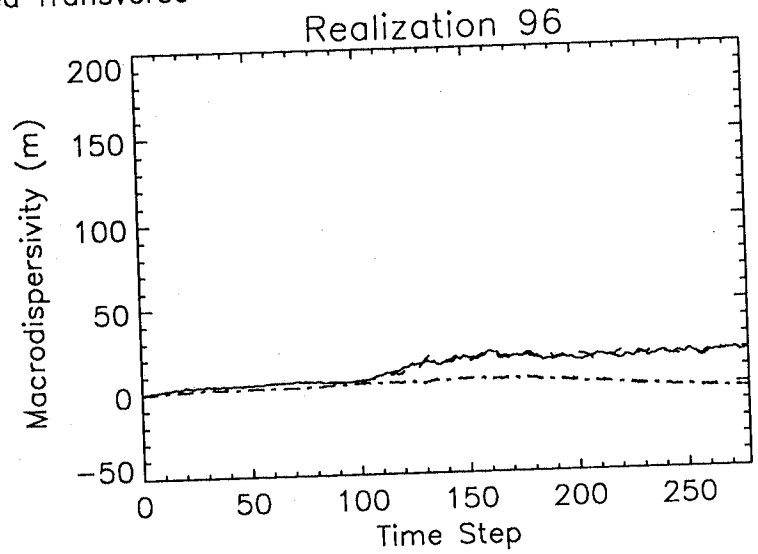


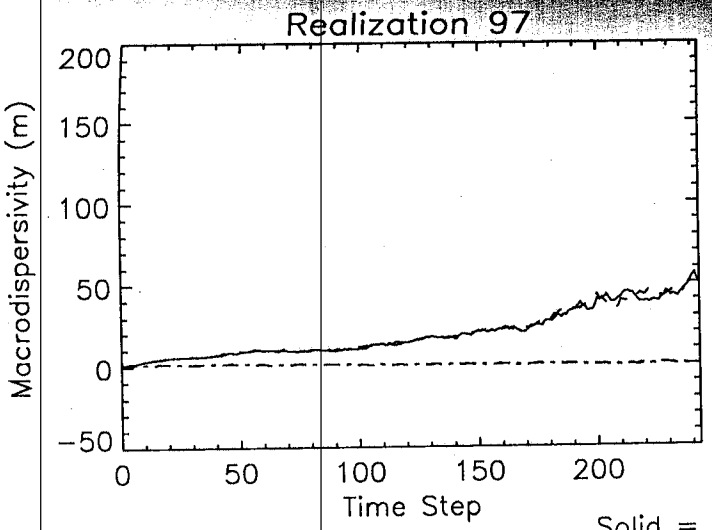


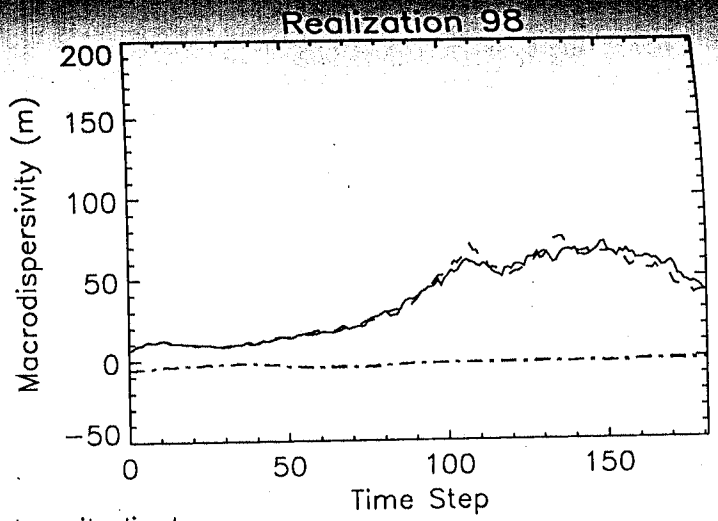


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

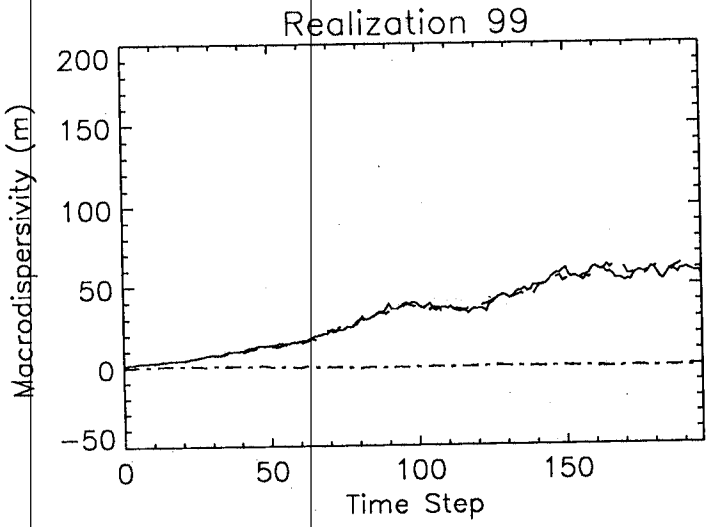


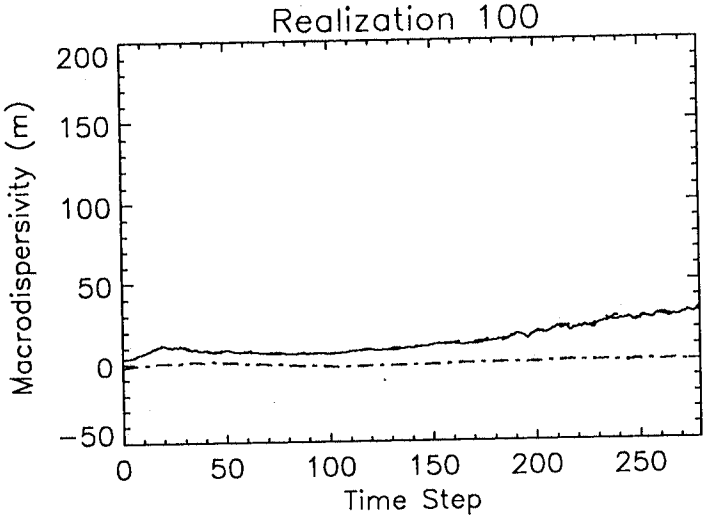






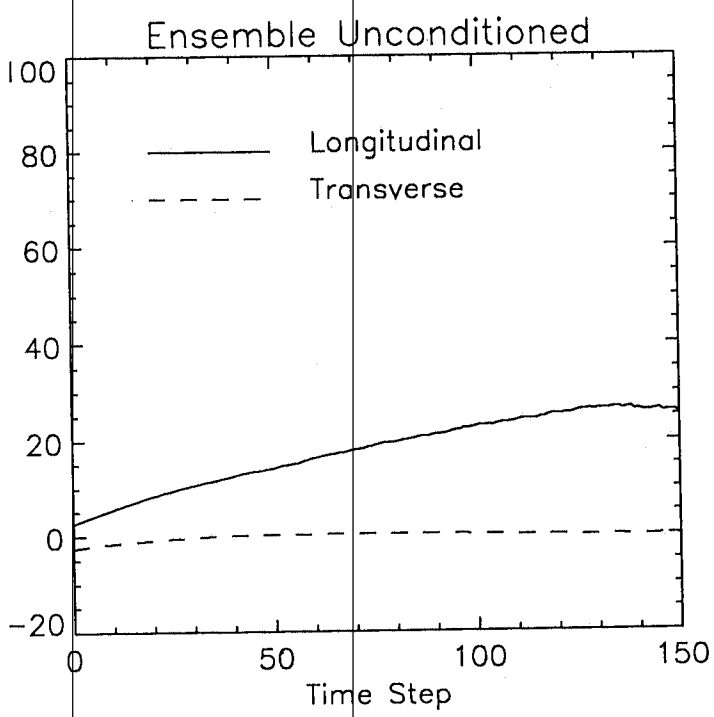
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

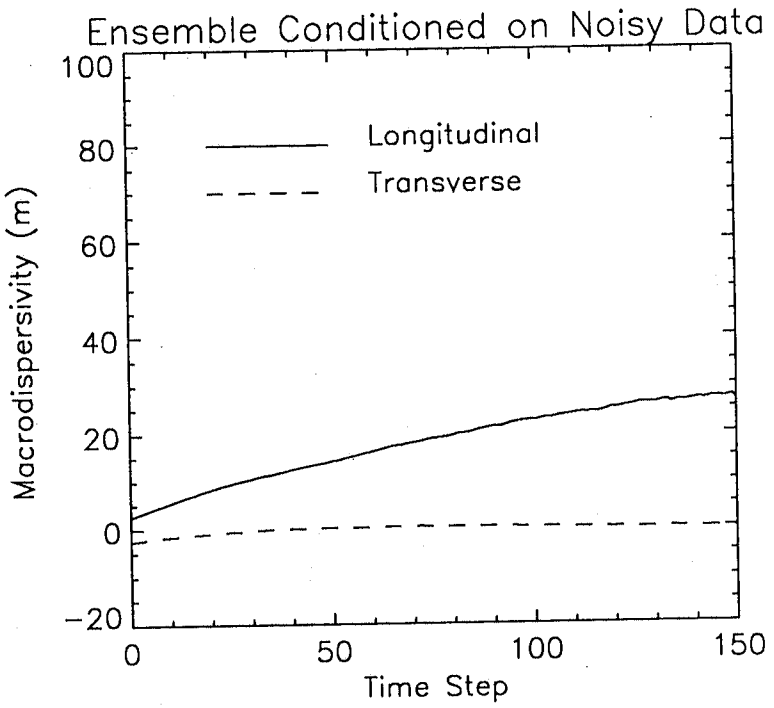




Appendix A: Macrodisersivity vs. Time Plots for Information Transfer Experiment

Figure A.2.
Ensemble Macrodispersivity vs. Time for 100 Single-Scale Realizations.


Ensemble Unconditioned
Longitudinal
Transverse
100
80
60
40
20
0
-20
0
50
100
150
Time Step


Ensemble Conditioned on Noisy Data
Longitudinal
Transverse
Macrodispersivity (m)
100
80
60
40
20
0
-20
0
50
100
150
Time Step

Appendix A: Macrodisersivity vs. Time Plots for Information Transfer Experiment

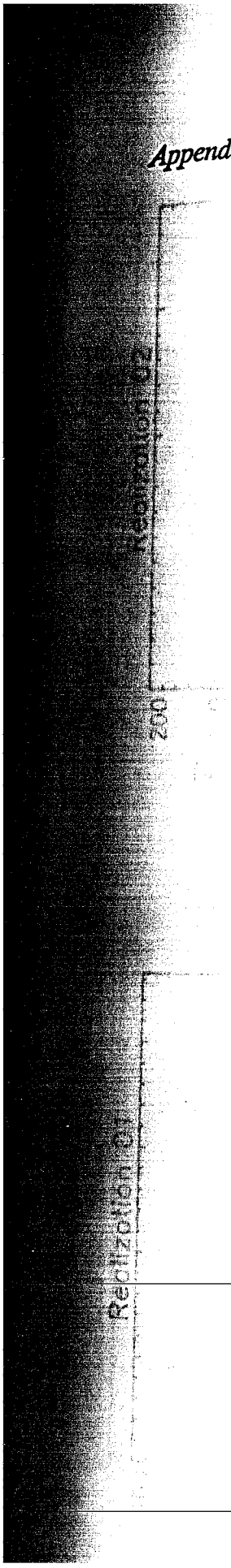
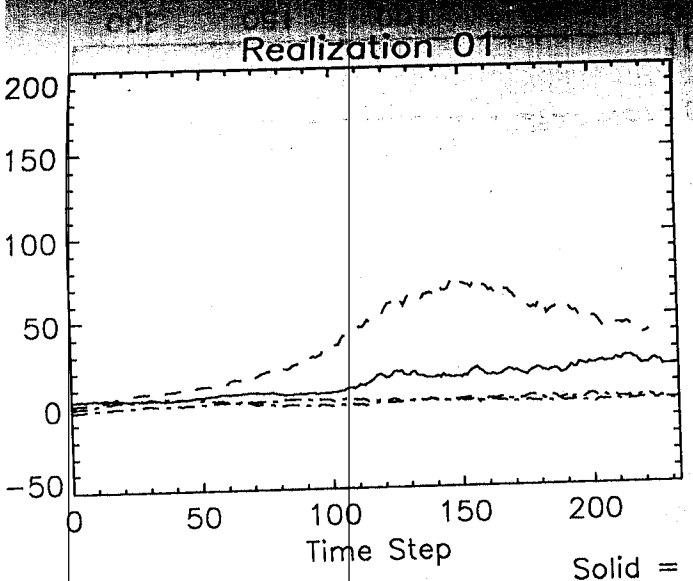
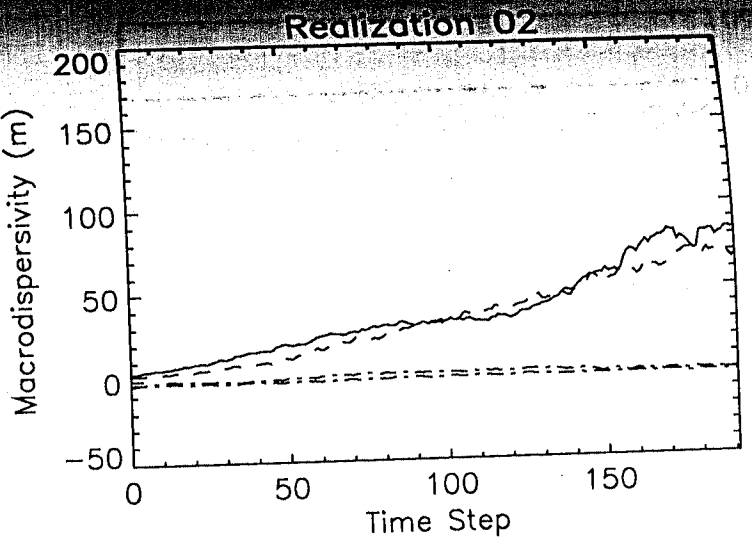
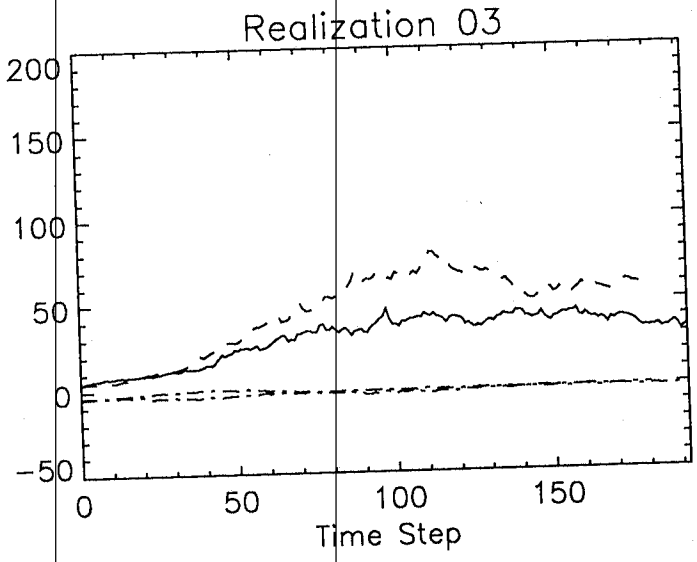


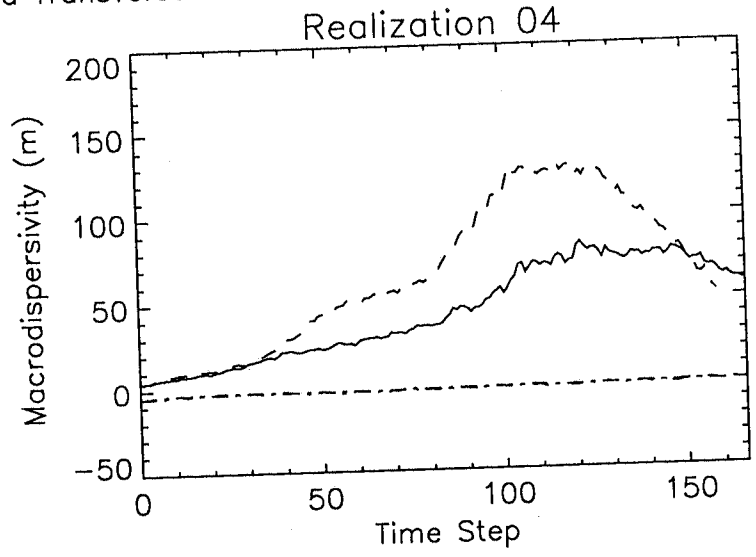
Figure A.3. Macrodispersivity vs. Time for 100 Two-Scale Realizations.

