

© Copyright by Patrick Michael Reed, 2002

**STRIKING THE BALANCE: LONG-TERM GROUNDWATER  
MONITORING DESIGN FOR MULTIPLE CONFLICTING OBJECTIVES**

BY

**PATRICK MICHAEL REED**

B.S., University of Missouri at Rolla, 1997  
M.S., University of Illinois at Urbana-Champaign, 1999

**THESIS**

Submitted in partial fulfillment of the requirements  
for the degree of Doctor of Philosophy in Environmental Engineering  
in the Graduate College of the  
University of Illinois at Urbana-Champaign, 2002

Urbana, Illinois

## **ABSTRACT**

This research develops a highly adaptable multiobjective long term monitoring (LTM) design methodology that aids the negotiation process by enabling decision makers to directly assess the tradeoffs among a variety of performance objectives. The monitoring methodology combines quantile kriging and the Nondominated Sorted Genetic Algorithm-II (NSGA-II) to successfully balance four objectives: (1) minimizing sampling costs, (2) maximizing the quality of interpolated plume maps, (3) maximizing the relative accuracy of contaminant mass estimates, and (4) minimizing local estimation uncertainty. Visualization is used as an aid in selecting, understanding, and balancing these performance objectives en route to a single compromise solution.

Quantile kriging was selected based on a rigorous study of 6 nonstationary plume interpolation methods for scatter-point concentration data ranging in complexity from intrinsic kriging based on intrinsic random function theory to a traditional implementation of inverse-distance weighting. Quantile kriging was the most robust of the interpolation methods, showing the least bias from the variability of contaminant samples and preferential sampling. Additionally, the method's non-parametric uncertainty estimates successfully predicted zones of high estimation error for each test case. Although this study focuses specifically on nonstationary plume data, its findings are relevant to a broad array of fields ranging from mining to life sciences where nonstationary data interpolation plays a vital role.

The tradeoffs between the 4 LTM objectives were successfully captured by developing a design methodology for the NSGA-II that uses a multi-population approach to automate parameter specification for the algorithm. The design methodology fully exploits the efficiency of the NSGA-II to enable the solution of a new class of high order multiobjective applications in

which users from any discipline seek to select, understand, and balance more than two performance criteria.

The LTM design methodology developed in this research addresses the two most important problems LTM practitioners face in the design process: (1) selecting monitoring objectives and (2) balancing these objectives. Tools such as multiobjective genetic algorithms that are capable of high order multiobjective optimization (i.e., optimizing a system for more than 2 objectives) can serve as an interface between any physical system being designed and the human decision process. This thesis demonstrates that combining higher order Pareto optimization with visualization can allow designers in any field to assess the mathematical models used to represent their objectives, discover how their objectives are affecting designs, and negotiate a final design that balances their conflicting design preferences.

*To my wife Christy, my daughter Arista, and  
in loving memory of Linas H. Reed*

## **ACKNOWLEDGMENTS**

This research was funded by the United States Environmental Protection Agency's (USEPA) Science to Achieve Results (STAR) doctoral fellowship (grant number U-91554601-1). I would like to fully acknowledge and thank the USEPA's National Center of Environmental Research for providing a fellowship program that provides its participants ample support to fully explore their research topics.

I would like to express my gratitude to my advisor Professor Barbara Minsker for encouraging me through both the successes and failures inherent to research. I especially appreciate the intellectual freedom she afforded me, allowing me to actively pursue varied research interests that crossed disciplinary borders. Her countless hours spent guiding both my writing and research efforts will be invaluable for my future career ambitions.

I would like to thank each of my doctoral committee members, Professor Albert Valocchi, Professor Tim Ellsworth, and Professor David Goldberg for patiently investing their time in guiding me through my research. Professor Valocchi has helped me to shape my research interests in long-term groundwater monitoring design from my first days on the University of Illinois' campus. Professor Ellsworth's willingness to collaborate and his expertise in geostatistics were essential to my research. Likewise, Professor Goldberg's collaboration and guidance in genetic and evolutionary computation were pivotal to the overall success of this research.

I owe a debt of gratitude to the entire faculty of Environmental Engineering and Science program for always being accessible whether for my course work or my research. Specifically, I would like to thank Professor Wayland Eheart for encouraging me and sharing his perspectives

on my research. Also, I would like to thank Professors Robert Sanford and Charles Werth for their time spent helping me to improve my presentation skills.

Beyond the borders of the University of Illinois campus, I would like to thank Dr. Reed Maxwell for providing the simulation data used to develop the test cases presented in my research. Also, I would like to thank Dr. Donna Rizzo for sharing her perspectives on both my research and groundwater monitoring design as a whole.

I would have fallen prey to the effects of sensory deprivation during my many long days and nights on the 4<sup>th</sup> floor of Newmark had it not been for the uplifting and helpful interactions I had with my fellow research group members and graduate students. Dr. Yong Liu, you blazed the trail I was to follow and inspired me with your work ethic. Felipe Espinoza, I owe you for the imagery of “confluence” and appreciate our many conversations on art and the general state of the world. I would also like to thank Glenn Hammond for his friendship and for entertaining me with his animated approach to research programming.

Finally, I would like to express my gratitude to my entire family for their continued support and encouragement. My parents Randy and Cindy Clavin have always supported my ambitions and encouraged me to chase my dreams. I would like to thank my wife Christy for always pressing me to perform to the best of my abilities. Her sacrifices, love, and steadfast support have brought about all of my successes.

# TABLE OF CONTENTS

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
1.1	Objectives and Scope .....	2
1.2	Summary of Research.....	2
1.2.1	<i>Chapter 2: Literature Review .....</i>	4
1.2.2	<i>Chapter 3: Optimization in Pareto Space.....</i>	4
1.2.3	<i>Chapter 4: Simplifying Optimization in Pareto Space .....</i>	5
1.2.4	<i>Chapter 5: Spatial Interpolation Methods for Plume Data.....</i>	5
1.2.5	<i>Chapter 6: Balancing Performance Criteria .....</i>	6
<b>2</b>	<b>LITERATURE REVIEW .....</b>	<b>8</b>
2.1	Geostatistical Approaches .....	8
2.1.1	<i>Minimization of Local Estimation Variance.....</i>	9
2.1.2	<i>Minimization of Global Estimation Variance.....</i>	9
2.1.3	<i>Minimization of Modeling Uncertainty.....</i>	10
2.2	Hazardous Waste Detection Monitoring .....	11
2.3	Spatial and Temporal Redundancy Analysis .....	12
<b>3</b>	<b>OPTIMIZATION IN PARETO SPACE .....</b>	<b>14</b>
3.1	Introduction.....	14
3.1.1	<i>Previous Work.....</i>	15
3.1.2	<i>Motivation and Scope.....</i>	16
3.2	Monitoring Application.....	17
3.2.1	<i>Problem Formulation .....</i>	19
3.2.2	<i>Nonlinear Spatial Interpolation .....</i>	20
3.3	The Basics of the NSGA .....	22
3.3.1	<i>Pareto Dominance.....</i>	23
3.3.2	<i>Niching &amp; Sharing.....</i>	23
3.4	Multiobjective Search & Optimization.....	25
3.4.1	<i>Performance Considerations .....</i>	25
3.4.2	<i>Defining a Measure of Relative Performance .....</i>	35
3.5	Results & Discussion.....	37
3.5.1	<i>Performance of the NSGA &amp; the Elitist NSGA .....</i>	37
3.5.2	<i>Performance Under Increased Selection Pressure.....</i>	44
3.6	Conclusions .....	50
<b>4</b>	<b>SIMPLIFYING OPTIMIZATION IN PARETO SPACE .....</b>	<b>53</b>
4.1	Introduction.....	53
4.2	Monitoring Application .....	53
4.2.1	<i>Test Case Data.....</i>	53
4.2.2	<i>Problem Formulation .....</i>	55
4.3	The NSGA-II Basics.....	56
4.4	Simplifying Multiobjective Search & Optimization .....	59
4.4.1	<i>Initial Considerations .....</i>	60

4.4.2	<i>Step 1: Preliminary Problem Analysis</i> .....	60
4.4.3	<i>Step 2: Balancing Innovation and Disruption</i> .....	64
4.4.4	<i>Step 3: Multi-Population Trial Runs</i> .....	66
4.5	Conclusions .....	70
<b>5</b>	<b>SPATIAL INTERPOLATION METHODS .....</b>	<b>71</b>
5.1	Introduction.....	71
5.1.1	<i>Motivation &amp; Scope</i> .....	71
5.1.2	<i>Previous Work in the Estimation of Nonstationary Phenomena</i> .....	74
5.1.3	<i>Test Cases</i> .....	78
5.2	Methods.....	81
5.2.1	<i>Geostatistical Methods</i> .....	81
5.2.2	<i>Deterministic Methods</i> .....	89
5.2.3	<i>Defining Estimation Neighborhoods</i> .....	91
5.2.4	<i>Performance Measures</i> .....	92
5.3	Results & Discussion.....	94
5.3.1	<i>Large Test Case</i> .....	95
5.3.2	<i>Medium Test Case</i> .....	104
5.3.3	<i>Small Test Case</i> .....	108
5.4	Conclusions.....	114
<b>6</b>	<b>BALANCING PERFORMANCE CRITERIA.....</b>	<b>116</b>
6.1	Introduction.....	116
6.1.1	<i>Motivation &amp; Scope</i> .....	116
6.1.2	<i>Background on LTM Design</i> .....	118
6.2	Test Case Data.....	120
6.3	Methodology .....	122
6.3.1	<i>Generalized Multiobjective Formulation</i> .....	123
6.3.2	<i>Design Objectives Suite</i> .....	124
6.3.3	<i>Plume Interpolation using Quantile Kriging</i> .....	127
6.3.4	<i>Multiobjective Search &amp; Optimization</i> .....	127
6.4	Results & Discussion.....	132
6.4.1	<i>Understanding the Interactions of Design Objectives</i> .....	132
6.4.2	<i>Negotiating a Final Design</i> .....	143
6.5	Conclusions.....	147
<b>7</b>	<b>CONCLUDING REMARKS .....</b>	<b>149</b>
7.1	Thesis Contributions.....	149
7.2	Future Research.....	151
<b>REFERENCES.....</b>		<b>155</b>
<b>APPENDIX A. GEOSTATISTICAL METHODS.....</b>		<b>170</b>
<b>APPENDIX B. STRUCTURAL ANALYSIS .....</b>		<b>177</b>

<b>APPENDIX C. CROSS VALIDATION.....</b>	<b>198</b>
<b>APPENDIX D. FREQUENCY ANALYSIS WITH ZEROS .....</b>	<b>210</b>
<b>VITA.....</b>	<b>213</b>

## LIST OF FIGURES

Figure 3.1 The 50 potential sampling locations (designated by the x's above) within a 20 well multi-level monitoring network .....	17
Figure 3.2 Actual tradeoff between the squared relative estimation error and cost. Each objective has been scaled to fall within the interval [0, 1]. The maximum and minimum SREE values found in $\Omega$ were 0 and 13,000,000, respectively. The maximum and minimum cost values found in $\Omega$ were \$18,420 and \$1473, respectively. ....	36
Figure 3.3 Selection of the proper population size for both the NSGA and the Elitist NSGA using the best RSM value attained from a single generation .....	37
Figure 3.4 NSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing .....	39
Figure 3.5 Elitist NSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing.....	40
Figure 3.6 Analysis of the NSGA's performance for the full range of possible values that the parameters controlling niching and elitism can be assigned when using (a) genotypic sharing and (b) phenotypic sharing. The arrows designate the algorithm's performance when these parameters are set equal to the recommended niche radius attained from the relationships presented by Deb & Goldberg (1989). .....	42
Figure 3.7 Performance of both the Elitist NSGA and the NSGA as functions of generation....	43
Figure 3.8 Performance of both the Elitist NSGA and the NSGA as a function of generation after rescaling the fitness assignments .....	45
Figure 3.9 sNSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing .....	46
Figure 3.10 Elitist sNSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing.....	47
Figure 3.11 Analysis of the sNSGA's performance for the full range of possible values that the parameters controlling niching and elitism can be assigned when using (a) genotypic and (b) phenotypic sharing. The arrows designate the algorithm's performance when these parameters are set equal to the recommended niche radius attained from the relationships presented by Deb & Goldberg (1989).....	48
Figure 4.1 The enumerated tradeoff between Cost and SREE .....	56
Figure 4.2 (a) Offline results for Run 1 (b) Combined offline results for Runs 1 & 2 .....	67
Figure 4.3 (a) Combined offline results for Runs 1, 2, & 3 (b) Combined offline results for Runs 1, 2, 3, & 4 .....	68
Figure 5.1 Test case monitoring networks .....	80
Figure 5.2 Illustration of rank-order transformation used in quantile kriging .....	85
Figure 5.3 RMSE values from cross-validation for the large test case (mg/m <sup>3</sup> ).....	96
Figure 5.4 RMSE values from interpolation for the large test case (mg/m <sup>3</sup> ).....	99
Figure 5.5 Spatial distributions of estimation errors for the large test case (mg/m <sup>3</sup> ).....	100
Figure 5.6 RMSE values from cross-validation for the medium test case (mg/m <sup>3</sup> ).....	104
Figure 5.7 RMSE values from interpolation for the medium test case (mg/m <sup>3</sup> ).....	106
Figure 5.8 Spatial distribution of estimation errors for the medium test case (mg/m <sup>3</sup> ) .....	107
Figure 5.9 RMSE values from cross-validation for the small test case (mg/m <sup>3</sup> ) .....	109
Figure 5.10 RMSE values from interpolation for the small test case (mg/m <sup>3</sup> ) .....	110

Figure 5.11 Spatial distribution of estimation errors for the small test case (mg/m <sup>3</sup> ).....	112
Figure 5.12 Spatial distributions of local uncertainty for QK (mg/m <sup>3</sup> ) .....	113
Figure 6.1 (a) 3-dimensional view of the monitoring test case (b) XY-plane view of monitoring wells .....	121
Figure 6.2 (a) Cost—SREE tradeoff (b) Cost—Mass tradeoff (c) Cost—Uncertainty tradeoff	134
Figure 6.3 Sampling locations corresponding to solutions from the Cost—SREE tradeoff (a) 19 sample solution (b) 31 sample solution (c) 44 sample solution.....	135
Figure 6.4 Sampling locations corresponding to the 31 sample solution from the Cost—Mass tradeoff.....	136
Figure 6.5 Sampling locations corresponding to solutions from the Cost—Uncertainty tradeoff (a) 19 sample solution (b) 31 sample solution (c) 44 sample solution .....	137
Figure 6.6 Three dimensional cross sectional views of how SREE changes for (a) the SREE(31) solution (b) the Mass(31) solution (c) the Uncertainty(31) solution .....	139
Figure 6.7 Three-dimensional cross sectional views of how Uncertainty changes for (a) the SREE(31) solution (b) the Mass(31) solution (c) the Uncertainty(31) solution.....	142
Figure 6.8 Objective conflicts between (a) SREE and Uncertainty (b) Mass Error and Uncertainty (c) SREE and Mass Error for the 44 sample solutions .....	144
Figure 6.9 Compromise solution's (a) XY-plane view of sampling locations (b) three dimensional cross section of how SREE changes (c) three dimensional cross section of how Uncertainty changes .....	146
Figure B.1 Large case variogram model for the ordinary kriging system in the X-direction....	177
Figure B.2 Large case variogram model for the ordinary kriging system in the Y-direction....	178
Figure B.3 Large case variogram model for the ordinary kriging system in the Z-direction....	179
Figure B.4 Large case variogram model for the Multigaussian kriging system in the X-direction .....	180
Figure B.5 Large case variogram model for the Multigaussian kriging system in the Y-direction .....	181
Figure B.6 Large case variogram model for the Multigaussian kriging system in the Z-direction .....	182
Figure B.7 Large case variogram model for the quantile kriging system in the X-direction ....	183
Figure B.8 Large case variogram model for the quantile kriging system in the Y-direction ....	184
Figure B.9 Large case variogram model for the quantile kriging system in the Z-direction.....	185
Figure B.10 Large case Isatis output specifying the generalized covariance model for the intrinsic kriging system .....	186
Figure B.11 Large case Isatis output specifying the trend function for the intrinsic kriging system .....	187
Figure B.12 Medium case isotropic variogram model for the ordinary kriging system.....	188
Figure B.13 Medium case isotropic variogram model for the Multigaussian kriging system....	189
Figure B.14 Medium case isotropic variogram model for the quantile kriging system .....	190
Figure B.15 Medium case Isatis output specifying the generalized covariance model for the intrinsic kriging system.....	191
Figure B.16 Medium case Isatis output specifying the trend function for the intrinsic kriging system .....	192
Figure B.17 Small case pure nugget variogram model for the ordinary kriging system .....	193
Figure B.18 Small case pure nugget variogram model for the Multigaussian kriging system... ..	194
Figure B.19 Small case pure nugget variogram model for the quantile kriging system.....	195

Figure B.20 Small case Isatis output specifying the generalized covariance model for the intrinsic kriging system .....	196
Figure B.21 Small case Isatis output specifying the trend function for the intrinsic kriging system .....	197
Figure C.1 Cross validation analysis of the anisotropic variogram model used for ordinary kriging in the large case .....	198
Figure C.2 Cross validation analysis of the anisotropic variogram model used for Multigaussian kriging in the large case .....	199
Figure C.3 Cross validation analysis of the anisotropic variogram model used for quantile kriging in the large case .....	200
Figure C.4 Cross validation analysis of the isotropic variogram model used for intrinsic kriging in the large case .....	201
Figure C.5 Cross validation analysis of the isotropic variogram model used for ordinary kriging in the medium case .....	202
Figure C.6 Cross validation analysis of the isotropic variogram model used for Multigaussian kriging in the medium case .....	203
Figure C.7 Cross validation analysis of the isotropic variogram model used for quantile kriging in the medium case .....	204
Figure C.8 Cross validation analysis of the nugget variogram model used for intrinsic kriging in the medium case .....	205
Figure C.9 Cross validation analysis of the nugget variogram model used for ordinary kriging in the small case .....	206
Figure C.10 Cross validation analysis of the nugget variogram model used for Multigaussian kriging in the small case .....	207
Figure C.11 Cross validation analysis of the nugget variogram model used for quantile kriging in the small case .....	208
Figure C.12 Cross validation analysis of the isotropic variogram model used for intrinsic kriging in the small case .....	209

## **LIST OF TABLES**

Table 5.1 Sample data statistics for each test case.....	79
Table 5.2 Nonlinear least squares fitting parameter values .....	94
Table 5.3 Interpolation grid specifications for each test case.....	98
Table 5.4 Percentage of true data falling in 95th confidence interval .....	102
Table 6.1 Interpolation grid definition.....	124

# 1 INTRODUCTION

The *National Research Council* (NRC) has estimated that 300,000 to 400,000 sites in the United States have contaminated groundwater (1997). The estimated cost of remediating these sites ranges from \$480 billion to \$1 trillion, or an average cost of \$8,000 per household in the United States (NRC 1993, 1997). Early legislative efforts leading to the Resource Conservation and Recovery Act (RCRA) and the Comprehensive Environmental Response, Compensation, and Liability Act (CERCLA) focused on restoring these resources to their natural states. Cost and technology limitations have since resulted in a shift in the design paradigm for groundwater remediation from resource recovery to long-term risk management.

Risk-based Corrective Action (RBCA) is a direct result of this shifting emphasis [*American Society for Testing and Materials (ASTM)* 1995]. RBCA uses a tiered approach to remediation, where the cost and level of remediation efforts are controlled by the human health and ecological risks posed by the contaminated resources. The increasing use of RBCA is expected to result in more contaminants being left in place that will require long-term monitoring [NRC, 1999]. Long-term monitoring (LTM) is particularly important for monitored natural attenuation, in which contaminants are mitigated by the natural processes of dilution, dispersion, and degradation. LTM at many sites can require decades of expensive sampling at tens or even hundreds of existing monitoring wells, resulting in hundreds of thousands or millions of dollars for sampling and data management per year.

Developing efficient and effective LTM sampling plans can be difficult when numerous options exist. In any given monitoring period, the number of possible sampling plans is  $2^n$ , where  $n$  is the product of the number of wells and the number of possible constituents to be measured. A site with 10 wells where up to 3 constituents can be measured ( $n = 30$ ) requires the decision

makers to identify a sampling plan among the more than one billion that exist while also trying to balance cost and other performance objectives for each sampling scheme.

## 1.1 Objectives and Scope

In developing long-term monitoring plans, regulators and stakeholders must negotiate monitoring objectives and decision parameters while accounting for technical, social, and regulatory considerations. The primary objective of this research is to develop a highly adaptable multiobjective LTM design methodology that aids this negotiation process by enabling decision makers to discover, understand, and balance tradeoffs among a variety of performance objectives. Three steps are required in the proposed methodology:

- Step 1: Selection and understanding of performance criteria
- Step 2: Selection of a plume interpolation method
- Step 3: Quantify design tradeoffs using multiobjective genetic algorithms (GAs).

The subsequent chapters of this dissertation address the specific challenges practitioners face when completing steps 1 thru 3 of the LTM design methodology.

## 1.2 Summary of Research

This thesis is the first monitoring research (see Chapter 2) that successfully addresses the two most challenging problems that face monitoring network designers: (1) selecting monitoring objectives and (2) balancing these objectives. These challenges were addressed using the 3-step problem decomposition discussed above in Section 1.1. Problem decomposition guided the selection and development of tools that were used in a multiobjective optimization framework. The optimization framework serves as an interface between the monitoring system being designed and the human decision process. The optimization framework requires effective plume

interpolation for evaluating LTM designs and the efficient use of multiobjective GAs for quantifying LTM objective tradeoffs.

Chapter 5 of this thesis is the first groundwater plume interpolation research to directly illustrate how preferential sampling and the highly skewed nature of groundwater contamination can combine to severely bias performance rankings of interpolation methods. Quantile kriging was the most robust of 6 groundwater plume interpolation methods, showing the least bias from both preferential sampling and the variability of contaminant data. The findings of Chapter 5 warrant further studies into the applicability of quantile kriging to data sets from other fields ranging from mining to life sciences where nonstationary interpolation also plays a vital role. Chapters 3 and 4 develop the first design methodologies for using evolution-based strategies to efficiently solve a new class of high order multiobjective applications (i.e., applications with more than 2 objectives). These methodologies are then used in Chapter 6 to solve the first application of evolutionary multiobjective optimization algorithms to a real-world problem with 4 objectives. The optimization framework developed in this chapter combines quantile kriging with high order multiobjective optimization to select, understand, and balance LTM performance criteria en route to a final negotiated design.

This thesis demonstrates that combining higher order Pareto optimization with visualization can allow designers in any field to assess the mathematical models used to represent their objectives, discover how their objectives are affecting designs, and negotiate a final design that balances their conflicting design preferences. The methods developed in this thesis are powerful tools for enhancing the design of LTM systems. Sections 1.2.1 thru 1.2.5 summarize each of the individual chapters of this thesis in more detail below.

### **1.2.1 Chapter 2: Literature Review**

Chapter 2 summarizes the extensive previous work in groundwater monitoring network design. Previous studies have primarily focused on two problems: (1) the use of geostatistics to augment or design monitoring networks for site characterization (for a review, see *ASCE Task Committee on Geostatistical Techniques* 1990b) and (2) the use of optimization and numerical simulation for screening monitoring plans for plume detection at landfills and hazardous waste sites (for a review, see *Loaiciga et al.* 1992). Recently, a third problem has emerged that seeks to reduce spatial and temporal redundancies in pre-existing well networks for sites undergoing long term monitoring. The LTM design methodology proposed in this dissertation combines elements of the geostatistical characterization approaches with spatial redundancy analysis to balance sampling costs, uncertainty, the quality of plume maps, and the accuracy of contaminant mass estimates (see Chapter 6 for more details).

### **1.2.2 Chapter 3: Optimization in Pareto Space**

Chapter 3 provides practitioners with a design methodology for the Nondominated Sorted Genetic Algorithm (NSGA). This portion of the research represents an extension of the simple GA design methodology presented by *Reed et al.* (2000b) to computationally intensive, multiobjective water resources applications. The NSGA design methodology is demonstrated using an LTM application, in which the tradeoffs between sampling costs and local concentration estimation errors in an existing groundwater monitoring network were quantified. This chapter shows that with proper design and parameterization, the NSGA is able to accurately quantify 2 dimensional tradeoffs.

### **1.2.3 Chapter 4: Simplifying Optimization in Pareto Space**

Chapter 4 extends the design methodology presented in Chapter 3 to the Nondominated Sorted Genetic Algorithm-II (NSGA-II). NSGA-II is a second generation evolutionary multiobjective (EMO) genetic algorithm that significantly improves upon the original NSGA. NSGA-II improves upon the NSGA [see *Deb et al.* 2000] by (1) invoking a more efficient nondomination sorting algorithm, (2) eliminating the sharing parameter, and (3) adding an implicitly elitist selection method that greatly aids in capturing high order Pareto surfaces. Chapter 4 builds on the NSGA design methodology of Chapter 3 and *Lobo* (2000) to introduce a multi-population approach that automates parameter specification for the NSGA-II and significantly reduces the computational costs associated with solving LTM applications. The methodology successfully solved the same LTM application as was solved in Chapter 3 using 80 percent fewer function evaluations (i.e., sampling design evaluations). The combined efficiency of the NSGA-II and design methodology presented in this chapter allows for more challenging higher order Pareto optimization problems (i.e., problems with more than 2 objectives) to be solved [see Chapter 6].

### **1.2.4 Chapter 5: Spatial Interpolation Methods for Plume Data**

Plume interpolation consists of estimating contaminant concentrations at unsampled locations using the available contaminant data surrounding those locations. The goal of groundwater plume interpolation is to maximize the accuracy in estimating the spatial distribution of the contaminant plume given the data limitations associated with sparse monitoring networks with irregular geometries. Beyond data limitations, contaminant plume interpolation is a difficult task because contaminant concentration fields are highly heterogeneous, anisotropic, and nonstationary phenomena. This chapter provides a

comprehensive performance analysis of 6 interpolation methods for scatter-point concentration data, ranging in complexity from intrinsic kriging based on intrinsic random function theory to a traditional implementation of inverse-distance weighting. High resolution simulation data of perchloroethylene (PCE) contamination in a highly heterogeneous alluvial aquifer were used to generate 3 test cases, which show how each interpolation method performs as a function of the amount of available sample data. Overall, the variability of PCE samples and preferential sampling in the source area controlled how well each of the interpolation schemes performed. Quantile kriging was the most robust of the interpolation methods, showing the least bias from both of these factors. Additionally, the method's non-parametric uncertainty estimates successfully predicted zones of high estimation error for each test case. This chapter provides guidance to practitioners balancing opposing theoretical perspectives, ease-of-implementation, and effectiveness when choosing a plume interpolation method.

### ***1.2.5 Chapter 6: Balancing Performance Criteria***

This chapter integrates the tools developed in the previous two chapters into a multiobjective optimization framework that can serve as an interface between the physical system being designed and the human decision process. This chapter demonstrates the use of high order Pareto optimization (i.e., optimizing a system for more than 2 objectives) in a highly adaptable optimization methodology. The methodology is implemented on an LTM application that combines quantile kriging [see Chapter 5] and the NSGA-II [see Chapter 4] to successfully balance four objectives: (1) minimizing sampling costs, (2) maximizing the quality of interpolated plume maps, (3) maximizing the relative accuracy of contaminant mass estimates, and (4) minimizing local estimation uncertainty. Optimizing the LTM application with respect to these objectives reduced the decision space of the problem from a total of 500 million designs to

the set of 1156 designs identified on the Pareto surface. Visualization of a total of 8 designs aided in understanding and balancing the application's objectives en route to a single compromise solution. This study shows that high order Pareto optimization holds significant potential as a tool that can be used in the balanced design of water resources systems.

<b>1</b>	<b>INTRODUCTION.....</b>	<b>1</b>
<b>1.1</b>	<b>Objectives and Scope .....</b>	<b>2</b>
<b>1.2</b>	<b>Summary of Research.....</b>	<b>2</b>
<b>1.2.1</b>	<b><i>Chapter 2: Literature Review .....</i></b>	<b>4</b>
<b>1.2.2</b>	<b><i>Chapter 3: Optimization in Pareto Space.....</i></b>	<b>4</b>
<b>1.2.3</b>	<b><i>Chapter 4: Simplifying Optimization in Pareto Space .....</i></b>	<b>5</b>
<b>1.2.4</b>	<b><i>Chapter 5: Spatial Interpolation Methods for Plume Data.....</i></b>	<b>5</b>
<b>1.2.5</b>	<b><i>Chapter 6: Balancing Performance Criteria .....</i></b>	<b>6</b>

## 2 LITERATURE REVIEW

Monitoring network design has been studied extensively in the past, but previous studies have primarily focused on two problems: (1) the use of geostatistics to augment or design monitoring networks for site characterization (for a review, see *ASCE Task Committee on Geostatistical Techniques* 1990b) and (2) the use of optimization and numerical simulation for screening monitoring plans for plume detection at landfills and hazardous waste sites (for a review, see *Loaiciga et al.* 1992). Recently, a third problem has emerged that seeks to reduce spatial and temporal redundancies in pre-existing well networks for sites undergoing long term monitoring. Previous work addressing each of these problems is described in sections 2.1 – 2.3.

### 2.1 Geostatistical Approaches

The first approach to monitoring network design utilizes geostatistical estimation procedures (kriging) to evaluate alternative sampling schemes. Geostatistics is a stochastic approach that provides minimum error estimates of a phenomenon at unsampled locations using linear combinations of samples values (for an introduction see *Journel & Huijbregts* 1978, *Kitanidis* 1997, *Goovaerts* 1997, *Chilès & Delfiner* 1999). In addition to providing the expected value of a phenomenon, the estimation variance is computed that represents the uncertainty of estimates at unsampled locations. Estimation variances can be computed both *locally* and *globally*. A *local* estimation variance accounts for the uncertainty of an estimate at a single unsampled location, whereas a *global* estimation variance represents the uncertainty accumulated over several unsampled locations within a subset of the interpolation domain. Estimation variances are independent of the sampled data and vary only as a function of the monitoring well network's geometry, enabling monitoring network designs to be evaluated for their ability to characterize a site before any sampled data are available. The previous monitoring design

methodologies that have used geostatistics are classified as being either local or global approaches depending on the form of the estimation variance used.

### ***2.1.1 Minimization of Local Estimation Variance***

The local approach to monitoring design for site characterization uses geostatistics to augment monitoring networks by analyzing the effectiveness of adding sampling points to the network (*ASCE Task Committee on Geostatistical Techniques* 1990b). Additional sampling points are added based on an analysis of which locations will maximally decrease the estimation variance attained in geostatistical interpolation (*Rouhani* 1985). *Rouhani & Hall* (1988) build on this localized approach by modifying the objective to also include the magnitude of contaminant concentrations within the modeled domain. Both of these approaches design monitoring networks considering only spatial uncertainties. Extensions of the local approach have incorporated stochastic fate-and-transport modeling into estimation methodologies enabling estimation or error variances to be analyzed in both space and time (*Graham and McLaughlin* 1989a, b; *Herrera de Olivares* 1998, *Rizzo et al.* 2000). These approaches are based on Kalman filtering, which is theoretically equivalent to geostatistical estimation methods using advanced forms of kriging (see *Herrera de Olivares* 1998).

### ***2.1.2 Minimization of Global Estimation Variance***

The global approach to site characterization uses geostatistics to evaluate the performance of potential monitoring network designs still in their planning stage (*ASCE Task Committee on Geostatistical Techniques* 1990a,b). These studies consider several spatial configurations and densities of monitoring wells, each of which are evaluated using the global estimation variance for each potential monitoring network (*Olea*, 1984; *Yfantis et al.*, 1987; and *Christakos & Olea*, 1988). These methods can exactly compute a unique estimation variance for

well clusters with regular geometries and rapidly evaluate the expectation of the variance for random configurations (*ASCE Task Committee on Geostatistical Techniques* 1990b). The global estimation variance is minimized within these studies to improve the accuracy of potential network designs. The global approaches have primarily been used in the design of networks that do not yet exist or to select additional sampling points to augment pre-existing monitoring networks.

### ***2.1.3 Minimization of Modeling Uncertainty***

The local and global estimation variance methods assume that the estimation uncertainty is *representative* of both the sample data and the underlying spatially variable phenomenon being studied. Although the estimation variance computations are independent from the actual values of the sample data, the uncertainty estimates based on these computations are heavily dependent on the underlying covariance (or spatial structure) of the phenomenon being studied. In geostatistics, the spatial structures of studied phenomena are computed in a manner similar to traditional statistics where the theoretical second moment is estimated with a discrete approximation using sampled data. Geostatistical estimation methods require that a continuous, positive-definite model be fit to the discrete approximation of the covariance to represent the “true” underlying structure of the phenomenon of interest.

Several site characterization methodologies have directly addressed the *representative* assumption implicit to the estimation variance methods by focusing on the design of monitoring networks to minimize the error associated with modeling the “true” underlying covariance structure of a spatially variable phenomenon. Early studies analyzed the effect of sampling network geometry on the accuracy of covariance structure estimates (*Russo 1984, Warrick & Myers 1987, Russo & Jury 1988*). Later works considered multiobjective approaches that

balanced cost considerations with the model estimation error (*Knopman & Voss* 1989, *Wagner* 1995). *Bogaert & Russo* (1999) extend the earlier studies by considering both the network geometry and the covariance model in a least squares approach to minimizing the spatial structure modeling errors. Many of these approaches require the assumption that the spatially variable phenomenon of interest is second order stationary, which restricts their applicability for designing monitoring networks for sites with groundwater contamination.

## 2.2 Hazardous Waste Detection Monitoring

The second monitoring network design problem that has been studied in the literature involves screening monitoring plans for plume detection at landfills and hazardous waste sites, primarily using optimization and/or numerical simulation. Much of the research in this area builds on the work of *Massmann and Freeze* (1987a,b), who present a guide for landfill monitoring design that uses a conditional simulation approach to calculate the probability of detection for a particular monitoring network. *Meyer and Brill* (1988) built upon this work by combining Monte Carlo simulation with an optimization model, which was formulated as a maximal covering location problem. Analytical optimization formulations have since been used in several studies to identify monitoring network designs at landfill sites that maximize the likelihood of intercepting plumes from leaking landfill leachate (*Hsueh and Rajagopal* 1988, *Loaiciga* 1989, *Hudak and Loaiciga* 1993).

*Hudak and Loaiciga* [1992] present a heuristic approach using facility location theory to augment pre-existing monitoring networks. The approach was proposed as an alternative to the geostatistical methods described above. More recent studies have incorporated uncertainty and/or multi-objective optimization to design cost-effective monitoring networks that maximize the probability of plume detection at landfills and hazardous waste sites undergoing remediation

(Meyer *et al.* 1994, James and Gorelick 1994, Cieniawski *et al.* 1995, and Storck *et al.* 1997).

Montas *et al.* (2000) combined stochastic fate-and-transport and partial enumeration to estimate the spatial moments characterizing a contaminant plume subject to cost constraints.

### 2.3 Spatial and Temporal Redundancy Analysis

Redundancy analysis of pre-existing LTM networks has emerged as a third important research area in the optimal monitoring design literature. Redundancies are defined as data that contribute minimally to estimation of the plume's extent in space or time. As the number of contaminated sites requiring long-term monitoring has increased over the previous decade, sampling-and-analysis costs have been recognized as significant contributors to the overall cost of remediating sites with groundwater contamination. Evidence of the emerging importance of redundancy analysis can be seen in the Federal Remediation Technologies Roundtable's (a multi-agency consortium that disseminates important innovations in the environmental field) focus on tools that directly account for redundancies in monitoring data to reduce sampling costs (see <http://www.frtr.gov/optimization/monitoring/>).

Johnson *et al.* (1996) present a method that uses a trial-and-error technique for analysis of time series at individual monitoring wells, which was able to identify temporal redundancies in the sampling data at the Savannah River Site. Reed *et al.* (2000a) combined numerical simulation, ordinary kriging, and a simple GA to reduce sampling costs by identifying spatial redundancies at the Hill Air Force Base site in Utah. The methodology showed that sampling costs could be reduced by as much as 60 percent while maintaining accurate global contaminant mass estimates. Software packages are available that combine interpolation methods such as kriging or Thiessen polygons in combination with traditional statistical regression analysis to allow practitioners to identify spatial and temporal redundancies in monitoring networks

(Cameron & Hunter 2000, Aziz *et al.* 2000). None of these methods, however, are capable of quantifying the inherent tradeoffs in LTM design, which is the focus of this thesis.

<b>2 LITERATURE REVIEW.....</b>	<b>8</b>
<b>2.1 Geostatistical Approaches .....</b>	<b>8</b>
<b>2.1.1     <i>Minimization of Local Estimation Variance.....</i></b>	<b>9</b>
<b>2.1.2     <i>Minimization of Global Estimation Variance.....</i></b>	<b>9</b>
<b>2.1.3     <i>Minimization of Modeling Uncertainty.....</i></b>	<b>10</b>
<b>2.2 Hazardous Waste Detection Monitoring.....</b>	<b>11</b>
<b>2.3 Spatial and Temporal Redundancy Analysis .....</b>	<b>12</b>

## 3 OPTIMIZATION IN PARETO SPACE

### 3.1 Introduction

Chapters 1 and 2 showed that LTM design often requires the exploration of vast decision spaces while balancing conflicting design objectives (e.g., reducing sampling costs while maintaining an acceptable picture of the groundwater contamination plume). The size and complexity of these decision spaces motivated the use of multiobjective GAs, primarily in this thesis because their population-based approach to multiobjective search and optimization has been shown to be more efficient than other more traditional optimization techniques (see *Deb* 2001). The purpose of this chapter is to demonstrate that multiobjective genetic algorithms are capable of navigating LTM decision spaces while accurately quantifying design tradeoffs. Moreover, this chapter demonstrates how theoretical work from the genetic and evolutionary field can be used to improve real-world applications of multiobjective GAs by following the recommendations of *Goldberg* (1998).

*Goldberg* (1998) compels GA users to seek “...GAs that solve hard problems quickly, reliably, and accurately—through a combination of effective (1) *design methodology*, (2) *design theory*, and (3) *design*.” *Goldberg* (1998) also describes the *control map* concept for GAs where theoretical relationships for population sizing, selection pressure, crossover, and mutation are used to find values for the control parameters that identify a performance “sweet spot” where efficient search and optimization is enabled. The *design theory* for efficient GAs and the concept of the control map for GA performance are largely derived from *Goldberg et al.* (1992a) and *Thierens* (1995), respectively. Although these studies provided a theoretical basis for the design of GAs, practitioners still face the difficulty of using relationships that require the specification of parameters that are not readily identifiable in practice such as building block (BB) size and

order. BB's blocks are highly relevant subsets of the binary digits representing designs and are used by the GA to construct optimal solutions.

This difficulty was first addressed by the simple GA *design methodology* presented by *Reed et al.* (2000b) for single objective applications. This chapter extends *Reed et al.* (2000b) to multiobjective applications, presenting a *design methodology* for the Nondominated Sorted Genetic Algorithm (NSGA) that enabled the algorithm to accurately quantify 2 dimensional tradeoffs. This chapter has been excerpted from *Reed et al.* (2001a).

### **3.1.1 Previous Work**

Evolution-based multiobjective optimization (EMO) methods are pragmatic tools for solving problems with large decision spaces and conflicting objectives. *Schaffer* (1984) provided the seminal work within the EMO field in which a vector evaluated genetic algorithm (VEGA) was designed to search decision spaces for the optimal tradeoffs among a vector of objectives. Subsequent innovations in EMO have resulted in a rapidly growing field with a variety of solution methods that have been used successfully in a wide range of applications (for reviews see *Fonseca & Fleming* 1995, *Coello* 1999, *Van Veldhuizen* 1999). These solution methods have garnered increased attention over the past decade and have been applied in a variety of contexts within the water resources field.

*Cieniawski* (1993) is one of the earliest studies in water resources to utilize EMO methods. The study is an empirical comparison of the performance of VEGA relative to niching-based techniques from *Goldberg & Richardson* (1987) for identifying a monitoring network to detect potential contaminant leaks from a hazardous waste landfill. *Cieniawski* (1993) and *Cieniawski et al.* (1995) clearly espouse the efficiency of EMO methods in quantifying tradeoffs between maximizing a groundwater-monitoring network's reliability in detecting contaminants

and minimizing the area of contaminated aquifer at the time of first detection. Ritzel *et al.* (1994) compared the performances of VEGA, a domination ranking-based genetic algorithm (Pareto GA), and mixed integer chance constrained programming (MICCP) in solving a multiobjective, groundwater pollution containment application. Halhal *et al.* (1997) successfully incorporated Pareto domination ranking into the messy genetic algorithm (Goldberg *et al.* 1989) to quantify the tradeoffs in rehabilitating water distribution networks. Gupta *et al.* (1998) combined a downhill simplex method with an evolutionary search strategy implementing Pareto ranking to seek tradeoff solutions when calibrating hydrologic models. These studies show how EMO methods have been adapted to solve a variety of water resource applications. This chapter focuses on the Nondominated Sorted Genetic Algorithm (NSGA) because Zitzler *et al.* (2000) showed that the NSGA performed as well or better than a representative sampling of EMO methods on a suite of test problems with properties similar to our application.

### **3.1.2 Motivation and Scope**

One of the difficulties in applying EMO methods is identifying parameter settings that ensure efficient navigation of the decision space and adequate coverage of the Pareto frontier (*Van Veldhuizen & Lamont* 2000, *Cieniawski* 1993). Most practitioners use trial-and-error runs to identify the best parameter settings, but this approach is quite time consuming, particularly for applications with computationally intensive fitness functions. A major goal of this chapter is to develop guidelines for using theoretical relationships from the genetic and evolutionary computation literature to ensure that the NSGA efficiently navigates the problem's decision space. These guidelines are then applied to quantify the tradeoffs implicit in designing sampling strategies for a long-term groundwater-monitoring network. Additionally, a niching-based elitist

enhancement of the NSGA is also presented, which substantially improves coverage of the Pareto frontier.

### 3.2 Monitoring Application

The test case developed in this chapter uses data drawn from a 38 million-node flow-and-transport simulation performed by *Maxwell et al.* (2000). The simulation provided realistic historical data for the migration of a hypothetical perchloroethylene (PCE) plume in a highly heterogeneous alluvial aquifer.

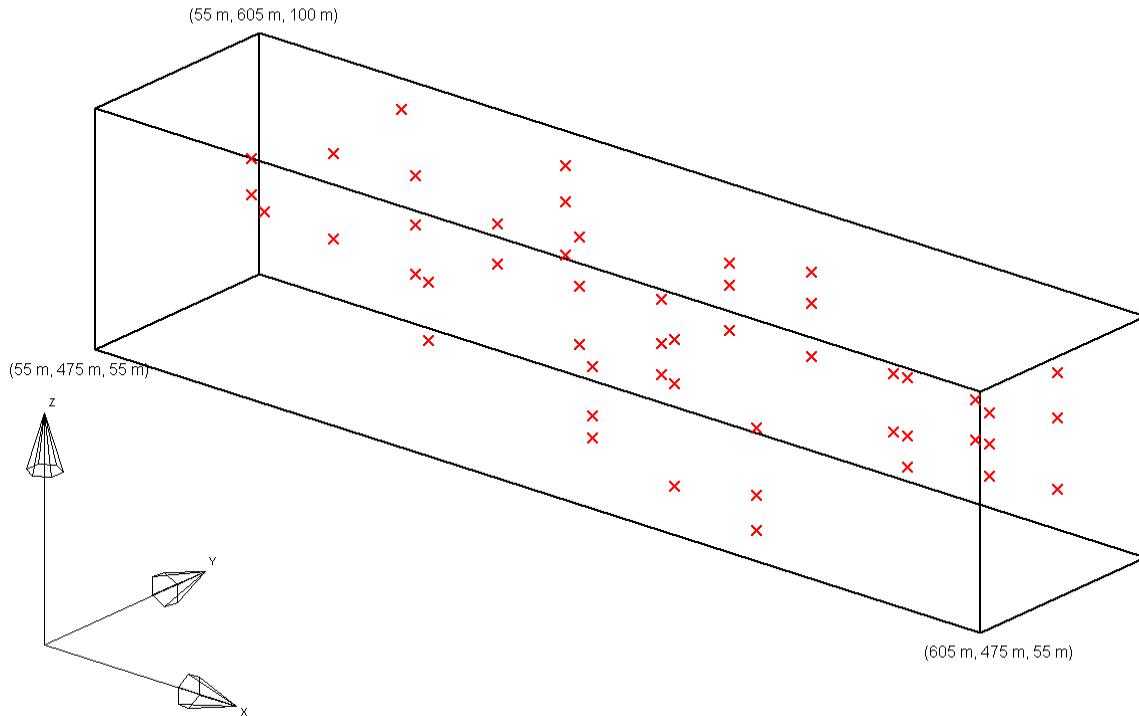


Figure 3.1 The 50 potential sampling locations (designated by the x's above) within a 20 well multi-level monitoring network

The hydrogeology of the test case is based on an actual site located at the Lawrence Livermore National Laboratory in Livermore, California, currently being managed under the United States' Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) program.

Data were provided for a total of 50 hypothetical sampling locations within the 20-well multi-level monitoring network shown in Figure (3.1). The data represent a snapshot in time, 10 years after a continuous point source began contaminating the aquifer system. The monitoring wells can sample from 1 to 3 locations along their vertical axis and have a minimum spacing of 60-m between wells in the horizontal plane.

The site is assumed to be undergoing long-term monitoring, in which groundwater samples are used to assess the effectiveness of current remediation strategies. During this long-term monitoring phase of a remediation, sampling and laboratory analysis can be a controlling factor in the costs of remediating a site. Quarterly sampling of the entire network in Figure (3.1) has a potential cost of over \$70,000 annually for PCE testing alone, which could translate into millions of dollars if the site had a typical life span of 20 to 30 years.

The significance of these costs has motivated the development of several approaches for reducing the fiscal burden posed by long term monitoring by identifying redundant wells in groundwater monitoring networks that can be omitted from future sampling periods (*Cameron & Hunter 2000, Aziz et al. 2000, Reed et al. 2000a, Rizzo et al. 2000*). These methods define sampling points to be redundant when they minimally affect interpolation-based plume estimates. They employ a variety of single objective optimization techniques ranging from a simple genetic algorithm to trial-and-error heuristics. The objective of these methods is to minimize sampling costs while incorporating performance objectives associated with plume estimates as constraints. The management model presented in this work builds on these previous methods by introducing a sampling design methodology that explicitly identifies the tradeoffs encountered when reducing monitoring costs.

### 3.2.1 Problem Formulation

To identify which wells are redundant, this chapter employs a local concentration approach with the intention of attaining the best-interpolated picture of the PCE plume for the least cost. Equation (3.1) gives the multiobjective problem formulation for quantifying the tradeoff between sampling costs and maintenance of a high quality interpolated picture of the plume.

$$\text{Minimize } F(\bar{x}_\kappa) = [f_1(\bar{x}_\kappa), f_2(\bar{x}_\kappa)], \quad \forall \kappa \in \Omega$$

$$f_1(\bar{x}_\kappa) = \sum_{i=1}^{n_{well}} C_S(i)x_{\kappa i}$$

$$f_2(\bar{x}_\kappa) = \sum_{j=1}^{n_{est}} \left( c_{all}^*(\bar{u}_j) - c_{est}^\kappa(\bar{u}_j) \right)^2 \quad (3.1)$$

$F(\bar{x}_\kappa)$  is a vector valued objective function whose components  $[f_1(\bar{x}_\kappa), f_2(\bar{x}_\kappa)]$  represent the cost and squared relative estimation error (SREE), respectively, for the  $\kappa^{\text{th}}$  monitoring scheme  $\bar{x}_\kappa$  taken from the collection of all possible sampling designs  $\Omega$ . Equation (3.2) defines the binary decision variables representing the  $\kappa^{\text{th}}$  monitoring scheme.

$$x_{\kappa i} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ well is sampled} \\ 0, & \text{otherwise} \end{cases}, \quad \forall \kappa, i \quad (3.2)$$

If the  $i^{\text{th}}$  well is sampled it is assumed that all available locations along the vertical axis of that well will be sampled at a cost of  $C_S(i)$ .  $C_S(i)$  ranged from \$365 to \$1095 for 1 to 3 samples analyzed for PCE solely (Rast 1997). Sampling all available levels within each well reduces the size of  $\Omega$  from  $2^{50}$  to  $2^{20}$  where 50 and 20 represent the total number of sampling locations and monitoring wells ( $n_{well}$ ), respectively. Reducing the size of  $\Omega$  enabled the entire

decision space of this application to be enumerated. Enumeration was employed to identify the true Pareto frontier so that the performance of the NSGA under different parameter settings could be rigorously tested.

The SREE provides a measure of how the interpolated picture of the plume using data only from wells included in the  $\kappa^{\text{th}}$  sampling plan compares to the result attained using data from all available sampling locations. The measure is computed by summing the squared deviations between the PCE estimates using data from all available sampling locations,  $c_{\text{all}}^*(\bar{u}_j)$ , and the estimates based on the  $\kappa^{\text{th}}$  sampling plan  $c_{\text{est}}^\kappa(\bar{u}_j)$  at each location  $\bar{u}_j$  in the interpolation domain. Each  $\bar{u}_j$  specifies the coordinates for the  $j^{\text{th}}$  grid point in the interpolation domain. The interpolation domain consisted of a total of 3300 grid points ( $n_{\text{est}}$  in equation (3.1)). The PCE estimates used in the calculation of the SREE for each of the sampling designs were attained using the nonlinear spatial interpolation method described below.

### **3.2.2 Nonlinear Spatial Interpolation**

The interpolation method used in this chapter is a variant of the scheme used by *Barry and Sposito* (1990) in their analysis of tracer plumes at the Borden site located in Ontario, Canada. The *Barry and Sposito* (1990) interpolation method was selected because it requires minimal modeling assumptions and it has been successfully applied to three-dimensional historical data. *Barry and Sposito* (1990) interpolated the Borden tracer data using 7 fitting parameters with nonlinear least squares. This chapter simplified the interpolation scheme to use 3 fitting parameters in a nonlinear least squares version of inverse distance weighting. Neglecting 4 of the 7 fitting parameters did not appreciably affect the cross-validation residuals and improved the algorithm's stability. Additionally, using fewer fitting parameters reduced the size of the

Jacobian matrix and the number of iterations required to evaluate each sampling design, enabling  $\Omega$  to be enumerated.

Equation (3.3) shows the interpolation function used in this chapter, which estimates of PCE concentration  $c_{est}^\kappa(\bar{u}_j)$  at each unknown location  $\bar{u}_j$  by the weighted sum of the  $nsamp(\kappa)$  total samples  $c(\bar{u}_\psi)$  taken in the  $\kappa^{\text{th}}$  sampling scheme.

$$c_{est}^\kappa(\bar{u}_j) = \sum_{\psi=1}^{nsamp(\kappa)} w(\bar{u}_j, \bar{u}_\psi) c(\bar{u}_\psi) / w_t \quad (3.3)$$

The weights for each of the samples in the  $\kappa^{\text{th}}$  sampling plan are calculated as a function of the distance between the  $\psi^{\text{th}}$  sample and the  $j^{\text{th}}$  grid point in the interpolation domain as shown in equation (3.4).

$$w(\bar{u}_j, \bar{u}_\psi) = 1/D_\kappa(\bar{u}_j, \bar{u}_\psi)$$

$$D_\kappa(\bar{u}_j, \bar{u}_\psi) = \left[ (u_{j1} - u_{\psi1})^2 / \alpha_1 + (u_{j2} - u_{\psi2})^2 / \alpha_2 + (u_{j3} - u_{\psi3})^2 / \alpha_3 \right]^{0.5} \quad (3.4)$$

The  $w_t$  factor is the sum of the  $nsamp(\kappa)$  weights calculated using equation (3.4) and serves to scale the system such that the weights sum to one. The parameters  $\alpha_1$ ,  $\alpha_2$ , and  $\alpha_3$  are fitting parameters that were used to minimize the cross-validation residuals for each of the sampling designs considered in this chapter.

Equation (3.5) gives the cross-validation estimation method used in this chapter to fit the  $\alpha$ -fitting parameters shown in equation (3.4) to the data provided by each sampling plan (*Barry and Sposito 1990*).

$$\text{Minimize } S(\kappa) = \sum_{\psi=1}^{nsamp(\kappa)} \left\{ c(\bar{u}_\psi) - \sum_{h=1, h \neq \psi}^{nsamp(\kappa)} w(\bar{u}_\psi, \bar{u}_h) c(\bar{u}_h) / (w_t - w(\bar{u}_\psi, \bar{u}_h)) \right\}^2 \quad (3.5)$$

The cross-validation method minimizes the sum of the squared residuals  $S(\kappa)$  between each of the actual PCE concentrations  $c(\bar{u}_\nu)$  and their interpolated estimates based on the  $(nsamp(\kappa) - 1)$  remaining data  $c(\bar{u}_h)$  in the  $\kappa^{\text{th}}$  sampling plan. Equation (3.5) was solved for every sampling plan considered in this chapter using the Levenberg-Marquardt nonlinear least squares solution method (Moré 1977). The Levenberg-Marquardt solution technique implicitly requires that the number of data points in the system must exceed the number of fitting parameters. This requirement necessitated that the monitoring designs considered in this chapter have a minimum of four PCE samples.

### 3.3 The Basics of the NSGA

The NSGA utilizes the Darwinian process of natural selection to effectively search for solutions that are optimal across a vector of objectives. The algorithm is very similar in form to the simple genetic algorithm (sGA), in that it exploits the operators of selection, crossover, and mutation when building a set of optimal solutions. The performance of both algorithms can be described using the building block (BB) theory presented by *Holland* (1975) and *Goldberg* (1989). For both the NSGA and sGA, highly fit designs have a higher probability of being selected to mate and pass their traits (or BBs) to succeeding generations. Stochastic remainder selection was used in this chapter as recommended in *Srinivas & Deb* (1995). The operators of crossover and mutation are identical for the two algorithms. Crossover (occurring with probability  $p_c$ ) exerts an innovative force on the system by allowing favorable traits from parent designs to be juxtaposed in offspring that possess higher relative fitness values (*Thierens & Goldberg* 1993). Mutation locally refines solutions by randomly changing bit values (with a probability of  $p_m$ ) from 0 to 1 or vice versa within a design's genotype. The difference between the NSGA and the sGA lies in how fitness is assigned. Unlike the sGA, the NSGA evaluates

sampling designs in terms of a vector of objectives. A sampling design cannot be assessed in terms of its performance in any single objective because it may perform poorly with respect to the remaining objectives. The NSGA employs the concepts of *Pareto dominance* and *niching* to assign fitness values to sampling designs in two steps described below (*Srinivas & Deb* 1995).

### 3.3.1 Pareto Dominance

The first step in fitness assignment employs the *Pareto dominance* concept defined in equation (3.6), using notation adapted from *Van Veldhuizen & Lamont* (2000).

$$F(\bar{x}) \prec F(\bar{x}'), \text{ iff } (\forall \beta, f_\beta(\bar{x}) \leq f_\beta(\bar{x}') \wedge \exists \beta \in (1, n_{obj}) : f_\beta(\bar{x}) < f_\beta(\bar{x}')) \quad (3.6)$$

Equation (3.6) states that a design  $\bar{x}$  dominates another design  $\bar{x}'$  (represented by  $F(\bar{x}) \prec F(\bar{x}')$ ) if and only if it performs as well as  $\bar{x}'$  in all  $n_{obj}$  objectives and better in at least one (assuming minimization of all objectives). The NSGA identifies nondominated individual designs within the current population and assigns an arbitrary dummy fitness value to each of them. These individuals are all initially assigned the same fitness value to ensure that they have an equal likelihood of being selected to pass their traits unto latter generations; these designs compose the first nondominated front.

### 3.3.2 Niching & Sharing

The second step in fitness assignments for the individuals in the first front utilizes the concept of *niching* (for more details see *Goldberg & Richardson* 1987, *Deb & Goldberg* 1989, *Mahfoud* 1995, *Horn* 1997) to ensure that the NSGA finds a diverse set of solutions defining the entire extent of the tradeoff (or Pareto front) among the objectives. Extending the “natural selection” analogy to include the phenomena of *niching* and *speciation* helps to satisfy this goal. *Horn* (1997) defines niching as a “form of cooperation” where there is “... the localization of competition around finite, limited resources (niches), resulting in the lack of competition

between such areas, and causing the formation of species for each niche.” Niching allows the NSGA to form stable subpopulations of sampling designs (species) each of which are well adapted to search for nondominated solutions specific to subspaces (niches) in  $\Omega$ . The NSGA elicits niche formation by treating dummy fitness as a limited resource, which is shared by sampling designs using the relationship given in equation (3.7) (*Goldberg & Richardson 1987, Srinivas & Deb 1995*).

$$Sh[d(\kappa, \kappa')] = \begin{cases} 1 - \left( \frac{d(\kappa, \kappa')}{\sigma_{share}} \right)^2, & \text{if } d(\kappa, \kappa') < \sigma_{share} \\ 0, & \text{otherwise} \end{cases} \quad (3.7)$$

The value of the fitness sharing function  $Sh[d(\kappa, \kappa')]$  ranges between 0 and 1 for two sampling designs  $\kappa$  and  $\kappa'$  depending on the ratio of their distance from one another  $d(\kappa, \kappa')$  and the niche size defined by the niche radius  $\sigma_{share}$ . The sharing function measures the similarity between designs ranging from completely dissimilar ( $Sh = 0$ ) to being identical ( $Sh = 1$ ). The distance term  $d(\kappa, \kappa')$  can measure similarity in several ways that are discussed in the next section.

The third step in assigning the fitness values to the nondominated set identified in step 1 consists of dividing each individual’s dummy fitness by the sum of all of its sharing function values (termed the niche count) as shown in equation (3.8).

$$\text{niche count}(\kappa) = \sum_{\kappa'=1}^N Sh[d(\kappa, \kappa')], \forall \kappa \quad (3.8)$$

When these three steps are completed, the individuals in the first front are removed from consideration and the remaining population members undergo nondomination ranking and sharing to yield the second nondominated front. These individuals are assigned the minimum

shared fitness value of the first front decremented by a value  $\Delta D_f$ . This process continues until the entire population is assigned to the  $n_f^{\text{th}}$  front and given a shared dummy fitness value. After shared dummy fitness values have been assigned, the NSGA is operated the same as the sGA, constructing the Pareto front using the traditional operators of selection, crossover, and mutation. See *Goldberg* (1989) for more background on these operators.

## 3.4 Multiobjective Search & Optimization

### 3.4.1 Performance Considerations

Construction of the Pareto frontier requires that the NSGA be effectively designed to navigate a problem's decision space. Effective design of the NSGA entails careful consideration of the factors controlling the algorithm's performance. *Zitzler et al.* (2000) concluded that population sizing and elitism are the most influential factors influencing the performance of EMO methods. Elitist operators provide a means of ensuring that the best individuals are identified and allowed to pass their traits to latter generations. Unlike sGA applications, EMO methods cannot simply pass a single individual with the current best objective function value into the next generation. Multiobjective optimization requires that some fraction of the solutions along the current nondominated front be passed on to the next generation. The next four sections of this chapter describe an NSGA design methodology that integrates several relationships from the genetic and evolutionary literature to identify effective population sizes and determine the niching parameters.

This methodology is meant to serve as a guide for applying the NSGA to other applications. Additionally, an elitist strategy is presented with the aim of using niching to guide the selection of elite population members and maximize *online* performance (the term online performance means that only the nondominated individuals within a single generation are

considered when assessing the NSGA's performance). Performance is finally discussed in the context of selection pressure in the fourth and final section of the NSGA design methodology.

### 3.4.1.1 Population Sizing: Ensuring Genetic Diversity

The first step in designing an efficient NSGA is to perform initial problem analysis to determine a range of population sizes. The goal of this initial step is to provide a means of selecting the best population size for our application with respect to a single random seed. Computationally intensive objective functions that occur in many water resources applications require that effective parameters settings for the NSGA be identified using a minimum number of runs, because a single run can take days or even weeks. The range of population sizes considered in this chapter were attained using relationships from *Mahfoud* (1995), who derived population-sizing relationships for genetic algorithms employing niching to solve multimodal problems. The relationships used in this chapter were designed for applications with multiple optima that have *identical fitness values*. Recall, the NSGA employs niching using members of the same nondominated set with *identical dummy fitness values*, which makes the population-sizing relationships in *Mahfoud* (1995) relevant to NSGA applications.

Equation (3.9) gives population size estimates assuming that crossover will not disrupt the traits (or BBs) required to assemble optimal solutions.

$$N \geq q \left[ -\ln \left( \frac{1 - \gamma^{\frac{1}{g}}}{q} \right) \right] \quad (3.9)$$

The relationship was derived as the minimum population size  $N$  required to maintain  $q$  niches for  $g$  generations with reliability  $\gamma$  for the special case when the niches have identical fitness values. Disruption adversely affects the NSGA's performance when crossover between a pair of parent strings destroys traits that are essential to the evolution of nondominated individuals in

subsequent generations (for more details see *Mahfoud* 1995, *Zitzler et al.* 2000, *Van Veldhuizen & Lamont* 2000). Equation (3.10) accounts for the potential disruptive effects of crossover in population size estimates (*Mahfoud* 1995).

$$N \geq 2 \ln \left( \frac{1 - \gamma^{\frac{1}{g}}}{q} \right) \left/ \ln \left[ \left( 1 - p_c \right) \left( 1 - \frac{1}{q} \right)^2 + p_c \left( 1 - \frac{1}{q^2} \right) - 2 p_c (1 - p_d) \frac{1}{q} \left( 1 - \frac{1}{q} \right) \right] \right. \quad (3.10)$$

Equation (3.10) represents the minimum population size  $N$  required to maintain  $q$  niches for  $g$  generations with reliability  $\gamma$  given a probability of crossover  $p_c$  and probability of disruption  $p_d$ . The probability of disruption  $p_d$  can be conservatively estimated using equation (3.11), which assumes the maximally disruptive operator of uniform crossover. In uniform crossover, the binary values at each bit position in two individual strings is swapped with probability  $p_c$ .

$$p_d = 1 - \left( \frac{1}{2} \right)^{O(m)} \quad (3.11)$$

For the binary-coded NSGA, the decision variables ( $\bar{x}_\kappa$  in equation (3.1)) representing a potential design (termed its *phenotype*) are converted to binary representations and concatenated into a string of binary variables (termed the design's *genotype*) of length  $l$ . The traits or (BBs) are actually small subsets of binary digits from each design's genotype. Equation (3.11) computes the probability that the BBs will be disrupted as a function of their order  $O(m)$ , which is the smallest number of fixed value digits in a design's genotype that are relevant to the final solutions of the problem.

As was the case in the sGA design methodology presented by *Reed et al.* (2000b), the actual size of BBs in an engineering application,  $O(m)$  are unknown, requiring the practitioner to employ conservative assumptions and problem-specific information to calculate a potential range of population sizes from equations (3.9) and (3.10). The  $p_d$  was set by assuming that the BB

order  $O(m)$  may range from 1 to 5 because higher order BBs will be disrupted by the operators of selection, crossover, and mutation (*Goldberg* 1989), following a similar approach presented by *Reed et al.* (2000b) for designing sGAs.

The remaining parameters in equations (3.9) and (3.10) were set as follows. The population size calculations assumed a reliability of 85 percent ( $\gamma = 0.85$ ). The number of generations until convergence  $g$  can be estimated to fall within the range  $[2l, l\ln l]$  where  $l$  is the binary string length of each design (*Thierens et al.* 1998, *Thierens & Goldberg* 1994). *Mahfoud* (1995) states that equations (3.9) and (3.10) are not sensitive to the parameter value  $g$  and recommends that the user set the value of this parameter to be greater than or equal to the number of generations they expect to run the algorithm. For these reasons,  $g$  was set equal to  $2l$  (40 generations) for this chapter. A majority of EMO applications use a value for  $p_c$  falling within the range [0.6, 0.9] (*Fonseca & Fleming* 1995, *Coello* 1999, *Van Veldhuizen* 1999, *Horn* 1997). In this application,  $p_c$  was set equal to 0.6 to reduce the potential for disruption and the resulting population size estimates.

The number of niches  $q$  was set using information specific to the tradeoff being sought between cost and SREE in this application. The discussion of the spatial interpolation stated that the number of PCE samples  $c(\bar{u}_\psi)$  ranged between a minimum value of 4 and a maximum of 50, yielding a tradeoff with a maximum of 46 discreet cost levels. The goal of this application is to find the minimum SREE for each of these discreet costs levels. The number of niches was set equal to 40, which represents a goal of attaining over 85 percent of the points in the Pareto frontier.

Using these parameters in equations (3.9) – (3.11) yielded six population size estimates ranging from a minimum value of 370 to a maximum of 870. The lower bound population size

estimate of 370 assumes that crossover will not disrupt BBs. The remaining five population size estimates account for disruption of BBs with orders,  $O(m)$ , ranging from 1 to 5. Given this range of population sizes, the computational complexity of using the NSGA to solve this problem was estimated by multiplying  $N$  by the number of generations  $g$  to attain the total number of function evaluations. Function evaluations took an average time of 0.04 s on a Dell XPS T800r running Windows NT, yielding estimated total run times between 10 and 25 minutes.

Finally, it should be noted that both equations (3.9) and (3.10) assume that mutation is minimally disruptive. This assumption was accounted for by setting  $p_m$  to be equal to  $1/N$  (*DeJong* 1975, *Schaffer et al.* 1989), which limits the number of mutations to  $l$  and reduces the operator's influence as  $N$  increases. In this application, the optimal population was determined by running the NSGA for each of the six estimates and choosing the run that best defined the full extent of the tradeoff between cost and SREE.

#### *3.4.1.2 Sharing: Sizing of the Niche*

An important component of ensuring that the NSGA converges to the Pareto frontier is properly setting the size of niches represented by  $\sigma_{share}$  in equation (3.7). *Deb & Goldberg* (1989) provide guidance for setting this parameter in both phenotypic and genotypic space. Recall that a sampling design's phenotype is the floating-point representation of each of its decision variables. Additionally, the design's genotype is its concatenated binary representation (or its chromosome). Equation (3.7) requires that the distance  $d(\kappa, \kappa')$  between 2 individuals be compared to  $\sigma_{share}$  in order to compute the sharing function values used to assign fitness. If the individuals are within  $\sigma_{share}$  distance of each other, their fitness is penalized to encourage future generations to spread along the entire Pareto frontier.

Two different distance metrics are used to compare the relative magnitudes of  $d(\kappa, \kappa')$  and  $\sigma_{share}$ . In genotypic space,  $d(\kappa, \kappa')$  is equal to the Hamming distance between the chromosomes representing individuals  $\kappa$  and  $\kappa'$ . The Hamming distance is simply the total number of positions in the binary strings that have different values. In phenotypic space, the Euclidean distance metric shown in equation (3.12) is used to calculate the distance between individuals.

$$d(\kappa, \kappa') = \sqrt{\sum_{i=1}^{npar} (x_{i,\kappa} - x_{i,\kappa'})^2}, \forall \kappa \quad (3.12)$$

For this application, the total number of decision variables ( $npar$ ) is equal to the number of monitoring wells,  $nwell$  in equation (3.1).

To determine the appropriate size of  $\sigma_{share}$  for a particular problem, three approaches can be taken. These approaches are described below in order of increasing domain-specific knowledge required. *Deb & Goldberg* (1989) derived equation (3.13) as a guide for practitioners for sizing niches in genotypic space by deriving  $\sigma_{share}$  to represent "...the maximum bits of difference allowed between the strings to make  $q$ -subspaces [or niches] in the solution space".

$$\frac{1}{2^l} \sum_{b=0}^{\sigma_{share}} \binom{l}{b} = \frac{1}{q} \quad (3.13)$$

Equation (3.13) assumes that the niches are uniformly spaced with each niche apportioned  $1/q$  of the decision space (*Deb & Goldberg* 1989). The equation represents a binomial distribution with a probability of 0.5, which can be solved for  $\sigma_{share}$  using the cumulative binomial distribution tables when string lengths  $l$  are less than or equal to 25 bits (indexed by  $b$  above). In this chapter, the chromosomes have a length of  $l = 20$  and the number of niches  $q = 40$ , which resulted in a genotypic  $\sigma_{share}$  equal to 5. For longer string lengths, the binomial distribution can

be approximated with the normal distribution. In this case, equation (3.14) from *Deb & Goldberg* (1989) can be used, where the “...  $z^*$  corresponding to the fraction  $1/q$  may be found from a cumulative normal distribution chart”.

$$\sigma_{share} = \frac{1}{2} \left( l + z^* \sqrt{l} \right) \quad (3.14)$$

The second approach for selecting  $\sigma_{share}$  is for phenotypic space. *Deb & Goldberg* (1989) derive  $\sigma_{share}$  in phenotypic space where “... each niche is enclosed in a  $p$ -dimensional hypersphere of radius  $\sigma_{share}$  such that each sphere encloses  $1/q$  of the volume of the space”. Equation (3.15) is the resulting expression for the phenotypic niche radius, which was found to have a value of 1.859 in this chapter.

$$\sigma_{share} = \sqrt{\sum_{i=1}^{npar} (x_{i,max} - x_{i,min})^2} / 2 q^{\frac{1}{npar}} \quad (3.15)$$

Phenotypic sharing has been shown to outperform genotypic sharing in most previous applications (*Fonseca & Fleming* 1995, *Coello* 1999, *Van Veldhuizen* 1999, *Horn* 1997, *Mahfoud* 1995).

Finally, an additional form of phenotypic sharing has been employed successfully in other applications (*Fonseca & Fleming* 1995, *Horn* 1997) where the distance metric between individuals is calculated in objective space. *Horn* (1997) provides guidance to practitioners for sizing niches in objective space using a geometric approach similar to the *Deb & Goldberg* (1989) phenotypic niche sizing method given in equation (3.15). The method requires that the maximum and minimum values of the objectives be known a priori. This approach was not used in this chapter because of the discrete nature of the sampling costs. Moreover, the maximum value of the SREE could not be calculated a priori.

The niche sizing relationships given by *Deb & Goldberg* (1989) and described above have been shown to be robust for a variety of problems (*Goldberg et al.* 1992b, *Horn* 1997) and will be analyzed in greater detail in the results section of this chapter.

### 3.4.1.3 Elitism: Seeking the King of the Niche

*Zitzler et al.* (2000) showed that elitism is one of the most important factors affecting the performance of EMO methods. A variety of elitist strategies have been used previously, usually consisting of maintenance of a population of nondominated solutions outside of the normal operators of the given EMO method being employed (for more details see *Ishibuchi & Murata* 1996, *Bäck* 1996, *Parks & Miller* 1998, *Zitzler & Thiele* 1999). *Zitzler et al.* (2000) state the primary question practitioners must answer when using elitist strategies as: “When and how are which members of the elite set re-inserted into the population?”. The elitist strategy employed in this chapter was designed to use previously derived niching relationships to answer this question and maximize the online performance of the NSGA. In an elitist sGA, the best member in the population at generation  $t$ , if not present in the new population resulting from selection, crossover, and mutation at generation  $(t+1)$ , randomly replaces one member of the population. For the NSGA, sharing provides niches that represent stable subpopulations that search for nondominated solutions in subspaces of  $\Omega$ . Conceptually, the elitist strategy proposed in this chapter is very similar to the sGA, in that the current best individual in a given niche at generation  $t$ , if not present in generation  $(t+1)$ , is inserted into that subpopulation, ensuring that its traits are available for subsequent search for the Pareto front.

This strategy was implemented by defining  $\sigma_{elite}$  or the elite radius, which is a parameter that allows the user to easily manipulate the amount of elitism. The elite radius defines the distance (either genotypic or phenotypic) beyond which members of the current nondominated

set are considered independent from one another. Only independent members of the nondominated set are considered for insertion in the next generation. For this application,  $\sigma_{elite} \approx \sigma_{share}$  which means that only one representative of each niche in the current nondominated set is considered for elitist reproduction into the next generation. The elitist solutions were selected in the four steps shown below from the nondominated set (or first front) at each generation  $t$ .

*Step 1:* Randomly select an objective  $f_\beta$  for  $\beta$  equal 1 to  $n_{obj}$

*Step 2:* Flip a coin to determine whether to start with either the member in the current nondominated set with the maximum value of  $f_\beta$  or the member with the minimum value.

*Step 3:* Identify the next point in the nondominated set that satisfies the following conditions:

- (1) Is a distance greater than  $\sigma_{elite}$  from the current solution
- (2) Is the closest member of the nondominated set to the current position

If none exist, then elitist reproduction is ceased or not performed at all.

*Step 4:* Repeat *Step 3* until elitist reproduction is ceased.

This approach identifies a niched elitist set by systematically stepping through the current nondominated front from one end to the other. After the elitist set of solutions is selected using the above steps, those members who are not represented in generation  $(t+1)$  randomly replace individuals within that population. Setting the elite radius equal to the niche radius worked well for this application, but the elite radius parameter allows the practitioner to directly manipulate the elitist selection pressure for other applications where this rule-of-thumb may not work as effectively.

The importance of elitism in the performance of the NSGA *Zitzler et al.* (2000) relates directly to the concepts of *genetic drift* and *selection pressure*. Genetic drift occurs when promising solutions in the population do not experience sufficient selection pressure and converge to non-optimal values (drift stall) under the influence of crossover and mutation. Elitism increases selection pressure by ensuring that the traits of nondominated individuals remain in the population for use in later generations.

#### 3.4.1.4 Selection Pressure: Avoiding Drift Stall

The importance of elitism in the performance of the NSGA motivated a further analysis of the role that selection pressure has on the algorithm's performance. Stochastic remainder selection selects a particular individual using the ratio of the individual's fitness and the average fitness of the current population in generation  $t$ . Recall from the introduction to the NSGA above that selection is based on dummy fitness values, which are decremented by a constant value  $\Delta D_f$  for each of the  $n_f$  successive fronts within the population. For example, analysis of the initial random population of 760 individuals used in the elitist runs discussed in the results section, shows that the expected number of copies of members in the nondominated set in the next generation had an average value of 1.7. The expected number of members from the 10<sup>th</sup> and 20<sup>th</sup> fronts in the next generation was 1.45 and 1.25, respectively. Although members of dominated fronts are expected to receive fewer copies than the nondominated front, the relative difference between 1.7 and 1.25 is small, which gives rise to an increased potential for drift stall to occur.

This increased potential for drift stall motivated an analysis of the effect of scaling the fitness of successive fronts on the NSGA's performance. The dummy fitness values for successive fronts are scaled using a scaling coefficient  $S_c$  whose value is less than one, such that

the minimum fitness in each front is guaranteed to be at least  $(1-S_c)$  percent higher than the maximum fitness in the front that immediately succeeds it. The scaling-based fitness assignments replace the constant decrement  $\Delta D_f$  assignments used previously in this work.  $S_c$  was set to be equal to 0.9, which ensures that the minimum fitness for front ( $n_f - 1$ ) is at least 10 percent greater than the maximum fitness in the  $n_f^{th}$  front. For the random population of 760 designs discussed above, the expected numbers of individuals in the next generation from the 1<sup>st</sup>, 10<sup>th</sup>, and 20<sup>th</sup> fronts will then be 6.2, 2.27, and 0.78 respectively. Note that scaling the system in this manner exponentially decreases the fitness of the members of the fronts succeeding the nondominated set. Setting  $S_c$  requires striking a balance between maintaining a diverse population and ensuring that selection pressure is sufficient to prevent drift stall. Further discussion on this issue is given in the results section below.

### ***3.4.2 Defining a Measure of Relative Performance***

To compare performance of the NSGA under different parameter settings, a measure of the algorithm's performance must be defined. In this work, the performance of the NSGA as a function of its parameters was measured relative to the Pareto frontier for cost and SREE. The frontier is shown in Figure (3.2), which shows the 36 sampling designs that compose the Pareto optimal set identified using enumeration of the more than 1 million potential designs in  $\Omega$ . The performance of the NSGA was quantified using a relative scoring metric (RSM) that measures the deviation of the nondominated set in generation  $t$  from the true front using equation (3.16).

$$Deviation(\eta) = \begin{cases} |SREE_{true}(\bar{x}_\eta) - SREE(\phi(t))| / SREE_{max}, & \text{if } f_1(\bar{x}_\eta) = f_1(\phi(t)) \forall \eta, \phi(t) \\ 1, & \text{otherwise} \end{cases}$$

$$RSM = 1 - \left( \sum_{\eta=1}^{36} Deviation(\eta) / 36 \right) \quad (3.16)$$

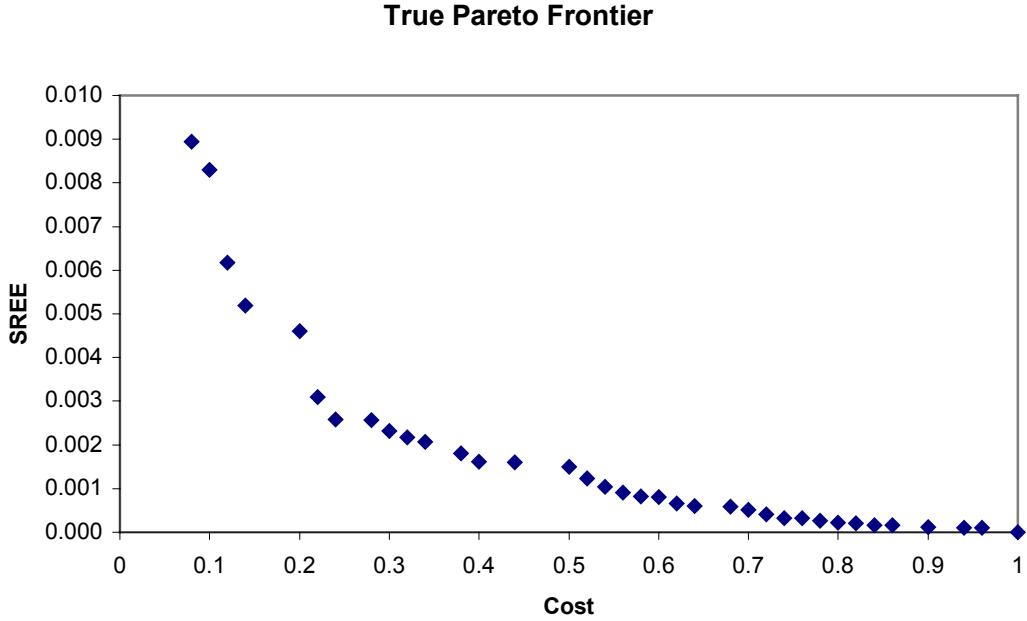


Figure 3.2 Actual tradeoff between the squared relative estimation error and cost. Each objective has been scaled to fall within the interval [0, 1]. The maximum and minimum SREE values found in  $\Omega$  were 0 and 13,000,000, respectively. The maximum and minimum cost values found in  $\Omega$  were \$18,420 and \$1473, respectively.

Although the maximum SREE value ( $SREE_{\max}$ ) cannot be calculated a priori, its maximum value was recorded while performing the enumeration of the decision space. Equation (3.16) requires the maximum value because the SREE values have to be scaled such that they fall within the interval [0,1]. Equation (3.16) defines the deviation between the  $\eta^{\text{th}}$  member of the enumerated Pareto optimal set and the  $\phi^{\text{th}}$  member of the current nondominated set in generation  $t$  to be equal to the absolute difference of their SREE values ( $SREE_{\text{true}}(\bar{x}_\eta) - SREE(\phi(t))$  shown above) if the designs have the same cost. If a cost level present in the Pareto optimal set is not represented in the nondominated set at generation  $t$  then equation (3.16) assumes a maximum deviation of one. The RSM was used to monitor the performance of the NSGA in order to evaluate the effectiveness of the guidelines presented in this section for a realistic application.

### 3.5 Results & Discussion

The guidelines presented in the previous section identify a range of potential population sizes and compute the niching parameters used by the NSGA to search a diverse set of stable subpopulations for the Pareto frontier. Additionally, the niching parameters are integral to properly setting the elitist selection pressure. The subsequent sections analyze the effectiveness of these guidelines by presenting the results of over 400 trial runs elucidating the performances of both the NSGA and the Elitist NSGA as a function of their parameters for the long-term monitoring design application.

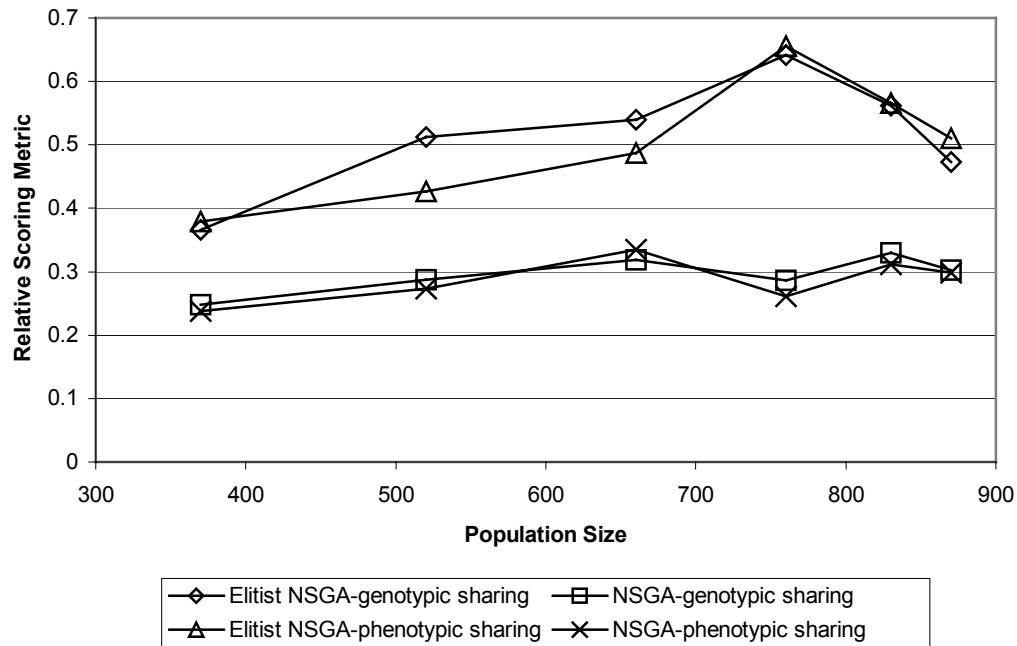


Figure 3.3 Selection of the proper population size for both the NSGA and the Elitist NSGA using the best RSM value attained from a single generation

#### 3.5.1 Performance of the NSGA & the Elitist NSGA

Figure (3.3) shows the performances of both the NSGA and the Elitist NSGA for the range of population sizes attained from equations (3.9) – (3.11) when implementing either

genotypic or phenotypic sharing. It is readily apparent from the figure that elitism greatly improves the online performance of the NSGA regardless of the population size used. Figure (3.3) was used to select the optimal population size for each of the four forms of the NSGA being considered. The performance trends for the NSGA and the Elitist NSGA are nearly identical for both sharing methods. The NSGA’s population sizes were set equal to 830 and 660 for genotypic and phenotypic sharing, respectively. These population sizes are consistent with previous studies, which have found that phenotypic sharing outperforms genotypic sharing. *Deb & Goldberg* (1989) argue that the reduced performance of genotypic sharing is caused by increased sensitivity of Hamming distance calculations to the assumption that each niche is uniformly apportioned  $1/q$  of the decision space, as required by equation (3.13). *Mahfoud* (1995) states that in addition to the sensitivity of genotypic sharing to the uniformity assumption, genetic drift and population sizing are influential in the performance differences between the two sharing schemes. Larger population sizes are required to ensure that important subpopulations (or niches) receive sufficient selection pressure and are not lost to the “noisy discrimination of genotypic sharing” (*Mahfoud* 1995).

Figure (3.3) and Figure (3.4) confirm that for the NSGA to have comparable performance under both sharing schemes the algorithm requires larger population sizes for genotypic sharing. Figure (3.4) shows that the NSGA has relatively poor online performance for discerning the true tradeoffs between costs and SREE for the monitoring application presented in this chapter, regardless of the sharing scheme invoked. It is important to note that genotypic sharing requires the least amount of problem-specific information of the sharing methods discussed in this chapter, but the inherent tradeoff of the NSGA using this form of sharing lies in the increased computational demands required due to the increased population size necessary for sharing in

genotypic space. This tradeoff is not present for the Elitist NSGA where Figure (3.3) clearly shows that the optimal population size is clearly defined by the peak when  $N$  is equal to 760 regardless of the sharing method used.

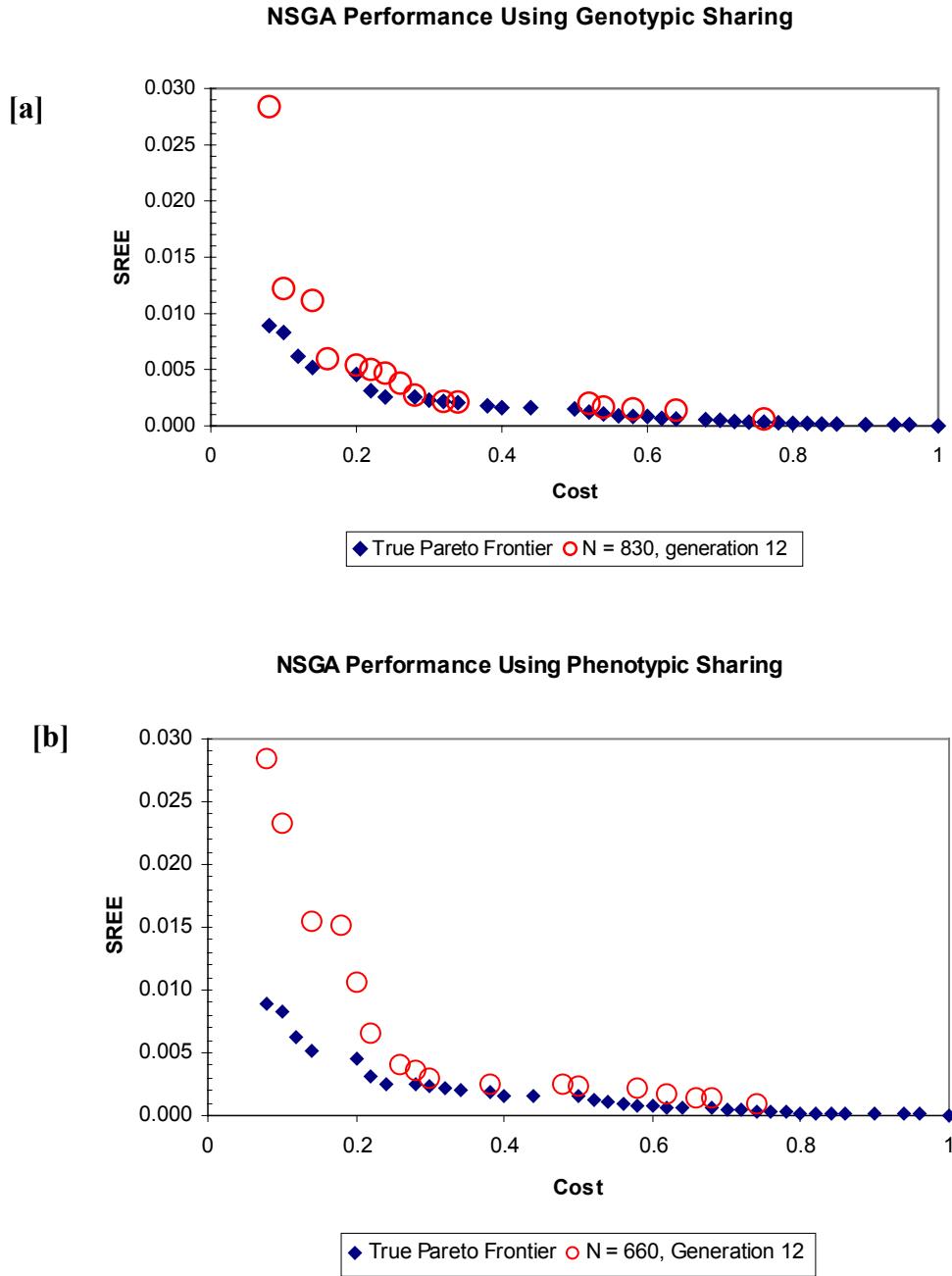


Figure 3.4 NSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing

Figure (3.5) shows that the niching-based elitist strategy presented in this chapter greatly improved the performance of the NSGA. The algorithm found solutions along the full extent of the Pareto front and was able to identify 17 of 36 members of the Pareto optimal set exactly, regardless of the sharing method used.

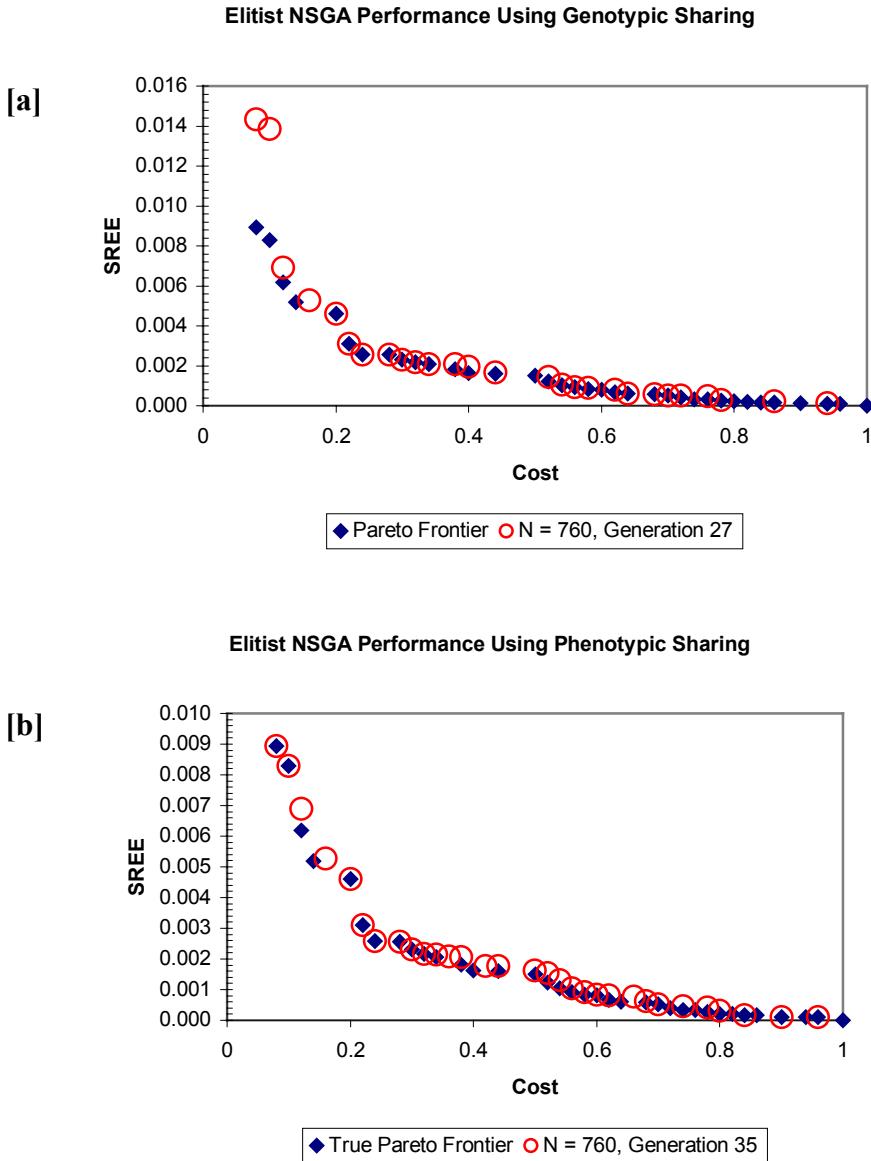


Figure 3.5 Elitist NSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing

Figure (3.3) and Figure (3.5) show that elitism served to both improve the online performance of the algorithm and help to overcome the “noisy discrimination of genotypic sharing” relative to sharing in phenotypic space.

Figure (3.6) shows the performance of the Elitist NSGA for the full range of possible values that both  $\sigma_{elite}$  and  $\sigma_{share}$  can be assigned under both sharing schemes (when  $N$  equals 760). Several observations are apparent from analysis of these plots. First, note that decreasing  $\sigma_{elite}$  from its maximum to its minimum value shows the NSGA’s behavior between the extremes of having no elitism (at the maximum value) to the case when all nondominated individuals are selected for elitist reproduction (at the minimum value). Both plots show that the NSGA’s performance improves with increasing elitist selection pressure. Figure (3.6) also confirms the noisy nature of sharing in genotypic versus phenotypic space. Under genotypic sharing, Figure (3.6a) shows several small peaks in performance for relatively sporadic combinations of niching and elitist parameter settings. In contrast, Figure (3.6b) shows that sharing in phenotypic space yields a smoother surface with a well-defined peak relative to the surface in (3.6a). Both plots serve to verify the guiding relationships discussed in the methods sections for population sizing, niching, and elitist selection. The arrows in the figure designate the algorithm’s performance when both  $\sigma_{elite}$  and  $\sigma_{share}$  are set using equations (3.12) and (3.14) taken from *Deb & Goldberg* (1989).

Figure (3.4) and Figure (3.5) show that both the NSGA and the Elitist NSGA solutions have gaps in the extreme regions of the Pareto frontier. These gaps occur because the subpopulations or niches representing these portions of the frontier were not able to survive for successive generations and are symptomatic of genetic drift stall.

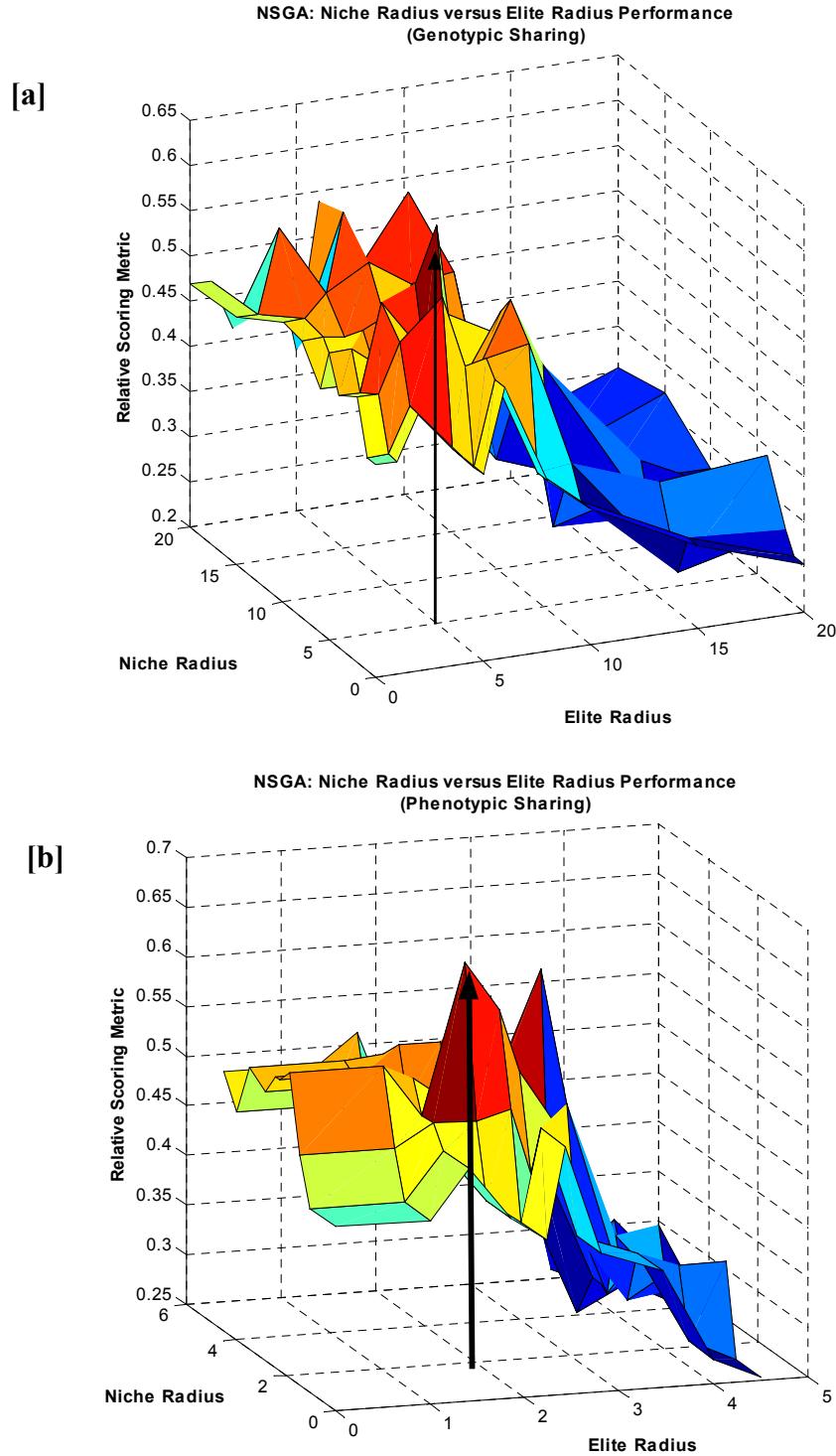


Figure 3.6 Analysis of the NSGA's performance for the full range of possible values that the parameters controlling niching and elitism can be assigned when using (a) genotypic sharing and (b) phenotypic sharing. The arrows designate the algorithm's performance when these parameters are set equal to the recommended niche radius attained from the relationships presented by Deb & Goldberg (1989).