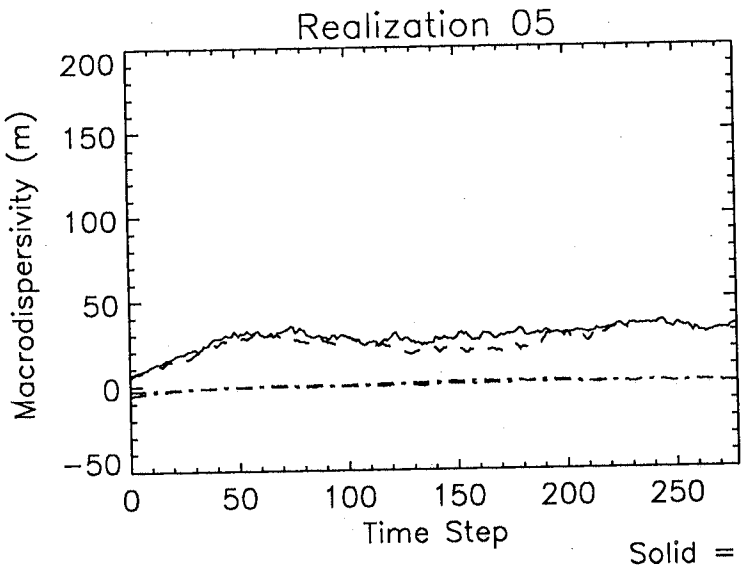


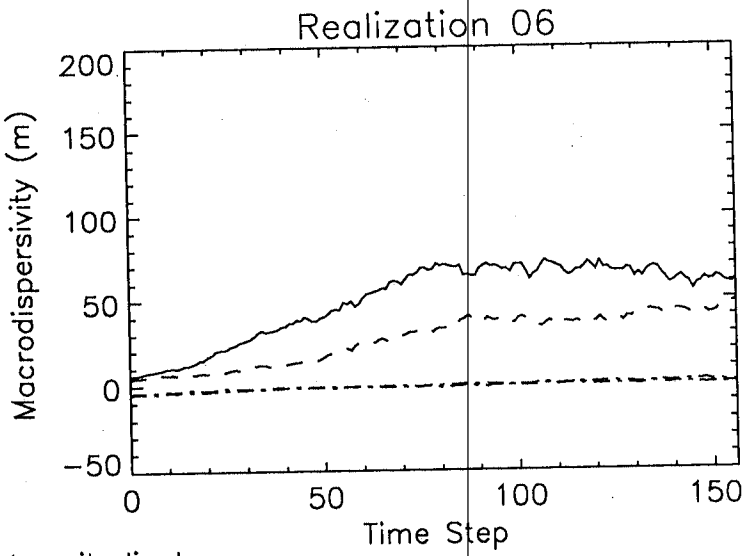


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

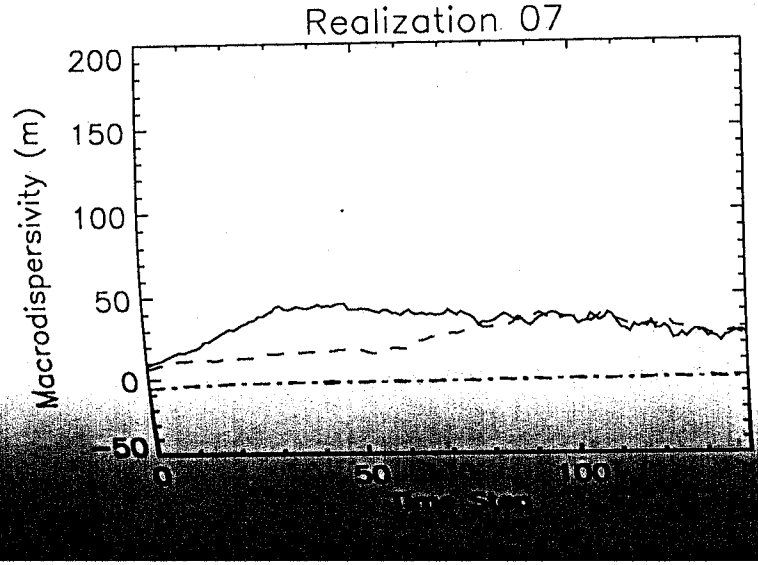


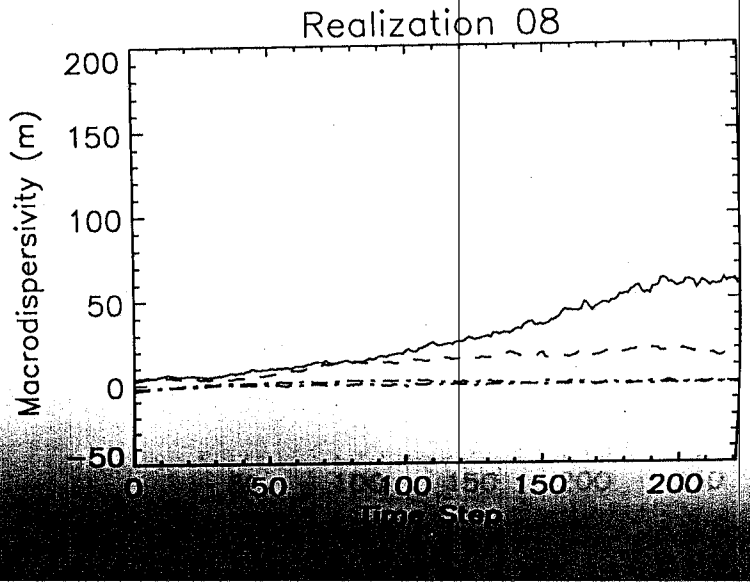


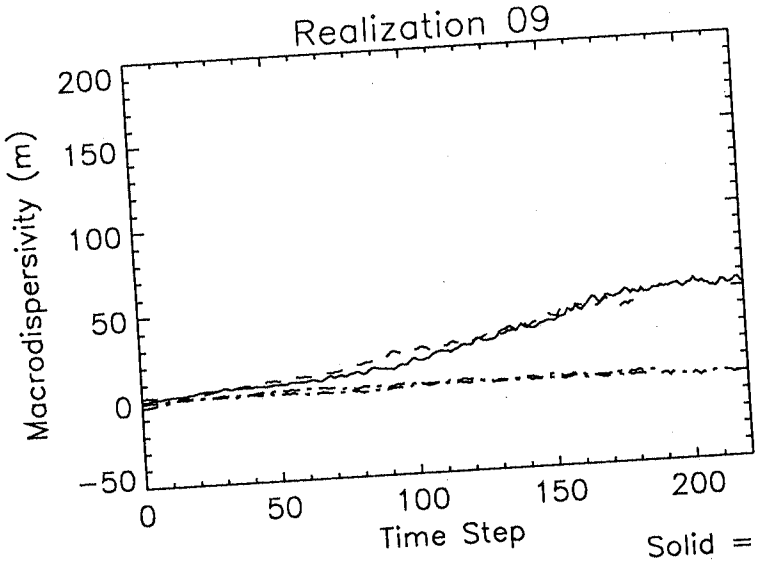


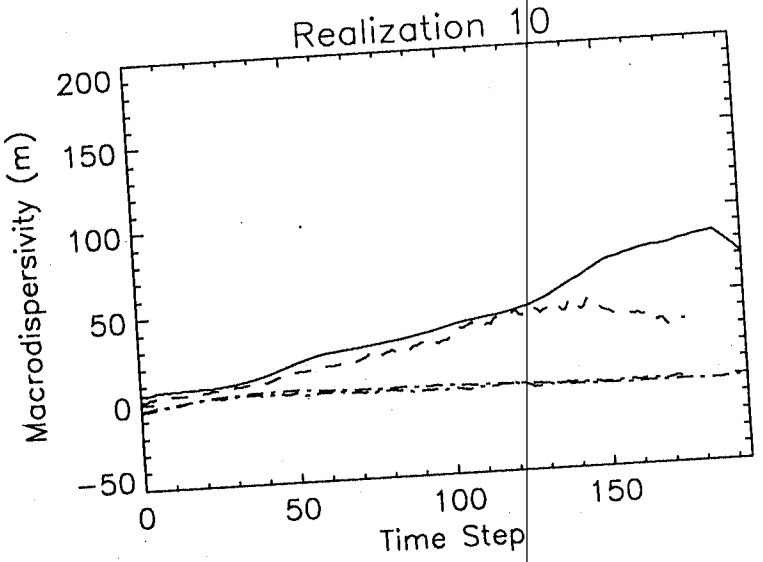


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

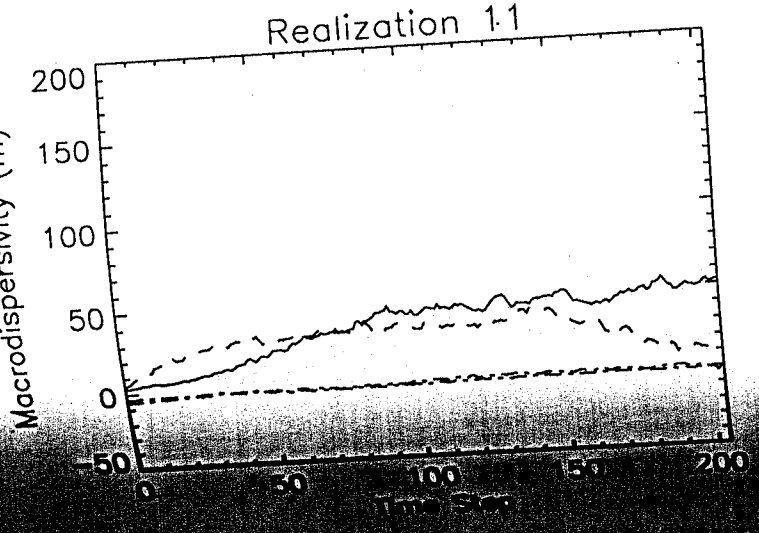


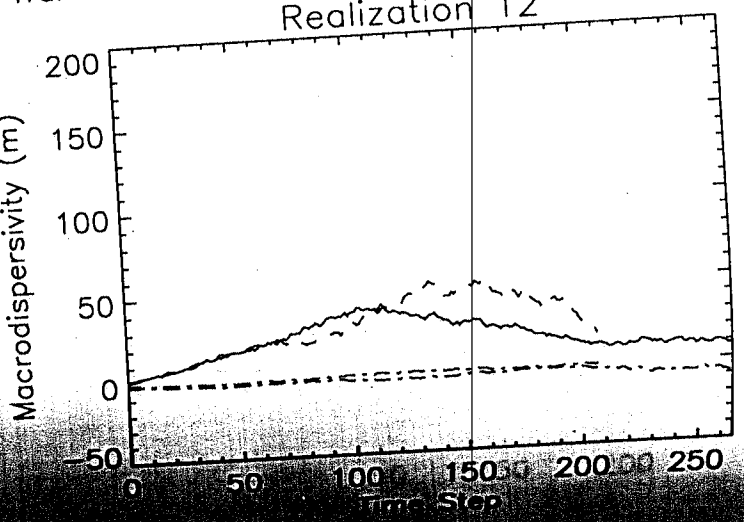


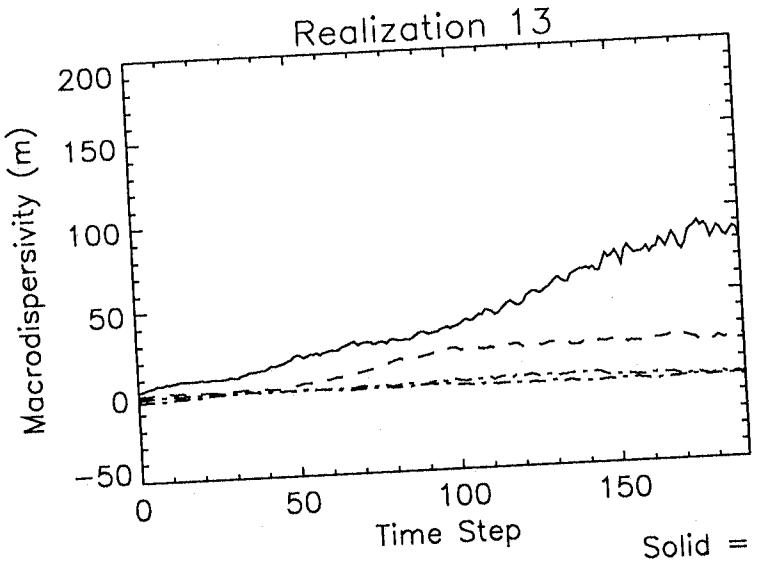


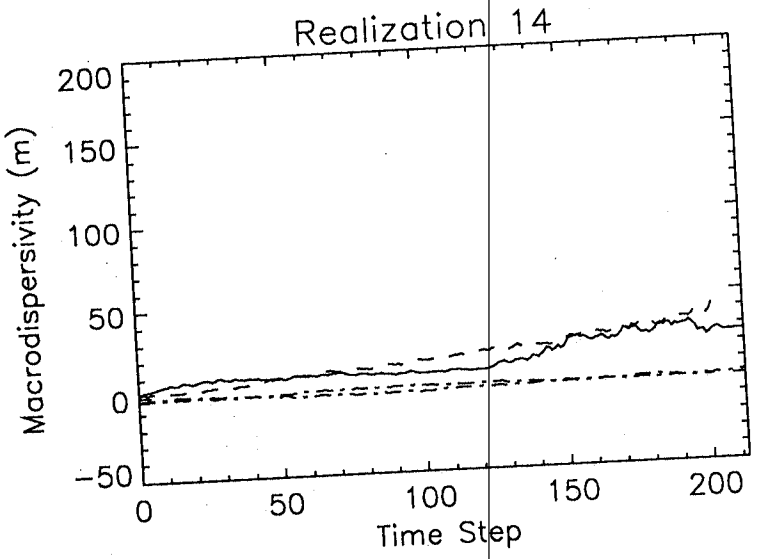


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

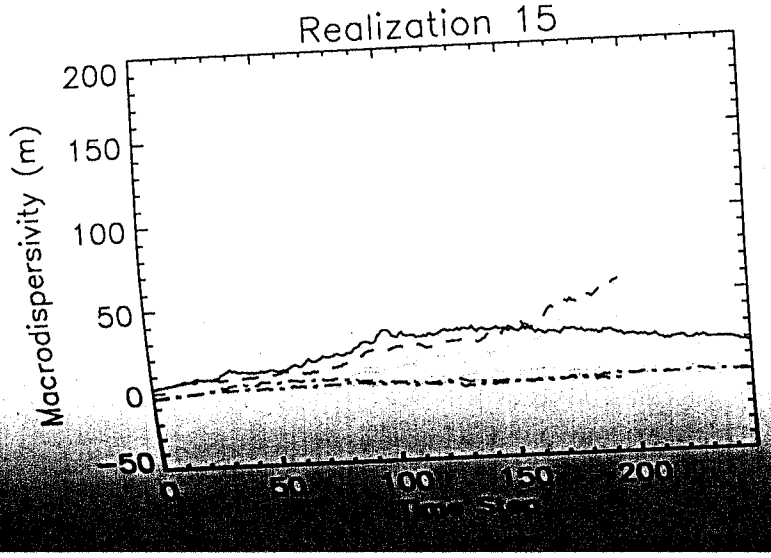


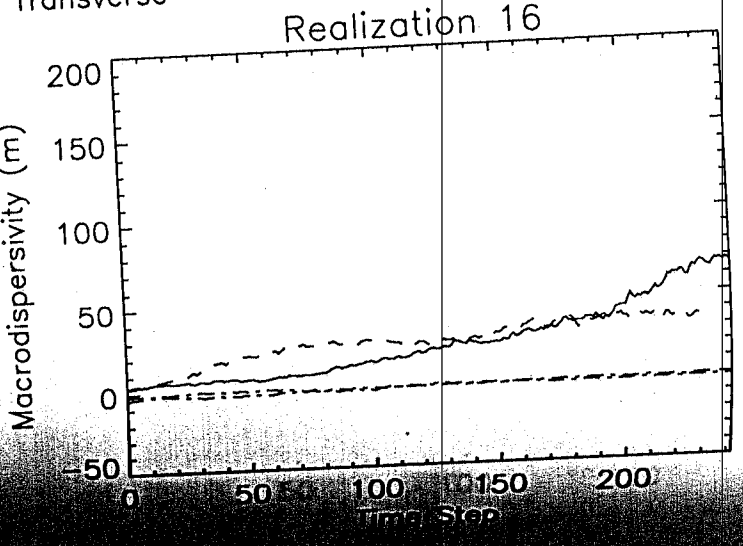


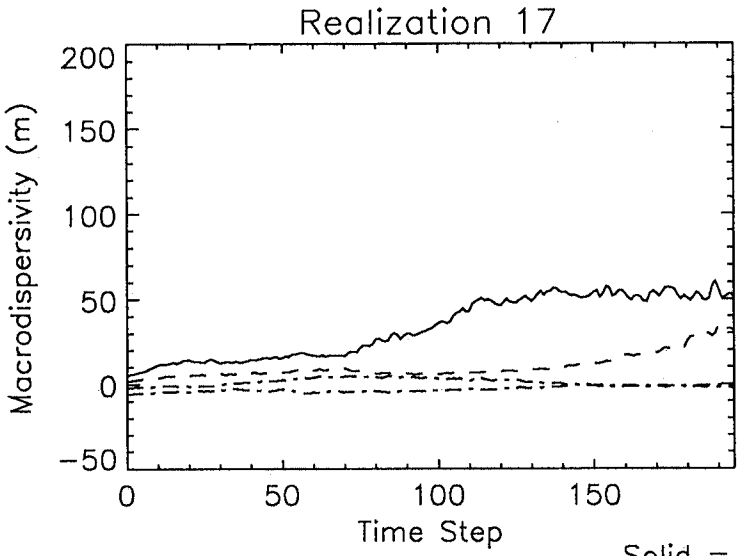


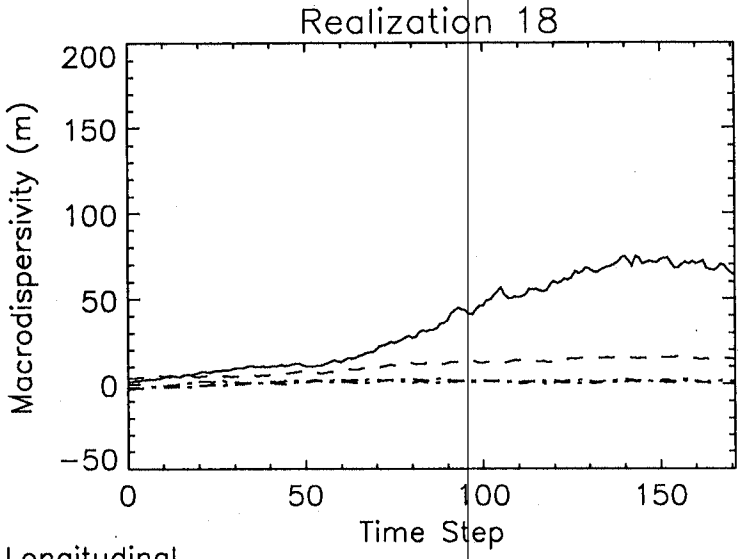


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

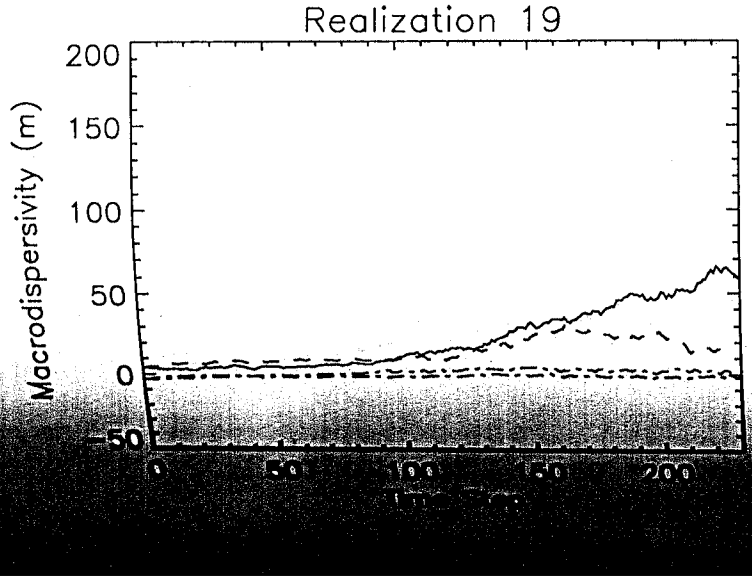


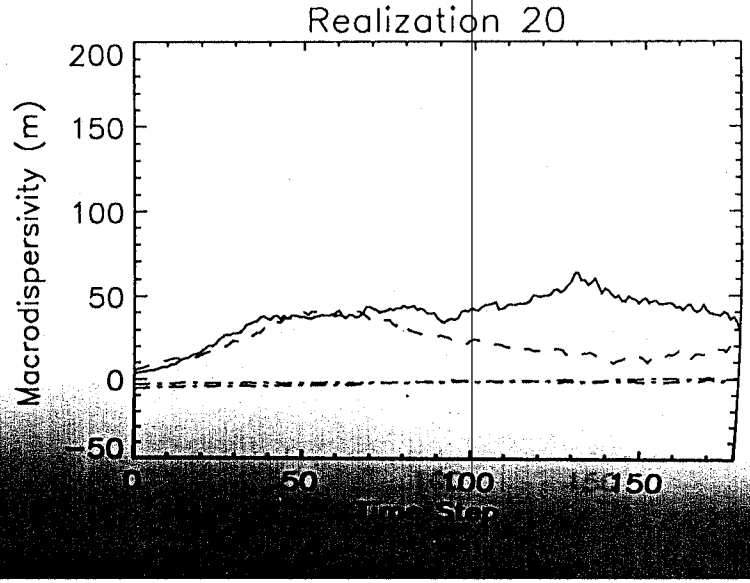


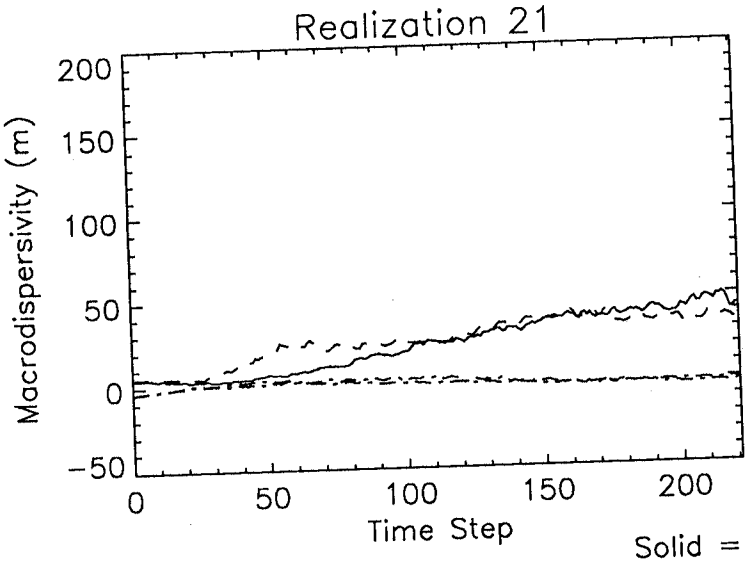


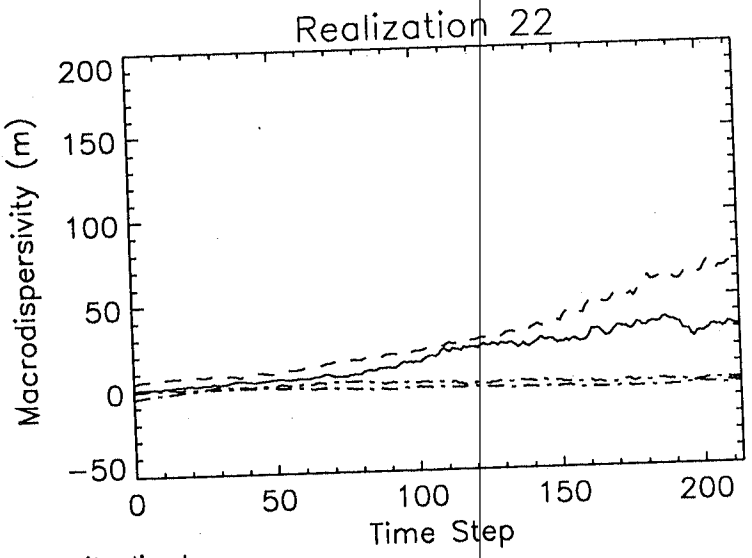


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

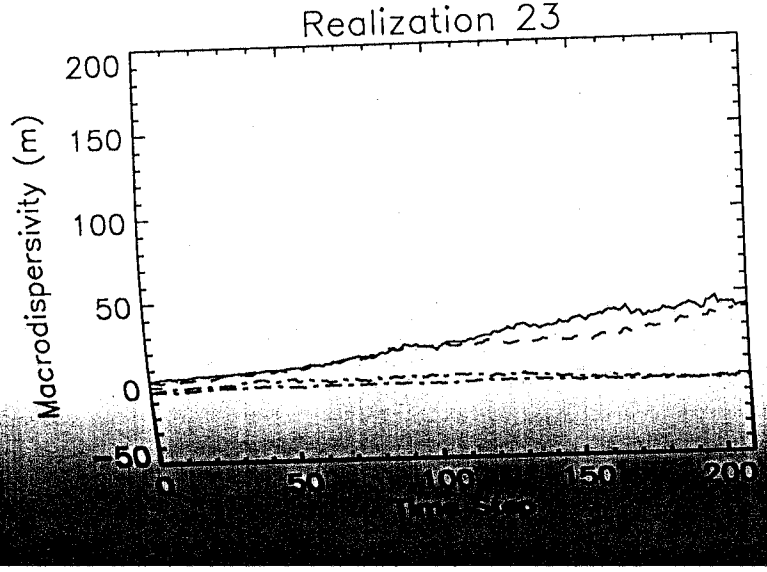


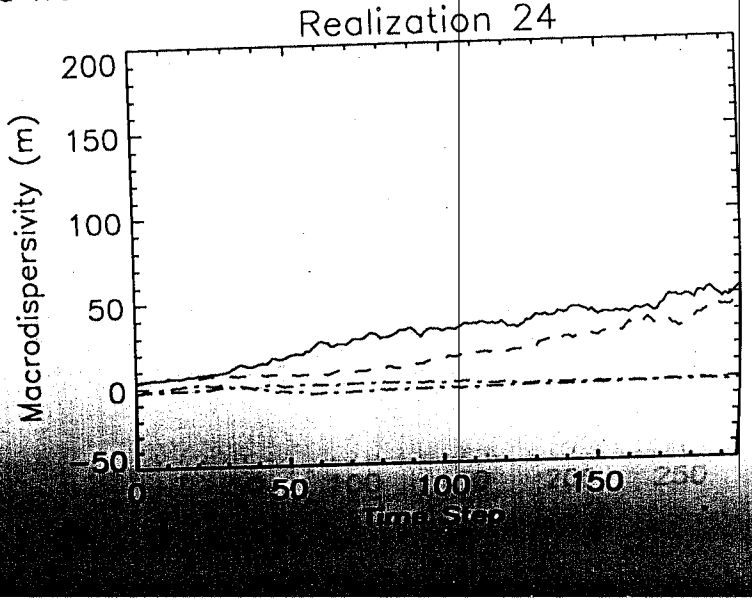


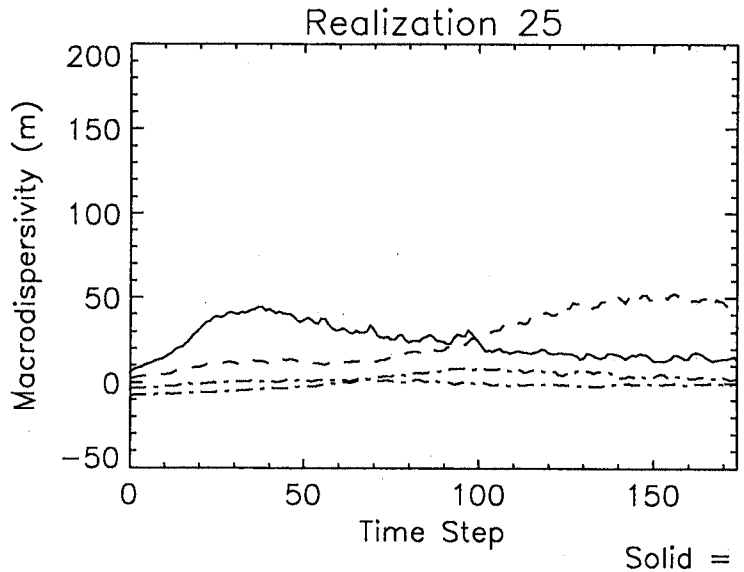


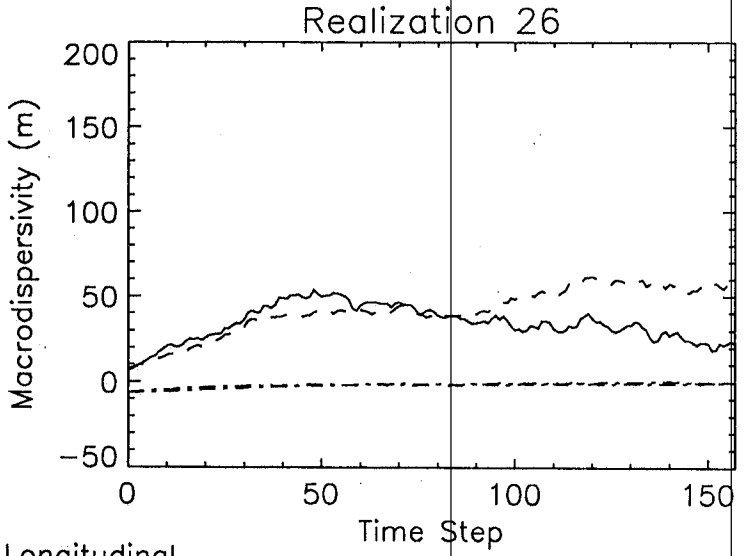


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

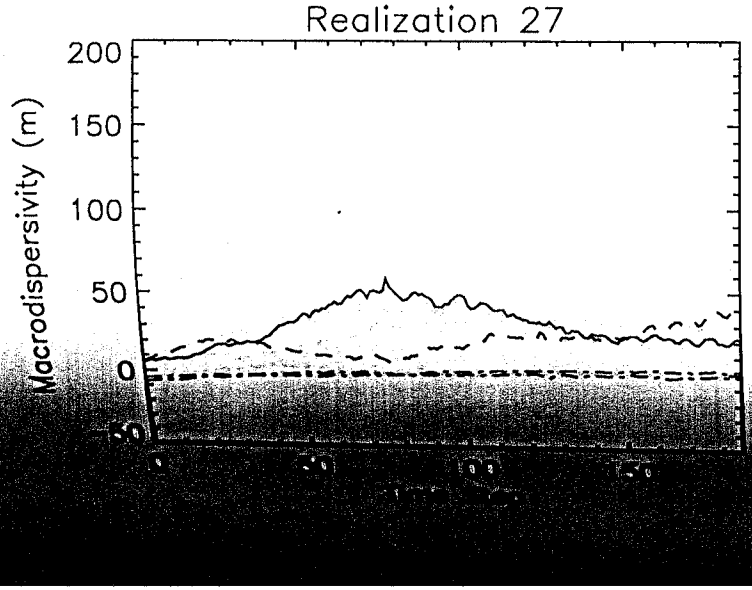


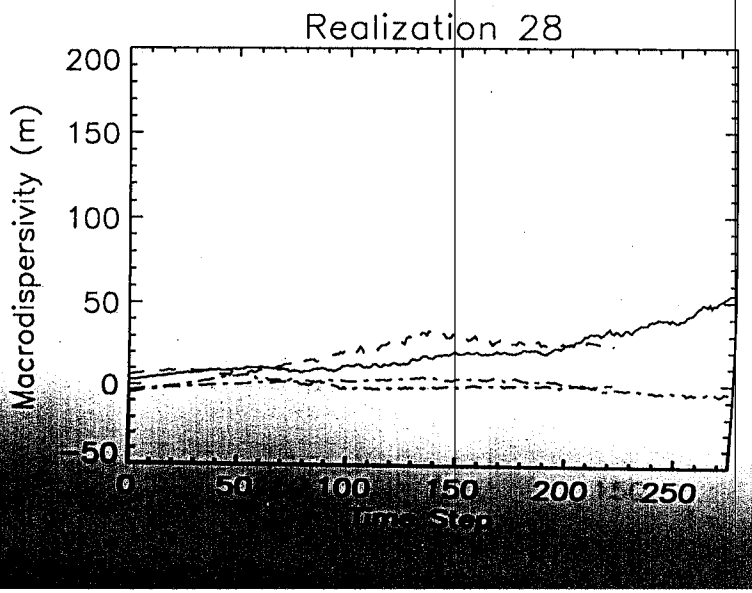


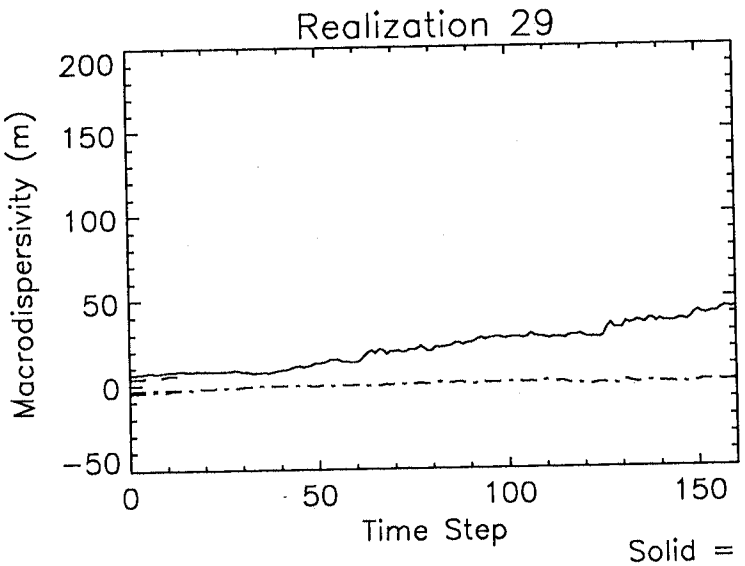


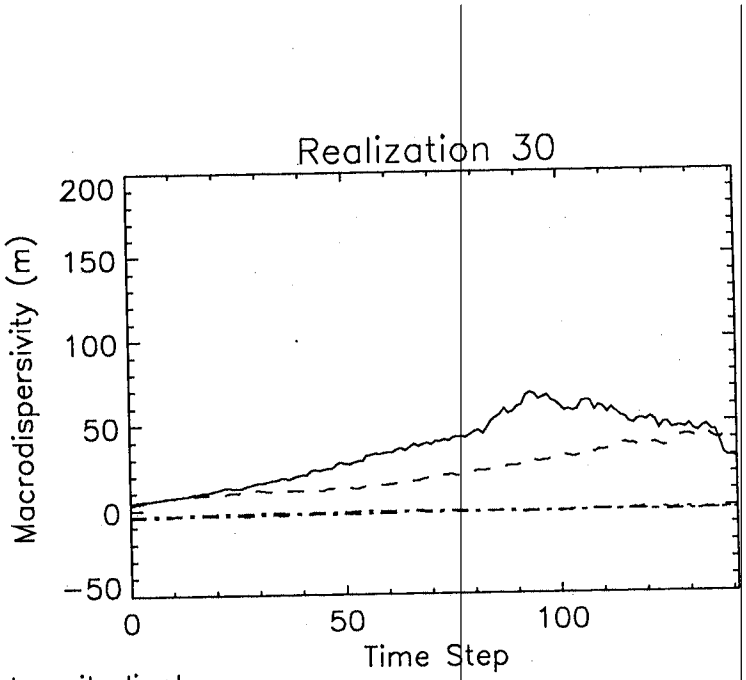


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

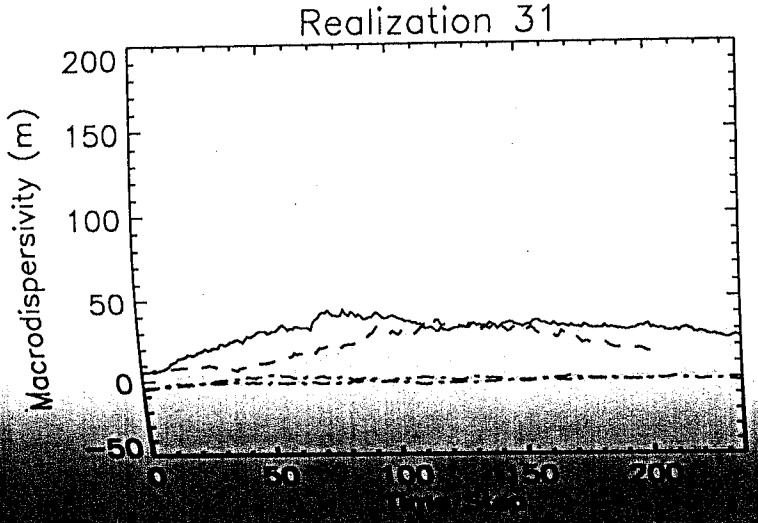


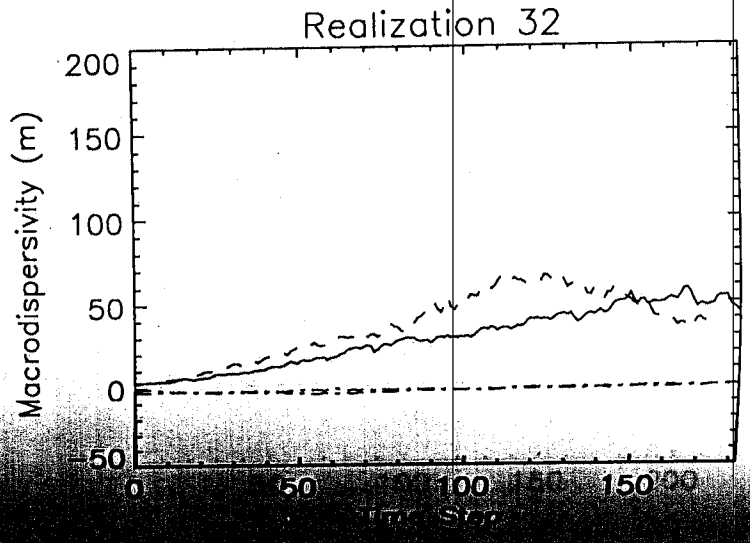


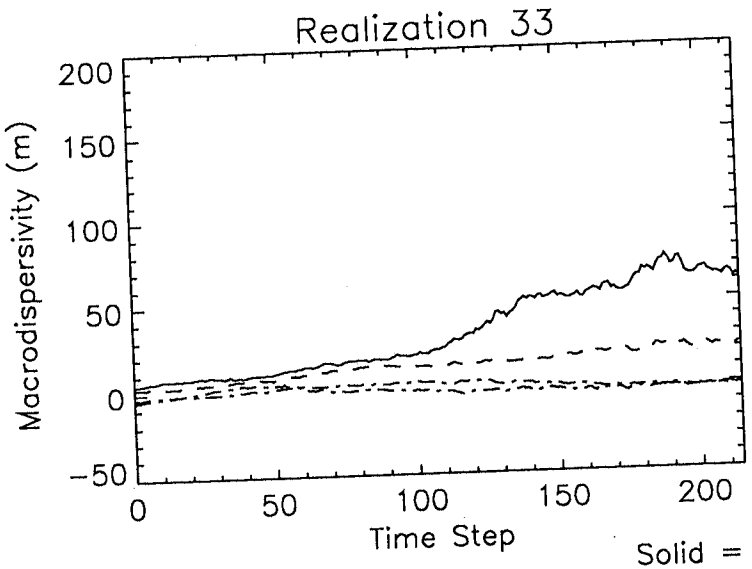


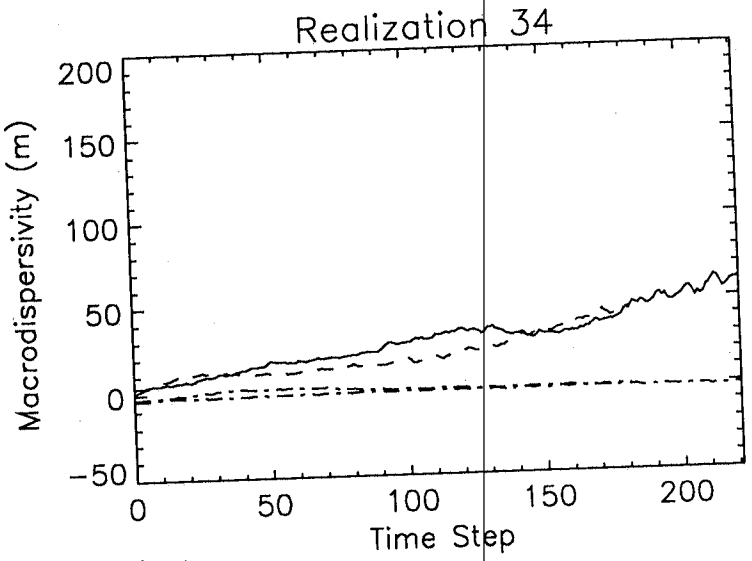


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

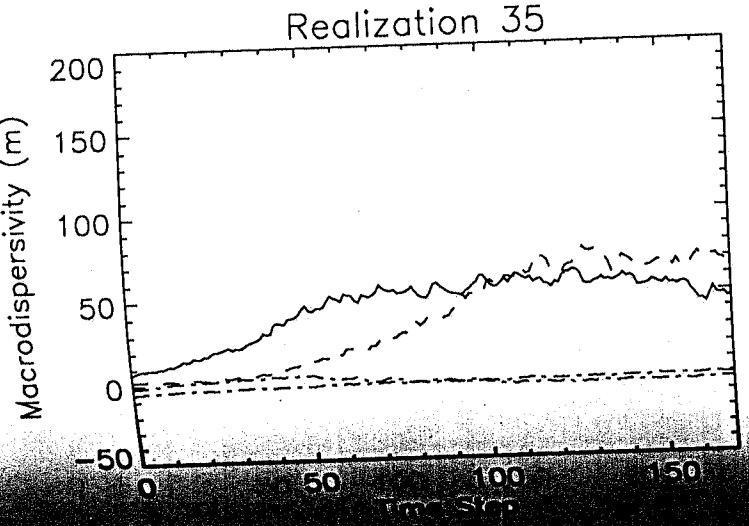


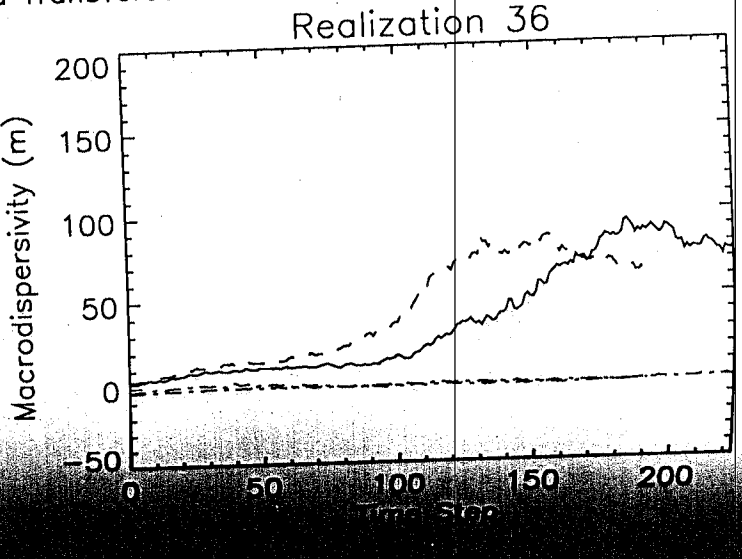


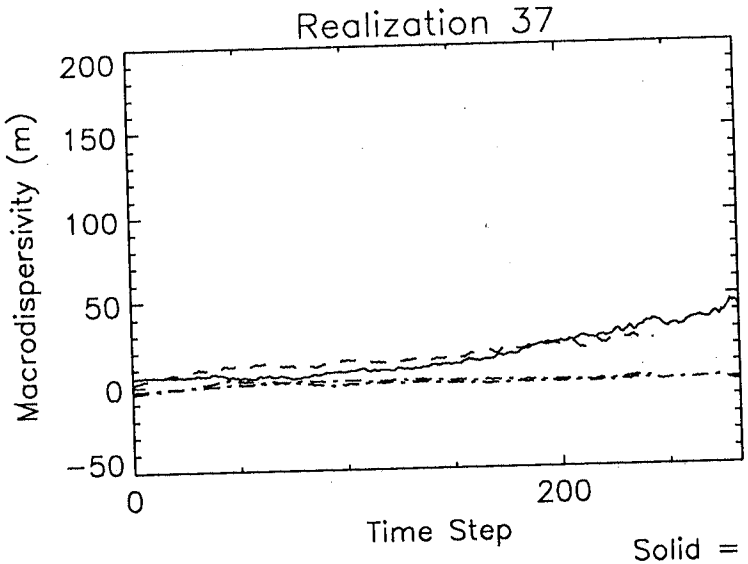


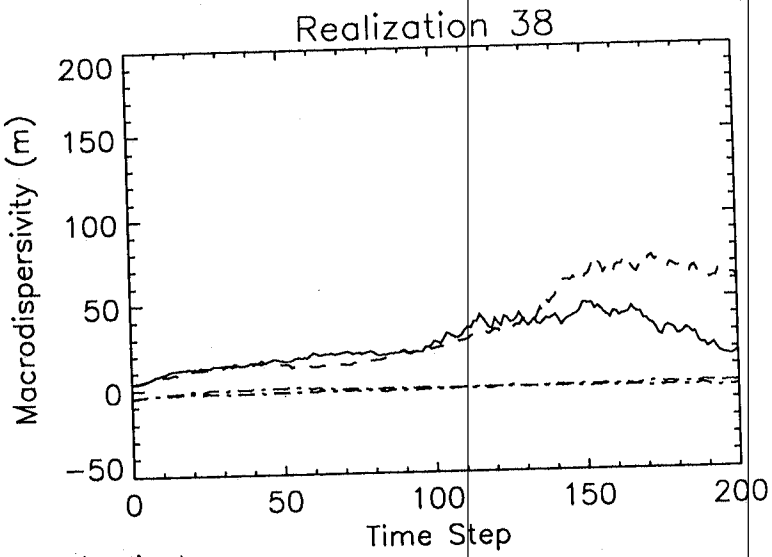


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

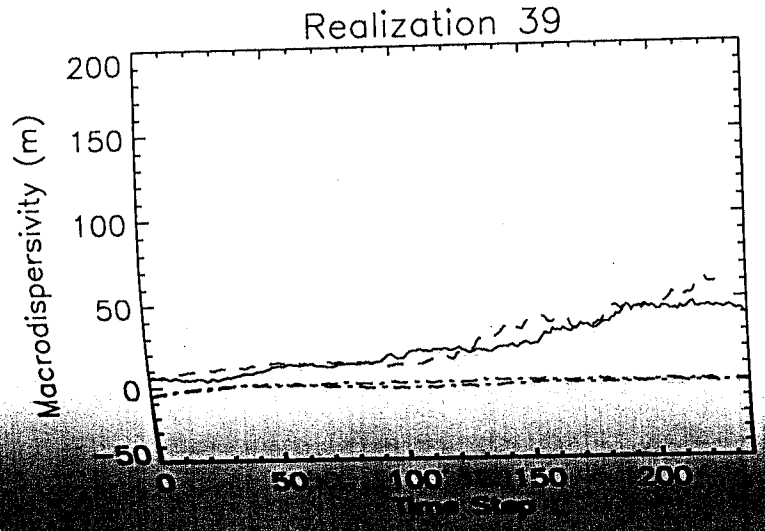


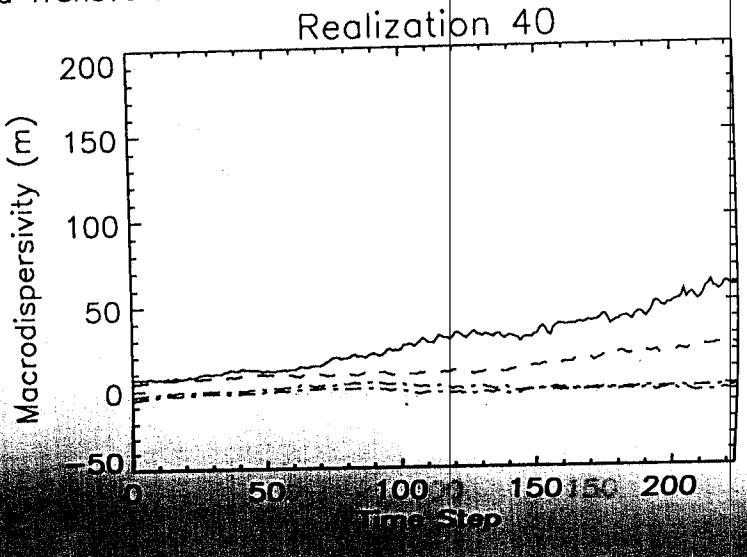


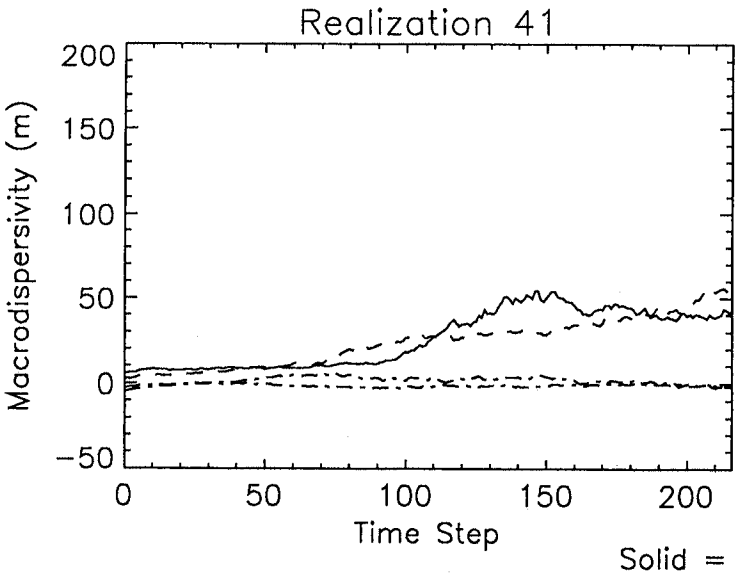


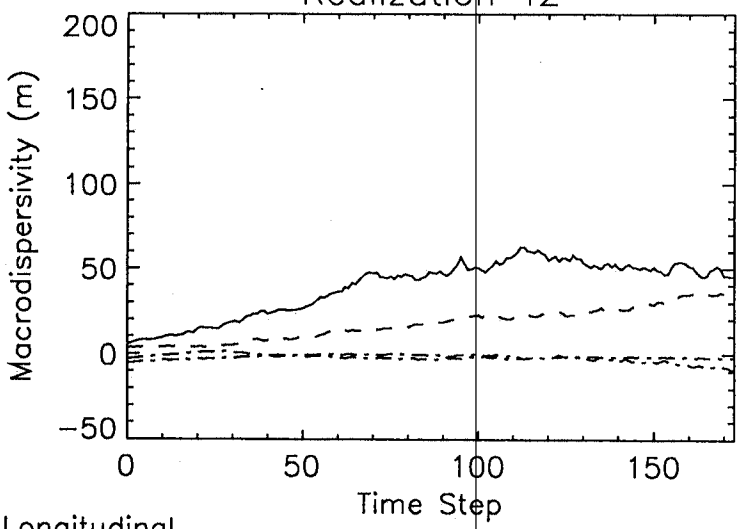


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

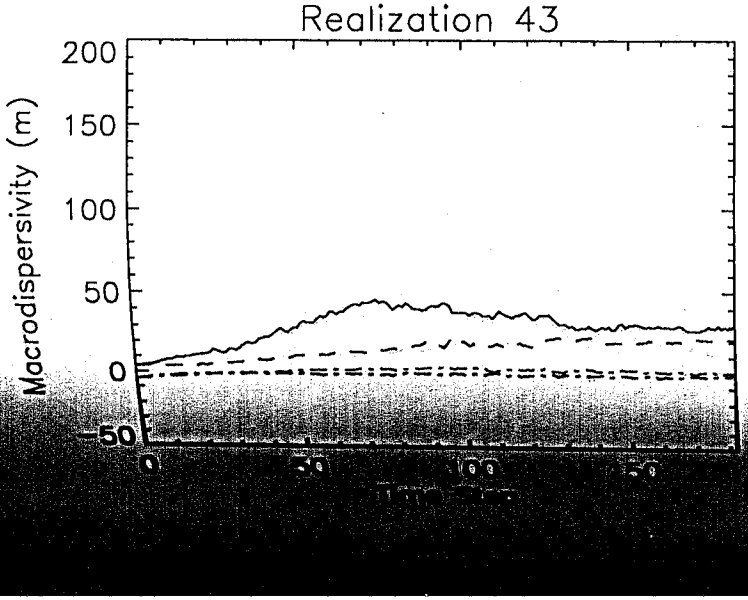


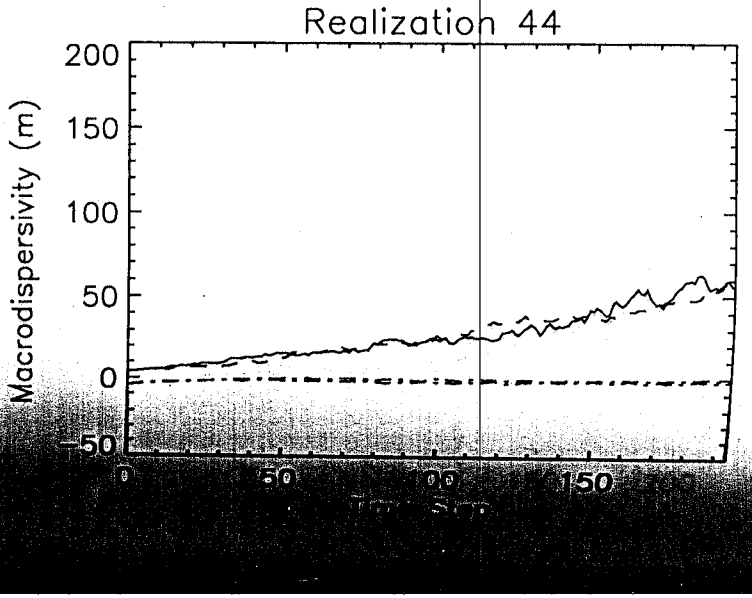


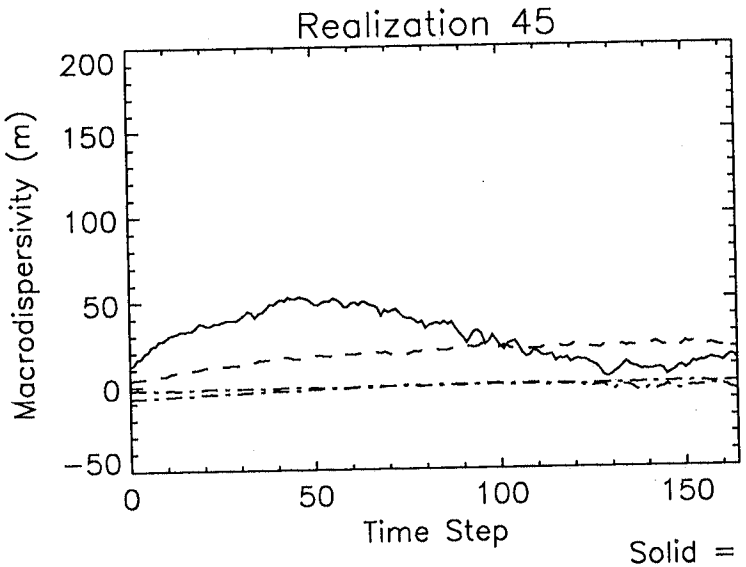


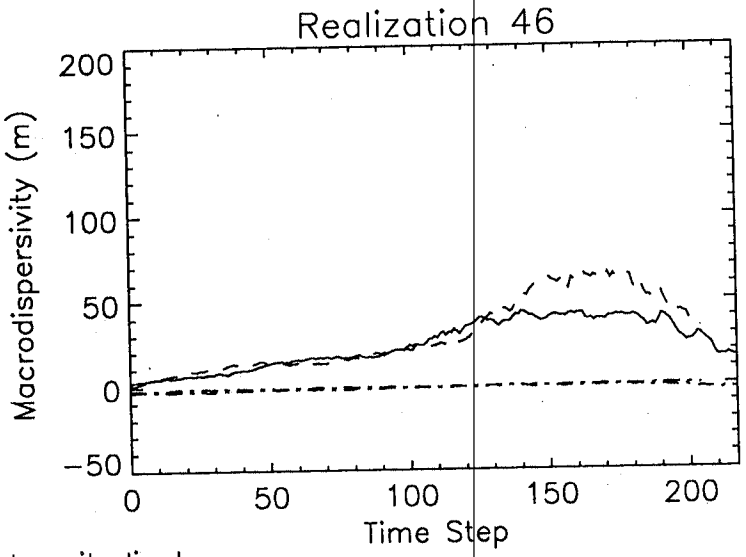


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

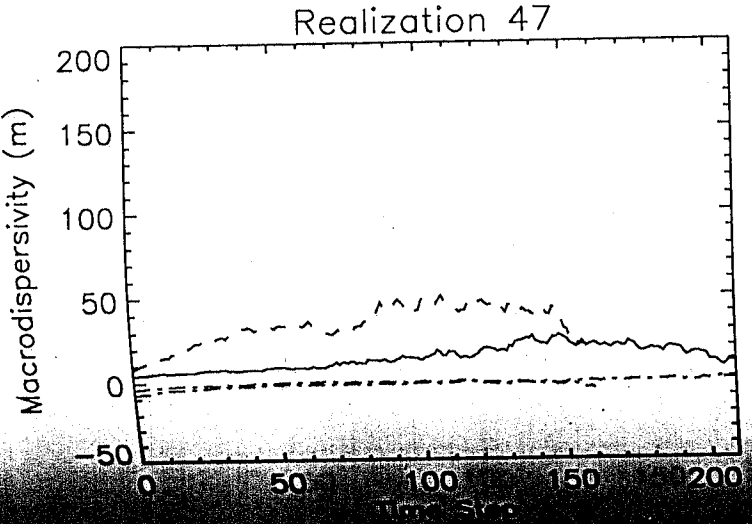


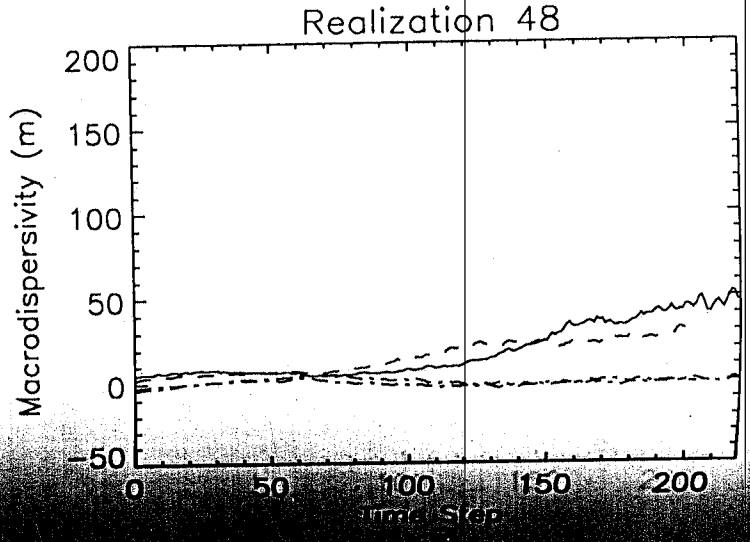


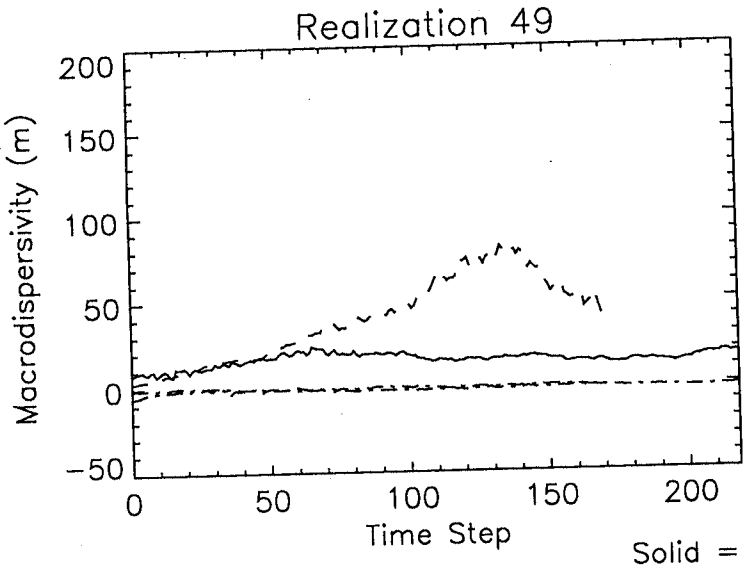


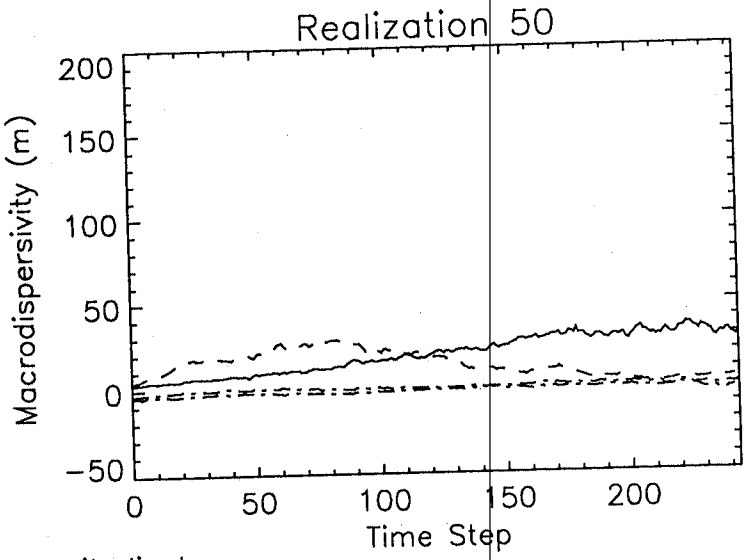


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

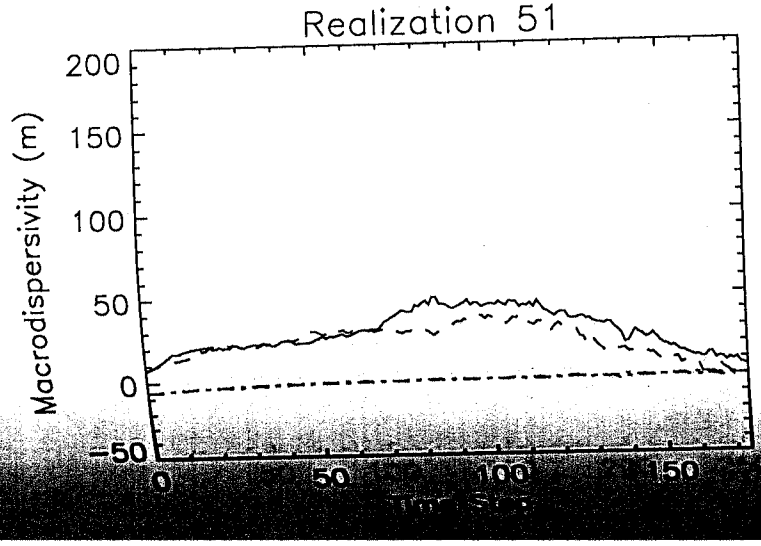


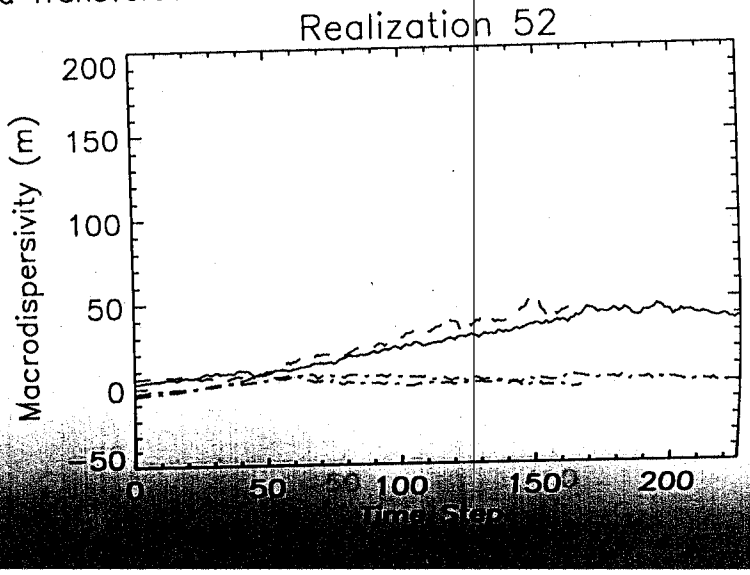


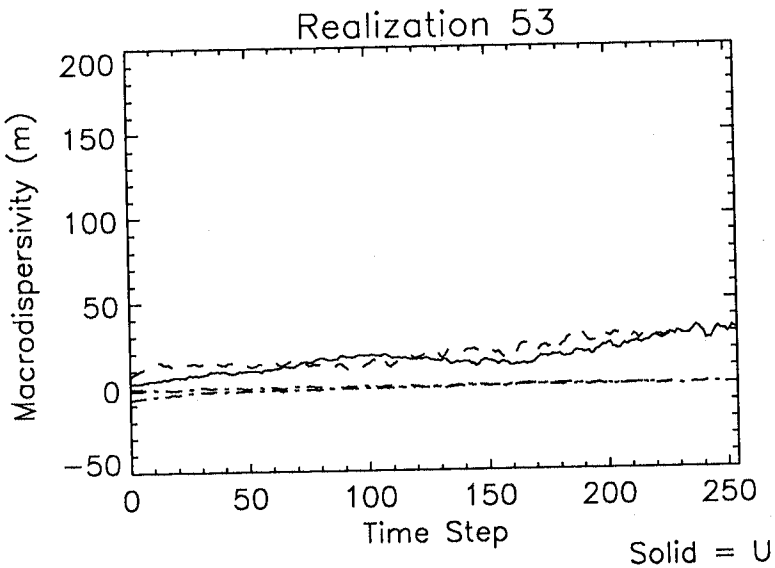


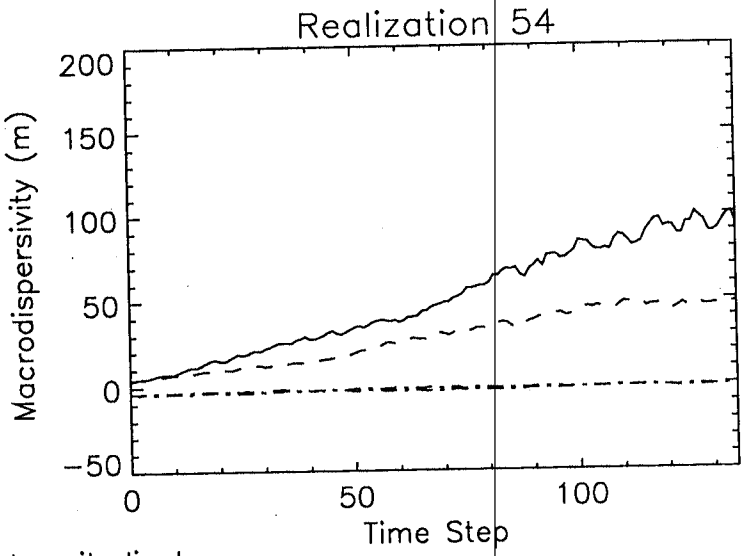


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

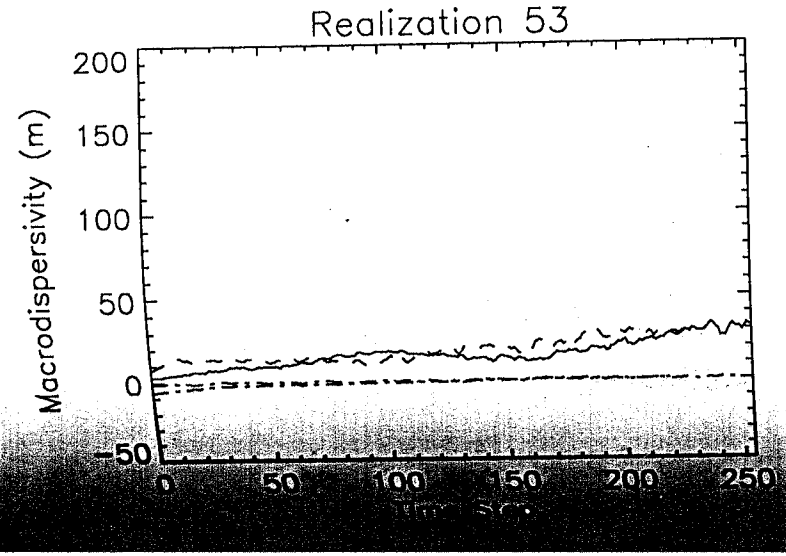


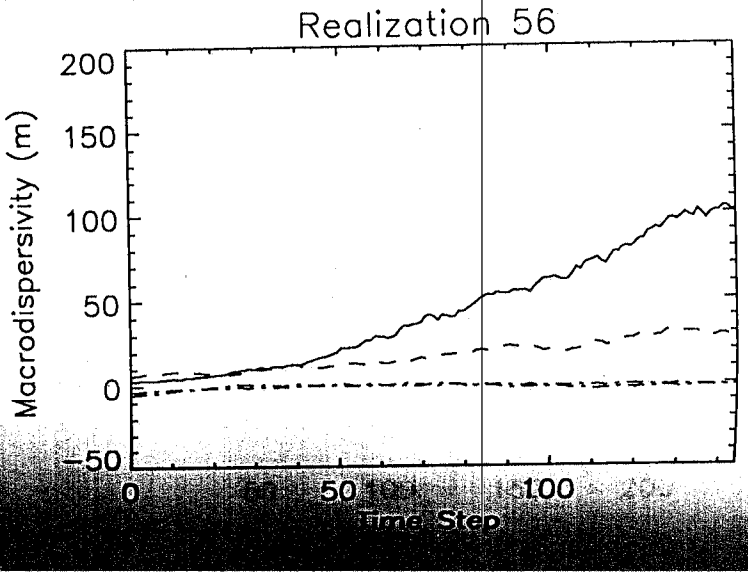


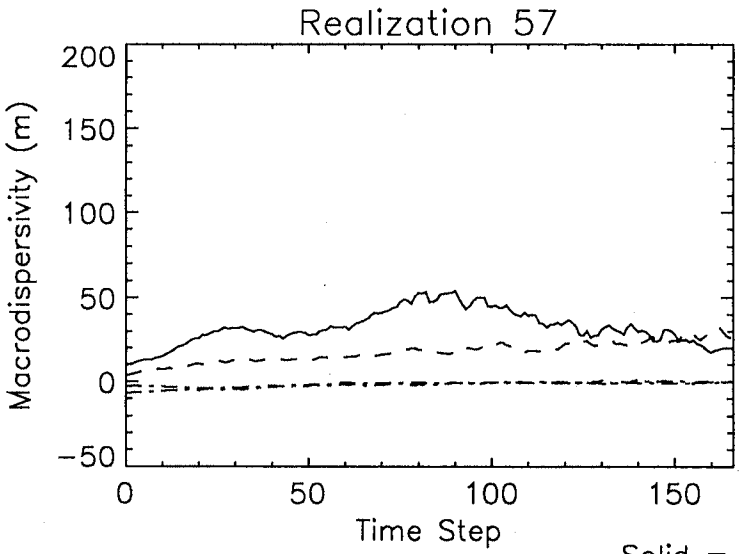


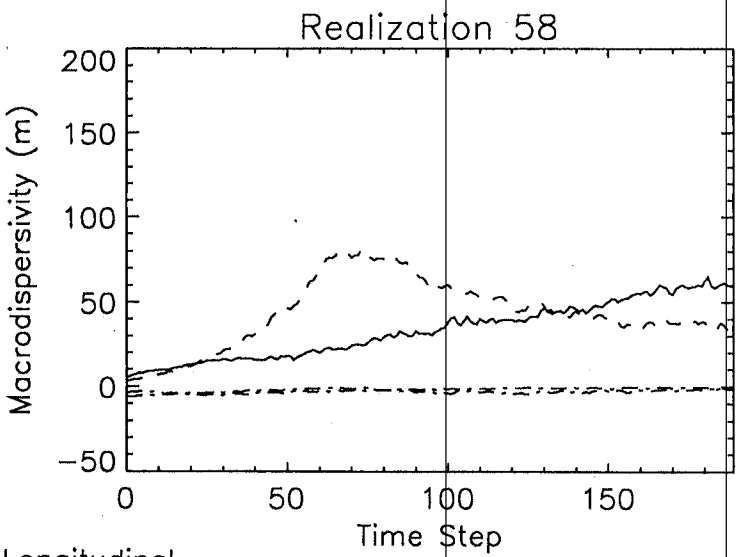


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

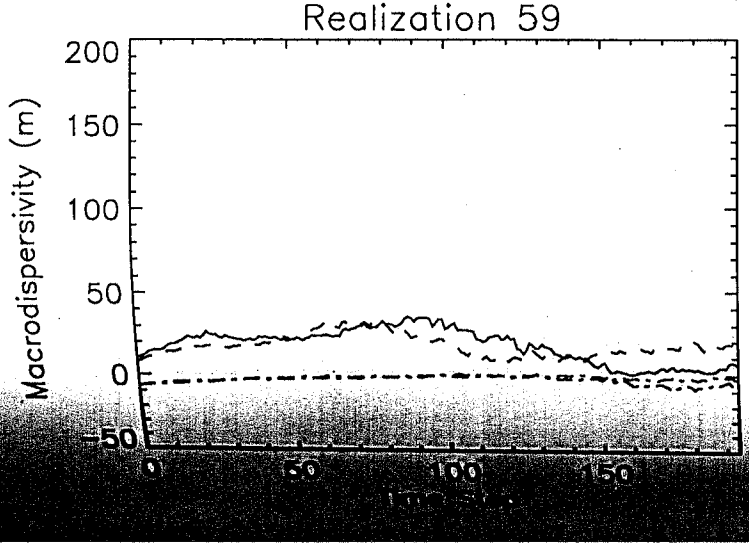


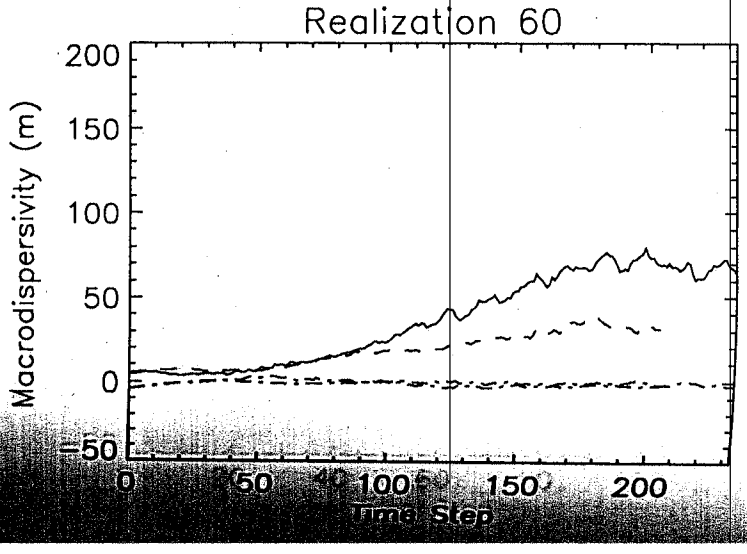


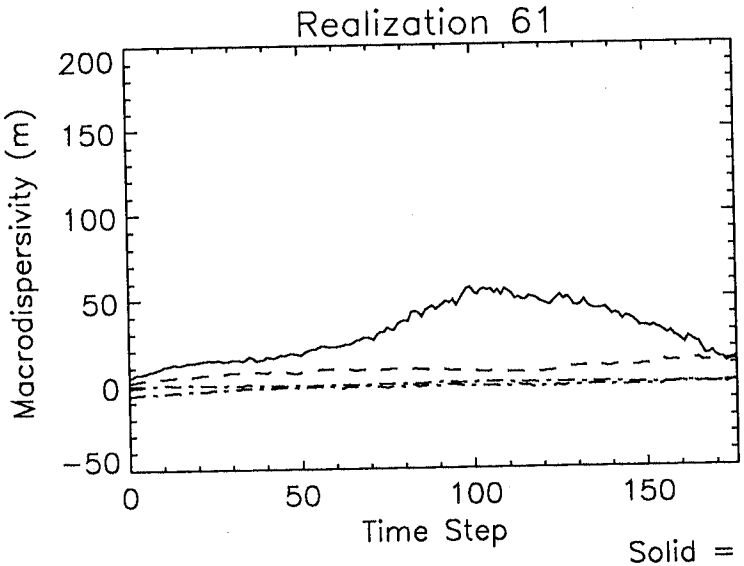


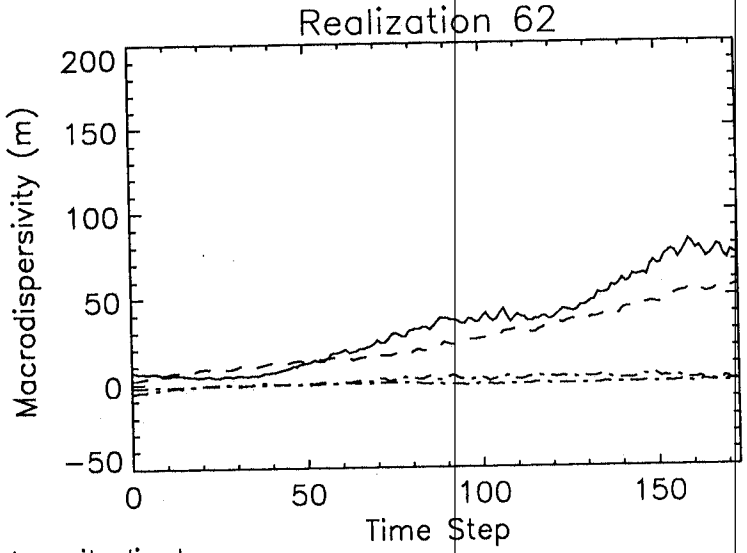


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

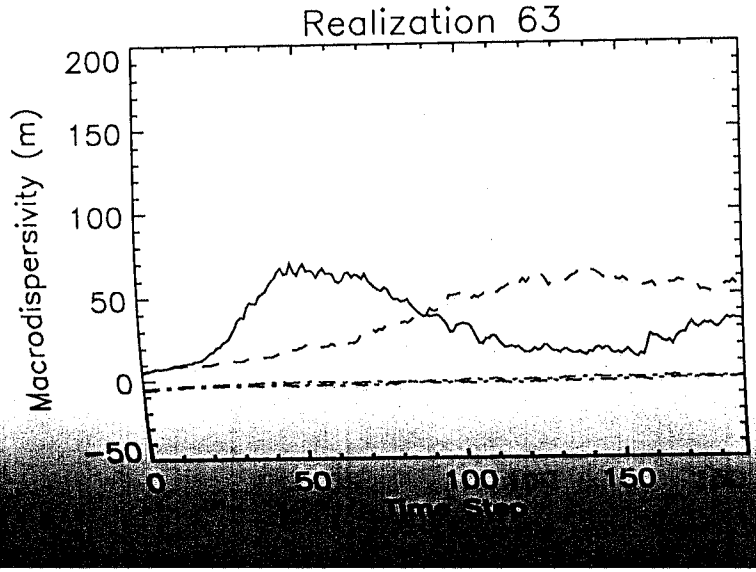


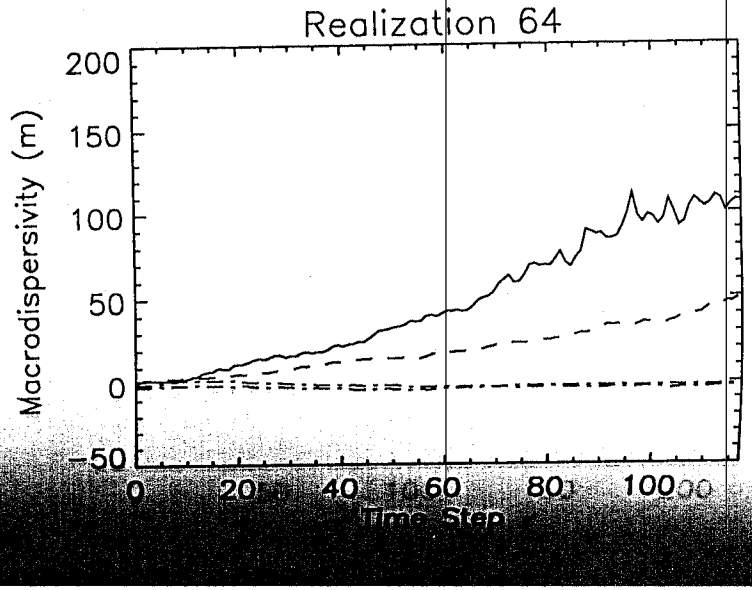


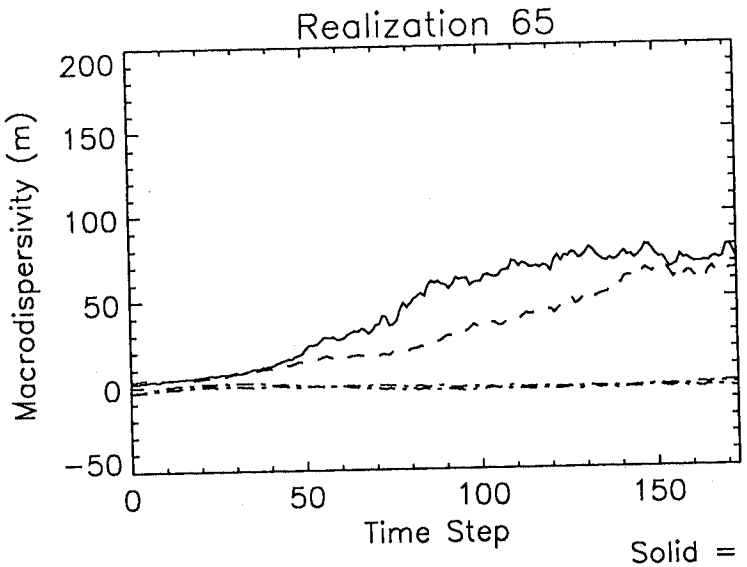


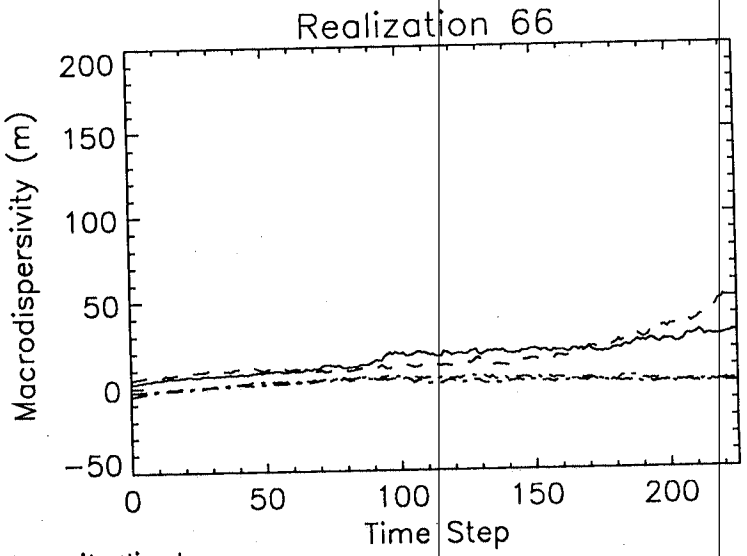


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

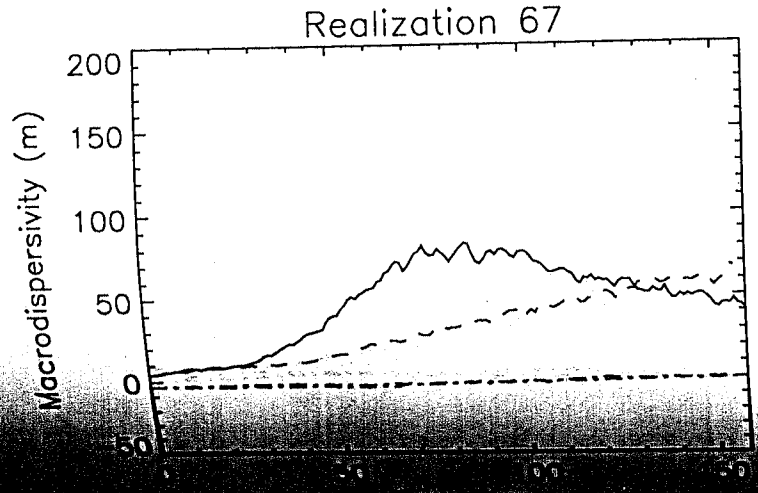


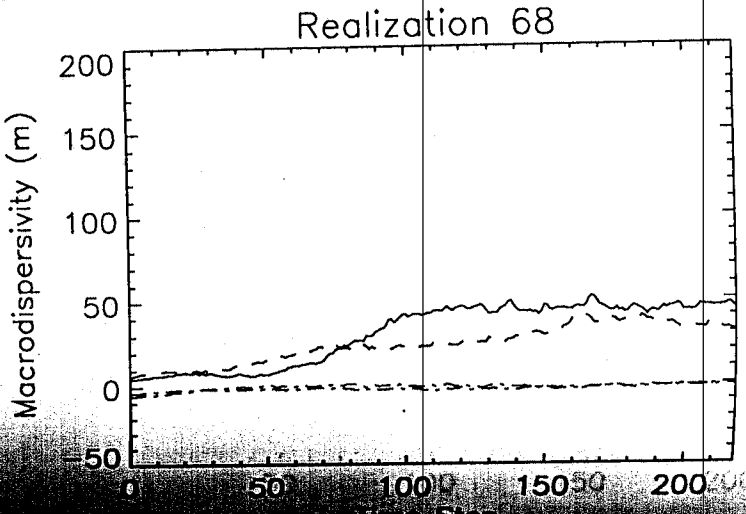


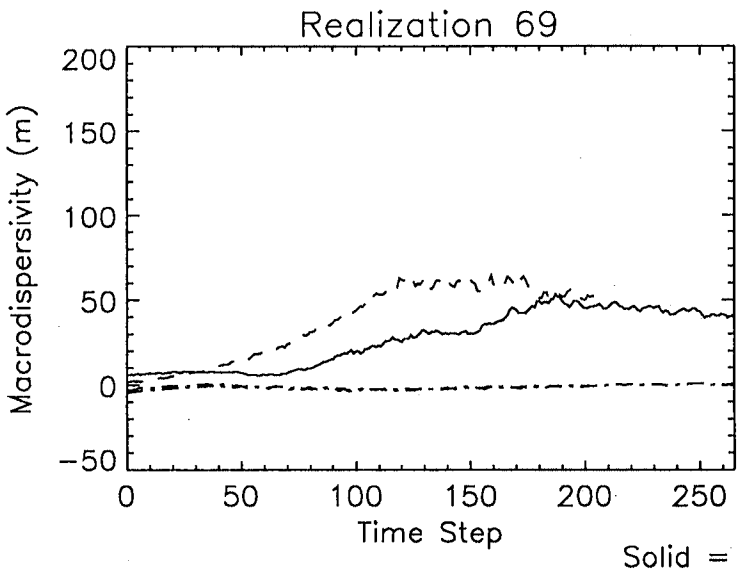


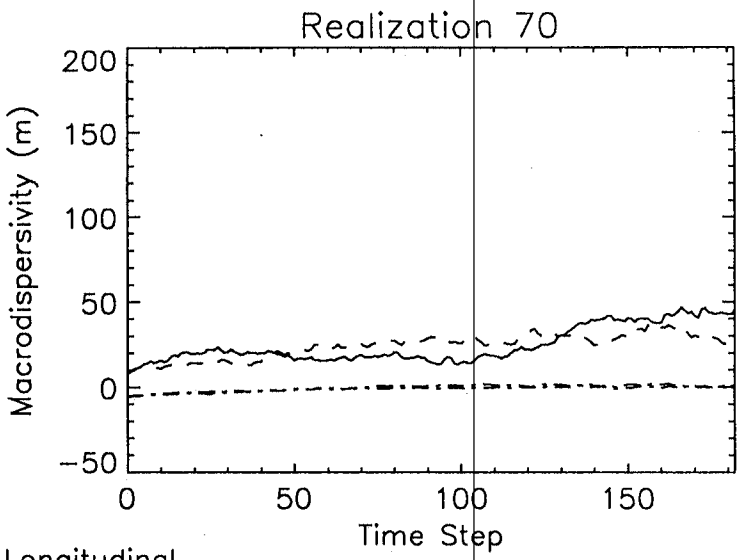


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

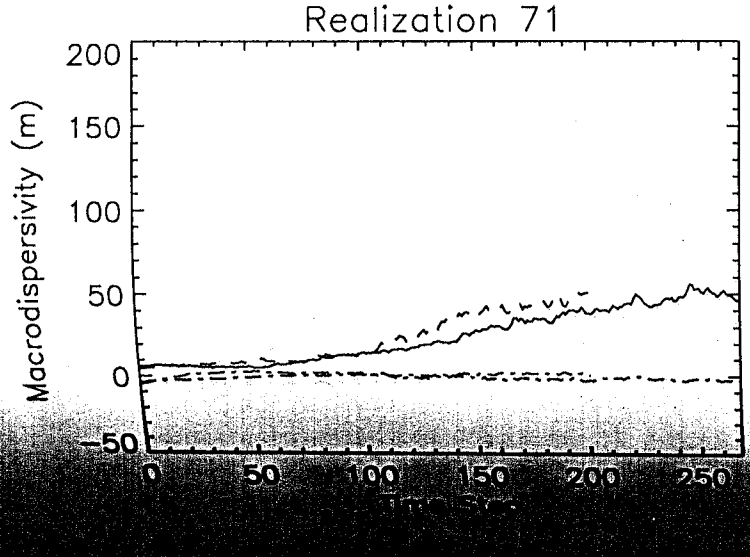


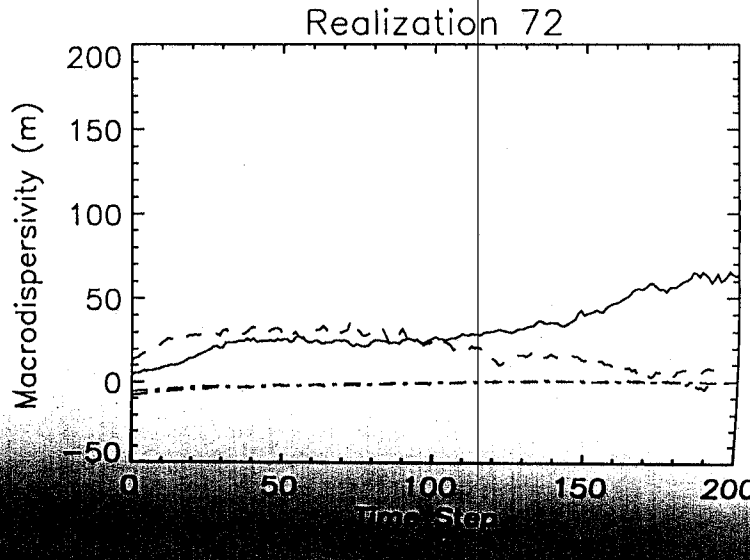


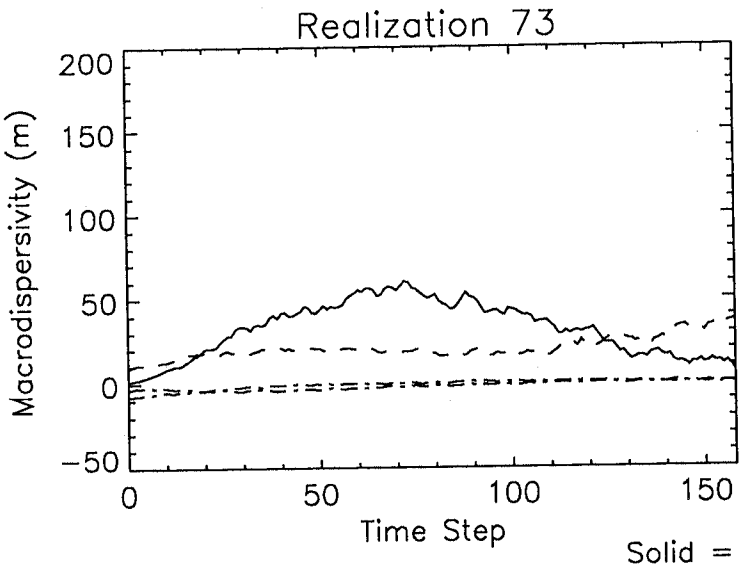


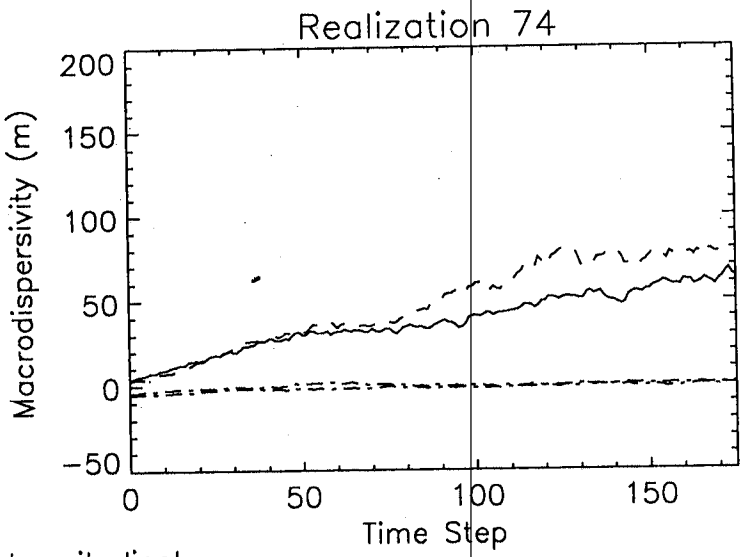


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

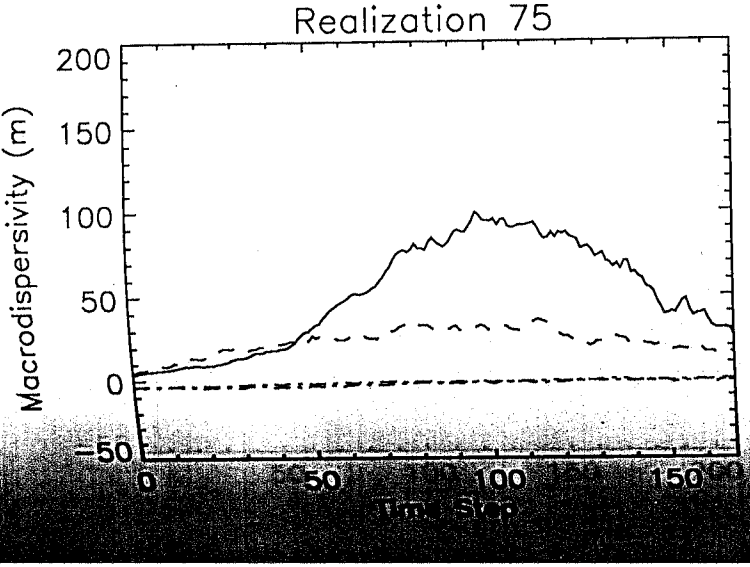


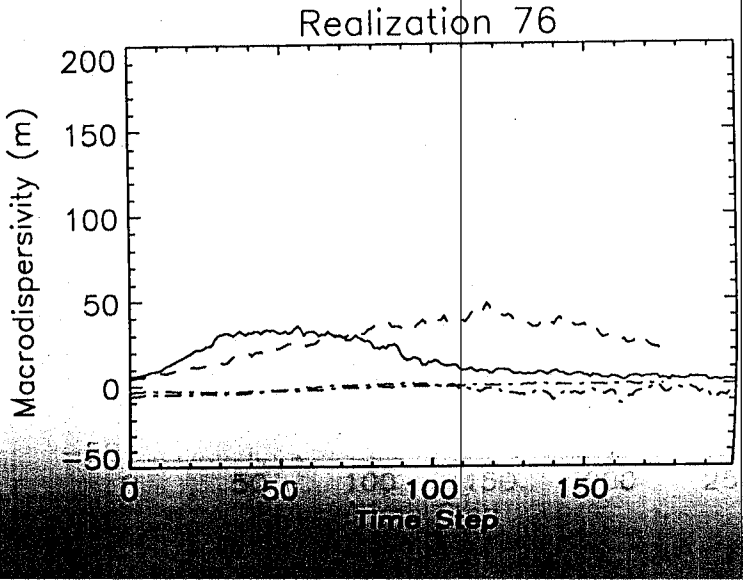


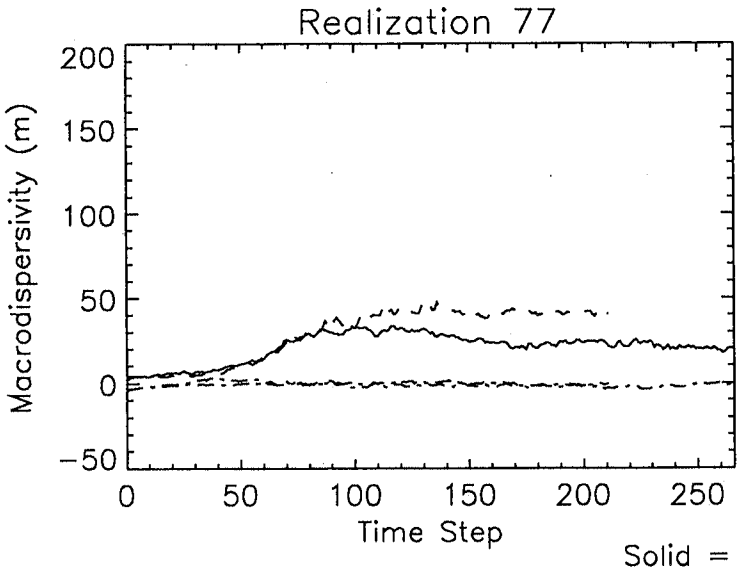


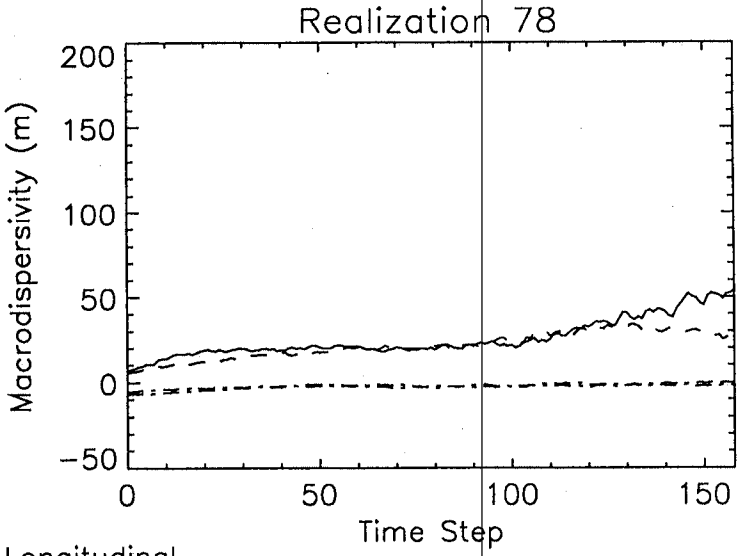


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

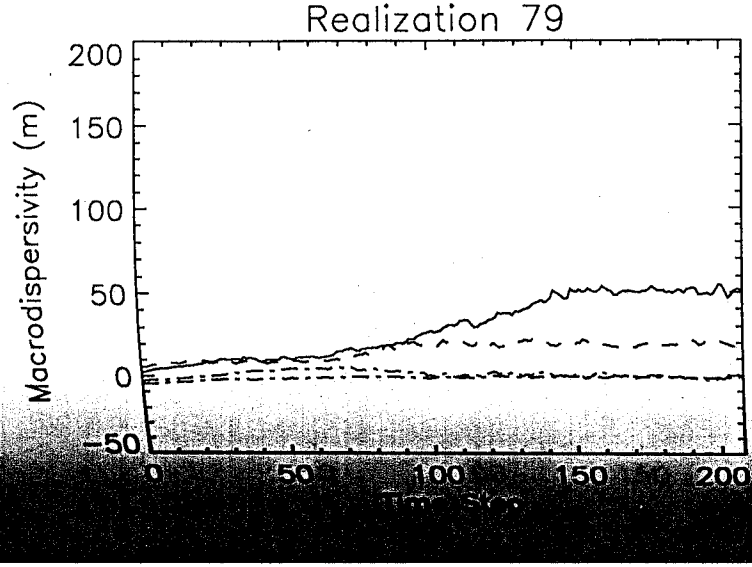


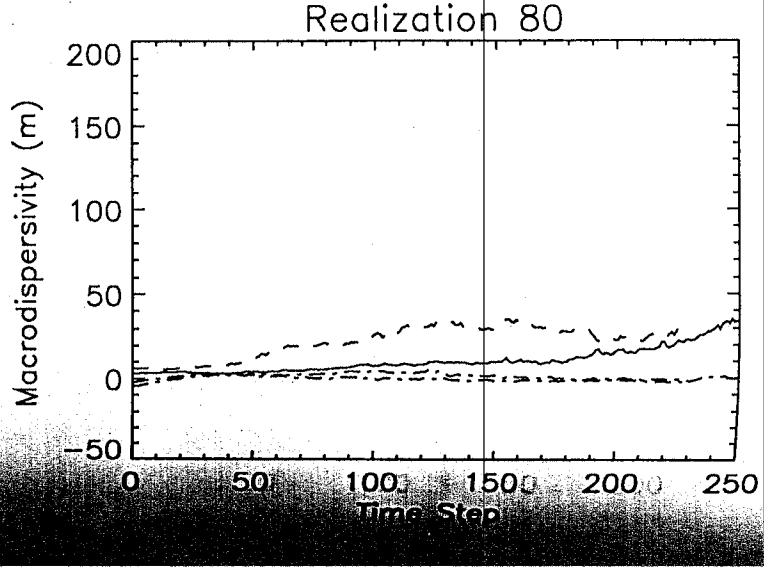


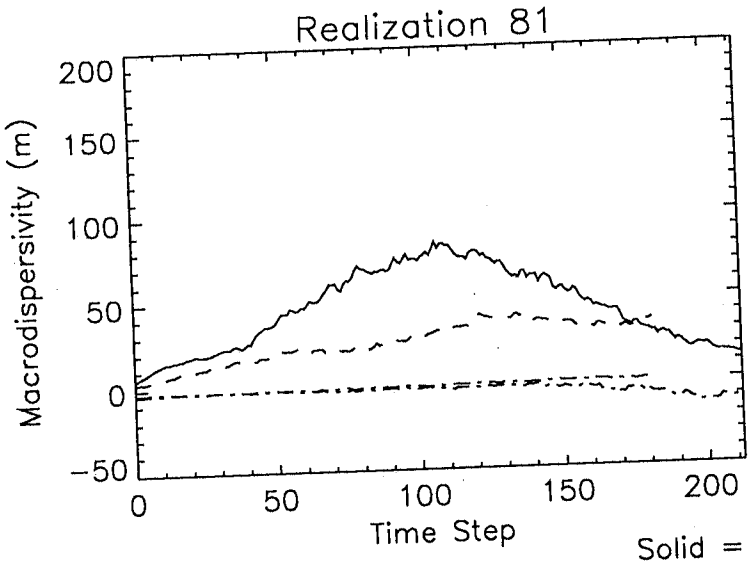


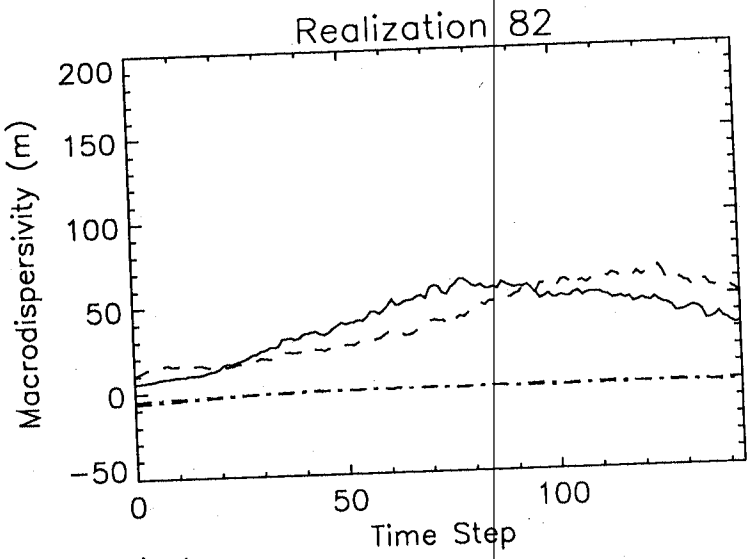


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

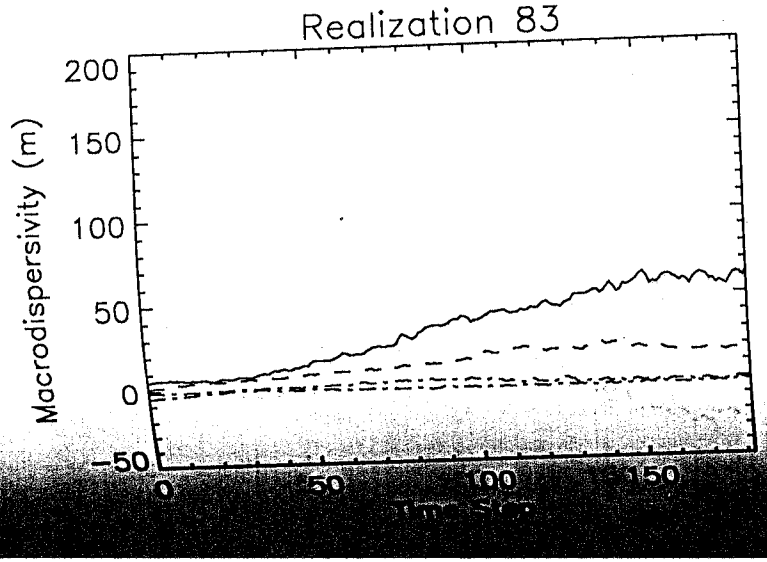


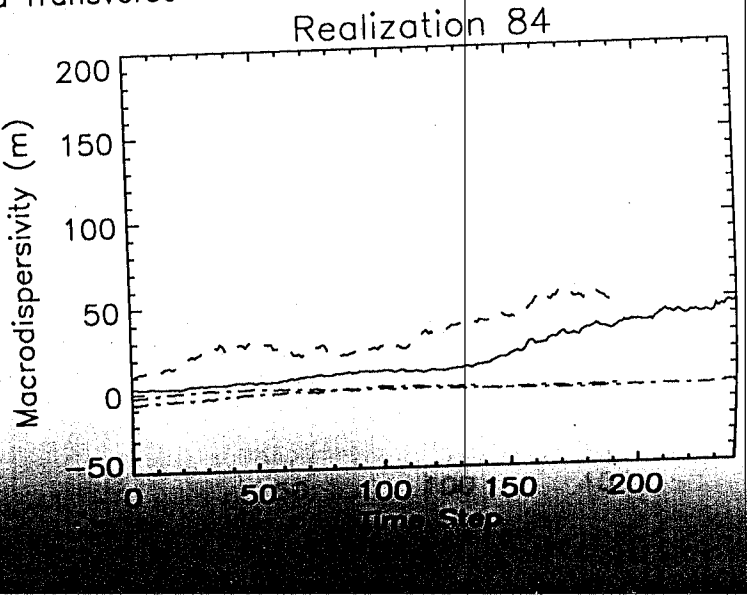


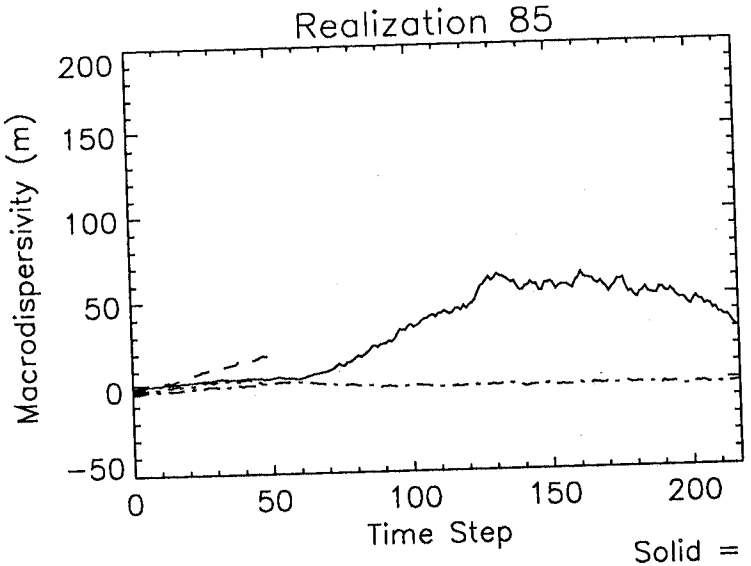


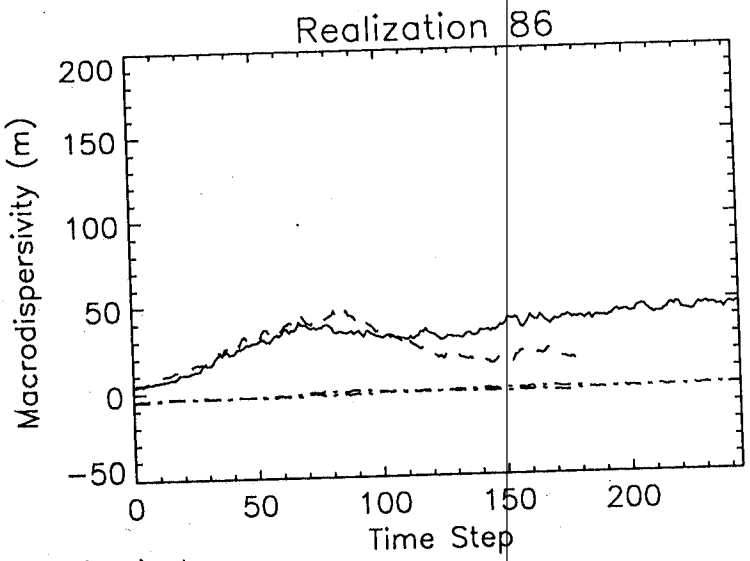


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

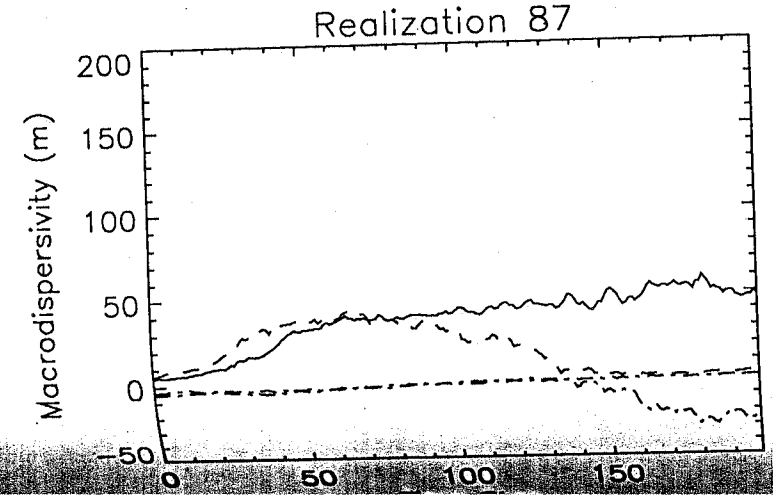


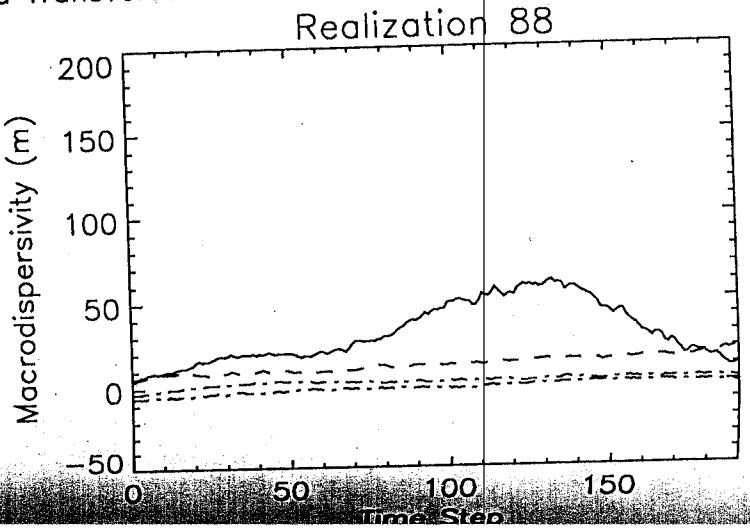


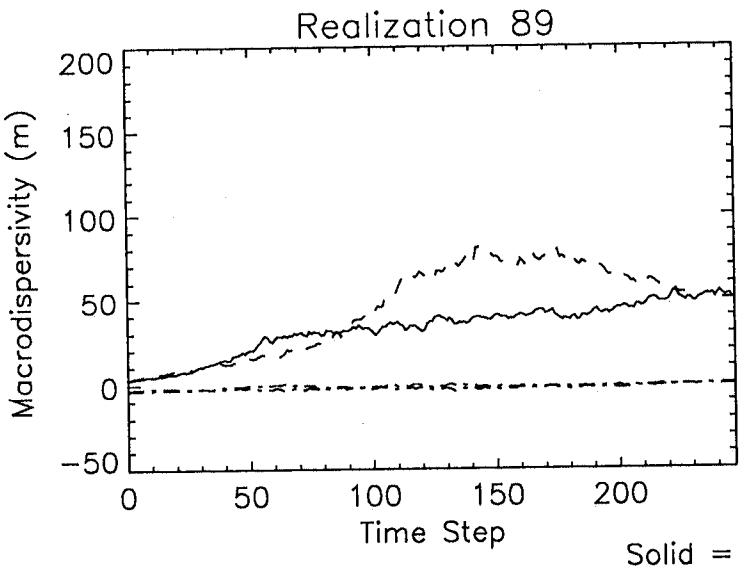


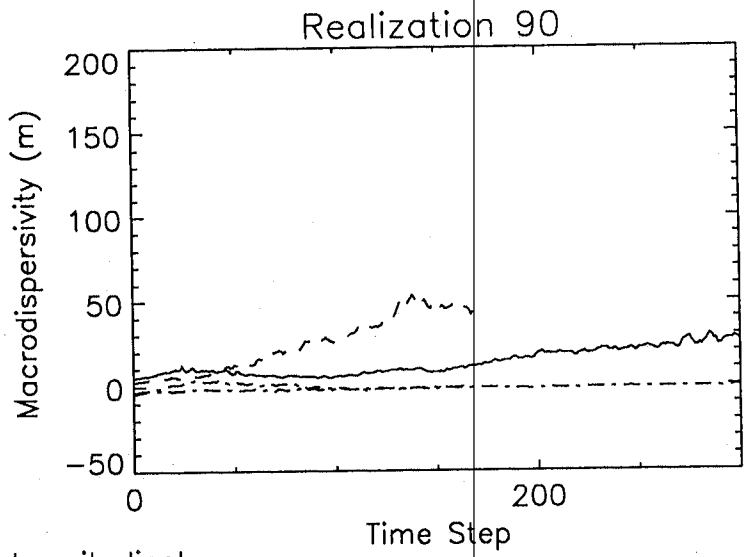


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

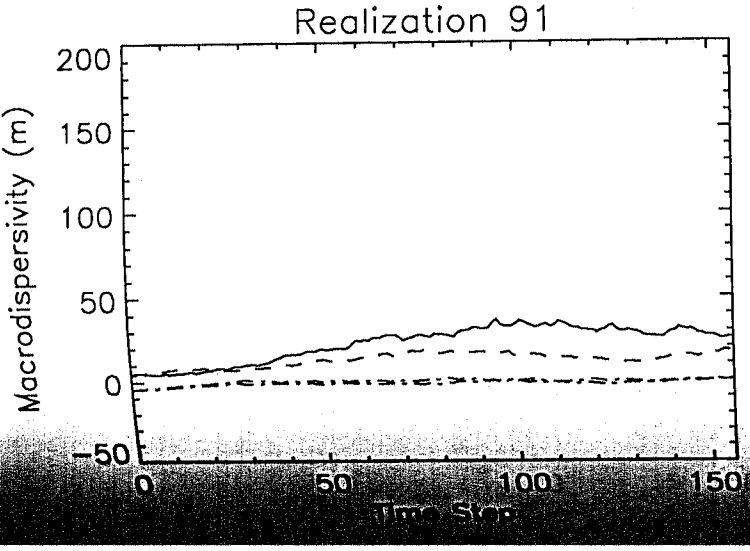


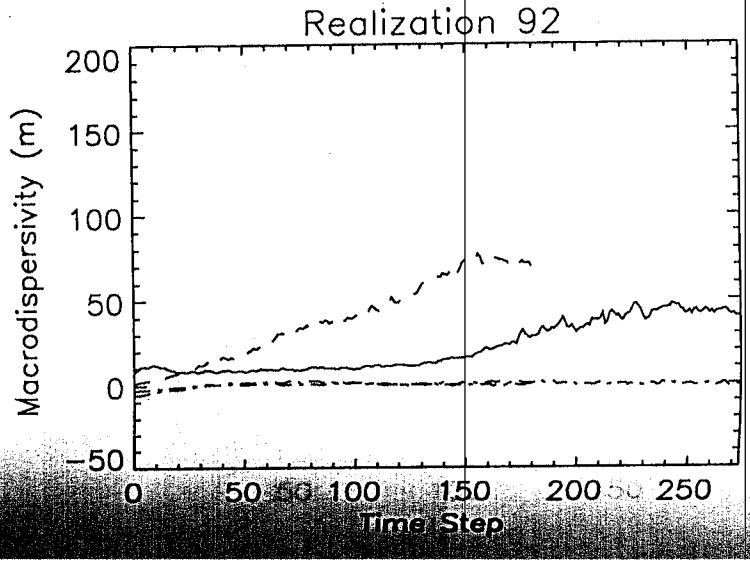


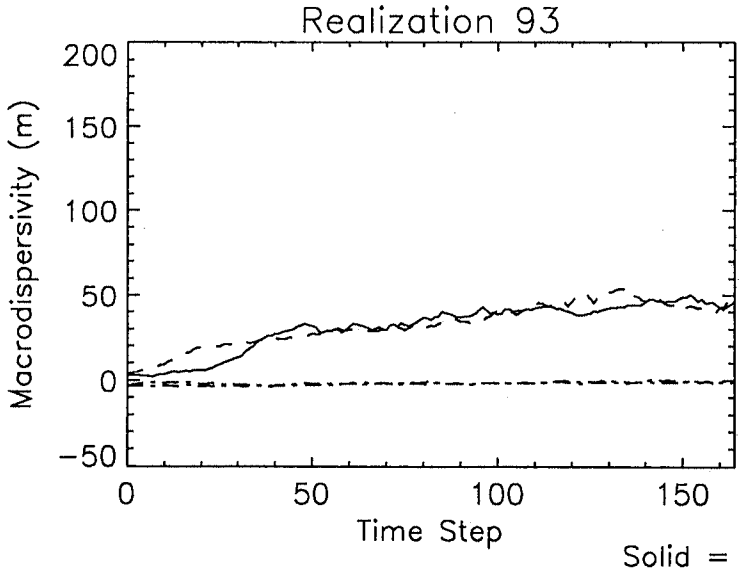


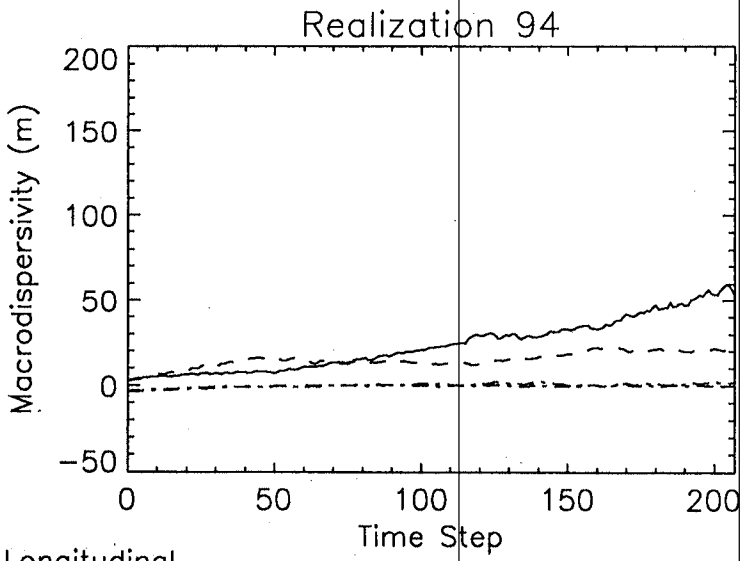


Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

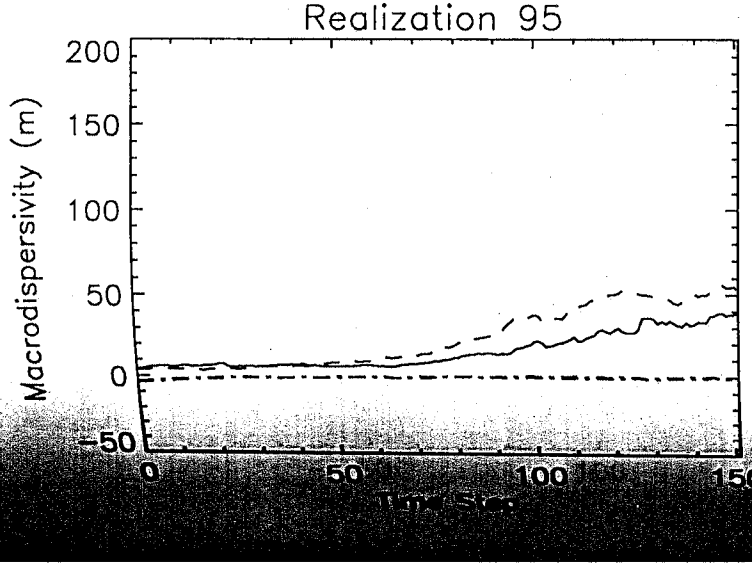


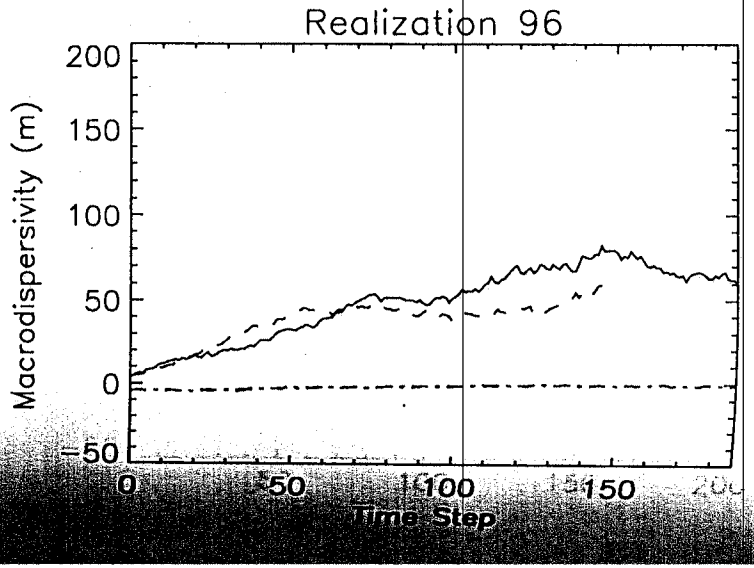






Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse





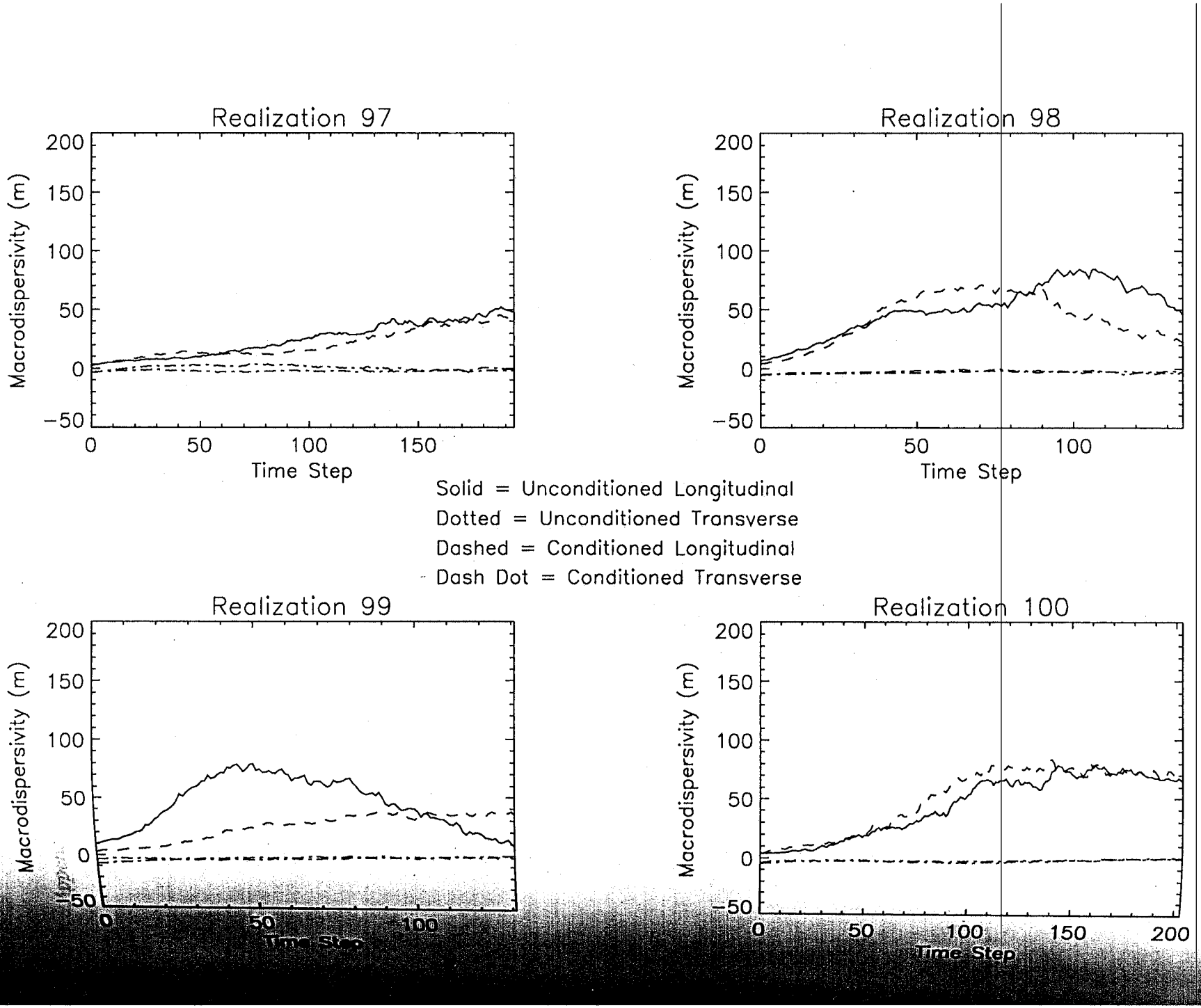
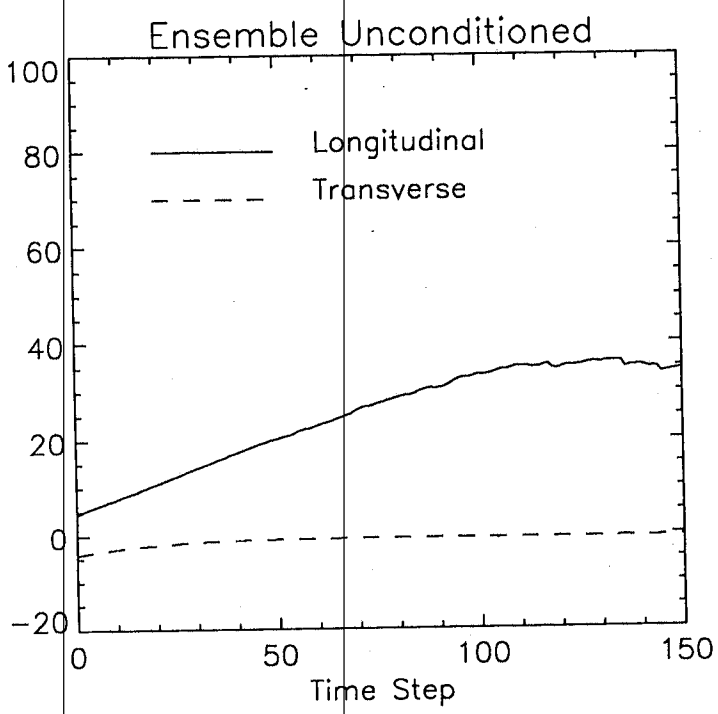
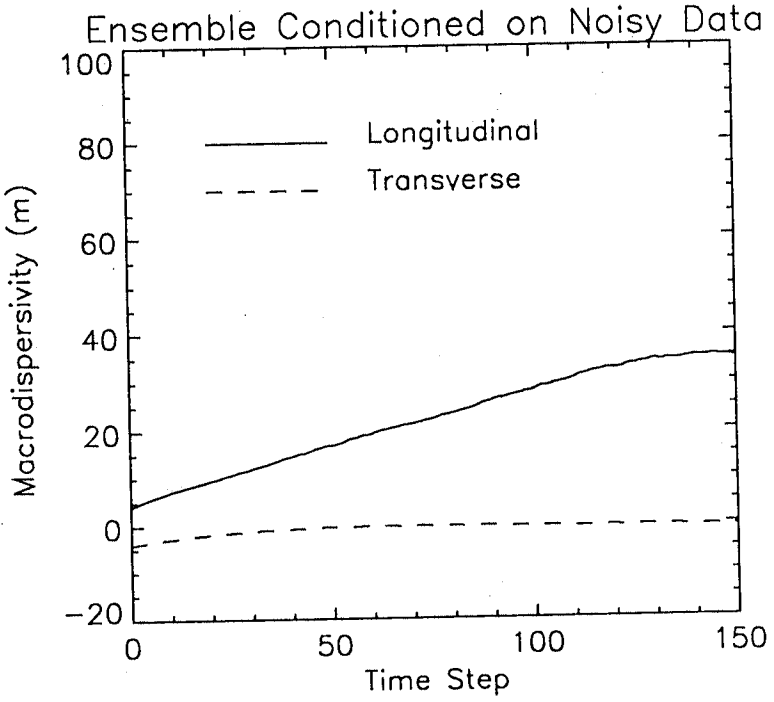

Realization 97
Realization 98
Realization 99
Realization 100
Macrodispersivity (m)
Time Step
200
150
100
50
0
−50
0
50
100
150
Solid = Unconditioned Longitudinal
Dotted = Unconditioned Transverse
Dashed = Conditioned Longitudinal
Dash Dot = Conditioned Transverse

Figure A.4.
Ensemble Macrodispersivity vs. Time for 100 Two-Scale Realizations.


Ensemble Unconditioned
Longitudinal
Transverse
100
80
60
40
20
0
-20
0
50
100
150
Time Step


Ensemble Conditioned on Noisy Data
Longitudinal
Transverse
Macrodispersivity (m)
100
80
60
40
20
0
-20
0
50
100
150
Time Step

Appendix B.
Derivation of Log Likelihood Derivatives

The modified-Newton gradient search requires both first- and second-order derivatives of log likelihood $\mathcal{L}(k)$, given by equation [3.23]. Taking the first-derivative of $\mathcal{L}(k)$ with respect to the parameter α in energy potential function equation [3.17] yields:

$$\frac{\partial \mathcal{L}(k)}{\partial \alpha} = \frac{\partial \sum_{i=1}^{M(k)} \left[U'(i) - \ln \sum_{l=1}^{L} e^{U'(i,l)} \right]}{\partial \alpha}$$

$$= \sum_{i=1}^{M(k)} \left[- U_\alpha'(i) - \frac{1}{\sum_{l=1}^{L} e^{U'(i,l)}} \sum_{l=1}^{L} \frac{\partial e^{U'(i,l)}}{\partial \alpha} \right]$$