Genetic drift also appears to be prevalent in Figure (3.4), which shows that the NSGA (in the absence of elitism) is unable to identify more than 3 members of the Pareto optimal set regardless of the sharing method invoked. Figure (3.7) shows the performance of all forms of the NSGA as a function of time.

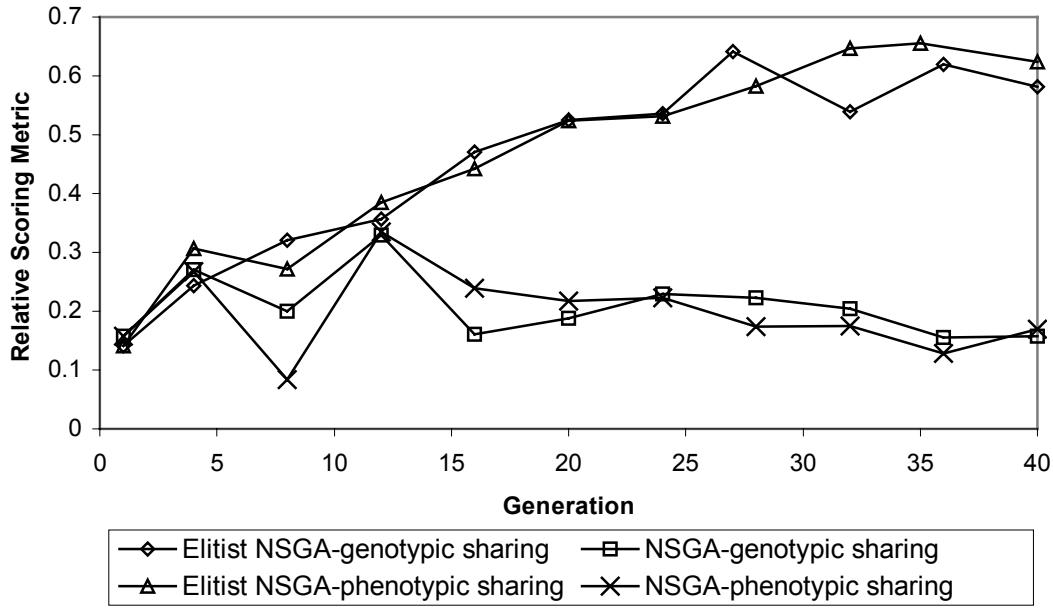


Figure 3.7 Performance of both the Elitist NSGA and the NSGA as functions of generation

The figure shows that elitism allows the NSGA to steadily improve performance over the duration of the run. In the absence of elitism, the sporadic peaks in performance are steadily degraded over the course of the runs, confirming that drift stall is occurring. Recall that the NSGA decrements the dummy fitness values for successive fronts by a constant value  $\Delta D_f$  and that stochastic remainder selection selects population members based on the ratio of their dummy fitness relative to the population's average dummy fitness. Drift stall is occurring because the constant decrement  $\Delta D_f$  does not properly scale the ratio of the nondominated individuals' fitness values relative to the population's average fitness, making it impossible for the NSGA to

distinguish and propagate these individuals into successive generations. These observations motivated further analysis of the influences of selection pressure and drift stall on the algorithm’s performance in the subsequent section.

### ***3.5.2 Performance Under Increased Selection Pressure***

The results of the previous section showed that both the NSGA and the Elitist NSGA converged to nondominated fronts with gaps in the extreme portions of the Pareto frontier. Additionally, Figure (3.7) shows that the NSGA’s performance only sporadically improves and generally degrades over the duration of the runs regardless of the sharing method considered. These observations confirm that several of the niches are either being lost or converging to non-optimal values due to the absence of sufficient selection pressure. Furthermore, it is readily apparent that elitism significantly improves the algorithm’s performance confirming the results of *Zitzler et al.* (1999). Elitism serves to increase the selection pressure on the niches and reduce the influence of genetic drift on the system, which in effect is analogous to rescaling the relative fitness values of the population. Modifying the NSGA and Elitist NSGA such that successive fronts have exponentially decreasing fitness using the scaling coefficient  $S_c$  (as described in the methods section) enabled further analysis of the influence of increased selection pressure on performance. For the monitoring application, analysis of the NSGA’s performance for scaling coefficient values ranging from 0.7 to 0.95 showed that  $S_c$  should be set to approximately 0.9. Values for  $S_c$  lower than 0.9 greatly reduced the diversity of the populations and caused premature convergence (due to excessive selection pressure). Scaling coefficient values above 0.9 resulted in significant gaps in the extreme regions of the Pareto frontier due to insufficient selection pressure and genetic drift.

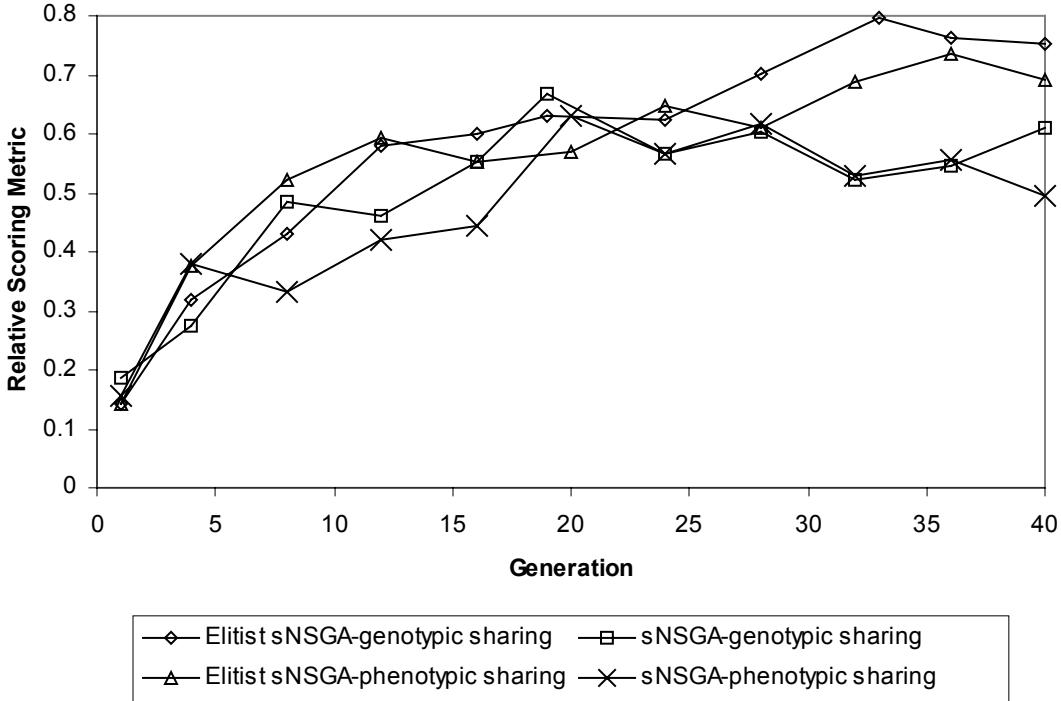


Figure 3.8 Performance of both the Elitist NSGA and the NSGA as a function of generation after rescaling the fitness assignments

Figure (3.8) shows the performances of the scaled NSGA (sNSGA) and the Elitist sNSGA as a function of time when  $S_c$  equals 0.9. The population sizes used in these runs remained the same as those in the previous section for all forms of the sNSGA except the non-elitist, genotypic sNSGA where a population size of 870 was found to be optimal, which represents only a slight increase from the previous section. Rescaling was not able to reduce the sensitivity of genotypic sharing to non-uniformities in the niche spacing, although elitism successfully closed the performance gap between the two sharing methods for the Elitist sNSGA. It is immediately obvious from Figure (3.8) that rescaling the fitness assignments greatly improved the performance of the sNSGA, which no longer shows degradation in performance with time and instead steadily improves towards an upper bound in performance. Note that the combination of rescaling and elitism resulted in the Elitist sNSGA exceeding the previous results

shown in Figure (3.7) for the Elitist NSGA. The phenotypic, Elitist NSGA was able to reach a maximum RSM value of 0.66 whereas the genotypic, Elitist sNSGA attained a maximum RSM value of 0.8.

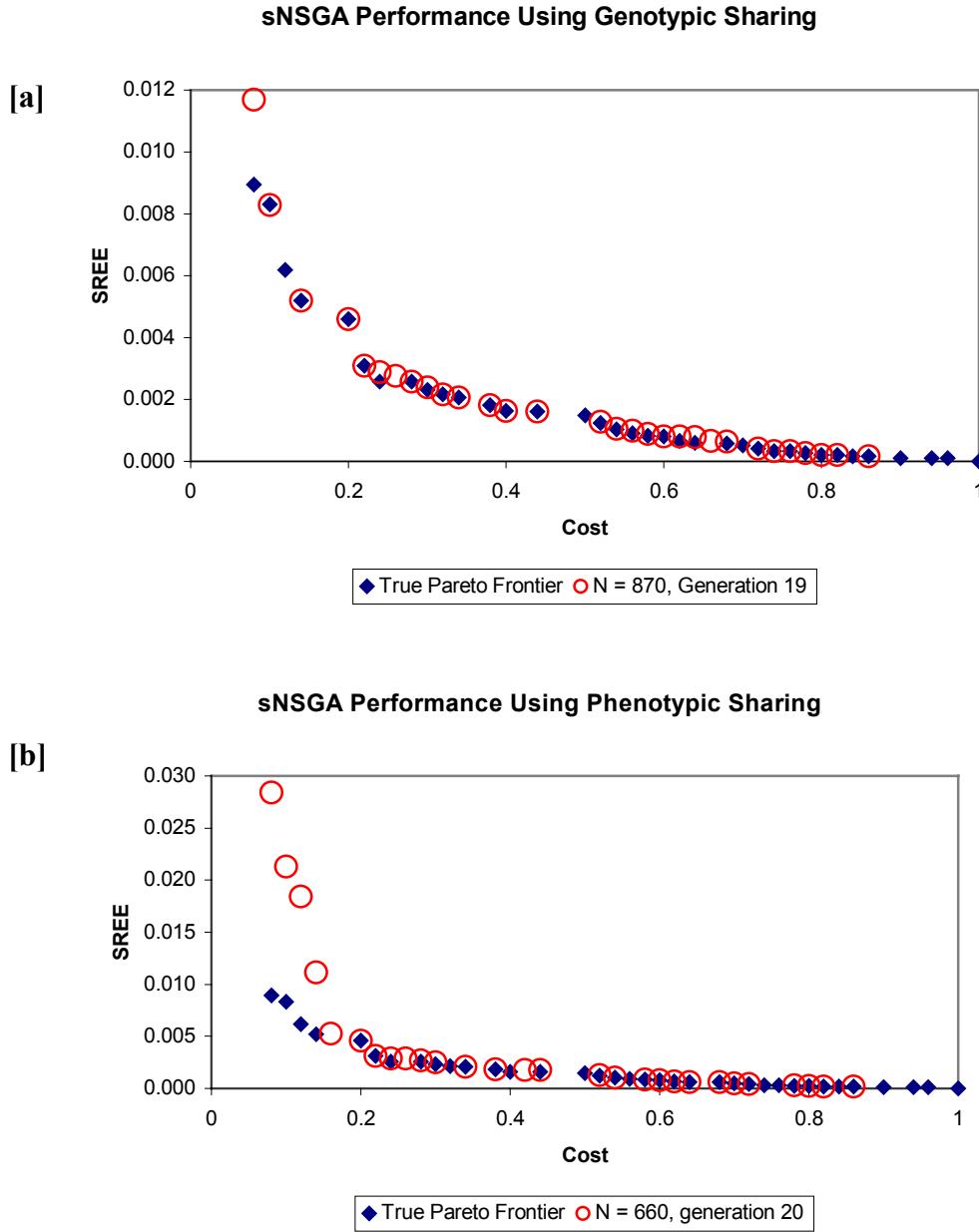


Figure 3.9 sNSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing

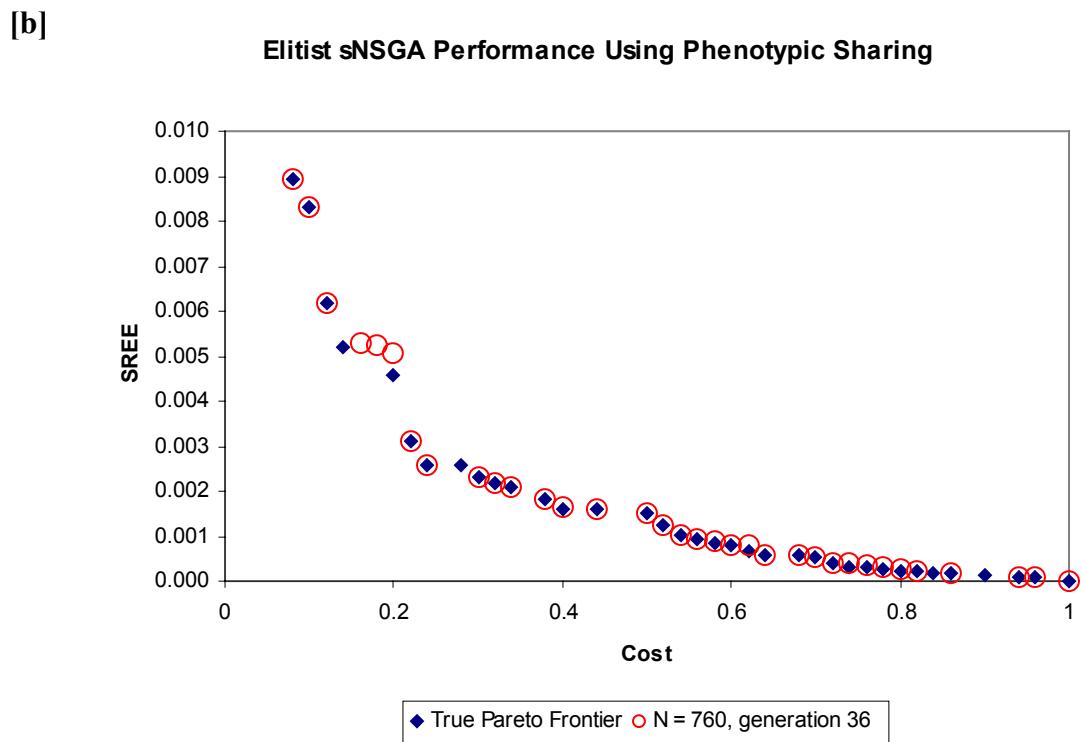
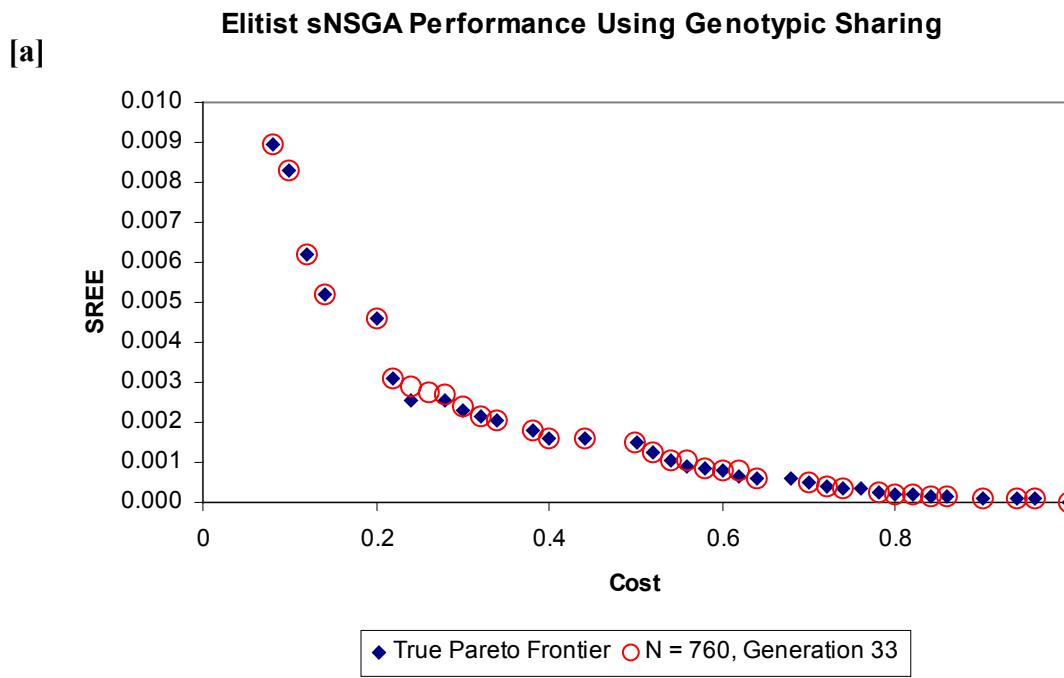


Figure 3.10 Elitist sNSGA's performance relative to the true Pareto front using both (a) genotypic and (b) phenotypic sharing

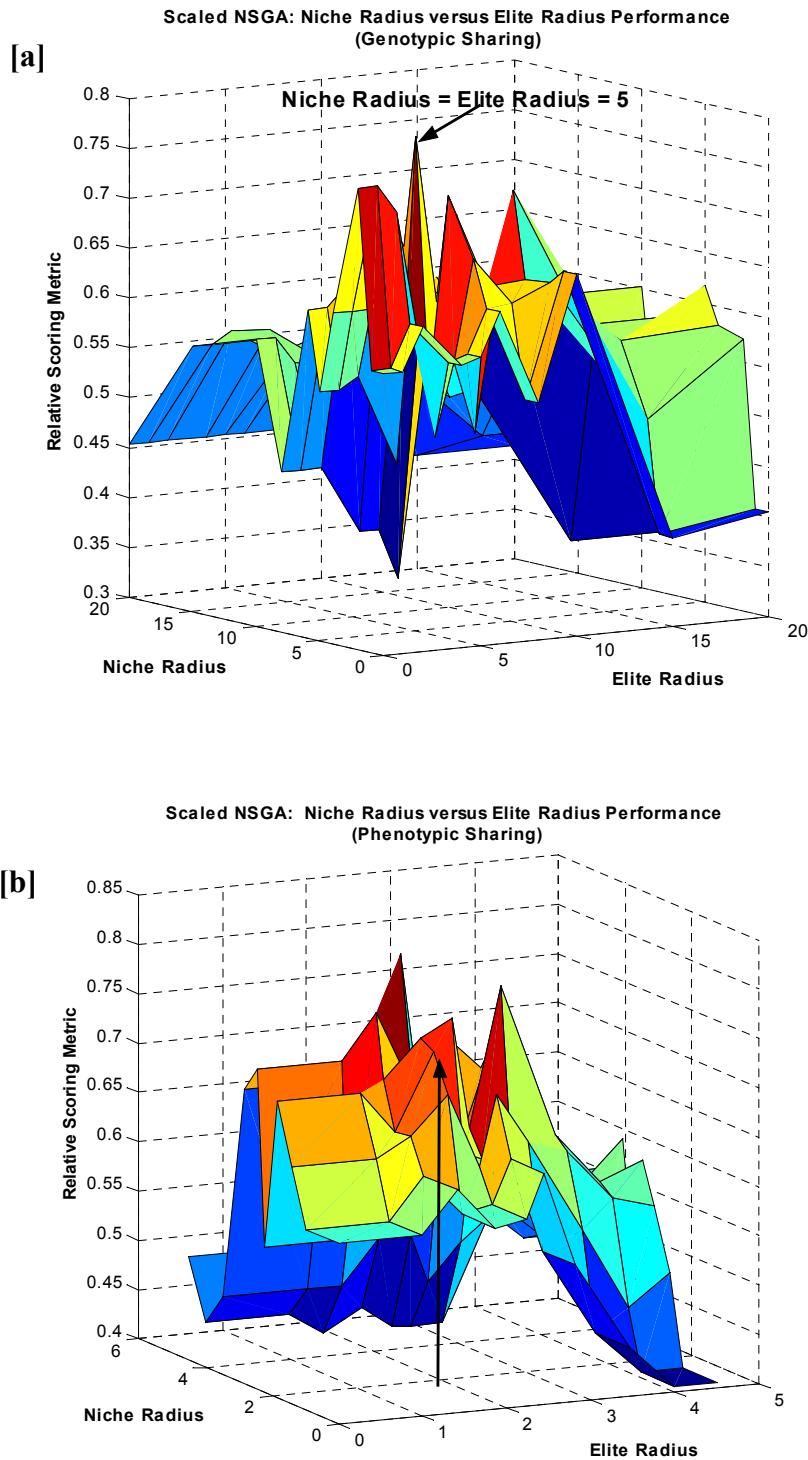


Figure 3.11 Analysis of the sNSGA's performance for the full range of possible values that the parameters controlling niching and elitism can be assigned when using (a) genotypic and (b) phenotypic sharing. The arrows designate the algorithm's performance when these parameters are set equal to the recommended niche radius attained from the relationships presented by Deb & Goldberg (1989).

Figure (3.9) and Figure (3.10) show the improved performances of all forms of the NSGA that result from rescaling the fitness assignments. Recall that previously the non-elitist NSGA was able to find a maximum of 3 members of the Pareto optimal set. Rescaling the shared dummy fitness values resulted in finding 21 and 17 members of the Pareto optimal set when implementing genotypic and phenotypic sharing, respectively. The additional use of elitism enabled the sNSGA to almost replicate the Pareto frontier for both sharing methods, as shown in Figure (3.10). The genotypic, Elitist sNSGA provided the most complete representation of the frontier by exactly finding 29 members of the Pareto optimal set while also finding close representations of all but 2 of the remaining 36 members. The phenotypic, Elitist sNSGA found either exact or close representations of 31 of the 36 members of the Pareto optimal set. The slight decrease in performance of the phenotypic, Elitist sNSGA can be explained with analysis of Figure (3.11).

Figure (3.11) shows the performance of the sNSGA using both sharing schemes for the full range of values  $\sigma_{elite}$  and  $\sigma_{share}$  that can be assigned. The plot again shows that increasing the elitist selection pressure by decreasing  $\sigma_{elite}$  generally improves performance. The slight difference in performance between genotypic and phenotypic sharing was caused by the assumption used in this chapter that  $\sigma_{elite} \approx \sigma_{share}$ . Figure (3.11a) shows that this assumption results in finding the highest peak in performance for the sNSGA under genotypic sharing, whereas Figure (3.11b) shows that this assumption was not able to exactly find the highest performance peak in phenotypic space. It should be noted from Figure (3.6) and Figure (3.11), however, that assuming  $\sigma_{elite} \approx \sigma_{share}$  for the monitoring application generally resulted in finding peak or very near peak performances for the various forms of the NSGA in both sharing spaces, which is probably sufficient for most applications. Figure (3.11a) again shows that genotypic

space was generally very noisy relative to phenotypic space, which explains why the sNSGA still required larger population sizes to be used in genotypic space in order for the two sharing schemes to attain comparable results in the absence of elitism. Additionally, the increased noise present in genotypic space after rescaling, as shown in Figure (3.11a) relative to the previous case in Figure (3.6a), explains why  $N$  had to be increased from 830 to 870 as discussed above. Overall, the guiding relationships and considerations discussed in the methods section of this chapter were able to successfully guide the design of the NSGA. Figure (3.10) shows that the algorithm was able to find 95 percent of the Pareto optimal set when the influences of population sizing, elitism, and genetic drift were carefully considered in its design.

### 3.6 Conclusions

The two dimensional tradeoff between cost and SREE for the groundwater monitoring application presented in this work was accurately quantified by ensuring that the NSGA was properly designed to navigate the problem's decision space. Preliminary problem analysis identified a range of potential population sizes and the potential computational complexity of solving the problem. Additionally, relationships from *Deb & Goldberg* (1989) effectively sized the niches required to maintain stable subpopulations, each of which actively sought different sections of the Pareto frontier. Using the recommended niche radius from *Deb & Goldberg* (1989) also effectively identified the niched elitist set of solutions for successive generations in each of the runs performed in this chapter. Elitism greatly improved the efficacy of the NSGA and helped to reduce performance differences between genotypic and phenotypic sharing. Analysis of the algorithm's performance as a function of time showed the adverse effects of genetic drift and motivated further study of the role of selection pressure in the NSGA. Rescaling the shared dummy fitness assignments of the NSGA such that successive fronts had

exponentially decreasing fitness values greatly improved the algorithm's performance. Combining elitism with rescaled fitness assignments resulted in identifying all but 2 members of the Pareto optimal set.

This chapter provides practitioners with a methodology that identifies the proper parameter settings for the NSGA by: (1) identifying the most appropriate population size, (2) properly sizing niches for fitness-based sharing, (3) correctly setting the elitist selection pressure, and (4) careful performance analysis to avoid genetic drift stall. Parameter settings for the NSGA attained using a total of 10 runs resulted in peak or near peak performance of the algorithm. Six of the trial runs were necessary to select the proper population size. Elitism greatly improved the NSGA's performance and narrowed any performance differences between sharing in genotypic versus phenotypic space, a clear advantage when problem-specific information required for phenotypic sharing is not available. When such information is available, though, phenotypic sharing is generally preferred because of the increased noise associated with selection in genotypic space in most EMO applications (*Fonseca & Fleming* 1995, *Horn* 1997), as was observed in this application. Finally, the use of the niche radius to set the elitist selection pressure proved to be very effective in attaining peak or near peak performance from the NSGA when quantifying the 36 member Pareto optimal set of designs that compose the Cost—SREE tradeoff.

Rescaling the fitness assignments to further increase selection pressure and reduce the influence of genetic drift requires consideration of the potential tradeoff between improving the algorithm's performance and reducing diversity within the population. Additional analysis was necessary to properly address this tradeoff and select the appropriate scaling coefficient. The

final selection of a scaling coefficient required 4 additional trial runs after the six trial runs used to set the population and niching parameters.

Although the NSGA was able to successfully capture the Cost—SREE tradeoff, the user should note that this success required significant user interaction, a detailed analysis of 10 runs, and the evaluation of more than 280,000 sampling designs. The NSGA design methodology presented in this chapter requires significant user expertise and a tremendous investment of computational resources that precludes the solution of more challenging multiobjective problems (e.g., problems with more than 2 objectives). Chapter 4 overcomes these limitations by introducing a design methodology for the NSGA-II, a second generation EMO genetic algorithm (see *Deb et al.* 2000) that improves upon the NSGA. The NSGA-II design methodology in the next chapter simplifies Pareto optimization by automating parameter specification and reducing the computational effort required to solve multiobjective applications. The design methodology enables the automatic solution of a new class of high order multiobjective applications in which users can select, understand, and balance more than two performance criteria [see Chapters 4 and 6].



## 4 SIMPLIFYING OPTIMIZATION IN PARETO SPACE

### 4.1 Introduction

Chapter 3 showed that with proper design and parameterization, the NSGA is able to accurately quantify 2 dimensional tradeoffs. This chapter builds on Chapter 3 by introducing a design methodology for the NSGA-II, a second generation EMO genetic algorithm (see *Deb et al.* 2000) that improves upon the NSGA by (1) invoking a more efficient nondomination sorting algorithm, (2) eliminating the sharing parameter (i.e., the niche radius), and (3) adding an implicitly elitist selection method that greatly aids in solving high order problems (i.e., problems with more than 2 objectives). The NSGA-II is the focus of this work because the algorithm has been shown to perform as well or better than the other second generation EMO algorithms on difficult, high order problems (see *Zitzler et al.* 2001 and *Deb et al.* 2001). The NSGA-II design methodology presented in this chapter builds on previous GA design methodologies (*Reed et al.* 2000b, *Lobo* 2000, and *Reed et al.* 2001) to introduce a multi-population approach that automates parameter specification while significantly reducing the computational effort required to solve multiobjective applications. The design methodology fully exploits the efficiency of the NSGA-II to enable the automatic solution of a new class of high order multiobjective applications in which users can select, understand, and balance more than two performance criteria (see Chapter 6). Readers familiar with Chapter 3 can skip forward to Section 4.3 because Section 4.2 describes the same test case and problem formulation.

### 4.2 Monitoring Application

#### 4.2.1 Test Case Data

The NSGA-II design methodology presented in this chapter is demonstrated on the same long-term monitoring (LTM) application from Chapter 3. The test was developed using data

drawn from a 38 million-node flow-and-transport simulation performed by *Maxwell et al.* (2000). The simulation provided realistic historical data for the migration of a hypothetical perchloroethylene (PCE) plume in a highly heterogeneous alluvial aquifer. The hydrogeology of the test case is based on an actual site located at the Lawrence Livermore National Laboratory in Livermore, California, currently being managed under the United States' Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) program. Data were provided for a total of 50 hypothetical sampling locations within the 20-well multi-level monitoring network. The data represent a snapshot in time, 10 years after a continuous point source began contaminating the aquifer system. The monitoring wells can sample from 1 to 3 locations along their vertical axis and have a minimum spacing of 60-m between wells in the horizontal plane.

The site is assumed to be undergoing long-term monitoring, in which groundwater samples are used to assess the effectiveness of current remediation strategies. During this long-term monitoring phase of a remediation, sampling and laboratory analysis can be a controlling factor in the costs of remediating a site. Quarterly sampling of the entire network has a potential cost of over \$70,000 annually for PCE testing alone, which could translate into millions of dollars if the site had a typical life span of 20 to 30 years. To reduce these costs, the application seeks to identify *redundant* sampling locations (i.e., points that minimally affect interpolation-based plume estimates when not sampled). Specifically, the application seeks to quantify the tradeoff between LTM sampling costs and the accuracy of the plume interpolation estimates. These objectives are discussed in more detail in Section 4.2.2, below.

#### 4.2.2 Problem Formulation

Equation (4.1) gives the multiobjective problem formulation for quantifying the tradeoff between sampling costs and maintenance of a high quality interpolated picture of the plume.

$$\begin{aligned}
 & \text{Minimize } F(\bar{x}_\kappa) = [f_1(\bar{x}_\kappa), f_2(\bar{x}_\kappa)], \quad \forall \kappa \in \Omega \\
 & f_1(\bar{x}_\kappa) = \sum_{i=1}^{n_{well}} C_S(i)x_{\kappa i} \\
 & f_2(\bar{x}_\kappa) = \sum_{j=1}^{n_{est}} \left( c_{all}^*(\bar{u}_j) - c_{est}^\kappa(\bar{u}_j) \right)^2
 \end{aligned} \tag{4.1}$$

$F(\bar{x}_\kappa)$  is a vector valued objective function whose components  $[f_1(\bar{x}_\kappa), f_2(\bar{x}_\kappa)]$  represent the cost and squared relative estimation error (SREE), respectively, for the  $\kappa^{\text{th}}$  monitoring scheme  $\bar{x}_\kappa$  taken from the collection of all possible sampling designs  $\Omega$ . Equation (4.2) defines the binary decision variables representing the  $\kappa^{\text{th}}$  monitoring scheme.

$$x_{\kappa i} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ well is sampled} \\ 0, & \text{otherwise} \end{cases}, \quad \forall \kappa, i \tag{4.2}$$

If the  $i^{\text{th}}$  well is sampled it is assumed that all available locations along the vertical axis of that well will be sampled at a cost of  $C_S(i)$ .  $C_S(i)$  ranged from \$365 to \$1095 for 1 to 3 samples analyzed for PCE solely (Rast 1997). Sampling all available levels within each well reduces the size of  $\Omega$  from  $2^{50}$  to  $2^{20}$ , where 50 and 20 represent the total number of sampling locations and monitoring wells ( $n_{well}$ ), respectively. Reducing the size of  $\Omega$  enabled the entire decision space of this application to be enumerated. Enumeration was employed to identify the true tradeoff (or Pareto frontier) between cost and SREE, shown in Figure (4.1), to aid in assessing the performance of the NSGA-II design methodology presented in this chapter.

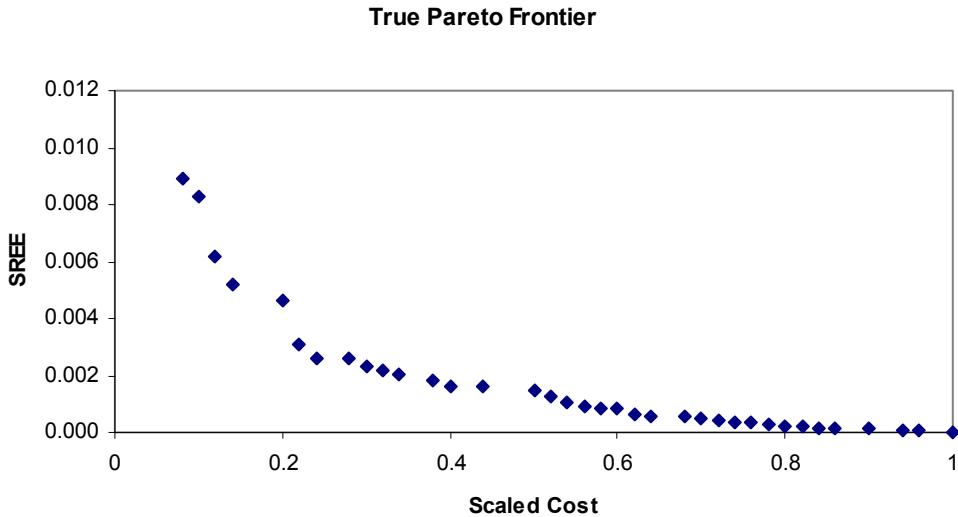


Figure 4.1 The enumerated tradeoff between Cost and SREE

The SREE provides a measure of how the interpolated picture of the plume using data only from wells included in the  $\kappa^{\text{th}}$  sampling plan compares to the result attained using data from all available sampling locations. The measure is computed by summing the squared deviations between the PCE estimates using data from all available sampling locations,  $c_{\text{all}}^*(\bar{u}_j)$ , and the estimates based on the  $\kappa^{\text{th}}$  sampling plan  $c_{\text{est}}^\kappa(\bar{u}_j)$  at each location  $\bar{u}_j$  in the interpolation domain. Each  $\bar{u}_j$  specifies the coordinates for the  $j^{\text{th}}$  grid point in the interpolation domain. The interpolation domain consisted of a total of 3300 grid points ( $n_{\text{est}}$  in equation (4.1)). Nonlinear least squares inverse distance weighting was used to interpolate plume concentrations in this chapter (for more details see Chapter 3 or Chapter 5).

### 4.3 The NSGA-II Basics

Basic principals and operators used by the NSGA-II in quantifying design tradeoffs are discussed in this section. Genetic algorithms search a decision space using a process that is analogous to Darwin's "natural selection". The decision variables associated with the

optimization model to be solved are encoded as 0-1 binary strings of length  $l$  (termed chromosomes). The fitness of each member of a randomly generated initial population of these strings is determined by how well the design satisfies the objectives of an application. After each individual is assigned a fitness value, GAs find optimal solutions using three basic operators: (1) selection, (2) crossover (mating), and (3) mutation.

The NSGA-II uses a two-step selection process, which combines both binary tournament selection and  $(\mu + \lambda)$  selection. First, binary tournament selection allows only the fittest individual from two strings randomly drawn from the current population to be placed into the mating population. Next, the crossover operator couples members of the mating population to mate. Mating was done using uniform crossover in which the strings exchange each of their component bit values with a specified probability of crossover ( $P_c$ ). Tournament selection and crossover are repeated until a population of  $N$  children has been produced. Mutation then randomly flips binary bits from 1 to 0 or vice versa within the new child population with a given probability of mutation ( $P_m$ ). Lastly, the NSGA-II uses  $(\mu + \lambda)$  selection to choose which of the parent and child designs will survive. In this selection scheme, the populations of  $N$  parent designs and  $N$  child solutions in the current generation  $t$  are combined to yield a selection pool of  $2N$  individuals, from which the  $N$  best individuals are allowed to pass to generation  $t+1$ . This selection method aids the algorithm in quantifying high order Pareto surfaces because it is implicitly elitist (i.e., the best designs are guaranteed to survive into the next generation).

These three operators act to create a new population (or generation) of individual sampling plans with improved average fitness. The Schema Theorem is the general theory describing how these three operators combine to evolve high quality near-optimal solutions [see *Goldberg, 1989* or *Holland, 1975* for more information]. It states that highly fit strings are

composed of small chunks of information (or building blocks) that are relevant to the solution of the problem. The GA exerts a *selection pressure* where only highly fit members are allowed to pass their traits or building blocks to the next generation. The phrase “selection pressure” refers to the probability of survival for individuals in a given population. For example, if population members are picked at random without being compared with one another in terms of their fitness values than each population member has an equal probability of surviving and the system has low selection pressure. Alternatively in a system with high selection pressure, a population member must compete with several other population members and only the fittest design has a high probability of survival. Highly fit parent strings are allowed to mate, yielding offspring that inherit building blocks from both parents. It is in this manner that the GA assembles optimal or near-optimal solutions to a problem.

The primary difference between the NSGA-II and single objective GAs is in how fitness is assigned. The NSGA-II evaluates sampling designs in terms of a vector of objectives. A sampling design cannot be assessed in terms of its performance in any single objective because it may perform poorly with respect to the remaining objectives. The NSGA-II employs the concepts of *Pareto dominance* and *crowding* to assign fitness values to sampling designs in the two steps described below (*Deb et al. 2000*).

The first step in fitness assignment employs the concept of *Pareto dominance* in which a design  $\bar{x}$  dominates another design  $\bar{x}'$  if and only if it performs as well as  $\bar{x}'$  in all  $n_{obj}$  objectives and better in at least one. In the NSGA-II’s improved nondomination sorting, the algorithm first ranks each design by the number of designs that dominate it. The second step in fitness assignments utilizes the concept of *crowding* (for more details see *Deb et al. 2000*) in which the average distance between an individual design and those designs nearest to it within

the population (in terms of their objectives' values). The crowding distance represents "...the largest cuboid enclosing the [ $\kappa^{\text{th}}$  design] without including any other point in the population" (*Deb et al.* 2000). The crowding operator eliminates the sharing parameter calculations that were required for the original form of the algorithm.

After these two steps, the fitness value of an individual design is assigned either its rank or its crowding distance. In cases where two designs have different ranks, the individual with the lower rank is preferred (i.e., individuals that are dominated by fewer solutions). Alternatively, if both solutions possess the same rank then the individual with larger crowding distance is preferred (i.e., the individual that adds the most diversity to the population). A diverse population ensures that the NSGA-II will find solutions along the full extent of the Pareto surface.

#### 4.4 Simplifying Multiobjective Search & Optimization

The NSGA-II design methodology presented in this chapter combines concepts from previous GA design methodologies (*Reed et al.* 2000b, *Reed et al.* 2001) and the "parameter-less GA" methodology presented by *Lobo* (2000). The NSGA-II requires the specification of the four following parameters: (1) population size, (2) probability of crossover, (3) probability of mutation, and (4) run length. This set of 4 parameters represents an immediate simplification relative to the NSGA design methodology presented in Chapter 3, which also required the user to set a niche radius, an elite radius, and the appropriate selection pressure. In Chapter 3, the NSGA design methodology required a total of 10 runs and substantial user interaction to ensure the algorithm was able to capture a 2-dimensional tradeoff between costs and SREE. The NSGA-II design methodology presented in this chapter reduces the complexity of solving

multiobjective problems by introducing 3-step approach that can be used to completely automate parameter selection, minimize user interaction, and substantially reduce computational costs.

Step 1 consists of a preliminary analysis where the user sets bounds for the population size and the total time of computation. Step 2 utilizes control map theory for genetic algorithms (see *Thierens* 1995) to automatically set the probabilities of crossover and mutation. Finally, in Step 3 the algorithm automatically increases population sizes in successive runs until either *offline analysis* shows the NSGA-II sufficiently quantified the nondominated set (i.e., the tradeoff surface) or the user-specified maximum time of computation has been reached. *Offline analysis* keeps track of the best collection of nondominated individuals from all of the designs evaluated in successive runs. Each of these steps is described in more detail in Sections 4.4.2-4.4.4 below.

#### **4.4.1 Initial Considerations**

The design methodology assumes that computationally intensive fitness functions for water resources applications preclude identifying parameter settings for a distribution of initial random number seeds and instead focuses on finding optimal parameter settings for a single random number seed. Additionally, the method assumes that the user has successfully formulated their problem such that the NSGA-II will converge to a feasible nondominated set.

#### **4.4.2 Step1: Preliminary Problem Analysis**

The first step in this design methodology requires users to answer the four following questions:

- What is your initial goal for the number of nondominated solutions,  $R_{ND}$  that is acceptable?
- What is the minimum percentage change in the number of nondominated solutions,  $\Delta_{ND}$ , for two successive runs to be considered identical?

- What is the average time required to evaluate a design,  $T_{eval}$ ?
- What is the maximum acceptable run time,  $T_{tot}$ ?

In the first question,  $R_{ND}$  defines the user's preliminary goal for the number of nondominated solutions she or he would want to obtain on the tradeoff surface.  $R_{ND}$  is then used to calculate a lower bound population size  $N_0$  for the NSGA-II using the empirical rule-of-thumb presented in equation (4.3).

$$N_0 \approx 2R_{ND} \quad (4.3)$$

Users should not concern themselves with exactly specifying  $R_{ND}$  because it is generally impossible to know a priori the exact number of Pareto optimal solutions that exist for a given application.

Since it is impossible to know a priori the exact number of Pareto optimal solutions that exist, the user must consider problem specific information to set this value. For the LTM application presented in this chapter, the NSGA-II is seeking to find the minimum SREE value for each level of cost. Theoretically, the nondominated set can be composed of no more than 58 solutions because the cost objective is a discrete, linear function of the number of sampling locations (which ranges between 1 and 58).  $R_{ND}$  was set equal to 30, which is approximately 50 percent of the theoretical maximum number of nondominated solutions that could exist.  $R_{ND}$  was set to a relatively small number of nondominated solutions because only in rare cases will the nondominated set be composed of the theoretical maximum number of individuals. Figure (4.1) demonstrates this fact, showing that only 36 nondominated solutions exist for the LTM application because some of the designs at lower cost levels dominate higher cost designs (i.e., they have both a lower cost and a lower SREE value). Setting  $R_{ND}$  to a reasonably small value

has the additional advantage of ensuring that the NSGA-II can be tested for small population sizes, which can significantly reduce the total number of designs that must be evaluated if the algorithm successfully quantifies the tradeoff surface using these reduced population sizes. Moreover, the successive increases in population size described in Step 3 of this methodology will correct for an undersized population automatically.

The second question directly addresses the issue of *solution exactness*, by requiring the user to specify the minimum percentage change in the number of nondominated individuals  $\Delta_{ND}$  for two successive runs to be considered identical. This parameter is best explained using an illustrative example. Consider two successive runs of the NSGA-II in which the first run uses a population of  $N$  designs to evolve a nondominated set composed of  $A$  individuals while the second run uses a population of  $2N$  designs to evolve a nondominated set of  $K$  individuals. The results of these runs are used in Equation (4.4) to define which of the two following courses of action will be taken: (1) population size is again doubled, resulting in  $4N$  individuals to be used in an additional run of the NSGA-II or (2) the algorithm stops because the nondominated set has been quantified to sufficient accuracy.

$$\begin{aligned} & \text{if } \Delta_{ND} < \left( \frac{|K - A|}{A} \right) 100 \text{ then double } N \text{ and continue search} \\ & \text{else stop search} \end{aligned} \quad (4.4)$$

In equation (4.4), if the run using a population size of  $2N$  does not increase the size of the nondominated set by more than  $\Delta_{ND}$ -percent then the search has reached a point of diminishing return where further population size increases are no longer justified. Equation (4.4) is relevant to both discrete and continuous problems because in both cases the NSGA-II is building a discrete approximation to the true Pareto surface.  $\Delta_{ND}$  was set to 10 percent for the LTM application presented in this chapter, implying that a close approximation of the true Pareto front

shown in Figure (4.1) is being sought. Setting  $\Delta_{ND}$  directly addresses the tradeoff between *solution exactness* and the *computational complexity* of an application. By setting  $\Delta_{ND}$  to any value above zero, the user is accepting a less exact representation of the nondominated set to achieve a decrease in computing time.

The third and fourth questions require the user to specify how long he or she is willing to wait for a solution (i.e.,  $T_{tot}$ ), which allows the upper bound population size to be computed given how long each function evaluation takes on average,  $T_{eval}$ . Note that any timing variables presented in this chapter with a capital  $T$  are in terms of clock time and those presented in lower case  $t$  are in terms of the number of generations the NSGA-II is run. For the LTM application,  $T_{tot}$  was set equal to 4 hours to allow the NSGA-II ample opportunity to quantify the nondominated solution set given that each design took an average of 0.044 seconds to evaluate on a Dell XPS T800r running Windows 2000. The maximum number of designs,  $E_{tot}$ , that can be evaluated given both  $T_{eval}$  and  $T_{tot}$  can be computed using equation (4.5) and had a value of 327,000 for the LTM application.

$$E_{tot} = \frac{T_{tot}}{T_{eval}} \quad (4.5)$$

The results of equation (4.5) are used in equation (4.6) to compute the maximum number of times,  $\beta_{MAX}$ , the population size can be doubled beyond its lower bound value  $N_0$  computed using equation (4.3).

$$\sum_{\beta=0}^{\beta_{MAX}} 2^\beta N_0 t \leq E_{tot} \quad (4.6)$$

In equation (4.6),  $t$  represents the total number of the generations used in each run of the NSGA-II to search for the nondominated set. The run length  $t$  is estimated to be approximately equal to  $2l$ , where  $l$  is the length of the binary strings that represent designs, as was recommended by *Reed et al.* (2000b). The LTM designs had a string length  $l = 20$  yielding a run time of 40 generations. Conservatively, this estimate assumes that the combination of  $(\mu + \lambda)$  and binary tournament selection in the NSGA-II converges as fast as a system undergoing pure binary tournament selection (for details see *Thierens et al.* 1998). The quantity  $N_0 t$  represents the total number of designs that are evaluated when the lower bound population size is used for  $t$  generations and had a value of 2400 for the LTM application. Given the maximum number of designs that can be evaluated ( $E_{tot} = 327,000$ ), equation (4.6) shows that  $N_0$  can be doubled a total of six times (i.e.,  $\beta_{MAX} = 6$ ), yielding an upper bound population size of 3840.

#### **4.4.3 Step 2: Balancing Innovation and Disruption**

Step 1 of the methodology requires users to supply information for performing population sizing and specifying run length, which are two of the four parameters required for the NSGA-II. This step explains how the two remaining parameters, the probability of crossover  $P_c$  and the probability of mutation  $P_m$ , are set in this methodology. There are two forces that must be balanced when setting these parameters' values: (1) disruption and (2) innovation. Disruption occurs when excessive crossover or mutation "disrupts" pertinent building blocks within solutions, which makes it impossible for GAs to assemble optimal solutions. Equation (4.7) presents the disruption boundary relationship developed in *Thierens* (1995) and used in the simple GA design methodology of *Reed et al.* (2000b).

$$P_c \leq \frac{s-1}{s} \quad (4.7)$$

The parameter  $s$  represents the total number of individuals that compete in tournament selection. Equation (4.7) is intended to protect pertinent building blocks from being destroyed due to excessive crossover. Since the child populations in the NSGA-II are selected using binary tournaments (i.e.,  $s = 2$ ), the probability of crossover must be set less than or equal to 0.5 or 50 percent. As the number of members that compete to survive [i.e., tournament size  $s$  in equation (4.7)] increases, selection pressure also increases because it becomes increasingly more difficult for unfit population members to survive. Equation (4.7) shows that the disruption boundary is less restrictive (i.e., higher  $P_c$  values can be used) when selection pressure is increased. The disruption boundary value of  $P_c \leq 50$  percent attained from equation (4.7) is conservative for the NSGA-II because the algorithm is guaranteed to always have higher selection pressure relative to a system undergoing pure binary tournament selection.

The potential disruptive effect of crossover must be balanced with the positive effect of this operator. *Thierens and Goldberg* (1993) show that selection and crossover combine to exert an innovative force on a system that guides GAs towards promising solutions. This force of innovation should be maximized by using the largest  $P_c$  possible below the upper bound defined by equation (4.7), which predicts solution instability for  $P_c$  near and above 50 percent. Because the methodology presented in this chapter is self-adaptive, if  $P_c = 0.5$  is initially too low for the NSGA-II to successfully mix or discover promising building blocks, then doubling the population size in a successive run would overcome this problem because both the population's diversity and the expected number of crossovers that the algorithm can use to successfully assemble optimal solutions would double. Doubling the population size from  $N$  to  $2N$  also increases the size of the selection pool from  $2N$  to  $4N$ , greatly increasing selection pressure

(designs must compete with far more individuals to survive) and decreasing the probability of disruption due to crossover and mutation.

The probability of mutation  $P_m$  is set using equation (4.8) based on the recommendations of the empirical studies of *De Jong* (1975) and *Schaffer et al.* (1989).

$$P_m \approx \frac{1}{N} \quad (4.8)$$

Selection and mutation combine to locally refine promising solutions. Equation (4.8) ensures that mutation is minimally disruptive.

#### **4.4.4 Step 3: Multi-Population Trial Runs**

Steps 1 and 2 of this design methodology provide all of the parameters required by the NSGA-II to perform trial runs. In a manner analogous to *Lobo* (2000) and *Reed et al.* (2000b), the NSGA-II automatically initiates trial runs for successively doubled population sizes. These trials runs use offline performance analysis to *sufficiently* quantify tradeoffs. From Step 1, tradeoffs are defined to be sufficiently quantified when the minimum percentage change in the number of individuals in the nondominated set for successive runs is less than 10 percent, defining the point of diminishing returns when successive increases in population size are no longer justified.

The design methodology presented in this chapter requires a minimum of two runs using two successively doubled population sizes to determine if the nondominated set has been sufficiently captured (i.e.,  $\Delta_{ND} \leq 10$  percent for this case). Figure (4.2a) shows the NSGA-II's performance for Run 1 using the lower bound population size of 60 members running for 40 generations. The figure shows that the NSGA-II is able to closely approximate 19 members of the nondominated set, most of which are “compromise solutions” in the center of the Cost—

SREE tradeoff. Figure (4.2a) illustrates the importance of allowing the NSGA-II to search for the nondominated set with small population sizes. Often practitioners are primarily focused on “compromise solutions” in the central region of tradeoffs for water resources applications because extreme solutions have a low likelihood of being implemented (although extreme solutions can provide insight into how individual objectives are affecting designs).

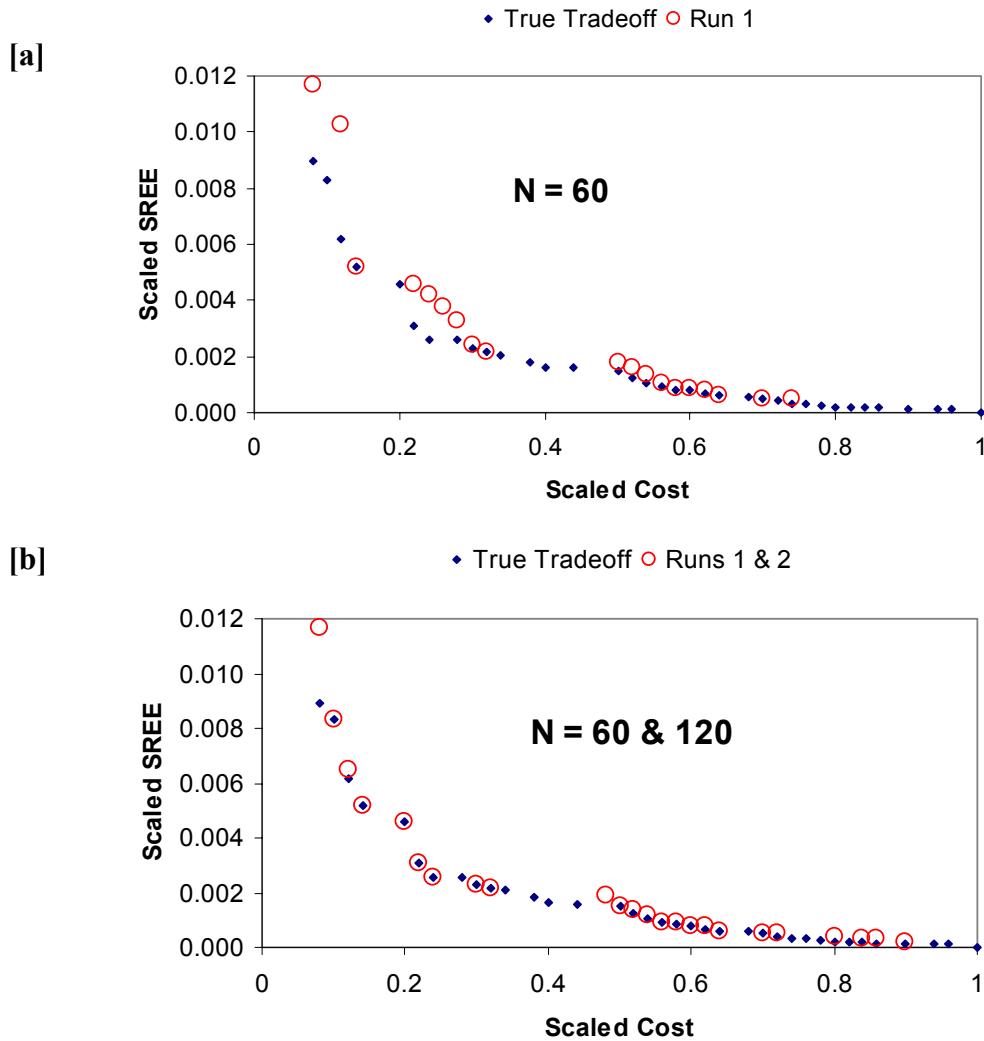


Figure 4.2 (a) Offline results for Run 1 (b) Combined offline results for Runs 1 & 2

Figure (4.2b) shows the combined offline results of Runs 1 and 2, in which all of the nondominated solutions found using  $N = 60$  are combined with those found using  $N = 120$  to

form an improved approximation of the Cost—SREE tradeoff. Combined offline analysis for multiple runs takes full advantage of every function evaluation required to solve an application and guarantees that the NSGA-II’s performance will only improve for successive runs. In Figure (4.2b), the nondominated set grew by 5 members relative to Run 1 shown in Figure (4.2a), a 26 percent increase.

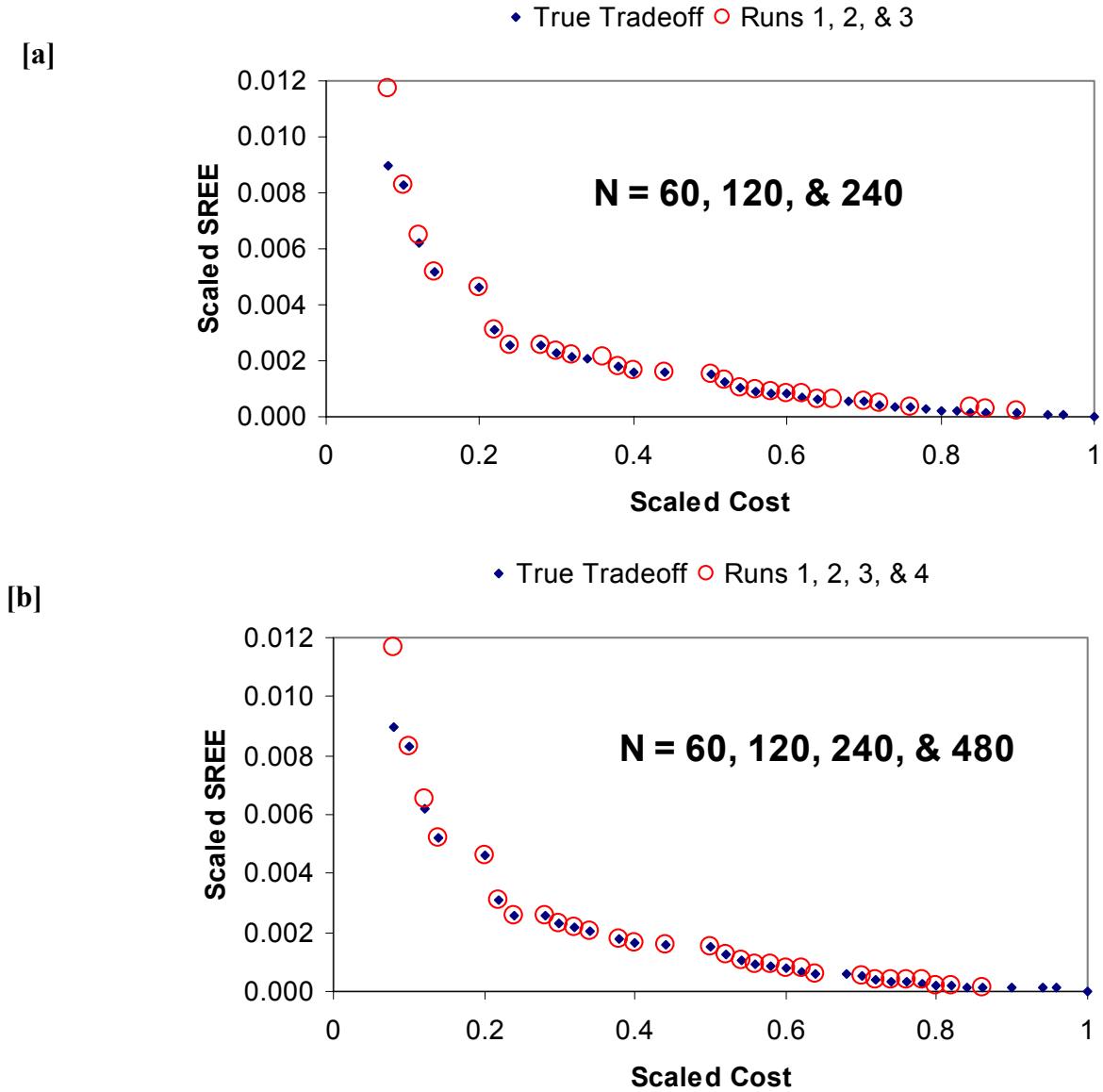


Figure 4.3 (a) Combined offline results for Runs 1, 2, & 3 (b) Combined offline results for Runs 1, 2, 3, & 4

Figure (4.2b) shows that the NSGA-II is more closely approximating solutions along the full extent of the Cost—SREE tradeoff; a third run is then initiated with a population size  $N = 240$  because  $\Delta_{ND}$  is less than 26 percent [see equation (4.4)]. Figure (4.3a) shows the combined offline results of Runs 1, 2, and 3 in which the NSGA-II identified 29 nondominated solutions. The figure shows that the NSGA-II has nearly captured the Pareto optimal set of solutions, but the tradeoff still has not been proven to be sufficiently quantified because the nondominated set increased by 20 percent [i.e.,  $\Delta_{ND} < 20$  percent in equation (4.4)]. An additional run using a population size of  $N = 480$  increased the nondominated set by only 1 solution, representing a 3 percent change. It is clear from both plots in Figure (4.3) that the point of diminishing return has been reached when subsequent increases in population size are no longer justified. The only task that remains for the user is to inspect the quality of the nondominated set quantified by the NSGA-II to ensure her or his expectations have been met successfully. If the NSGA-II does not satisfy the user’s expectations then a more exact approximation of the nondominated set can be sought by decreasing  $\Delta_{ND}$  and continuing to increase population size within the user-specified computational limits.

A significant advantage of this design methodology is that it requires only minimal initial user interaction, with successive runs completed automatically. Additionally, the design methodology of this chapter exploits the efficiency of the NSGA-II to substantially reduce computational costs. Consider that the NSGA design methodology of Chapter 3 required more than 160,000 function evaluations to meet the performance of the NSGA-II shown in Figure (4.3) for the same LTM application. The 4 trial runs discussed in this chapter required a total of 38,000 function evaluations, representing an 80 percent reduction in the computational costs relative to the previous NSGA design methodology.

## 4.5 Conclusions

The NSGA-II design methodology presented in this chapter builds on previous GA design methodologies (*Reed et al.* 2000, *Lobo* 2000, and *Reed et al.* 2001) to introduce a multi-population approach that automates parameter specification for the algorithm. The design methodology simplifies Pareto optimization using the NSGA-II into a simple 3 step process. The first step only requires users to answer 4 questions on their performance expectations and run time. Step 2 utilizes GA control map theory to automatically set the probabilities of crossover and mutation. The final step of the design methodology successively doubles population sizes until the NSGA-II has sufficiently identified the nondominated set or acceptable run time is exceeded. The methodology successfully solved the LTM design application using 80 percent fewer function evaluations than the NSGA design methodology presented in Chapter 3. The combined efficiency of the NSGA-II and the design methodology presented in this chapter allows more challenging higher order Pareto optimization problems (i.e., problems with more than 2 objectives) to be solved [see Chapter 6].



## 5 SPATIAL INTERPOLATION METHODS

### 5.1 Introduction

Chapters 3 and 4 focused on capturing monitoring design tradeoffs using genetic algorithms. This chapter provides guidance to practitioners in choosing a plume interpolation method for any groundwater application by explicitly addressing opposing theoretical perspectives, ease-of-implementation, and effectiveness. More specific to the goals of this dissertation, Chapter 5 provides guidance in choosing the best estimation technique for evaluating the LTM performance criteria examined in Chapter 6.

#### 5.1.1 *Motivation & Scope*

Plume interpolation consists of estimating contaminant concentrations at unsampled locations using the available contaminant data surrounding those locations. The goal of groundwater plume interpolation is to provide the best possible picture of the contaminant plume given the data limitations associated with sparse monitoring networks with irregular geometries. Beyond data limitations, contaminant plume interpolation has proven to be a difficult task because contaminant concentration fields are highly heterogeneous, anisotropic, and nonstationary phenomena (i.e., their mean and covariance vary widely across a given site). Contaminant concentrations commonly vary over several orders of magnitude within relatively short distances and in most cases are preferentially sampled in areas of high concentration. Preferential sampling and the highly skewed nature of groundwater contamination can combine to severely bias interpolation estimates and any subsequent decisions that must be made using these estimates (*Kitanidis* 1997, *Goovaerts* 1997, and *Chilès & Delfiner* 1999). Contaminant fate-and-transport simulation, inverse parameter estimation, site-characterization, and long-term monitoring (LTM) network design are a few examples of applications within which plume

interpolation plays a vital role (see *ASCE 1990b*, *Loaiciga* 1992, *Reed et al.* 2000 for more detailed reviews).

Historically, there have been two primary approaches to address the problem of contaminant plume interpolation. The first approach uses deterministic methods such as inverse distance weighted sums to estimate concentrations within the interpolation domain (*Gambolati & Volpi* 1979, *Gotway et al.* 1996). The second approach uses the stochastic framework of regionalized variable theory and geostatistics to obtain contaminant plume estimates *Matheron* (1965, 1971). *Chilès & Delfiner* (1999) define a *regionalized variable* "...to designate a numerical function [ $c(x)$ ] depending on a continuous space index  $x$ , and combining high irregularity of detail with spatial correlation".

Of the two interpolation approaches, geostatistics has been the primary focus of previous studies within the water resources area (*Cooper & Istok* 1988a,b,c; *ASCE* 1990a,b). Additionally, there are several recent texts available that provide an array of theoretical perspectives on the application of geostatistics to both stationary and nonstationary phenomena (*Isaaks & Srivastava* 1989, *Kitanidis* 1997, *Goovaerts* 1997, and *Chilès & Delfiner* 1999). However, there have been very few studies that compare both deterministic and geostatistical interpolation methods in the water resources literature (*Gambolati & Volpi* 1979 and *Hughes & Lettenmaier* 1981). Given that spatial structure identification in geostatistics applications improves markedly as the number of data increases, and that many sites have limited data sets, it is important to consider both deterministic and geostatistical interpolation methods. Selecting an appropriate method can be a difficult task for the practitioner because of opposing theoretical perspectives given in the interpolation literature. For example, some geostatisticians recommend the use of more traditional ordinary kriging (OK) within local neighborhoods (*Journel & Rossi*

1989, *Goovaerts* 1997) when interpolating nonstationary phenomena (e.g. contaminant concentrations), while others recommend more advanced methods such as intrinsic kriging (IK) (*Chilès & Delfiner* 1999).

The attractive properties of geostatistical approaches, relative to deterministic methods, are their ability to: (1) incorporate knowledge of the underlying spatial structure of the phenomenon of interest into “unbiased” estimates and (2) provide estimation variances. Estimation variances are considered to be measures of the quality of the estimates (thus a measure of “local” uncertainty). As discussed in the next section, several studies have shown that estimating nonstationary phenomena such as contaminant concentrations with limited data is subject to bias for all forms of kriging. Moreover, as the level of available data decreases for a given site the estimation variances attained using geostatistical approaches significantly degrade in their ability to represent the local uncertainty in estimates. Identifying an optimal method for plume interpolation is a particularly difficult task given both conflicting perspectives in the literature and the limited availability of comprehensive performance comparisons of deterministic and geostatistical methods in the literature.

*Gotway et al.* (1996) and *Crawford & Hergert* (1997) discuss these difficulties in the context of mapping nonstationary soil properties for a regularly gridded data set. This series of studies presents an empirical comparison of OK within local neighborhoods, kriging with a trend (KT) [also termed *universal kriging*], median polish kriging (a transformation-based kriging approach for regularly gridded data), and inverse-distance weighting. Although these studies did not make definitive recommendations as to which interpolation approach was the best, they present a careful synopsis of their relative benefits and limitations to help guide the practitioner in her or his application. This chapter extends their comparison of interpolation methods to

scatter-point (i.e., non-gridded data) groundwater plume interpolation and provides a comparison of the following interpolation approaches:

- Ordinary Kriging
- Multigaussian Kriging (MGK)
- Intrinsic Kriging (ItK)
- Quantile Kriging (QK)
- Nonlinear Least Squares Inverse Distance Weighting (NLS)
- Inverse Distance Weighting of Power 2 (ID).

This chapter compares plume interpolation methods as a function of data availability. Additionally, it illustrates the consequences of each method's underlying assumptions. The methods were selected solely for addressing the problem of estimating contaminant concentrations at unknown spatial locations at a single snapshot in time and not for spatio-temporal interpolation.

### **5.1.2 Previous Work in the Estimation of Nonstationary Phenomena**

The plume interpolation methods discussed in this chapter are based on the underlying premise that the values of contaminant concentrations at any two points become increasingly similar as the distance between the points decreases. *Gambolati & Volpi* (1979) presented a deterministic analysis of the geostatistical estimation method *kriging* and demonstrated its close relationship to inverse distance weighting and least squares methods because of this underlying premise. Studies comparing kriging methods and inverse distance-based interpolation schemes have generally found that the superiority of kriging is dependent on the number of available data, the level of anisotropy of the underlying phenomenon, and the presence of a trend (*Gambolati & Volpi* 1979, *Hughes & Lettenmaier* 1981, *Gotway et al.* 1996). These studies compared the

performance of inverse-distance methods with only one of the several possible forms of kriging and generally concluded that the primary strength of the geostatistical approach is that each estimate has a corresponding estimation variance (a measure of estimation error).

Kriging nonstationary data is a difficult task that has led to extensions of the original kriging estimator. However, there has been considerable debate about the relative merits of these extensions. The earliest forms of kriging, termed *simple kriging* (SK) and *ordinary kriging*, are based on an assumption of stationarity, first with respect to the mean and also with respect to the covariance or spatial increments (intrinsic hypothesis). The mean is assumed constant either throughout the entire domain or within local neighborhoods composing the domain, respectively. *Chilès & Delfiner* (1999) state that *Matheron* (1969) recognized the limitations of these assumptions and formulated the *kriging with a trend* (KT) approach for nonstationary phenomena.

The KT approach models the regionalized variable  $c(x)$  as the sum of a smoothly varying mean component  $m(x)$  termed a trend or drift and a second order stationary residual component  $r(x)$ . The KT estimator requires specification of a trend function to model  $m(x)$  and knowledge of the covariance structure of the residual component  $r(x)$ . Several studies have subsequently criticized the method primarily due to the "chicken-and-egg" (*Armstrong* 1984) conundrum practitioners face where the exact form of  $m(x)$  must be known to accurately model the spatial structure of  $r(x)$  while the exact form of  $m(x)$  cannot be determined without knowing the true spatial structure of  $r(x)$  (*Volpi & Gambolati* 1978, *Hughes & Lettenmaier* 1981, *Russo & Jury* 1987, *Journel & Rossi* 1989, *Crawford & Hergert* 1997). These studies show that the common practice of specifying  $m(x)$  as a second order polynomial can introduce significant bias in both

the kriged estimates and their corresponding estimation variances. For this reason, this method is not included in this chapter.

Recognizing the limitation of the KT approach, *Matheron* (1973) and *Delfiner* (1976) formulated the *intrinsic random function of order k* (IRF-k) theory for *intrinsic kriging* (see *Chilès & Delfiner* 1999, *Hughes & Lettenmaier* 1981). IRF-k theory was formulated to model nonstationary phenomena where the division of  $c(x)$  into  $m(x)$  and  $r(x)$  has no meaning. The ItK approach utilizes the concept of the *generalized increment of order k* in which a linear combination of sampled data is selected that will intrinsically filter a polynomial trend of order  $k$  without the need for explicitly specifying the trend as in the KT approach (*Chilès & Delfiner* 1999). Although IRF-k theory provides a theoretically elegant alternative to the KT approach, its application has been limited because determining the spatial structure of the underlying phenomena requires significantly more expertise (*Delfiner* 1976, *Kitanidis* 1983, *Marshall & Mardia* 1985, *Chilès & Delfiner* 1999).

*Hughes & Lettenmaier* (1981) compared the performance of ItK and traditional least squares approaches using hydrologic data and found that the relative superiority of ItK estimates is dependent on having a sample size greater than 50. Additionally, the study showed that for small data sets inference of spatial structure is subject to a bias that "...will generally lead the investigator to believe his [or her] estimates are more precise than they really are" (*Hughes & Lettenmaier* 1981). *Journel & Rossi* (1989) and *Goovaerts* (1997) recommend using OK within neighborhoods rather than KT and ItK when interpolating nonstationary natural phenomena. Note that these findings directly contradict the methodological recommendations given by *Chilès & Delfiner* (1999) who recommend the use of ItK in place of OK or KT for nonstationary phenomena.

In addition to nonstationarity, the distributional dependence of the estimation variance has led to additional kriging approaches. *Journel & Rossi* (1989) explicitly show that the estimation variance is a “model dependent ranking” of sampling configurations and is reflective of local accuracy only in cases where the studied phenomenon is multivariate Gaussian (i.e., normally distributed). The lognormal transform (i.e.,  $y = \ln[c(x)]$ ) is one of the most commonly applied transformations within water resources applications to try to normalize spatial phenomena such as transmissivities or contaminant concentrations (see *Cooper & Istok* 1988a and *ASCE* 1990a). Several issues require careful consideration when using the lognormal transform or log-kriging approach. The method implicitly assumes that the contaminant concentrations are multi-variate lognormal. *ASCE* (1990) state that this assumption cannot be verified using real-world data sets (i.e., data are often too sparse to even adequately evaluate the bivariate spatial distribution, let alone 3 or 4 point distributions). *Deutsch & Journel* (1998) warn that the back transform requires bias-correction to ensure  $y$  is an unbiased estimator of  $c(x)$ . Even with the bias-correction, the back-transform exponential function amplifies estimation errors (*Deutsch & Journel* 1998, *ASCE* 1990a). These drawbacks have led to a shift towards the transformations that are briefly described below.

Multigaussian kriging uses a normal score transformation of the sample data (i.e., mapping highly asymmetrical contaminant samples to a corresponding normal distribution) to improve its approximation of the multivariate Gaussian distribution with the intention of making estimation variances more reflective of the local accuracy of estimates (*Goovaerts* 1997). The assumptions and limitations of MGK are discussed in more detail in the methods section of this chapter. Other extensions of the OK system have used a variety of data transforms to formulate nonparametric (or distribution free) kriging systems that quantify local accuracy based on the

data's empirical cumulative distribution function (cdf) (*Journel* 1983, *Journel* 1988, *Chiueh et al.* 1997, *Journel & Deutsch* 1997, *Juang et al.* 2001). In addition to making the kriging approach distribution free, non-parametric kriging methods reduce estimation bias caused by highly variable phenomena with severely asymmetrical distributions such as contaminant concentrations. This chapter focuses on the rank-order transform (or quantile kriging) proposed by *Journel & Deutsch* (1998) and successively applied to a soil contamination application by *Juang et al.* (2001). As with MGK, the details of the QK approach are discussed in greater detail in the methods section of this chapter. Although transformation methods can render contaminant data to be more amenable to structure identification and kriging, *Goovaerts* (1997) warns that practitioners should carefully consider the effects of transformations before they apply them by paying close attention to any potential side effects (e.g. inflation of estimation error) incurred in such applications. The performance of each interpolation method considered in this chapter was explored using the three test cases described in the next section.

### **5.1.3 Test Cases**

The test case data used in this research are drawn from a 50 million-node flow-and-transport simulation (for more details see *Maxwell et al.* 2000). The simulation provided realistic historical data for a steady-state hypothetical perchloroethylene (PCE) plume in a highly heterogeneous alluvial aquifer. The hydrogeology of the test case is based on an actual site located at the Lawrence Livermore National Laboratory in Livermore, California, currently being managed under the United States' Comprehensive Environmental Response, Compensation and Liability Act (CERCLA) program. One of the objectives of this chapter is to show how each of the interpolation approaches performs as a function of the amount of available data. To satisfy this objective, the test case data from *Maxwell et al.* (2000) were used to

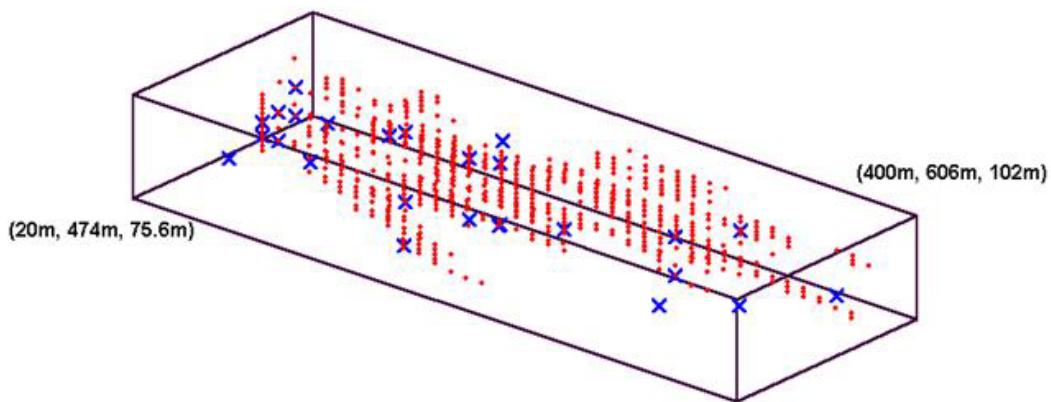
develop the three monitoring networks shown in Figure (5.1). The smallest network, shown in Figure (5.1a), consists of 20 monitoring wells that sample a total of 26 sampling locations within 1 million cubic meters of contaminated aquifer. The intermediate network, shown in Figure (5.1b), has 29 monitoring wells that sample 58 locations within 6 million cubic meters of contaminated aquifer. Finally, the largest network, shown in Figure (5.1c), has 59 wells that sample 124 sampling locations within 16 million cubic meters of contaminated aquifer.

	PCE Mean (mg/m <sup>3</sup> )	PCE Median (mg/m <sup>3</sup> )	PCE Coefficient of Variation
<b>Large Test Case</b>	29	1	11.4
<b>Intermediate Test Case</b>	164	11	3.6
<b>Small Test Case</b>	355	9	1.8

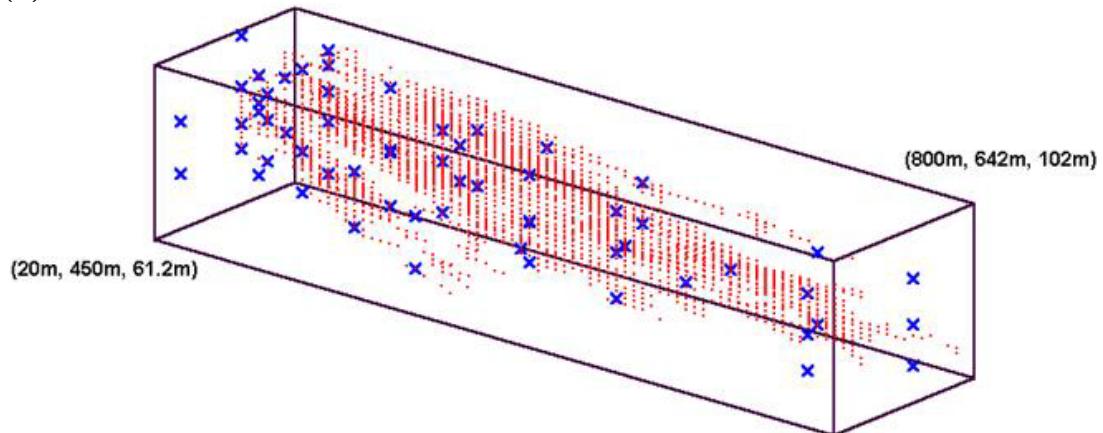
Table 5.1 Sample data statistics for each test case

This chapter makes the assumption that as the size and maturity of a contaminant plume increases, more monitoring points will be sampled for mapping contaminant concentrations. Table (5.1) presents the means, medians, and coefficients of variation computed using the PCE sample data attained from the test cases' monitoring networks. The statistics shown in Table (5.1) confirm that the sample data from these monitoring networks are both heavily skewed and highly variable. Moreover, Figure (5.1) shows that the scatter-point sampling locations are arbitrarily clustered in the body of plume, and preferentially sample the source area of the plumes. These test data sets are designed to be as representative as possible of real-world sites. Further, the use of the high resolution simulation data from *Maxwell et al.* (2000) enables comparison of the interpolation methods' performance at all estimated locations with the true values at these locations, while still using a realistic data set.

**(a) Small Test Case**



**(b) Medium Test Case**



**(c) Large Test Case**

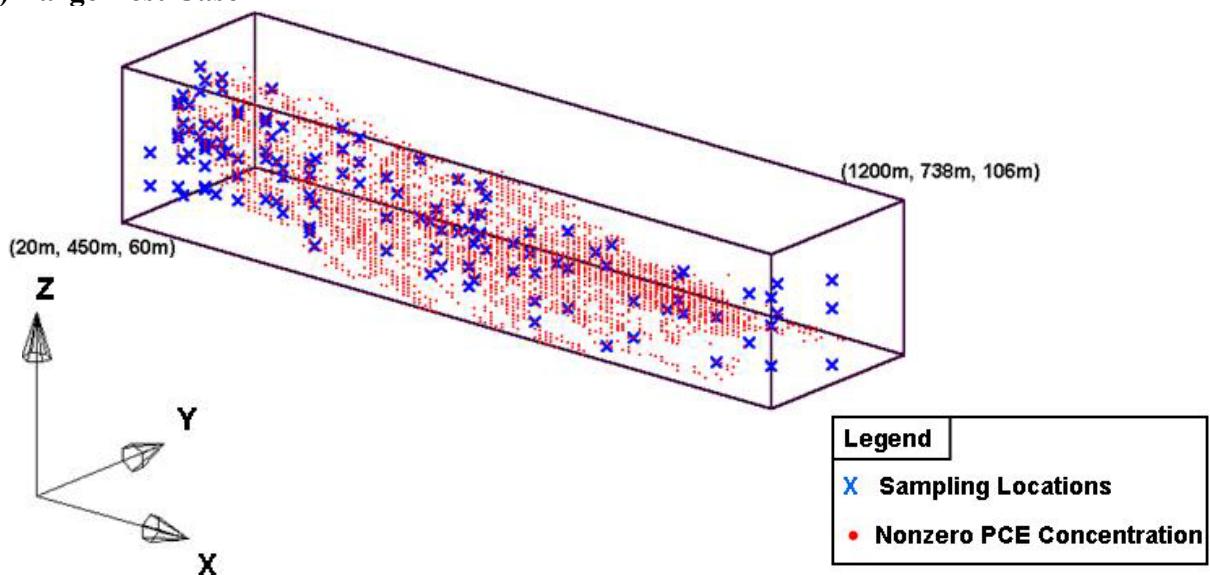


Figure 5.1 Test case monitoring networks

## 5.2 Methods

### 5.2.1 Geostatistical Methods

The geostatistical modeling approach requires both structural analysis and specification of a variogram model to represent the theoretical spatial structure. Structural analysis consists of analyzing available concentration data to determine the correlational structure. The analysis uses historical contaminant data to discretely approximate the *theoretical variogram*  $\gamma(\bar{h})$ , given in equation (5.1), where  $\bar{h} = \bar{x}_0 - \bar{x}_1$  is a vector of separation distances in each spatial dimension between two locations  $\bar{x}_0$  and  $\bar{x}_1$ .

$$\frac{1}{2} E[c(\bar{x}_0) - c(\bar{x}_1)]^2 = \gamma(\bar{h}) + [m(\bar{x}_0) - m(\bar{x}_1)]^2 \quad (5.1)$$

In equation (5.1),  $c(\bar{x})$  is a regionalized variable representing concentration that can be decomposed into its mean,  $m(\bar{x})$ , and random,  $r(\bar{x})$ , components. Assuming that the random component  $r(\bar{x})$  is stationary with a zero mean, half the expected value of the squared difference between concentrations at locations  $\bar{x}_0$  and  $\bar{x}_1$  (i.e.,  $1/2 E[c(\bar{x}_0) - c(\bar{x}_1)]^2$ ) simplifies to yield the sum of the theoretical variogram  $\gamma(\bar{h})$  and the squared difference between the mean concentrations at these two locations. The theoretical variogram  $\gamma(\bar{h})$  represents the average dissimilarity between two concentrations separated by a distance  $\bar{h}$ . In the OK approach, the mean component  $m(x)$  is assumed to be constant (or locally stationary), in which case the trend (or bias) term,  $[m(\bar{x}_0) - m(\bar{x}_1)]^2$ , in equation (5.1) is equal to zero and the theoretical variogram can be deduced directly from concentration samples. For nonstationary phenomena, the "chicken-and-egg" (Armstrong 1984) conundrum exists because the trend term is nonzero and a proper model for the theoretical variogram cannot be directly identified solely using

concentration data without making an assumption about the functional form of the trend. The next four sections of this chapter discuss different forms of the kriging system and describe the implicit assumptions that the practitioner must make when using each of the methods to interpolate nonstationary contaminant concentrations.

#### 5.2.1.1 Ordinary Kriging within Local Neighborhoods

The OK estimator is formulated to provide the Best Linear Unbiased Estimate (or BLUE) for contaminant concentration at an unsampled location (for a detailed derivation see *Kitanidis* 1997 or *Goovaerts* 1997). The OK approach invokes the *intrinsic hypothesis*, which assumes that the mean or trend terms given in equation (5.1) are locally constant but unknown within neighborhoods surrounding the current unsampled location (i.e., the trend is modeled using a zero<sup>th</sup> order polynomial). *Crawford & Hergert* (1997) state that, theoretically, locally defined neighborhoods of sample points should be used to discretely approximate the theoretical variogram [ $\gamma(\bar{h})$  in equation (5.1)]. *Journel & Rossi* (1989) recommend using either data perpendicular to the general trend or local neighborhoods with radii defined to have lengths equal to 25 percent of the total length of the interpolated domain for approximating the variogram. The local approximation of the theoretical variogram is not done in practice because the limited number of data available in most applications would make discrete or experimental variograms unreliable (*Crawford & Hergert* 1997, *Goovaerts* 1997).

The OK approach assumes that the estimation error is normally distributed when computing a confidence interval for every estimate. Recall that the estimation variance is only a function of the theoretical model used to represent the variogram and the geometrical configuration of the sampling data (i.e., a function of the kriging weights only). *Journel & Rossi* (1989) showed that estimation variance is only directly reflective of true estimation error when

the underlying phenomenon is multivariate Gaussian. They also showed that asymmetric data distributions can significantly bias estimates of local uncertainty. The next two sections describe transformation-based kriging approaches that preserve the relative ranks and spatial structure of contaminant data while addressing estimation inaccuracies caused by highly asymmetric data distributions.

#### 5.2.1.2 Multigaussian Kriging within Local Neighborhoods

The MGK approach to estimation utilizes the normal score transform of the sample data and models the contaminant concentrations as a multivariate Gaussian random function,  $T(\bar{x})$  (for details on the transform see *Goovaerts* 1997). The normal score (NSCORE) transform maps the asymmetrical discrete sample distribution to a  $N(0,1)$  normal distribution. *Deutsch & Journel* (1998) provide several programs that will perform all of the required steps for the transformation in the GSLIB software library. In the MGK approach, estimation is performed using OK on the transformed data, again assuming a zero<sup>th</sup> order polynomial trend. The NSCORE estimates can then be back-transformed into concentration-space by using the same mapping discussed above.

The multivariate Gaussian assumption should be carefully validated using the methods described by *Goovaerts* (1997) because it has direct consequences on the estimation process. This transformation assumes extreme values have reduced correlation in space (relative to median values). In plume interpolation, often the extreme concentrations are pivotal in properly assessing risks and uncertainty for site management decisions. Assuming a reduced correlation structure for these extreme values can significantly affect risk and uncertainty calculations and lead to serious misrepresentations of the site conditions if the Gaussian assumption is inappropriate.

### 5.2.1.3 Quantile Kriging within Local Neighborhoods

The distributional dependence of OK and MGK, as well as their sensitivity to highly skewed data, led to the development of non-parametric or distribution-free kriging approaches using indicator (*Journel* 1983) or rank-order (*Journel & Deutsch* 1997) transformations of the data. Indicator kriging uses a binary transformation of sample data where the data value is set to either 1 or zero depending on whether it is greater than or equal to a user-specified threshold value. The indicator approach requires  $n_q$  threshold indicator transforms, each of which requires its own variogram and kriging system. *Goovaerts* (1997) states that at least five thresholds are required (i.e.,  $n_q = 5$ ) to quantify the local uncertainty of indicator kriging estimates. Correctly quantifying variograms for extreme concentration thresholds is generally not possible in plume interpolation applications due to data limitations (see *Chilès & Delfiner* 1999). Moreover, solving the OK system of equations for each of the thresholds for every point in the interpolation domain represents a severe computational limitation of the method.

These limitations motivated the development of quantile kriging (*Journel & Deutsch* 1997), which transforms concentrations into standardized ranks (or quantiles) using equation (5.2) and computes estimates in quantile space using OK (see *Journel & Deutsch* 1997 and *Juang et al.* 2001).

$$c(\bar{x}) = \frac{\text{rank of sample } i}{N + 1} \quad (5.2)$$

The transform consists of ranking the concentration data in ascending order and dividing these ranks by one plus the total number of sample data  $N$  (for cases with multiple zero measures see the frequency analysis discussion in Appendix D). Figure (5.2) graphically illustrates quantile or standardized ranks transform. The  $i^{th}$  sample is assigned the probability that  $c(\bar{x})$  is less than or equal to its concentration value [ $F(c)$  in Figure (5.2)]. Note that quantiles [i.e., the probabilities

from the empirical cdf  $F(c)$ ] are known to be uniformly distributed; this fact can be used in conjunction with the estimation variances to estimate the local uncertainty at unsampled locations in the interpolation domain. The empirical cdf of the sample data is also used to back transform estimates from quantile space to concentration space (for details see *Juang et al.* 2001).

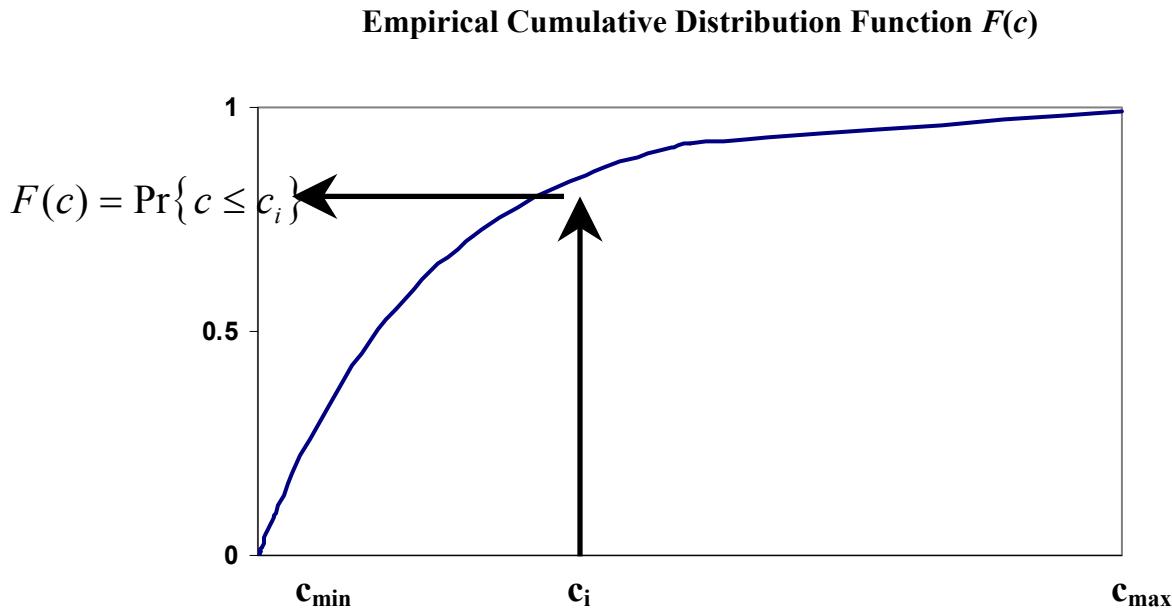


Figure 5.2 Illustration of rank-order transformation used in quantile kriging

This kriging approach has the following limitations: (1) it assumes that the empirical (or sample) cdf attained from available data is reflective of the true distribution of concentrations throughout the interpolation domain and (2) the approach probabilistically models a deterministic quantile space. As with any probabilistic statistic, the quality of the empirical cdf as a representation of the true concentration distribution in space varies as a function of the number of data (i.e., given infinite samples the empirical cdf should in theory approach the true concentration cdf). Note that regardless of its theoretical underpinnings, the primary criteria for evaluating geostatistical approaches such as quantile kriging is how well they work in practice.

This is in line with *Gambolati & Volpi* (1979), who show that the geostatistical approach can be viewed as a deterministic interpolation scheme.

Quantile kriging and MGK are both extensions of the OK approach, which assumes that the trend term in equation (5.1) is modeled as a zero<sup>th</sup> order polynomial. The next section discusses both the KT and ItK approaches, which generalize the kriging system to consider higher order polynomial trends.

#### 5.2.1.4 *Intrinsic Kriging*

The ItK approach was motivated by perceived theoretical shortcomings in both the OK and kriging with a trend (KT) systems. The KT and ItK approaches are closely linked, making it necessary to set the stage for the ItK description with a brief discussion of the KT method and its shortcomings. This is particularly relevant since KT is still actively applied in practice. The KT approach models concentrations as a regionalized variable,  $c(\bar{x})$  that is composed of the sum of a smoothly varying mean (trend) component  $m(\bar{x})$  and a second order stationary residual  $r(\bar{x})$ . The KT estimator for contaminant concentration requires the specification of a linear combination of known polynomial functions of spatial coordinates to model the trend component of the concentration [for a detailed derivation see *Goovaerts* 1997]. Given the assumed trend function, the residual of the concentration data and the trend are calculated and structural analysis of the residual is used to derive a residual covariance function  $K_r(\bar{h})$ . Inference of the residual covariance or variogram is difficult because the concentration data are not directly reflective of the residual function. *Goovaerts* (1997) recommends extending the recommendations of *Journel & Rossi* (1989) for the OK system to the KT approach by using either data perpendicular to the general trend or local neighborhoods with radii defined to have lengths equal to 25 percent of the total length of the interpolated domain. Again, this approach is

not feasible for most real-world contaminant data sets due to data limitations. The OK system is a special case of the KT system where the trend is assumed to be a zero<sup>th</sup> order polynomial and is set equal to 1. *Goovaerts* (1997) shows that the only difference between the OK and KT systems results from the practitioner's arbitrary decision to explicitly model the local trend as either a constant or a higher order polynomial.

The KT and variations of OK discussed above all require the practitioner to accept the dichotomy of the regionalized variable representing contaminant concentrations into a smoothly varying mean component and a stochastic fluctuation. Several studies have shown that this assumption is often an arbitrary choice that can significantly bias both estimates and their respective estimation variances (*Volpi & Gambolati* 1978, *Hughes & Lettenmaier* 1981, *Russo & Jury* 1987, *Journel & Rossi* 1989, *Crawford & Hergert* 1997). Intrinsic random function of order k (IRF-k) theory avoids the above dichotomy by not requiring the explicit specification of a model for the mean trend. Instead, the ItK approach defines allowable linear combinations of the sample data, termed *generalized increments*, that are implicitly capable of filtering trends from the data, facilitating a more direct measure of the underlying spatial structure of contaminant concentrations. The approaches discussed in previous sections model spatial structure using pairs of sample data to deduce the variogram shown in equation (5.1). Alternatively, IRF-k theory defines *generalized increments of order k* to be  $(k+1)$  point increments that are capable of filtering polynomial trends of degree less than or equal to  $k$  (*Chilès & Delfiner* 1999). Generalized increments, as the name implies, generalize the inference of spatial structure from a 2-point to a  $(k+1)$  point measure of spatial correlation termed the *generalized covariance*  $G(\bar{h})$ .