$$= \sum_{i=1}^{M(k)} \left[- U_\alpha'(i) + \sum_{l=1}^{L} \frac{U_\alpha'(i,l)\, e^{U'(i,l)}}{\sum_{l=1}^{L} e^{U'(i,l)}} \right] \qquad [B.1]$$

Appendix B: Derivation of Log Likelihood Derivatives

where for the case of pair potential interaction energy, $U'_\alpha(i)$ is equal to the total of current potential differences between site i and all sites in the neighborhood weighted by α, and $U'_\alpha(i, l)$ is the total of potential differences between site i and its α-weighted neighbors for all possible site states at site i.

Taking the derivative of equation [B.1] with respect to α to obtain the second-order derivative,

$$\frac{\partial^2 \mathcal{L}(k)}{\partial \alpha^2} = \sum_{i=1}^{M(k)} \left[\frac{- \sum_{l=1}^{L} e^{U'(i,l)} \sum_{l=1}^{L} U_\alpha'^2(i,l) e^{U'(i,l)} + \sum_{l=1}^{L} U_\alpha'(i,l) e^{U'(i,l)} \sum_{l=1}^{L} U_\alpha'(i,l) e^{U'(i,l)}}{\left[\sum_{l=1}^{L} e^{U'(i,l)} \right]^2} \right]$$

$$= \sum_{i=1}^{M(k)} \left[\frac{\left[\sum_{l=1}^{L} U_\alpha'(i,l) e^{U'(i,l)} \right]^2 - \left[\sum_{l=1}^{L} e^{U'(i,l)} \sum_{l=1}^{L} U_\alpha'^2(i,l)\, e^{U'(i,l)} \right]}{\left[\sum_{l=1}^{L} e^{U'(i,l)} \right]^2} \right] \qquad [B.2]$$

Finally, taking the derivative of equation [B.1] with respect to the remaining first-order MRF parameter β:

Appendix B: Derivation of Log Likelihood Derivatives

$$\frac{\partial^2 \mathcal{L}(k)}{\partial\alpha\partial\beta} = \sum_{i=1}^{M(k)} \left[\frac{- \sum_{l=1}^{L} e^{U'(i,l)} \sum_{l=1}^{L} U_\alpha'(i,l)\, U_\beta'(i,l) e^{U'(i,l)} + \sum_{l=1}^{L} U_\beta'(i,l) e^{U'(i,l)} \sum_{l=1}^{L} U_\beta'(i,l) e^{U'(i,l)}}{\left[\sum_{l=1}^{L} e^{U'(i,l)} \right]^2} \right]$$

$$= \sum_{i=1}^{M(k)} \left[\frac{\left[\sum_{l=1}^{L} U_\beta'(i,l) e^{U'(i,l)} \right]^2 - \left[\sum_{l=1}^{L} e^{U'(i,l)} \sum_{l=1}^{L} U_\alpha'(i,l)\, U_\beta'(i,l) e^{U'(i,l)} \right]}{\left[\sum_{l=1}^{L} e^{U'(i,l)} \right]^2} \right]$$

[B.3]

Appendix C.
Derivation of Adjoint Equation Derivative Terms

The derivatives of $\mathbf{g}(\bar{\sigma},\mathbf{Y})$ with respect to the state variables, $\nabla_{\sigma}\mathbf{g}(\bar{\sigma},\mathbf{Y})$, are required for solution of the adjoint-state equation [4.23] for $(\bar{\psi}_C^*)^T$. In addition, the derivative $\nabla_{\mathbf{Y}}\mathbf{g}(\bar{\sigma}, \mathbf{Y})$ is required to solve equation [4.24] for the monitoring well concentration sensitivity $\partial C/\partial \mathbf{Y}$. For purposes of minimizing matrix size, only two transport time steps were considered — 'initial' concentration $\bar{c}_o$ associated with the time immediately prior to the time step where concentration sensitivities are of interest, and final or current concentration $\bar{c}$ at the time of interest. For transient design, all time steps included in the design horizon should be incorporated into the problem.

Derivatives of $\mathbf{g}(\bar{\sigma}, \mathbf{Y})$ with respect to $\bar{\sigma}$ and $\mathbf{Y}$ were obtained analytically from the matrix equation:

$$\mathbf{g}(\bar{\sigma}, \mathbf{Y}) = \begin{bmatrix} g_{flow}(\bar{h}, \mathbf{Y}) \\ g_{darcy}(\bar{h}, \mathbf{Y}) \\ g_{tran0}(\bar{v}, \bar{c}_0, \mathbf{Y}) \\ g_{tran}(\bar{v}, \bar{c}_0, \bar{c}, \mathbf{Y}) \end{bmatrix} = \begin{bmatrix} \mathbf{A}_{flow}(\mathbf{Y})\,\bar{h} - \bar{b}_{flow}(\mathbf{Y}) \\ \mathbf{I}\bar{v} - \mathbf{A}_{darcy}(\mathbf{Y})\,\bar{h} \\ \mathbf{A}_{trans}(\bar{v})\,\bar{c}_0 + \mathbf{A}_{trant}\,\bar{c}_{-1} - \bar{b}_{to}(\mathbf{Y}) \\ \mathbf{A}_{trans}(\bar{v})\,\bar{c} + \mathbf{A}_{trant}\,\bar{c}_0 - \bar{b}_t(\mathbf{Y}) \end{bmatrix} = \begin{bmatrix} 0 \\ 0 \\ 0 \\ 0 \end{bmatrix} \quad [C.1]$$

Appendix C: Derivation of Adjoint Equation Derivative Terms

where $\bar{c}_o$ represents the vector of nodal concentrations at the time step immediately prior to the time step where monitoring well concentration sensitivity is of interest, and not necessarily the initial concentration. These concentrations, along with final concentrations $\bar{c}$, are determined by solving the primary transport problem using a particle tracker. Values of head $\bar{h}$ and velocity $\bar{v}$ are likelwise determined from solution of the corresponding forward steady-state flow equation and Darcy's law.

C.1 Derivation of the $\nabla_\sigma g(\bar{\sigma}, Y)$ Matrix

From equation [C.1], differentiation of $\mathbf{g_{flow}}(\bar{\sigma}, \mathbf{Y})$ with respect to head $\bar{h}$ yields:

$$\nabla_h \mathbf{g_{flow}}(\bar{\sigma}, \mathbf{Y}) = \mathbf{A_{flow}}(\mathbf{Y}) \qquad [C.2]$$

where $\mathbf{A_{flow}}(\mathbf{Y})$ is evaluated at the current estimate of $\mathbf{Y}$. From inspection of equation [C.1], it is evident that derivatives of $\mathbf{g_{flow}}(\bar{\sigma}, \mathbf{Y})$ with respect to other state variables velocity $\bar{v}$, prior concentration $\bar{c}_o$, and current concentration $\bar{c}$ are identically equal to zero vectors.

Differentiation of $\mathbf{g_{darcy}}(\bar{\sigma}, \mathbf{Y})$ in equation [C.1] with respect to velocity $\bar{v}$ yields the identity matrix $\mathbf{I}$, with the head derivative of $\mathbf{g_{darcy}}(\bar{\sigma}, \mathbf{Y})$ equal to:

Appendix C: Derivation of Adjoint Equation Derivative Terms

$$\nabla_h \mathbf{g_{darcy}}(\bar{h}, \mathbf{Y}) = -\mathbf{A_{darcy}}(\mathbf{Y}) \qquad [C.3]$$

where, again, the matrix is evaluated at the current estimate of **Y**.

FInally the derivative of $\mathbf{g_{tran0}}(\bar{\sigma}, \mathbf{Y})$ and $\mathbf{g_{tran}}(\bar{\sigma}, \mathbf{Y})$ with respect to velocity is equal to:

$$\nabla_v \mathbf{g_{tran0}}(\bar{v}, \bar{c}_0, \mathbf{Y}) = \frac{\mathbf{dA_{trans}}(\bar{v})}{\mathbf{d}\bar{v}} \bar{c}_o \qquad [C.4]$$

$$\nabla_v \mathbf{g_{tran}}(\bar{v}, \bar{c}_0, \bar{c}, \mathbf{Y}) = \frac{\mathbf{dA_{trans}}(\bar{v})}{\mathbf{d}\bar{v}} \bar{c} \qquad [C.5]$$

with the matrix derivatives evaluated at the current estimate of nodal velocity $\bar{v}$ predicted via the darcy equation and nodal concentrations $\bar{c}_0$ and $\bar{c}$ obtained from solution of the primary transport problem.

The derivative of $\mathbf{g_{tran0}}(\bar{\sigma}, \mathbf{Y})$ and $\mathbf{g_{tran}}(\bar{\sigma}, \mathbf{Y})$ with respect to prior and current concentrations is:

$$\nabla_{c_0} \mathbf{g_{tran0}}(\bar{v}, \bar{c}_0, \mathbf{Y}) = \mathbf{A_{trans}}(\bar{v}) \qquad [C.6]$$

$$\nabla_{c_0} \mathbf{g_{tran}}(\bar{v}, \bar{c}_0, \bar{c}, \mathbf{Y}) = \mathbf{A_{trant}} \qquad [C.7]$$

$$\nabla_c \mathbf{g_{tran}}(\bar{v}, \bar{c}_0, \bar{c}, \mathbf{Y}) = \mathbf{A_{trans}}(\bar{v}) \qquad [C.8]$$