In the ItK approach, concentration is modeled as an IRF-k where the generalized increments of  $c(\bar{x})$  are assumed to be a stationary random function with a zero mean and a covariance equal to  $G(\bar{h})$ . It has been shown that the ItK system is identical to the KT system except for the substitution of the generalized covariance function  $G(\bar{h})$  for the 2-point residual covariance function  $K_r(\bar{h})$  (for details see *Chilès & Delfiner* 1999). The ItK method was formulated to avoid the estimation biases caused by the explicit specification of a functional trend model. Although the IRF-k approach elegantly unifies geostatistical theory, ItK requires significant expertise to implement relative to the more traditional approaches discussed above. Also, given the common occurrence of limited data sets with irregular sample network geometries for plume characterization, inference of a model for the theoretical generalized covariance function is a difficult process that again introduces bias into estimates and their corresponding estimation variances (see *Hughes & Lettenmaier* 1981 and *Goovaerts* 1997).

With scatter-point data, the traditional graphical approaches used to identify variograms cannot be used for ItK and instead identification of the appropriate generalized covariance model requires the use of “black-box” automated fitting procedures (see *Delfiner* 1976, *Kitanidis* 1983, *Marshall & Mardia* 1985, *Chilès & Delfiner* 1999). In this chapter, the *Delfiner* (1976) method was implemented using the Geovariances ISATIS version 3.4.0 advanced geostatistical software package. *Chilès & Delfiner* (1999) state that this approach “...has the advantage of being completely general, since it does not require the data points to be evenly distributed nor the underlying random function [i.e., contaminant concentrations] to be Gaussian”. The method consists of the following tasks: (1) fitting a polynomial of degree less than or equal to  $k$  and (2) identifying the generalized covariance model with the best cross-validation scores. Cross-validation is performed by removing a “known” data point and estimating its values using the

remaining sample data. Residuals can then be computed as the difference between the data point's true value and its estimate (for more details see the discussion of equation (5.4) in the next section). Readers interested in the details of this fitting algorithm should see both *Delfiner* (1976) and *Chilès & Delfiner* (1999).

### 5.2.2 Deterministic Methods

Traditional inverse-distance weighting and least squares approaches have been compared to the kriging methods described above. The deterministic methods do not incorporate the underlying spatial structure of the phenomenon of interest and have generally been shown to be more sensitive to asymmetrical data distributions and bias-based estimation error (*Gambolati & Volpi* 1979, *Hughes & Lettenmaier* 1981, *Gotway et al.* 1996). These studies have also found that the relative superiority of kriging approaches is largely a function of the number of available data because, below a certain threshold, spatial structure (i.e., the spatial covariance) cannot be determined.

This chapter considers an array of plume interpolation methods that will enable the practitioner to define the most appropriate estimation technique given the data limitations of his or her site. Deterministic estimation techniques can be used for LTM design applications for smaller sites with limited data sets. *Reed et al.* (2000) also found that the inverse distance weighting techniques are excellent screening tools, which can be used to justify the added cost and analyses required for geostatistical approaches to LTM design. The next two sections present the deterministic interpolation methods considered in this chapter.

#### 5.2.2.1 Nonlinear Least Squares Inverse Distance

The nonlinear least squares (NLS) method studied in this research is a variant of the scheme used by *Barry & Sposito* (1990) in their analysis of tracer plumes at the Borden site

located in Ontario, Canada. The *Barry & Sposito* (1990) interpolation method shown in equation (5.3) was selected because it requires minimal modeling assumptions and it has been successfully applied to three-dimensional historical contaminant data.

$$w(\bar{x}_j, \bar{x}_k) = 1/D(\bar{x}_j, \bar{x}_k)$$

$$D(\bar{x}_j, \bar{x}_k) = [ (x_{j1} - x_{k1})^2 / \alpha_1 + (x_{j2} - x_{k2})^2 / \alpha_2 + (x_{j3} - x_{k3})^2 / \alpha_3 ]^P$$

$$c_{est}(\bar{x}_j) = \sum_{\psi=1}^N w(\bar{x}_j, \bar{x}_k) c(\bar{x}_k) / w_t, \forall j \quad (5.3)$$

In equation (5.3), the  $j^{th}$  concentration estimate  $c_{est}(\bar{x}_j)$  at location  $\bar{x}_j$  is computed as a weighted sum of the available sample data. Each weight  $w(\bar{x}_j, \bar{x}_k)$  is computed as an inverse function of the distance between the  $j^{th}$  unsampled location and  $k^{th}$  sample raised to the power  $P$ . The alpha parameters in equation (5.3) scale the relative importance of each spatial dimension in the distance calculations. Equation (5.4) presents the cross-validation based parameter estimation method that was used to fit the  $\alpha$  parameters and distance exponent  $P$  shown in equation (5.3) to the contaminant data of each test case (*Barry & Sposito* 1990).

$$\text{Minimize } S = \sum_{k=1}^N \left\{ c(\bar{x}_k) - \sum_{i=1, i \neq k}^N w(\bar{x}_k, \bar{x}_i) c(\bar{x}_i) / (w_t - w(\bar{x}_k, \bar{x}_i)) \right\}^2 \quad (5.4)$$

The cross-validation method minimizes the sum of the squared residuals  $S$  between each of the actual PCE concentrations  $c(\bar{x}_k)$  and their interpolated estimates based on the  $(nsamp - 1)$  remaining data  $c(\bar{x}_i)$ . Equation (5.4) was solved using the Levenberg-Marquardt nonlinear least squares solution method (Moré 1977).

#### 5.2.2.2 Inverse Distance Power-2

The primary strength of inverse distance-based interpolation is its numerical simplicity. Equation (5.3) describes the inverse distance power-2 approach (ID) when the alpha fitting

parameters are set equal to 1 and the distance exponent  $P$  is set equal to 2. Additionally, equation (5.3) shows that ID is simply a matrix multiplication of the inverse distance weights  $w(\bar{x}_j, \bar{x}_k)$  by the vector of  $N$  sampled concentrations. Traditional inverse-distance estimation within local neighborhoods was tested in this chapter because it is commonly used in practice. Inclusion of this method enabled a comprehensive empirical performance analysis of interpolation methods ranging from advanced geostatistical estimators (i.e., the ItK approach) to one of the simplest interpolation methods used in practice.

### 5.2.3 Defining Estimation Neighborhoods

All of the methods considered in this chapter except for ItK and NLS require the specification of a neighborhood within which the sample data used to estimate concentration values are selected. Specification of an estimation neighborhood is particularly important for scatter-point contaminant data due to the ubiquity of preferential sampling at contaminated sites. Specifically, preferential sampling of areas of high concentration can lead to a bias that causes severe overestimation at unknown locations. In this chapter, each estimator used ellipsoid neighborhoods where, for each case, the axes were defined to be equal to one half of plume lengths along its transverse and longitudinal axes as recommended by *Cooper & Istok* (1988a) and *Goovaerts* (1997).

To reduce clustering effects, each of the ellipsoid neighborhoods were subdivided into octants within which a minimum of 1 and a maximum of 3 of the nearest sampled concentrations are selected to be used in estimation as recommended by *Chilès & Delfiner* (1999). This means that a maximum of 24 points were used to estimate concentrations at unknown locations. If there are no sample data within the estimation neighborhood surrounding an unknown location, then no estimate is computed for that location. The definition of the estimation neighborhood requires

that a balance be struck between the accuracy of estimates and the computational demands required to attain those estimates. For kriging systems, the computational complexity grows as a cubic function of the number of data used for estimation (e.g. doubling the number of data included in the neighborhood results in an 8-fold increase in the computational time required to solve the problem). Inverse distance approaches have the advantage that their computational complexity grows as a quadratic function of the number of data used in estimation. Each test case had a unique estimation neighborhood. To enable fair comparisons among the estimation methods, the same estimation neighborhood was used for all methods applied to a given test case.

#### **5.2.4 Performance Measures**

The performance of the interpolation methods was assessed using the following measures of their estimation errors: estimation error residual, standardized estimation error, confidence intervals, and root-mean-square error (RMSE). Each measure is summarized separately below.

##### *5.2.4.1 Estimation Error Residual*

The estimation error residual is shown in equation (5.5), which directly measures the difference between true and estimated concentration values. In this work estimation error residuals can be exactly calculated from the true simulation data. In practice, the evaluation of error residuals is approximated by using re-estimation scores, such as with cross-validation.

$$E_{rsd}(\bar{x}) = c_{true}(\bar{x}) - C_{est}(\bar{x}) \quad (5.5)$$

The kriging methods and NLS both seek to minimize these residuals to produce minimum error estimates of contaminant concentrations.

#### 5.2.4.2 Standardized Estimation Error

In addition to computing minimum error estimates, kriging methods also produce estimation variances that can be used to measure the “local uncertainty” of the estimates. The term “local uncertainty” is taken from *Goovaerts* (1997) because uncertainty estimates are computed considering only a single point in the interpolation domain. Estimation variances are used to *standardize* error residuals by dividing the errors computed at each location  $\bar{x}$  [shown in equation (5.5)] by the square root of its corresponding estimation variance, as shown in equation (5.6). Equation (5.6) was used to assess the cross-validation performance of each kriging method except QK by classifying estimates as being robust when their *standardized estimation error* falls within the interval [-2, 2], which is equivalent to a 95<sup>th</sup> percentile confidence interval for the normal distribution.

$$E_{Std}(\bar{x}) = \frac{c_{true}(\bar{x}) - c_{est}(\bar{x})}{\sigma(\bar{x})} \quad (5.6)$$

#### 5.2.4.3 Confidence Intervals

A *robust* estimate can be defined as one in which its true value falls within the 95<sup>th</sup> percentile confidence interval surrounding its estimate, which can be computed using equation (5.7).

$$c_{est}(\bar{x}) \pm 2\sigma(\bar{x}) \quad (5.7)$$

This uncertainty calculation makes the following assumptions: (1) that the estimation errors are normally distributed at location  $\bar{x}$  and (2) that  $c_{est}(\bar{x})$  and  $\sigma(\bar{x})$  represent the distribution mean and standard deviation and are independent. *Goovaerts* (1997) argues that these assumptions are “stringent” and rarely satisfied in practice due to highly asymmetrical sample distributions and spatially correlated estimation errors. Normality assumptions are avoided in the QK approach

because standardized ranks are known to be uniformly distributed allowing confidence intervals to be computed using equation (5.8) [see *Juang et al.* 2001].

$$c_{est}(\bar{x}) \pm \sqrt{3}\sigma(\bar{x}) \quad (5.8)$$

#### 5.2.4.4 Root-Mean-Square Error

Finally, the root-mean-square error (RMSE), shown in equation (5.9), is the last performance measure that was used to rank the relative performance of the estimation methods.

$$RMSE = \sqrt{\frac{1}{n_{est}} \sum_{j=1}^{n_{est}} [c_{est}(\bar{x}) - c_{true}(\bar{x})]^2} \quad (5.9)$$

RMSE represents the average error computed for  $n_{est}$  estimates.

## 5.3 Results & Discussion

The next sections compare the relative performances of the interpolation schemes as the number of available sample data decreases. The geostatistical approaches required both structural analysis and cross-validation to specify models of the spatial “structure” or correlation of the contamination data. Each test case required the specification of 4 variogram models, one for each of the kriging approaches (yielding a total of 12 models among the 3 test cases), which were then cross-validated using analyses of the histograms of standard errors, plots of estimates versus their true values, and measures of estimate robustness (for detailed presentation of these results see Appendices B and C). Additionally, NLS minimizes cross-validation squared residuals to specify the fitting parameters of equation (5.3). Table (5.2) gives the NLS fitting parameters’ values for each test case.

	$(\alpha_1, \alpha_2, \alpha_3)$	$P$
<b>Large Test Case</b>	(1.8, 0.1, 0.4)	30
<b>Intermediate Test Case</b>	(306, 304, 305)	0.5
<b>Small Test Case</b>	(36, 60, 0.5)	1.2

Table 5.2 Nonlinear least squares fitting parameter values

In addition to cross-validation scores, this chapter evaluates the accuracy of estimated PCE concentrations relative to their *true* values using the simulation data from *Maxwell et al.* (2000). For each test case, the interpolation schemes are ranked based on their RMSE values as well as a detailed analysis of the spatial distribution of estimation errors using three-dimensional cross sections along the vertical and horizontal axes of the each plume. These cross-sectional views provide the most comprehensive means of visualizing the spatial distribution of estimation errors. Finally, the discussion of each test case concludes with an analysis of how well the *local uncertainty* estimates attained from the kriging approaches predict areas of high estimation error.

### **5.3.1 Large Test Case**

The large test case represents a heavily sampled, mature PCE plume that extends over 1-km in the primary direction of groundwater flow within a highly heterogeneous alluvial aquifer. In this test case, 124 sample locations were available from 59 multi-level monitoring wells for the structural analysis required for each kriging system. Variogram models for the OK, MGK, and QK schemes were deduced using the standard trial-and-error approach of graphically fitting candidate models to experimental variograms followed by cross-validation to determine the models' performances.

The estimates from these kriging approaches were significantly improved when *geometric* anisotropies where incorporated into their variogram models (for detailed discussions of anisotropic variogram modeling see *Chilès & Delfiner* 1999 or *Goovaerts* 1997). These anisotropic models indicate that the PCE has shorter correlation ranges in the y- and z-directions (or in the transverse directions of the plume) relative to the x-direction (or along the longitudinal axis of the plume). Moreover, the vertical correlation range was shorter relative to both of the ranges specified in x- and y-directions as expected in an alluvial aquifer. For ItK, the second

order polynomial trend function ( $1+z+z^2$ ) and an isotropic generalized covariance model were specified by black-box fitting (*Delfiner* 1976). Due to both computational limits and improved cross-validation performance, ItK estimates were computed using local neighborhoods. The next section overviews the cross-validation results for the kriging approaches and NLS.

#### 5.3.1.1 Cross-Validation

Figure (5.3) shows the RMSE values attained from cross-validation for 120 of the 124 available sample data points. Four PCE concentrations, each of which exceeded 1000-mg/m<sup>3</sup> and were located in the source area of the plume are not included because these extremes severely impeded structural analysis for both OK and ItK. The calculation of the experimental variogram is very sensitive to extreme concentration values and in this case no spatial structure could be identified when these values where kept in the data set. The sensitivity of both variogram calculations and the kriging estimators to highly variable, strongly skewed concentration data is a motivating factor for the use of MGK and QK.

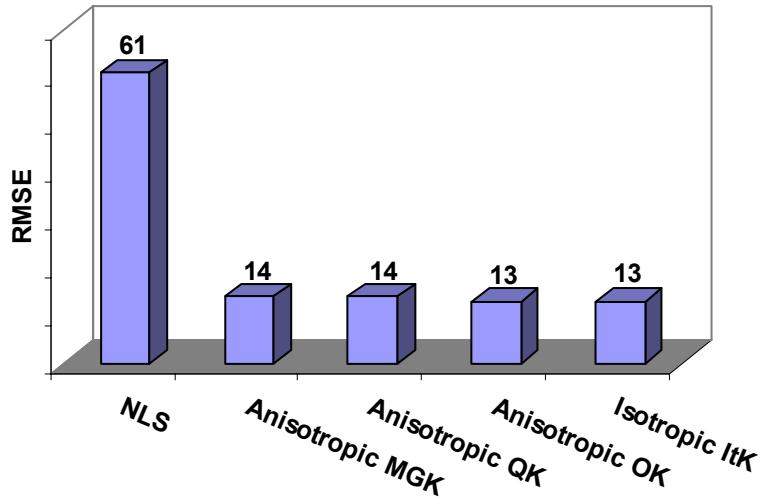


Figure 5.3 RMSE values from cross-validation for the large test case (mg/m<sup>3</sup>)

The cross-validation results for MGK and QK presented in Figure (5.3) exclude the 4 largest samples to allow these methods to be compared to OK and ItK. However, these samples

were included in both the structural analysis and cross-validation used to specify their variogram models. Figure (5.3) shows that there is only a slight difference between the RMSE values for all of the kriging methods. These cross-validation scores must be considered cautiously. Although OK and ItK have slightly smaller RMSE values, these methods could not utilize the full data set in their cross-validation estimates, in particular the maximum concentration values that can be important in plume interpolation. Moreover, 80-percent of the monitoring wells in this test case sample multiple locations in space, a majority of which fall within 10-meters of each other. The close proximity of these sampling locations to one another greatly reduced the mean distance between unknown and known locations in the cross-validation calculations. The magnitudes of cross-validation RMSE scores are directly proportional to the mean distance between known and unknown locations and in this case are underestimated.

The RMSE values in Figure (5.3) clearly show that NLS has the highest expected error of the five methods discussed in this section. Again, caution must be exercised before drawing the conclusion that NLS is the worst of the methods. Kriging is a *low-pass filter* that by means of statistical expectation (or averaging) tends to “filter” extreme values from its estimates. Note that the distance exponent  $P$  shown in Table (5.2) for the large test case is equal to 30, which causes NLS to behave essentially as a nearest neighbor estimator. This yielded a high degree of variability in the NLS estimates. The actual quality of NLS estimates are a function of how well the 124 concentration samples represent the true site conditions, which will be captured by the estimation accuracy analysis below.

The four kriging approaches were also assessed in terms of the robustness of their estimates (i.e., the percentage of the estimates with standardized errors falling in the interval [-2, 2]). Conservatively, over ninety-eight percent of the MGK and QK estimates were robust. Only

90-percent of the OK and ItK estimates were classified as being robust due to their sensitivity to extremes, even without the 4 highest PCE samples.

Overall, cross-validation scores failed to identify which interpolation method is superior. This failure resulted due to clustered, multi-level monitoring points that preferentially sample zones of high concentration causing the cross-validation RMSE to be significantly underestimated, especially for the interpolation methods that are sensitive to extreme PCE concentrations.

#### 5.3.1.2 Estimation Accuracy

The accuracy of interpolation estimates were tested relative to their *true* values for every point in the interpolation domains. Table (5.3) gives the grid dimensions of the interpolation domains.

	<b>(xmin, ymin, zmin)</b>	<b>(Δx, Δy, Δz)</b>	<b>(nx, ny, nz)</b>
<b>Large Test Case</b>	(20m, 450m, 60m)	(20m, 24m, 4.8m)	(59, 12, 9)
<b>Intermediate Test Case</b>	(20m, 474m, 61.2m)	(20m, 24m, 4.8m)	(39, 8, 8)
<b>Small Test Case</b>	(20m, 486m, 75.6m)	(10m, 12m, 4.8m)	(38, 11, 5)

Table 5.3 Interpolation grid specifications for each test case

Figure (5.4) gives the RMSE over the 6732 grid points in the large case domain for each interpolation method. The figure clearly shows the relative performances of the 6 interpolation methods considered in this chapter. The cross-validation RMSE values shown in Figure (5.3) predicted that the 4 kriging approaches should perform equally well, whereas Figure (5.4) shows that ItK and OK clearly fail to estimate PCE concentrations as accurately as QK and MGK. QK performed the best, with an absolute average error for PCE estimates of 8-mg/m<sup>3</sup>.

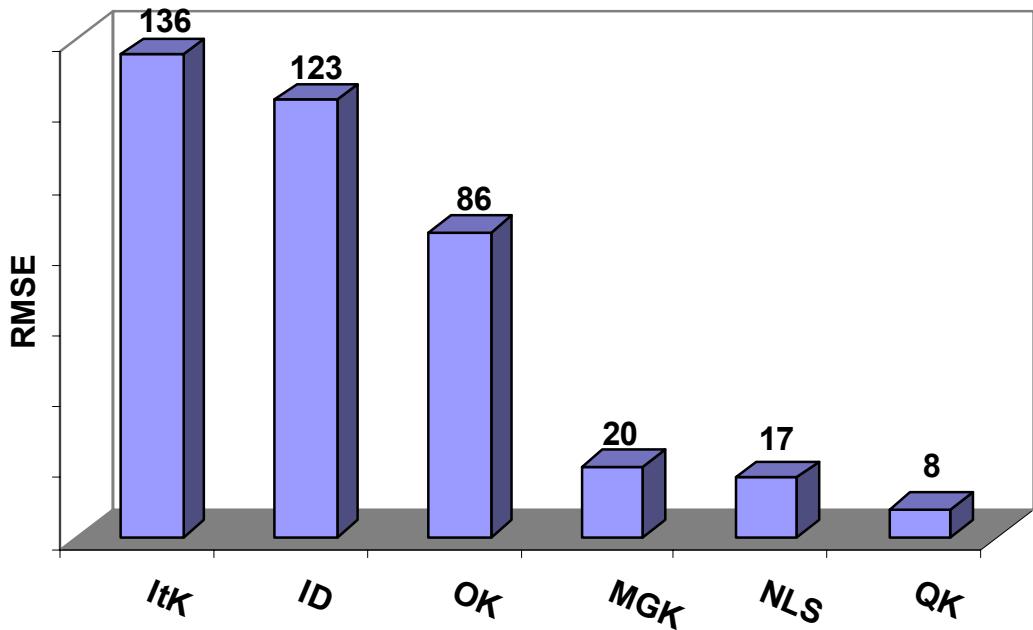


Figure 5.4 RMSE values from interpolation for the large test case ( $\text{mg}/\text{m}^3$ )

Figure (5.5) shows how the estimation errors are distributed throughout the interpolation domain for each of the interpolation methods shown in the order from the most to the least accurate approach. Figure (5.5a) shows that almost all of the QK estimates fell within  $\pm 10\text{-mg}/\text{m}^3$  of their true values. The figure shows that the zone of highest error occurs in the midsection of the plume centered at about 600-m in the x-direction. There are two primary factors contributing to the errors in this portion of the plume: (1) there is a reduced number of sample data available in this portion of the plume and (2) a group of monitoring wells clustered in this portion of the plume had PCE concentrations greater than  $50\text{-mg}/\text{m}^3$  but were not representative of the concentrations in this portion of the plume. Overall, the QK approach was fairly robust given the relatively poor spatial coverage of the monitoring wells in the midsection of the plume and preferential sampling of high concentrations throughout the plume.

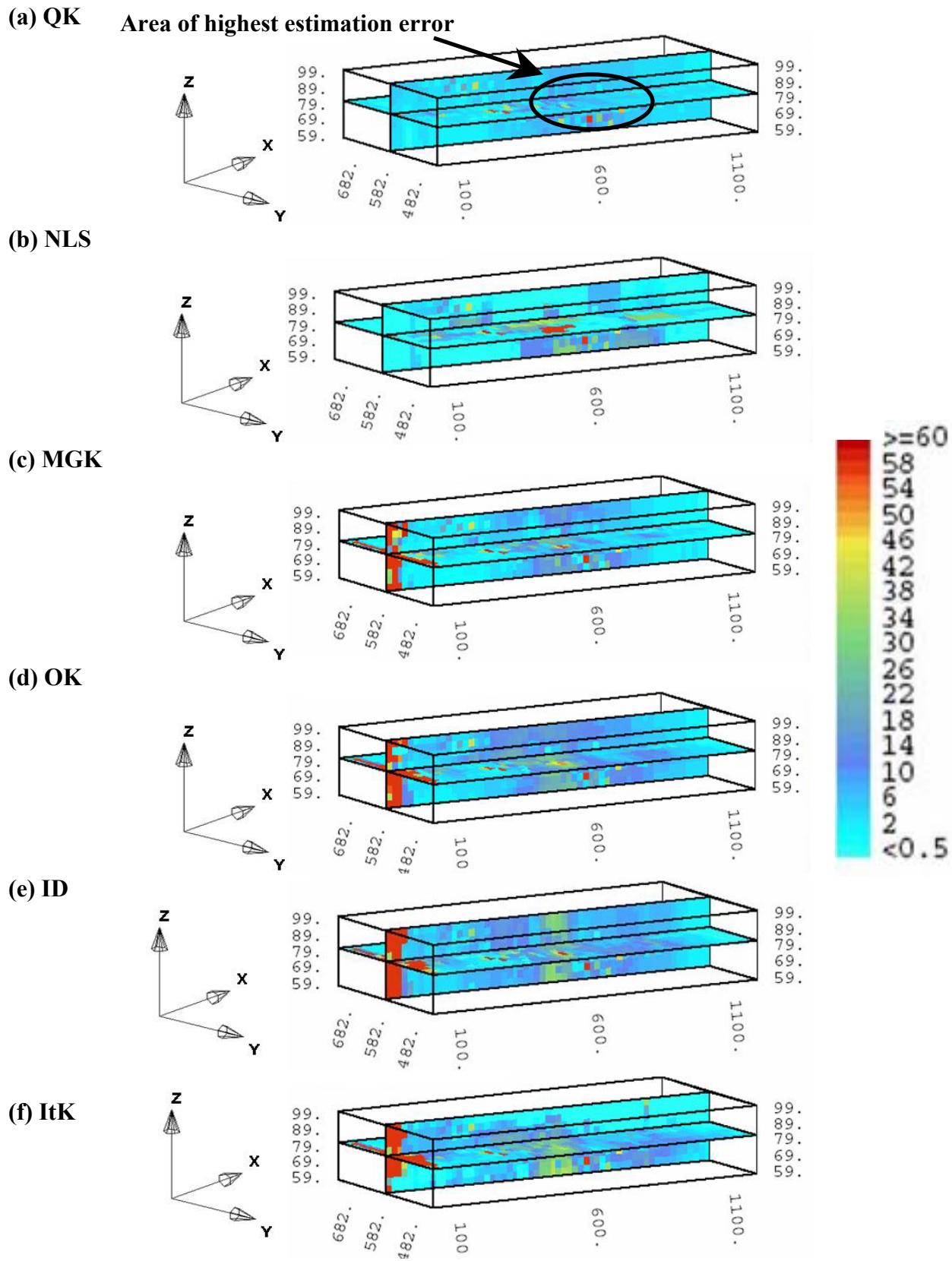


Figure 5.5 Spatial distributions of estimation errors for the large test case (mg/m<sup>3</sup>)

Figure (5.5b) shows that NLS actually outperforms QK in estimating the boundaries of the plume. The method's increased RMSE relative to QK resulted from an increased sensitivity to the poor spatial coverage and preferential sampling of high concentrations in the midsection of the plume. Figures (5.5a) and (5.5b) show NLS's sensitivity to these factors, clearly illustrating increased estimation errors in the midsection of the plume, many of which exceed  $\pm 20\text{-mg/m}^3$ .

The spatial distribution of errors for MGK illustrated in Figure (5.5c) shows increased estimation errors in the midsection of the plume relative to both NLS and QK. Far more troubling, the figure shows that the MGK approach fails to produce accurate estimates in the source area of the plume. The extreme values within the plume's source area were considered during both structural analysis and cross-validation for MGK. However, the high estimation errors shown in Figure (5.5c) within the source area of plume occur because MGK implicitly assumes that (1) the data are symmetrically distributed and (2) they display a *destructurization effect*, which means that the extreme values are not highly correlated in space. The first assumption fails because of the tremendous variability of the PCE concentration data; even after transformation the data set remained skewed and asymmetrically distributed. The second assumption fails because both tails of the data cdf (i.e., the zero concentrations and the source area concentrations) are highly correlated in certain regions of the plume. Using the available software in the GSLIB library of *Deutsch and Journel* (1997), the PCE concentration data sets from all 3 test cases failed to match the theoretically expected behavior of a Multi-Gaussian phenomenon. These results are particularly important given the common use of normal and lognormal transforms within the groundwater community.

Figure (5.4) shows that the three remaining interpolation schemes had significantly higher RMSE values relative to QK, MGK, and NLS. The 4 extreme PCE samples located

within the source zone severely impeded the ability of OK, ID, and ItK to accurately estimate PCE concentrations within this zone of the plume. This is rather unsettling given that the RMSE values for OK and ItK from cross-validation do not accurately reflect the true performance of these methods. Figures (5.5d) – (5.5f) show the distributions of errors in the plume for each method. Figure (5.4) shows that ItK had the highest expected error of the six interpolation methods. Figure (5.5f) shows that the increased RMSE for ItK results from the method's failure to accurately estimate PCE concentrations in the source area and midsection of the plume. Although ItK is elegantly formulated for working with nonstationary phenomena, it was heavily biased in the present case by extreme concentration values; this result concurs with *Hughes & Lettenmaier* (1981). Both deterministic interpolation schemes (NLS and ID) perform nearly as well or better than OK. Given the ease with which these methods can be implemented, they have significant potential for use as screening tools (see *Reed et al.* 2000 for an example in monitoring design), although they have the drawback of not providing any measure of estimate uncertainty.

### 5.3.1.3 Local Uncertainty Analysis

Table (5.4) summarizes how well the kriging approaches were able to capture the true PCE concentrations within the 95<sup>th</sup> confidence intervals computed using the estimation variance for each estimate as shown in equations (5.7) and (5.8).

	<b>QK</b>	<b>MGK</b>	<b>OK</b>	<b>ItK</b>
<b>Large Test Case</b>	99	95	92	92
<b>Intermediate Test Case</b>	100	100	66	76
<b>Small Test Case</b>	100	100	100	89

Table 5.4 Percentage of true data falling in 95th confidence interval

For the 6372 points in the large test case's interpolation domain, all of the kriging methods were able to capture greater than 90-percent of the true data values within the estimated 95<sup>th</sup> percentile

confidence intervals. OK and ItK fell slightly below the expected value of 95 percent while QK exceeded this value by capturing 99-percent of the true data in its uncertainty estimates, suggesting that the uncertainty estimates for QK were conservative.

In addition to the results of Table (5.4), Figure (5.12a) shows the spatial distribution of predicted uncertainty computed for QK for the large test case. The figure plots the *interquantile ranges*, which are equal to the difference between the upper and lower bound estimates for the 95<sup>th</sup> percentile confidence intervals at every point in the 3 interpolation domains considered in this chapter. The larger the distance between these values, the more uncertainty that is associated with an estimate. QK was used because the method was the least affected by extreme concentrations in both structural analysis and estimation. Moreover, QK uncertainty estimates have the advantage of being non-parametric. The purpose of the plots is to verify if zones of high predicted uncertainty correspond with zones of high estimation error. This comparison is important for applications such as site characterization or long-term monitoring where local uncertainty estimates are used to make regulatory or managerial decisions. Comparison of Figures (5.12a) and (5.5a) shows that the area of highest predicted uncertainty is centered at  $x = 600\text{m}$  [Figure (5.12a)] and corresponds well with the area of highest actual estimation errors [Figure (5.5a)]. The absolute estimation errors were highly correlated with the predicted estimation errors from kriging (i.e., the kriging estimation variances) for the 6372 grid points in the large case domain as evidenced by a rank correlation coefficient equal to 0.98 for the two quantities. The numerical values of kriging estimation variances did not predict the “actual” error values (see *Goovaerts 1997, Journel & Rossi 1989*), but they did successfully rank which areas of the plume have the highest expected error (i.e., areas with fewer sampling locations had both increased kriging estimation variances and actual estimation errors). In this test case, QK was

able to both successfully capture the true PCE concentrations and spatially predict the zone of highest estimation errors.

### 5.3.2 Medium Test Case

The medium test case shown in Figure (5.1b) consists of a total of 58 sample locations within a network of 29 multi-level monitoring wells. Reducing the data set size from 124 to 58 significantly impaired structural analysis. Anisotropy could no longer be identified for any of the kriging methods. Isotropic variogram models were specified for OK, MGK, and QK. The black-box fitting procedure used for ItK was unable to identify any correlation structure in the dataset and, as a result, was specified as a pure nugget variogram model. The fitting procedure failed to identify a model with structure due to an increase in the number of extreme PCE samples and a decrease in the overall number of data available for estimation.

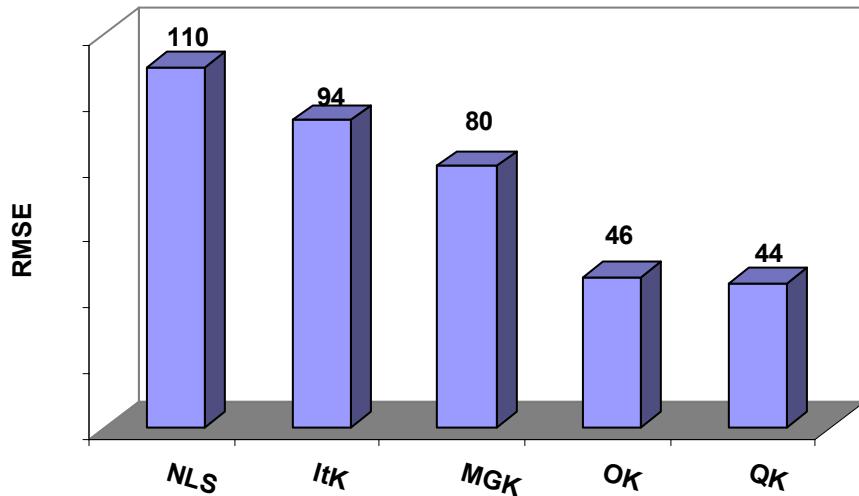


Figure 5.6 RMSE values from cross-validation for the medium test case ( $\text{mg}/\text{m}^3$ )

#### 5.3.2.1 Cross-Validation

The cross-validation RMSE values shown in Figure (5.6) are significantly higher than those shown in Figure (5.3) for the large test case. Three factors contributed to the increased

cross-validation errors for this test case: (1) the reduction in the spatial coverage and total number of the sample data, (2) the sample data set is more heavily skewed toward high PCE concentrations, and (3) the proportion of monitoring wells that sample multiple locations along their vertical axis decreased from 80-percent for the large case to 66-percent for this case, which causes the average distance between known and unknown locations to increase. OK and ItK were again severely affected by high PCE concentrations and required the exclusion of 7 samples with values ranging from 500-mg/m<sup>3</sup> to 6000-mg/m<sup>3</sup>. To facilitate performance comparisons with the 3 remaining interpolation methods, only the cross-validation residuals for 51 of the 58 available sample data points were used to compute the RMSE values shown in Figure (5.6). QK is clearly the superior interpolation scheme because the structural analysis performed for this approach did not require the exclusion of high PCE concentrations. Interestingly, the relative rankings of the 5 schemes changed significantly between large and medium test cases. Particularly, the RMSE scores for MGK and ItK increased significantly because both methods drastically overestimated concentrations for samples with no PCE that are in close proximity to high concentration samples in the source area of the plume. The next section shows that cross-validation again failed to properly rank the relative performances of the interpolation schemes.

#### 5.3.2.2 *Estimation Accuracy*

Figure (5.7) presents the absolute errors for each of the six interpolation methods averaged over the 2496 points in the interpolation domain [see Table (5.3)]. Comparing the medium case RMSE values from cross-validation and interpolation, it is evident that cross-validation is again underestimating the expected errors for all of the interpolation methods. As was observed for the large test case, clustered sampling locations in areas of high concentration are biasing cross-validation residual calculations. This is particularly true for NLS, OK, and ItK,

all of which were heavily biased by extreme PCE concentrations. QK and MGK performed better than the remaining methods because their data transformations reduce the influence of the extreme PCE values on variogram identification and spatial overestimation.

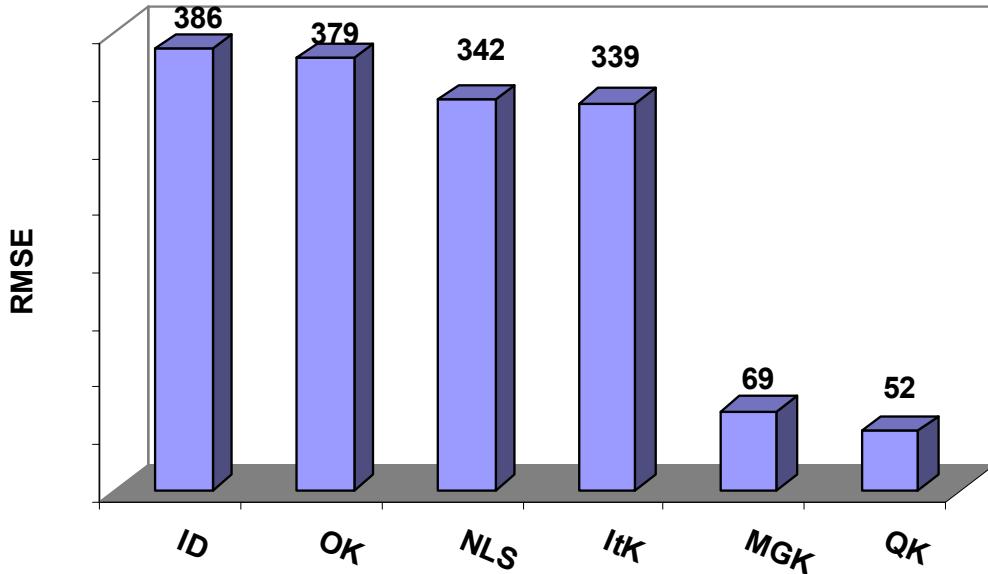


Figure 5.7 RMSE values from interpolation for the medium test case ( $\text{mg}/\text{m}^3$ )

Figures (5.8a)-(5.8f) show the spatial distributions of errors for each of the interpolation schemes. These figures clearly show that QK has the best distribution of error, with a majority of the errors falling below  $\pm 25 \text{-mg}/\text{m}^3$ . MGK again fails to accurately estimate concentrations within the source area of plume due to the inappropriateness of the Multi-Gaussian assumption. Figures (5.8c) – (5.8f) show that the distributions of error for ItK, NLS, OK, and ID are all very similar. The methods all failed to produce accurate estimates in the source area of the plume and severely overestimate PCE concentrations in the area centered at  $x = 500\text{m}$  located near the bottom of the interpolation domain. This bias towards overestimation for ItK and OK also degraded the quality of their local uncertainty estimates as will be shown in the next section.

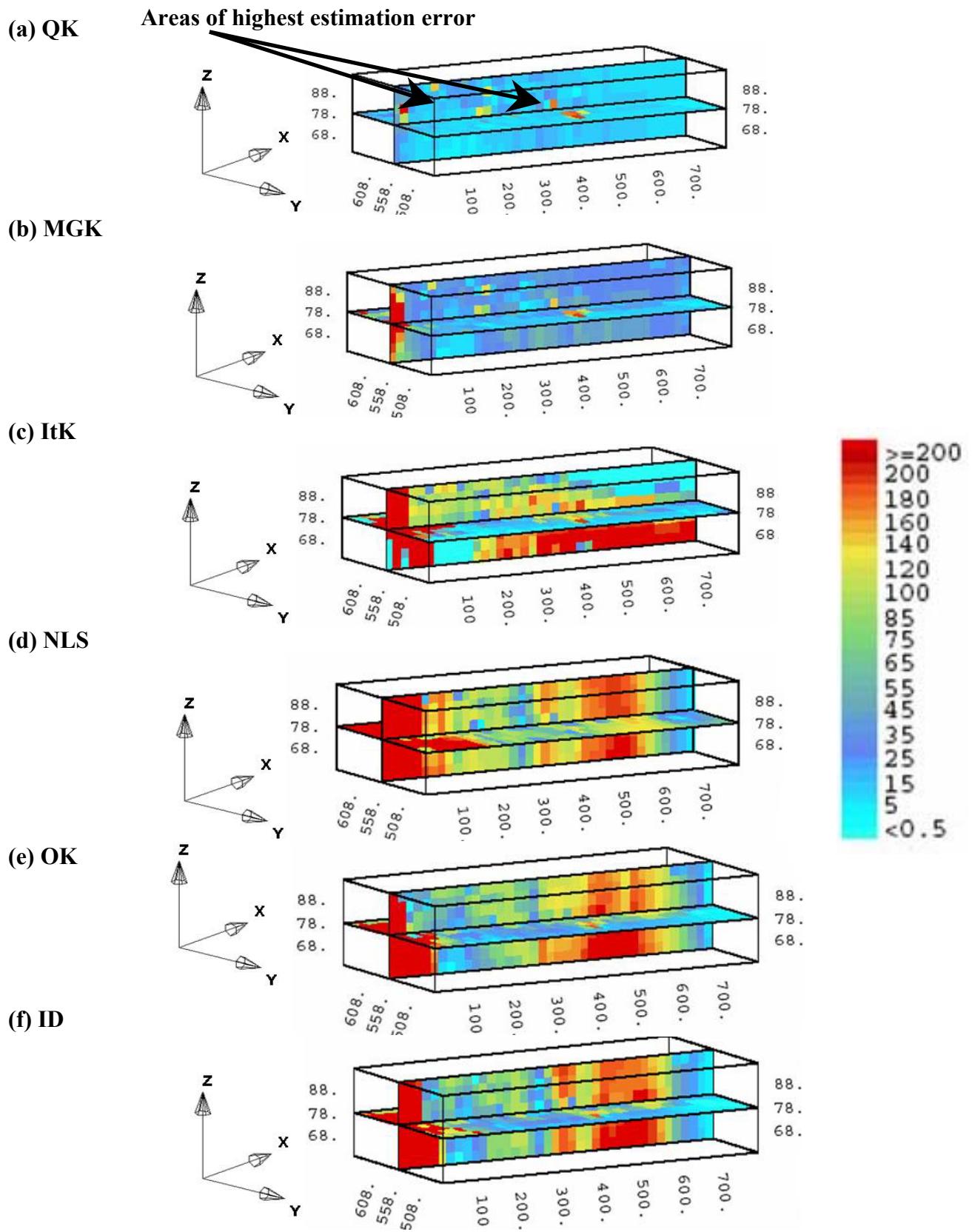


Figure 5.8 Spatial distribution of estimation errors for the medium test case (mg/m<sup>3</sup>)

### 5.3.2.3 Local Uncertainty Analysis

Table (5.4) shows that for this case, ItK and OK local uncertainty estimates were severely affected by estimation bias, capturing only 76- and 66-percent of the true data, respectively. Both methods underestimated the true extent of the uncertainty of their estimates. *Goovaerts* (1997) states that highly asymmetric data sets can significantly degrade local uncertainty estimates for kriging methods that require the assumption that estimation errors are normally distributed and spatially independent. The estimation errors for OK and ItK shown in Figures (5.8c) and (5.8e) were highly asymmetrical and correlated in space. Given that ItK and OK both failed to capture the true data with their local uncertainty estimates, these methods have no clear advantage over ID and NLS for this test case.

MGK and QK both produced conservative uncertainty estimates and captured all of the true data in their 95<sup>th</sup> percentile confidence intervals. Moreover, a comparison of Figures (5.12b) and (5.8a) again shows a close correspondence between areas projected to be highly uncertain and zones of high estimation error for QK with a rank correlation coefficient equal to 0.97.

### 5.3.3 Small Test Case

The small test case represents a sparsely sampled PCE plume extending approximately 400-m in the primary direction of groundwater flow, with a total of 20 multi-level monitoring wells sampling 26 locations. The high variability of the PCE concentrations and extremely limited size of the small test case data set combined to make identification of spatial structure impossible for OK, QK, and MGK. Pure nugget variogram models were specified for each of these methods. The black-box fitting procedure used for structural analysis in ItK was able to specify an isotropic general covariance model because this method did not use estimation neighborhoods (i.e., all 26 sample data were used in all of the fitting procedure's computations).

ItK could not use estimation neighborhoods for this test case because there were too few data in each neighborhood to fit a trend function. The limited size and heavily skewed nature of the small test case data set helped to identify which of the schemes were the most robust in facing these challenges.

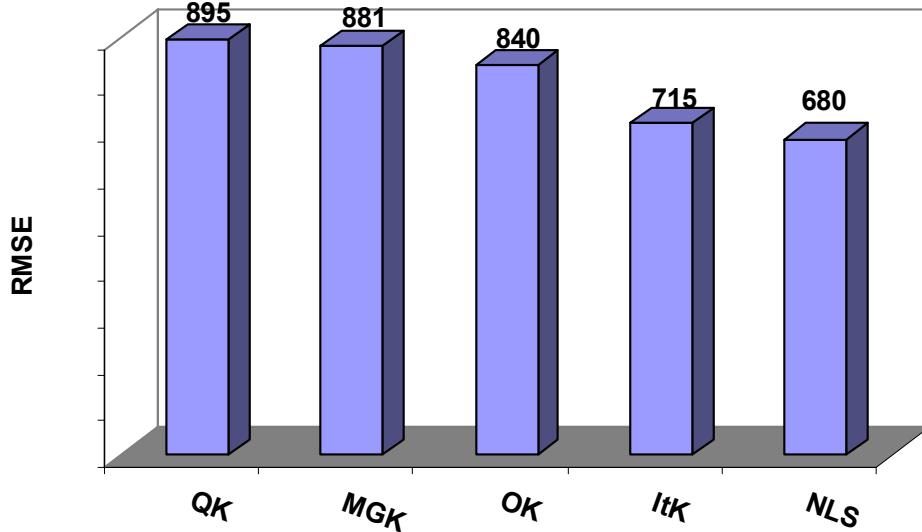


Figure 5.9 RMSE values from cross-validation for the small test case ( $\text{mg}/\text{m}^3$ )

#### 5.3.3.1 Cross-Validation

Figure (5.9) shows that all of the interpolation methods had significantly higher cross-validation RMSE values for the small test case relative to the previous cases. The increased RMSE scores for this case resulted because the size and spatial coverage of this data set decreased while the proportion of extreme PCE samples increased relative to the large and medium test cases. Additionally, only 30-percent of monitoring wells for this test case sample multiple locations in space, which greatly increases the average distance between unknown and known locations in cross-validation calculations relative to the previous two test cases. Recall that as this average distance increases, RMSE increases.

The relative ranking of the interpolation schemes given in Figure (5.9) agrees with previous studies that conclude that kriging methods have no significant advantage relative to NLS and ID for data sets with less than 50 samples (*Hughes & Lettenmaier* 1981, *Gotway et al.* 1996). Interestingly, ItK produced more accurate cross-validation estimates, but the method's estimates were less robust (i.e., its local uncertainty estimates failed to capture the true PCE concentrations) relative to MGK and QK. Recall *Hughes & Lettenmaier* (1981) found that ItK suffers from a bias that causes estimation variances to be smaller then they should be for reduced data sets. However caution must be exercised when assigning performance rankings using cross-validation, because the next section again shows that the cross-validation scores for this test case do not reflect the true performances of the interpolation schemes.

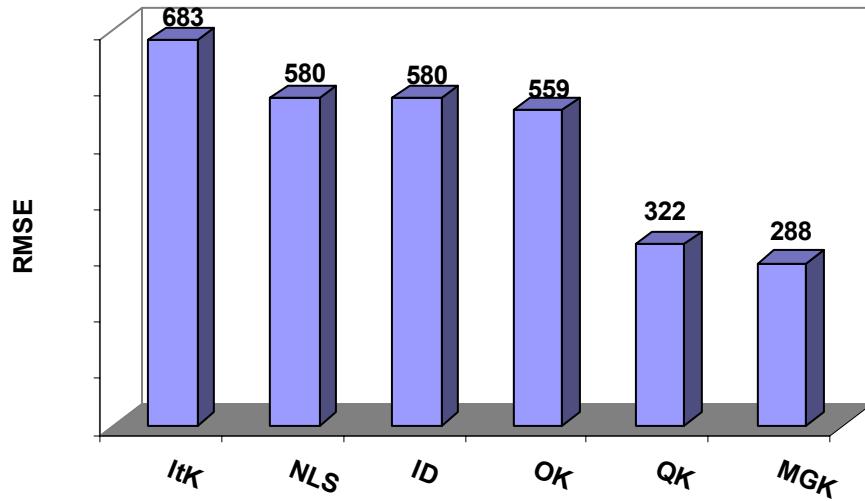


Figure 5.10 RMSE values from interpolation for the small test case (mg/m<sup>3</sup>)

### 5.3.3.2 Estimation Accuracy

Figure (5.10) presents the RMSE values computed relative to the true concentrations for the 2090 points in the small case interpolation domain. Comparison of the interpolation RMSE scores in Figure (5.10) with the cross-validation scores in Figure (5.9) shows that unlike the

previous cases, cross-validation actually overestimates the expected errors for all of the interpolation methods. The primary factor causing this shift from underestimation to overestimation lies in the reduction of the percentage of multi-level monitoring wells from 80-percent for the large case to 30-percent for the small case. Reducing the number of multi-level wells greatly increased the average distance between known and unknown locations in cross-validation computations, which resulted in a bias toward overestimating estimation error.