where $\mathbf{A_{trans}}(\bar{v})$ depends on the current estimate of velocity $\bar{v}$, and is the same in both equation [C.6] and [C.8] for steady-state flow.

Appendix C: Derivation of Adjoint Equation Derivative Terms

Equations [C.2] through [C.8] can be substituted into the adjoint-state equation [4.23]:

$$\begin{bmatrix} \nabla_h g_{flow} & \nabla_h g_{darcy} & \nabla_h g_{tran0} & \nabla_h g_{tran} \\ \nabla_v g_{flow} & \nabla_v g_{darcy} & \nabla_v g_{tran0} & \nabla_v g_{tran} \\ \nabla_{c0} g_{flow} & \nabla_{c0} g_{darcy} & \nabla_{c0} g_{tran0} & \nabla_{c0} g_{tran} \\ \nabla_c g_{flow} & \nabla_c g_{darcy} & \nabla_c g_{tran0} & \nabla_c g_{tran} \end{bmatrix} \begin{bmatrix} \nabla g_{flow} C \\ \nabla g_{darcy} C \\ \nabla g_{tran0} C \\ \nabla g_{darcy} C \end{bmatrix} = -\mathbf{i} \qquad [C.9]$$

to yield:

$$\begin{bmatrix} A_{flow}(Y)_{(MXM)} & -A_{darcy}(Y)_{(2MXM)} & 0_{(MXM)} & 0_{(MXM)} \\ 0_{(MX2M)} & I_{(2MX2M)} & \frac{dA_{trans}(\bar{v})}{dv}\bar{c}_o{}_{(MX2M)} & \frac{dA_{trans}(\bar{v})}{dv}\bar{c}_o{}_{(MX2M)} \\ 0_{(MXM)} & 0_{(2MXM)} & A_{trans}(\bar{v})_{(MXM)} & A_{trant}{}_{(MXM)} \\ 0_{(MXM)} & 0_{(2MXM)} & 0_{(MXM)} & A_{trans}(\bar{v})_{(MXM)} \end{bmatrix} (\psi_C{}^{*})^T = -\mathbf{i} \qquad [C.10]$$

Note that $A_{flow}(Y)$, $A_{trans}(\bar{v})$, and A_{trant} are of dimension M-by-M, where M is equal to the number of finite-difference blocks. Since velocities are required in both the x- and y-directions, $A_{darcy}(Y)$ is of dimension 2M-by-M, **I** is of dimension 2M-by-2M, and $\frac{dA_{trans}(\bar{v})}{d\bar{v}}\bar{c}_o$ and $\frac{dA_{trans}(\bar{v})}{d\bar{v}}\bar{c}$ are of dimension M-by-2M.

Appendix C: Derivation of Adjoint Equation Derivative Terms

Dimensions of the zero matrices vary, as shown in equation [C.10]. For a problem characterized by two transport time steps, the total dimension of the adjoint-state solution matrix in equation [C.10] is 5M-by-5M. Thus, even a small problem of physical dimension 5-by-5 requires a 5(25)-by-5(25) or a 125-by-125 solution matrix. Fortunately, the sparse nature of the matrix makes it possible to handle problems of very large physical dimension. The subroutine SPRSIN described in Press *et al.* (1992) was used in conjunction with the GMRES solver to solve equation [C.10] for $(\overline{\psi}_C^*)^T$ with minimal CPU and disk space requirements. As mentioned in Chapter 4, the adjoint equation can be decomposed into three seperate uncoupled problems when solved from the right. The first subproblem involves symmetric, banded, sparse matrix solution for the head adjoint state, followed by substitution of known head adjoint states and symmetric, banded, sparse matrix solution for velocity adjoints. Finally, the concentration adjoint states are determined in backward order by substituting velocity adjoint states, but in this case the transport adjoint matrix equation lacks symmetry and requires use of the GMRES solver. When the adjoint problem is solved from the right, however, such decoupling is not possible.

C.2 Derivation of the $\nabla_Y g(\overline{\sigma}, Y)$ Matrix

Using equation [C.1], the derivative of $\mathbf{g}(\overline{\sigma}, \mathbf{Y})$ with respect to $\mathbf{Y}$ can easily be derived and used, along with $\overline{\psi}_C^*$ from solution of equation [4.23], to solve equation [4.24] for dC/dY. For the case of specified flux boundary conditions used in this study,

Appendix C: Derivation of Adjoint Equation Derivative Terms

$\bar{b}_{flow}$ was not a function of **Y**. Thus, the derivative of $\mathbf{g_{flow}}(\bar{h}, \mathbf{Y})$ with respect to **Y** is:

$$\nabla_{\mathbf{Y}}\mathbf{g_{flow}}(\bar{h}, \mathbf{Y}) = \frac{d\mathbf{A_{flow}}(\bar{h}, \mathbf{Y})}{d\mathbf{Y}}\,\bar{h} \qquad [C.11]$$

where $\bar{h}$ is the steady-state head from solution of the forward flow equation.

The derivative of $\mathbf{g_{darcy}}(\bar{h}, \mathbf{Y})$ with respect to **Y** is:

$$\nabla_{\mathbf{Y}}\mathbf{g_{darcy}}(\bar{h}, \mathbf{Y}) = \frac{-\,d\mathbf{A_{darcy}}(\mathbf{Y})}{d\mathbf{Y}}\,\bar{h} \qquad [C.12]$$

Finally, for the case in which the plume is located sufficiently far from the boundaries to be unaffected by them, solute mass flux through the boundary will only be indirectly affected by **Y** through the darcy equation $\mathbf{g_{darcy}}(\bar{h}, \mathbf{Y})$. Thus, both $\bar{b}_{to}(\mathbf{Y})$ and $\bar{b}_{t}(\mathbf{Y})$ are independent of **Y**. Moreover, $\mathbf{A_{trans}}(\bar{v})$ is not directly dependent on **Y**, but rather depends on changes in **Y** through the velocity described by the darcy equation $\mathbf{g_{darcy}}(\bar{h}, \mathbf{Y})$. This, both $\nabla_{\mathbf{Y}}\,\mathbf{g_{tran0}}\,(\bar{v}, \bar{c}_0, \mathbf{Y})$ and $\nabla_{\mathbf{Y}}\,\mathbf{g_{tran}}\,(\bar{v}, \bar{c}_0, \bar{c}, \mathbf{Y})$ are equal to zero vectors.

Using equation [C.11] and [C.12], equation [4.24] reduces to:

Since both $\mathbf{A_{flow}}(\mathbf{Y})$ and $\mathbf{A_{darcy}}(\mathbf{Y})$ are expressed in terms of conductivity **K** in the primary flow and darcy simulation functions, equation [C.13] was restated as follows using the relation $d\mathbf{Y} = d\ln \mathbf{K} = d\mathbf{K}/\mathbf{K}$:

Appendix C: Derivation of Adjoint Equation Derivative Terms

$$\frac{dC}{dY} = \bar{\psi}_C^* \begin{bmatrix} \nabla_Y \, g_{flow}(\bar{h}, Y) \\ \nabla_Y \, g_{darcy}(\bar{h}, Y) \\ \nabla_Y \, g_{tran0}(\bar{v}, \bar{c}_0, Y) \\ \nabla_Y \, g_{tran}(\bar{v}, \bar{c}_0, \bar{c}, Y) \end{bmatrix} = \bar{\psi}_C^* \begin{bmatrix} \frac{dA_{flow}(Y)}{dY} \bar{h} \\ -\frac{dA_{darcy}(Y)}{dY} \bar{h} \\ 0 \\ 0 \end{bmatrix} \qquad [C.13]$$

$$\frac{dC}{dY} = \bar{\psi}_C^* \begin{bmatrix} \frac{dA_{flow}(K)}{dK/K} \bar{h} \\ -\frac{dA_{darcy}(K)}{dK/K} \bar{h} \\ 0 \\ 0 \end{bmatrix} = \bar{\psi}_C^* \begin{bmatrix} K\frac{dA_{flow}(K)}{dK} \bar{h} \\ -K\frac{dA_{darcy}(K)}{dK} \bar{h} \\ 0 \\ 0 \end{bmatrix} \qquad [C.14]$$

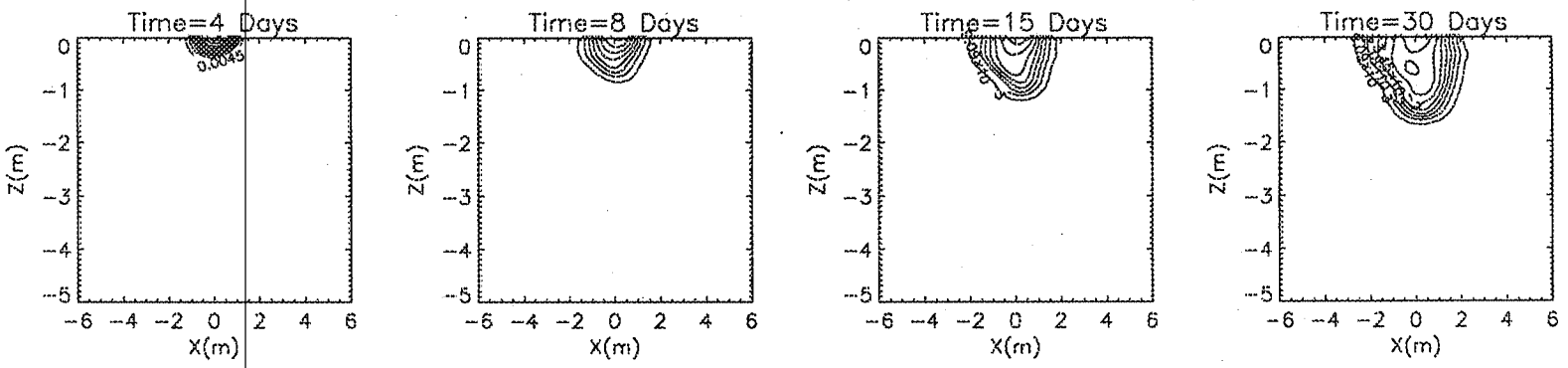
C.1 References

Press, W.H., Teukolsky, S.A. Vetterling, W.T., and Flannery, B.P., *Numerical Recipes: The Art of Scientific Computing*, Cambridge University Press, New York, 1992.

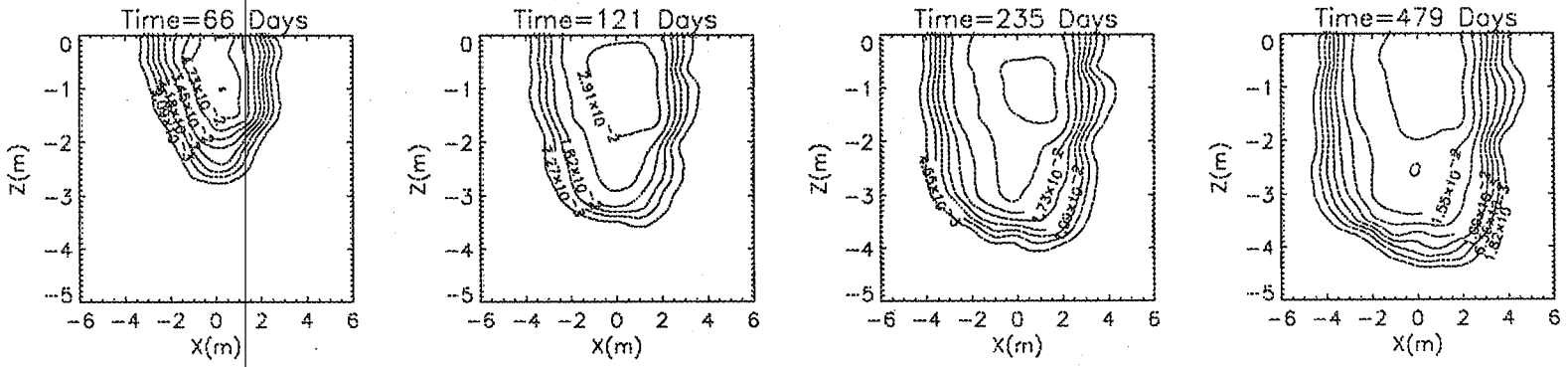
Appendix D.
Simulated Water Content Change and Bromide Concentration Realizations

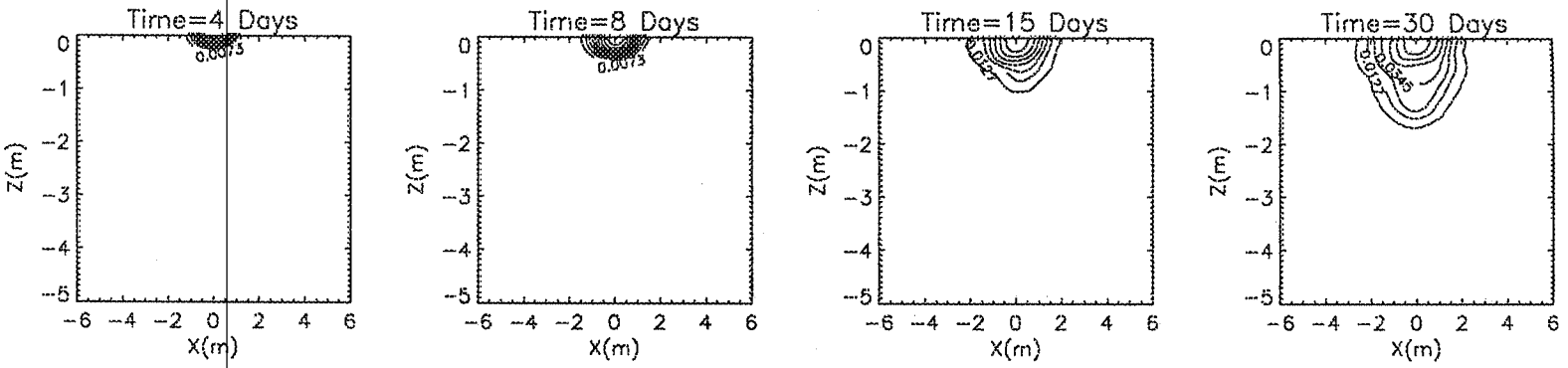
Appendix D: Simulated Water Content Change and Bromide Concentration Realizations

Figure D.1. Simulated Water Content Change Realizations For Single-Scale Permeability Model.