Figure (5.10) shows that MGK and QK were actually the best performing methods, in contrast to the worst performing methods as predicted in cross-validation [see Figure (5.9)]. Figure (5.11) provides a more detailed understanding of the relative performances of these methods, showing the spatial distributions of their estimation errors. Unlike the previous two test cases, QK had a higher RMSE value relative to MGK because the method overestimated zero concentrations near high concentrations in the source area of the plume as can be seen in Figure (5.11b). QK's performance was degraded because over 40-percent of the PCE sample data used to compute the empirical cdf had values exceeding 500-mg/m<sup>3</sup>. Despite this fact, Figures (5.11a) and (5.11b) show that QK was less prone to overestimation than MGK in the leading edge of the plume. Finally, Figures (5.10) and (5.11) show there was a minimal difference in the performance of OK, ID, and NLS. Given ID's ease of implementation, it would be the preferred estimation method for practitioners who do not require uncertainty estimates.

#### 5.3.3.3 Local Uncertainty Analysis

Table (5.4) shows that OK, QK, and MGK all captured 100-percent of the true data in the uncertainty estimates for the small test case. All of these methods produced very conservative uncertainty estimates because of their pure nugget variogram models (i.e., assumption of no spatial correlation), when spatial correlation is actually present in the spatial distribution of the

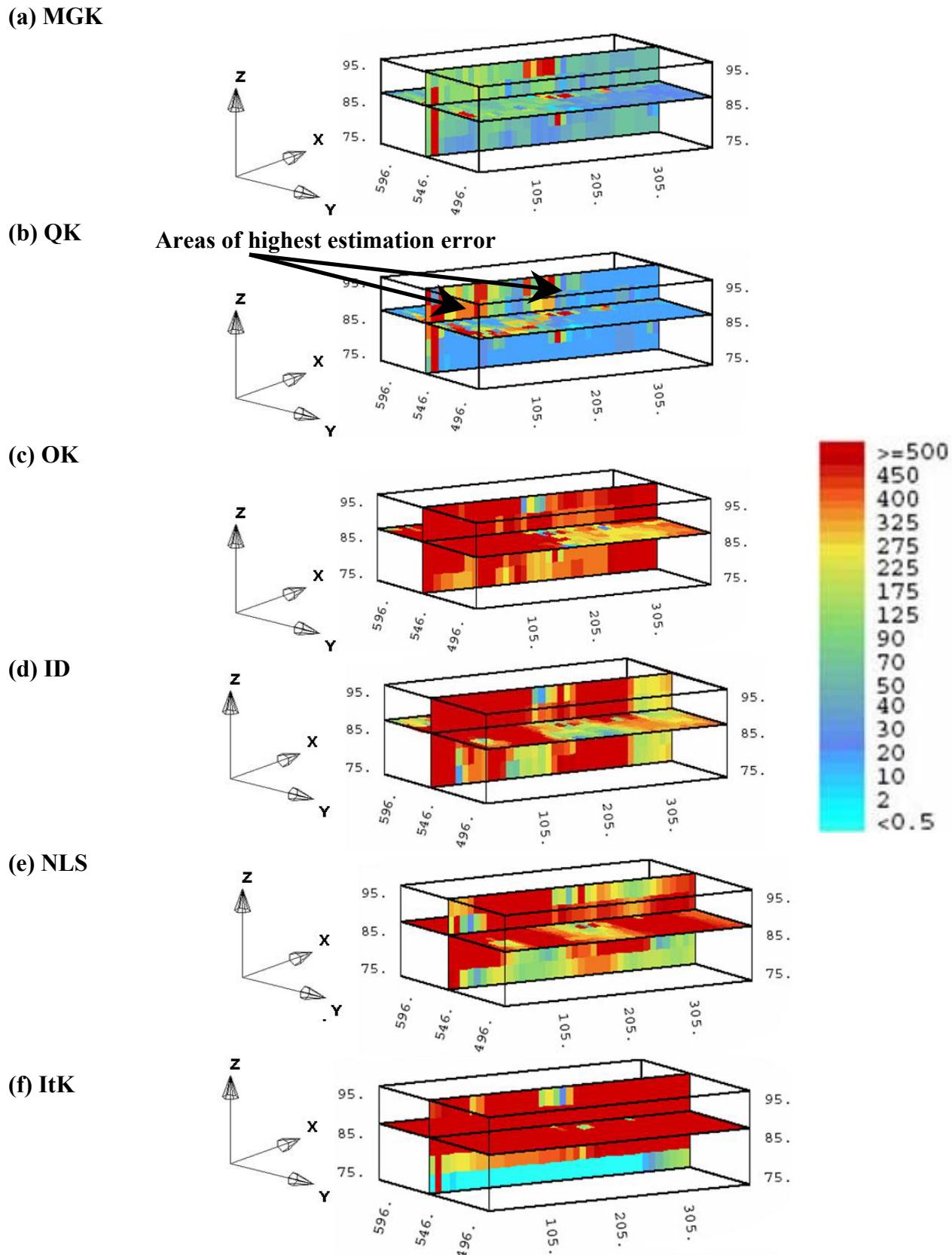
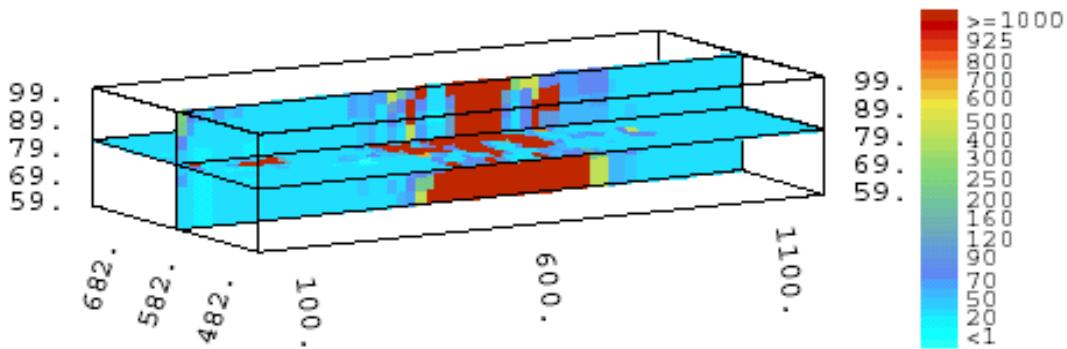
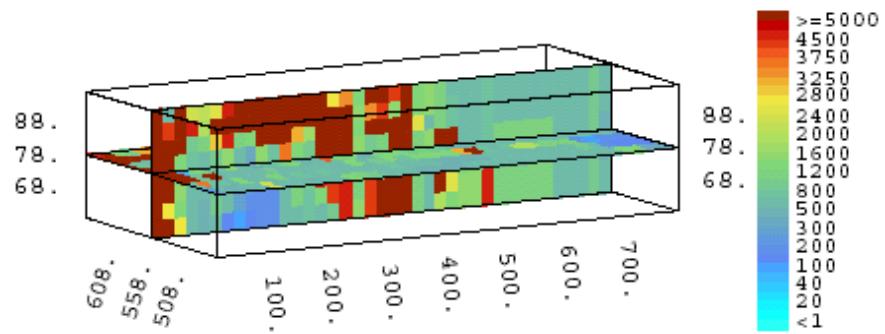


Figure 5.11 Spatial distribution of estimation errors for the small test case ( $\text{mg}/\text{m}^3$ )

**(a) Large Case Uncertainty:**



**(b) Medium Case Uncertainty:**



**(c) Small Case Uncertainty:**

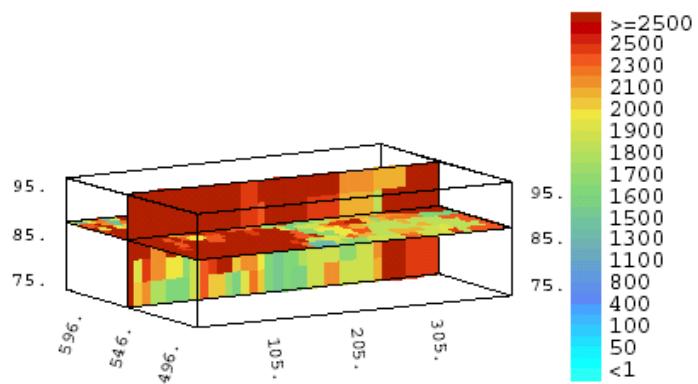


Figure 5.12 Spatial distributions of local uncertainty for QK (mg/m<sup>3</sup>)

plume but is not quantifiable from the sample data. ItK was only able to capture 89-percent of the true data due to its susceptibility to under predicting estimates' uncertainties for small data sets. Figure (5.12c) is particularly interesting because even for a small data set where PCE concentrations are assumed spatially uncorrelated, the predicted zones of highest uncertainty are still correlated with the actual estimation errors shown in Figure (5.11b) with a rank correlation coefficient between the two quantities equal to 0.97.

## 5.4 Conclusions

This chapter provides a detailed description of the motivations and assumptions underlying each of 6 interpolation methods and an analysis of their performance for plume interpolation. Results from the 3 test cases show how robust the interpolation methods are as the number of sample data decreases and illustrate the difficulties practitioners face when interpolating contaminant concentrations at highly heterogeneous sites. The use of simulation data allowed this chapter to examine the spatial distributions of estimation accuracy and uncertainty predictions. An important result from this chapter is that performance ratings from cross-validation are misleading and should be used cautiously. Specifically, the chapter shows that clusters of monitoring wells used to preferentially sample zones of high concentration heavily bias cross-validation scores.

Although QK and MGK were the methods least affected by preferential sampling of high PCE concentrations, MGK's normality assumptions were inappropriate for all of the heavily skewed concentration data sets as evidenced by the method's failure to accurately estimate source area concentrations for all three test cases. The only significant advantage OK and ItK had over ID and NLS was their ability to produce uncertainty estimates. ItK had two primary drawbacks relative to the other kriging methods: (1) traditional graphical methods for structural

analysis could not be used for scatter-point concentration data and (2) it underestimated the uncertainty of its estimates in sparsely sampled areas of the contaminant plume. Both ItK and OK were heavily biased by preferential sampling and extreme concentrations. For all of the data sets, the reduced computational complexity and ease-of-implementation of NLS and ID make these methods useful as screening tools to determine if the expense of a geostatistical study is warranted for a data set. When selecting between these two methods, it is important to consider that NLS was only able to outperform ID when more than 50 sample data were available. Overall, the variability of PCE samples and preferential sampling controlled how well each of the interpolation schemes performed. QK was the most robust of the interpolation methods, showing the least bias from both of these factors and has the additional advantage of being non-parametric. Moreover, the method's non-parametric uncertainty estimates successfully predicted zones of high estimation error for each test case. Quantile kriging was the most robust of 6 groundwater plume interpolation methods, showing the least bias from both preferential sampling and the variability of contaminant data. These findings warrant further studies into the applicability of quantile kriging to data sets from other fields ranging from mining to life sciences, where nonstationary interpolation also plays a vital role.



# 6 BALANCING PERFORMANCE CRITERIA

## 6.1 Introduction

This chapter builds upon the previous chapters by combining an effectively designed multiobjective GA [see Chapter 4] with quantile kriging [see Chapter 5] into a highly adaptable LTM design methodology that has significant potential for aiding stakeholders in selecting, understanding, and balancing monitoring objectives for contaminated sites.

### 6.1.1 *Motivation & Scope*

*Buras* (2001) contends that one of the unresolved issues in water resources is the inclusion of multiobjective formulations in the design of engineered systems. Multiobjective problem formulation implicitly requires decision makers to select, understand, and balance performance objectives for the physical systems being designed. The goal of this chapter is to demonstrate that tools such as multiobjective genetic algorithms (GAs) that are capable of high order Pareto optimization (i.e., optimizing a system for more than 2 objectives) can serve as an interface between the design of the physical system and the human decision process. *Buras* (2001) recommends that multi-attribute utility analysis (MAUA) be used in this role. MAUA implicitly requires a preferential weighting of design objectives as well as a conversion of the objective values into the single metric of “utility”. The major difficulties encountered when applying MAUA are (1) ensuring that only incommensurate, conflicting objectives are used and (2) assigning a unique utility function for group decision making (i.e., the Arrow paradox) [see *de Neufville* 1990 for more details]. Because of these difficulties, *de Neufville* (1990) states that multiobjective optimization is the preferred solution technique in cases where there are vast numbers of design alternatives. However, multiobjective optimization and MAUA can be combined sequentially; *Horn* (1997) recommends that Pareto optimization be used to quantify a

small set of optimal solutions from which a final solution can be selected using MAUA. This chapter will demonstrate that high order Pareto optimization can provide a means of selecting objectives, discovering objective conflicts, and helping stakeholders in the negotiation process, whether or not combined with MAUA. The highly adaptable optimization methodology proposed in this chapter has significant potential for aiding the design of water resources systems by enabling decision makers to discover, understand, and balance tradeoffs among their design objectives.

The optimization methodology is demonstrated in this chapter using a long-term monitoring (LTM) application. The application addresses the two most important problems LTM practitioners face in the design process: (1) selecting monitoring objectives and (2) balancing these objectives. Both the *ASCE Task Committee on Geostatistical Techniques* (1990b) and *Loaiciga et al.* (1992) concur that the selection of performance criteria is the most important component of any monitoring design methodology. The problem of selecting monitoring performance criteria requires stakeholders to abstract their design preferences into mathematical functions and understand how these functions affect sampling strategies. *Loaiciga et al.* (1992) state that “[o]ne of the key difficulties in the design of ground water monitoring networks via mathematical models is to choose objective functions that faithfully represent a [stakeholder’s] objective”. Moreover, stakeholders must be able to assess how these mathematical models interact and how these interactions affect the final design of a monitoring system.

For example, there is an obvious conflict between cost and uncertainty. As the number of sample locations used decreases, sampling costs also decrease but uncertainty increases. Now consider uncertainty and contaminant mass estimation error: both quantities increase as the number of sample locations decrease. Does a conflict exist between these objectives? Do both

objectives have a significant effect on the final design of a monitoring network? High order Pareto optimization can serve to answer these questions by enabling stakeholders and regulators to isolate and visualize a small number of sampling strategies that are optimal with respect to multiple objectives. Through visualization of these sampling strategies, stakeholders can discover how their objectives are affecting designs and select only those objectives that best fit their design preferences.

### **6.1.2 *Background on LTM Design***

Monitoring network design has been studied extensively in the past, but previous studies have primarily focused on two problems: (1) the use of geostatistics to augment or design monitoring networks for site characterization (for a review, see *ASCE Task Committee on Geostatistical Techniques* 1990a,b) and (2) the use of optimization and numerical simulation to site new monitoring points for contaminant plume detection at landfills and hazardous waste sites (for a review, see *Loaiciga et al.* 1992). Recently, a third problem has emerged that seeks to reduce spatial and temporal redundancies in pre-existing well networks for sites undergoing long term monitoring (LTM). The LTM design methodology proposed in this chapter combines concepts from the geostatistical design and the redundancy analysis studies described below.

The geostatistical approach to monitoring network design utilizes geostatistical estimation procedures (kriging) to evaluate alternative sampling schemes. Geostatistics provides minimum error estimates of contaminant concentrations at unsampled locations using linear combinations of sample values (for an introduction see *Kitanidis* 1997, *Goovaerts* 1997, *Chilès & Delfiner* 1999). In addition to providing the expected values of contaminant concentrations, the estimation variance is computed that represents the uncertainty of estimates at unsampled locations. Estimation variances are independent of the sampled data and vary only as a function

of the monitoring well network's geometry, enabling monitoring network designs to be evaluated for their ability to characterize a site before any sampled data are available.

*Hughes & Lettenmaier* (1981) recognized the potential of the estimation variance attained from the geostatistical estimation for guiding sampling design. Early studies focused on methodologies that used estimation variance to either augment or design monitoring networks for site characterization (for reviews see *ASCE Task Committee on Geostatistical Techniques* 1990b, *Loaiciga et al.* 1992). Additional sampling points are added based on an analysis of which locations will maximally decrease the estimation variance attained in geostatistical interpolation (*Rouhani* 1985, *Rouhani and Hall* 1988).

Spatial redundancy analysis has garnered more attention as the number of contaminated sites requiring LTM has increased over the previous decade. Sampling-and-analysis costs have been recognized as significant contributors to the overall cost of remediating sites with groundwater contamination. LTM at many sites can require decades of expensive sampling at tens or even hundreds of existing monitoring wells, resulting in hundreds of thousands or millions of dollars for sampling and data management per year. The tremendous costs associated with the collection and management of LTM data has motivated the development of design methodologies that seek to identify and remove sampling locations that contribute minimally to understanding the plume's extent in space, time, or both. Evidence of the emerging importance of redundancy analysis can be seen in the Federal Remediation Technologies Roundtable's focus on tools that directly account for redundancies in monitoring data to reduce sampling costs (see <http://www.frtr.gov/optimization/monitoring/>).

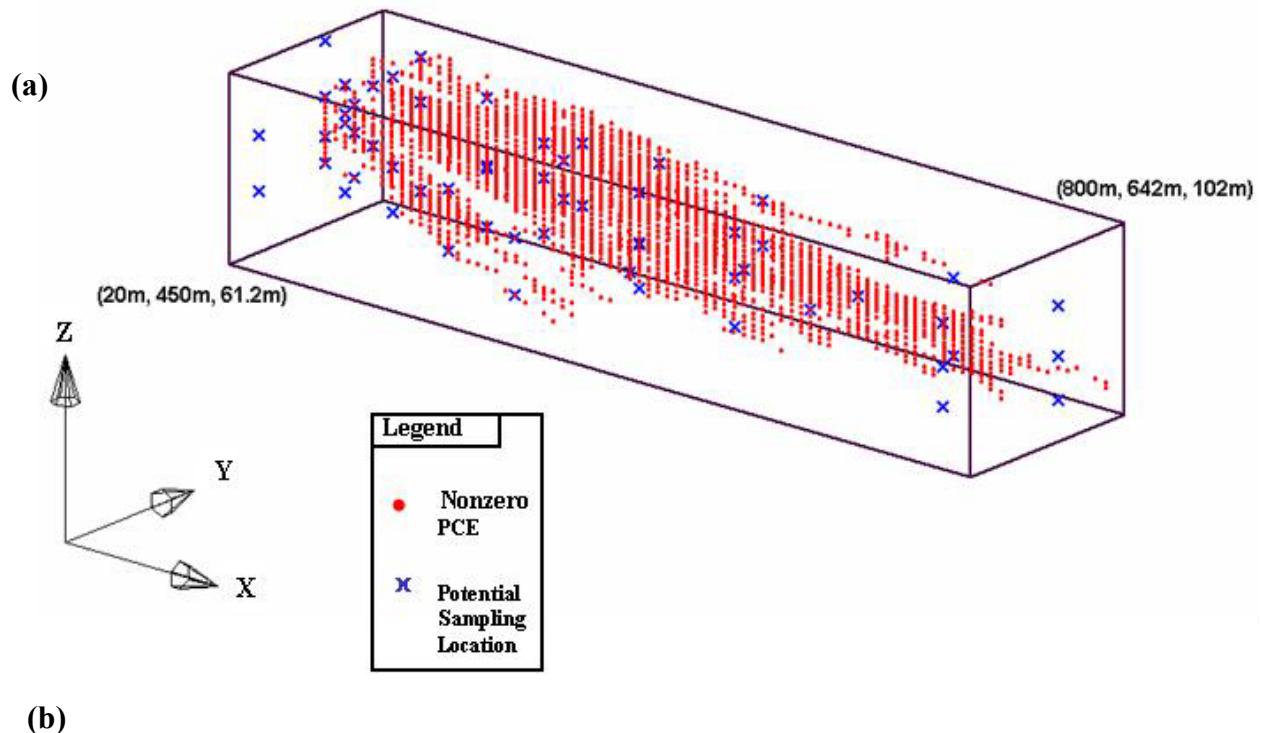
Previous spatial redundancy approaches can be classified as being either global or local in nature. *Reed et al.* (2000a) present a global redundancy analysis that combined a simple GA,

fate-and-transport simulation, and plume interpolation. The study presents a single objective methodology for monitoring design, in which minimum cost sampling plans are identified that quantify the total (global) contaminant mass as accurately as when all available monitoring locations are utilized. *Reed et al.* (2001) presents a local redundancy analysis that combined deterministic plume interpolation and a multiobjective genetic algorithm (GA) to seek the optimal tradeoff between sampling costs and the relative accuracy of local concentration estimates. Other local approaches have combined trial-and-error analysis with plume interpolation to eliminate spatial redundancies (see *Johnson et al.* 1996, *Cameron & Hunter* 2000, *Aziz et al.* 2000). None of these studies directly quantify how eliminating monitoring points increases uncertainty.

The LTM design methodology proposed in this chapter combines both the spatial redundancy and geostatistical approaches to monitoring design. Quantile kriging and the Nondominated Sorted Genetic Algorithm-II (NSGA-II) are combined to quantify the tradeoffs among the following four performance criteria: (1) cost, (2) the relative accuracy of local concentration estimates, (3) the relative accuracy of global mass estimates, and (4) local uncertainty as measured by kriging estimation variances. These four criteria not only combine spatial redundancy analysis with geostatistical approaches, but they also combine both the local and global spatial redundancy measures used separately in *Reed et al.* (2001) and in *Reed et al.* (2000a), respectively. The test case described in Section 6.2 corresponds to the medium test case described in Chapter 5. Readers familiar with the previous chapter can go directly to Section 6.3.

## 6.2 Test Case Data

The test case developed for this chapter uses data drawn from a 50 million-node flow-and-transport simulation performed by *Maxwell et al.* (2000). The simulation provided realistic



(b)

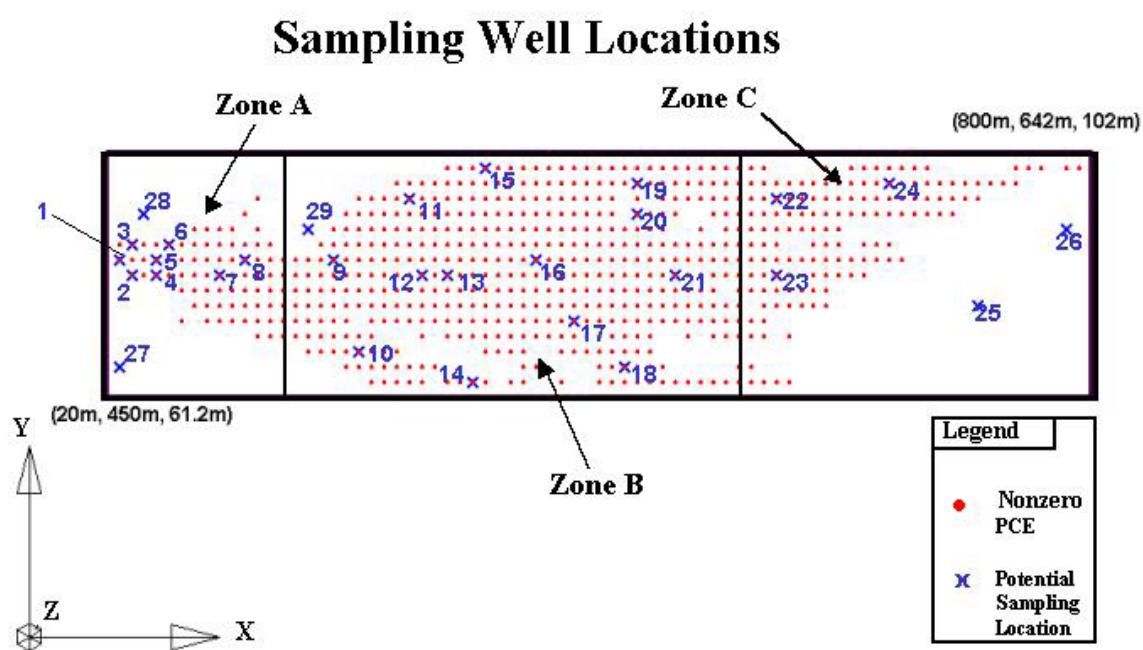


Figure 6.1 (a) 3-dimensional view of the monitoring test case (b) XY-plane view of monitoring wells

historical data for the migration of a hypothetical perchloroethylene (PCE) plume in a highly heterogeneous alluvial aquifer. The hydrogeology of the test case is based on an actual site located at the Lawrence Livermore National Laboratory (LLNL) in Livermore, California. Data were provided for a total of 58 hypothetical sampling locations within a 29-well multi-level monitoring network shown in Figures (6.1a) and (6.1b). Note that Figure (6.1b) and all subsequent XY-views of the plume are divided into zones A, B, and C to show how the sampling designs change in the source area, the midsection, and the leading edge of the plume, respectively. The data represent a snapshot in time, 8 years after an underground storage tank has continuously released contamination into the aquifer system. The monitoring wells can sample from 1 to 3 locations along their vertical axis and have a minimum spacing of 10 m between wells in the horizontal plane.

The site is assumed to be undergoing long-term monitoring, in which groundwater samples are used to assess the effectiveness of current remediation strategies. Quarterly sampling of the entire network has a potential cost of over \$85,000 annually for PCE testing alone, which could translate into millions of dollars if the site had a typical life span of 20 to 30 years (*Rast* 1997). This chapter addresses only spatial redundancy, assuming that the spatial sampling plan will be re-evaluated periodically as site conditions change. As noted previously, this type of approach has been applied in several trial-and-error field applications (*Johnson et al.* 1996, *Cameron & Hunter* 2000, *Aziz et al.* 2000).

### 6.3 Methodology

The methodology has 3 primary components: (1) a high order, constrained multiobjective problem formulation, (2) plume interpolation using quantile kriging, and (3) search and

optimization using the NSGA-II. Each of these components is described in more detail in the following sections of this chapter.

### 6.3.1 Generalized Multiobjective Formulation

Equation (6.1) presents the generalized multiobjective problem formulation used in this chapter to search for optimal sampling strategies.

$$\text{Minimize } F(\bar{x}_\kappa) = (f_1(\bar{x}_\kappa), f_2(\bar{x}_\kappa), \dots, f_\beta(\bar{x}_\kappa)), \quad \forall \kappa \in \Omega$$

*Subject to*

$$U(\bar{x}_\kappa) = 0 \quad (6.1)$$

In the above equation,  $F(\bar{x}_\kappa)$  is a vector-valued performance function in which the  $\beta$  components,  $f_1(\bar{x}_\kappa), f_2(\bar{x}_\kappa), \dots, f_\beta(\bar{x}_\kappa)$  are all minimized. Each of the component objectives are functions of the vector  $\bar{x}_\kappa$  which represents the  $\kappa^{\text{th}}$  sampling plan within the overall decision space  $\Omega$ . Equation (6.2) defines the  $i^{\text{th}}$  component of the decision variable vector  $\bar{x}_\kappa$  representing the  $\kappa^{\text{th}}$  monitoring scheme.

$$x_{\kappa,i}^{\text{samp}} = \begin{cases} 1, & \text{if the } i^{\text{th}} \text{ well is sampled} \\ 0, & \text{otherwise} \end{cases}, \quad \forall \kappa, i \quad (6.2)$$

The vector-valued performance function  $F(\bar{x}_\kappa)$  is minimized subject to the constraint shown in equation (6.1), which stipulates that the total number of unestimated points in the interpolation domain  $U(\bar{x}_\kappa)$ , described in Table 1 for this case chapter, is equal to zero. Kriging contaminant data requires the specification of maximum radii of search for known concentrations in the vicinity of an unestimated point. If no known concentrations are found within these radii of search, then the current grid point remains unestimated. The presence of unestimated points within the interpolation domain prevents reliable calculations of the relative accuracy of local concentration estimates, the relative accuracy of global mass estimates, and uncertainty.

<b>Minimum X-Coordinate</b>	40 meters
<b>Minimum Y-Coordinate</b>	474 meters
<b>Minimum Z-Coordinate</b>	70.8 meters
<b>Size of Grid Blocks in X-Direction</b>	20 meters
<b>Size of Grid Blocks in Y-Direction</b>	24 meters
<b>Size of Grid Blocks in Z-Direction</b>	2.4 meters
<b>Number of Grid Blocks in X-Direction</b>	34
<b>Number of Grid Blocks in Y-Direction</b>	7
<b>Number of Grid Blocks in Z-Direction</b>	7

Table 6.1 Interpolation grid definition

### 6.3.2 Design Objectives Suite

In this chapter,  $F(\bar{x}_\kappa)$  in equation (6.1) contains four component objectives: (1) cost, (2) relative error of local concentration estimates, (3) relative error of global mass estimates, and (4) local uncertainty. Details on each of these four objectives are given below.

#### 6.3.2.1 Cost

The cost of each sampling design is evaluated using equation (6.3) below, which shows that the cost is a discrete function of the number of wells sampled.

$$f_{COST}(\bar{x}_\kappa) = \sum_{i=1}^{n_{well}} C_S(i)x_{\kappa,i}^{samp} \quad (6.3)$$

Dollar-valued monitoring costs are computed by multiplying a constant coefficient defined in terms of dollars per sample by the total number of samples. In this chapter, costs have been normalized (i.e., the constant cost coefficient has been eliminated) and are presented in terms of the total number of samples required in a given sampling plan. In equation (6.3), if the  $i^{th}$  well is sampled then all available locations along the vertical axis of that well will be sampled. The normalized cost,  $C_S(i)$  of sampling the  $i^{th}$  well ranges between 1 and 3 depending on the number sampling locations along its vertical axis. The total cost of the  $\kappa^{th}$  sampling plan is determined

by summing the product of  $x_{\kappa,i}^{samp}$  from equation (6.2) and  $C_S(i)$  over the total of  $n_{well}$  monitoring wells. In this chapter, there are a total of 29 monitoring wells that sample 58 locations in space.

### 6.3.2.2 Relative Error of Local Concentration Estimates

The accuracy of local concentration estimates is quantified in terms of squared relative estimation error (SREE) using equation (6.4). SREE is the local redundancy measure that was used in Chapters 3 and 4.

$$f_{SREE}(\bar{x}_\kappa) = \sum_{j=1}^{n_{est}} (c_{all}^*(\bar{u}_j) - c_{est}^\kappa(\bar{u}_j))^2 \quad (4)$$

SREE provides a measure of how the interpolated picture of the plume using data only from wells included in the  $\kappa^{\text{th}}$  sampling plan compares to the result attained using data from all available sampling locations. The measure is computed by summing the squared deviations between the local concentration estimates attained using data from all available sampling locations,  $c_{all}^*(\bar{u}_j)$ , and the estimates based on the  $\kappa^{\text{th}}$  sampling plan  $c_{est}^\kappa(\bar{u}_j)$  at each location  $\bar{u}_j$  in the interpolation domain. Each  $\bar{u}_j$  specifies coordinates for the  $j^{\text{th}}$  grid point in the interpolation domain.

### 6.3.2.3 Relative Error of Global Mass Estimates

*Reed et al. (2000a)* analyzed spatial redundancies using global mass interpolation within a single objective LTM monitoring methodology. The study sought minimal cost sampling plans that would quantify the mass of dissolved groundwater contaminant as accurately as if all available sampling locations had been used. Quantifying the total mass of contaminant within the interpolated domain is equivalent to computing the zero<sup>th</sup> moment of the contaminant plume

at time  $t_s$  when samples are taken. Equation (6.5) gives relative mass estimation error measure ( $f_{MASS}(\bar{x}_\kappa)$ ) used both in this chapter and by *Reed et al.* (2000a).

$$f_{MASS}(\bar{x}_\kappa) = \left| \frac{Mass_{All} - Mass_\kappa}{Mass_{All}} \right| * 100 \quad (6.5)$$

$Mass_{All}$  is the best mass estimate based on sampling all available monitoring points in the model domain and has a value of 37 Kg for this application. Equation (6.5) is the absolute relative difference between  $Mass_{All}$  and the mass estimate,  $Mass_\kappa$ , computed using the data attained from the  $\kappa^{\text{th}}$  monitoring plan  $\bar{x}_\kappa$ .

#### 6.3.2.4 Local Uncertainty

Local uncertainty is quantified using estimation standard deviations (i.e., the square root of estimation variances) from kriging using equation (6.6).

$$f_{UNCERT}(\bar{x}_\kappa) = \sum_{j=1}^{nest} A\sigma(\bar{u}_j) \quad (6.6)$$

The local uncertainty measure  $f_{UNCERT}(\bar{x}_\kappa)$  is a weighted sum of the *nest* estimation standard error  $\sigma(\bar{u}_j)$  for location  $\bar{u}_j$  in the interpolation domain. For kriging-based estimation, *Journel & Rossi* (1989) showed that the estimation variance is a variogram model dependent ranking of sampling configurations. Chapter 5 shows that kriging estimation variances can successfully predict which areas of the plume have the highest estimation errors. The underlying goal of equation (6.6) is to identify solutions that sample subsets of the 58 total sampling locations while minimally increasing local uncertainty. Readers familiar with both quantile kriging and the NSGA-II should proceed directly to Section 6.3.4.3, which discusses how constraint violations were penalized.

### **6.3.3 Plume Interpolation using Quantile Kriging**

Quantile kriging was selected for plume interpolation in this chapter based on the findings of Chapter 5, which presents a comprehensive performance analysis of 6 interpolation methods for scatter-point concentration data, ranging in complexity from intrinsic kriging based on intrinsic random function theory to a traditional implementation of inverse-distance weighting. Quantile kriging was shown to be the most robust and least biased of the interpolation methods they studied. Additionally, the method's non-parametric uncertainty estimates successfully predicted zones of high estimation error for each test case.

Quantile kriging is a transformation-based variant of the kriging estimator in which the  $i^{th}$  concentration sample is assigned the probability that its value is less than or equal to all of the remaining concentration samples. These probabilities create an empirical cumulative distribution function for concentration for a given site and are known to be uniformly distributed (*Hogg and Tanis* 1997). For this reason, the constant  $A$  shown in equation (6.6) was assigned a value equal to  $2\sqrt{3}$  for computing local uncertainty estimates for a uniformly distributed quantity (for details see *Juang et al.* 2001). The term  $2\sqrt{3}\sigma(\bar{u}_j)$  in equation (6.6) then computes the distance between the upper and lower bound estimates from the 95<sup>th</sup> confidence interval for concentration at location  $\bar{u}_j$  within the interpolation domain. For more details on quantile kriging see *Journel & Deutsch* (1997), *Juang et al.* (2001), and Chapter 5. For more details on the spatial modeling used in the test case presented in this chapter, see Chapter 5 and Appendix B.

### **6.3.4 Multiobjective Search & Optimization**

NSGA-II is used to identify high order Pareto surfaces in the LTM methodology. NSGA-II is a second generation evolutionary multiobjective GA developed by *Deb et al.* (2000). It significantly improves upon the original NSGA by (1) invoking a more efficient

nondomination sorting algorithm, (2) eliminating the sharing parameter, and (3) adding an implicitly elitist selection method that greatly aids in capturing high order Pareto surfaces. *Zitzler et al.* (2001) and *Deb et al.* (2001) show that the NSGA-II performs as well or better than the other second generation evolutionary multiobjective algorithms on difficult, high order problems. The following section details basic principals and operators of the NSGA-II.

#### 6.3.4.1 NSGA(II) Basics

Genetic algorithms search a decision space using a process that is analogous to Darwin's "natural selection". The decision variables are first encoded as 0-1 binary strings, or chromosomes. The fitness of each member of a randomly generated initial population of these strings is determined by how well the design satisfies specified objectives and constraints. After each individual is assigned a fitness value, GAs find optimal solutions using three basic operators: (1) selection, (2) crossover (mating), and (3) mutation.

The NSGA-II uses a two-step selection process, which combines both binary tournament selection and  $(\mu + \lambda)$  selection. First, tournament selection allows only the fittest individual from a group of strings randomly drawn from the current population to be placed into the mating population. Next, the crossover operator couples members of the mating population to mate with a specified crossover probability ( $P_c$ ). Mating consists of randomly selecting one or more crossover points at which the strings exchange bit values with each other. Tournament selection and crossover are repeated until a population of  $N$  children has been produced. Mutation then randomly flips binary bits from 1 to 0 or vice versa within the new child population with a given probability of mutation ( $P_m$ ). Lastly, the NSGA-II uses  $(\mu + \lambda)$  selection to choose which of the parent and child designs will survive. In this selection scheme, the populations of  $N$  parent designs and  $N$  child solutions in the current generation  $t$  are combined to yield a selection pool of

$2N$  individuals, from which the  $N$  best individuals are allowed to pass to generation  $t+1$ . This selection method aids the algorithm in efficiently identifying high order Pareto surfaces because it is implicitly elitist (i.e., the best designs are guaranteed to survive into the next generation).

These three operators act to create a new population (or generation) of individual sampling plans with improved average fitness. The Schema Theorem is the general theory describing how these three operators combine to evolve high quality near-optimal solutions [see *Goldberg, 1989* or *Holland, 1975* for more information]. It states that highly fit strings are composed of small chunks of information (or building blocks) that are relevant to the solution of the problem. The GA exerts a selection pressure where only highly fit members are allowed to pass their traits or building blocks to the next generation. Highly fit parent strings are allowed to mate, yielding offspring that inherit building blocks from both parents. It is in this manner that the GA assembles optimal or near-optimal solutions to a problem.

The primary difference between the NSGA-II and single objective GAs is in how fitness is assigned. The NSGA-II evaluates sampling designs in terms of a vector of objectives. A sampling design cannot be assessed in terms of its performance in any single objective because it may perform poorly with respect to the remaining objectives. Instead, the concepts of *Pareto dominance* and *crowding* are used to assign fitness values to sampling designs in the two steps described below (*Deb et al. 2000*).

The first step in fitness assignment employs the concept of *Pareto dominance* in which a design  $\bar{x}$  dominates another design  $\bar{x}'$  if and only if it performs as well as  $\bar{x}'$  in all  $n_{obj}$  objectives and better in at least one. In the NSGA-II's improved nondomination sorting, the algorithm first ranks each design by the number of designs that dominate it. The second step in fitness assignments utilizes the concept of *crowding* (for more details see *Deb et al. 2001*) in

which the average distance between an individual design and those designs within the population that have been assigned the same rank is computed. The crowding operator eliminates the sharing parameter calculations (for more details see Chapter 3) that were required for the original form of the algorithm.

After these two steps, the fitness value of an individual design can be assigned either its rank or its crowding distance as follows. In cases where two designs have different ranks, the individual with the lower rank is preferred (i.e., individuals that are dominated by fewer solutions). Alternatively, if both solutions possess the same rank then the individual with larger crowding distance is preferred (i.e., the individual that adds the most diversity to the population). A diverse population ensures that the NSGA-II will find solutions along the full extent of the Pareto surface.

#### 6.3.4.2 Design & Parameterization

Chapter 4 introduces a multi-population approach for automating parameter specification for the NSGA-II. The methodology combines concepts from previous GA design methodologies (*Reed et al.* 2000b, *Reed et al.* 2001) and the “parameter-less GA” methodology presented by *Lobo* (2000). The methodology utilizes GA design theory to automatically set the probabilities of crossover and mutation as well as the maximum number of generations. The probabilities of crossover and mutation are set equal to 50 percent and  $1/N$ , respectively, where  $N$  is the population size. The maximum number of generations was set equal to 60. Four runs with increasing population sizes from 500 to 4000 members were completed to identify the nondominated set. The runs were halted automatically when further increases in population size resulted in less than a 10 percent increase in the number of nondominated solutions identified. See Chapter 4 for more details.

#### 6.3.4.3 Penalizing Constraint Violations

The constrained problem formulation given in equation (6.1) requires that the number of unestimated points  $U(\bar{x}_\kappa)$  be equal to zero. For any design with unestimated points, the penalty function shown in equation (6.7) was used in place of the fitness function in equation (6.1) to ensure that the solution would be dominated by feasible solutions (i.e., any solution with no unestimated points).

$$F_{PENALTY}(\bar{x}_\kappa) = \begin{pmatrix} f_{COST}^{PENALTY} = f_{COST}^{MAX} + f_{COST} \\ f_{SREE}^{PENALTY} = f_{SREE}^{MAX} + f_{SREE} \\ f_{MASS}^{PENALTY} = f_{MASS}^{MAX} + f_{MASS} \\ f_{UNCERT}^{PENALTY} = f_{UNCERT}^{MAX} + f_{UNCERT} \end{pmatrix}^T, \text{ given } U(\bar{x}_\kappa) > 0 \quad (6.7)$$

Equation (6.7) shows that the 4 component functions of  $F_{PENALTY}$  are the sum of the maximum possible values for the objectives and their actual values from equations (6.3) – (6.6), using only those points where estimates exist. The maximum values of every objective are used because every feasible design is guaranteed to have lower objective values and dominate these solutions. The actual objective values are important to allow the infeasible solutions to compete with one another, even though the estimates lack validity, because feasible solutions can be very near infeasible solutions in decision space and could be lost if the infeasible solutions are immediately removed from consideration.

The maximum value of the normalized cost in this chapter is equal to 58, which is the total number of sampling locations that are available. The use of quantile kriging made computing the maximum value of SREE,  $f_{SREE}^{MAX}$ , very easy because concentrations are transformed to have values in the interval [0, 1], so the maximum squared deviation at each grid point is 1. This results in a maximum SREE value equal to the number of nodes in the interpolation domain (or  $nest = 1666$ ).  $f_{UNCERT}^{MAX}$  was attained by assuming all kriging estimation

variances within the interpolation domain were equal to 1, their maximum possible value. This simplifies equation (6.6) to be the product of  $nest = 1666$  and the constant  $A = 2\sqrt{3}$ , resulting in a maximum value of uncertainty,  $f_{UNCERT}^{MAX}$ , equal to 5770. The maximum relative mass estimation error,  $f_{MASS}^{MAX}$ , was assumed to be 3300. This value was computed using  $Mass_{All}$  and the mass estimate that would be attained if every location in the interpolation domain had a concentration value equal to 6500 mg per m<sup>3</sup>, which is the maximum concentration in the monitoring dataset used in this chapter.

## 6.4 Results & Discussion

Recall that if the  $i^{th}$  monitoring well is selected for sampling then PCE is sampled at all the possible sampling locations along its vertical axis. This is equivalent to monitoring a single constituent at 29 monitoring wells, which results in a decision space of more than 500 million possible sampling designs (i.e.,  $2^{29}$  sampling designs). Using the NSGA-II to identify the subset of sampling designs that are optimal with respect to cost, SREE, mass estimation error, and uncertainty reduces the set of designs that must be considered from 500 million to 1156 designs identified on the Pareto surface. Although the 4-dimensional Pareto surface cannot be visualized, the set of 1156 designs can inform decision making as follows. Initially, interactions and conflicts among the design objectives are identified using a variety of visualizations derived from the full set of nondominated solutions. Once these interactions and conflicts are discovered, they are then used to identify acceptable objective bounds and negotiate a compromise design. This decision process is illustrated for the LTM design application in the following sections.

### 6.4.1 Understanding the Interactions of Design Objectives

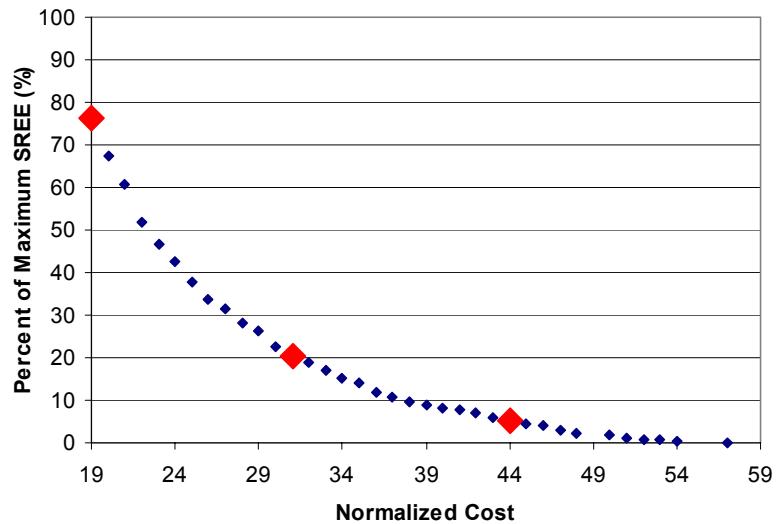
This section of the chapter illustrates the process of learning how cost, SREE, mass estimation error, and uncertainty interact and affect sampling designs. This process begins by

analyzing pairs of the objectives that are known to conflict. These 2-dimensional tradeoffs are subsets of the overall 1156 member nondominated set. These tradeoffs are found by identifying only those solutions that are nondominated in terms of cost and one other objective, independent of the remaining objectives' values. Three designs from each of these tradeoffs have been visualized to demonstrate their effect on sampling schemes as well as to promote the discovery of any additional objective conflicts. The next two sections demonstrate this process of discovery for the LTM application.

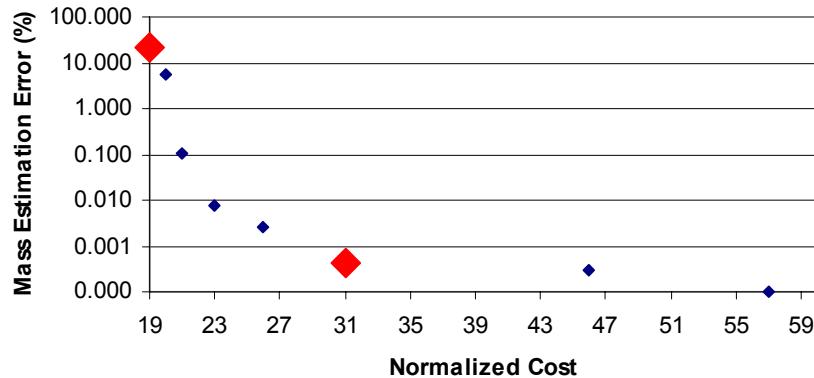
#### 6.4.1.1 Visualizing Known Objective Conflicts

Figures (6.2a) – (6.2c) present tradeoff curves for the following pairs of known conflicting objectives: (1) Cost—SREE, (2) Cost—Mass, and (3) Cost—Uncertainty. The solutions on these figures represent the lowest possible SREE, mass error, and uncertainty values attainable for each level of sampling cost. Both the Cost—SREE and Cost—Uncertainty tradeoffs in Figures (6.2a) and (6.2c) sample between 19 and 58 locations. Figure (6.2b) shows that there are significantly fewer nondominated solutions in the Cost—Mass tradeoff, which is presented using a log scale because mass estimation errors ranged over 5 orders of magnitude. The gaps present in Figure (6.2b) are not due to a shortcoming of the NSGA-II; they result because nondominated sorting considered only cost and mass estimation error. For example, the gap that exists between 31 and 46 sample solutions occurs because the 31 sample solution dominates all of the solutions that sample less than 46 locations (i.e., it is lower in both cost and mass estimation error). Only 21 sampling locations were required to attain a mass estimate that fell within 0.1 percent of  $Mass_{All}$ , showing that this objective is far less constraining than either SREE or uncertainty.

(a)



(b)



(c)

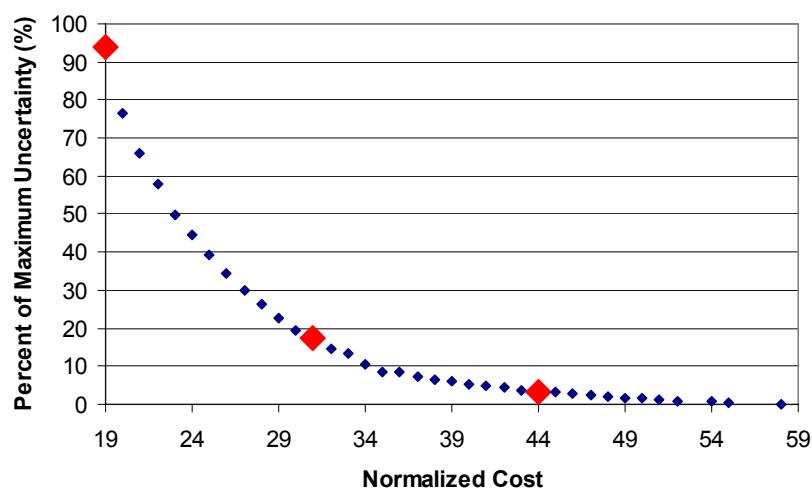


Figure 6.2 (a) Cost—SREE tradeoff (b) Cost—Mass tradeoff (c) Cost—Uncertainty tradeoff

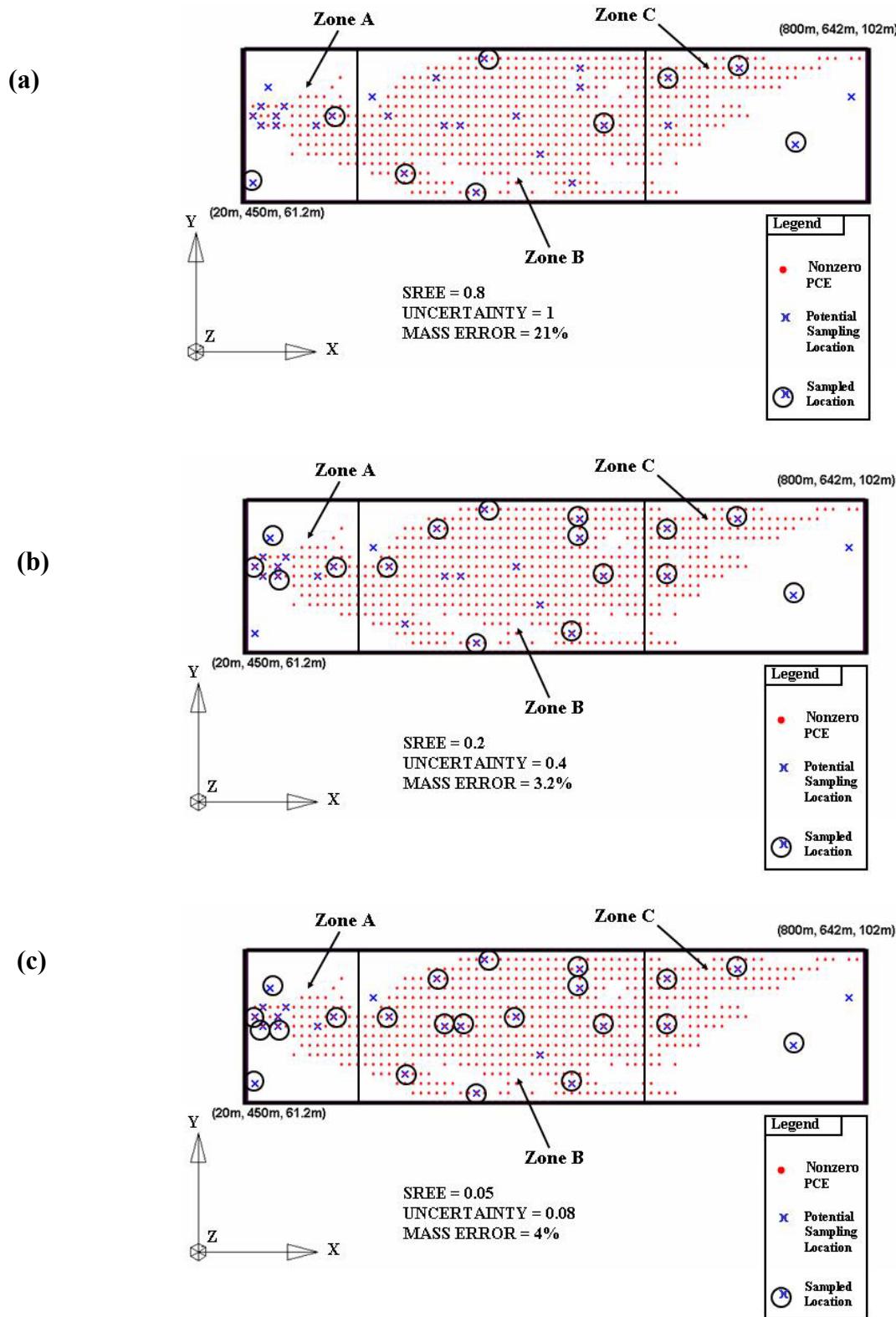


Figure 6.3 Sampling locations corresponding to solutions from the Cost—SREE tradeoff (a) 19 sample solution (b) 31 sample solution (c) 44 sample solution

The solutions highlighted in the tradeoffs of Figure (6.2) sample between 19 and 44 locations and are visualized in Figures (6.3) – (6.5) to investigate how the known objective conflicts are affecting sampling geometries. Comparison of the 19 sample solutions in Figures (6.3a) and (6.5a) show that the least cost solution for the Cost—Uncertainty tradeoff is very similar to the least cost solution for the Cost—SREE curve. The 19 sample solution for the Cost—Mass curve is identical to the 19 sample solution for the Cost—SREE curve and is not shown. The only differences between the solutions in Figures (6.3a) and (6.5a) lie in Zone B, where well 17 is sampled in place of well 15, and in Zone C where well 23 is sampled in place of well 22, resulting in a 6 percent reduction in the uncertainty measure. Chapter 5 shows for the same test case presented in this chapter that the zone of highest uncertainty is located near  $X = 400\text{m}$  and  $Y = 550\text{m}$ , very near wells 17 and 23, which explains why these sampling locations helped to reduce uncertainty.

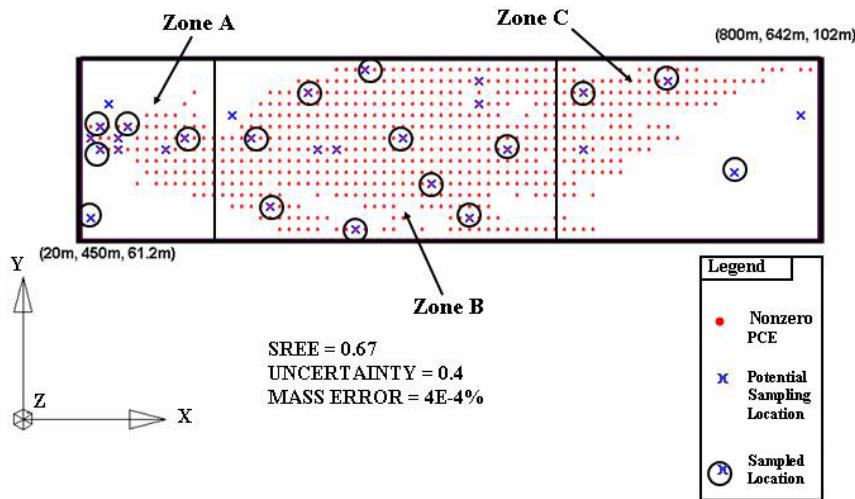


Figure 6.4 Sampling locations corresponding to the 31 sample solution from the Cost—Mass tradeoff

Figures (6.3b), (6.4), and (6.5b) present the 31 sample, nondominated solutions from the Cost—SREE, Cost—Mass, and Cost—Uncertainty tradeoffs, respectively. The 31 sample

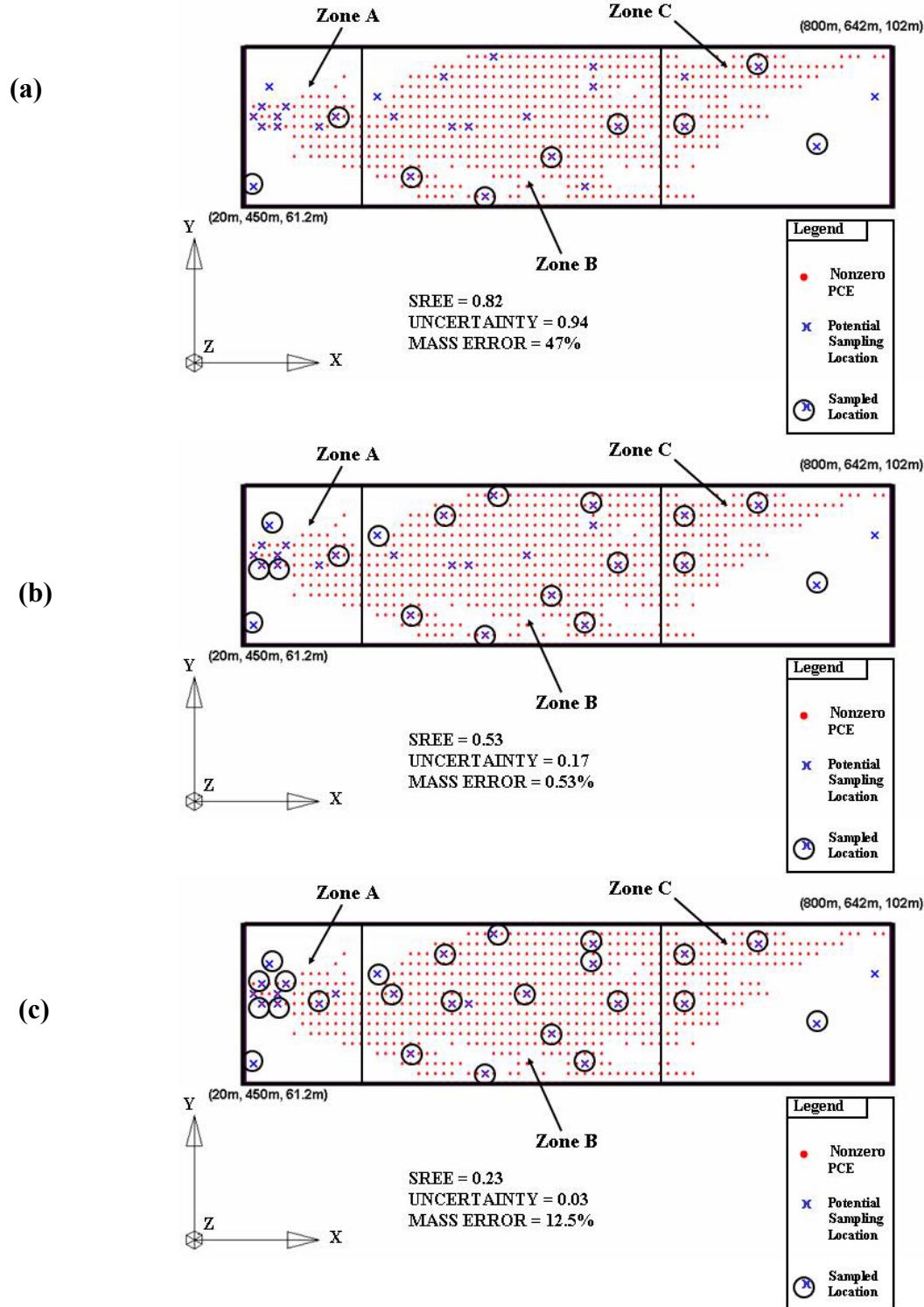


Figure 6.5 Sampling locations corresponding to solutions from the Cost—Uncertainty tradeoff  
 (a) 19 sample solution (b) 31 sample solution (c) 44 sample solution

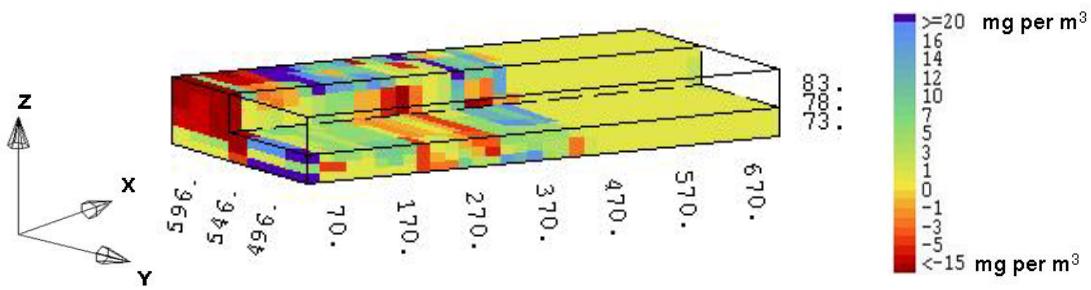
solutions all affirm that sampling wells along the plume's outer boundaries and near its leading edge are important for reducing SREE, mass error, and uncertainty. The 44 sample solutions presented in Figures (6.3c) and (6.5c) from the Cost—SREE and Cost—Uncertainty tradeoffs, respectively, show that both SREE and uncertainty are further reduced by sampling areas of high concentrations within Zone A and wells within the interior of Zone B.

As expected, the 44 sample solution shown in Figure (6.5c) uses more locations to decrease uncertainty, but note the surprising increase in mass estimation error relative to the 31 sample solution in Figure (6.5b). A very interesting multi-well interaction between well numbers 1, 27, and 28 [see Figure (6.1b)] causes this unexpected increase in mass estimation error. These wells sample the minimum and maximum concentrations within Zone A. Specifically, well 1 provides concentration values that exceed 4500 mg per m<sup>3</sup> while wells 27 and 28 sample locations where there is no PCE. All of the nondominated designs that sample 44 locations and include wells 27 and 28 but not well 1 resulted in mass estimation errors that exceed 9 percent. The increased mass estimation error results because wells 27 and 28 in the absence of well 1 cause the mass in Zone A to be severely underestimated. This result identifies an Uncertainty—Mass Error conflict. The next section demonstrates how visualization can be used to identify additional objective interactions.

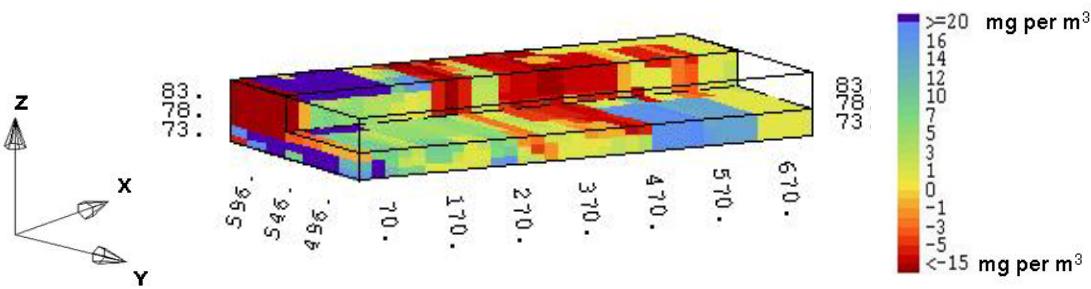
#### *6.4.1.2 Discovering and Understanding Additional Objective Interactions*

For the sake of illustration, the 31 sample cost level is used in this chapter to identify potential conflicts among SREE, mass error, and uncertainty. Figures (6.3b), (6.4), and (6.5b) presented the 31 sample, nondominated solutions from the Cost—SREE, Cost—Mass, and Cost—Uncertainty tradeoffs, respectively. To improve the clarity of the subsequent comparative discussion of these designs, they are referred to as SREE(31), Mass(31), and Uncertainty(31).

**(a) SREE = 20 percent of maximum value**



**(b) SREE = 67 percent of maximum value**



**(c) SREE = 53 percent of maximum value**

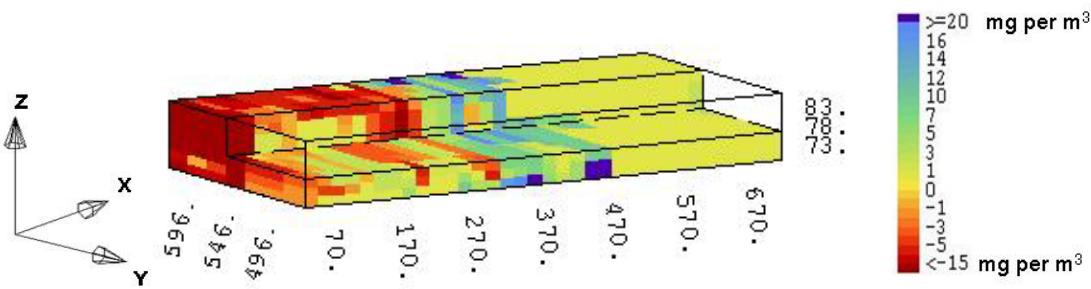


Figure 6.6 Three dimensional cross sectional views of how SREE changes for (a) the SREE(31) solution (b) the Mass(31) solution (c) the Uncertainty(31) solution

SREE increases from its minimum value of 0.2 in the SREE(31) solution to 0.67 and 0.53 for the Mass(31) and Uncertainty(31) solutions, respectively. Although the SREE performance measure provides a ranking of the sampling designs according to the relative accuracy of their local estimates, it does not provide a clear understanding of what these relative rankings

represent. Figures (6.6a) – (6.6c) provide a direct assessment of how the sampling geometries presented in Figures (6.3b), (6.4), and (6.5b) are affecting SREE. Figure (6.6) visualizes how the relative local errors in SREE [i.e.,  $- \left( c_{all}^*(\bar{u}_j) - c_{est}^\kappa(\bar{u}_j) \right)$  in equation (6.4)] vary throughout the interpolation domain in terms of milligrams per m<sup>3</sup>. Positive values represent the case where the PCE concentration estimate  $c_{est}^\kappa(\bar{u}_j)$  is overestimated relative to the estimate attained using all 58 available sampling locations. Conversely, negative values represent underestimates. For SREE, more than 50 percent of the  $c_{all}^*(\bar{u})$  values were less than 12 mg per m<sup>3</sup> and 70 percent where less than 20 mg per m<sup>3</sup>, hence any local errors exceeding  $\pm 5$  mg per m<sup>3</sup> represent significant increases. The 31 sample solutions visualized in Figure (6.6) clearly illustrate significant local errors, which motivated the selection of higher cost 44 sample designs in the negotiation stage of this methodology (described below) to reduce local errors within acceptable limits.

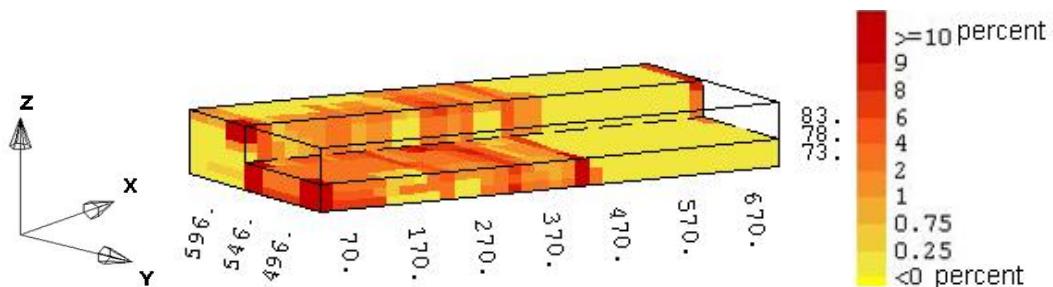
Figure (6.6a) shows that the majority of local errors for the SREE(31) solution are found in Zones A and B of the interpolation domain, while the local errors are nearly zero in the leading edge of the plume. The SREE(31)'s sampling plan in Figure (6.3b) shows 4 of the 5 the wells nearest to the leading edge of the plume are sampled while several wells within Zones A and B are unsampled, resulting in increased local estimation errors in these areas of the domain. Although this design has the lowest SREE value of the three 31 sample solutions being considered, it has the highest mass estimation error at 3.2 percent, showing a potential SREE—Mass tradeoff. This result is surprising because minimizing local estimation error throughout the interpolation domain would seem to implicitly result in accurate mass estimates. This contradiction can be understood by looking at the Mass(31) solution, which has negligible mass estimation error while having the highest SREE value of the three 31 sample solutions. Figure

(6.6b) visualizes the Mass(31) solution's local estimation errors and provides insight into how the SREE and mass error objectives conflict. The figure shows that local concentrations are significantly overestimated in the source area of the plume (Zone A) and severely underestimated in large portions of Zones B and C. In effect, the sampling plan shown in Figure (6.4) balances both overestimation and underestimation throughout the plume to yield a mass estimate with a coincidentally high relative accuracy while severely sacrificing the quality of the interpolated map of PCE (i.e., SREE). Severe overestimation and underestimation of contaminant concentrations can have very negative consequences in correctly assessing risk associated with a site. These results show that global measures such as mass should be counterbalanced with localized measures such as SREE and uncertainty.

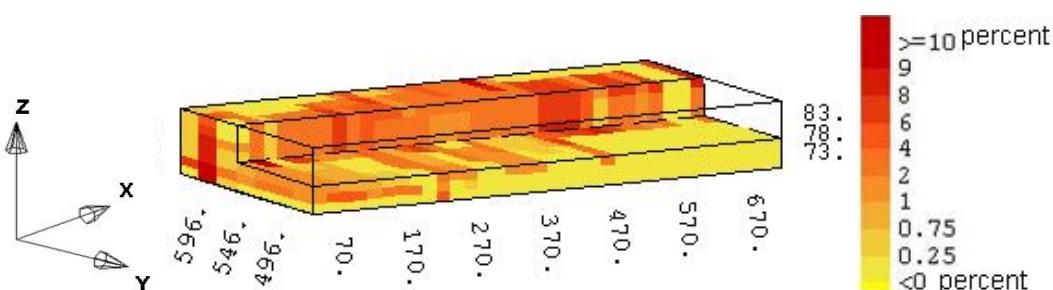
Figures (6.7a) – (6.7c) show how the local uncertainty estimates change in each of the three 31 sample solutions. These figures show the absolute percentage change of  $\sigma(\bar{u}_j)$  for the  $j^{th}$  location within the interpolation domain [see equation (6.6)]. It should be noted that a small percentage change in the kriging standard error [i.e.,  $\sigma(\bar{u}_j)$ ] for the  $j^{th}$  location yields potentially huge changes in uncertainty calculations. For example, consider a location  $\bar{u}_j$  with a mean concentration estimate equal to 7 mg per m<sup>3</sup> and kriging standard error  $\sigma(\bar{u}_j) = 0.2$  (typical for this case study). A 4 percent increase in  $\sigma(\bar{u}_j)$  causes the 95<sup>th</sup> percentile estimate for location  $\bar{u}_j$  to increase from 182 mg per m<sup>3</sup> to 482 mg per m<sup>3</sup>, representing a 260 percent increase.

Although the Uncertainty(31) solution has the minimum uncertainty value of the three 31 sample solutions, it has increased SREE and mass estimation errors relative to the SREE(31) and Mass(31) solutions. The SREE—Uncertainty and Mass—Uncertainty conflicts occur because both SREE and mass error are heavily affected by extreme concentrations, especially

**(a) Uncertainty = 40 percent of maximum value**



**(b) Uncertainty = 45 percent of maximum value**



**(c) Uncertainty = 17 percent of maximum value**

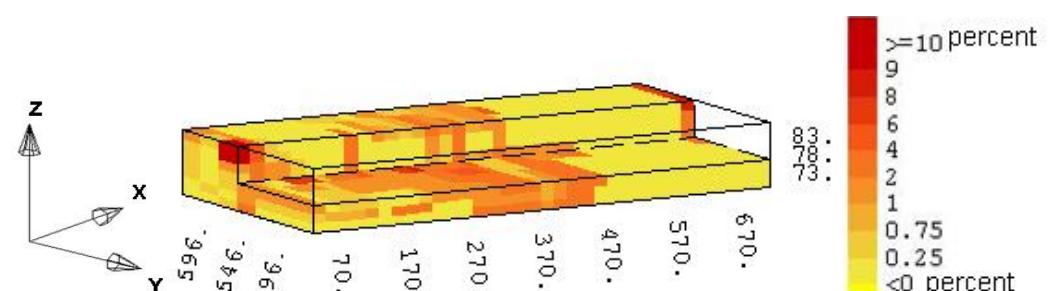


Figure 6.7 Three-dimensional cross sectional views of how Uncertainty changes for (a) the SREE(31) solution (b) the Mass(31) solution (c) the Uncertainty(31) solution

within the plume's source area, while uncertainty is independent of the concentration values and is solely a function of how well the sampling locations are distributed throughout the interpolation domain. Figure (6.7b) shows that the Mass(31) solution greatly increased uncertainty throughout the interpolation domain relative to the Uncertainty(31) solution in Figure (6.7c). Figures (6.6c) and (6.7a) confirm the SREE—Uncertainty conflict. The Uncertainty(31)

solution in Figure (6.6c) severely underestimates PCE concentrations within the source area of the plume, resulting in its increased SREE value of 0.53, while Figure (6.7a) shows that the SREE(31) solution significantly increased uncertainty within Zones A and B of the plume. The increased local uncertainties for the 31 sample solutions from the Cost—SREE and Cost—Mass tradeoffs resulted because they lack samples along the plume’s outer boundary or near the zone of highest uncertainty centered at X = 400m and Y = 550m [see Figures (6.3)-(6.5)].

#### ***6.4.2 Negotiating a Final Design***

Recall that the LTM application presented in this chapter has more than 500 million possible sampling designs. High order Pareto optimization using the NSGA-II reduced the set of designs to be considered to the 1,156-member nondominated set that are optimal in terms of cost, SREE, mass error, and uncertainty. The previous sections visualized a total of 7 sampling schemes to provide stakeholders and regulators with an improved understanding of their design objectives. The next two sections provide an illustrative example of how stakeholders and regulators can then negotiate a final compromise sampling scheme.

##### ***6.4.2.1 Bounding Stakeholder Expectations***

The first decision that faces stakeholders is selecting acceptable bounds on each objective. For this example, a conservative cost level of 44 sampling locations was selected to reduce costs by nearly 25 percent while minimally increasing the remaining objectives. Note that this example merely illustrates the negotiation process and that higher or lower cost levels could be selected. By considering only those designs at the 44 sample cost level, the set of potential designs further reduces from 1156 to 46 potential monitoring designs.

The next step in the negotiation process is to use the objective conflicts that occur in the 44 sample designs to bound stakeholder expectations and set “acceptable” upper bound values

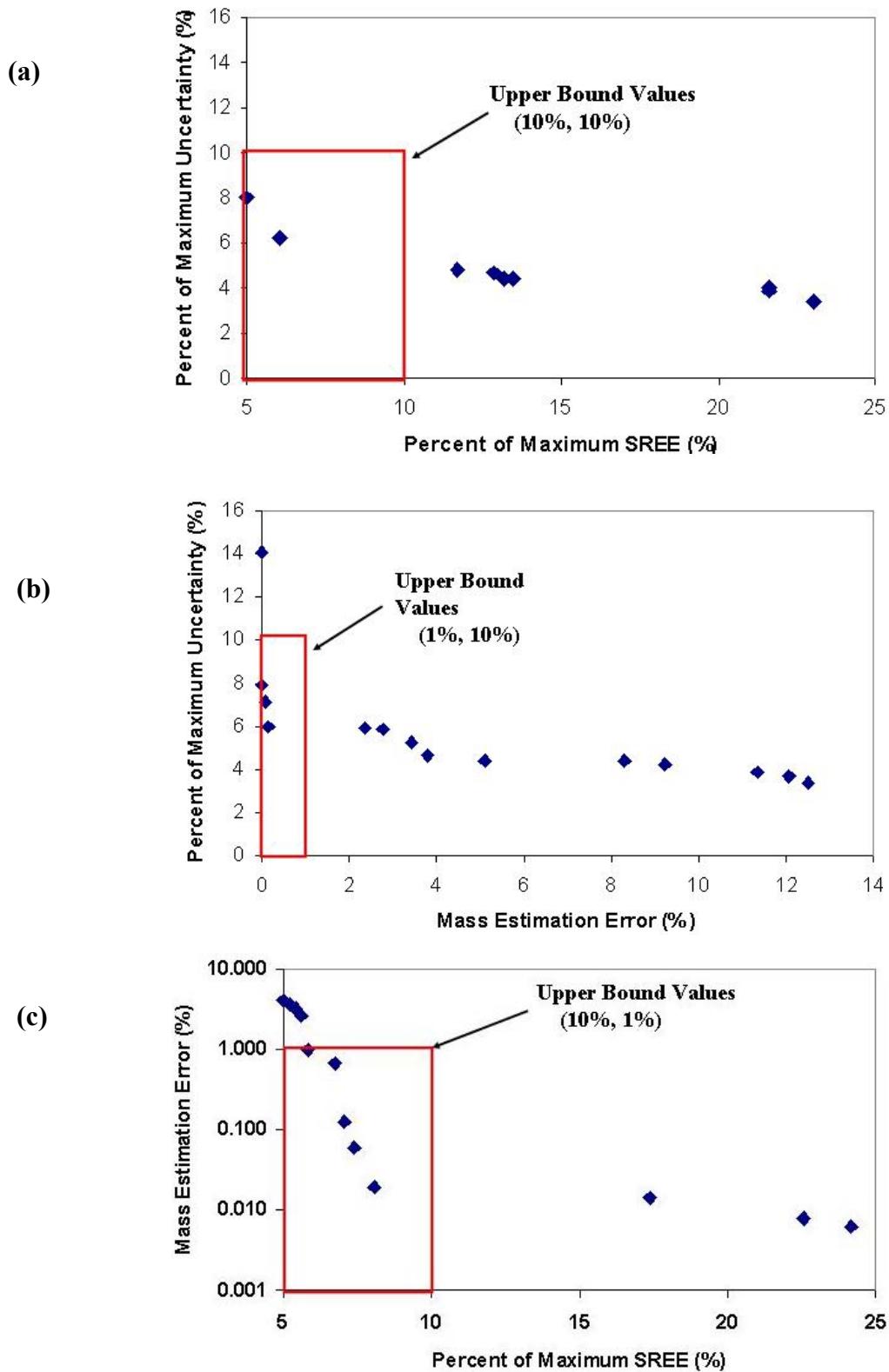


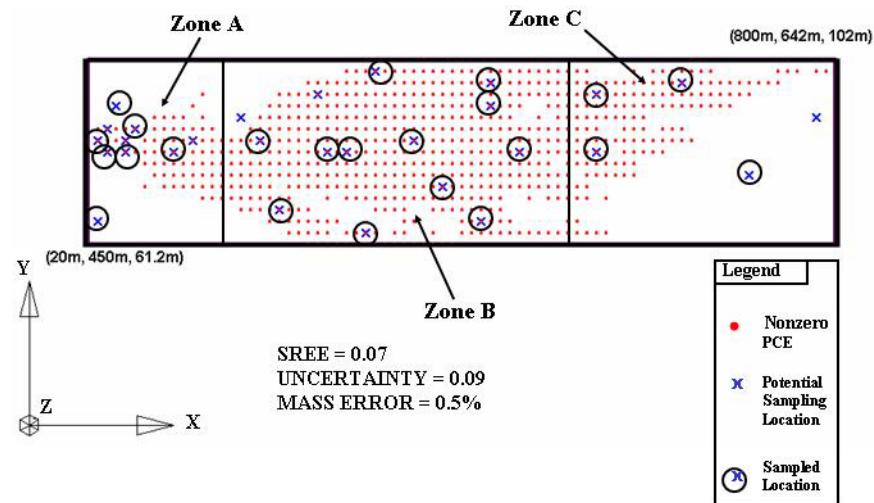
Figure 6.8 Objective conflicts between (a) SREE and Uncertainty (b) Mass Error and Uncertainty (c) SREE and Mass Error for the 44 sample solutions

for the remaining objectives. The SREE—Uncertainty tradeoff shown in Figure (6.8a) shows that increasing SREE beyond 10 percent of its maximum value in the nondominated set has a negligible effect on reducing uncertainty. The Mass—Uncertainty tradeoff presented in Figure (6.8b) shows that uncertainty can be significantly reduced with less than a 1 percent increase in mass estimation error. Lastly, the SREE—Mass tradeoff in Figure (6.8c) shows that a very significant decrease in mass estimation error can be attained for a relatively small increase in SREE. Given these findings, upper bound values for SREE, mass error, and uncertainty were set equal to 10%, 1%, and 10%, respectively, as shown in Figure (6.8).

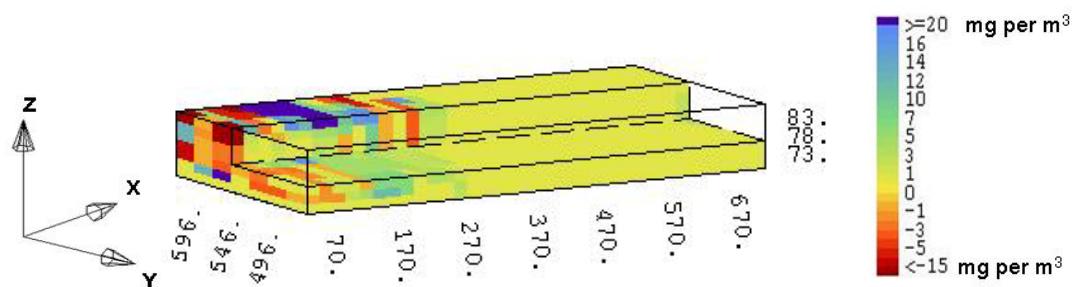
#### 6.4.2.2 *Striking the Balance*

The final step in the negotiation process is to search the nondominated set for designs that satisfy the objective bounds [i.e., (Cost = 44, SREE < 10%, Mass Error < 1%, Uncertainty < 10%)]. Setting these objective bounds reduces the number of possible sampling designs that must be considered from 46 to the single compromise solution illustrated in Figure (6.9). Although setting objective bounds will not always yield a single solution, it will vastly limit the number of designs that must be considered. The solution shown in Figure (6.9a) reduces sampling costs by nearly 25 percent in any given monitoring period while minimally increasing uncertainty, maintaining a high quality map of the plume, and accurately quantifying the mass of PCE within the subsurface. Figure (6.9b) shows that the majority of the local PCE estimation errors throughout the plume were less than 5 mg per m<sup>3</sup>. Figure (6.9c) shows that local uncertainty estimates remained unchanged for a large portion of the plume. Although these observations are important, the most important contribution of Figure (6.9) is how the figure provides regulators and stakeholders with a direct understanding of the effects of their negotiated objective bounds.

(a)



(b) SREE = 7 percent of maximum value



(c) Uncertainty = 9 percent of maximum value

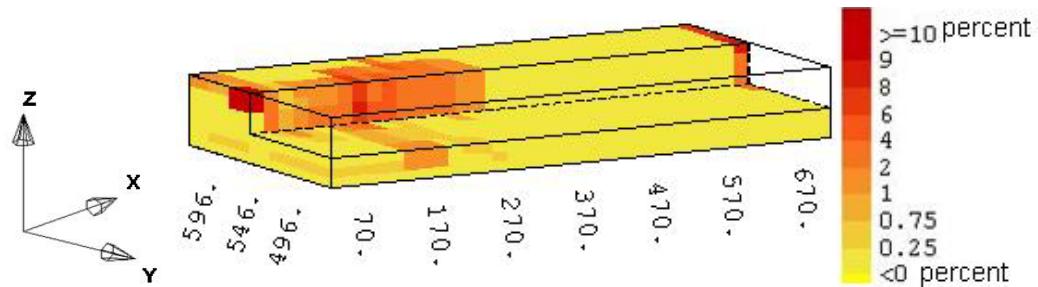


Figure 6.9 Compromise solution's (a) XY-plane view of sampling locations (b) three dimensional cross section of how SREE changes (c) three dimensional cross section of how Uncertainty changes

Current standard practice in redundancy analysis uses trial-and-error analysis to eliminate sampling locations (see *Johnson et al. 1996, Cameron & Hunter 2000, Aziz et al. 2000*). In these

methodologies, locations thought to be redundant are eliminated and visualization is used to determine the effect of these locations on the quality of the interpolated plume map. The process is repeated for tens if not hundreds of designs until the practitioner is satisfied. This time consuming process does not comprehensively search the decision space or account for multiple objectives. The single compromise solution shown in Figure (6.9) explicitly balances the stakeholders' objectives, required less than a day of computing time, and visualization of only 8 designs. Moreover, the optimization methodology used to attain the compromise solution provides practitioners with a better understanding of how their design preferences interact with the physical monitoring system.

## 6.5 Conclusions

The optimization methodology presented in this chapter demonstrates that algorithms such as the NSGA-II that are capable of high order Pareto optimization can serve as interfaces between the human decision process and engineered water resources systems. The LTM application used to demonstrate the methodology shows how multiobjective optimization combined with visualization can aid practitioners in selecting, understanding, and balancing these performance objectives when seeking a single compromise solution. The monitoring application successfully balances the following four objectives: (1) minimizing sampling costs, (2) maximizing the quality of interpolated plume maps (i.e., SREE), (3) maximizing the relative accuracy of contaminant mass estimates, and (4) minimizing estimation uncertainty.

These objectives result in a 4-dimensional Pareto surface that was explored using 2-dimensional tradeoffs between selected pairs of objectives. First, objective pairs that are known to conflict were explored through visualizing 7 sampling designs taken from the Cost—SREE, Cost—Mass, and Cost—Uncertainty tradeoffs. Visualization of these designs enabled

stakeholders to discover additional objective conflicts and their effects on the physical monitoring system. The final step in the methodology builds upon the improved stakeholder understanding of design objective interactions to negotiate acceptable bounds for all of the performance criteria used in a monitoring application. These bounds were then used to search the high dimensional Pareto set of optimal sampling strategies for a final compromise solution. The processes of discovery and negotiation demonstrated in this chapter through the use of high order Pareto optimization hold significant potential as tools that can be used in the balanced design of water resources systems.



## 7 CONCLUDING REMARKS

### 7.1 Thesis Contributions

This thesis contributes a highly adaptable groundwater monitoring design methodology that can reduce the fiscal burden of long term stewardship for contaminated sites while allowing stakeholders to select, understand, and balance their design objectives. The methodology has immediate potential for application at real world sites where it can be used to save millions of dollars that are currently spent on sampling spatially redundant data. The underlying structure of this thesis is based on decomposing the LTM problem into three steps: (1) select and balance design objectives, (2) choose a plume interpolation method, and (3) quantify design tradeoffs using multiobjective genetic algorithms (GAs). This thesis focused on the specific challenges practitioners will face in implementing the methodology proposed in this research.

The first step of the methodology involves specification, understanding, and negotiated balance of LTM design objectives to assess the “quality” of designs. Properly assessing the quality of designs is the most pivotal and challenging issue facing designers from any discipline because they must capture all of their preferences for a system. This is particularly challenging for LTM design and, in general, environmental design because quality is defined by the conflicting preferences of contaminated sites’ owners, multidisciplinary teams of experts, regulatory agents, and the public. This thesis suggests that high order Pareto optimization (i.e., optimizing a system for more than 2 objectives) combined with visualization can be used for selecting and balancing design objectives. Combining higher order Pareto optimization with visualization allows stakeholders to assess the mathematical models used to represent their objectives, discover how their objectives are affecting designs, and negotiate a final design that balances their conflicting design preferences.

In this thesis, plume interpolation is assumed to be required for both evaluating and visualizing design objectives. The second step of the LTM design methodology focuses on identifying which plume interpolation scheme should be used for these purposes. The basic goal of plume interpolation is to attain unbiased maps of contaminant concentrations and local uncertainties. Preferential sampling and the highly skewed nature of groundwater contamination can combine to severely bias interpolation estimates and any subsequent decisions that must be made using these estimates. This thesis directly illustrates that these factors bias the standard practice of using cross-validation to rank the relative performances of interpolation methods. These biased rankings can mislead the practitioner into using a less effective interpolation method and could adversely affect subsequent LTM design decisions. Quantile kriging was selected to evaluate the LTM designs in this research because it was the most robust of the interpolation methods, showing the least bias from both preferential sampling and the variability of contaminant data. These results warrant further studies into the applicability of quantile kriging to data sets from other fields ranging from mining to life sciences, where nonstationary interpolation also plays a vital role.

The final step of the LTM design methodology seeks to quantify high order Pareto surfaces using effectively designed multiobjective GAs. A multiobjective algorithm that is capable of high order Pareto optimization can be used as a tool of discovery for designers, serving as an interface between the human decision process and the design of a physical system. This capability was demonstrated in this thesis when the Nondominated Sorted Genetic Algorithm-II (NSGA-II) identified important system properties that had been previously unknown and would have remained unknown without the algorithm.

Real world applications of high order multiobjective optimization prior to this thesis have been constrained by computational limitations. This thesis overcomes these limitations by developing a design methodology for the NSGA-II that fully exploits the algorithm's efficiency to enable the automatic solution of this new class of problems. The NSGA-II design methodology was used in Chapter 6 to solve the first application of evolutionary multiobjective algorithms to a real-world problem with 4 objectives. The design methodology treats the NSGA-II as a system with discernable properties and utilizes valuable theoretical work from the field of genetic and evolutionary computation to discern these properties. Theoretical relationships allow practitioners to gain direct insights into the algorithm's performance as a function of its input parameters while also reducing the number of parameter combinations that must be considered relative to the trial-and-error methods that have traditionally been used in real world applications. Trial-and-error methods inherently treat evolution-based solution methods as randomized black boxes and obfuscate important trends in performance. In this thesis, recognizing the influence of population sizing and selection pressure on performance facilitated the selection of the NSGA-II and its efficient use as a tool for design and discovery. The NSGA-II design methodology developed in this thesis can be used by designers in any discipline to efficiently discover, understand, and balance tradeoffs among their design objectives.

## 7.2 Future Research

Although this thesis successfully addresses spatial redundancy analysis using a highly adaptive multiobjective methodology, several issues deserve further consideration:

- How can the methodology be extended to account for contaminant reaction and migration in both space and time?

- How can “soft” data (e.g., numerical modeling data) be combined with “hard” data (e.g., groundwater monitoring samples) to improve LTM design?
- How can multiple contaminants and multiple sampling technologies be integrated into LTM optimization?
- Can both LTM and remediation be optimized simultaneously at contaminated groundwater sites?
- Can the frequency of sampling for costly contaminant constituents be reduced using predictive relationships derived from lower cost measures of groundwater quality?
- When and how should negotiation be integrated into multiobjective design frameworks?

Spatial redundancy analysis assumes periodic re-assessment of the effectiveness of the monitoring network to determine whether locations previously classified as redundant still minimally affect LTM design objectives. Re-assessment reflects the temporal component of the LTM design problem, which results because the evolution of groundwater contaminant plumes occurs in both time and space. Existing redundancy studies have most often considered only time or only space. The few studies that have considered both space and time have generally used only single objective problem formulations. A potential extension of this thesis and previous studies lies in investigating a multiobjective framework that evaluates designs in both time and space. This framework would require the investigation of novel methods for creating spatiotemporal plume maps for single or multiple contaminant species that could range from statistical data-driven approaches to physics-based flow-and-transport simulation.

Flow-and-transport simulation is a tremendous, but costly source of spatiotemporal “soft” data that has significant potential for helping to address several of the outstanding LTM issues listed above. The first question LTM designers face is “Does the size and complexity of my site justify or preclude the costs of physics-based simulation?”. The answer to this question for large sites with tens or hundreds of monitoring wells is more straightforward because long-term stewardship will likely cost several million dollars. The maintenance of a simulation model would require a relatively small percentage of the overall cost and has the benefit of allowing stakeholders to jointly optimize remediation and monitoring.

For smaller sites, the cost of maintaining a simulation model may be too large to justify, although the persistence and number of contaminants at a site should be weighed as carefully as a site’s physical size and number of monitoring points. It is common for small sites in the United States to monitor several contaminants of concern. The cost of analyzing groundwater for these chemicals can range from tens to thousands of dollars per sample (*Rast 1997*). If the contaminants are persistent, LTM may be required for 30 years or more at a cost of several million dollars. The Department of Energy (DOE) has estimated that it will spend more than \$5 billion on long-term site management within the six year period between 2000 and 2006 (*DOE 2001*). Moreover, the DOE estimates that long-term site management will cost \$100 million per year for the next 70 years. The argument for using spatiotemporal LTM optimization techniques that integrate flow-and-transport simulation gains credence even for small sites when multiple, persistent contaminants are present.

After determining that fate-and-transport simulation is justified for a site, LTM designers then face the task of deciding how to combine model predictions with groundwater samples (i.e., hard data). Bayesian approaches hold significant potential for addressing this issue by providing

a “living model” approach where model predictions are continually conditioned to contaminant observations over the course of a site’s lifetime. The primary difficulty of using these methods in combination with multiobjective optimization is their computational complexity. The complexity of these methods grows markedly with the number of contaminants and data sources considered. In these cases, data mining tools such neural networks or decision trees maybe useful for reducing the computational complexity of predictions (see *Michael et al.*, 2002). Moreover, these tools could be used to develop approaches for limiting expensive contaminant observations by replacing them with lower cost groundwater quality measures.

The “living model” approach could greatly improve LTM design, especially if it used in the context of a spatiotemporal multiobjective optimization framework. All multiobjective frameworks for LTM design will result in tradeoff surfaces. Real-world implementations of these frameworks will require the negotiated selection of compromise solutions. Therefore a relevant question is “How should negotiation be integrated into environmental decision making?”. Visualization tools that provide a better understanding of how designs evolve in space and time as a consequence of stakeholders’ objectives could aid the negotiation process. Research is needed into the relative merits of normative approaches (e.g., multi-attribute utility analysis) versus descriptive approaches (e.g., fuzzy analysis) for aiding negotiation and environmental decision making. Improving our understanding of the intricacies of environmental negotiation could potentially revolutionize current industry policies and practices.

<b>7 CONCLUDING REMARKS.....</b>	<b>149</b>
<b>7.1 Thesis Contributions.....</b>	<b>149</b>
<b>7.2 Future Research .....</b>	<b>151</b>

## REFERENCES

- Aly, Alaa, H. and Richard C. Peralta 1999 Comparison of a genetic algorithm and mathematical programming to the design of groundwater cleanup systems, *Water Resources Research*, 35(8), 2415-2425.
- Armstrong, M. 1984 Problems with Universal Kriging, *Mathematical Geology*, 16(1), 101-108.
- ASCE Task Committee on Geostatistical Techniques 1990a Review of Geostatistics in Geohydrology I: Basic Concepts, *Journal of Hydraulic Engineering*, 116(5), 612-632.
- ASCE Task Committee on Geostatistical Techniques 1990b Review of Geostatistics in Geohydrology II: Applications, *Journal of Hydraulic Engineering*, 116(5), 633-658.
- ASTM 1995 *Standard Guide for Risk-Based Corrective Action Applied at Petroleum Release Sites* (E1739-95), American Society for Testing and Materials, West Conshohocken, PA.
- Aziz, J. J., C. J. Newell, H. S. Rifai, M. Ling, and J. R. Gonzales 2000 *Monitoring and Remediation Optimization System (MAROS): Software User's Guide*, Version 1, United States Air Force Center for Environmental Excellence, Brooks AFB: San Antonio, TX.
- Bäck, T. 1996 *Evolutionary Algorithms in Theory and Practice*, Oxford University Press: New York, NY.
- Barry, D. A. and G. Sposito 1990 Three-dimensional statistical moment analysis of the Stanford/Waterloo Borden tracer data. *Water Resources Research*, 26(8), 1735-1747.
- Bogaert, P. and D. Russo 1999 Optimal spatial sampling design for the estimation of the variogram based on a least squares approach, *Water Resources Research*, 35(4), 1275-1289.

Buras, Nathan 2001 Water Resources—Unresolved Issues, *Journal of Water Resources Planning and Management*, 127(6), 353.

Cameron, K. and P. Hunter 2000 Optimization of LTM Networks: Statistical Approaches to Spatial and Temporal Redundancy. Proceedings from the American Institute of Chemical Engineers, 2000 Spring National Meeting, Remedial Process Optimization Topical Conference, Atlanta, GA.

Chilès, J. P. and P. Delfiner 1999 *Geostatistics: Modeling Spatial Uncertainty*, Wiley Series in Probability and Statistics, John Wiley & Sons, Inc., New York, NY.

Chiueh, P. T., S. L. Lo, and C. D. Lee 1997 Prototype SDSS for using probability analysis in soil contamination, *Journal of Environmental Engineering*, 123, 514-519.

Christakos, G. and R. A. Olea 1988 A multiple-objective optimal exploration strategy, *Mathematical and Computer Modeling*, 11, 413-418.

Cieniawski, S. E. 1993 *An Investigation of the Ability of Genetic Algorithms to Generate the Tradeoff Curve of a Multi-Objective Groundwater Monitoring Problem*, Masters Thesis, University of Illinois: Urbana, IL.