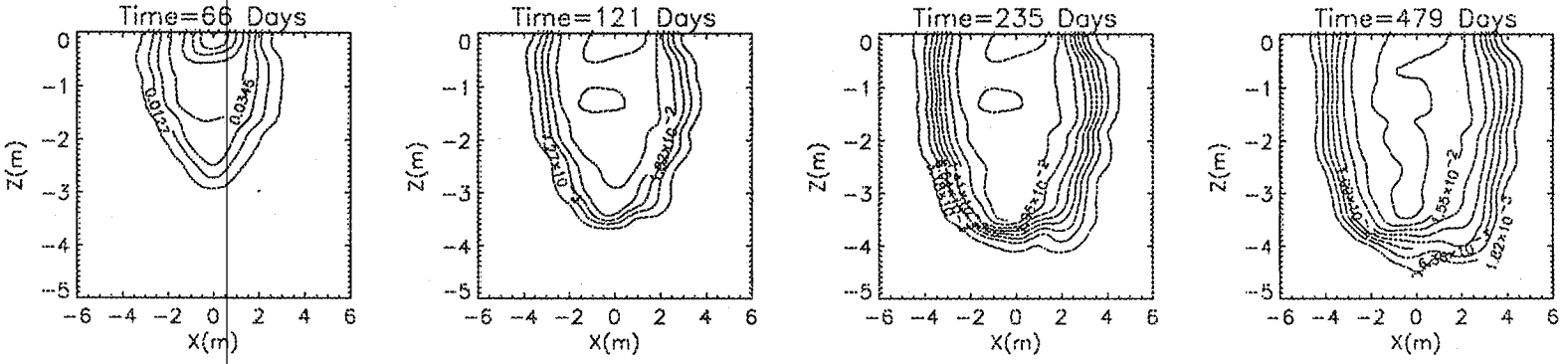


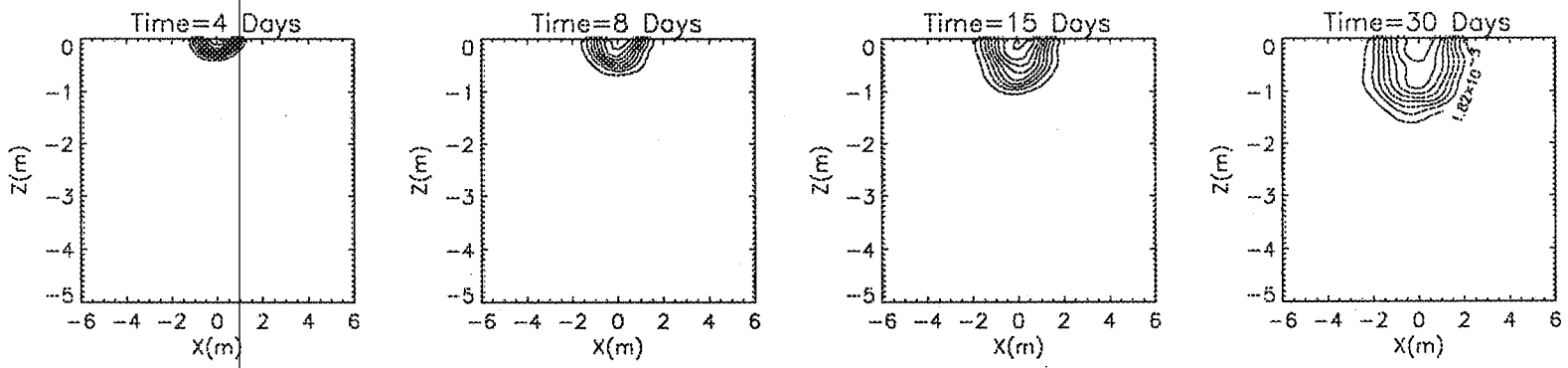
Realization 1



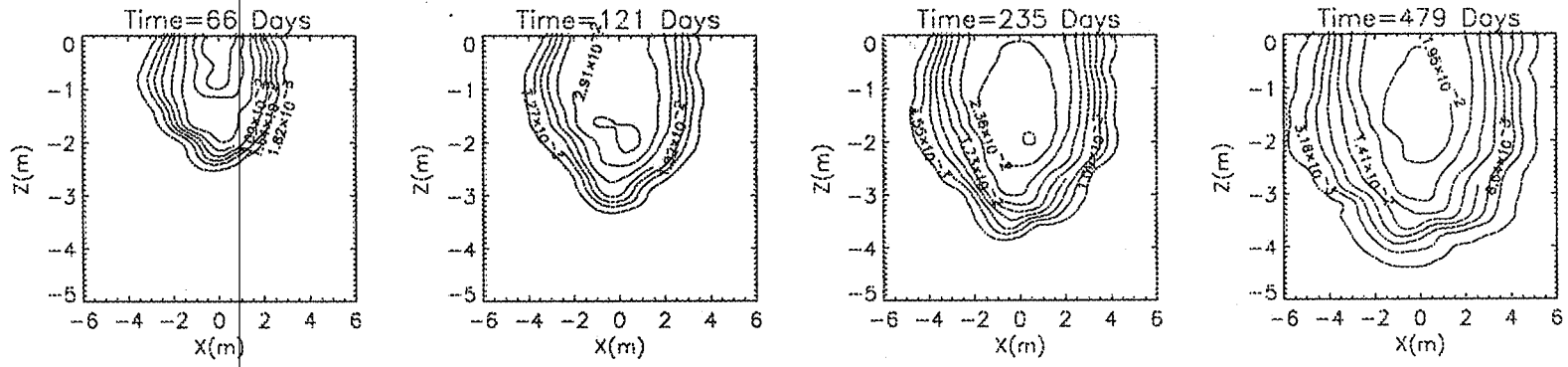


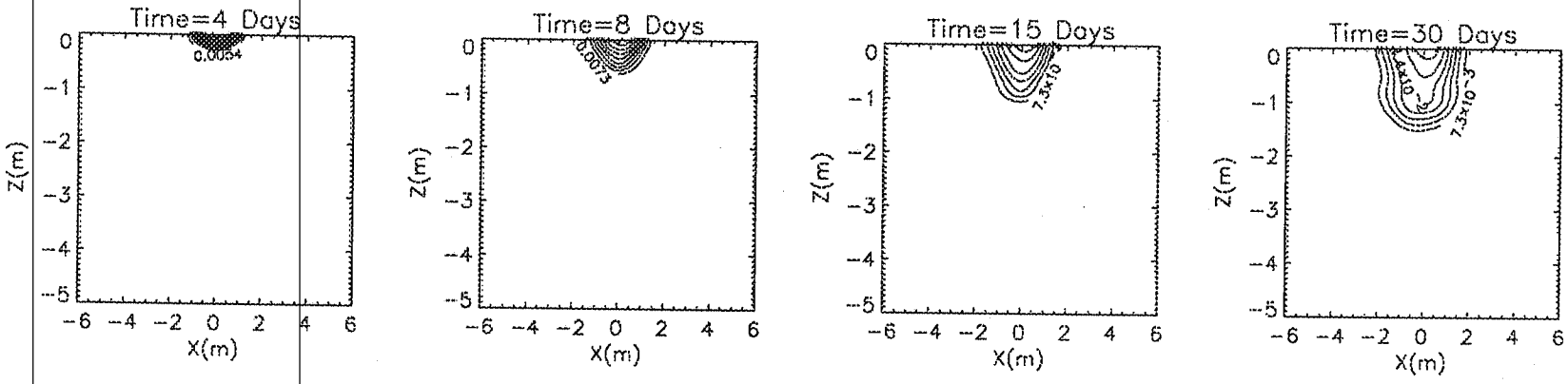
Realization 2



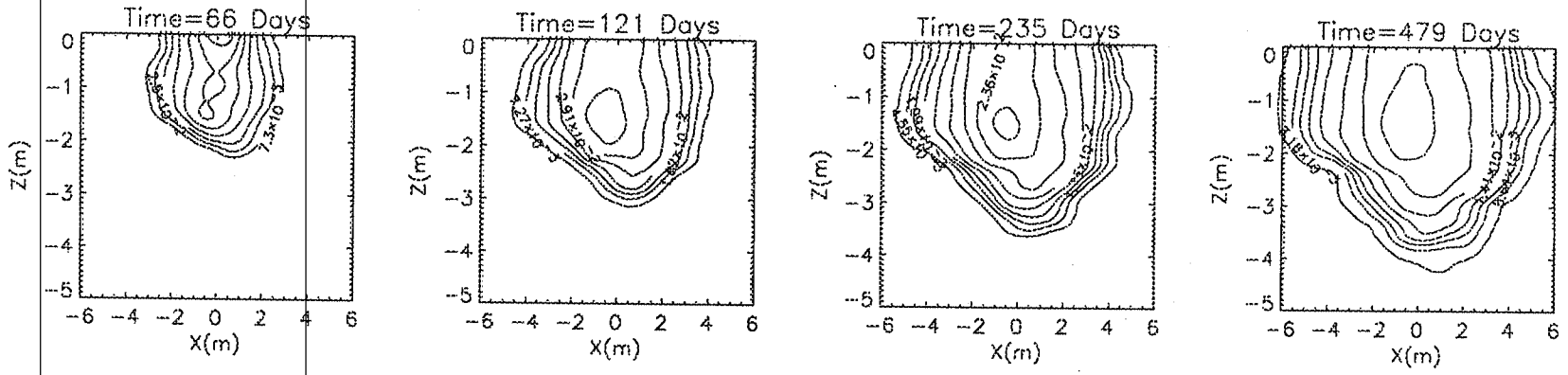


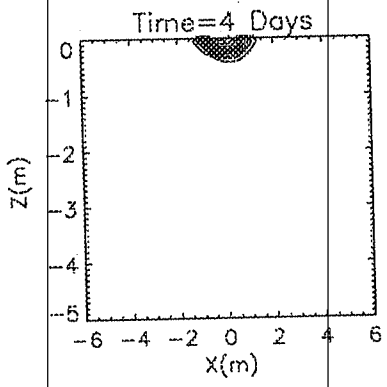
Realization 3

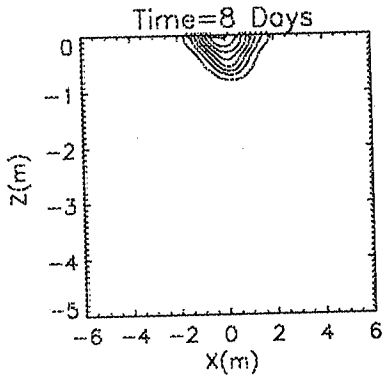


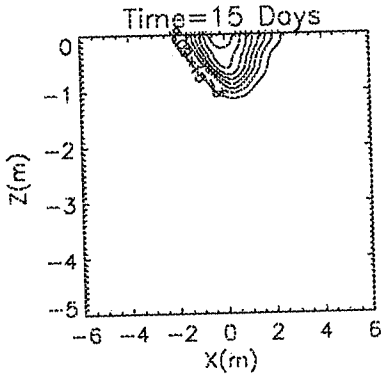


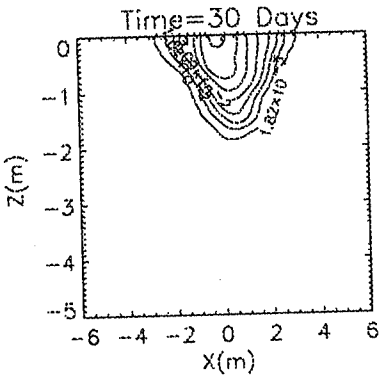
Realization 4



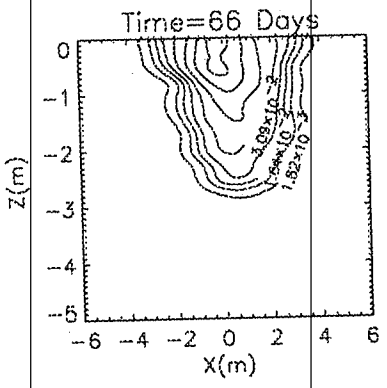


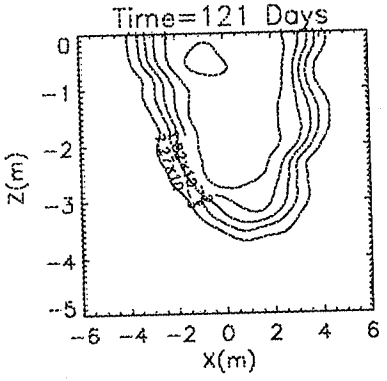


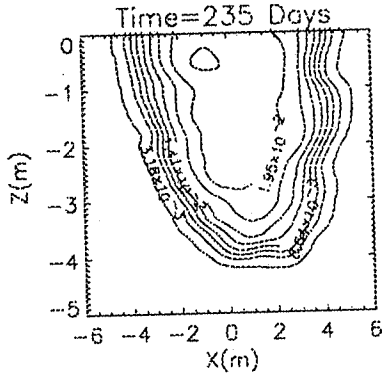


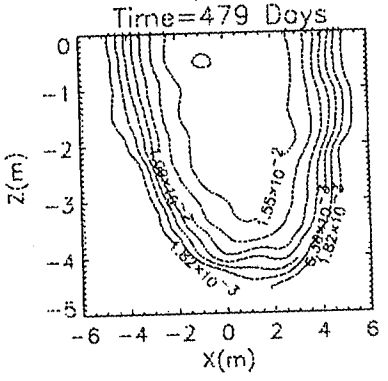


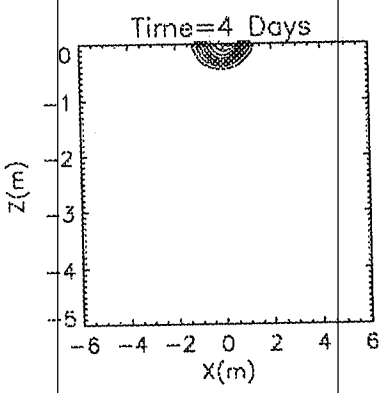
Realization 5

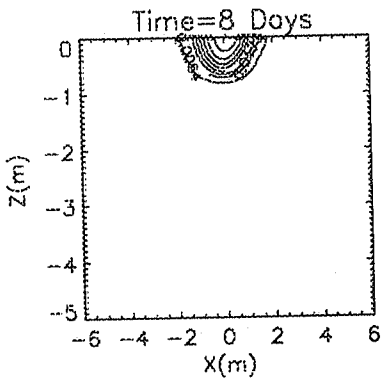


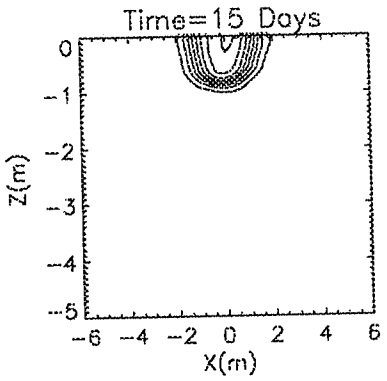


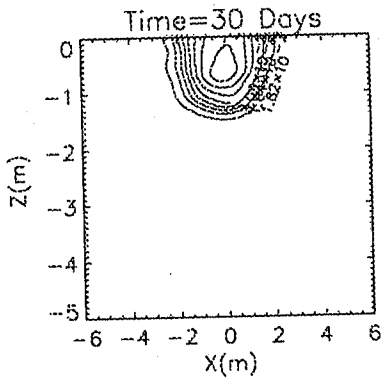




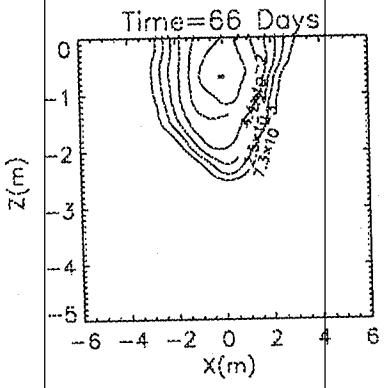


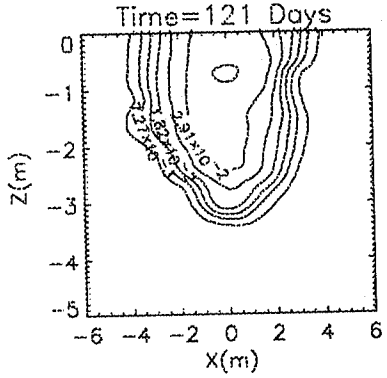


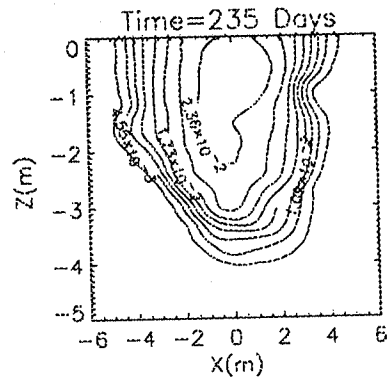


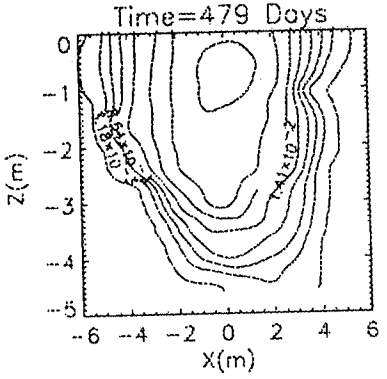


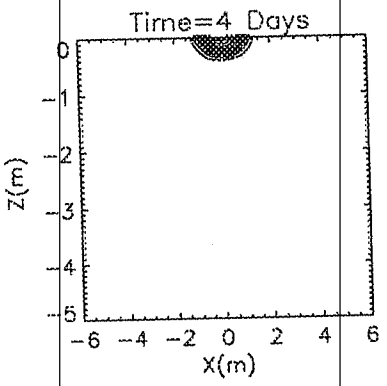
Realization 6

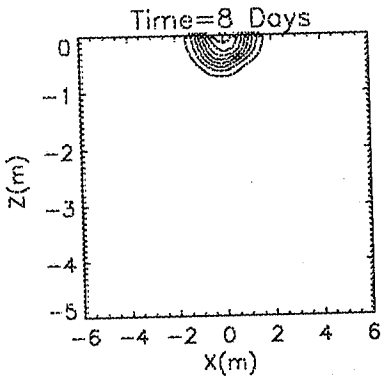


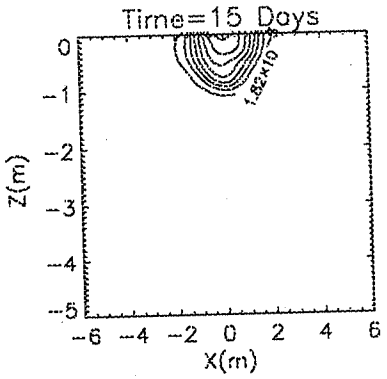


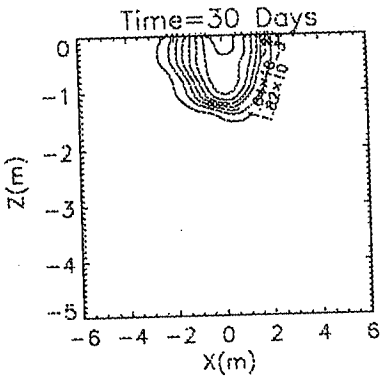




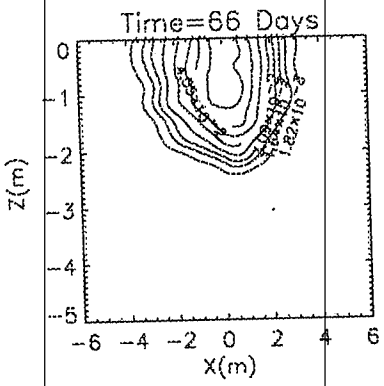


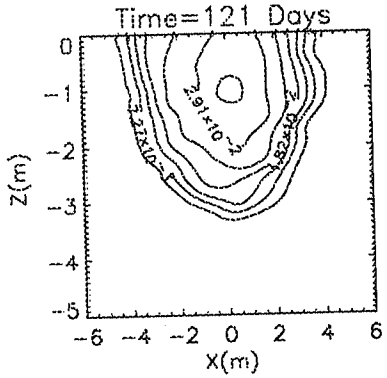


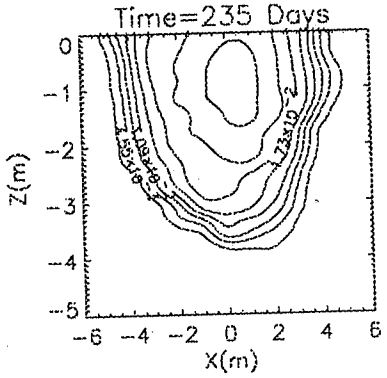


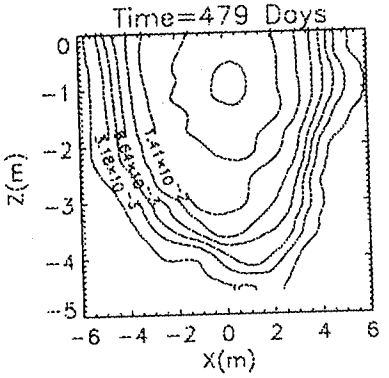


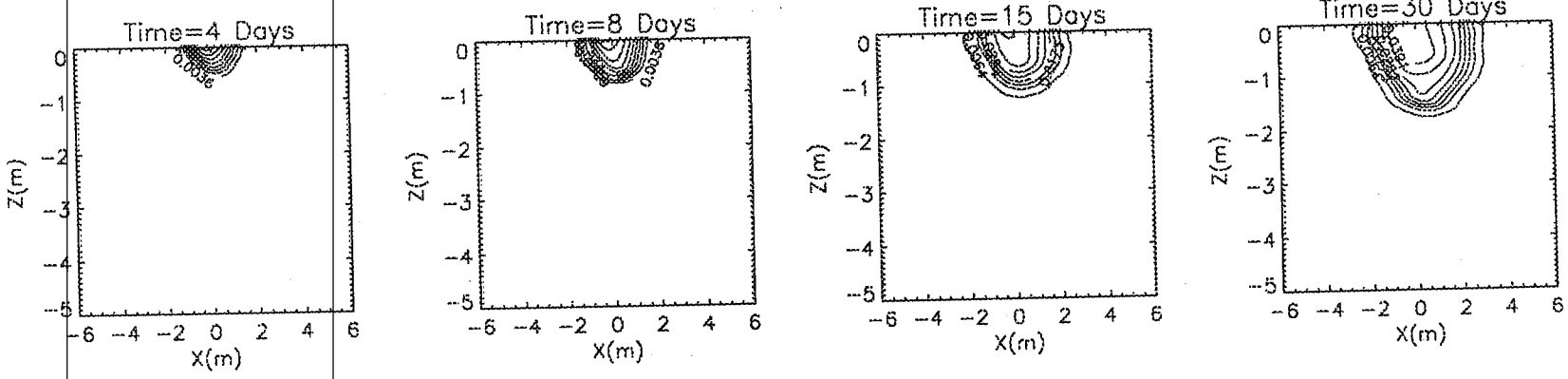
Realization 7



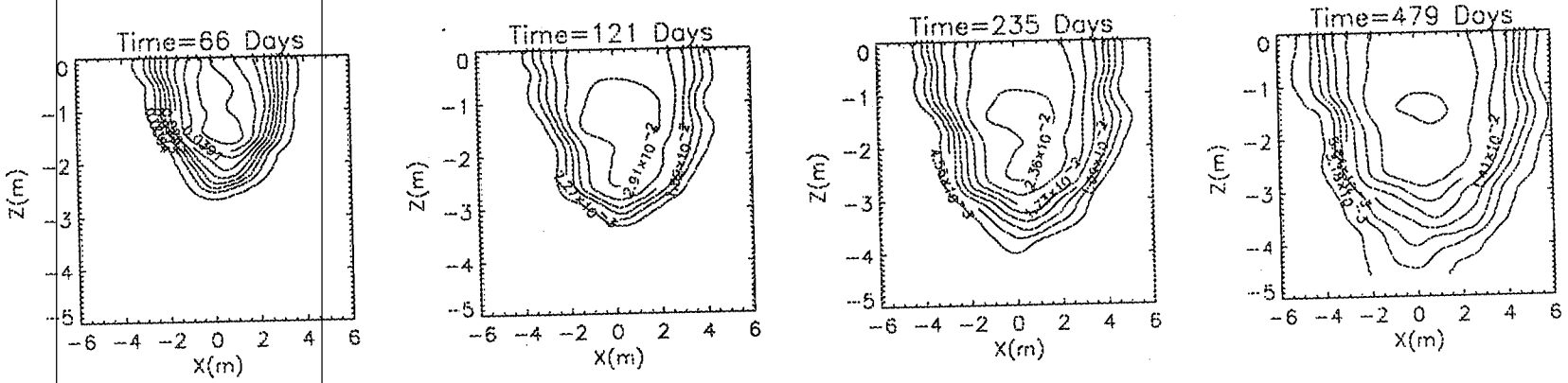


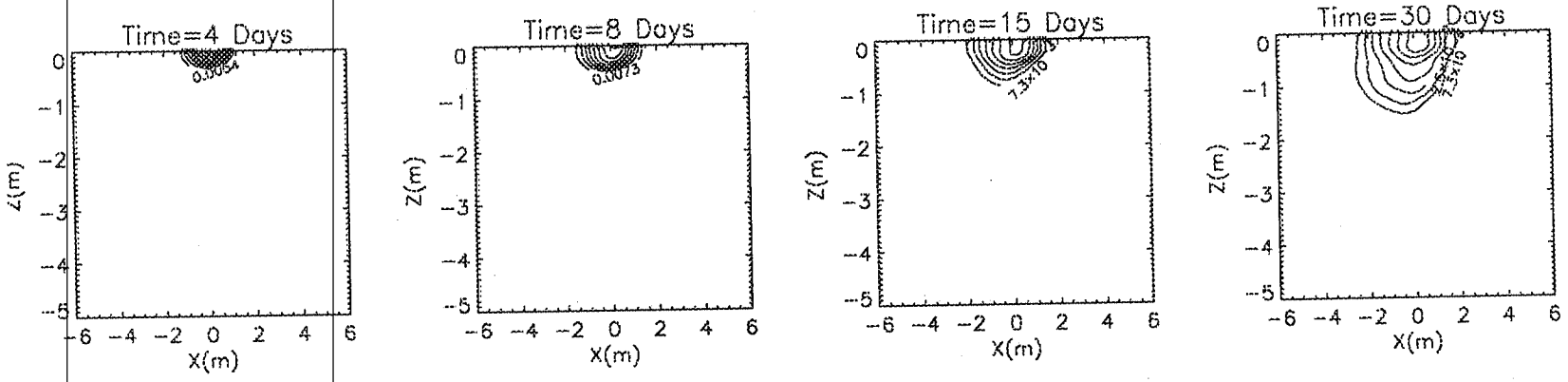




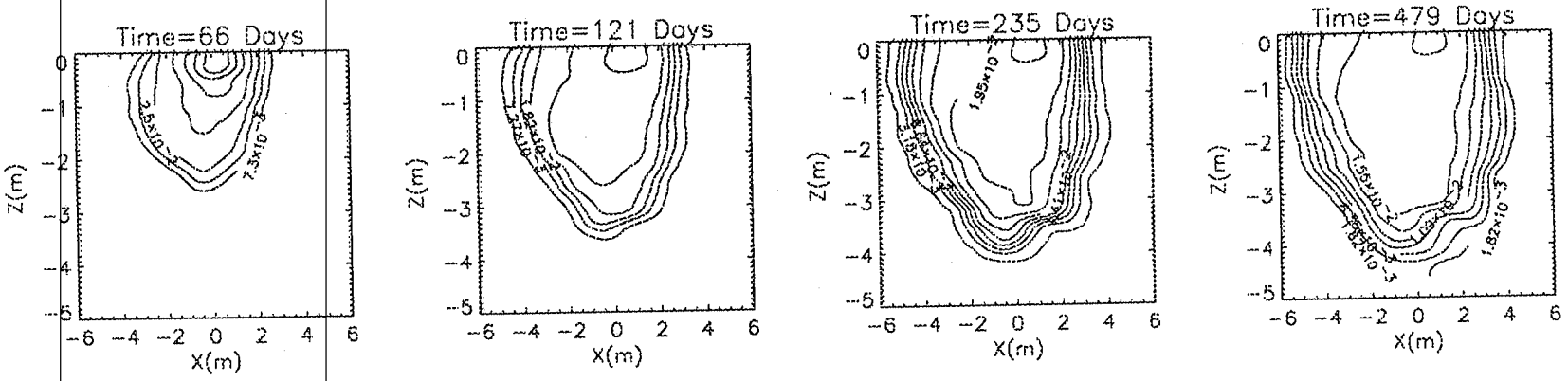


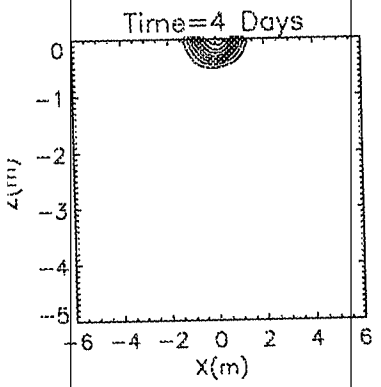
Realization 8

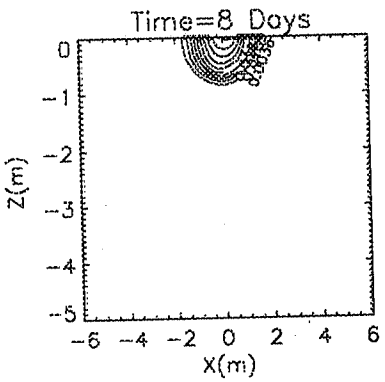


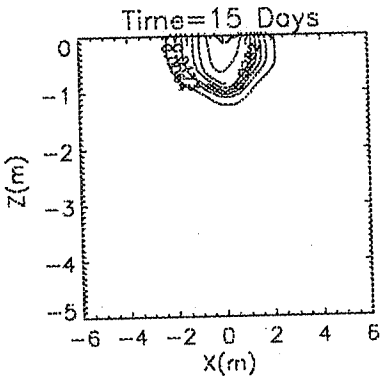


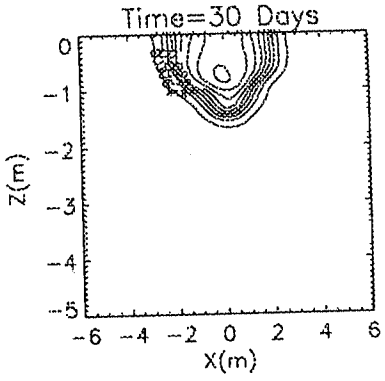
Realization 9



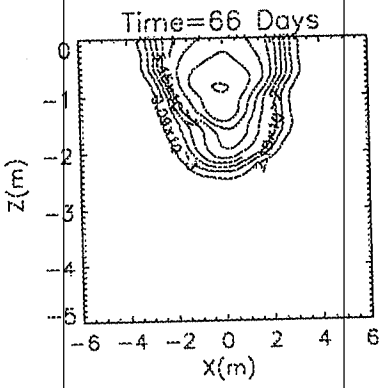


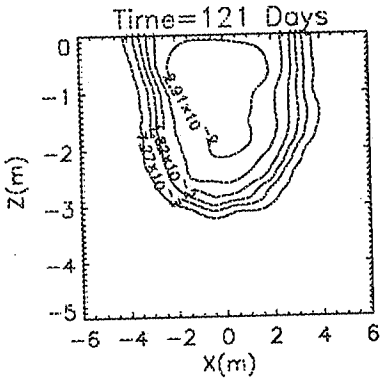


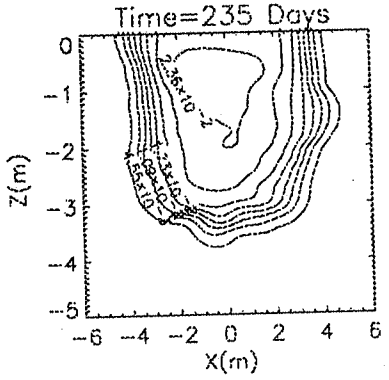


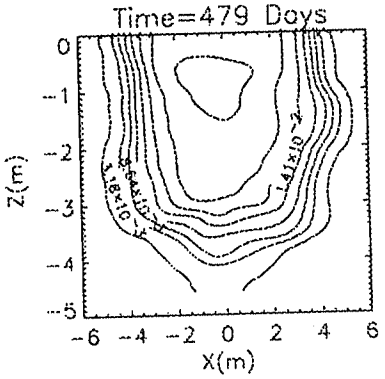


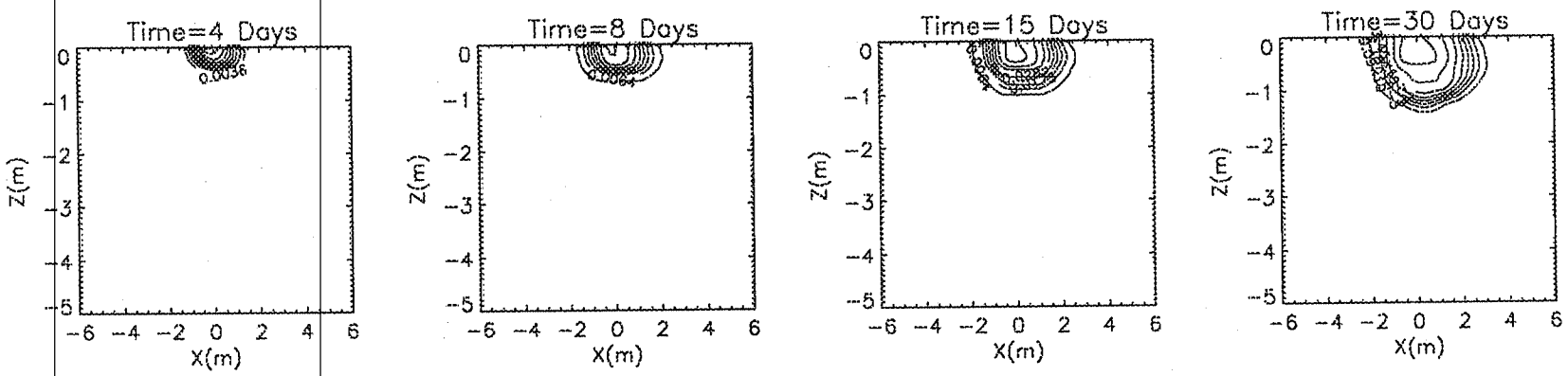
Realization 10



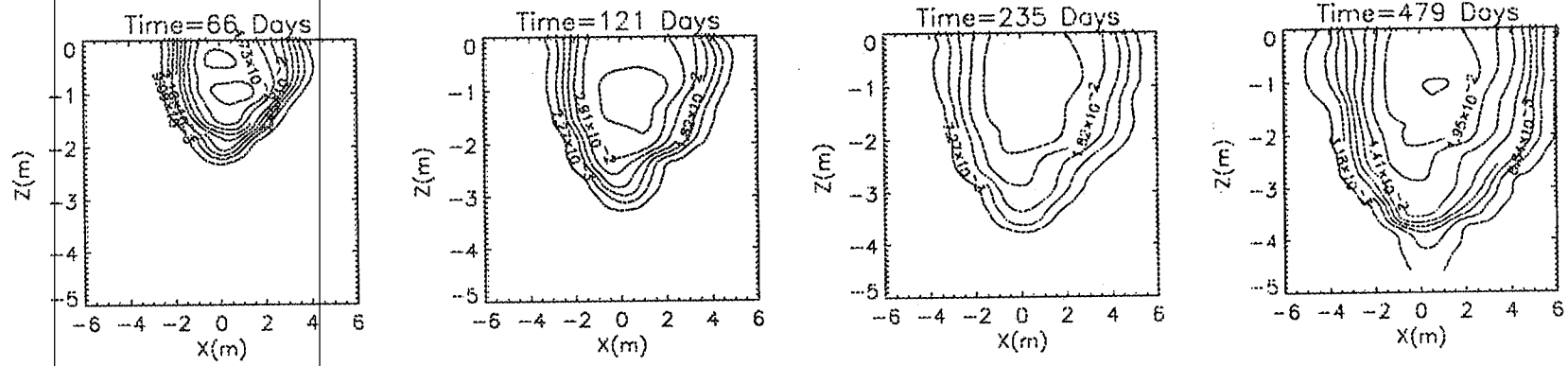


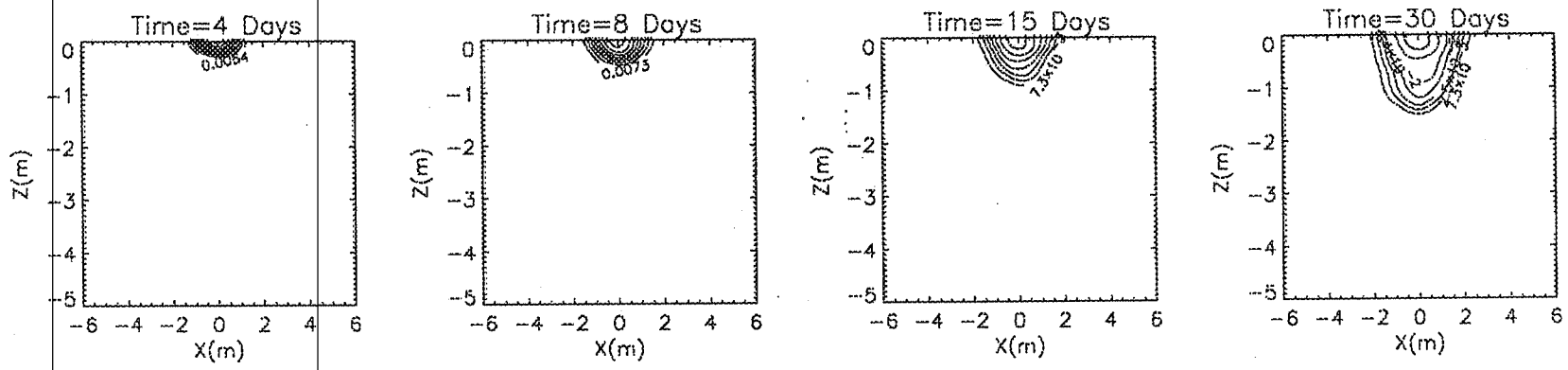




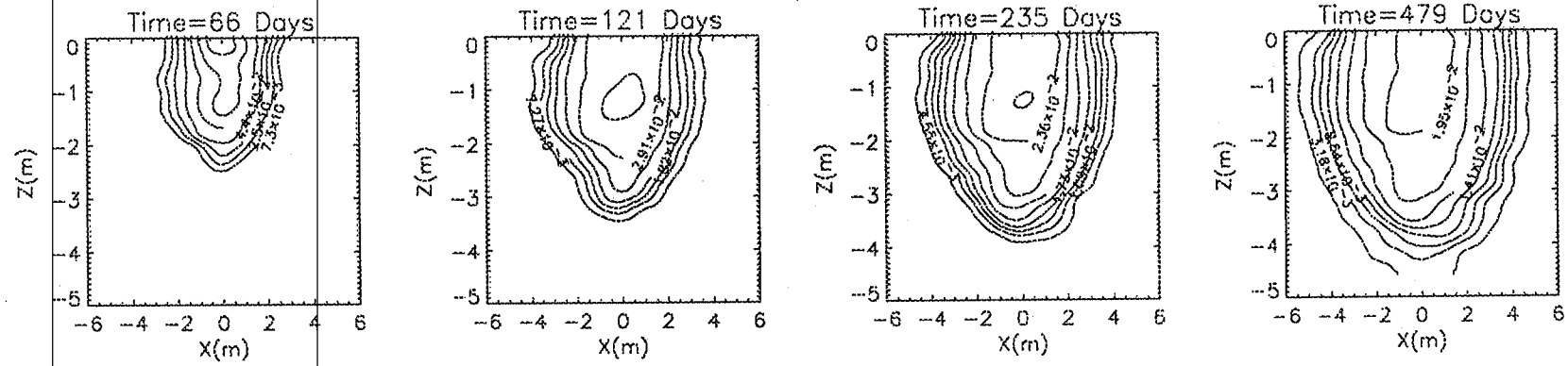


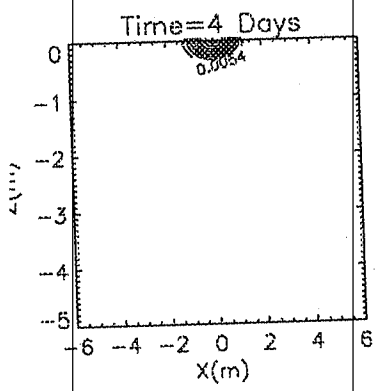
Realization 11

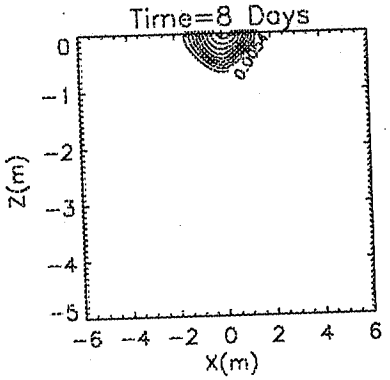


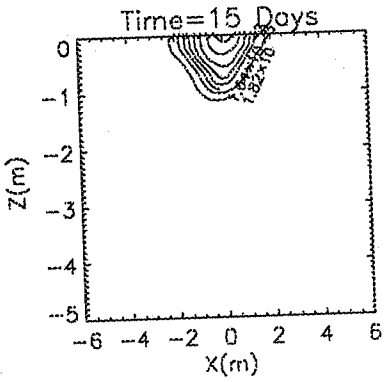


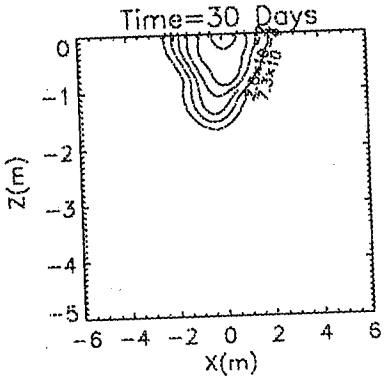
Realization 12



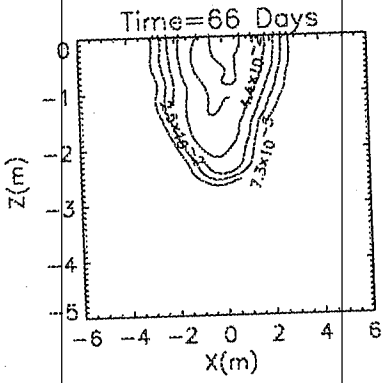


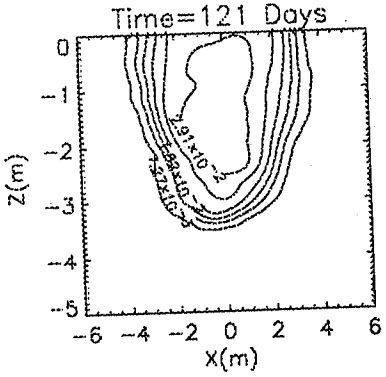


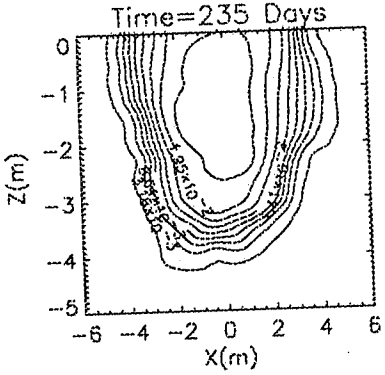


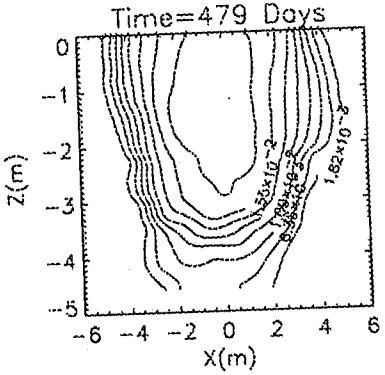


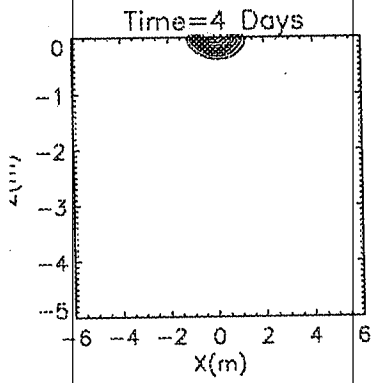
Realization 13

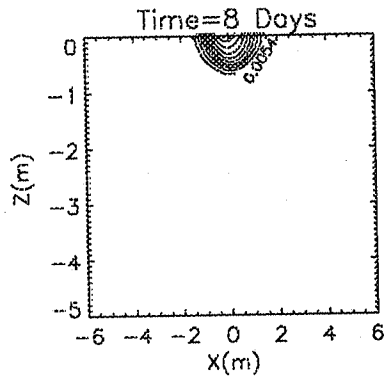


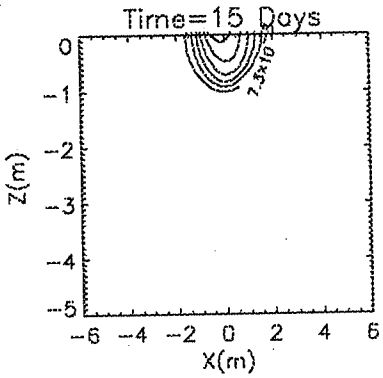


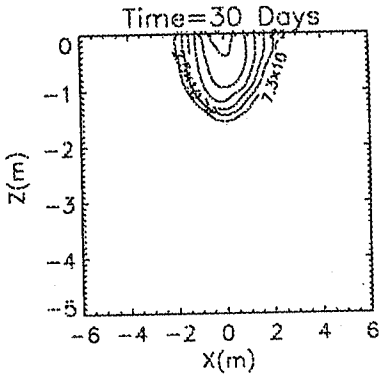




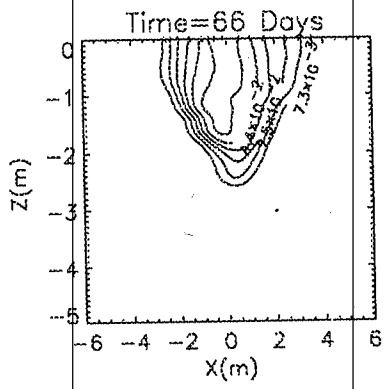


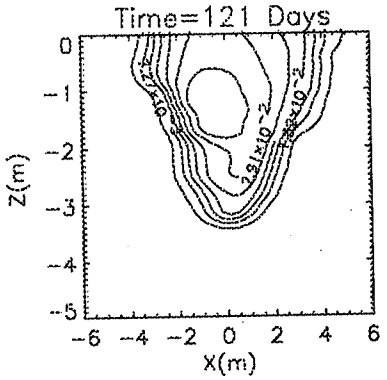


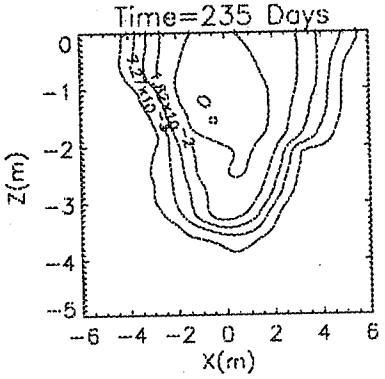


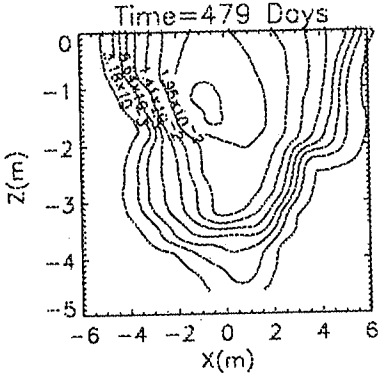


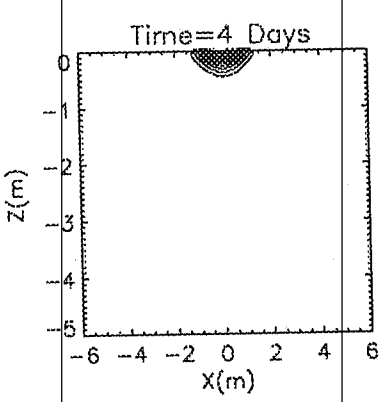
Realization 14

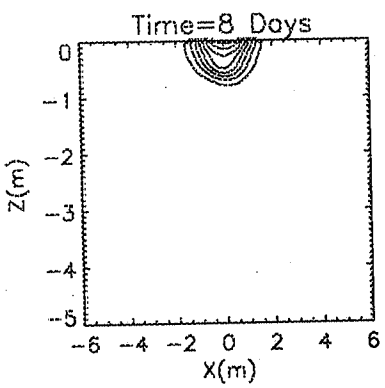


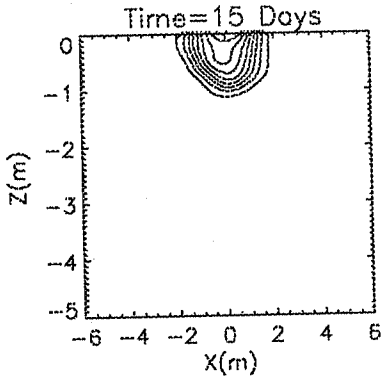


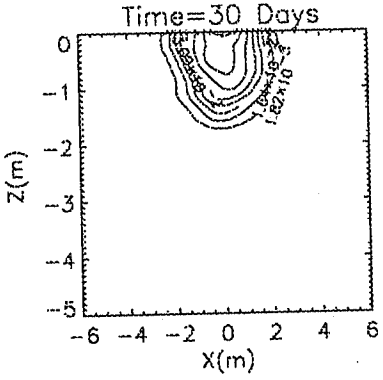




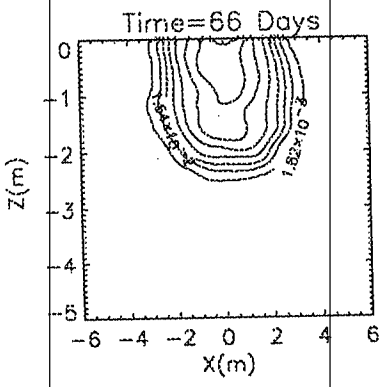


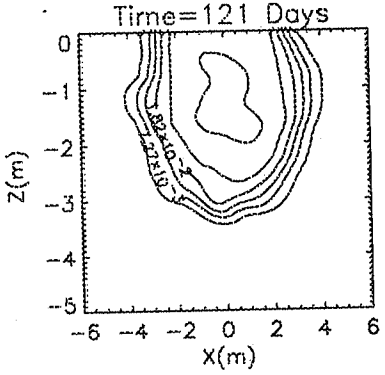


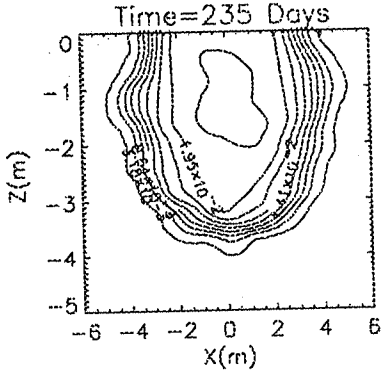


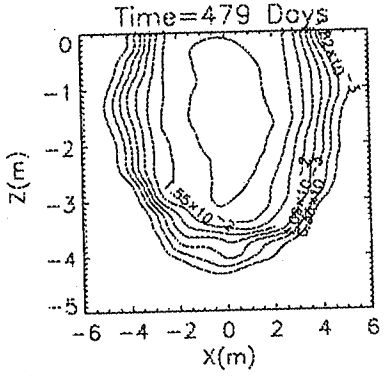


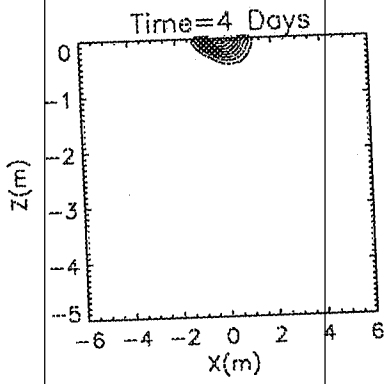
Realization 15

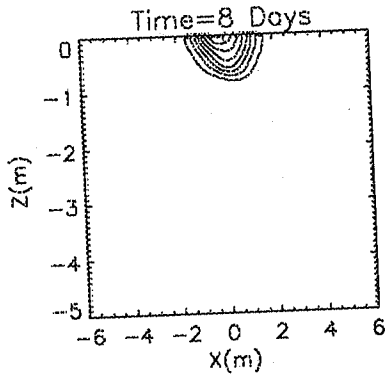


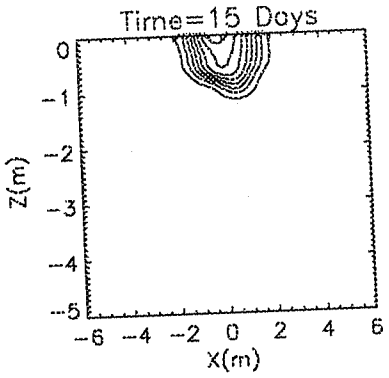


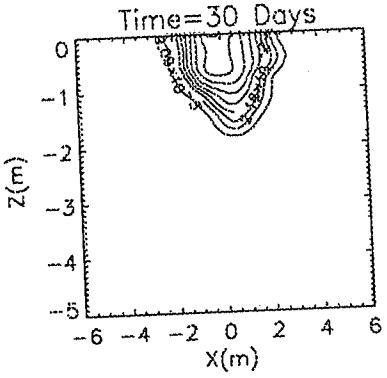




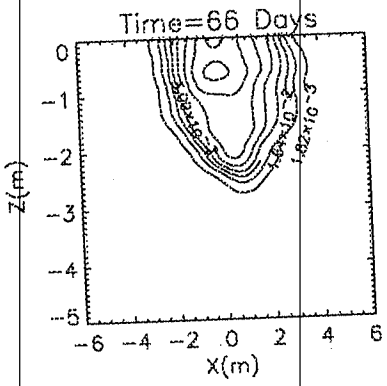


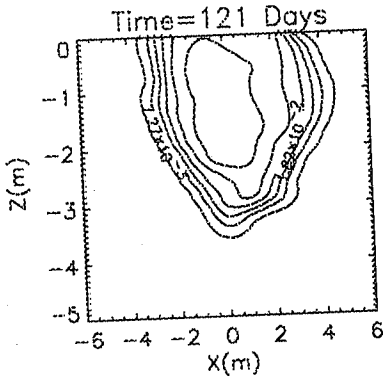


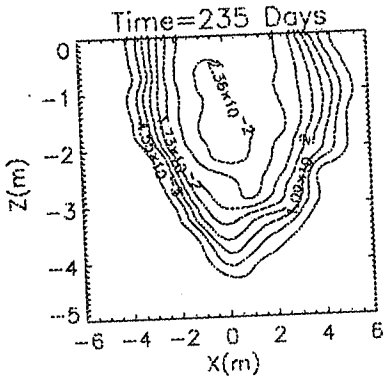


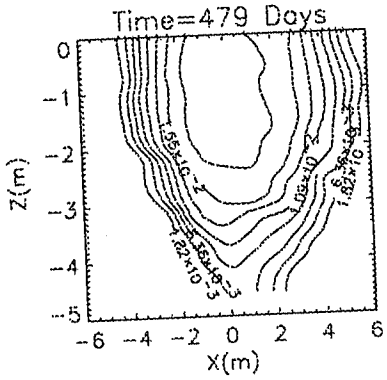


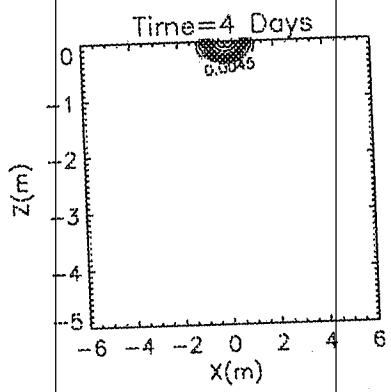
Realization 16

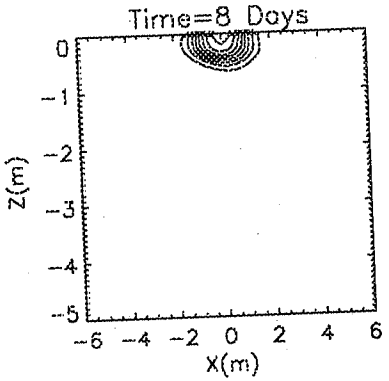


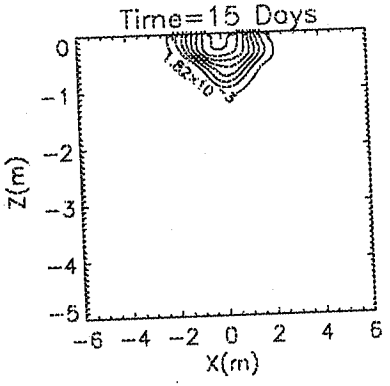


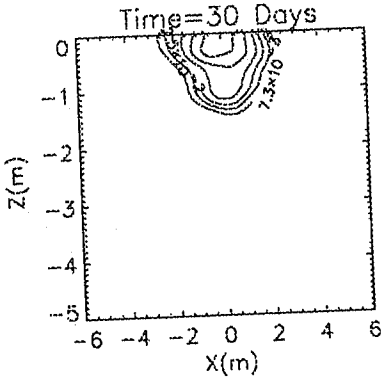




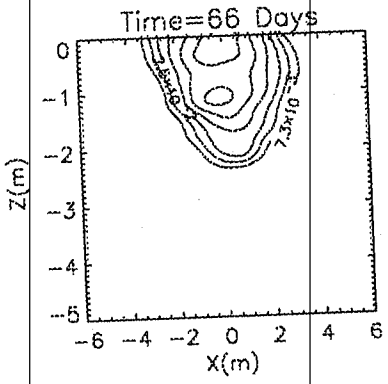


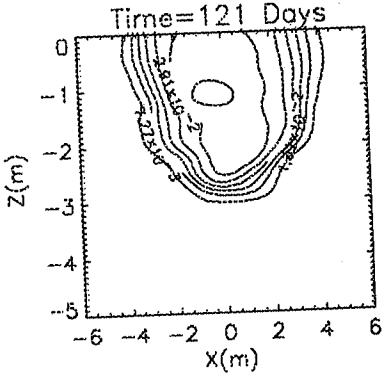


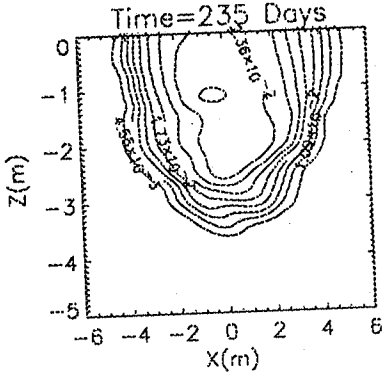


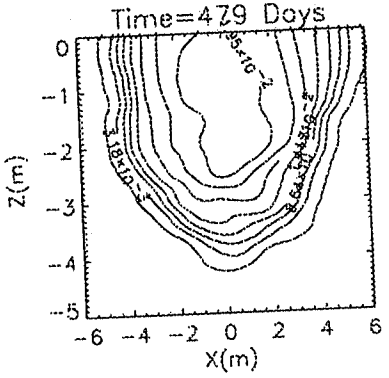


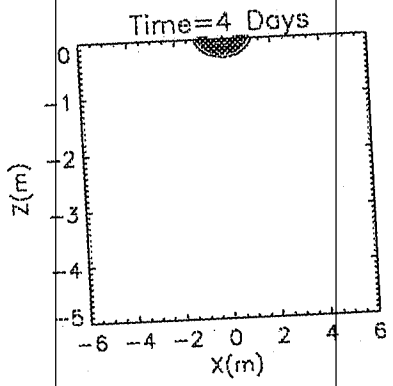
Realization 17

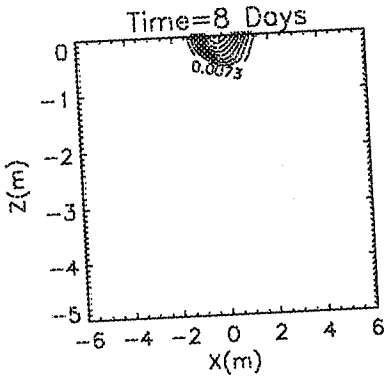


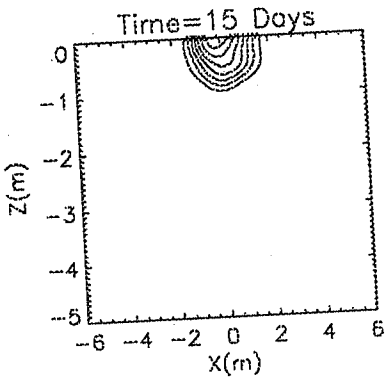


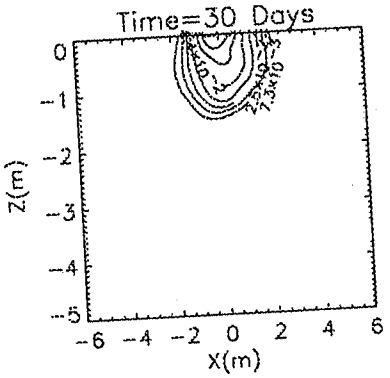




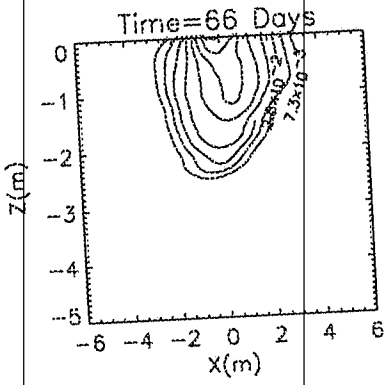


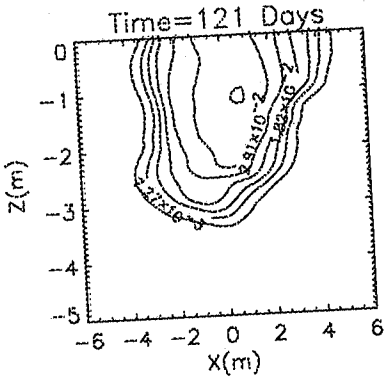


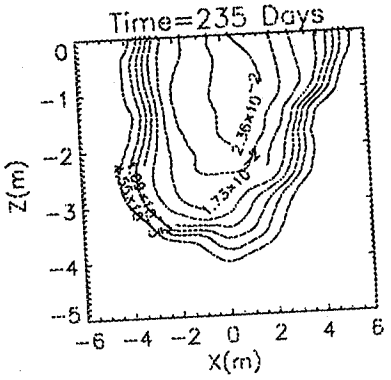


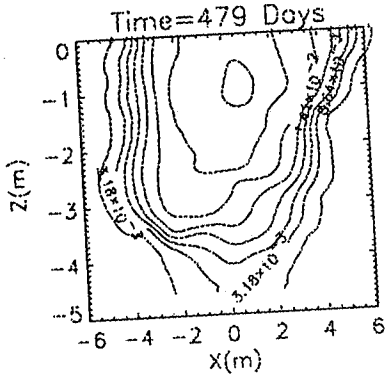


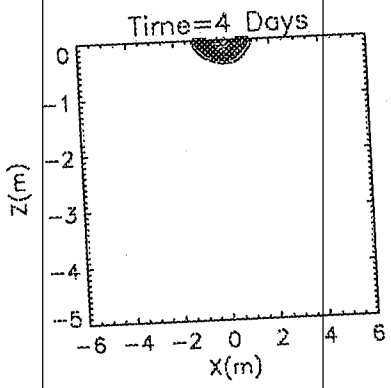
Realization 18

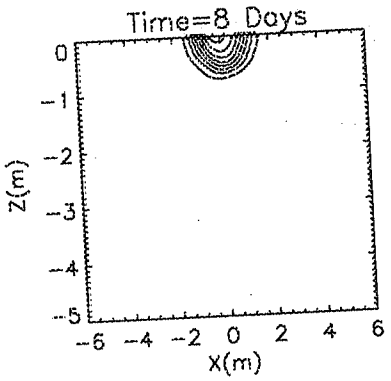


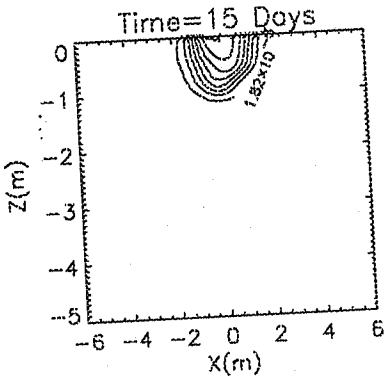


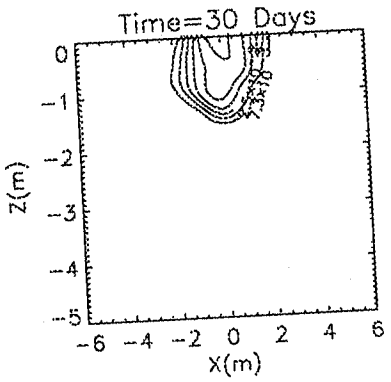




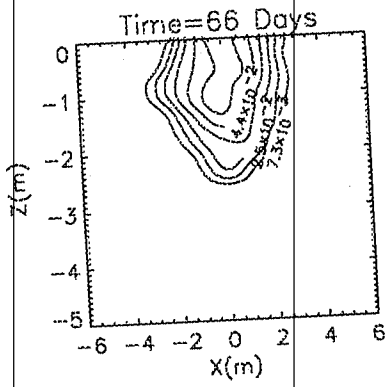


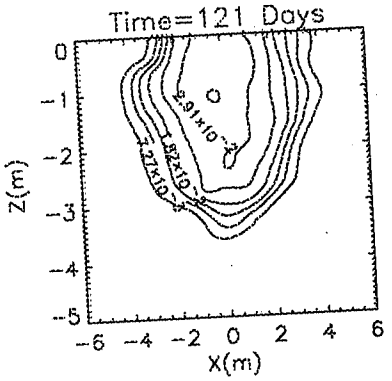


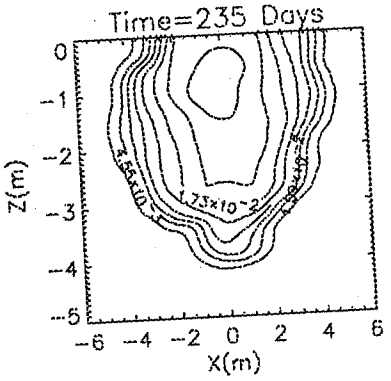


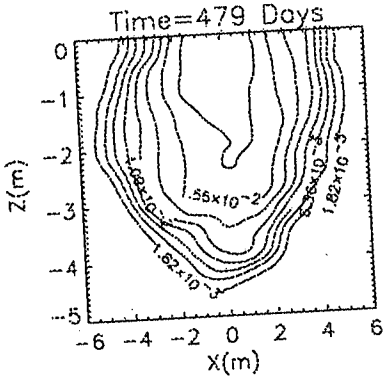


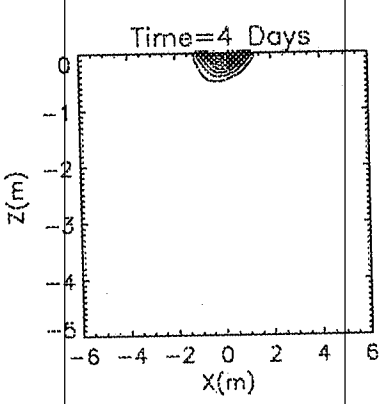
Realization 19

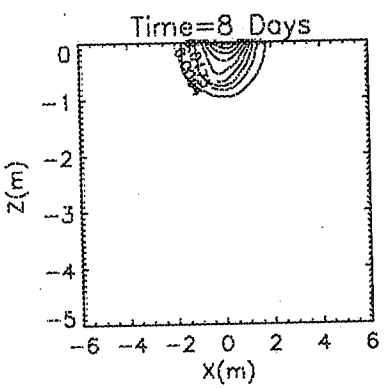


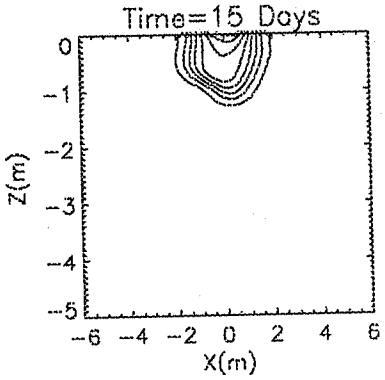


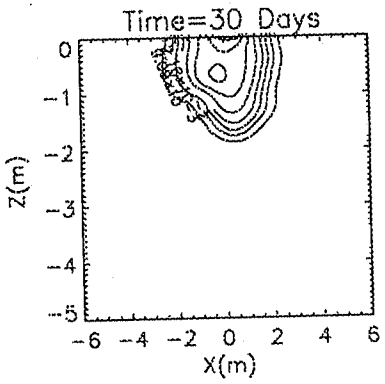




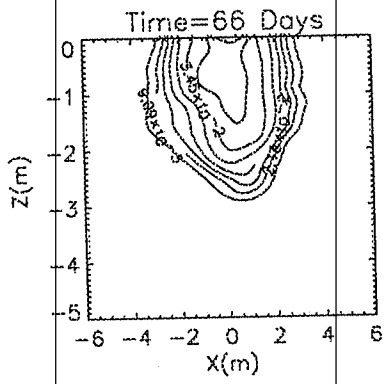


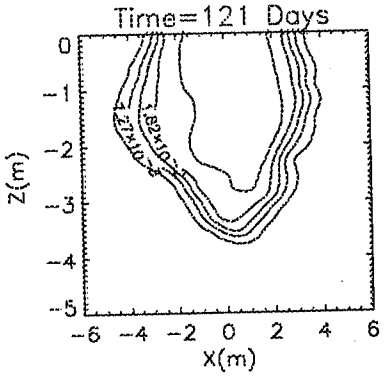


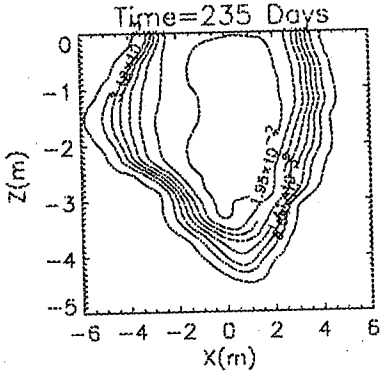


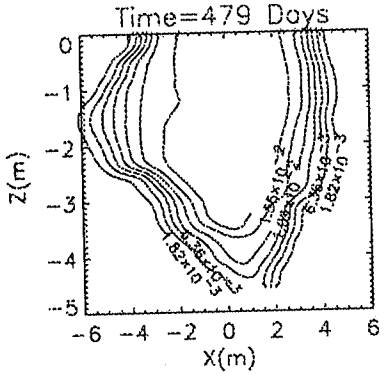


Realization 20



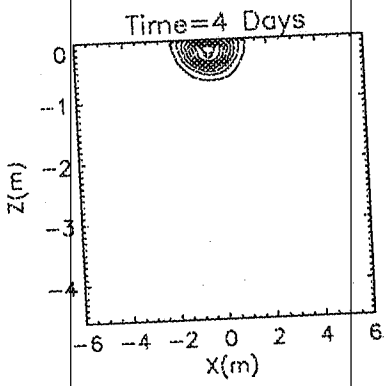


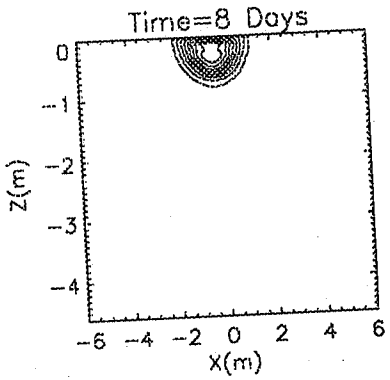


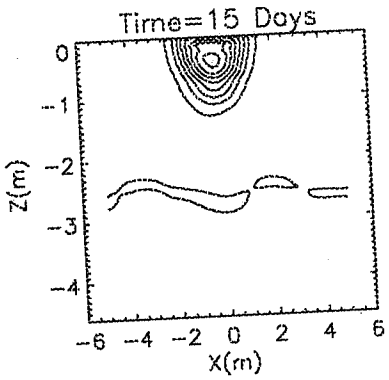


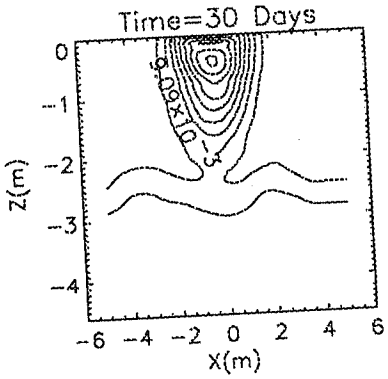
Appendix D: Simulated Water Content Change and Bromide Concentration Realizations

Figure D.2.
Simulated Water Content Change Realizations For Two-Scale Permeability Model.

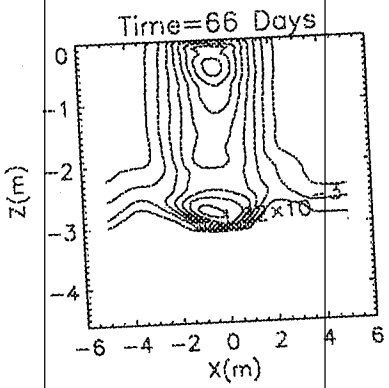


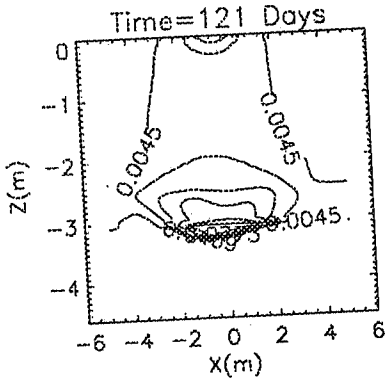


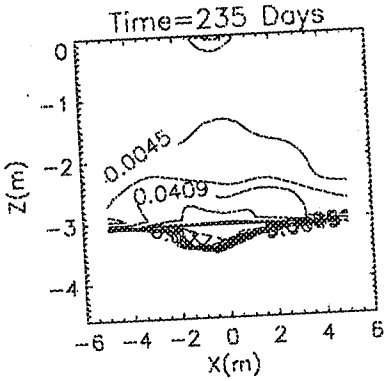


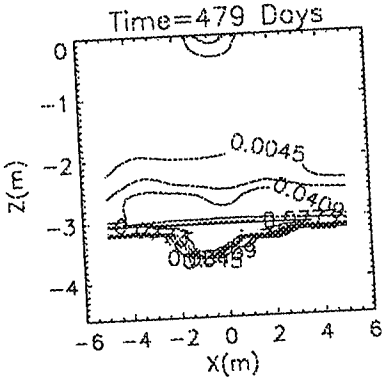


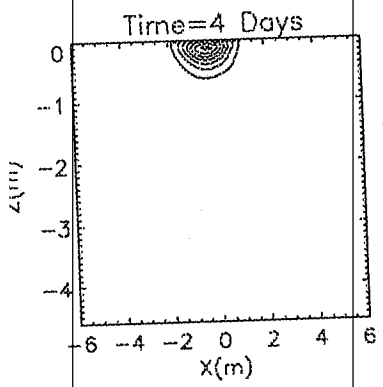
Realization 1

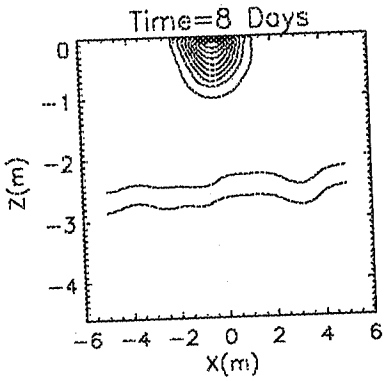


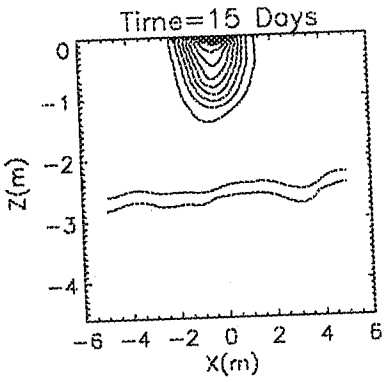


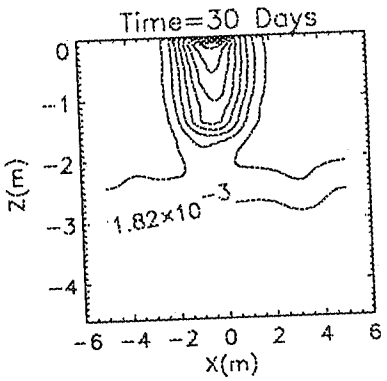




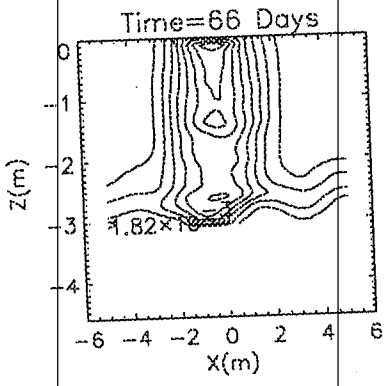


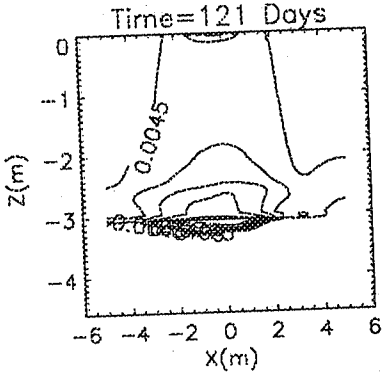


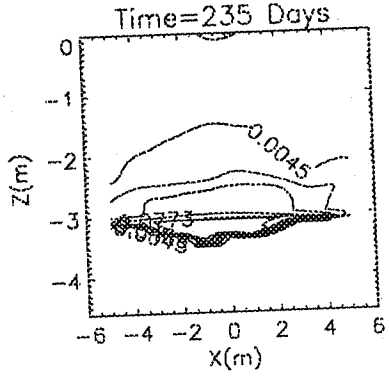


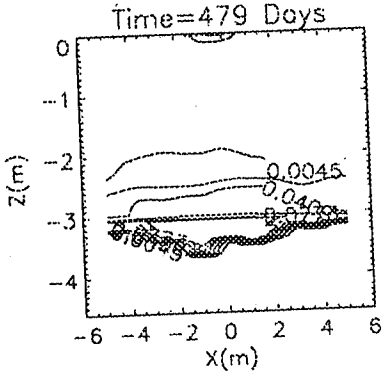


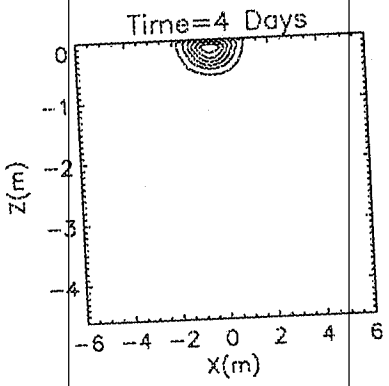
Realization 2

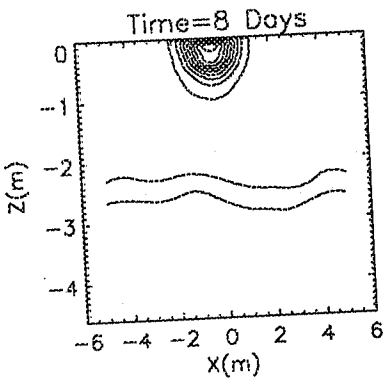


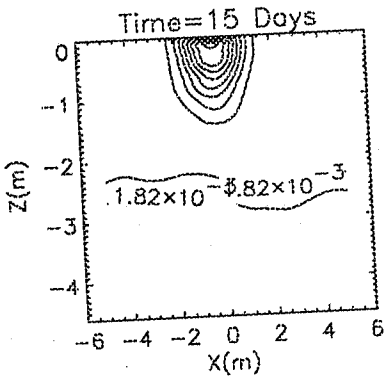


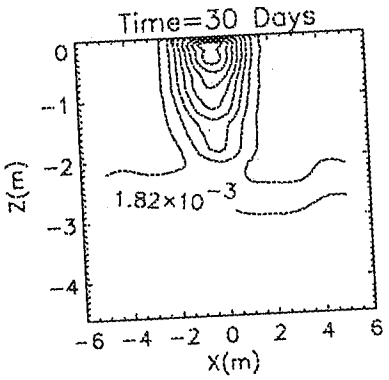




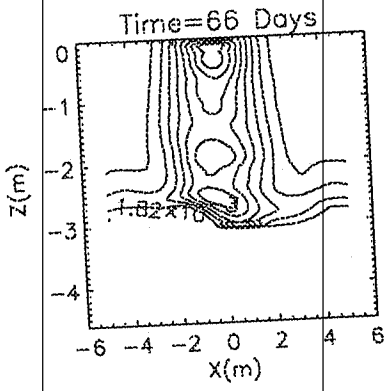


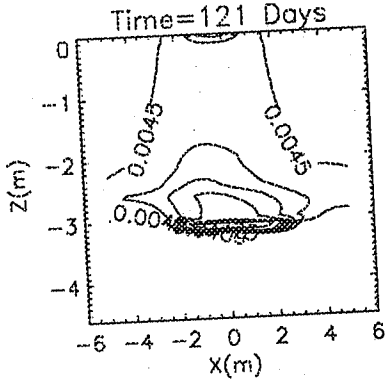


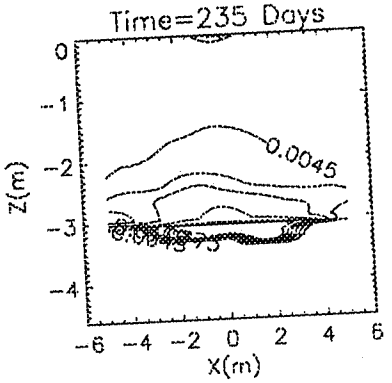


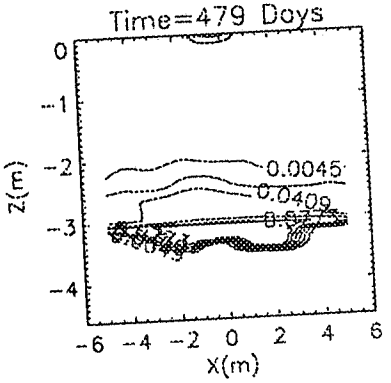


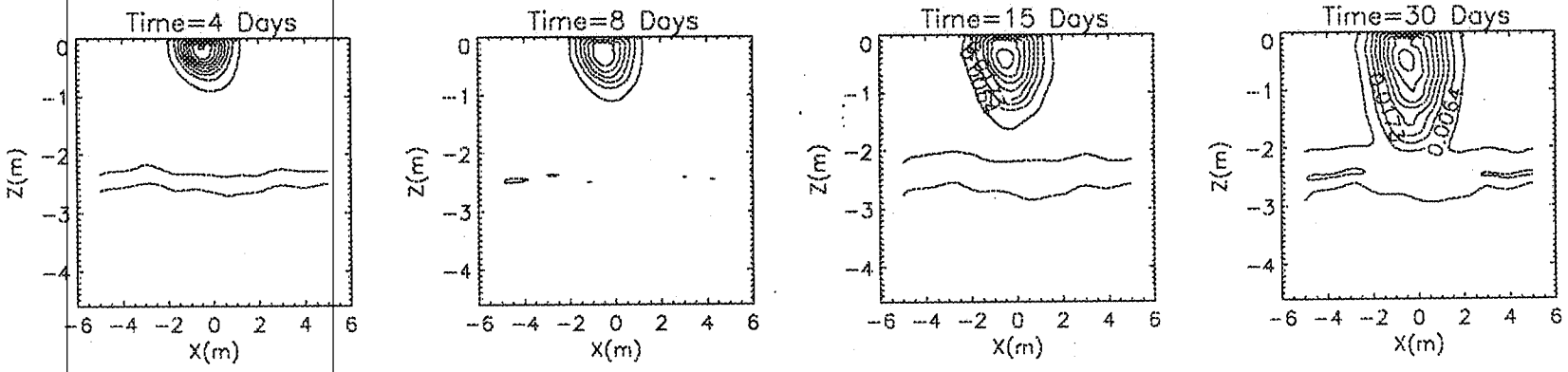
Realization 3



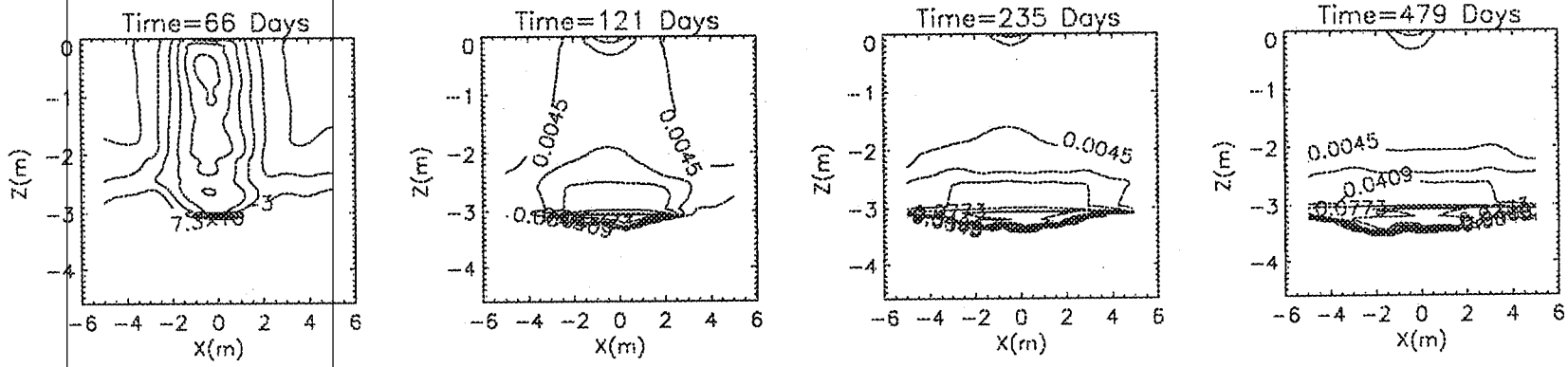


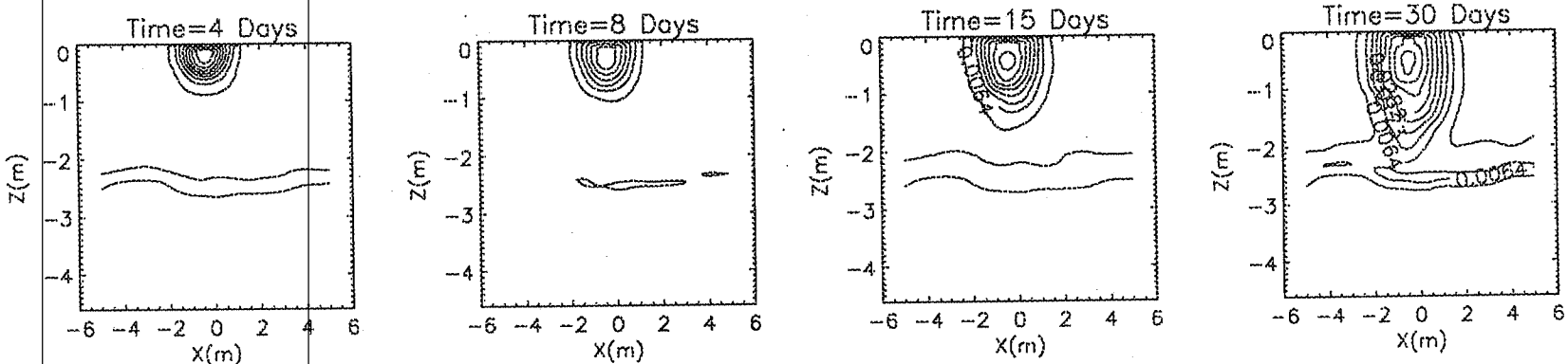




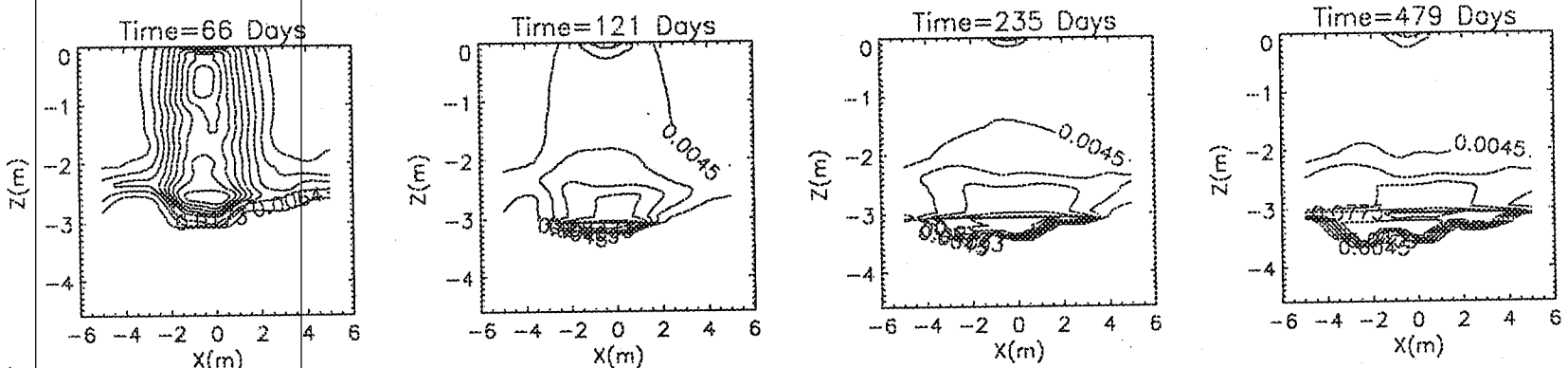


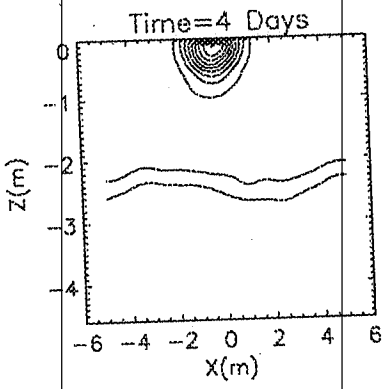
Realization 4

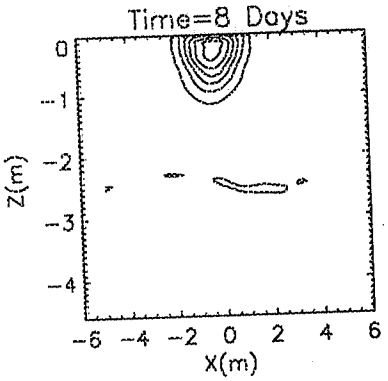


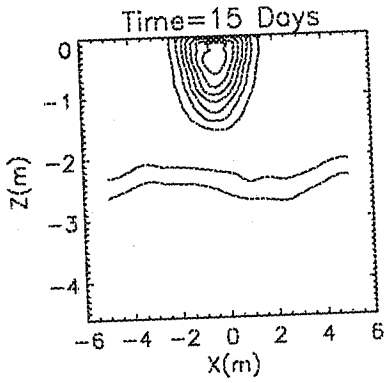


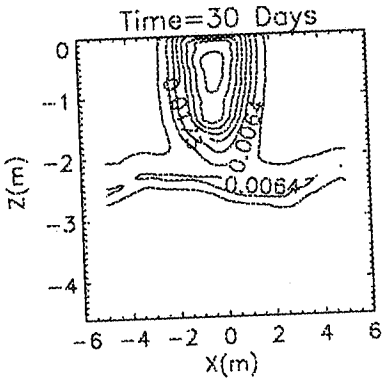
Realization 5



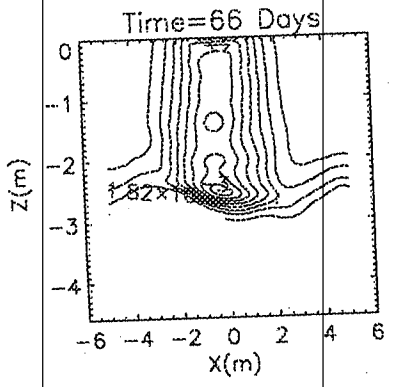


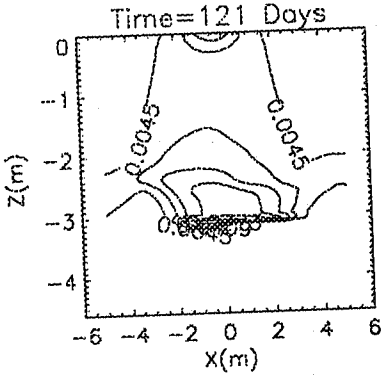


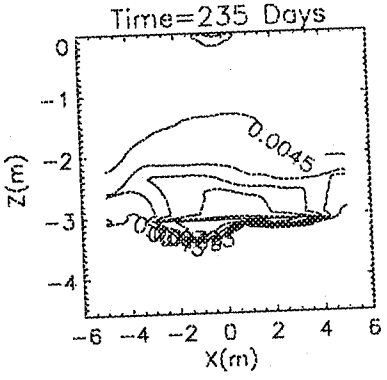


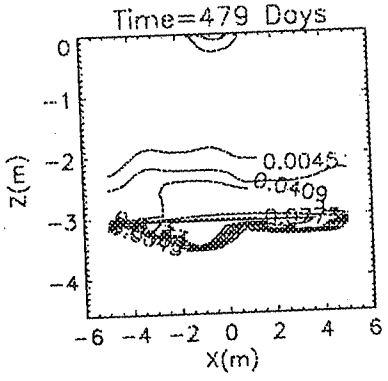


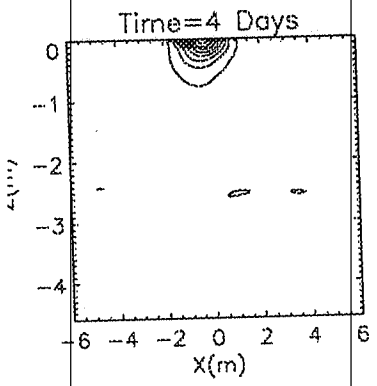
Realization 6

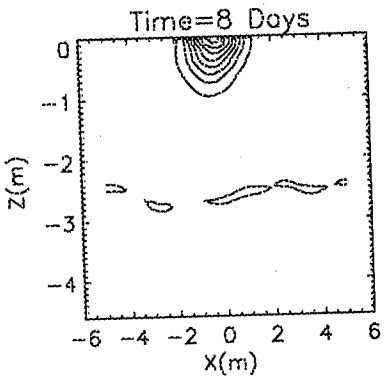


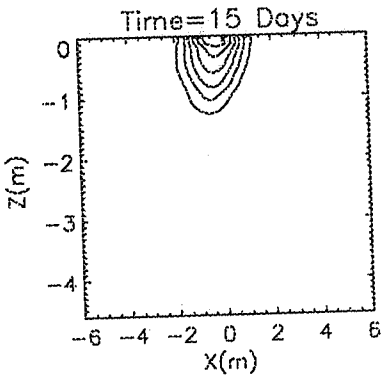


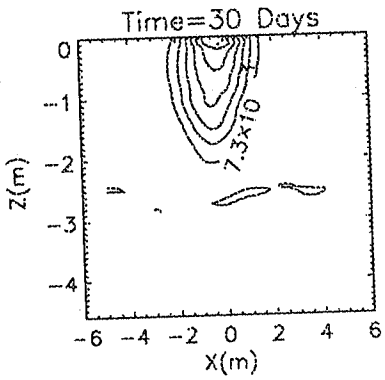




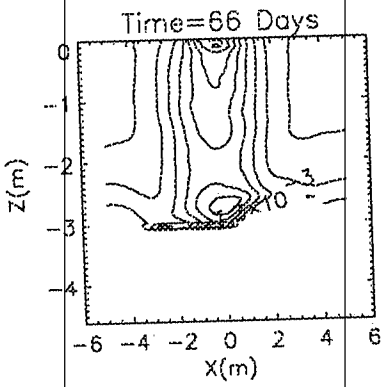


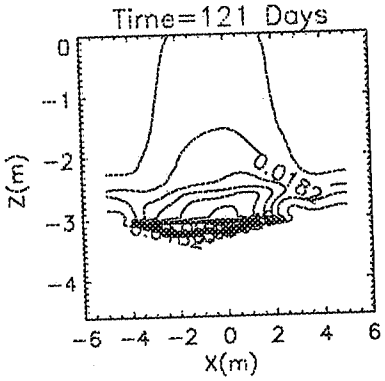


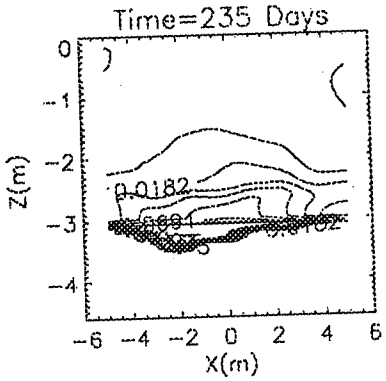


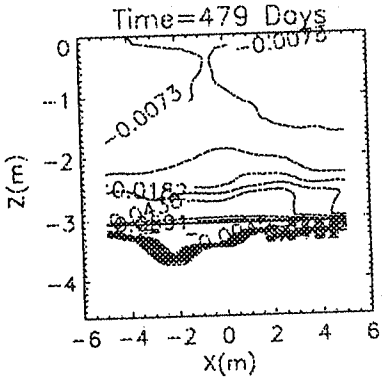


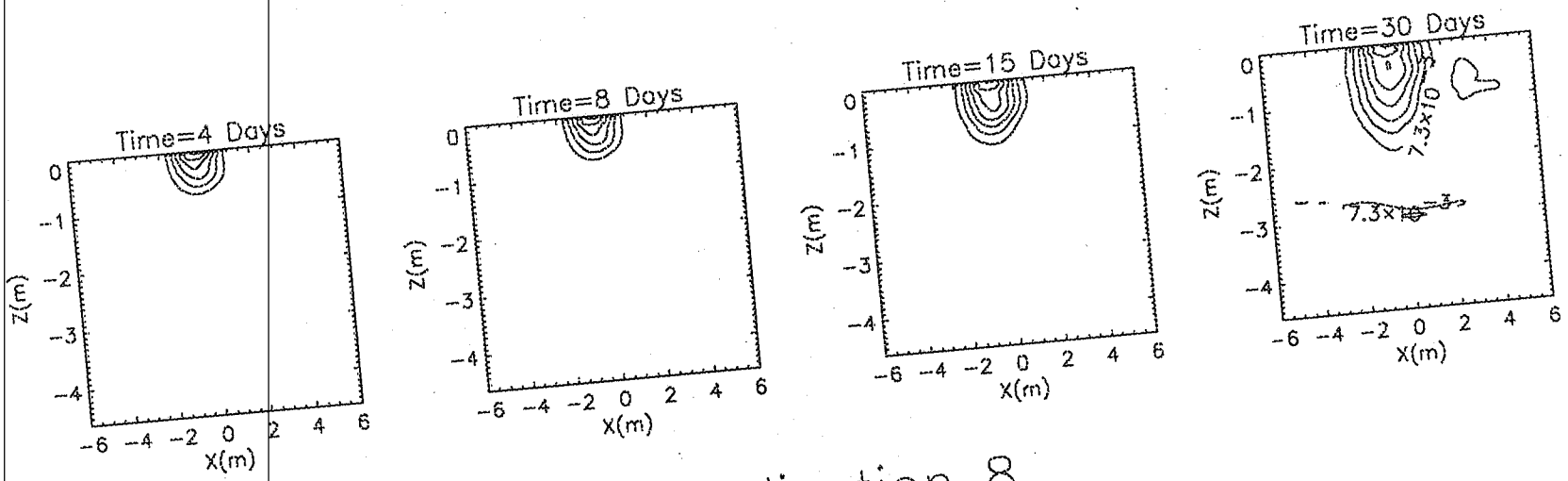
Realization 7



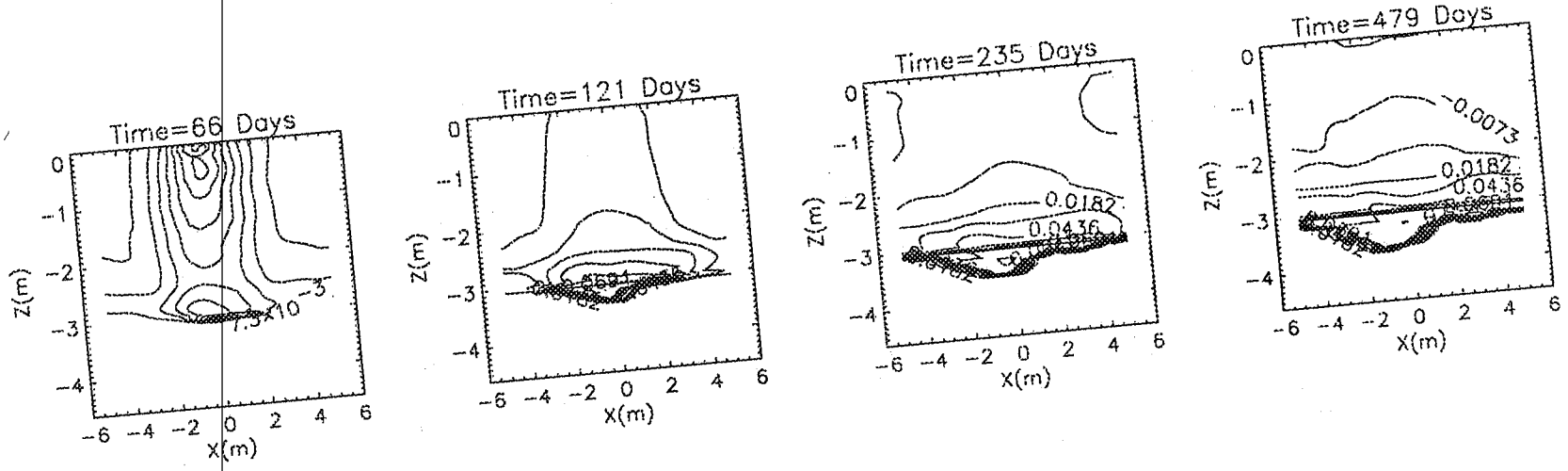


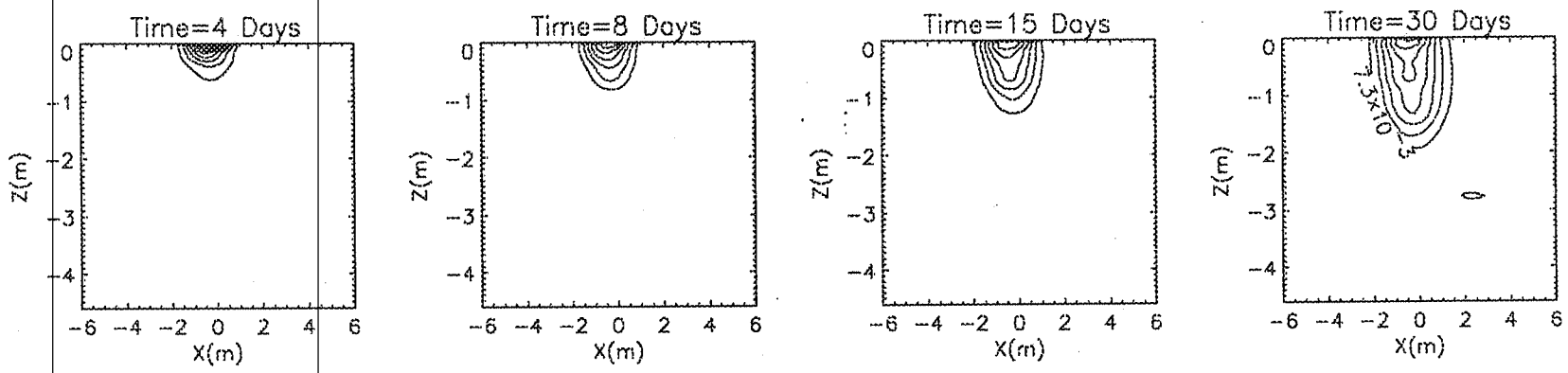




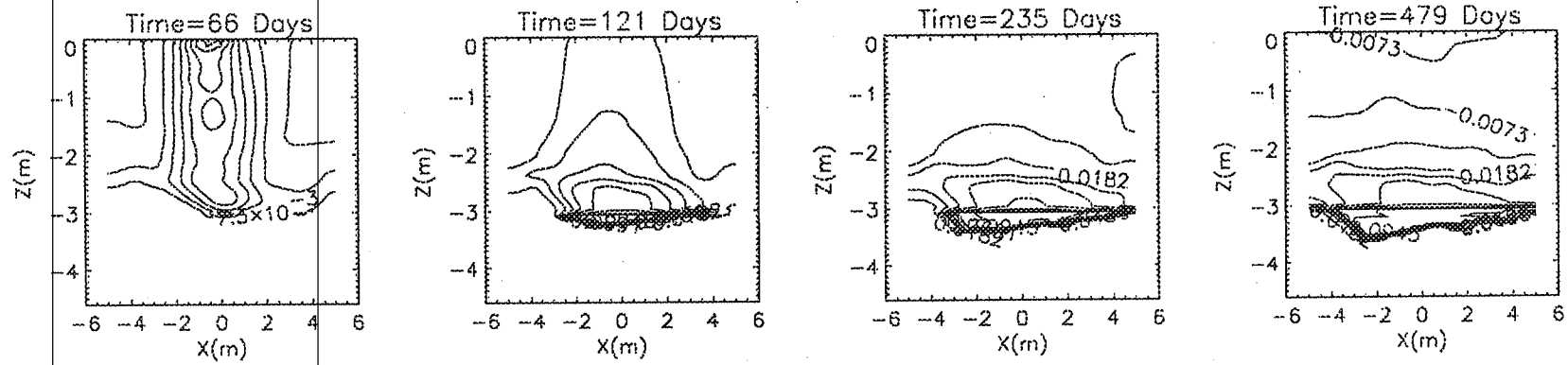


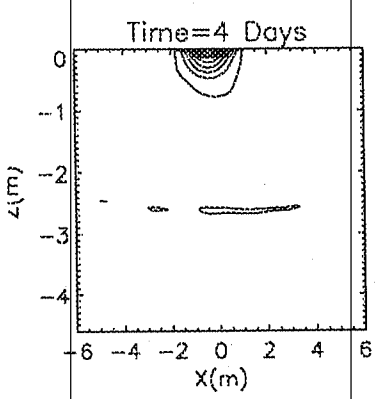
Realization 8

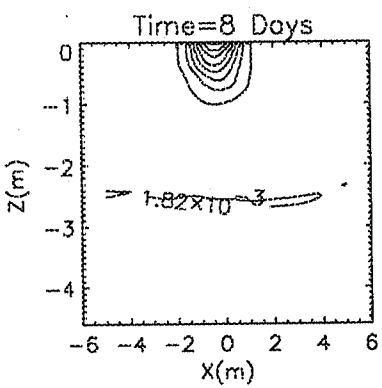


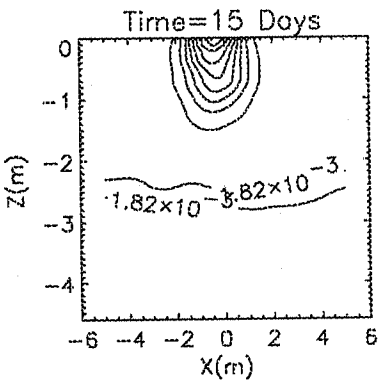


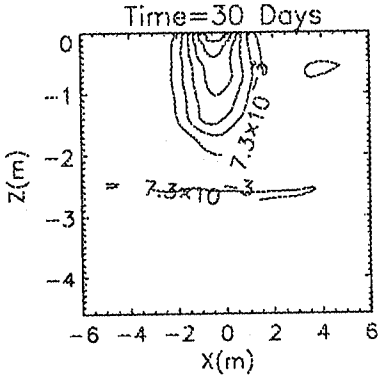
Realization 9



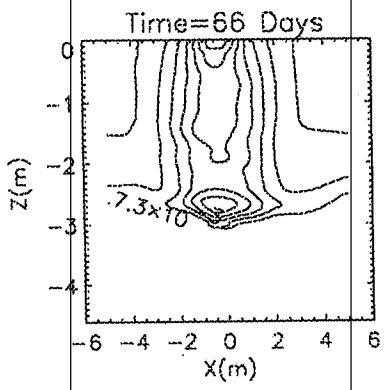


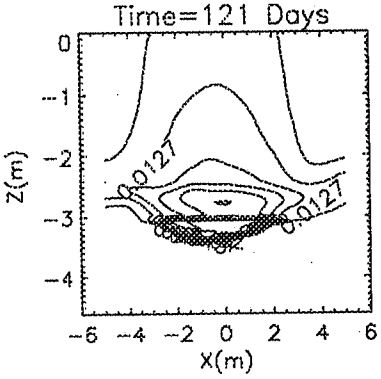


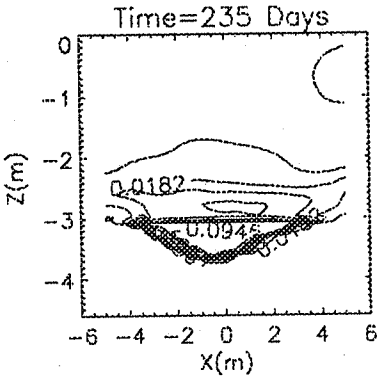


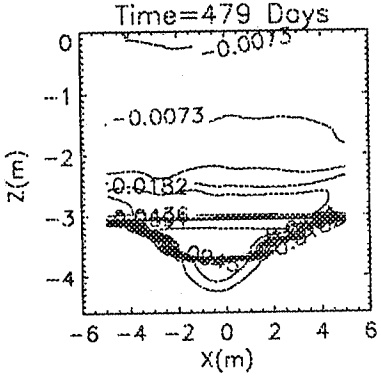


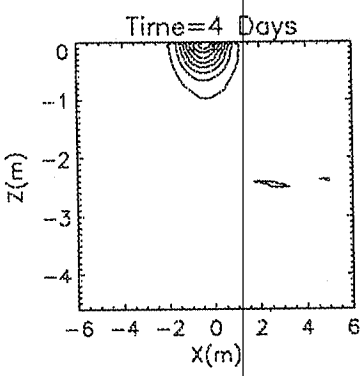
Realization 10

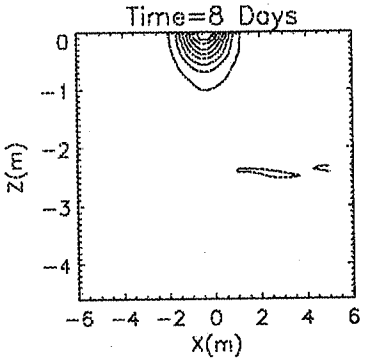


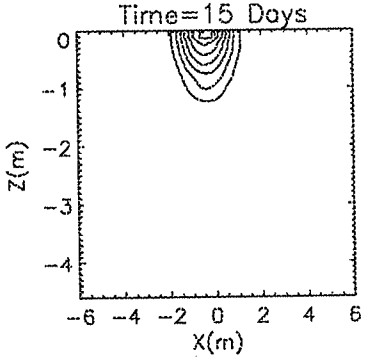


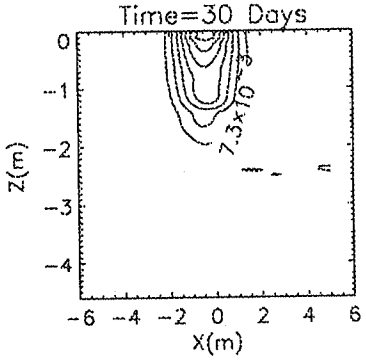




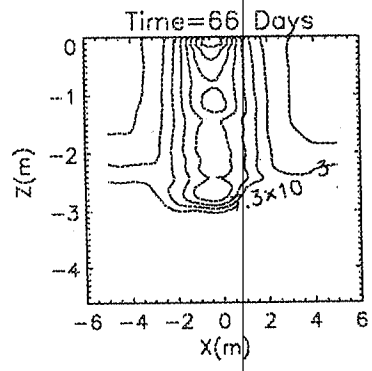


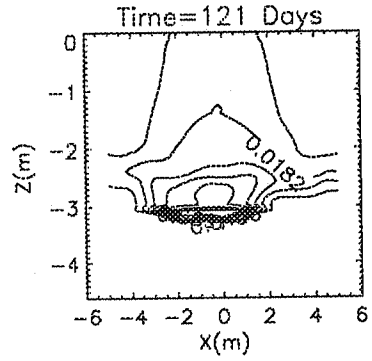


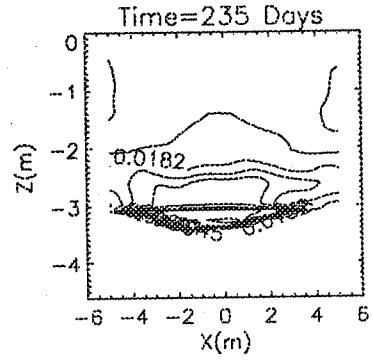


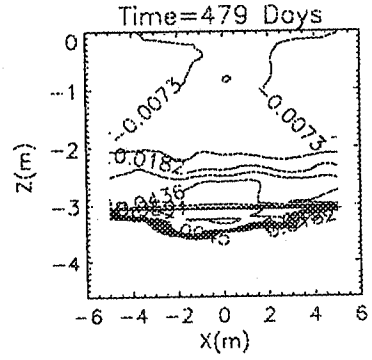


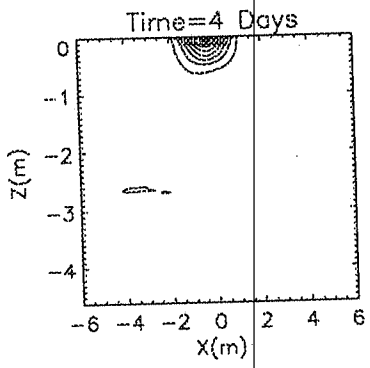
Realization 11

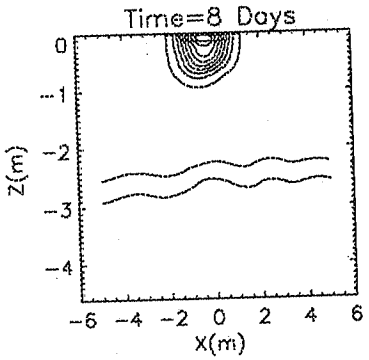


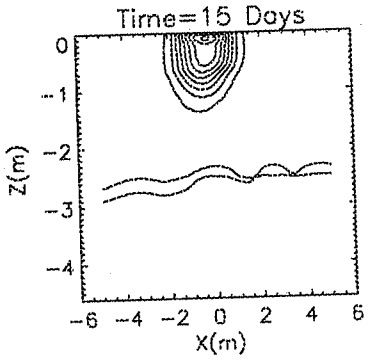


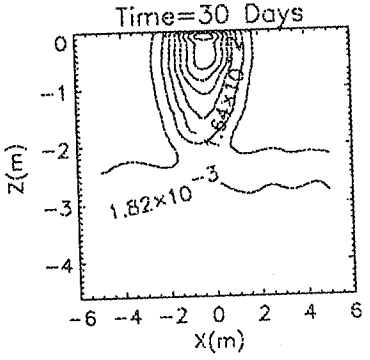




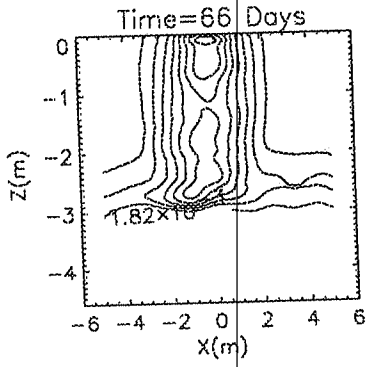


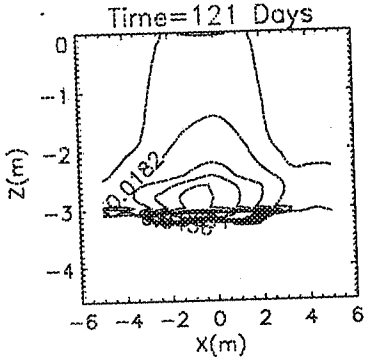


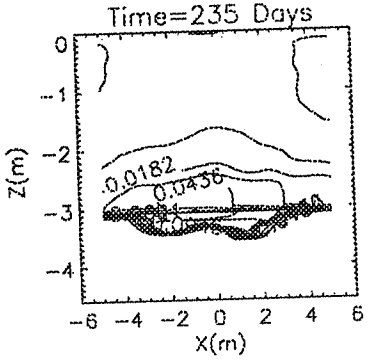


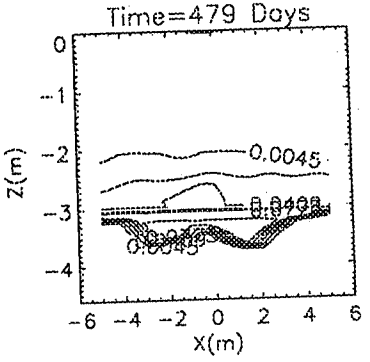


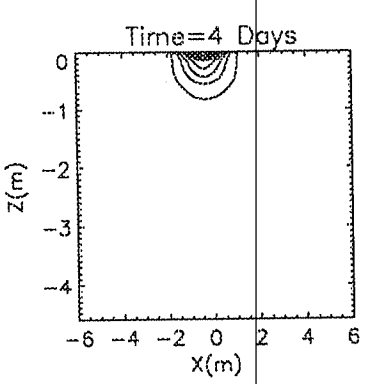
Realization 12

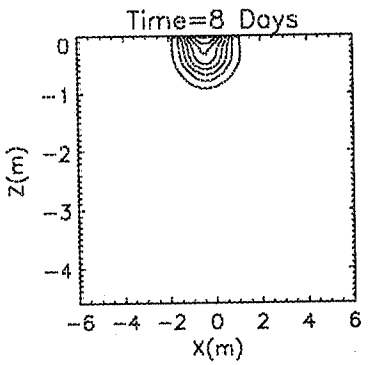


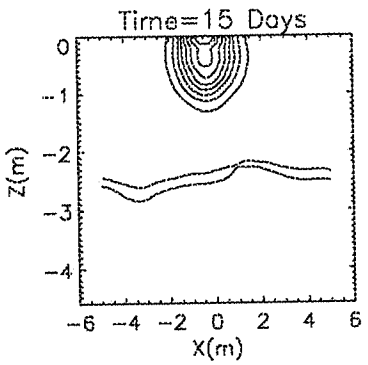


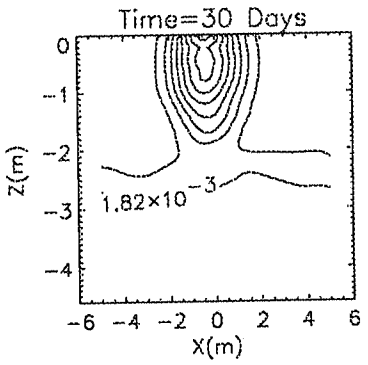




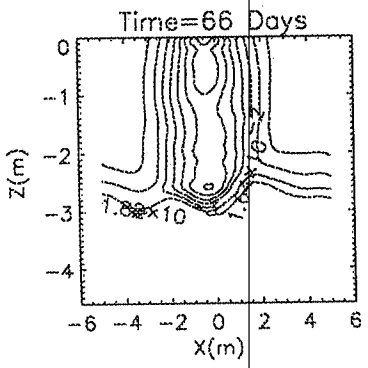


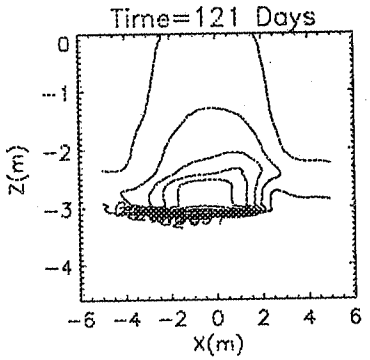


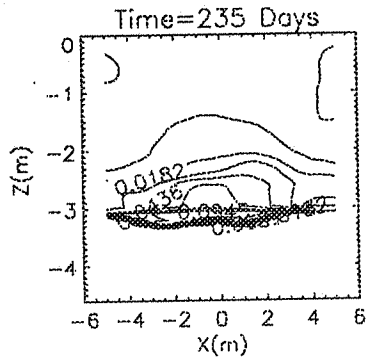


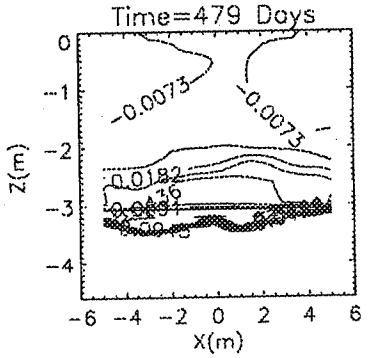


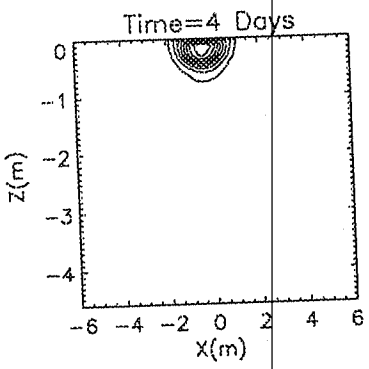
Realization 13

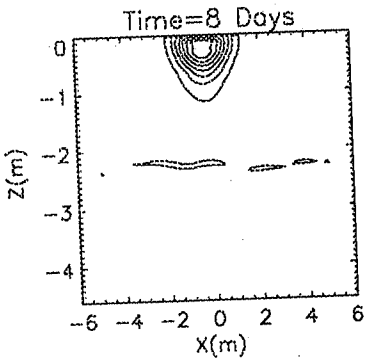


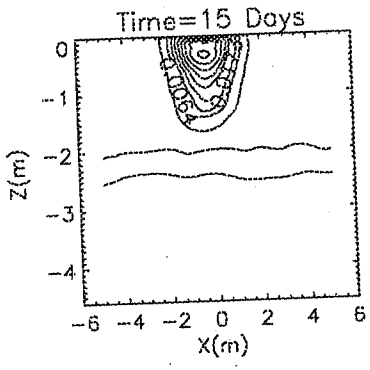


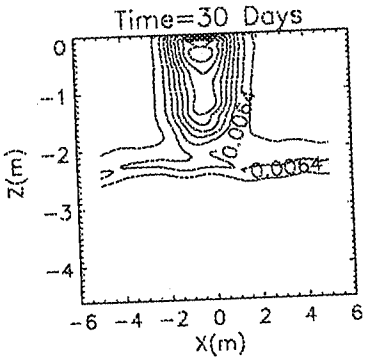




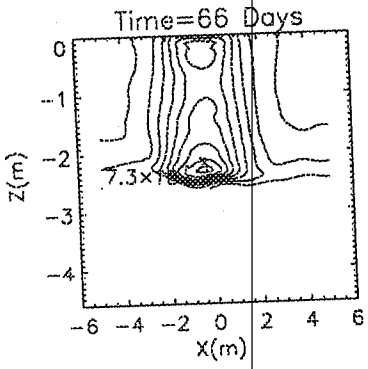


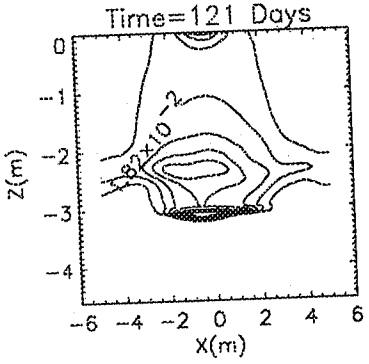


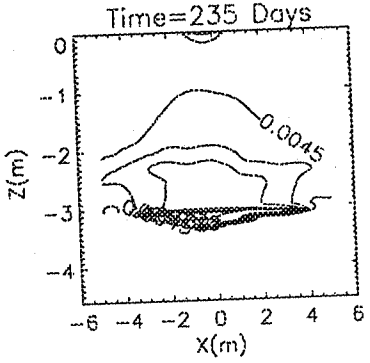


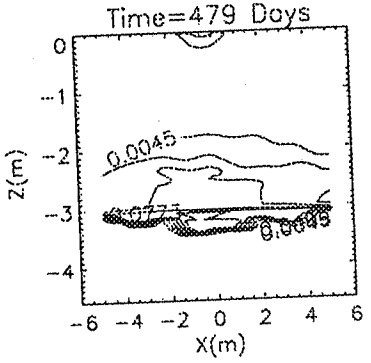


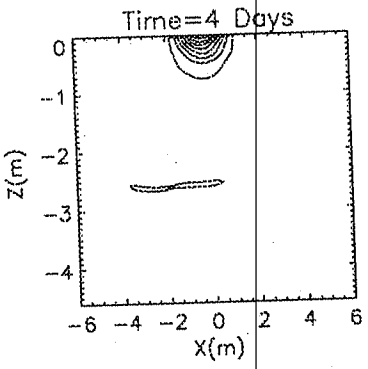
Realization 14

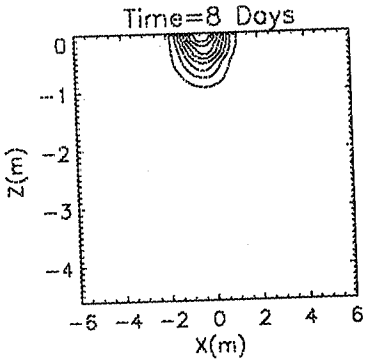


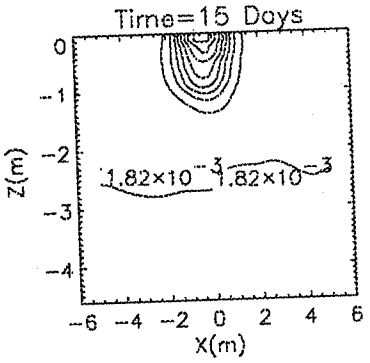


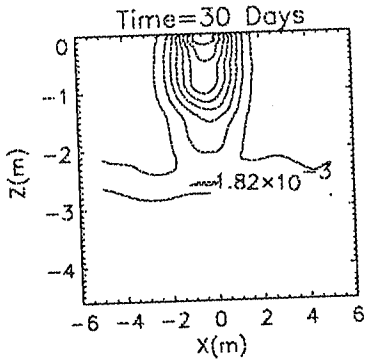




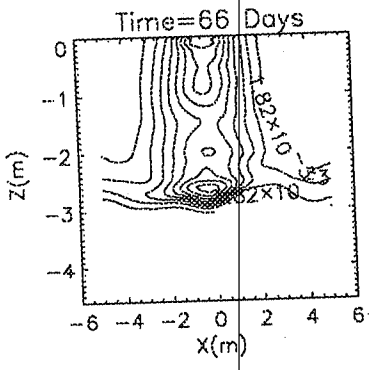


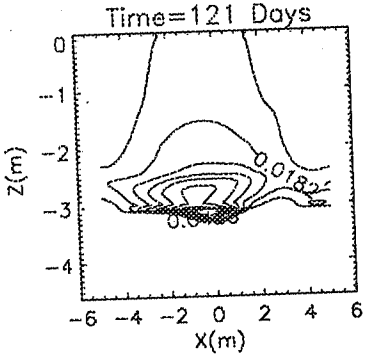


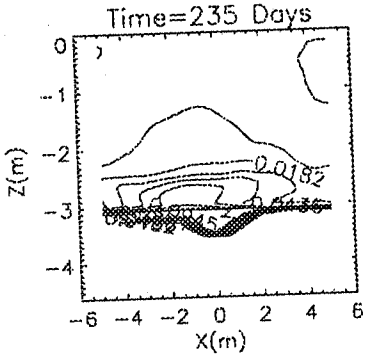


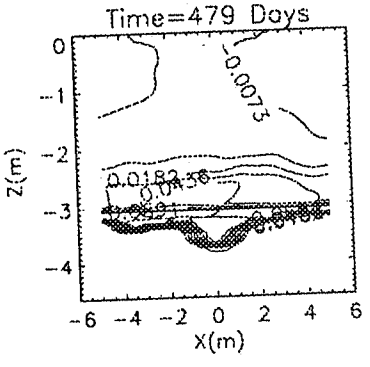


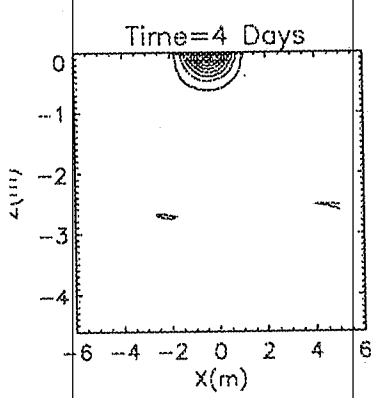
Realization 15

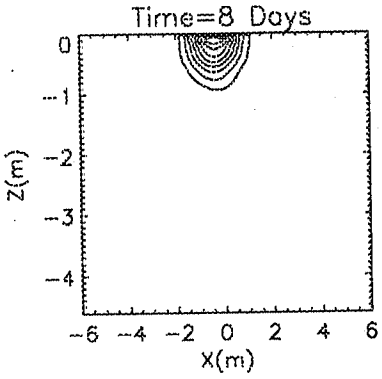


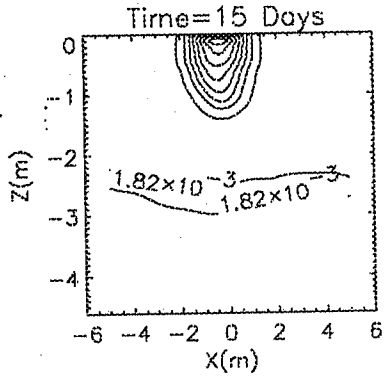


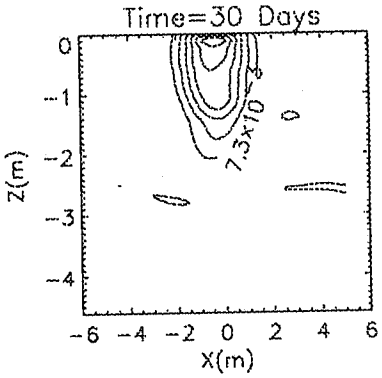




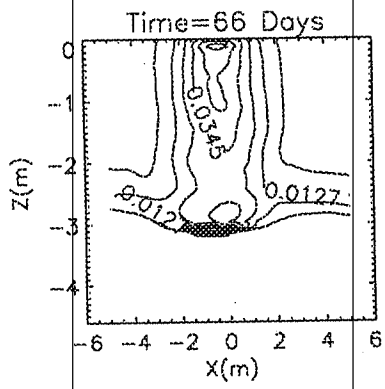


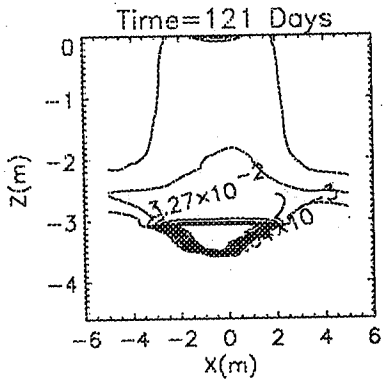


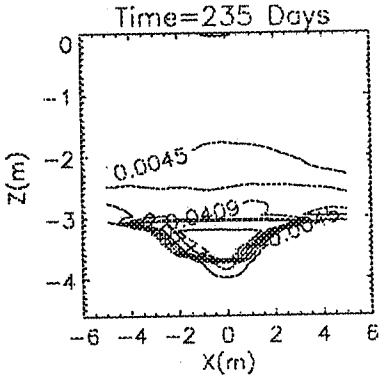


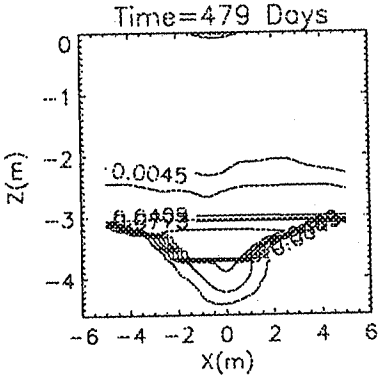


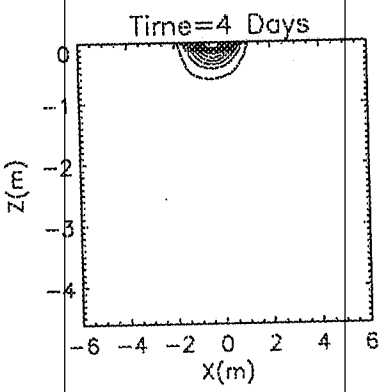
Realization 16

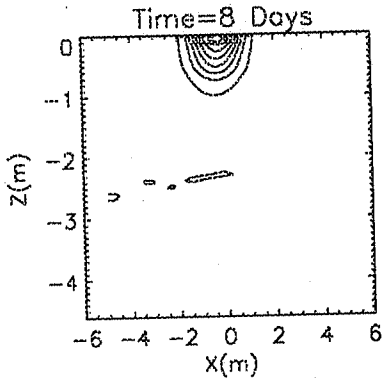


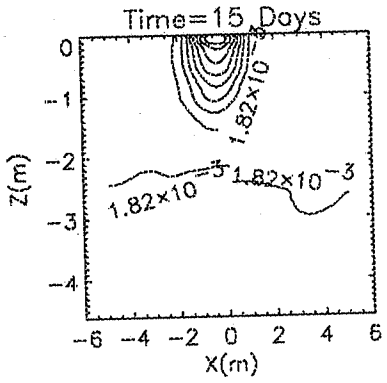


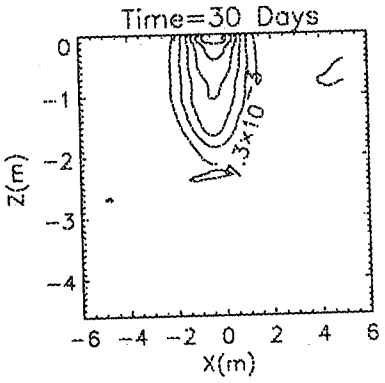




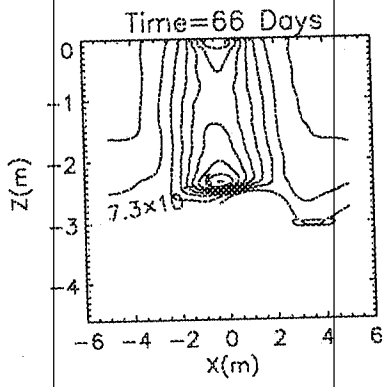


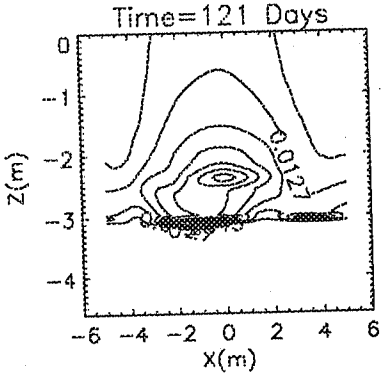


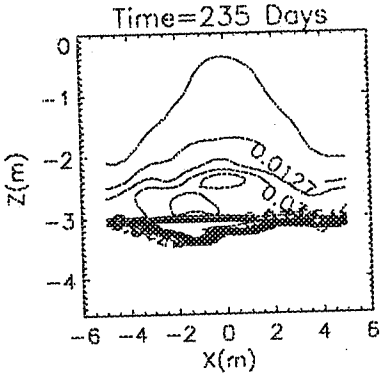


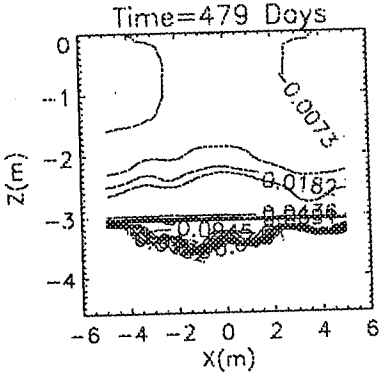


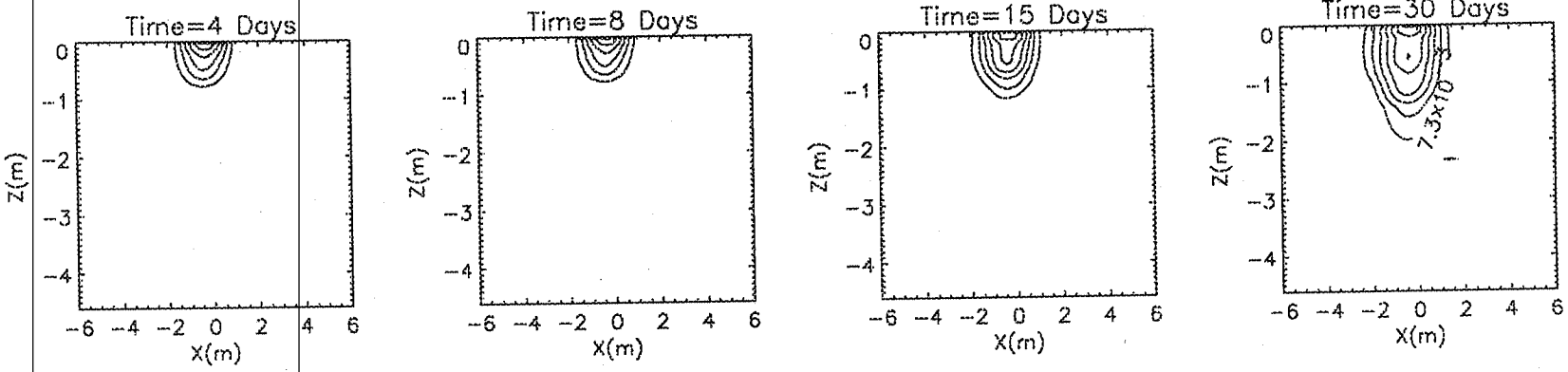
Realization 17



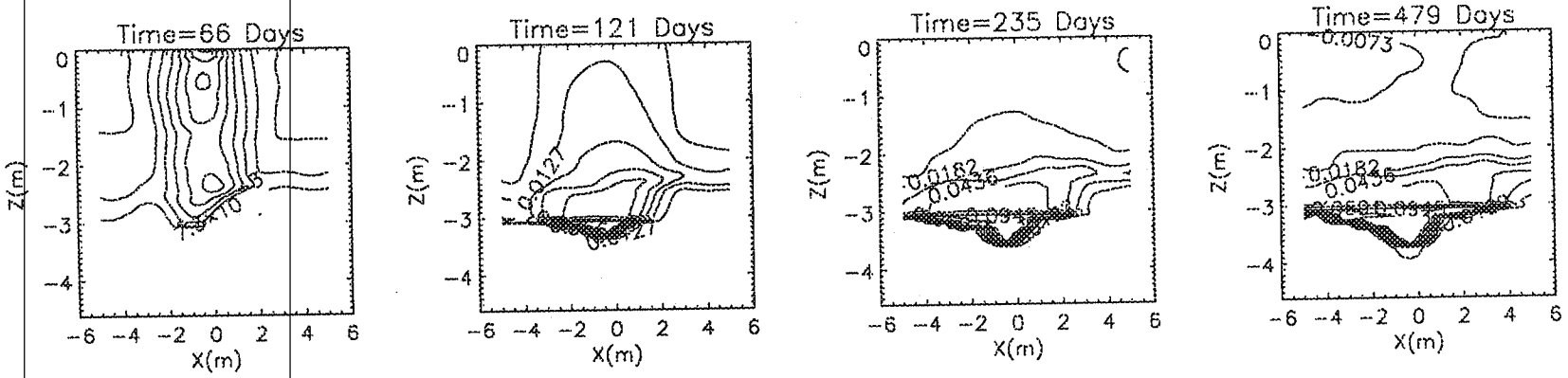


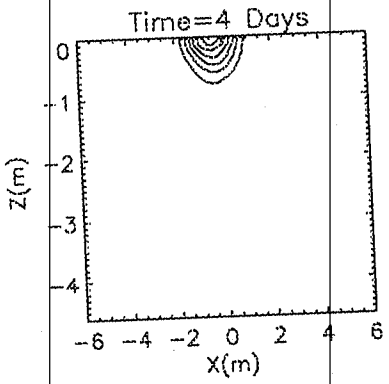


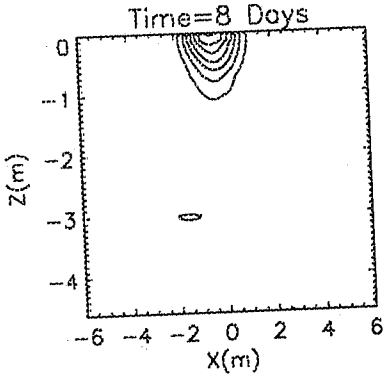


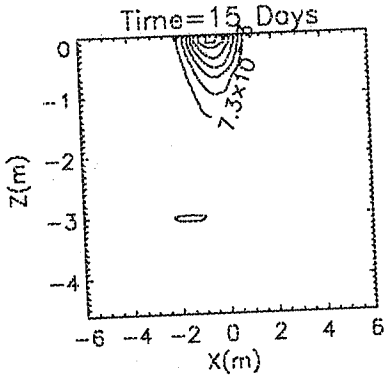


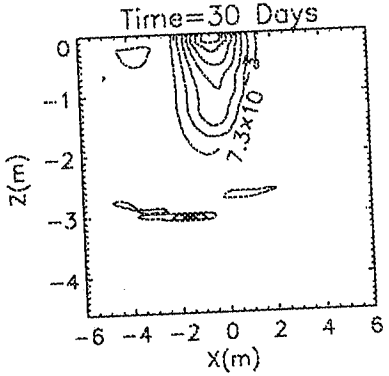
Realization 18



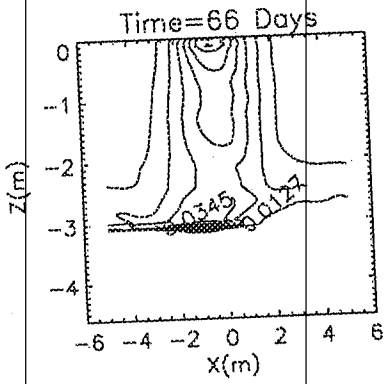


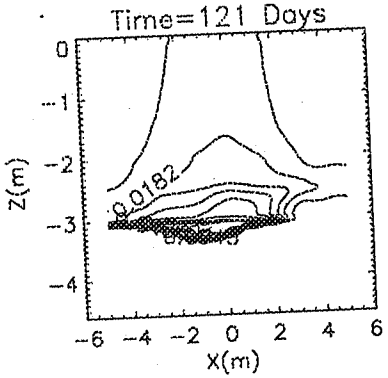


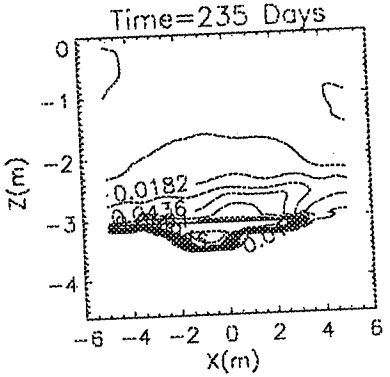


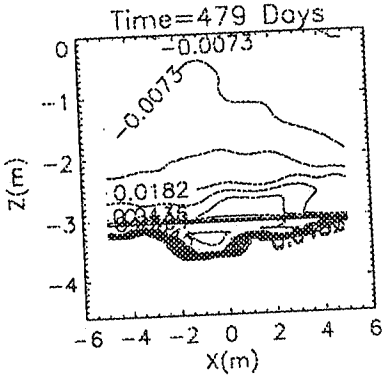


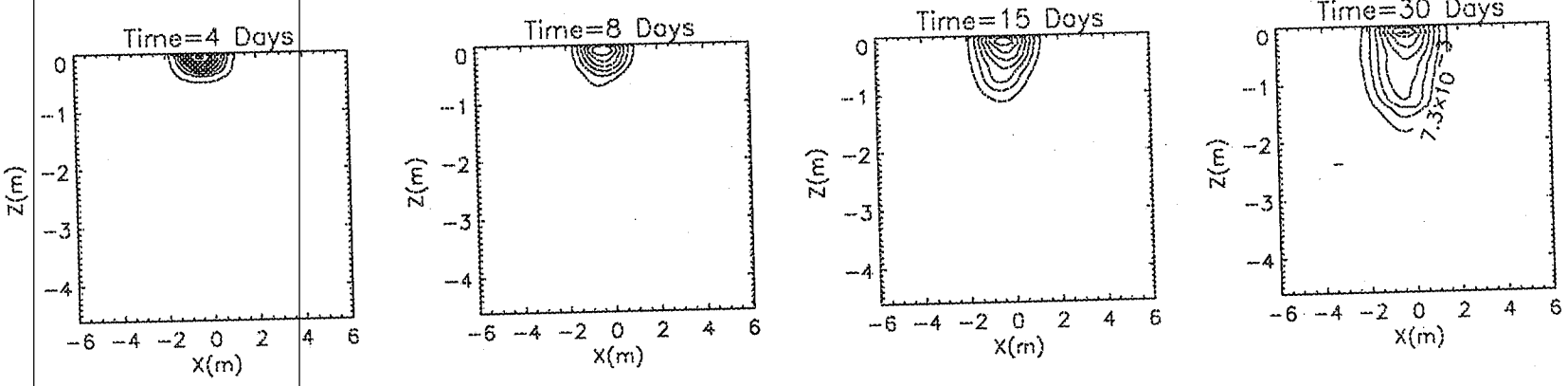
Realization 19



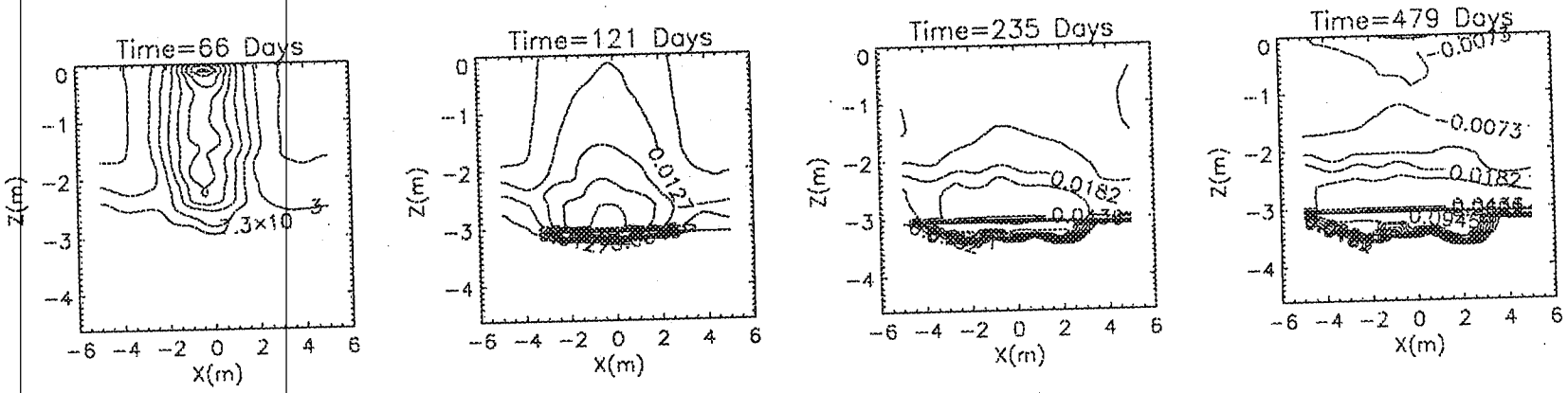






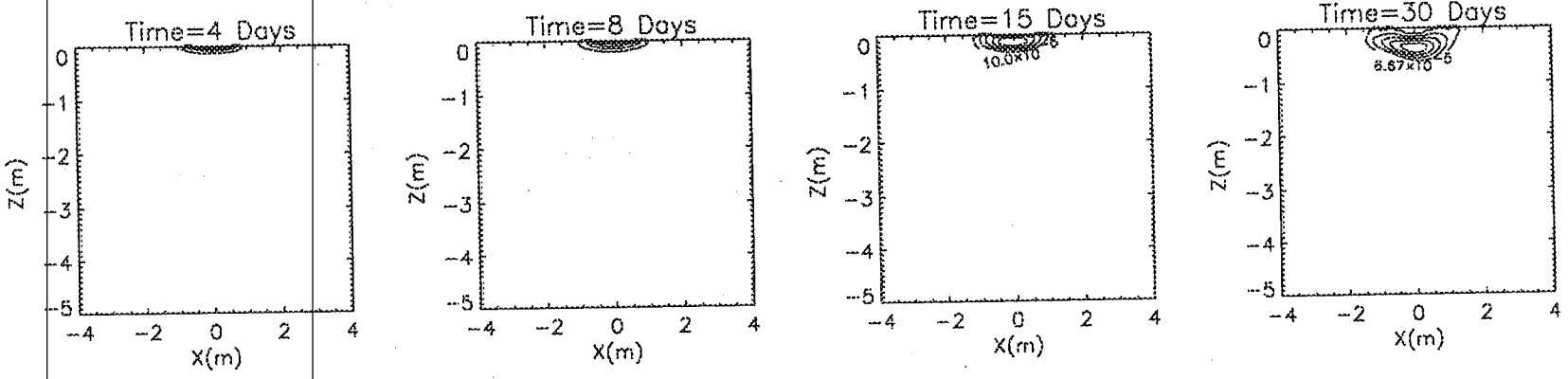


Realization 20

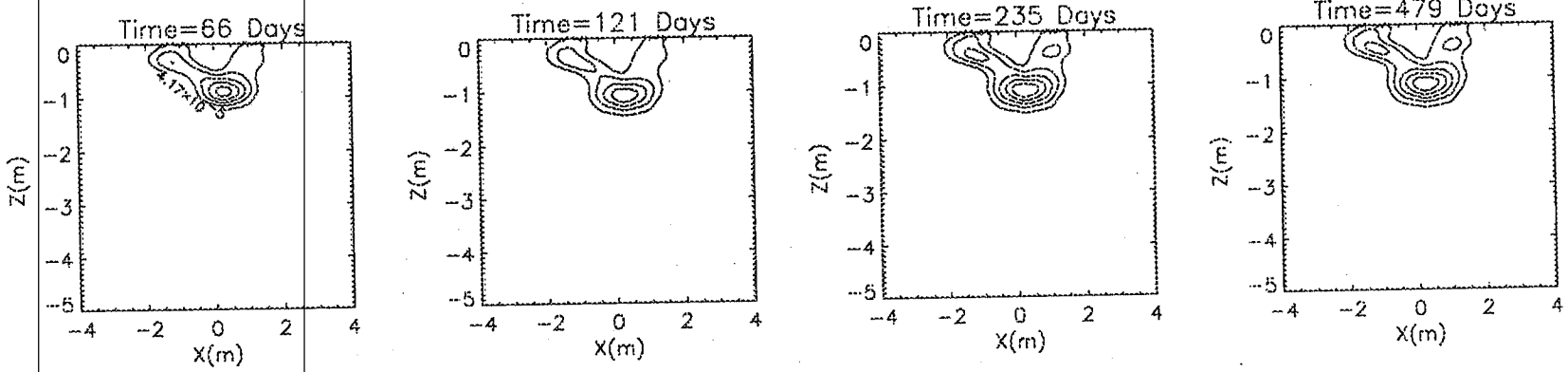


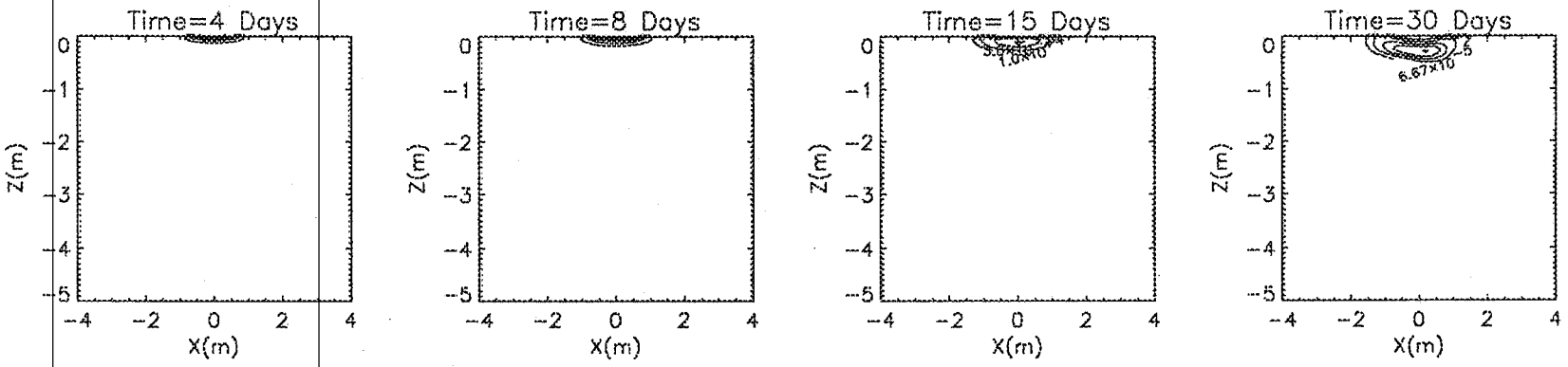
Appendix D: Simulated Water Content Change and Bromide Concentration Realizations

Figure D.3.
Simulated Bromide Concentration Realizations For Single-Scale Permeability Model.

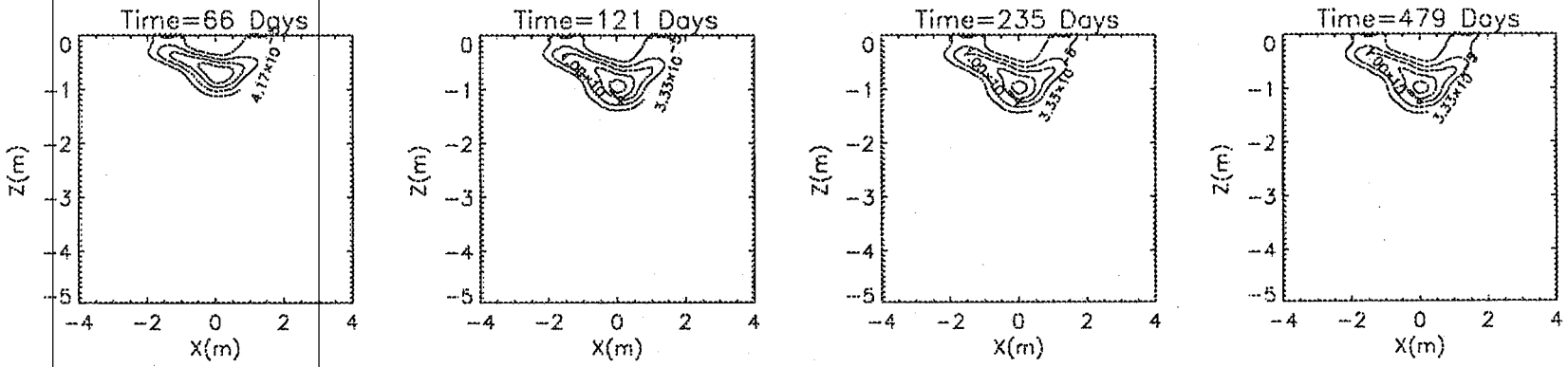


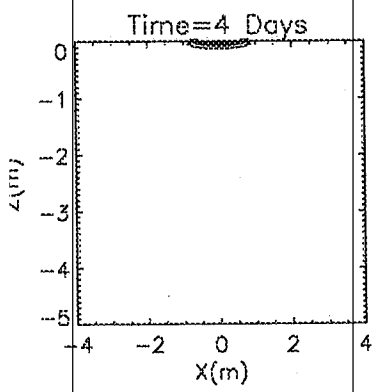
Realization 1

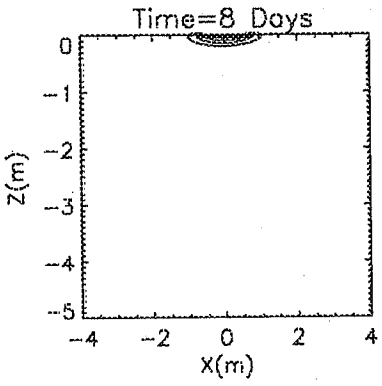


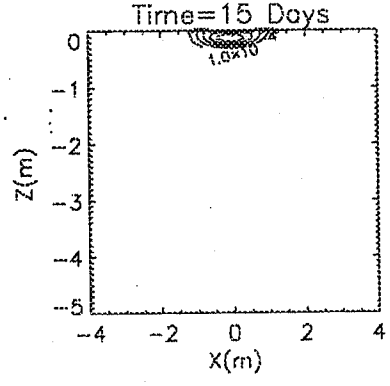


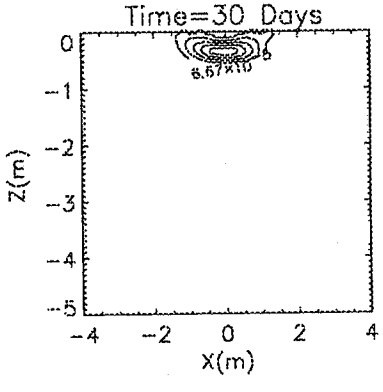
Realization 2



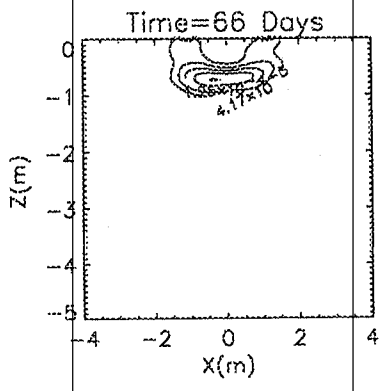


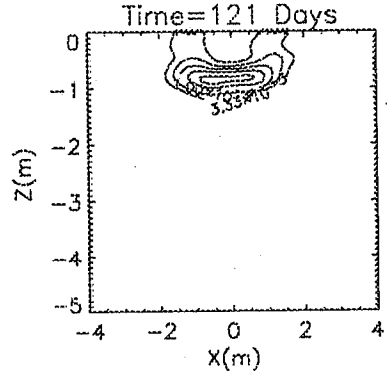


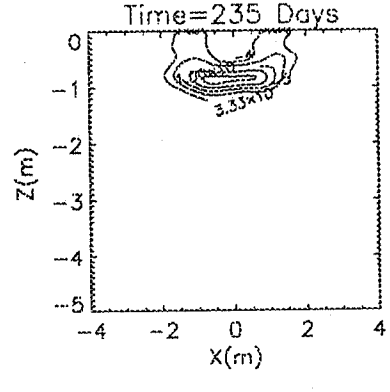


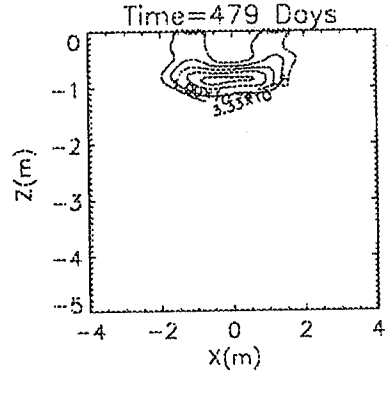


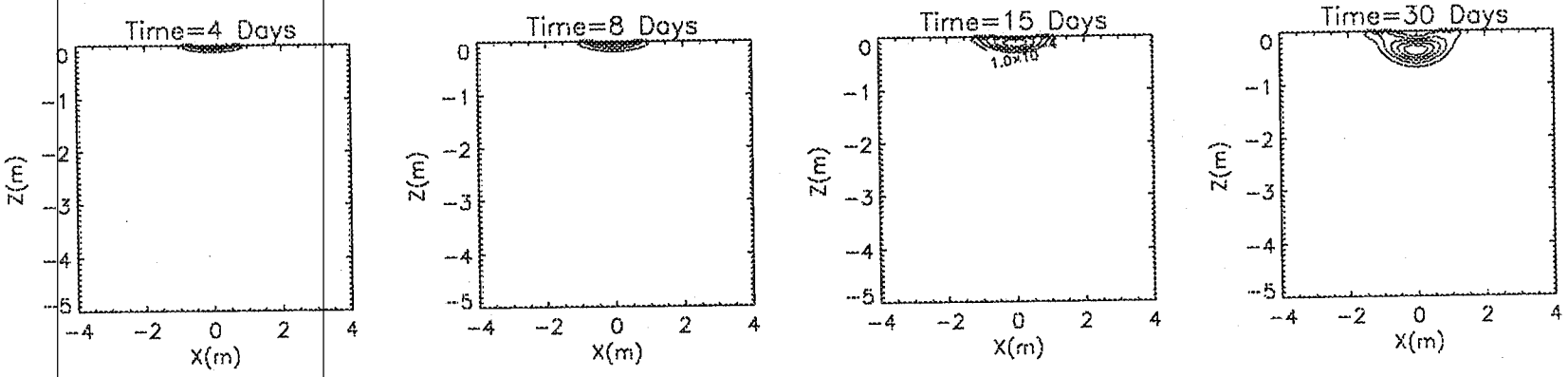
Realization 3



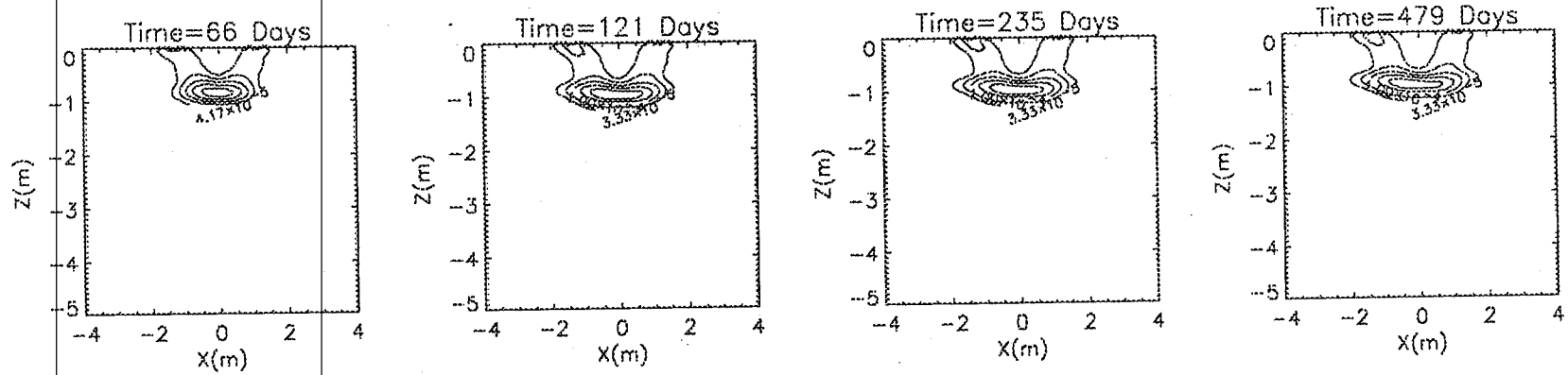


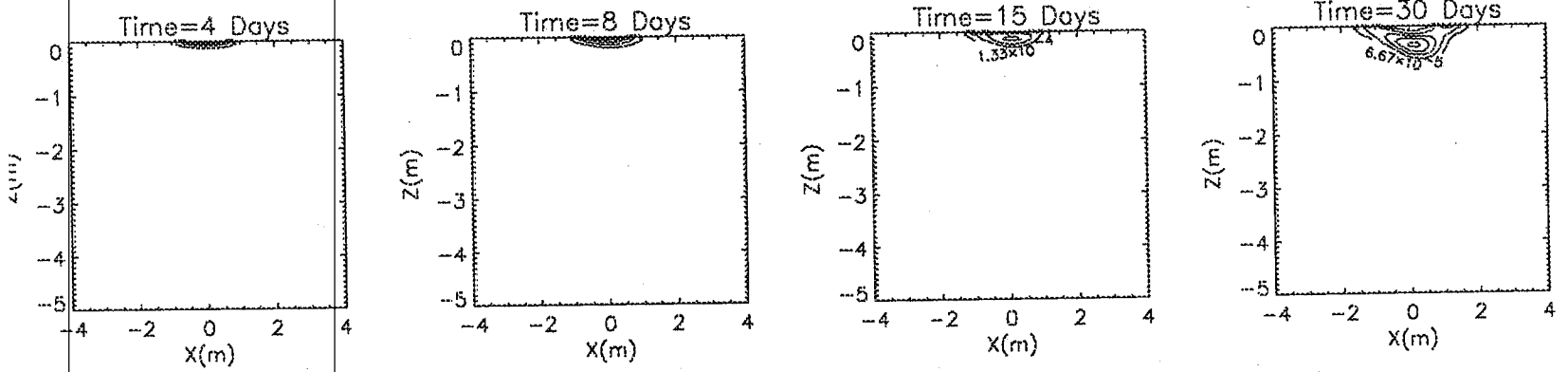




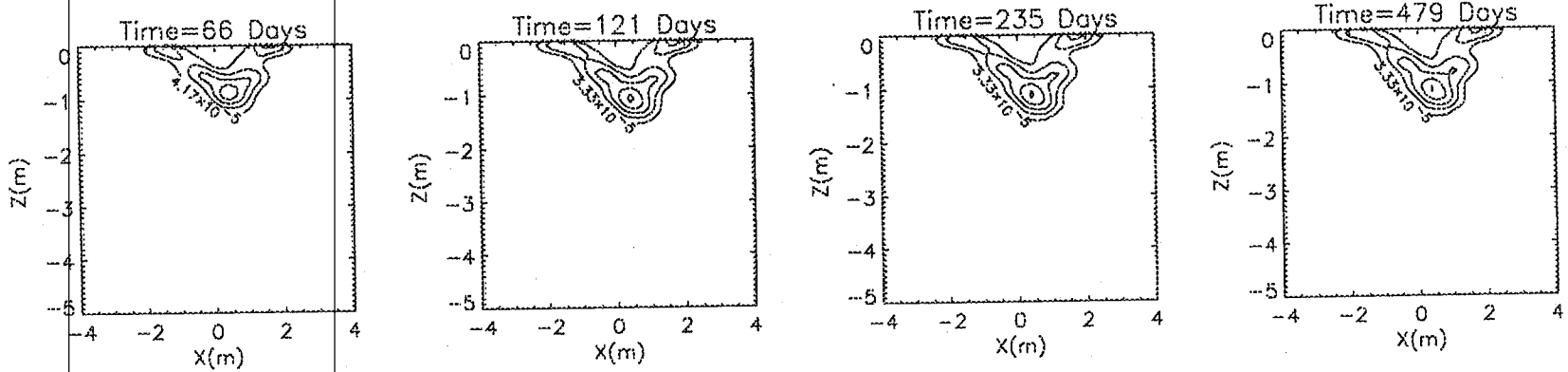


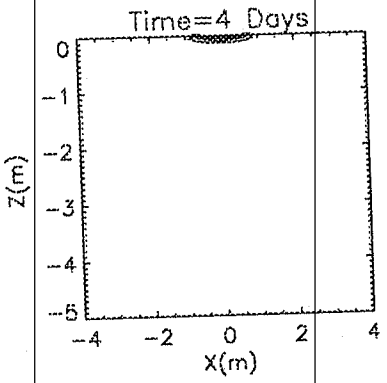
Realization 4

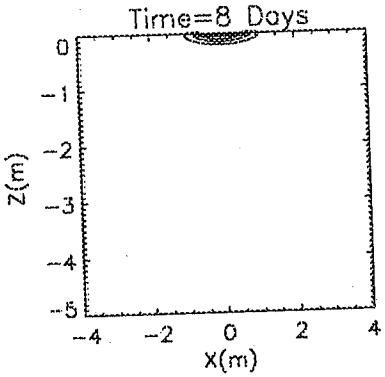


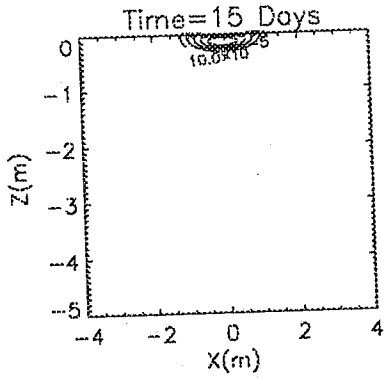


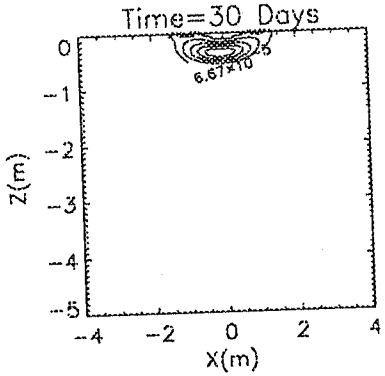
Realization 5



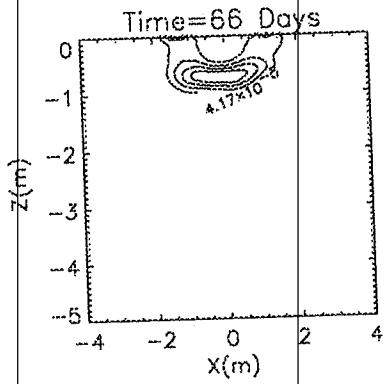


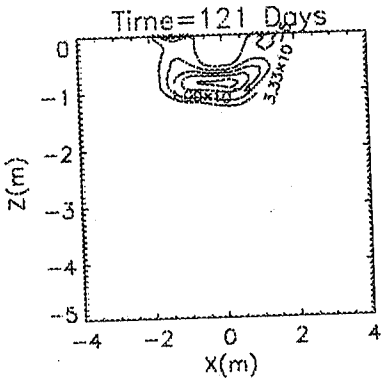


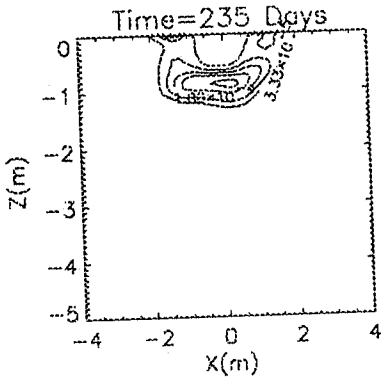


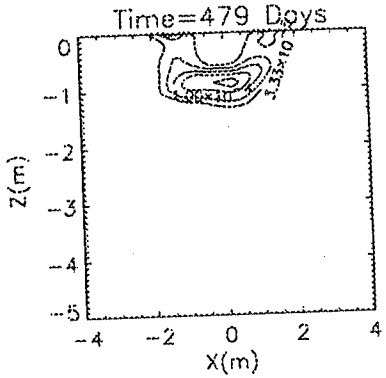


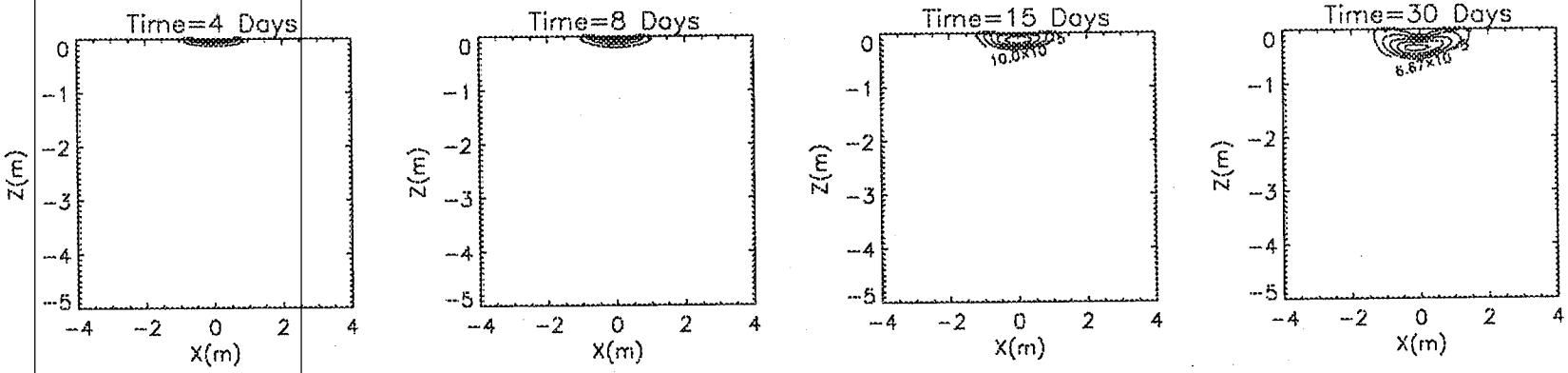
Realization 6



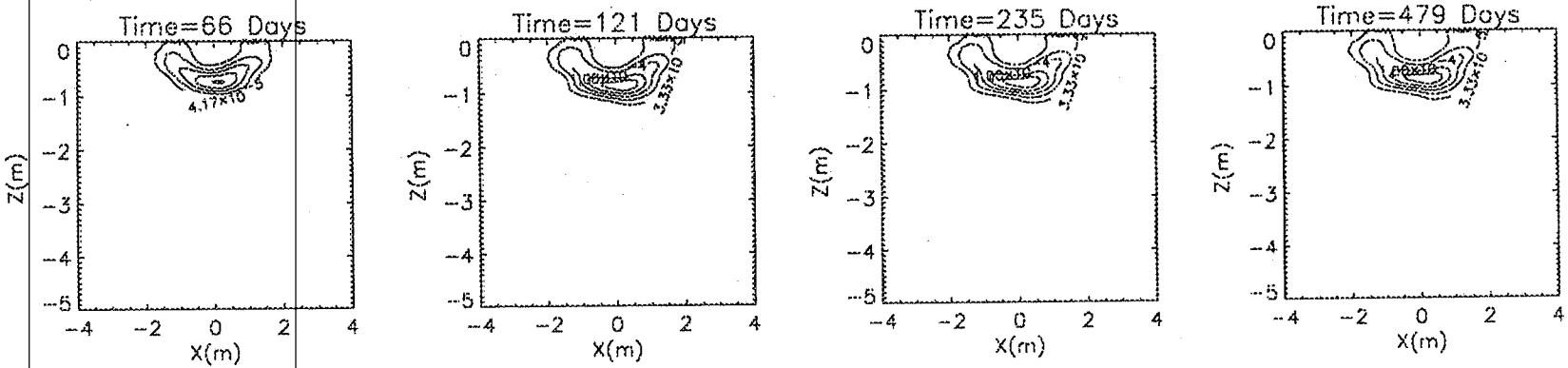


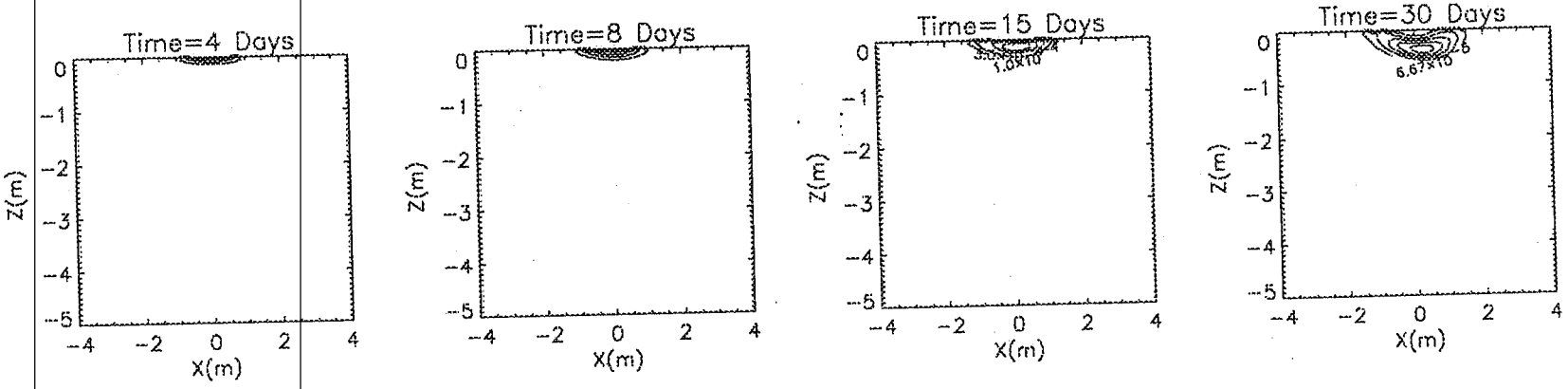




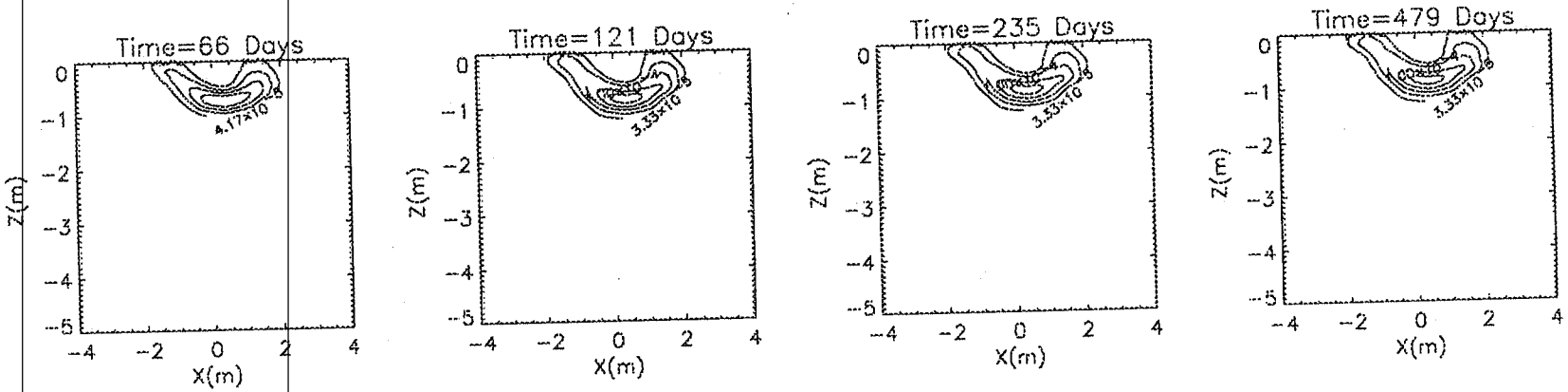


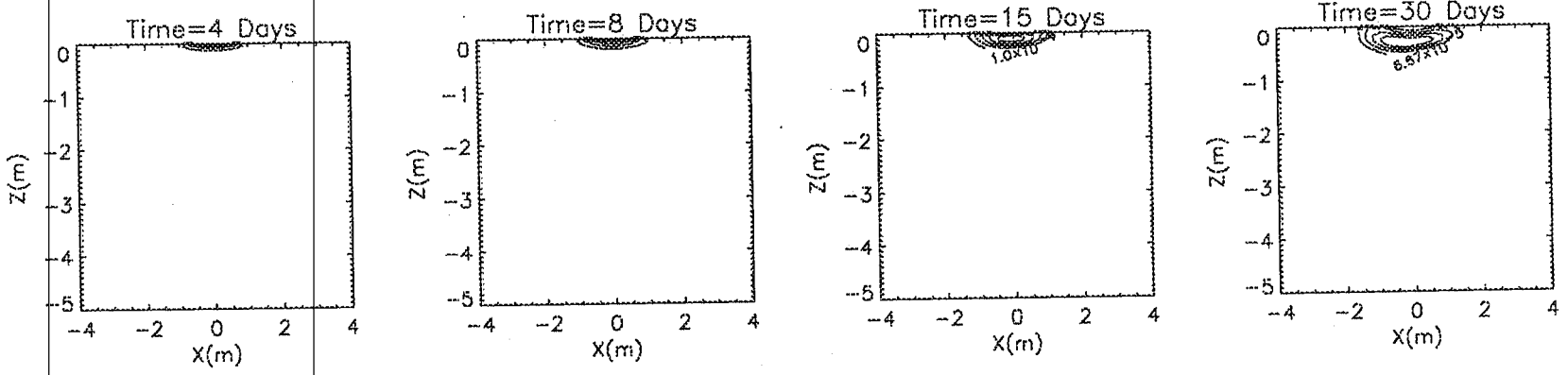
Realization 7



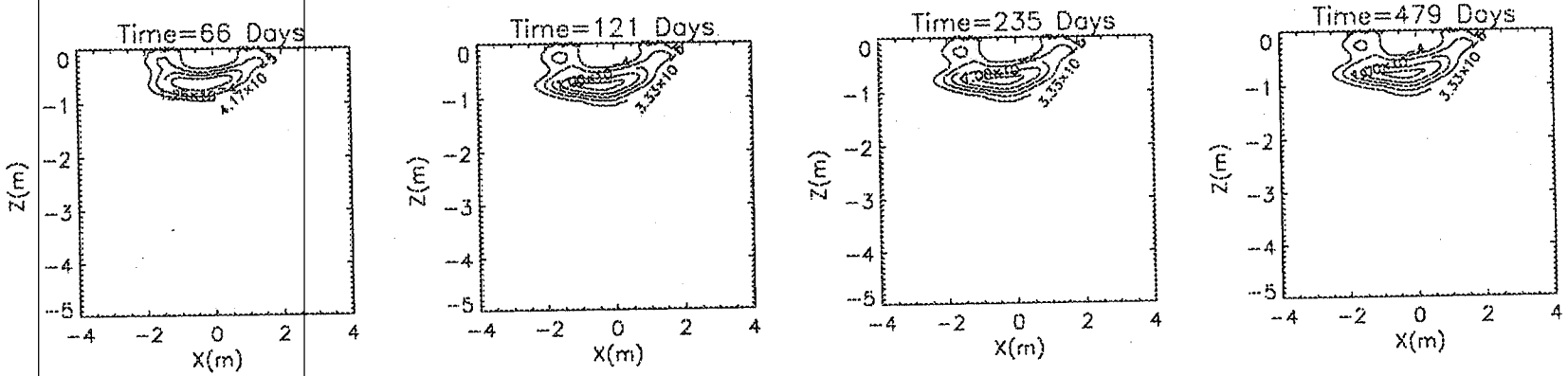


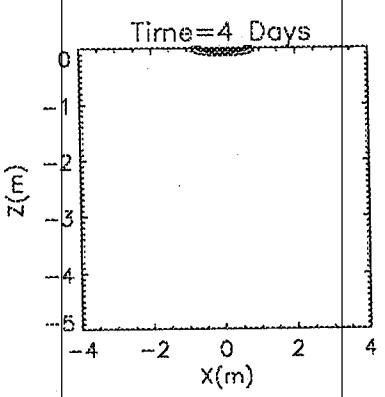
Realization 8

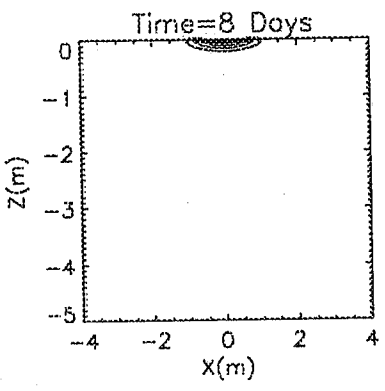


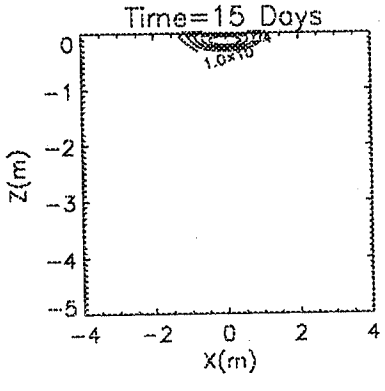


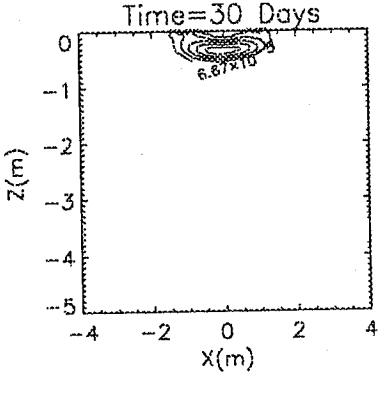
Realization 9



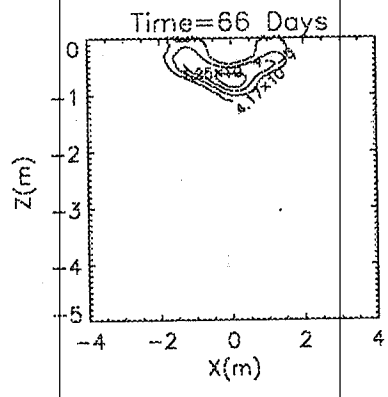


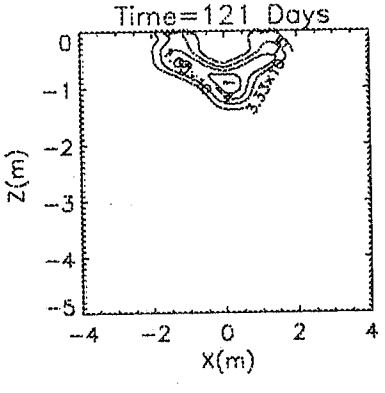


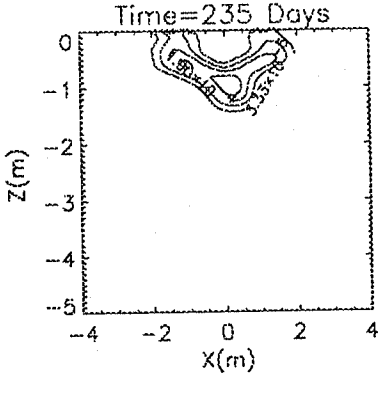


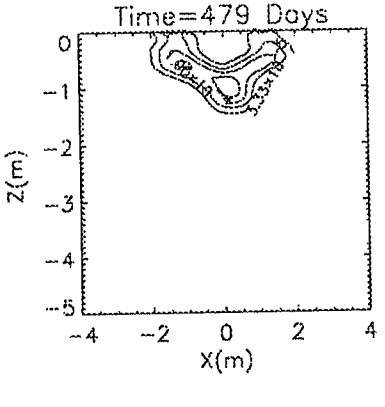


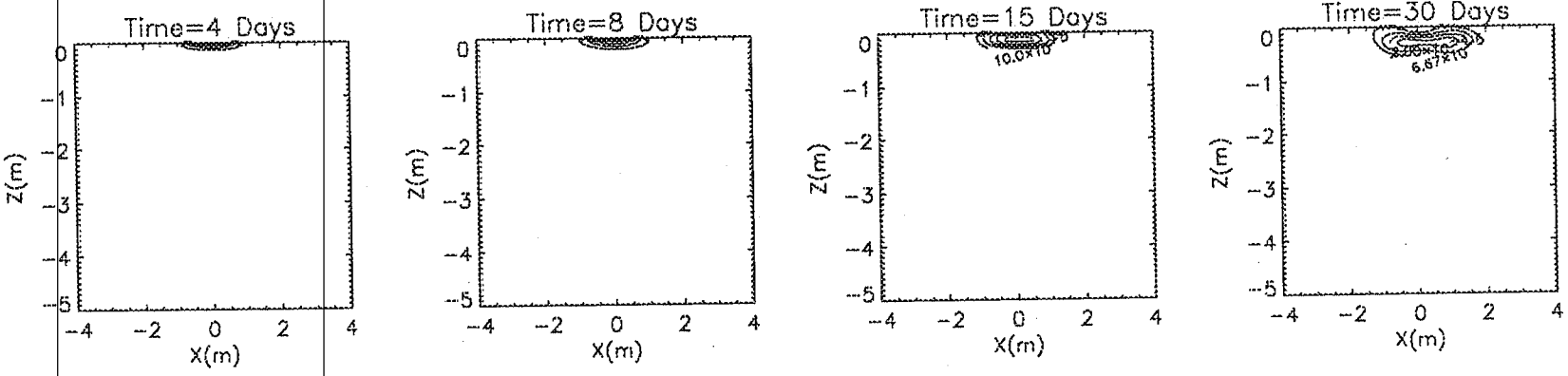
Realization 10



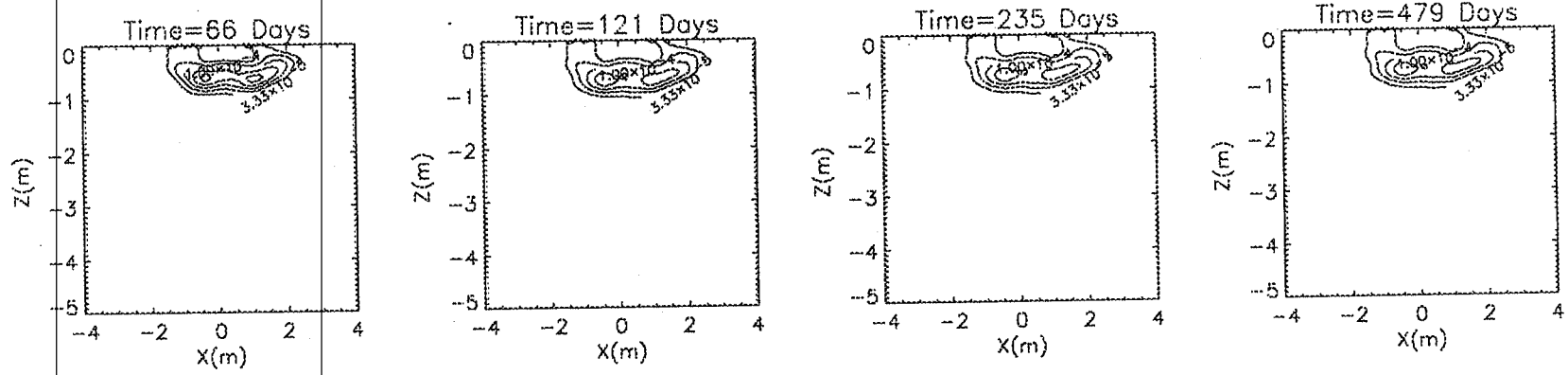


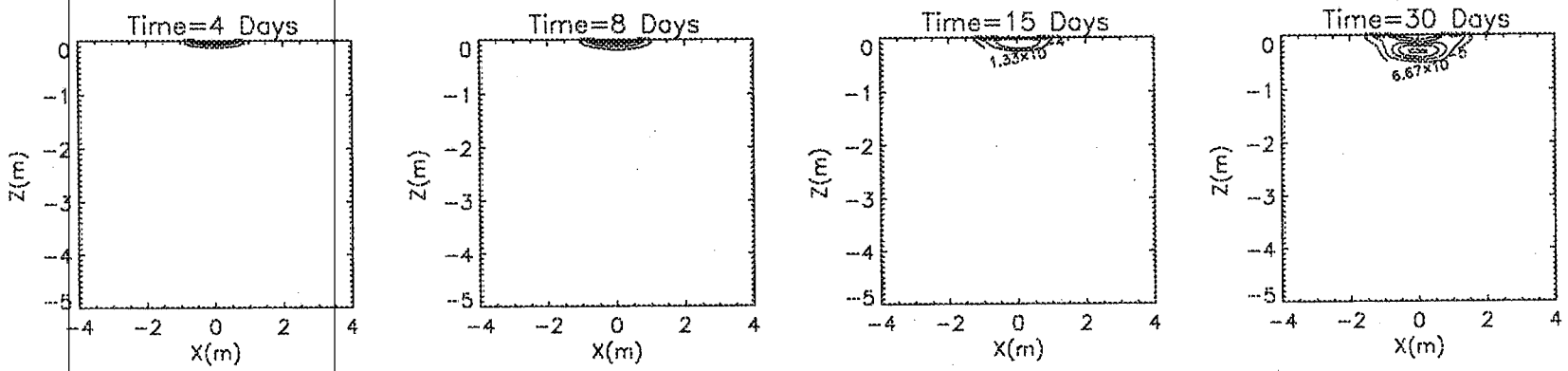




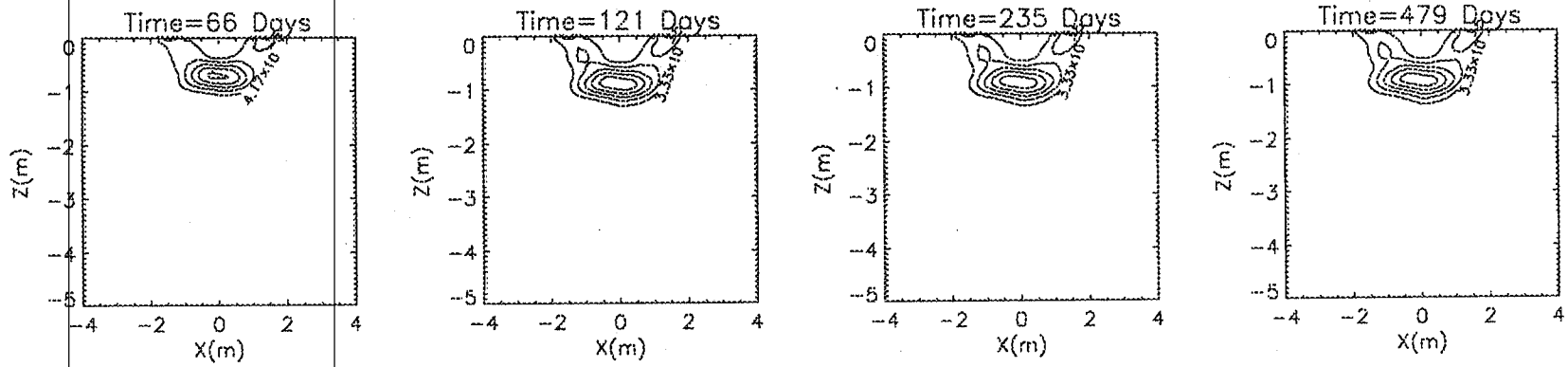


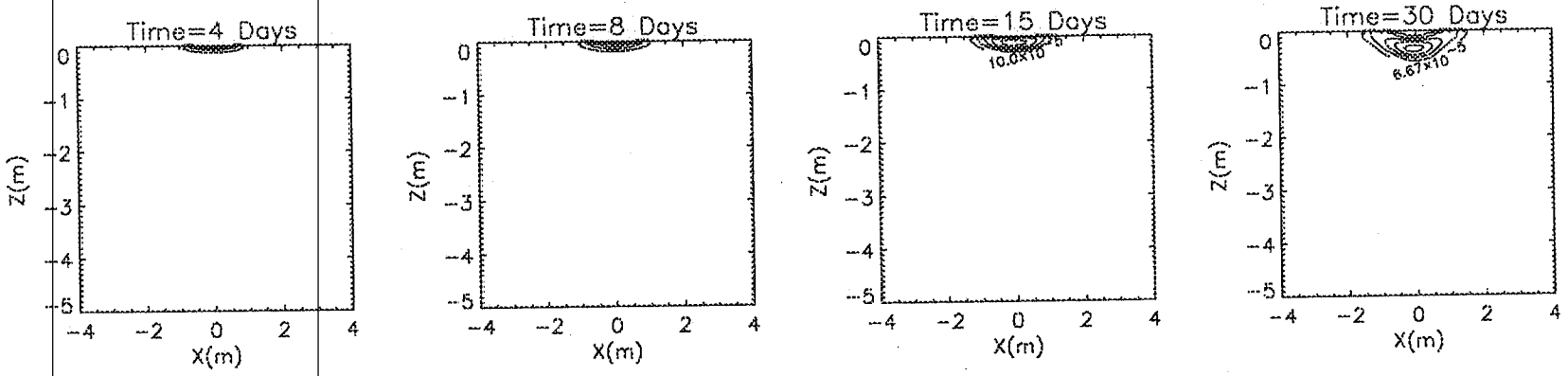
Realization 11



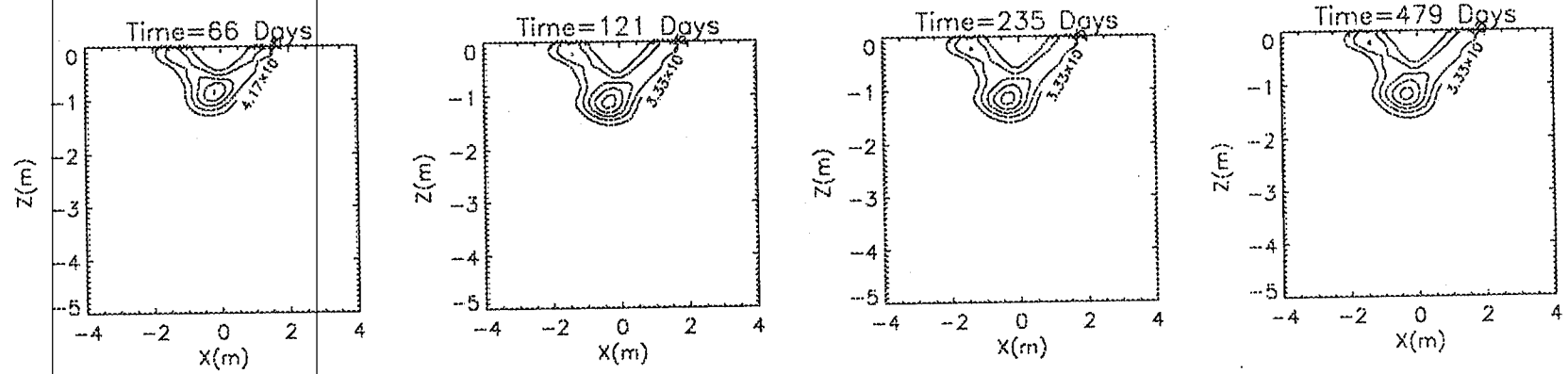


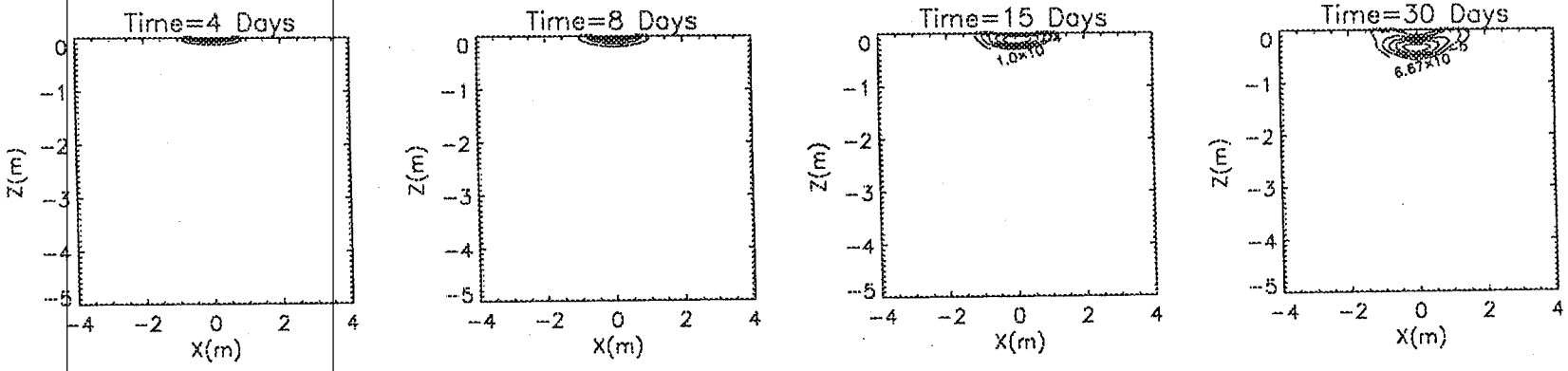
Realization 12



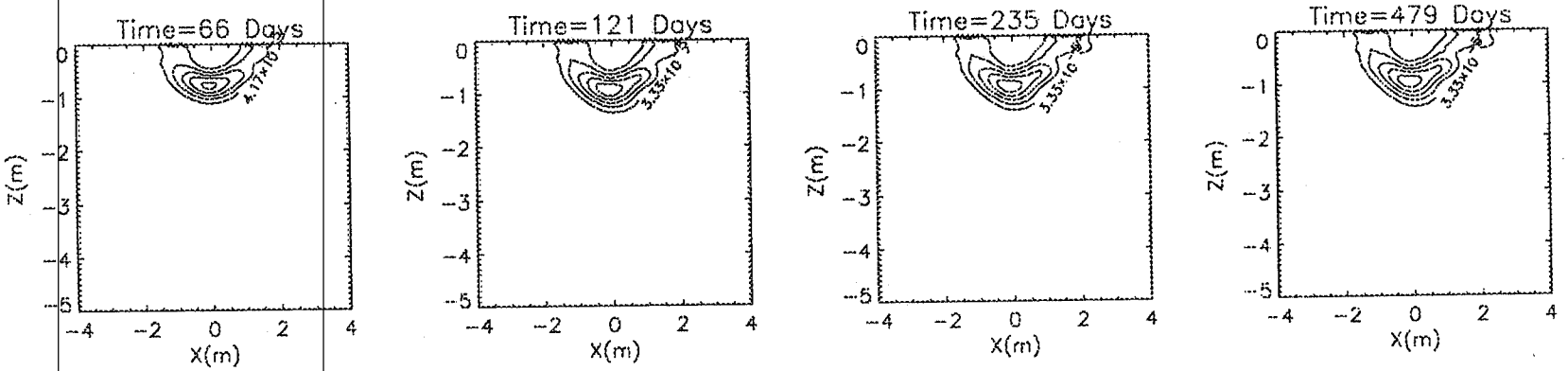


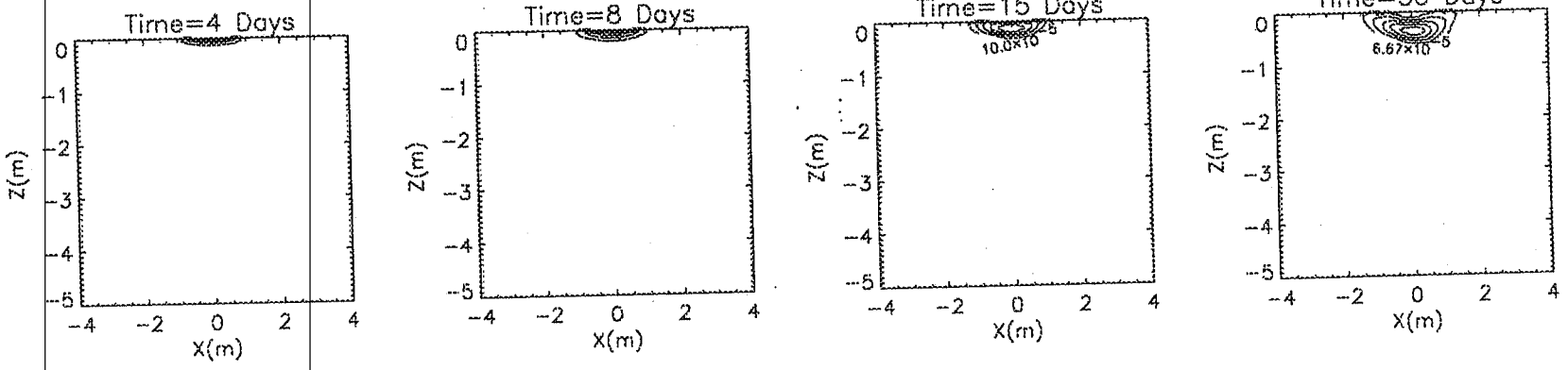
Realization 13



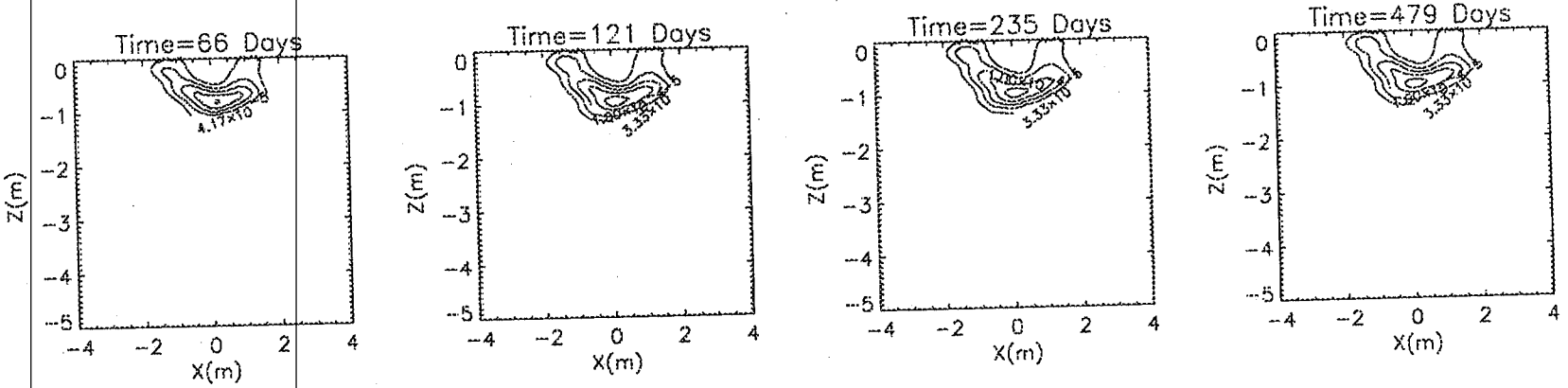


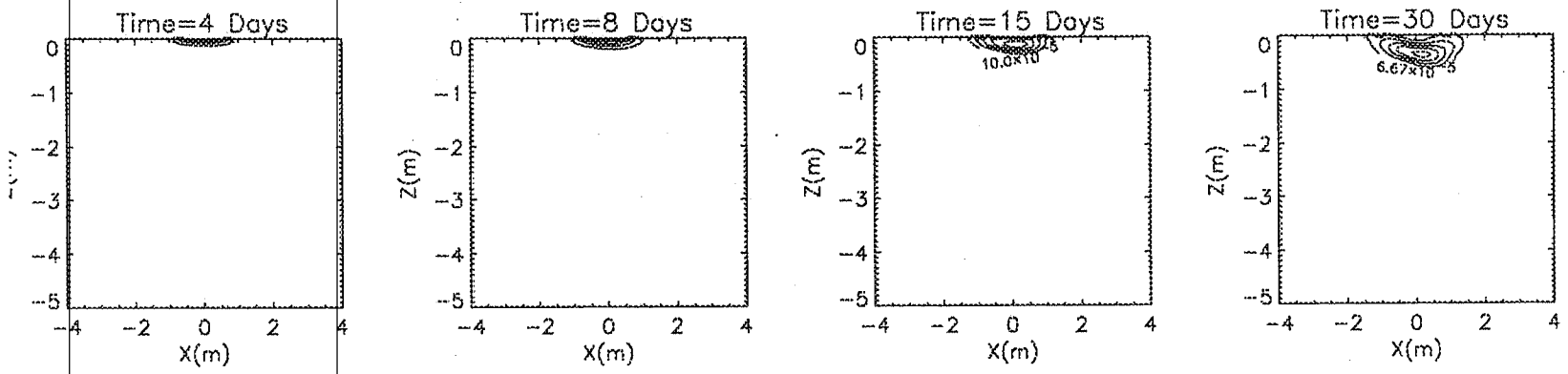
Realization 14



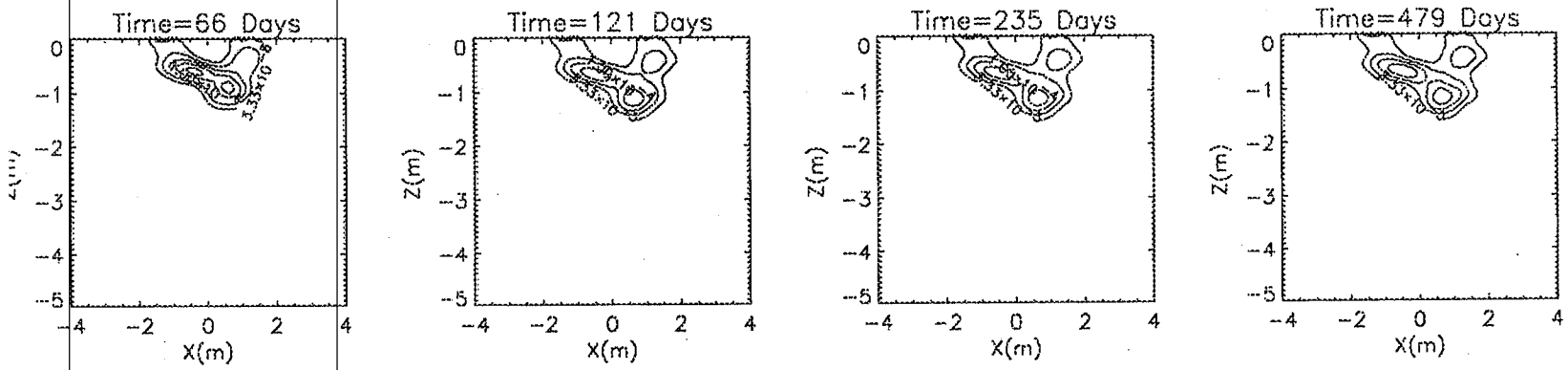


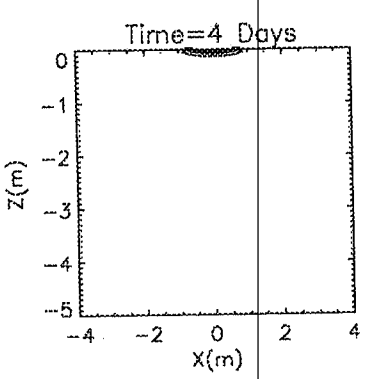
Realization 15

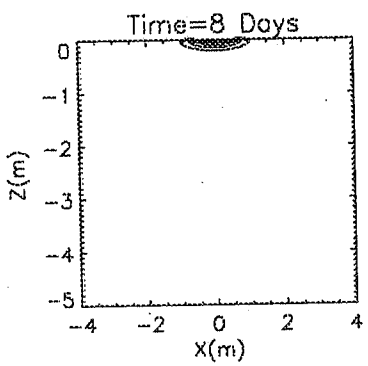


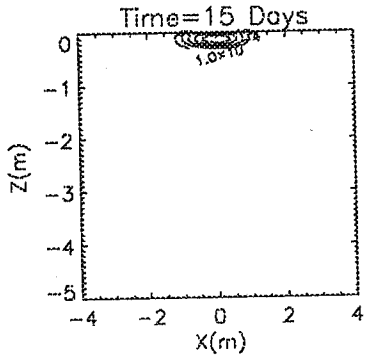


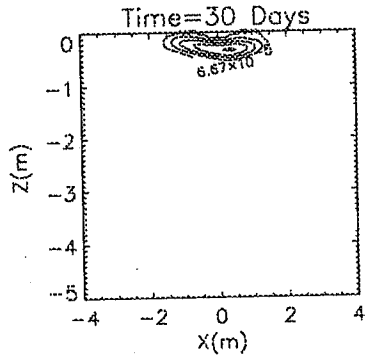
Realization 16



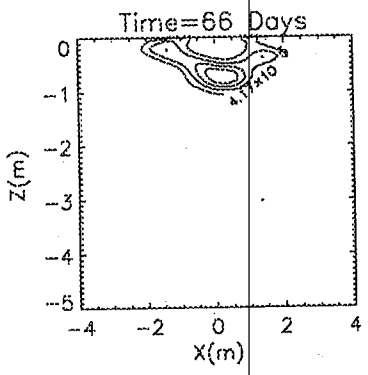


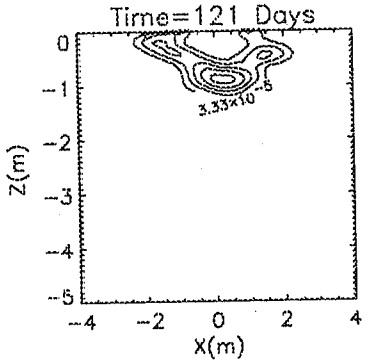


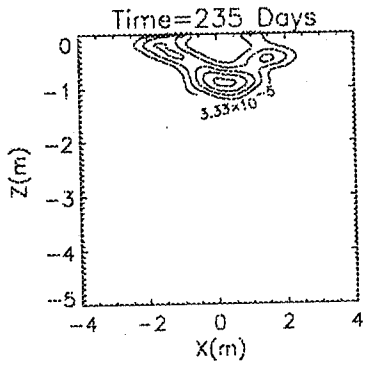


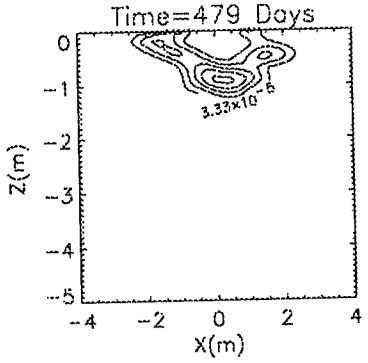


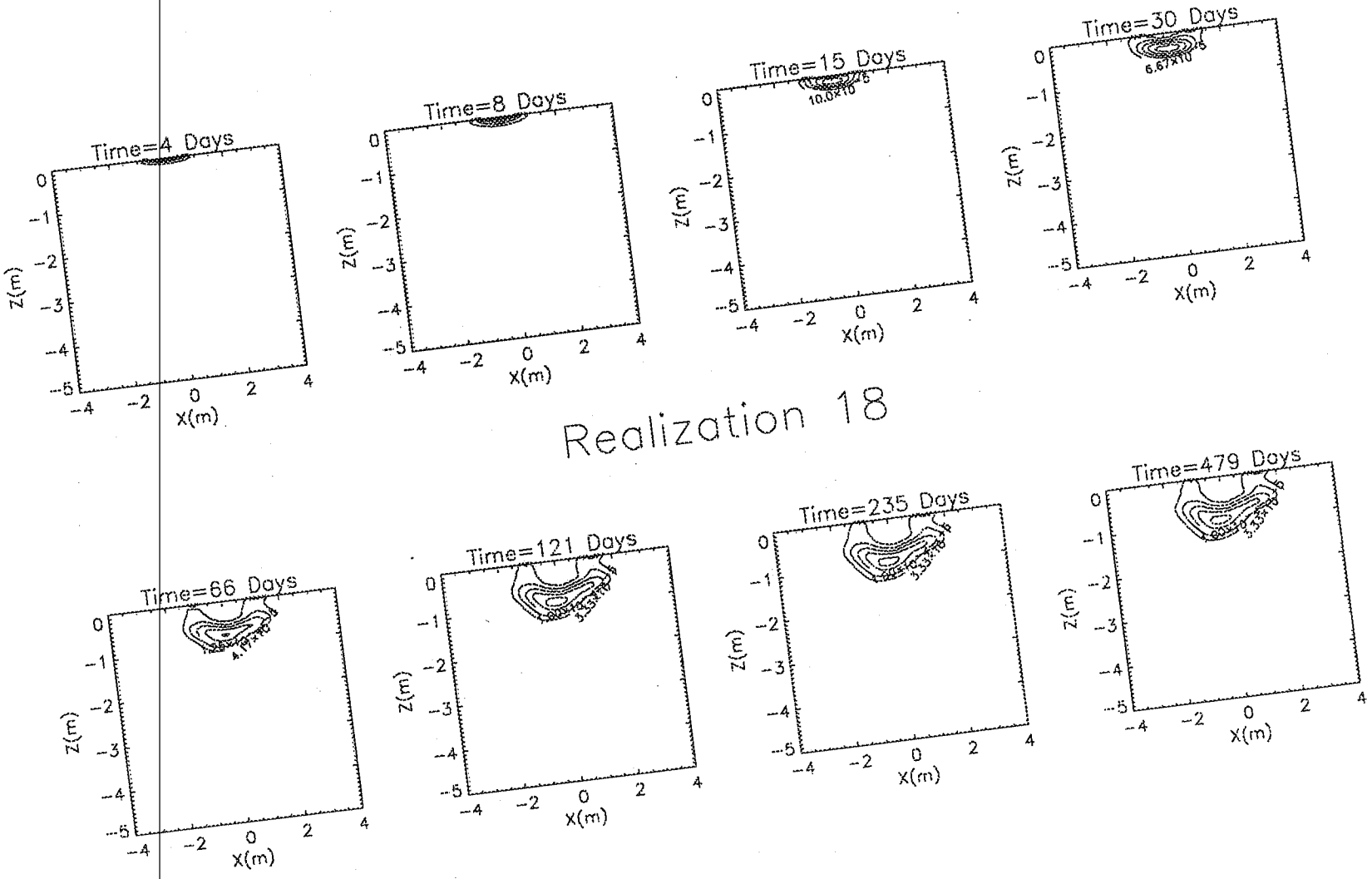
Realization 17

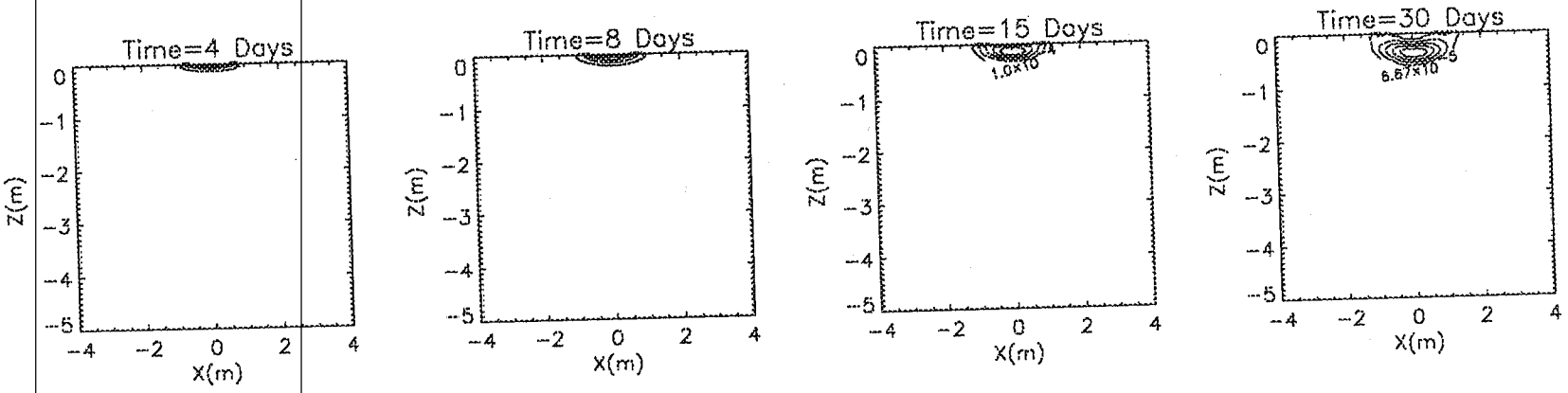




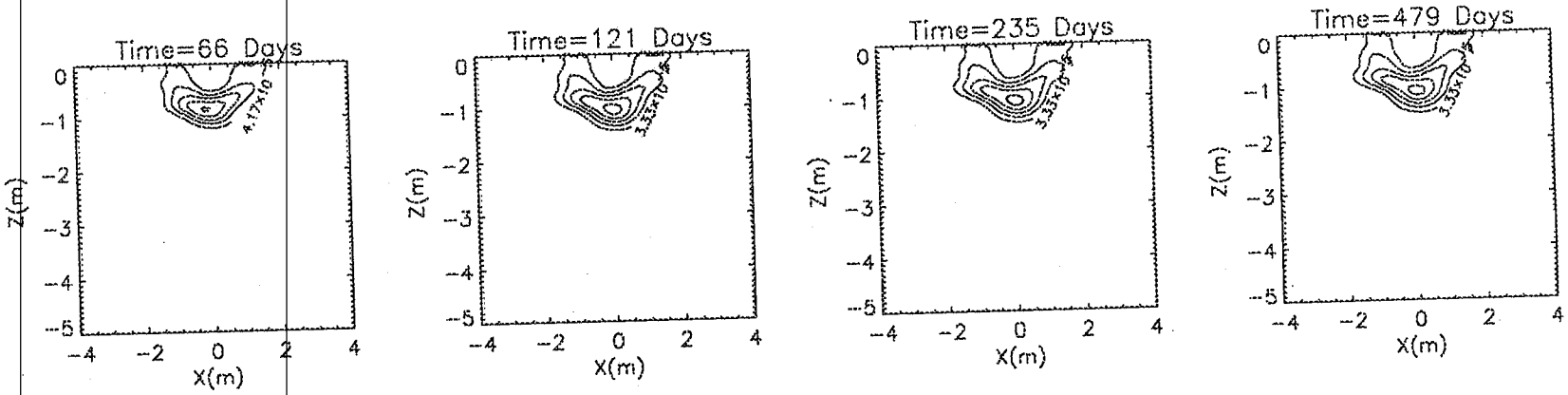


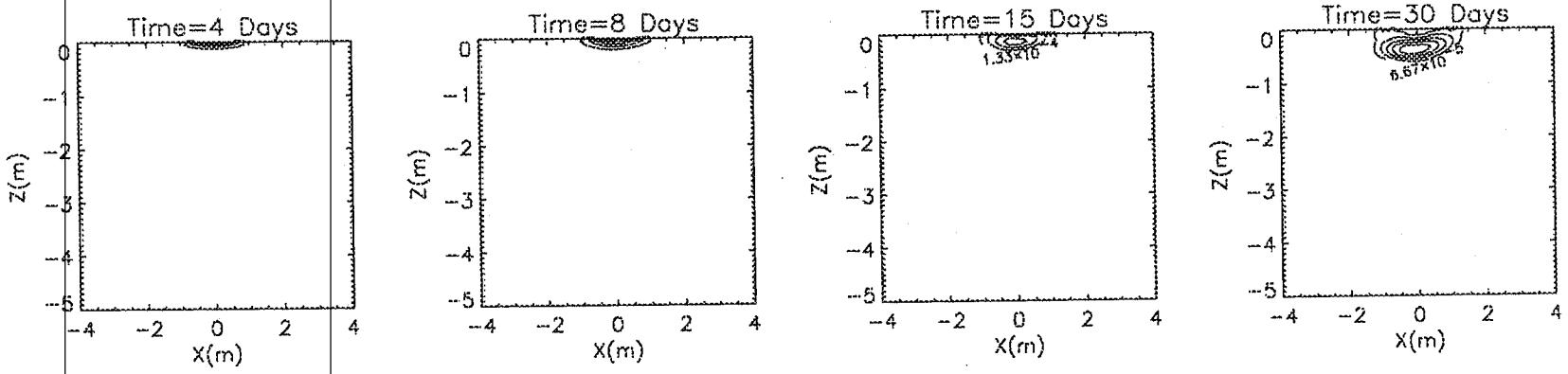



Time=4 Days
Time=8 Days
Time=15 Days
Time=30 Days
Realization 18
Time=66 Days
Time=121 Days
Time=235 Days
Time=479 Days
Z(m)
X(m)



Realization 19





Realization 20

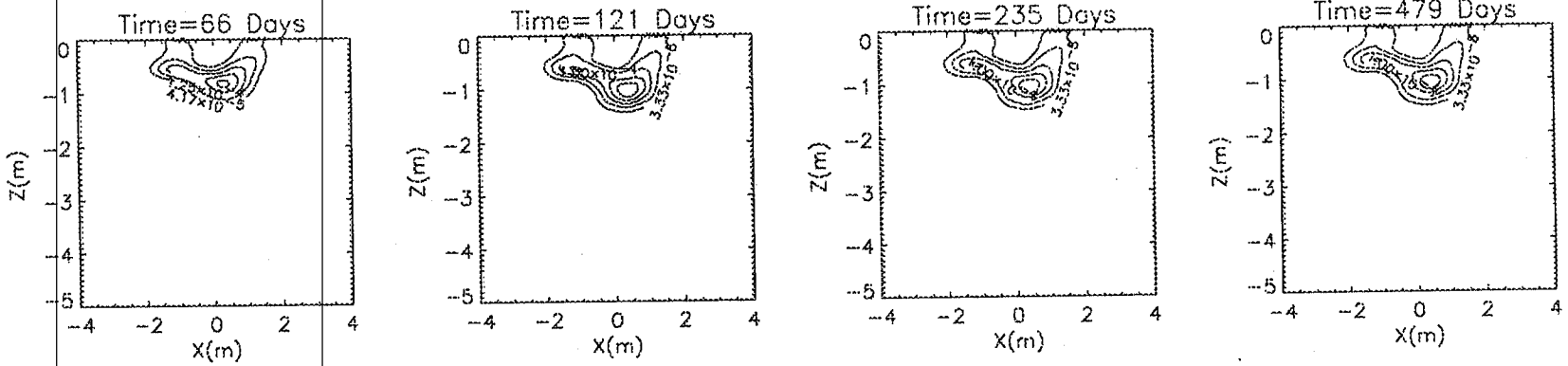
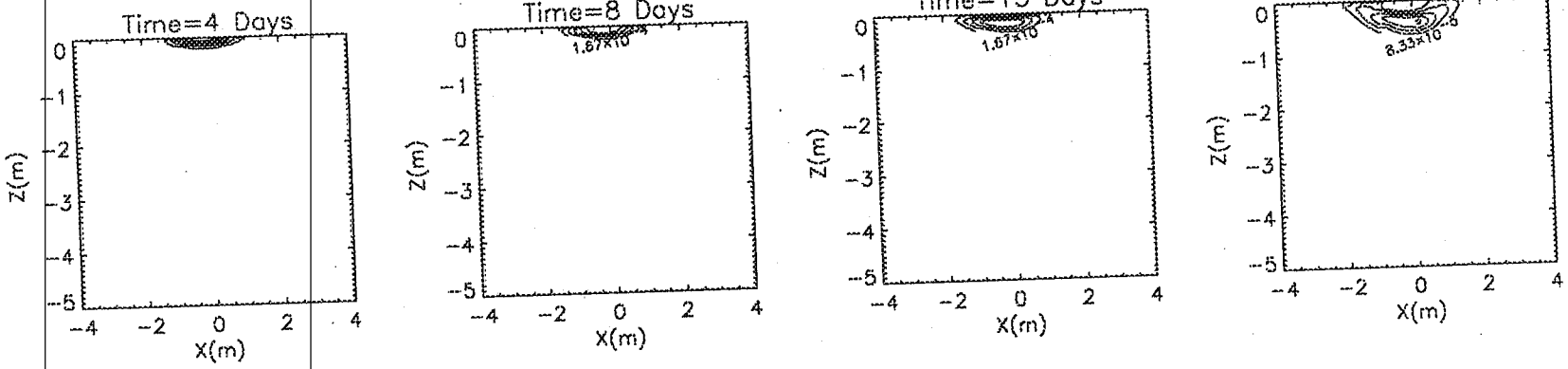
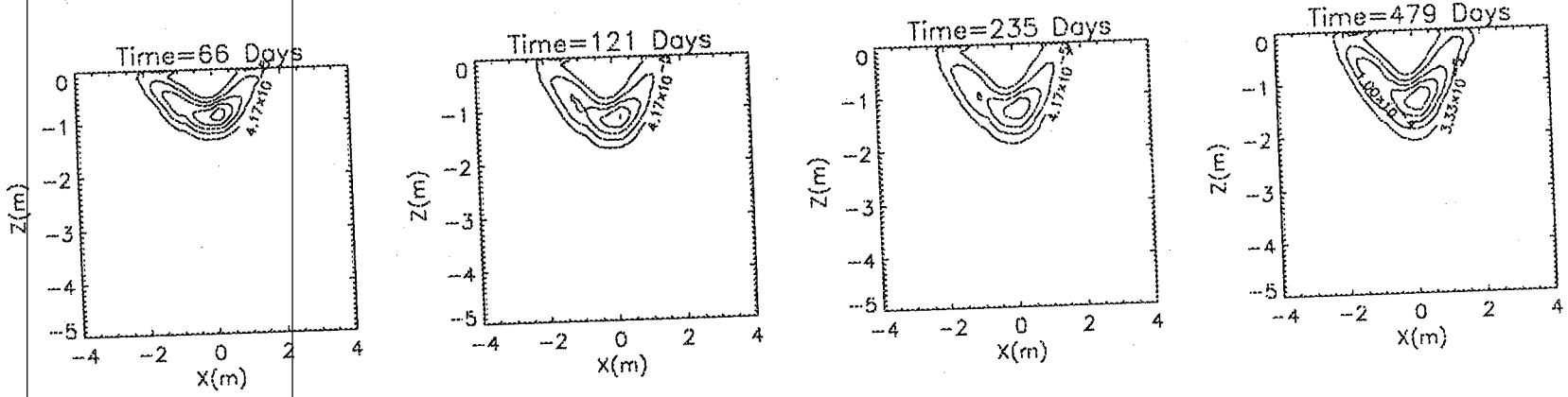
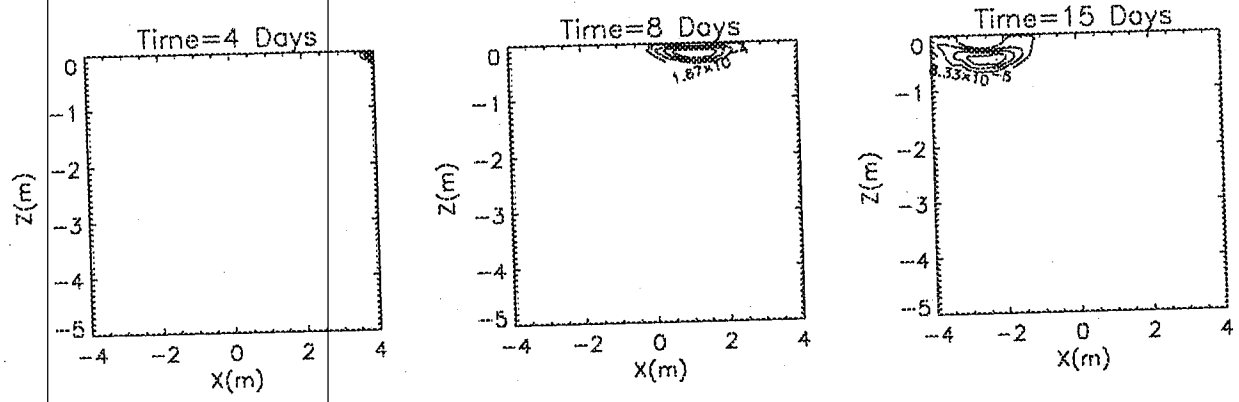


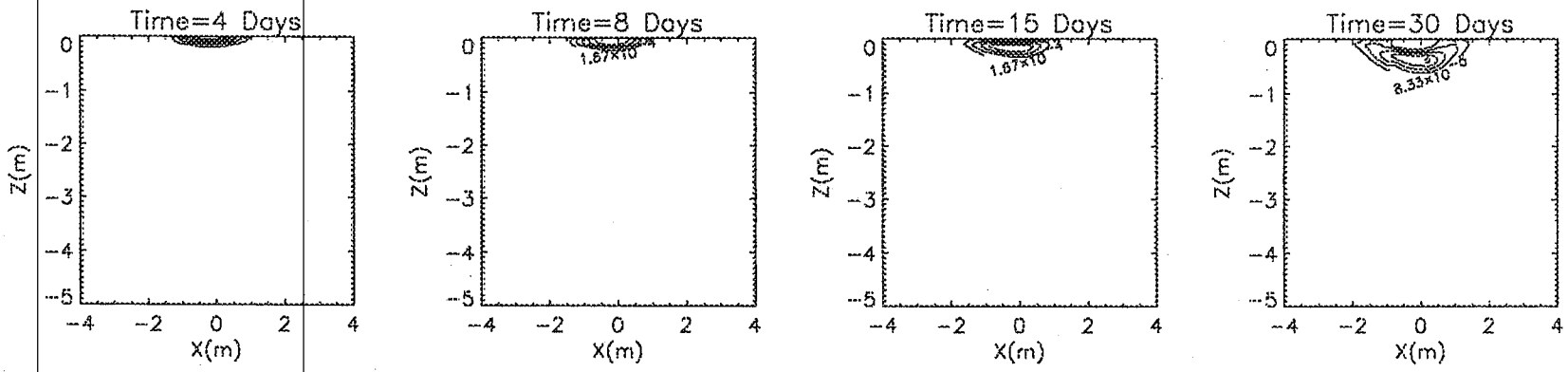
Figure D.4.
Simulated Bromide Concentration Realizations For Two-Scale Permeability Model.



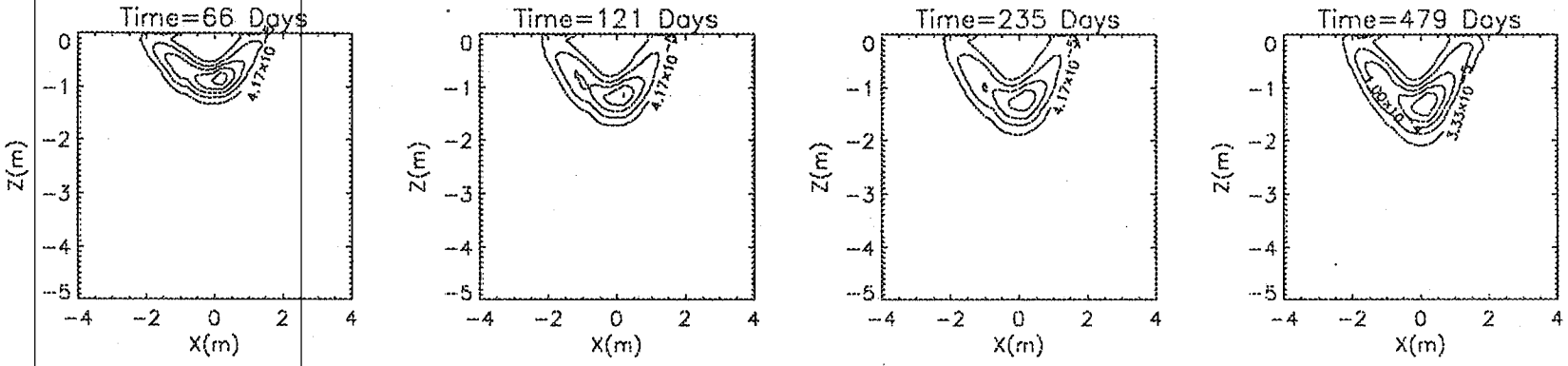
Realization 1

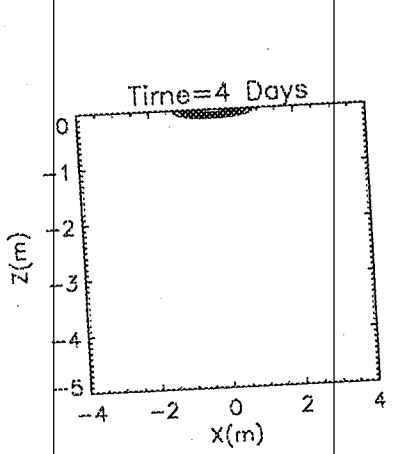


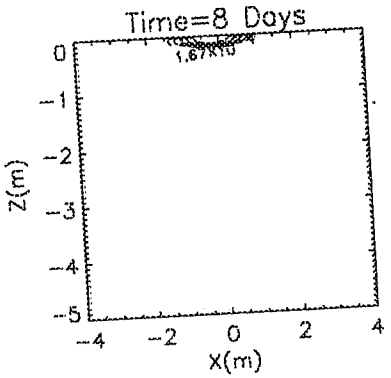

Time=4 Days
Time=8 Days
Time=15 Days
Z(m)
X(m)
0
−1
−2
−3
−4
−5
−4
−2
0
2
4
1.67×10⁻⁵
9.33×10⁻⁶

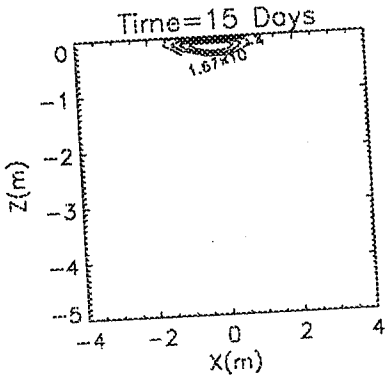


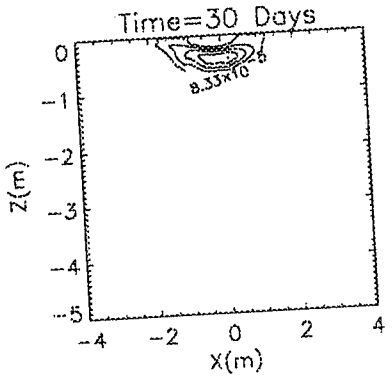
Realization 1



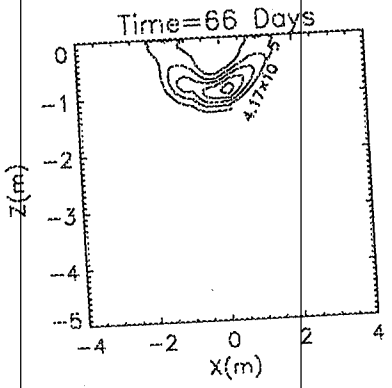


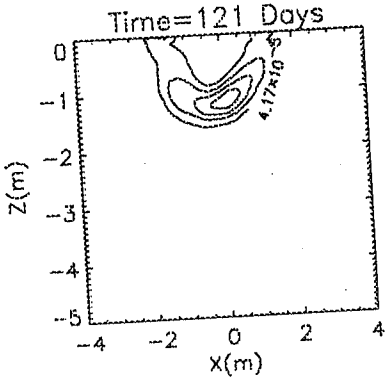


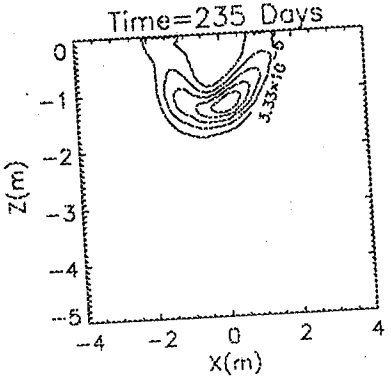


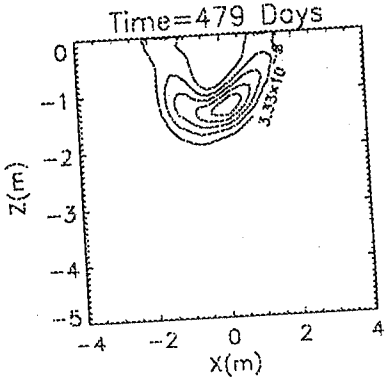


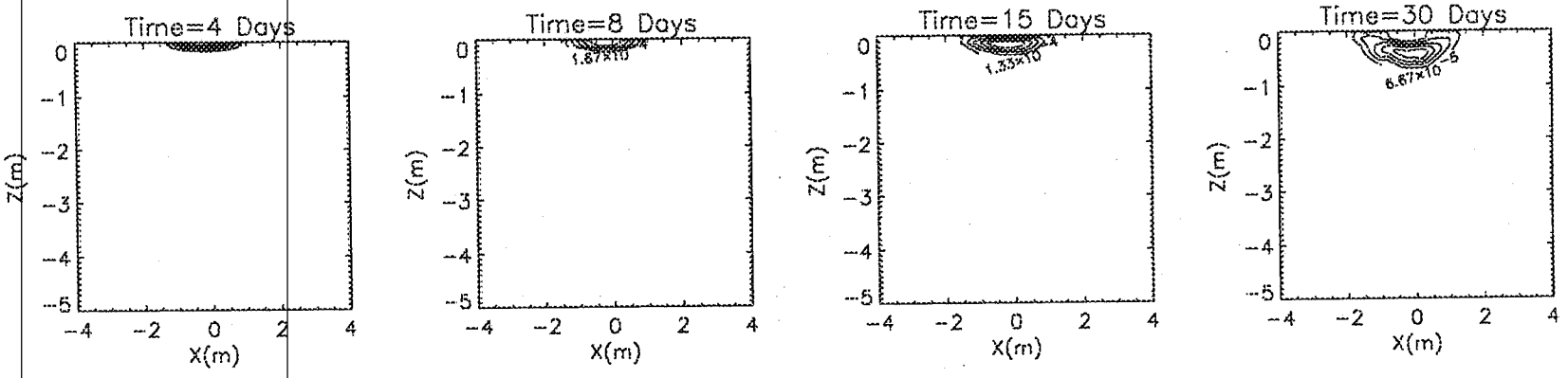
Realization 2



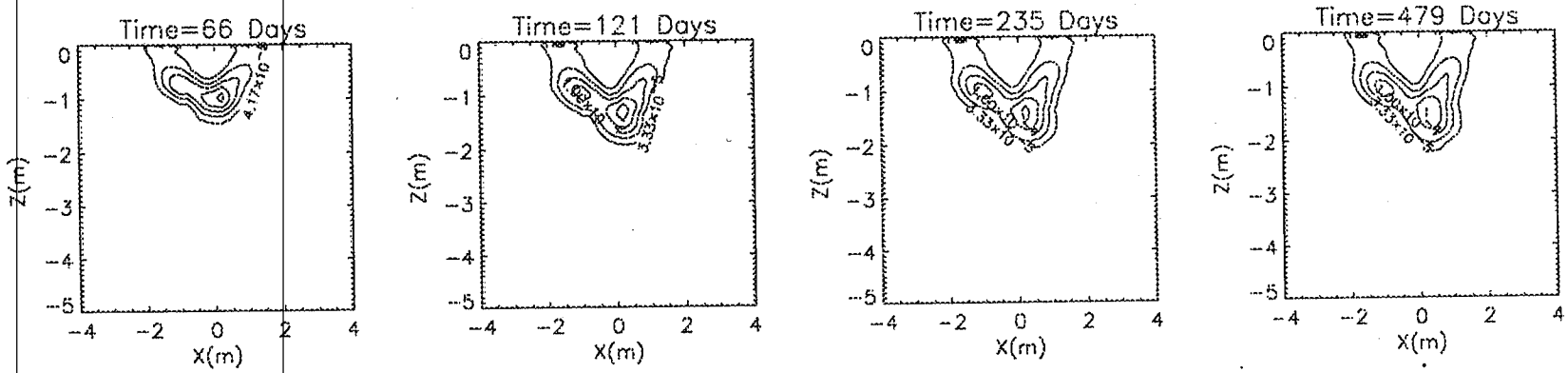


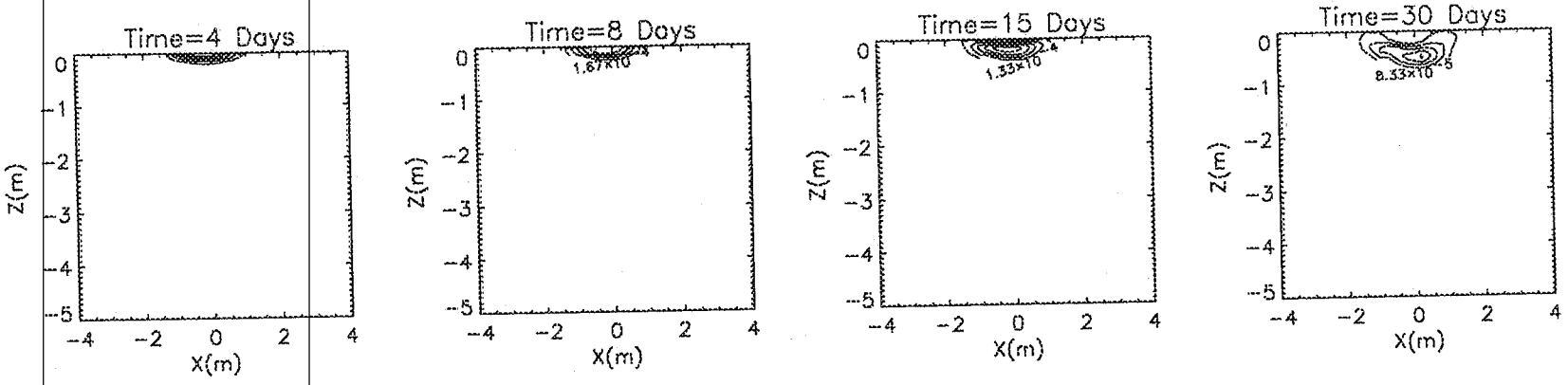




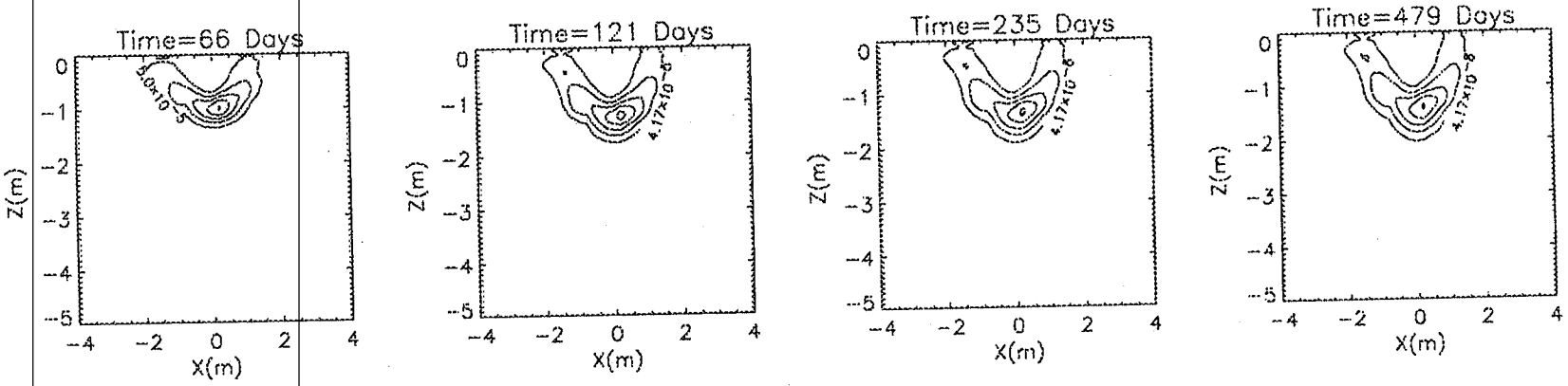


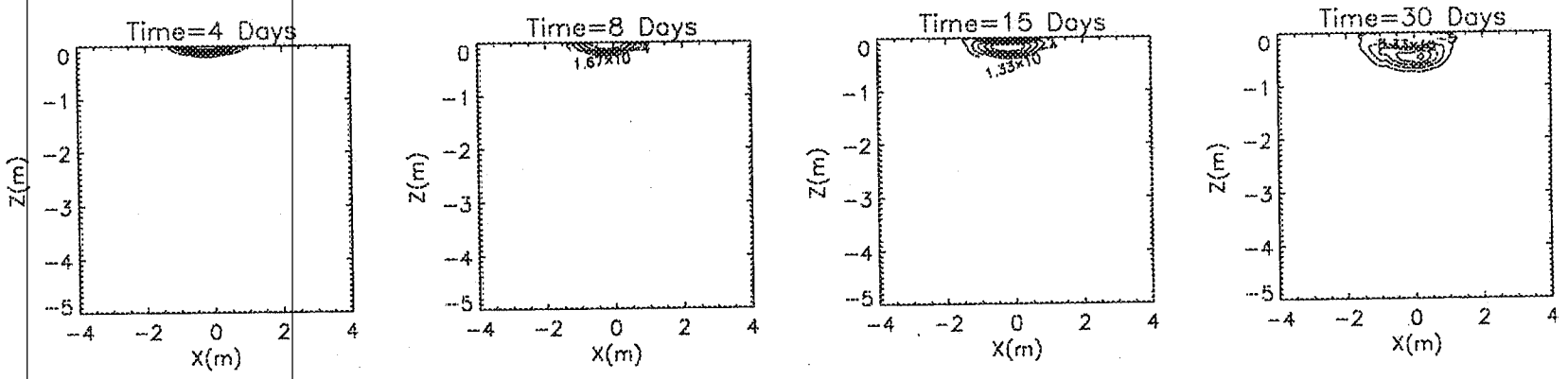
Realization 3



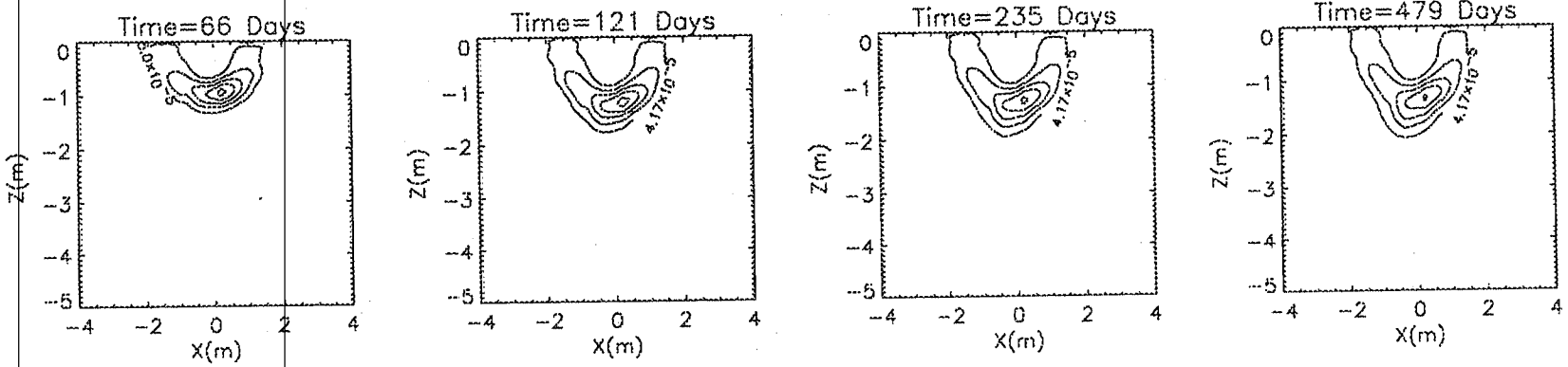


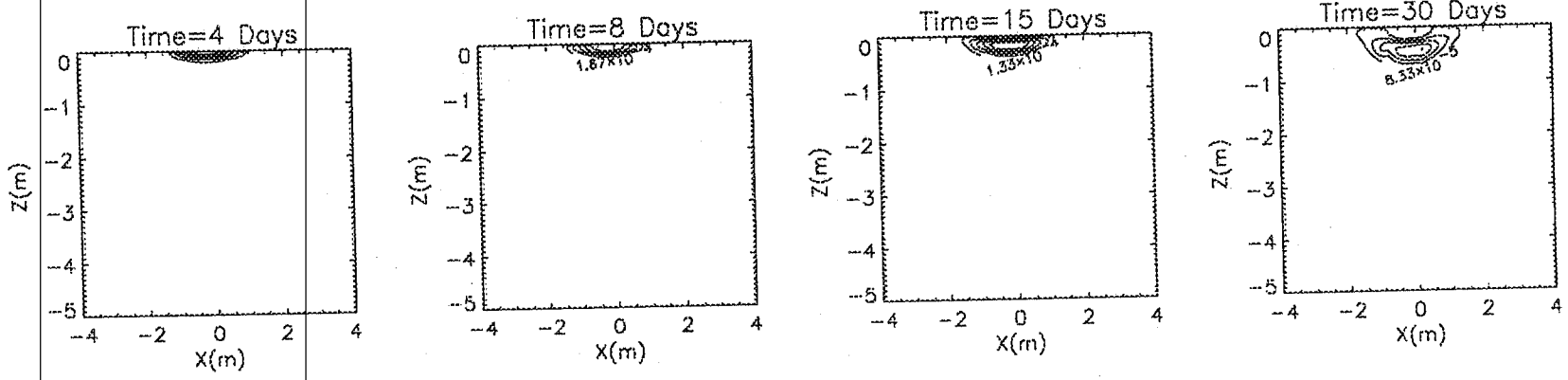
Realization 4



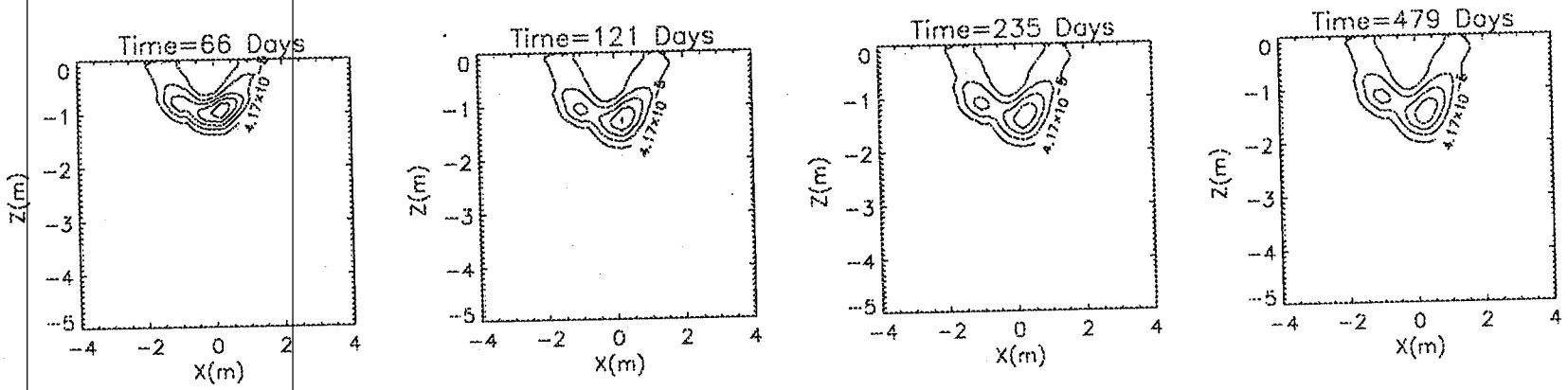


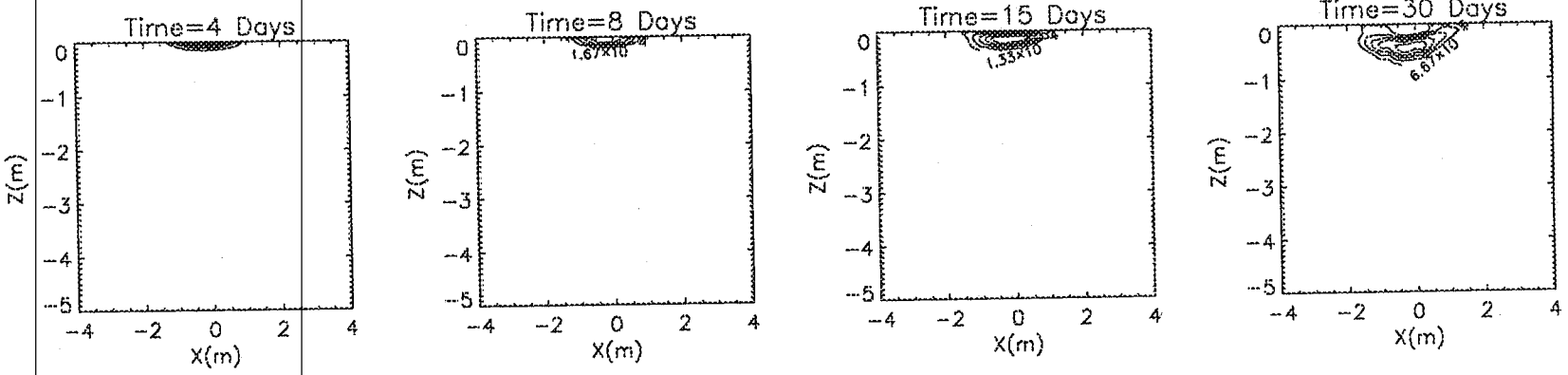
Realization 5



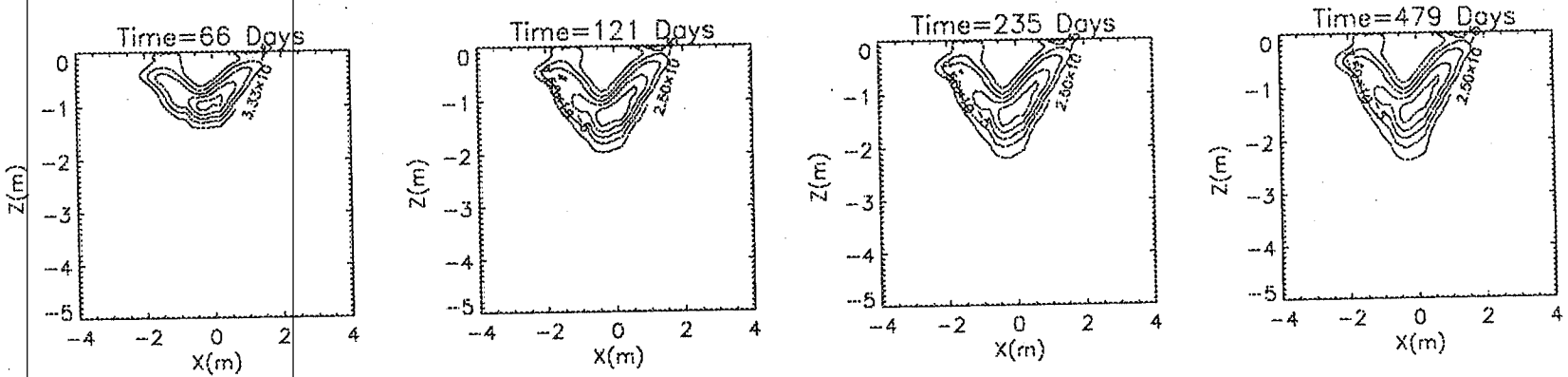


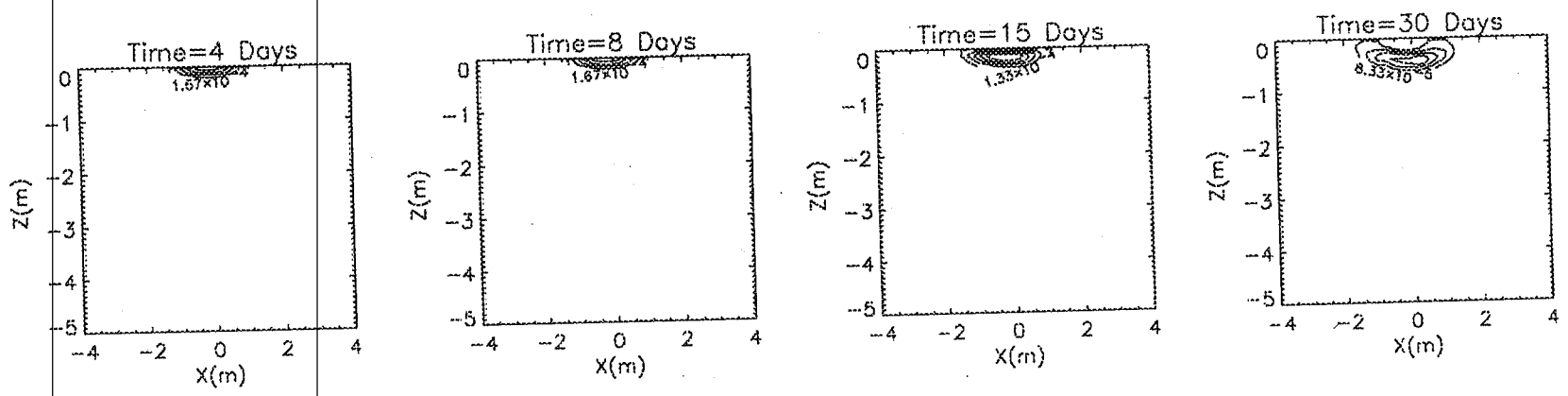
Realization 6



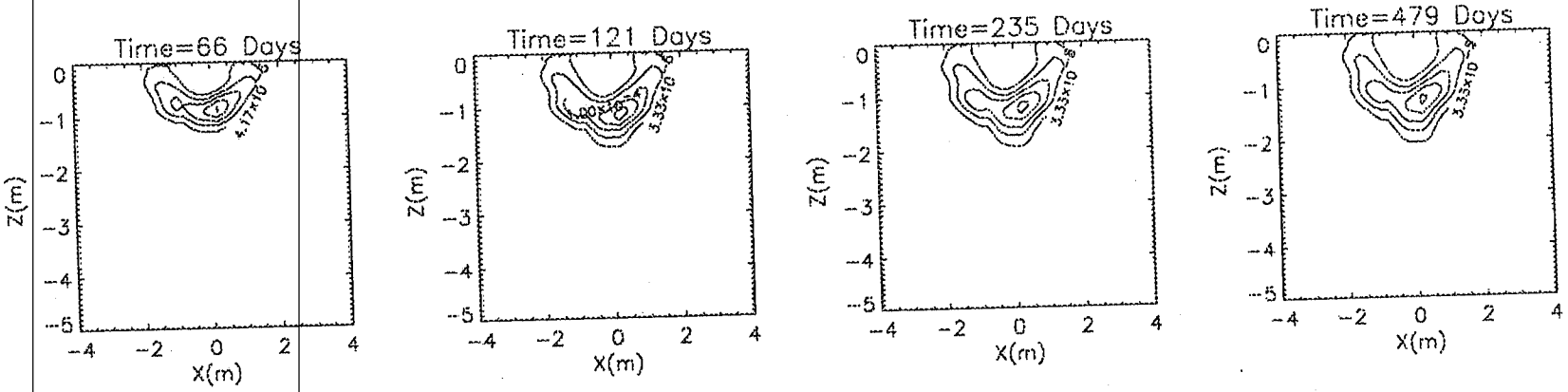


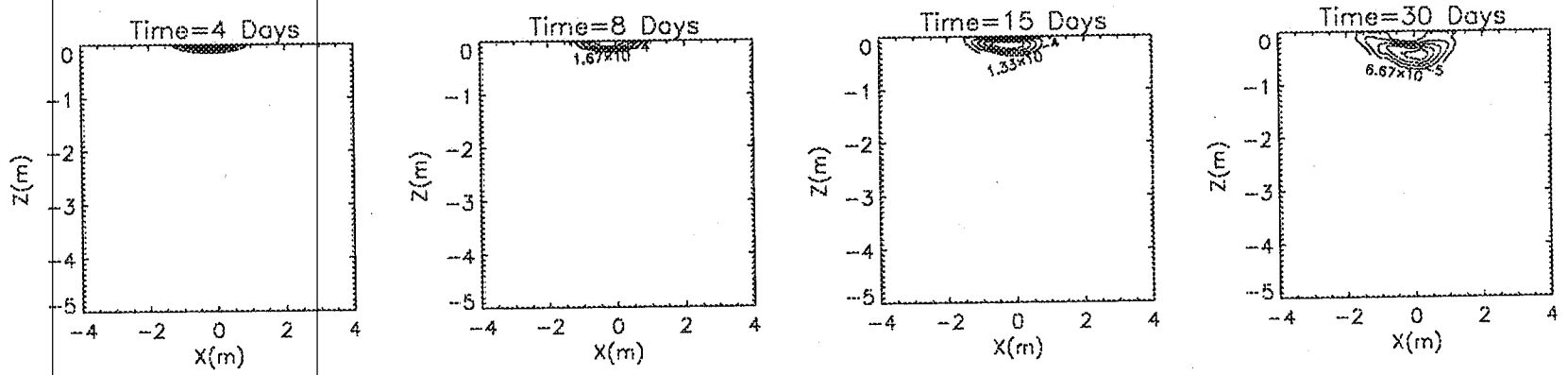
Realization 7



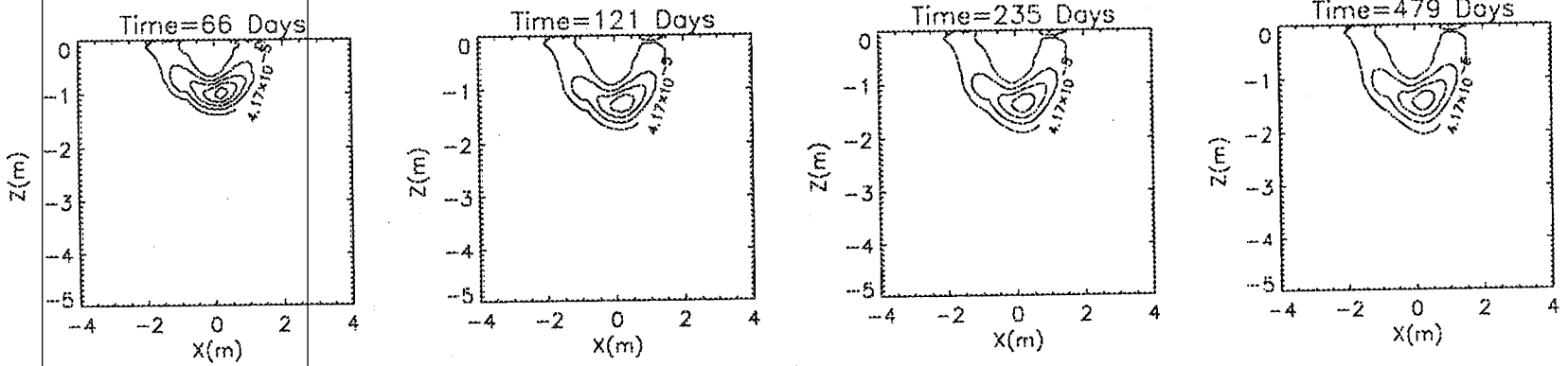


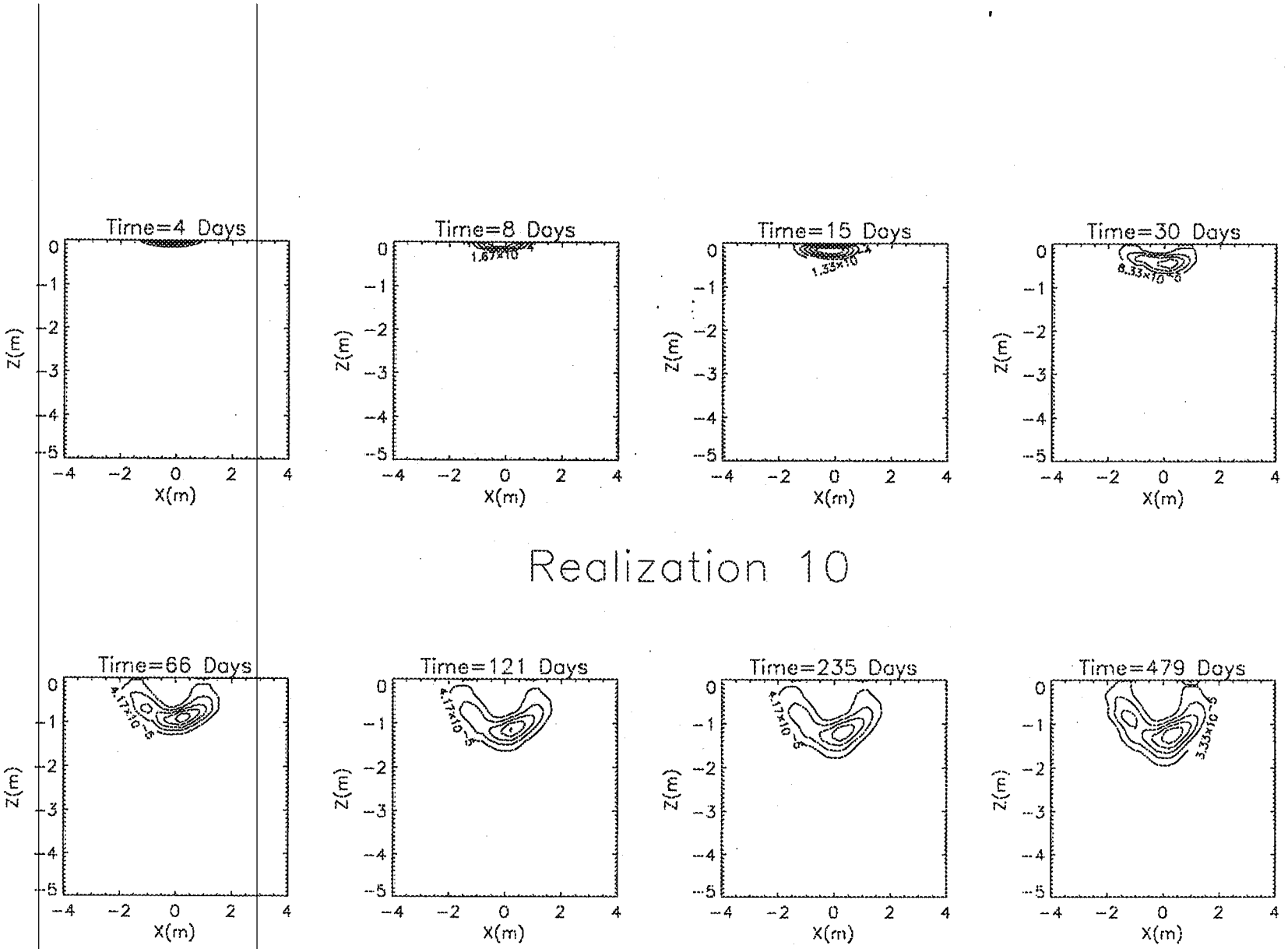
Realization 8

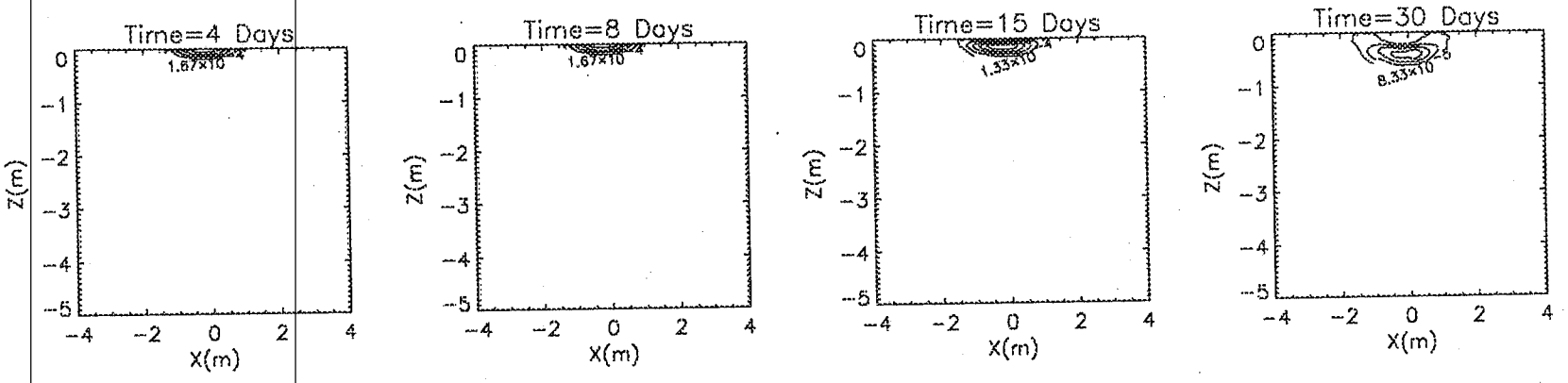




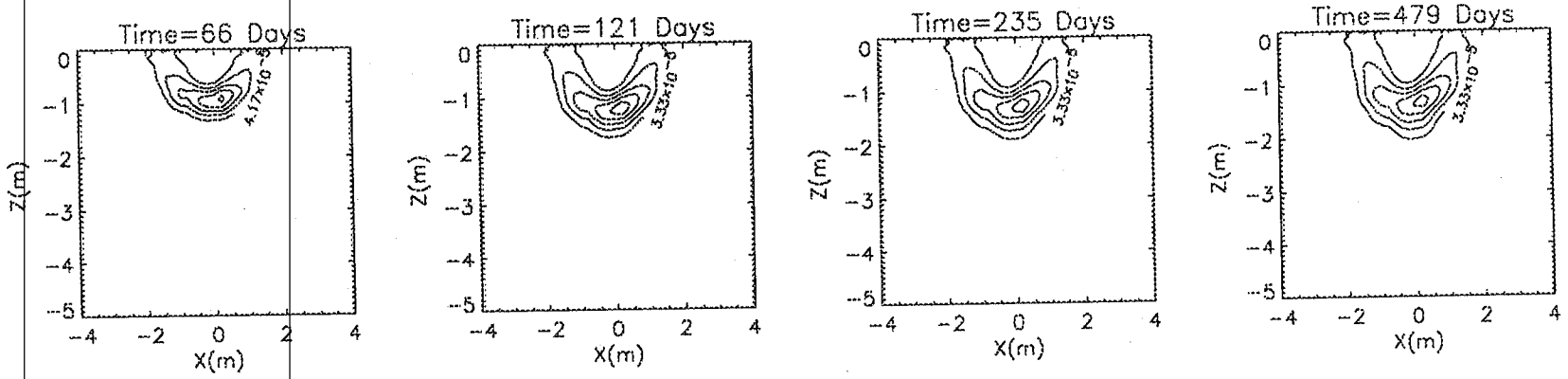
Realization 9

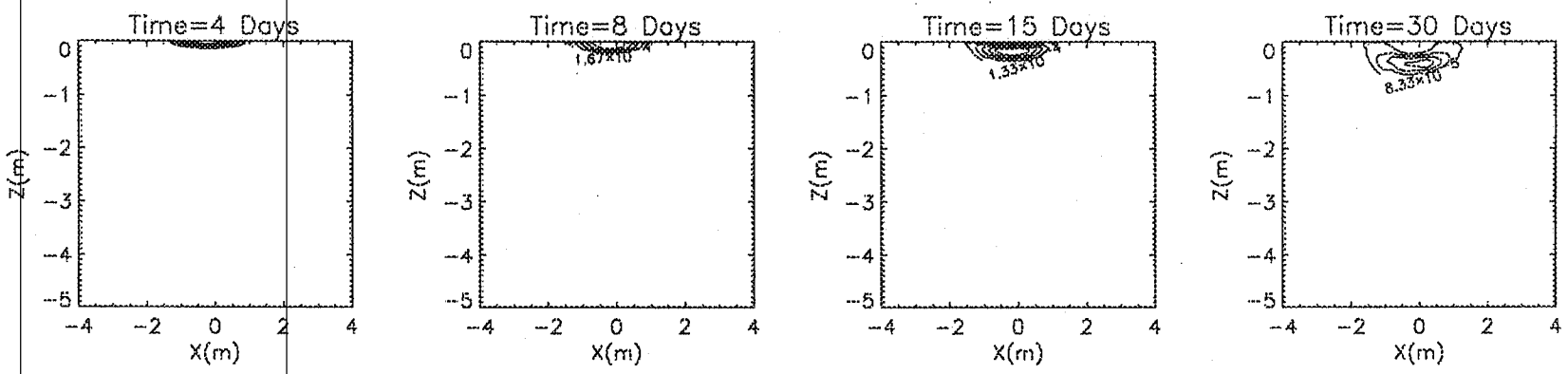



Time=4 Days
Time=8 Days
Time=15 Days
Time=30 Days
Realization 10
Time=66 Days
Time=121 Days
Time=235 Days
Time=479 Days
Z(m)
X(m)

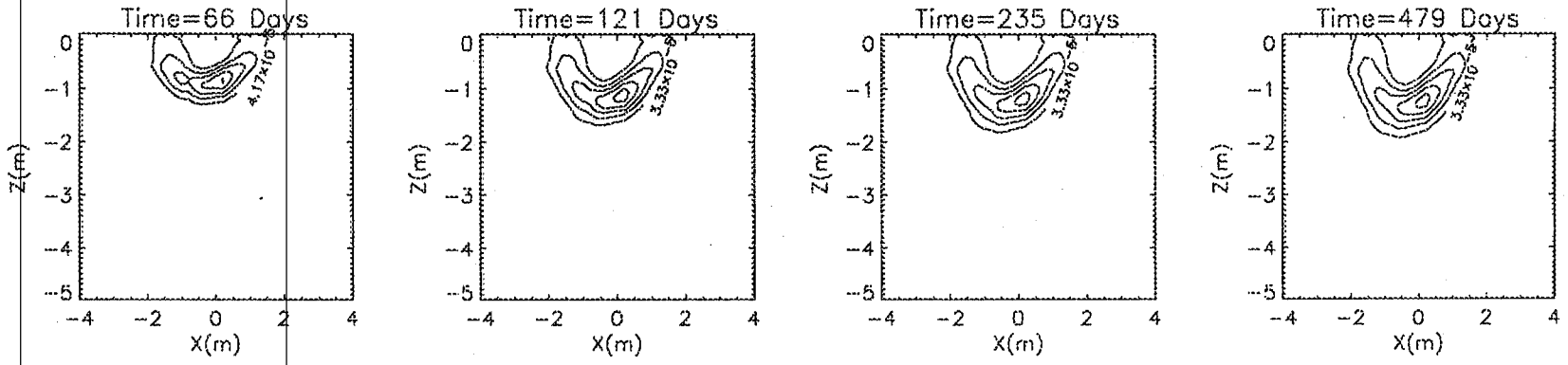


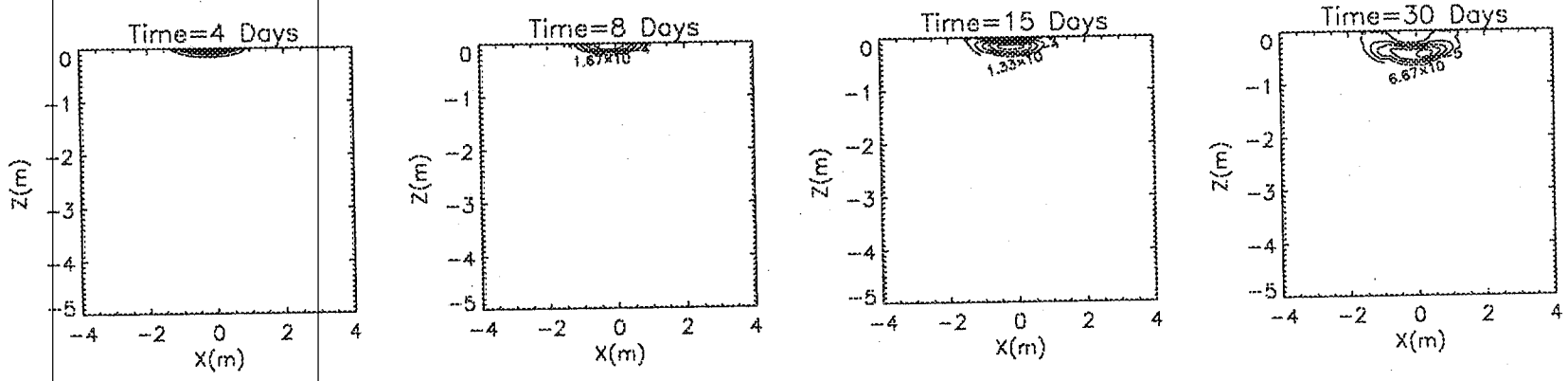
Realization 11



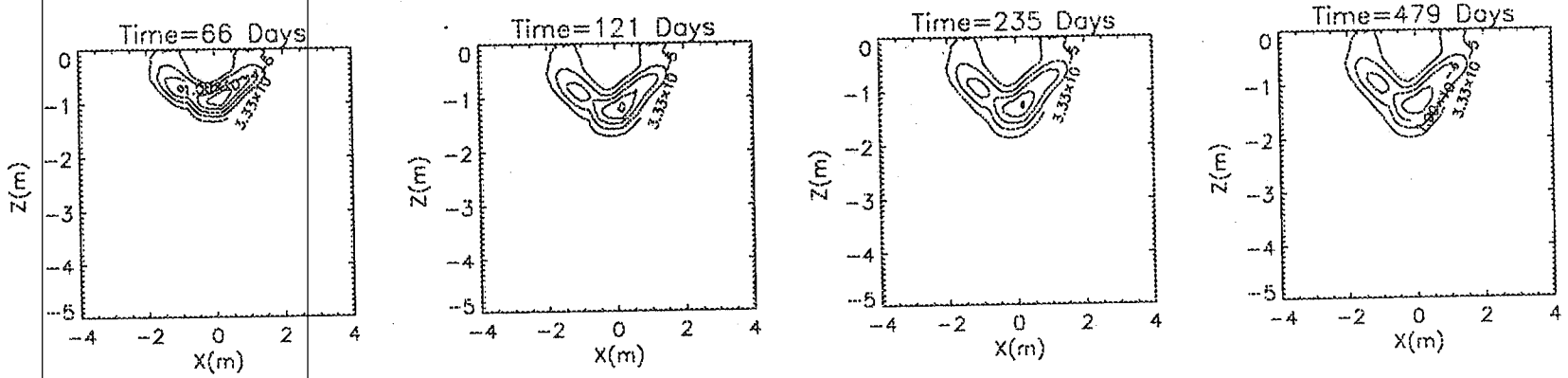


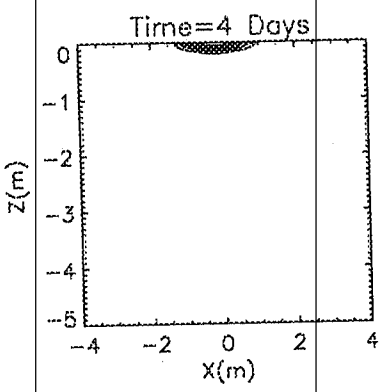
Realization 12

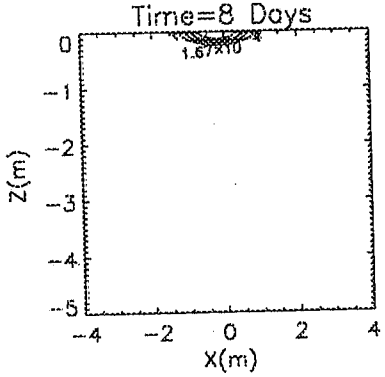


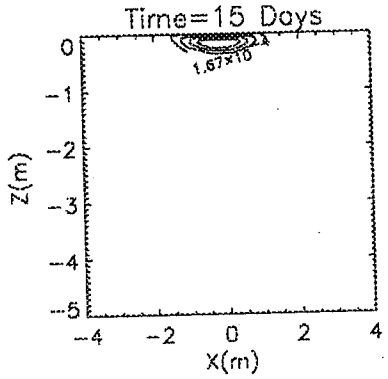


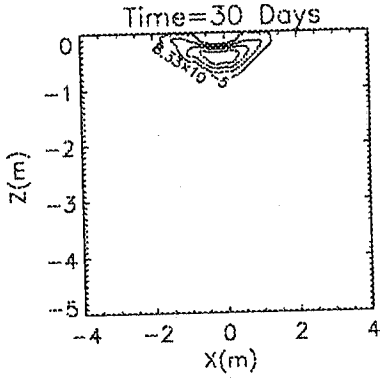
Realization 13



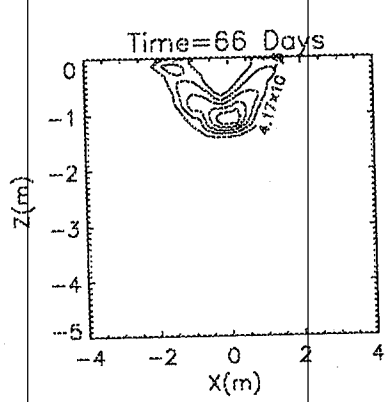


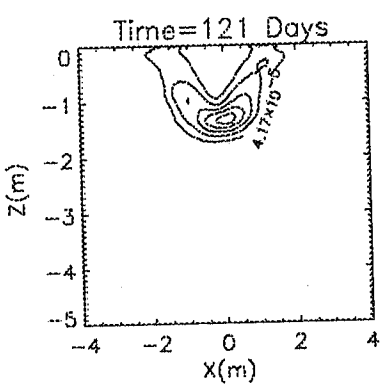


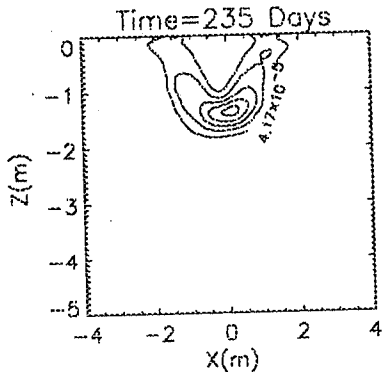


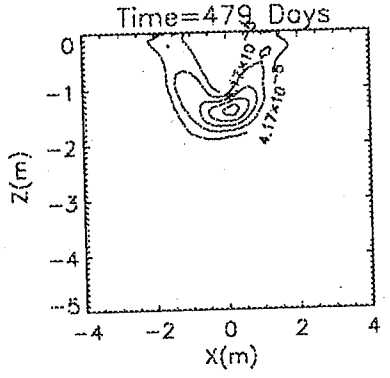


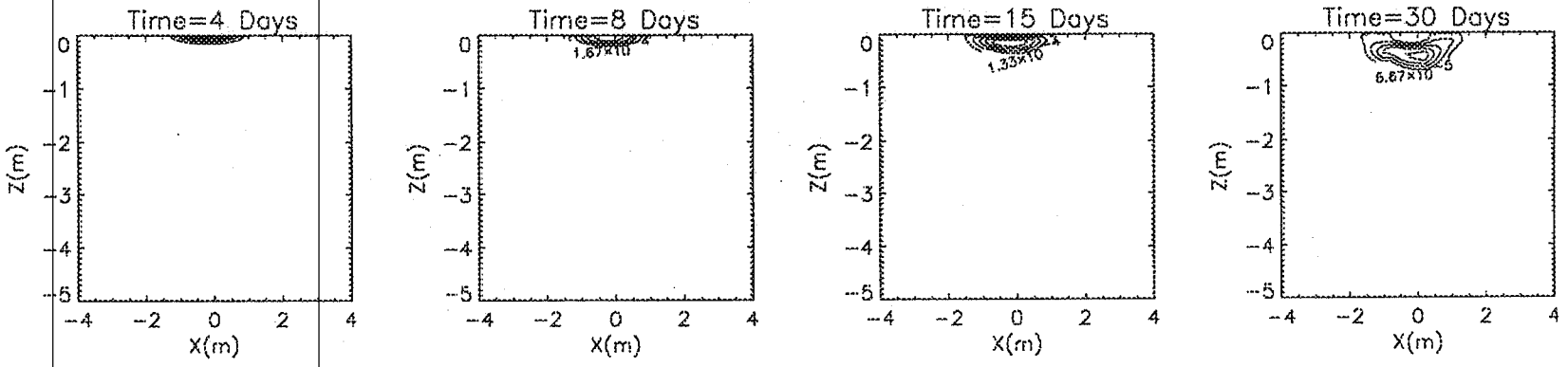
Realization 14



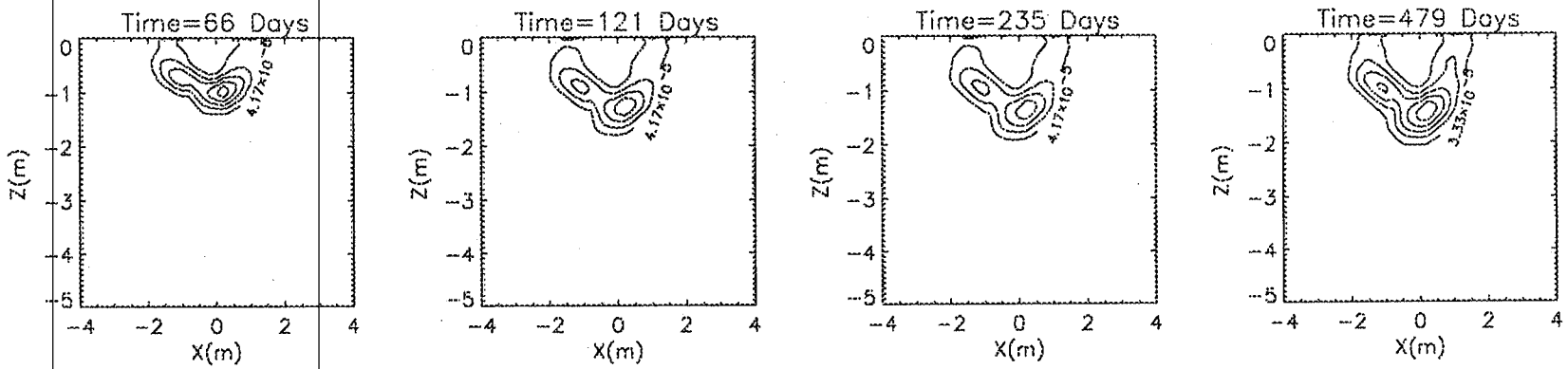


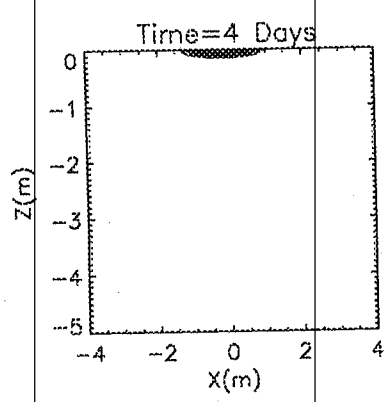


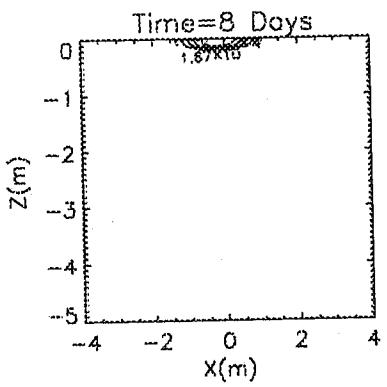


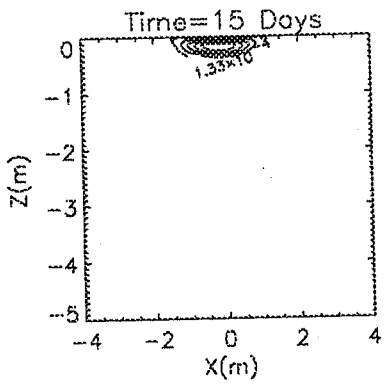


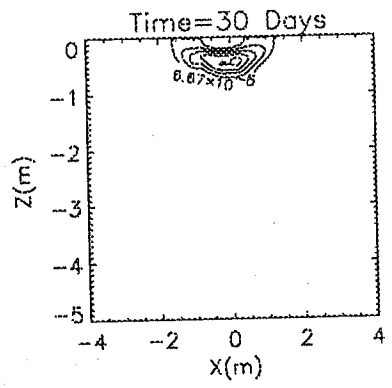
Realization 15



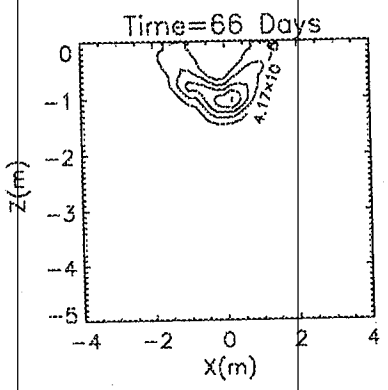


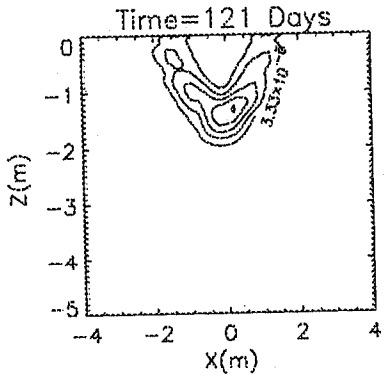


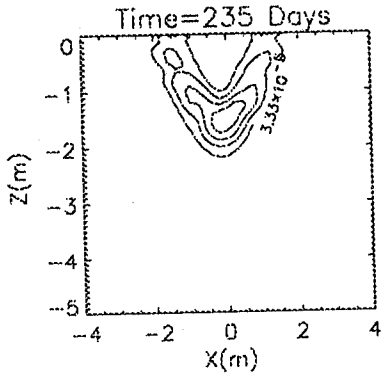


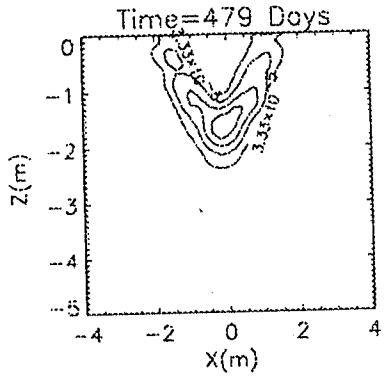


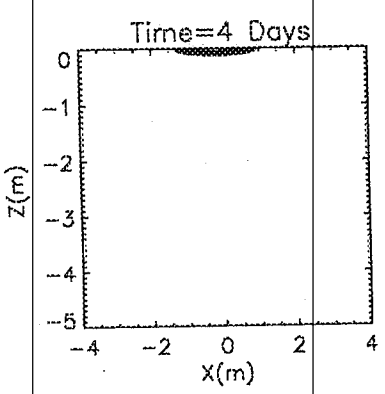
Realization 16

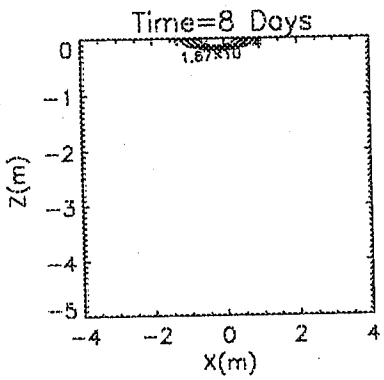


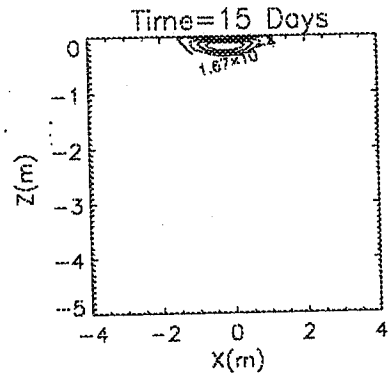


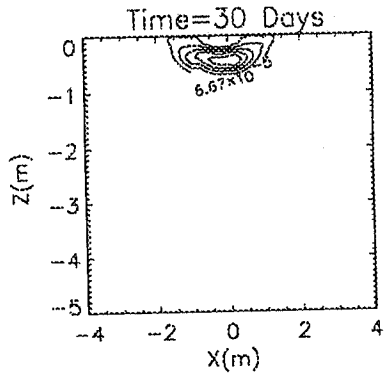




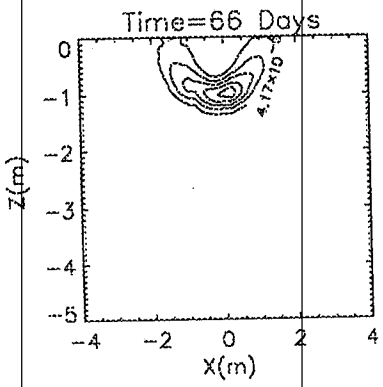


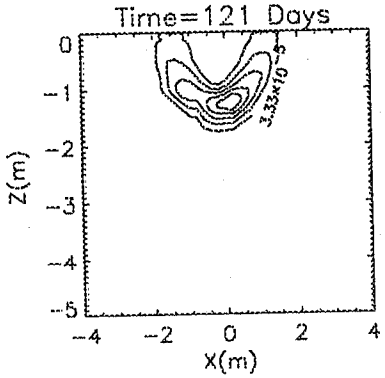


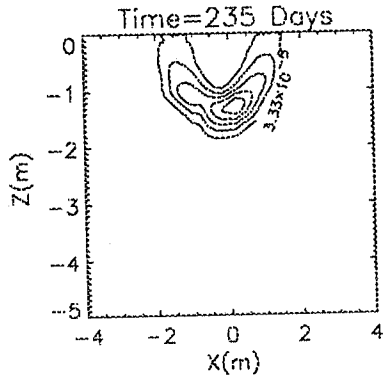


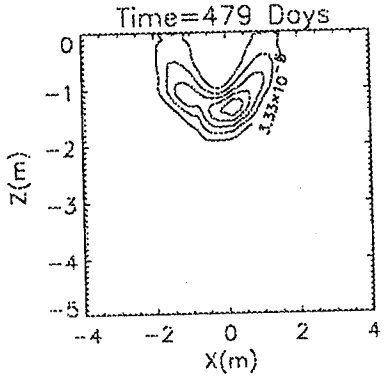


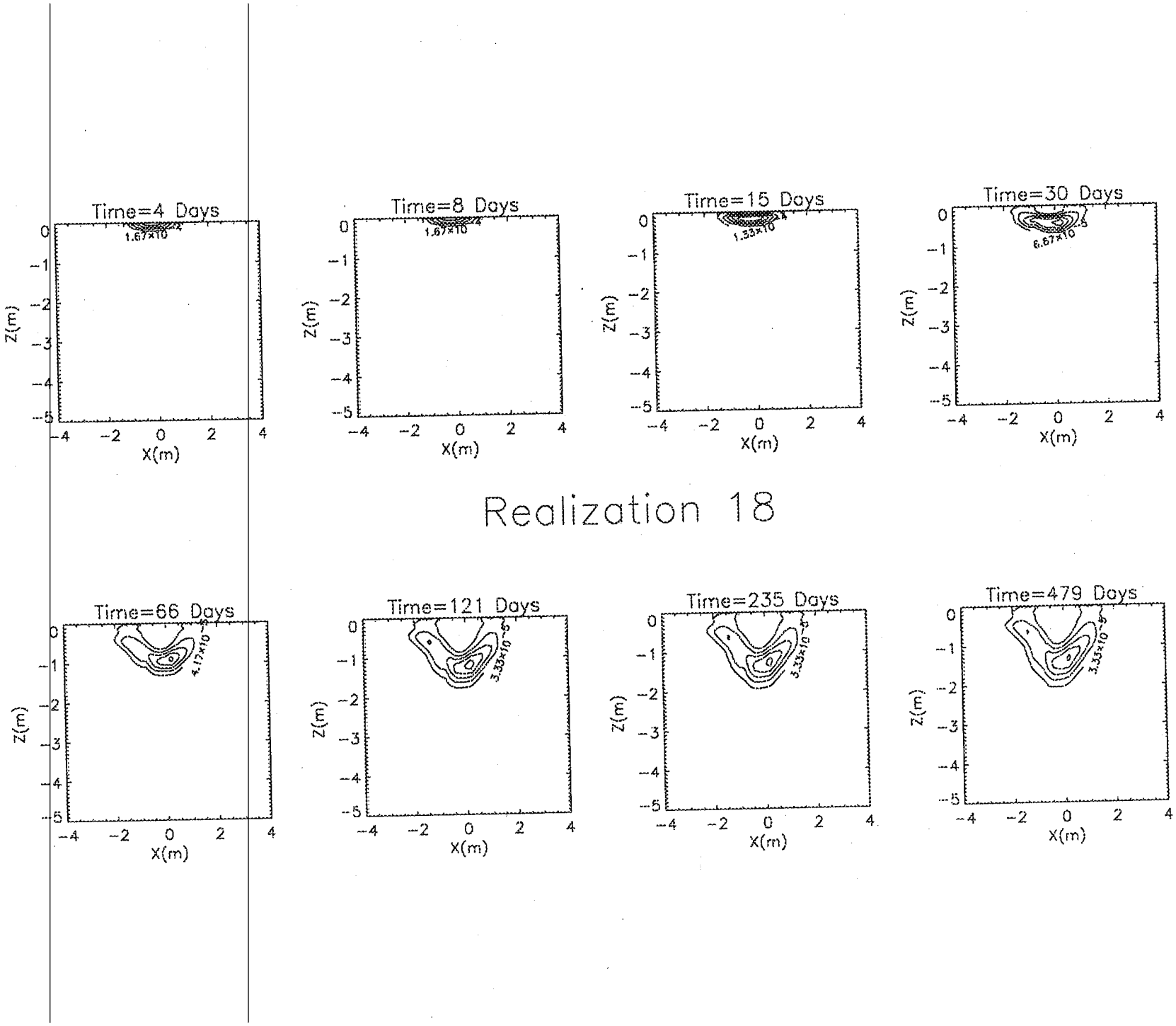
Realization 17

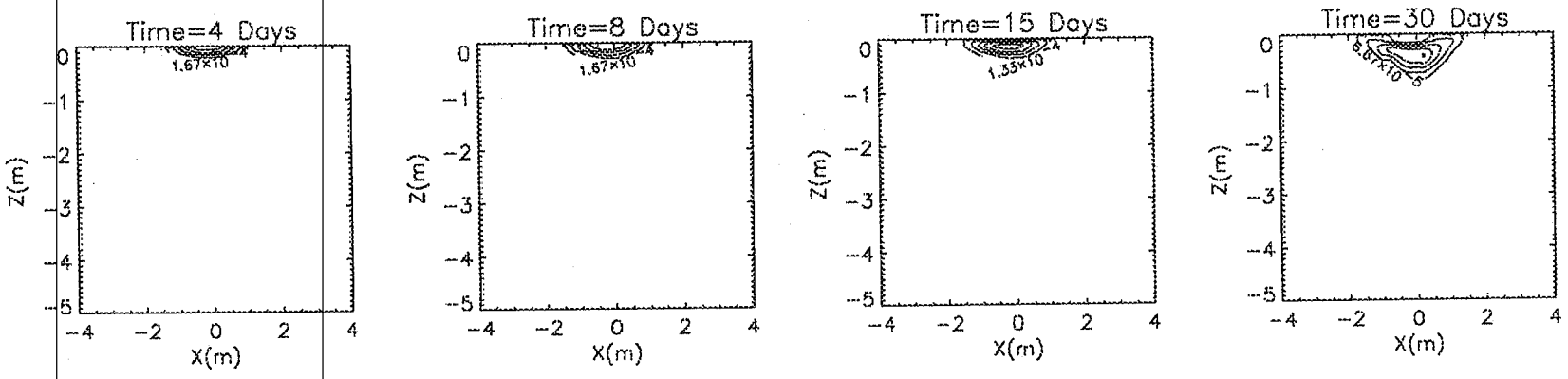




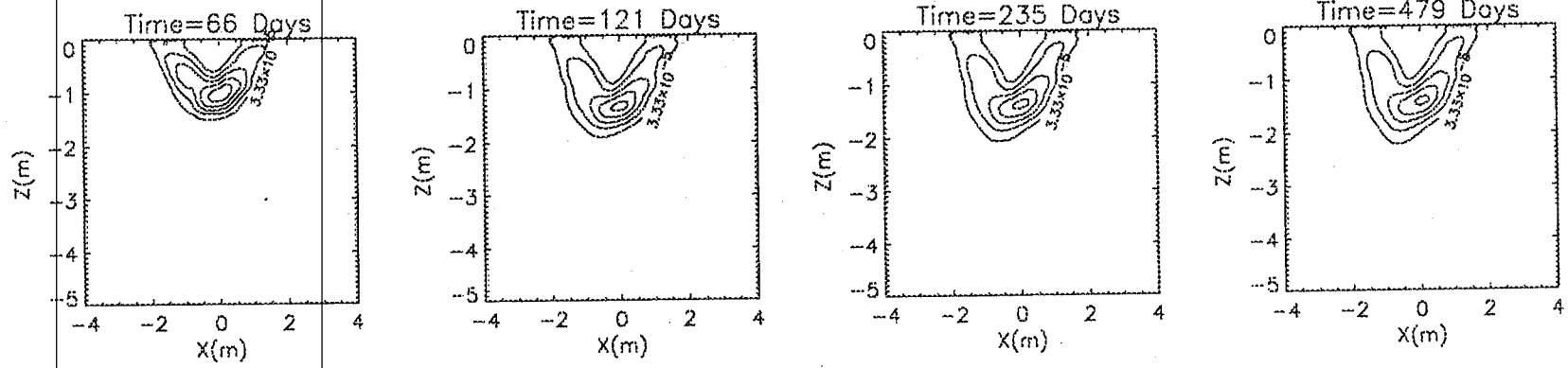


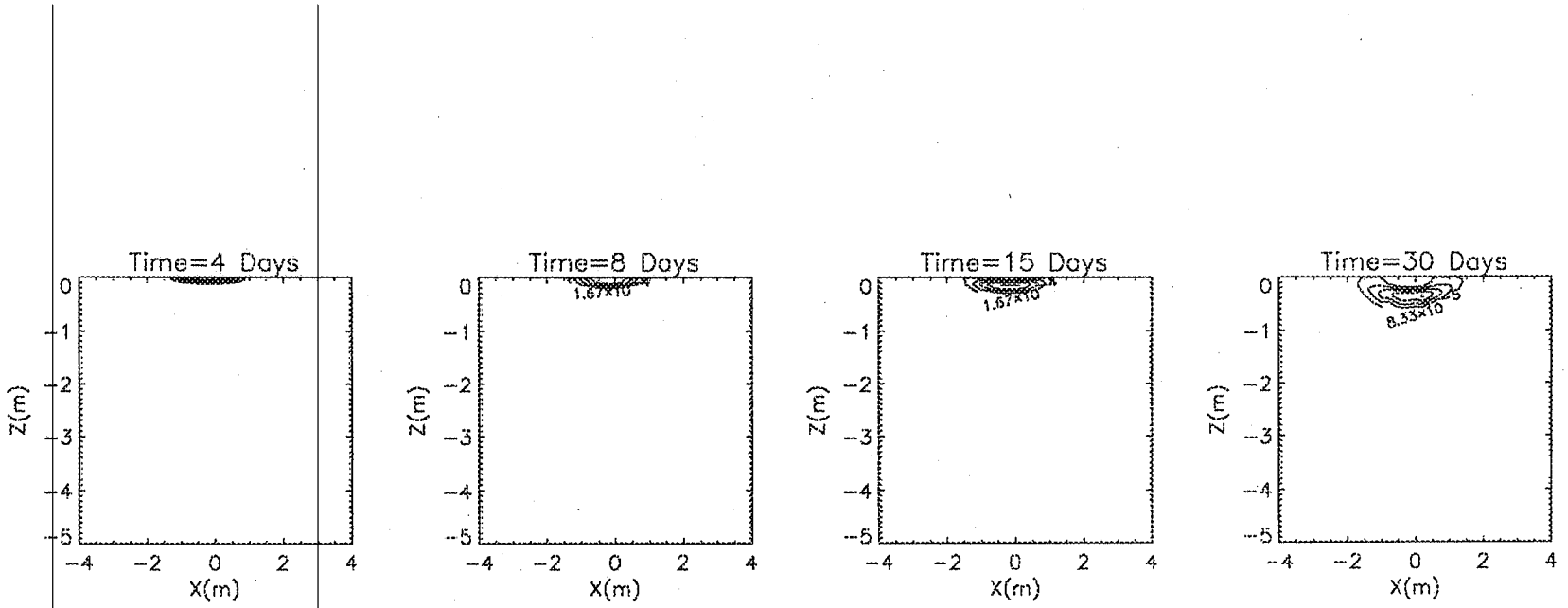



Time=4 Days
Time=8 Days
Time=15 Days
Time=30 Days
Realization 18
Time=66 Days
Time=121 Days
Time=235 Days
Time=479 Days
Z(m)
X(m)

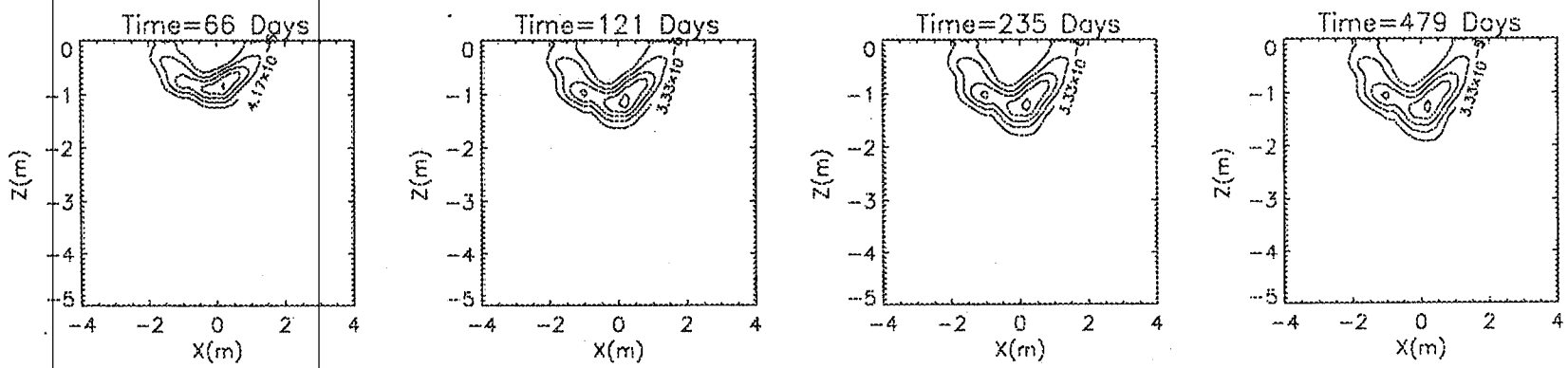


Realization 19





Realization 20



This dissertation is accepted on behalf of the faculty of the Institute by the following committee:

Allan Gutjahr
Adviser

H. H. Phillips

John L. Wilson

Bryan J. Travis

Jan 30, 1998
Date

I release this document to New Mexico Institute of Mining and Technology.

Suzan Columbo 1/28/98
Students Signature Date