Cieniawski, S. E., J. W. Eheart, and S. Ranjithan 1995 Using genetic algorithms to solve a multiobjective groundwater monitoring problem. *Water Resources Research*, 31(2), 399-409.

Coello, C. A. C. 1999 A comprehensive survey of evolutionary-based multiobjective optimization techniques. *Knowledge and Information Systems*, 1(3), 269-308.

Cooper, R. M., and J. D. Istok 1988a Geostatistics applied to groundwater contamination I: Methodology, *Journal of Environmental Engineering*, 114(2), 270-285.

Cooper, R. M., and J. D. Istok 1988b Geostatistics applied to groundwater contamination II: Application, *Journal of Environmental Engineering*, 114(2), 287-299.

Cooper, R. M., and J. D. Istok 1988c Geostatistics applied to groundwater pollution III: Global Estimates, *Journal of Environmental Engineering*, 114(4), 915-928.

Crawford, Carol Gotway and G. W. Hergert 1997 Incorporating Spatial Trends and Anisotropy in Geostatistical Mapping of Soil Properties, *Soil Sci. Soc. Am. J.*, 61, 298-309.

Deb, K. and D. E. Goldberg, 1989 An investigation of niche and species formation in genetic function optimization. In J. D. Schaffer, ed. Proceedings of the Third International Conference on Genetic Algorithms, pp. 42-50, Morgan Kaufmann: San Mateo, CA.

Deb, K., S. Agrawal, A. Pratap, and T. Meyarivan 2000 A Fast Elitist Non-Dominated Sorting Genetic Algorithm for Multi-objective Optimization: NSGA-II, Kanpur Genetic Algorithm Laboratory (KanGAL) Report 200001, Indian Institute of Technology, Kanpur, India.

Deb, K. 2001 *Multi-Objective Optimization using Evolutionary Algorithms*, John Wiley & Sons LTD: New York, NY.

Deb, K., L. Thiele, M. Laumanns, and E. Zitzler 2001 Scalable Test Problems for Evolutionary Multi-Objective Optimization, Computer Engineering and Networks Laboratory Report (TIK-112), Department of Electrical Engineering, Swiss Federal Institute of Technology, Zurich, Switzerland.

DeJong, K. A. 1975 *An analysis of the behavior of a class of genetic adaptive systems*, Doctoral dissertation, University of Michigan: Ann Arbor, MI.

Delfiner, P. 1976 Linear estimation of nonstationary spatial phenomena, in *Advanced Geostatistics in the Mining Industry*, (ed.) M. Guarascio, pp. 49-68, D. Reidel, Dordrecht, Holland.

De Neufville, Richard 1990 *Applied Systems Analysis: Engineering Planning and Technology Management*, McGraw-Hill Publishing Company, New York, NY.

Deutsch, Clayton V., and Andre G. Journel 1998 *GSLIB: Geostatistical Software Library and User's Guide*, Oxford University Press, New York, NY.

DOE 2001 *Adpative Sampling and Analysis Programs (ASAPs)*, Innovative Technology Summary Report, DOE/EM-0592, USDOE Office of Science and Technology.

Fonseca, C. M. and P. J. Fleming 1995 An overview of evolutionary algorithms in multiobjective optimization. *Evolutionary Computation*, 3(1), 1-16.

Gambolati, G. and G. Volpi 1979 A conceptual deterministic analysis of the kriging technique in hydrology, *Water Resources Research*, 15(3), 625-629.

Goldberg, D. E. and J. Richardson 1987 Genetic algorithms with sharing for multimodal function optimization. In J. J. Grefenstette, ed. *Genetic Algorithms and Their Applications: Proceedings of the Second International Conference on Genetic Algorithms*, pp. 41-49, Morgan Kaufmann: San Mateo, CA.

Goldberg, D. E. 1989 *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley: New York, NY.

Goldberg, D.E., K. Deb, and B. Korb 1989 Messy genetic algorithms: Motivation, analysis, and first results. *Complex Systems*, 3(4), 493-530.

Goldberg, D. E., K. Deb, and J. H. Clark 1992a Genetic algorithms, noise, and the sizing of populations, *Complex Systems*, 6, 333-362.

Goldberg, D. E., K. Deb, and J. Horn 1992b Massive multimodality, deception, and genetic algorithms. In R. Männer & B. Manderick, eds. *Parallel Problem Solving From Nature*, 2, Amsterdam: North-Holland.

Goldberg, D. E. 1998 *The Race, the Hurdle, and the Sweet Spot: Lessons from Genetic Algorithms for the Automation of Design Innovation and Creativity*, Department of General Engineering, University of Illinois at Urbana-Champaign, ILLIGAL Report No. 98007.

Goovaerts, Pierre 1997 *Geostatistics for Natural Resource Evaluation*, Oxford University Press, New York, NY.

Gotway, Carol A., R. B. Ferguson, G. W. Hergert, and T. A. Peterson 1996 Comparison of Kriging and Inverse-Distance Methods for Mapping Soil Parameters, *Soil Sci. Soc. Am. J.*, 60, 1237-1247.

Graham, Wendy and Dennis McLaughlin 1989 Stochastic analysis of nonstationary subsurface solute transport 1: Unconditional Moments, *Water Resources Research*, 25(2):215-232.

Graham, Wendy and Dennis McLaughlin 1989 Stochastic analysis of nonstationary subsurface solute transport 2: Conditional Moments, *Water Resources Research*, 25(11): 2331-2355.

Gray, H. L. and W. R. Schucany 1972 *The Generalized Jackknife Statistic*, STATISTICS: Textbooks and Monographs, Volume 1, Marcel Dekker, Inc., New York, NY.

Gupta, H. V., S. Sorooshian, and P. O. Yapo 1998 Toward improved calibration of hydrologic models: Multiple and noncommensurable measures of information. *Water Resources Research*, 34(4), 751-763.

Halhal, D., G. A. Walters, D. Ouazar, and D. A. Savic 1997 Water network rehabilitation with structured messy genetic algorithm. *Journal of Water Resources Planning and Management*, 123(3), 137-146.

Haan, Charles T. 1977 *Statistical Methods in Hydrology*, The Iowa State University Press, Ames IA.

Herrera de Olivares, G. 1998 *Cost Effective Groundwater Quality Sampling Network Design*, Doctoral dissertation, University of Vermont: Burlington, VT.

Hogg, R. V. and E. A. Tanis 1997 *Probability and Statistical Inference*, 5<sup>th</sup> Ed., Prentice Hall, Upper Saddle River, NJ.

Horn, J. 1997 *The Nature of Niching: Genetic Algorithms and the Evolution of Optimal, Cooperative Populations*, Doctoral dissertation, University of Illinois: Urbana, IL.

Holland, J. H. 1975 *Adaptation in Natural and Artificial Systems*, University of Michigan: Ann Arbor, MI.

Hsueh, Y. W., and R. Rajagopal 1988 Modeling ground-water quality decisions, *Ground-Water Monitoring Review*, 8(4), 121-134.

Hudak, P. F., and H. A. Loaiciga 1992 A location modeling approach for groundwater monitoring network augmentation, *Water Resources Research*, 28(3), 643-649.

Hudak, P.F., and H. A. Loaiciga 1993 An optimization method for monitoring network design in multilayered groundwater flow systems, *Water Resources Research*, 29(8), 2835-2845.

Hughes, J. P. and D. P. Lettenmaier 1981 Data Requirements for Kriging: Estimation and Network Design, *Water Resources Research*, 17(6), 1641-1650.

Ishibuchi, H. and T. Murata 1996 Multi-objective genetic local search algorithm. In Proceedings of 1996 IEEE International Conference on Evolutionary Computation, pp. 119-124, IEEE: Piscataway, NJ.

James, Bruce, R., and Steven M. Gorelick 1994 When enough is enough: The worth of monitoring data in aquifer remediation design, *Water Resources Research*, 30(12), 3499-3513.

Journel, A. G. 1983 Nonparametric estimation of spatial distributions, *Journal of the International Association for Mathematical Geology*, 15(3), 445-468.

Journel, A. G. 1988 Nonparametric geostatistics for risk and additional sampling assessment, in L. H. Keith (ed.) Principles of Environmental Sampling, pp. 45-72, American Chemical Society.

Journel, A. G. and C. V. Deutsch 1997 Rank Order Geostatistics: A proposal for a unique coding and common processing of diverse data, In E.Y. Baafi and N. A. Schofield (eds.) Geostatistics Wollongong '96, Volume 1, Proceedings of the 5<sup>th</sup> International Geostatistics Congress, Wollongong, Australia, Kluwer Academy Publ., Dordrecht, The Netherlands.

Journel, A. G. and J. Huijbregts 1978 *Mining Geostatistics*, Academic Press, New York, NY.

Journel, A. G. and M. E. Rossi 1989 When Do We Need a Trend Model in Kriging?, *Mathematical Geology*, 21(7), 715-739.

Johnson, V. M., R. C. Tuckfield, M. N. Ridley, and R. A. Anderson 1996 Reducing the sampling frequency of groundwater monitoring wells, *Environmental Science & Technology*, 30(1), 355-358.

Kai-Wei, Juang, Dar-Yuan Lee, and Timothy R. Ellsworth Delineation of Heavy Metal Contaminated Soils with Highly Skewed Data using Quantile Kriging, *Journal of Environmental Quality* (in press).

Kitanidis, Peter K. 1983 Statistical estimation of polynomial generalized covariance functions and hydrologic applications, *Water Resources Research*, 19(4), 909-921.

Kitanidis, Peter K. 1997 *Introduction to Geostatistics with Applications in Hydrogeology*, Cambridge University Press, New York, NY.

Knopman, D. S. and C. I. Voss 1989 Multiobjective sampling design for parameter estimation and model discrimination in groundwater solute transport, *Water Resources Research*, 25(10): 2245-2258.

Loaiciga, Hugo, A. 1989 An optimization approach for groundwater quality monitoring network design, *Water Resources Research*, 25(8), 1771-1780.

Loaiciga, Hugo, A., R. J. Charbeneau, L. G. Everett, G. E. Fogg, B. F. Hobbs, S. Rouhani 1992 Review of Ground-Water Quality Monitoring Network Design, *Journal of Hydraulic Engineering*, 118(1), 11-37.

Lobo, F., 2000 *The Parameter-less Genetic Algorithm: Rational and Automated Parameter Selection for Simplified Genetic Algorithm Operation*, Doctoral Dissertation, Universidad Nova de Lisboa, Lisboa, Portugal.

Mahfoud, S. 1995 *Niching Methods for Genetic Algorithms*, Doctoral dissertation, University of Illinois: Urbana, IL.

Marshall, R. and K. Mardia 1985 Minimum norm quadratic estimation of components of spatial covariance, *Journal of the International Association for Mathematical Geology*, 17(5), 517-525.

Massmann, J., and R. A. Freeze 1987a Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 1, Methodology, *Water Resources Research*, 23(2), 351-367.

Massmann, J., and R. A. Freeze 1987b Groundwater contamination from waste management sites: The interaction between risk-based engineering design and regulatory policy, 2, Results, *Water Resources Research*, 23(2), 368-380.

Matheron, G. 1965 *Les variables régionalisées et leur estimation. Une application de la théorie des fonctions aléatoires aux Sciences de la Nature*, Masson, Paris.

Matheron, G. 1971 *The theory of Regionalized Variables and Its Applications*, Ecole Des Mines, Fontainebleau, France.

Matheron, G. 1969. *Le krigeage universel*, Cahiers du Centre de Morphologie Mathématique de Fontainebleau, Fasc. 1, Ecole des Mines de Paris.

Matheron, G. 1973 The intrinsic random function and its application, *Advances in Applied Probability*, 5, 438-468.

Maxwell, Reed M., Carle, Steven F., and Tompson, F. B. 2000 Contamination, Risk, and Heterogeneity: On the Effectiveness of Aquifer Remediation, Lawrence Livermore National Laboratory Report, UCRL-JC-139664, Livermore, CA.

McAllister, P.M., and C. Y. Chiang 1994 A practical approach to evaluating natural attenuation of contaminants in ground water, *Ground Water Monitoring and Remediation*, 14(2), 161-173.

Meyer, P. D., and E. D. Brill 1988 A method for locating wells in a groundwater monitoring under conditions of uncertainty, *Water Resources Research*, 24(8), 1277-1282.

Meyer, P.D., Albert J. Valocchi, and J. Wayland Eheart 1994 Monitoring network design to provide initial detection of groundwater contamination, *Water Resources Research*, 30(9), 2647-2659.

Michael, W. J., D. K. Tcheng, B. S. Minsker, A. J. Valocchi, J. J. Quinn, and G. P. Williams 2002 Integrating Data Sources to Optimize Long-Term Monitoring, Operation, and Stewardship, The 3<sup>rd</sup> International Conference on Remediation of Chlorinated and Recalcitrant Compounds, Battelle, Monterey, CA, in press.

Montas, H. J., R. H. Mohtar, A. E. Hassan, F. A. AlKhal 2000, Heuristic space—time design of monitoring wells for contaminant plume characterization in stochastic flow fields, *Journal of Contaminant Hydrology*, 43, 271-301.

Moré, J. J. 1977 The Levenberg-Marquardt algorithm: Implementation and theory. In G. A. Watson, ed. *Numerical Analysis*, pp. 105-116, Springer-Verlag: New York.

National Research Council 1993 *In Situ Bioremediation: When Does It Work?*, National Academy Press, Washington, D.C.

National Research Council 1997 *Innovations in Ground Water and Soil Cleanup: From Concept to Commercialization*, National Academy Press, Washington, D.C.

National Research Council 1999 *Environmental Cleanup at Navy Facilities: Risk-Based Methods*, National Academy Press, Washington, D.C.

Nyer, Evan, Polly Mayfield, and Joseph Hughes 1998 Beyond the AFCEE Protocol for Natural Attenuation, *Ground Water Monitoring and Remediation*, 18(3), 70–77.

Olea, R. A. 1984 Sampling design optimization for spatial functions, *Mathematical Geology*, 16(4), 365-391.

Parks, G. T. and I. Miller 1998 Selective breeding in a multiobjective genetic algorithm. In A. E. Eiben, T. Bäck, M. Schoenauer, and H. Schwefel, eds. Fifth International Conference on Parallel Problem Solving from Nature, pp. 250-259, Springer: Berlin, Germany.

Rast, R. ed. 1997 *Environmental Restoration: Unit Cost Book*, R. S. Means: Kingston MA.

Reed, P., B. Minsker, and A. J. Valocchi 2000a Cost effective long-term groundwater monitoring design using a genetic algorithm and global mass interpolation. *Water Resources Research*, 36(12): 3731-3741.

Reed, P., B. Minsker, and D. E. Goldberg, 2000b Designing a competent simple genetic algorithm for search and optimization. *Water Resources Research*, 36(12): 3757-3761.

Reed, P., B. Minsker , and D. E. Goldberg 2001a A multiobjective approach to cost effective long-term groundwater monitoring using an Elitist Nondominated Sorted Genetic Algorithm with historical data, *Journal of Hydroinformatics*, 3(2), 71-90.

Reed, P., B. S. Minsker, and A. Valocchi, 2001b Why Optimize Long Term Groundwater Monitoring Design? A Multiobjective Case Study of Hill Air Force Base, "Bridging the Gap: Meeting the World's Water and Environmental Resources Challenges", Proceedings of the World Water and Environmental Resources Congress (ISBN 0-7844-0569-7),

Orlando, FL, ed. Don Phelps and Gerald Sehlke, American Society of Civil Engineers, Washington, DC.

Reed, P., T. R. Ellsworth, and B. Minsker, 2002 Spatial Interpolation Methods for Nonstationary Plume Data, *Ground Water*, In Review.

Ritzel, B. J., Eheart, J. E., & Ranjithan, S. 1994 Using genetic algorithms to solve a multiple objective groundwater pollution containment problem. *Water Resources Research*, 30(5), 1589-1603.

Rizzo, D. M., D. E. Dougherty, and M. Yu 2000 An adaptive long-term monitoring and operations system (aLTMOs) for optimization in environmental management, In ASCE 2000 Joint Conference on Water Resources Engineering and Water Resources Planning and Management, Minneapolis, MN.

Rouhani, S. 1985 Variance reduction analysis, *Water Resources Research*, 21(6), 837-846.

Rouhani, S., and T. J. Hall 1988 Geostatistical schemes for groundwater sampling, *Journal of Hydrology*, 103, 85-102.

Russo, D. 1984 Design of an optimal sampling network for estimating the variogram, *Soil Sci. Soc. Am. J.*, 45, 682-687.

Russo, D. and W. A. Jury 1987 A theoretical study of the estimation of the correlation scale in spatially variable fields 2: Nonstationary Fields, *Water Resources Research*, 23(7), 1269-1279.

Russo, D. and W. A. Jury 1988 Effect of the sampling network on estimates of the covariance function of stationary fields, *Soil Sci. Soc. Am. J.*, 52, 1228-1234.

Storck, Pascal, J. Wayland Eheart, and Albert J. Valocchi 1997 A method for the optimal location of monitoring wells for detection of groundwater contamination in three-dimensional aquifers, *Water Resources Research*, 33(9), 2081-2088.

Schaffer, J. D. 1984 *Some experiments in machine learning using vector evaluated genetic algorithms*, Doctoral dissertation, Vanderbilt University: Nashville, TN.

Schaffer, J. D., R. A. Caruana, L. J. Eshelman, and R. Das 1989 A study of control parameters affecting online performance of genetic algorithms for function optimization, In Schaffer, J. D. (Ed.), Proceedings of the Third International Conference on Genetic Algorithms, p. 51-60, Morgan Kaufmann, San Mateo, CA.

Smalley, J., B. Minsker, and D. E. Goldberg 2000 Risk-based in situ bioremediation design using a noisy genetic algorithm, *Water Resources Research*, 36(10), 3043-3052.

Srinivas, N. and K. Deb 1995 Multiobjective optimization using nondominated sorting in genetic algorithms, *Evolutionary Computation*, 2(3), 221-248.

Thierens, D. and D. E. Goldberg 1993 Mixing in genetic algorithms. In S. Forrest, ed. Proceedings of the Fifth International Conference on Genetic Algorithms, pp. 38-45, Morgan Kaufmann Publishers: San Mateo, CA.

Thierens, D. and D. E. Goldberg 1994 Convergence Models of Genetic Algorithm Selection Schemes, In Y. Davidor, Hans-Paul Schwefel, and Reinhard Manner, eds. Proceedings of the Third Conference on Parallel Problem Solving from Nature, pp. 119-129, Springer-Verlag: New York, NY.

Thierens, D. 1995 *Analysis and design of genetic algorithms*, Doctoral dissertation, Katholieke Universiteit Leuven, Leuven, Belgium.

Thierens, D., D. E. Goldberg, and A. G. Pereira 1998 Domino Convergence, Drift, and the Temporal-Salience Structure of Problems, In The 1998 IEEE International Conference on Evolutionary Computation Proceedings, pp. 535-540, IEEE Press: New York, NY.

Van Veldhuizen, D. A. 1999 Multiobjective Evolutionary Algorithms: Classifications, Analyses, and New Innovations, Doctoral dissertation, AFIT/DS/ENG/99-01, Air Force Institute of Technology: Wright-Patterson AFB, Ohio.

Van Veldhuizen, D. A. and G. B. Lamont 2000 Multiobjective evolutionary algorithms: analyzing the state-of-the-art. *Evolutionary Computation*, 8(2), 125-147.

Volpi, G. and G. Gambolati 1978 On the use of a main trend for the kriging technique in hydrology, *Advances in Water Resources*, 1, 345-349.

Wagner, Brian J. 1995 Sampling design methods for groundwater modeling under uncertainty, *Water Resources Research*, 31(10):2581-2591.

Warrick, A. and D. E. Myers 1987 Optimization of sampling locations for variogram calculations, *Water Resources Research*, 23(3): 496-500.

Yfantis, E. A., G. T. Flatman, and J. V. Behar 1987 Efficiency of kriging estimation for square, triangular, and hexagonal grids, *Mathematical Geology*, 19(3), 183-205.

Zitzler, E. and L. Thiele 1999 Multiobjective evolutionary algorithms: a comparative case study and the strength Pareto approach. *IEEE Transactions on Evolutionary Computation*, 3(4), 257-271.

Zitzler, E., K. Deb, and L. Thiele 2000 Comparison of multiobjective evolutionary algorithms: empirical results. *Evolutionary Computation*, 8(2), 173-195.

Zitzler, E., M. Laumanns, and L. Thiele 2001 SPEA2: Improving the Strength Pareto Evolutionary Algorithm, Computer Engineering and Networks Laboratory Report (TIK-103), Department of Electrical Engineering, Swiss Federal Institute of Technology, Zurich, Switzerland.

**REFERENCES.....155**

## APPENDIX A. GEOSTATISTICAL METHODS

### A.1 Chicken or the Egg?

The geostatistical modeling approach requires both structural analysis and specification of a positive definite covariance model (or a conditionally negative definite variogram model) to represent the theoretical spatial structure. Structural analysis consists of analyzing available concentration data to determine the correlational structure. The analysis uses historical contaminant data to discretely approximate the *theoretical variogram*  $\gamma(\bar{h})$  given in equation (A.1) where  $\bar{h}$  is a distance separation vector (or increment) for two locations  $\bar{u}_\nu$  and  $\bar{u}_{\nu'}$ .  $C(\bar{u}_\nu)$  is a regionalized variable representing concentration that can be decomposed into its mean,  $m(\bar{u}_\nu)$ , and zero mean random fluctuation,  $r(\bar{u}_\nu)$ , components.

$$\begin{aligned} E[C(\bar{u}_\nu) - C(\bar{u}_{\nu'})]^2 &= \text{var}[C(\bar{u}_\nu) - C(\bar{u}_{\nu'})] + [m(\bar{u}_\nu) - m(\bar{u}_{\nu'})]^2 \\ &= 2\gamma(\bar{h}) + [m(\bar{u}_\nu) - m(\bar{u}_{\nu'})]^2 \end{aligned} \quad (\text{A.1})$$

The theoretical variogram  $\gamma(\bar{h})$  represents the average dissimilarity between two concentrations separated by a distance  $\bar{h}$ . In the ordinary kriging approach, the mean component is assumed to be constant (or locally stationary) in which case the trend (or drift) term,  $[m(\bar{u}_\nu) - m(\bar{u}_{\nu'})]^2$ , in equation (A.1) is equal to zero and the theoretical variogram can be deduced directly from concentration samples. For nonstationary phenomena, the "chicken-and-egg" (Armstrong 1984) conundrum exists because the trend term is nonzero and a proper model for the theoretical variogram cannot be directly identified using concentration data solely without making an assumption on the functional form of the trend [i.e., specifying a mathematical function to model  $[m(\bar{u}_\nu) - m(\bar{u}_{\nu'})]^2$  in equation (A.1)]. The next sections of this appendix give a detailed

description of the five kriging systems important to this thesis. Note that the kriging systems are presented in terms of covariance  $K(\bar{h})$ , which can be derived from the theoretical variance  $K(0)$  and the variogram  $\gamma(\bar{h})$  using equation (A.2).

$$\gamma(\bar{h}) = K(0) - K(\bar{h}) \quad (\text{A.2})$$

#### A.1.1 Ordinary Kriging (OK) in Local Neighborhoods

The OK estimator presented in equation (A.3) is formulated to provide the Best Linear Unbiased Estimate (or BLUE) for contaminant concentration at an unsampled location  $\bar{u}_j$ .

$$c_{est}^{OK}(\bar{u}_j) = \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{OK}(\bar{u}_j) c(\bar{u}_{\psi}) \text{ where } \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{OK}(\bar{u}_j) = 1, \forall j \quad (\text{A.3})$$

The ordinary kriging weights are found using equation (A.4), which can be derived from equation (A.3) by minimizing the expected estimation errors using the Lagrange multiplier method (for a detailed derivation see *Kitanidis* 1997, *Goovaerts* 1997). The first equation of (A.4) represents the unbiasedness constraint for the OK system and  $\mu_{OK}(\bar{u}_j)$  is the Lagrange multiplier corresponding to this constraint.

$$\begin{aligned} & \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{OK}(\bar{u}_j) = 1, \forall j \\ & \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{OK}(\bar{u}_j) K(\bar{h}_{\psi}) + \mu_{OK}(\bar{u}_j) = K(\bar{u}_{\psi} - \bar{u}_j), \forall j \end{aligned} \quad (\text{A.4})$$

Solving the linear system represented by equation (A.4) yields the kriging weights used to compute each OK estimate and its corresponding estimation variance using equations (A.4) and (A.5), respectively.

$$\sigma_{OK}^2(\bar{u}_j) = K(0) - \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{OK}(\bar{u}_j) K(\bar{h}_{\psi}) - \mu_{OK}(\bar{u}_j), \forall j \quad (\text{A.5})$$

Note as stated previously, the estimation variance is only a function of the theoretical model used to represent the variogram and the geometrical configuration of the sampling data (i.e. a function of the kriging weights only). The OK approach invokes the *intrinsic hypothesis*, which assumes that mean or trend terms given in equation (A.1) are locally constant but unknown within neighborhoods surrounding the current unsampled location (i.e.,  $[m(\bar{u}_\nu) - m(\bar{u}_{\nu'})]^2 \approx 0$  in local neighborhoods). This assumption is equivalent to modeling the local trend using a zero<sup>th</sup> order polynomial.

#### **A.1.2 Multigaussian Kriging (MGK) in Local Neighborhoods**

The MGK approach to estimation utilizes the normal score transform of the sample data and models the contaminant concentrations as a multivariate Gaussian random function,  $T(\bar{u})$  (for details on the transform see *Goovaerts* 1997). The normal score transform maps the asymmetrical discrete sample distribution to a zero-one normal distribution. Estimation is performed using equations (A.3) thru (A.5) on the transformed data while again assuming a zero<sup>th</sup> order polynomial trend.

#### **A.1.3 Quantile Kriging (QK) in Local Neighborhoods**

The distributional dependence of OK and MGK, as well as their sensitivity to highly skewed sample data, led to the development of non-parametric or distribution-free kriging approaches using indicator (*Journel* 1983) or rank-order (*Journel & Deutsch* 1997) transformations of the data. Indicator kriging uses a binary transformation of sample data where the data value is set to either 1 or zero depending on whether it is greater than or equal to a user specified threshold value. The indicator approach requires  $n_q$  threshold indicator transforms each of which requires its own variogram and kriging system. Correctly quantifying  $n_q$  variograms is generally not possible in plume interpolation applications due to data limitations

(see *Chilès & Delfiner* 1999). Moreover, solving equations (A.4) and (A.5) for each of the  $n_q$  cutoffs for every point in the interpolation represents a severe computational limitation of the method.

These limitations motivated the development of quantile kriging (*Journel & Deutsch* 1997), which transforms concentrations into standardized ranks (or quantiles) using equation (A.6) and computes estimates in quantile space using OK (see *Journel & Deutsch* 1997 and *Juang et al.* 2001).

$$c(\bar{u}) = \frac{\text{rank of sample } i}{N + 1} \quad (\text{A.6})$$

The transform consists of ranking the concentration data in ascending order and dividing these ranks by one plus the total number of sample data  $N$ . Figure (5.2) graphically illustrates quantile or standardized ranks transform. The  $i^{\text{th}}$  sample is assigned the probability that  $c(\bar{x})$  is less than or equal to its concentration value [ $F(c)$  in Figure (5.2)]. Estimates and estimation variances are computed in quantile space using equations (A.4) and (A.5). The empirical cdf of the sample data is also used to back transform estimates from quantile space to concentration space (for details see *Juang et al.* 2001). Since quantiles are known to have a uniform distribution, this fact can be used in conjunction with the estimation variances to compute non-parametric local uncertainty estimates at unsampled locations in the interpolation domain (see Chapter 5).

#### **A.1.4 Kriging with a Trend (KT)**

The KT approach (also termed “universal kriging”) models concentrations as a regionalized variable,  $c(\bar{u})$ , that is composed of the sum of a smoothly varying mean component  $m(\bar{u})$  termed a trend and a second order stationary residual fluctuation  $r(\bar{u})$ . The KT estimator for contaminant concentration at an unsampled location  $\bar{u}_j$  is shown in equation (A.7) below.

$$c_{est}^{KT}(\bar{u}_j) = \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{KT}(\bar{u}_j) c(\bar{u}_{\psi}) \text{ where } \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{KT} p_s(\bar{u}_{\psi}) = p_s(\bar{u}_j) \quad \forall j \text{ and } s = 0, \dots, S \quad (\text{A.7})$$

The KT estimator requires the specification of a linear combination of known functions of spatial coordinates,  $p_s(\bar{u})$ , to model the trend component of the concentration. Typically the functions represent the sum of  $s$ -order monomials composing the overall  $S^{\text{th}}$  order polynomial used to model the trend. The kriging weights  $\lambda_{\psi}^{KT}$  are computed using the linear system shown in equation (A.8), which is derived by minimizing the expected estimation error for equation (A.7).

$$\begin{aligned} & \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{KT}(\bar{u}_j) = 1, \quad \forall j \\ & \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{KT}(\bar{u}_j) p_s(\bar{u}_{\psi}) = p_s(\bar{u}_j), \quad \forall j \text{ and } s = 0, \dots, S \\ & \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{KT}(\bar{u}_j) K_r(\bar{h}) + \sum_{s=0}^S \mu_s^{KT}(\bar{u}_j) p_s(\bar{u}_{\psi}) = K_r(\bar{u}_{\psi} - \bar{u}_j), \quad \forall j \end{aligned} \quad (\text{A.8})$$

Note that this system has  $S+1$  constraints to enforce the unbiasedness condition and the stipulation that the monomials of the second equation of (A.8) are filtered from the system. After the kriging weights are computed using (A.8), equation (A.9) is used to compute the estimation variance at each unsampled location  $\bar{u}_j$ .

$$\sigma_{KT}^2(\bar{u}_j) = K_r(0) - \sum_{\psi=1}^{nsamp} \lambda_{\psi}^{KT}(\bar{u}_j) K_r(\bar{h}) - \sum_{s=0}^S \mu_s^{KT}(\bar{u}_j) p_s(\bar{u}_j), \quad \forall j \quad (\text{A.9})$$

The KT approach requires the practitioner to explicitly specify a functional form for the trend term  $[m(\bar{u}_{\psi}) - m(\bar{u}_{\psi'})]^2$  in equation (A2.1) and subsequent structural analysis of the residual to attain its covariance function  $K_r(\bar{h})$ . Inference of the proper variogram is difficult

because the concentration data are not directly reflective of the residual function  $r(\bar{u})$  thereby causing the “chicken-and-egg” (Armstrong 1984) conundrum. The proper model for the theoretical variogram cannot be directly identified using concentration data solely without knowing the functional form of the trend, which itself cannot be deduced without knowing the proper model for the theoretical variogram. Note the OK system in equation (A.3) can be directly derived from the KT system by modeling the trend as a zero<sup>th</sup> order polynomial ( $s = 0$ ) set equal to 1 in equation (A.7). Goovaerts (1997) shows that the only difference between the OK and KT systems results from the practitioner’s arbitrary decision to explicitly model the local trend as either a constant or a polynomial of order  $S$ .

#### A.1.5 *Intrinsic Kriging (ItK)*

The KT and variations of OK discussed above all require the practitioner to accept the dichotomy of the regionalized variable representing contaminant concentrations into a smoothly varying mean component and a stochastic fluctuation. Several studies have shown that this assumption is often an arbitrary choice that can significantly bias both estimates and their respective estimation variances (Volpi & Gambolati 1978, Hughes & Lettenmaier 1981, Russo & Jury 1987, Journel & Rossi 1989, Crawford & Hergert 1997). Intrinsic random function of order k (IRF-k) theory avoids the above dichotomy by not requiring the explicit specification of a model for the mean trend. Instead, the ItK approach defines allowable linear combinations of the sample data, termed *generalized increments*, that are implicitly capable of filtering trends from the data, facilitating a more direct measure of the underlying spatial structure of contaminant concentrations. The approaches discussed in previous sections model spatial structure using pairs of sample data to deduce the variogram shown in equation (A.1). Alternatively, IRF-k theory defines *generalized increments of order k* to be ( $k+1$ ) point

increments that are capable of filtering polynomial trends of degree less than or equal to  $k$  (*Chilès & Delfiner* 1999). Generalized increments, as the name implies, generalize the inference of spatial structure from a 2-point to a  $(k+1)$  point measure of spatial correlation termed the *generalized covariance*  $G(\bar{h})$ .

In the ItK approach, concentration is modeled as an IRF-k, which for the proper generalized increment (or allowable linear combination of sample data)  $C(\bar{u})$  is a stationary random function with a zero mean and a covariance equal to  $G(\bar{h})$ . It has been shown that the ItK system is identical to the UK system presented in equations (A.7) thru (A.9) except for the substitution of the generalized covariance function  $G(\bar{h})$  for the 2-point residual covariance function  $K_r(\bar{h})$  (for details see *Chilès & Delfiner* 1999). The ItK method was formulated with the goal of avoid the estimation biases caused by the explicit specification of a functional trend model.



## APPENDIX B. STRUCTURAL ANALYSIS

### B.1 Large Test Case Results

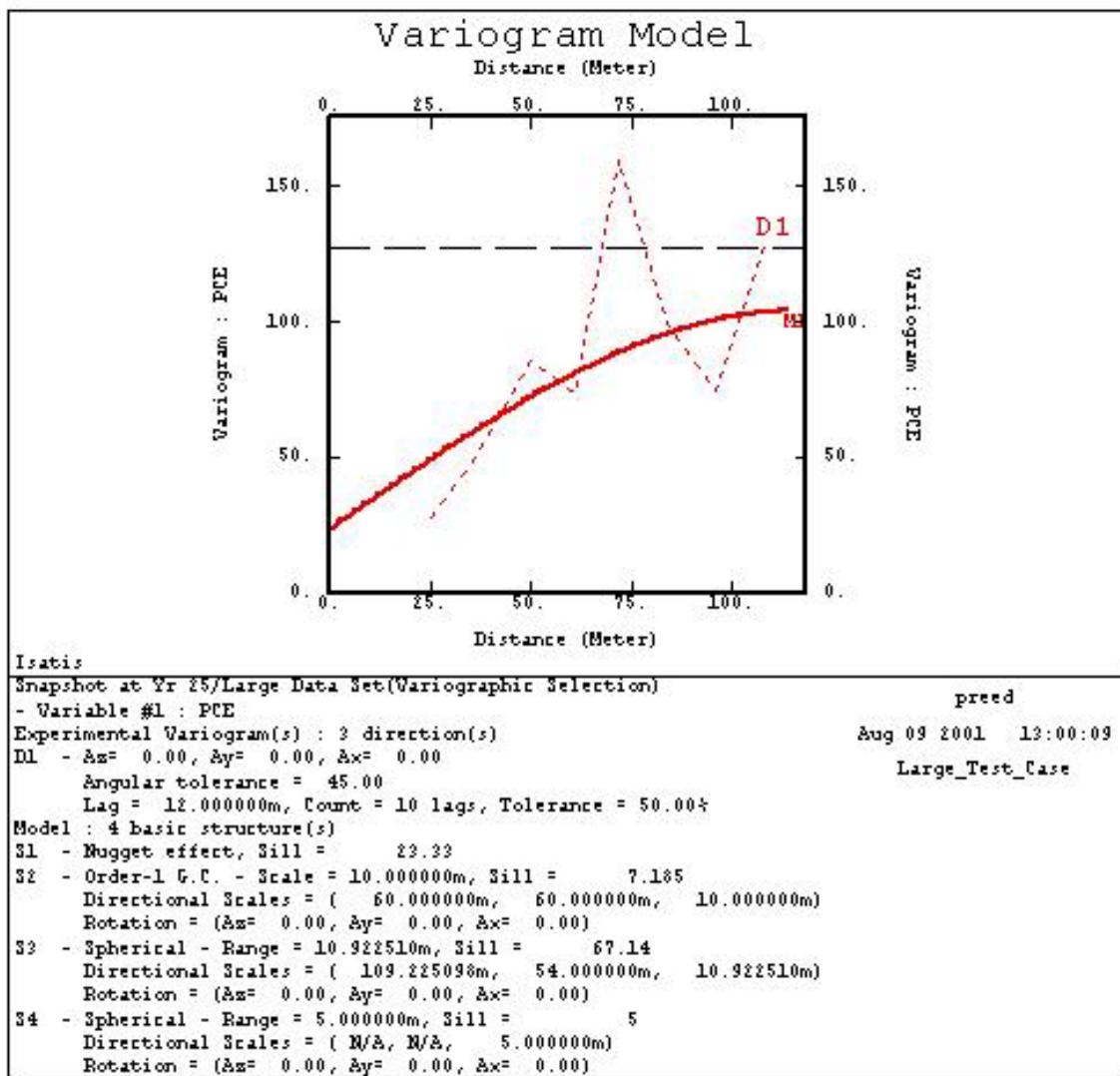


Figure B.1 Large case variogram model for the ordinary kriging system in the X-direction

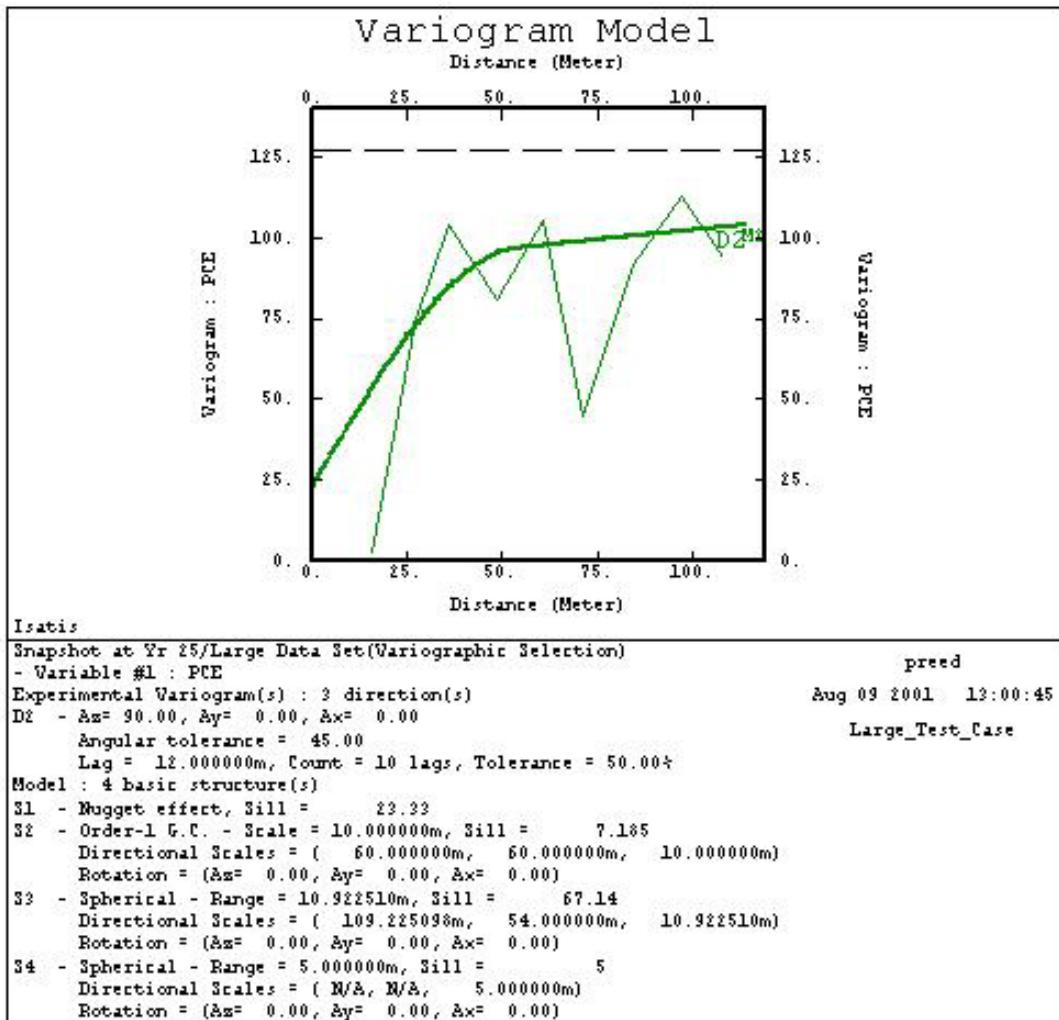


Figure B.2 Large case variogram model for the ordinary kriging system in the Y-direction

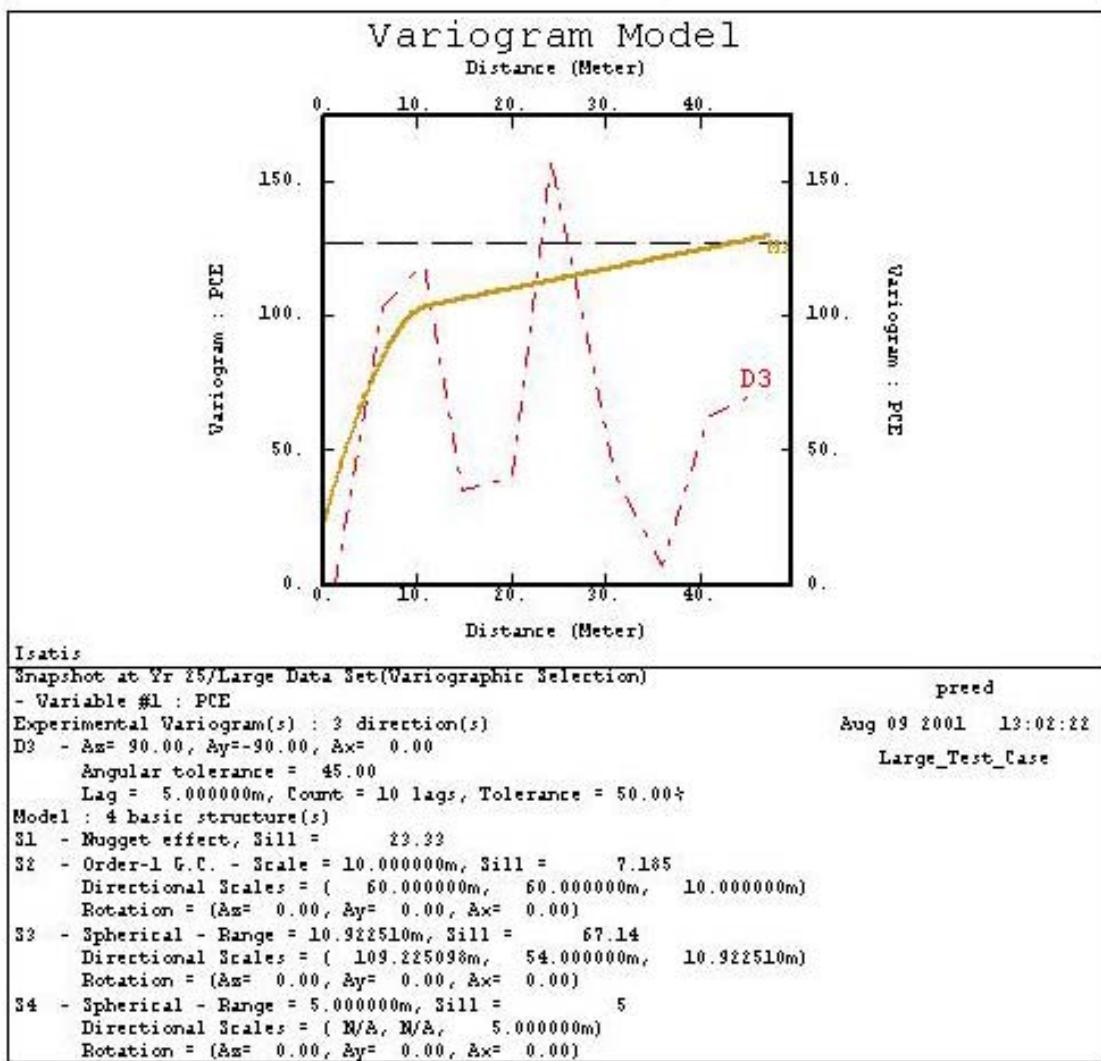


Figure B.3 Large case variogram model for the ordinary kriging system in the Z-direction

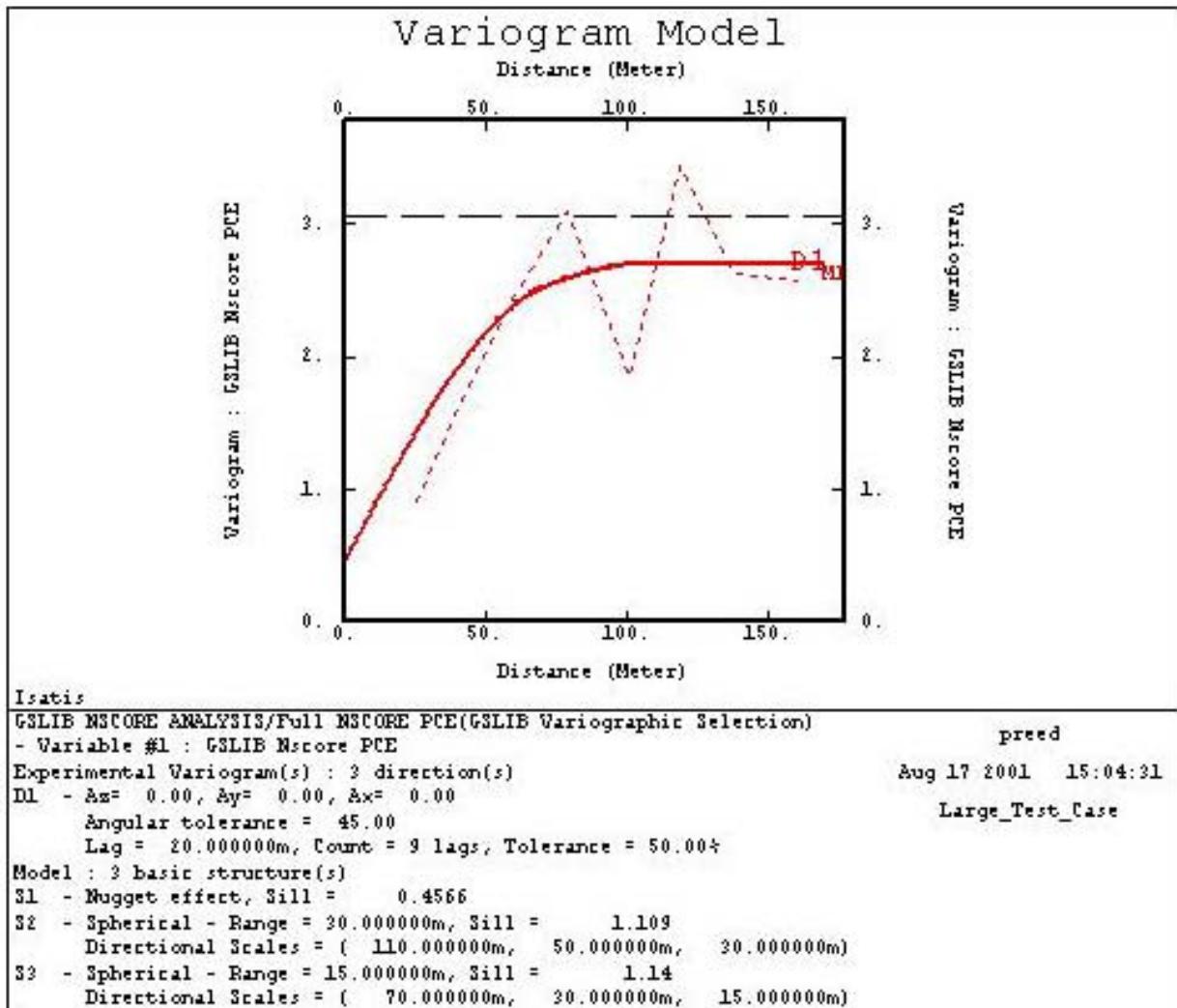


Figure B.4 Large case variogram model for the Multigaussian kriging system in the X-direction

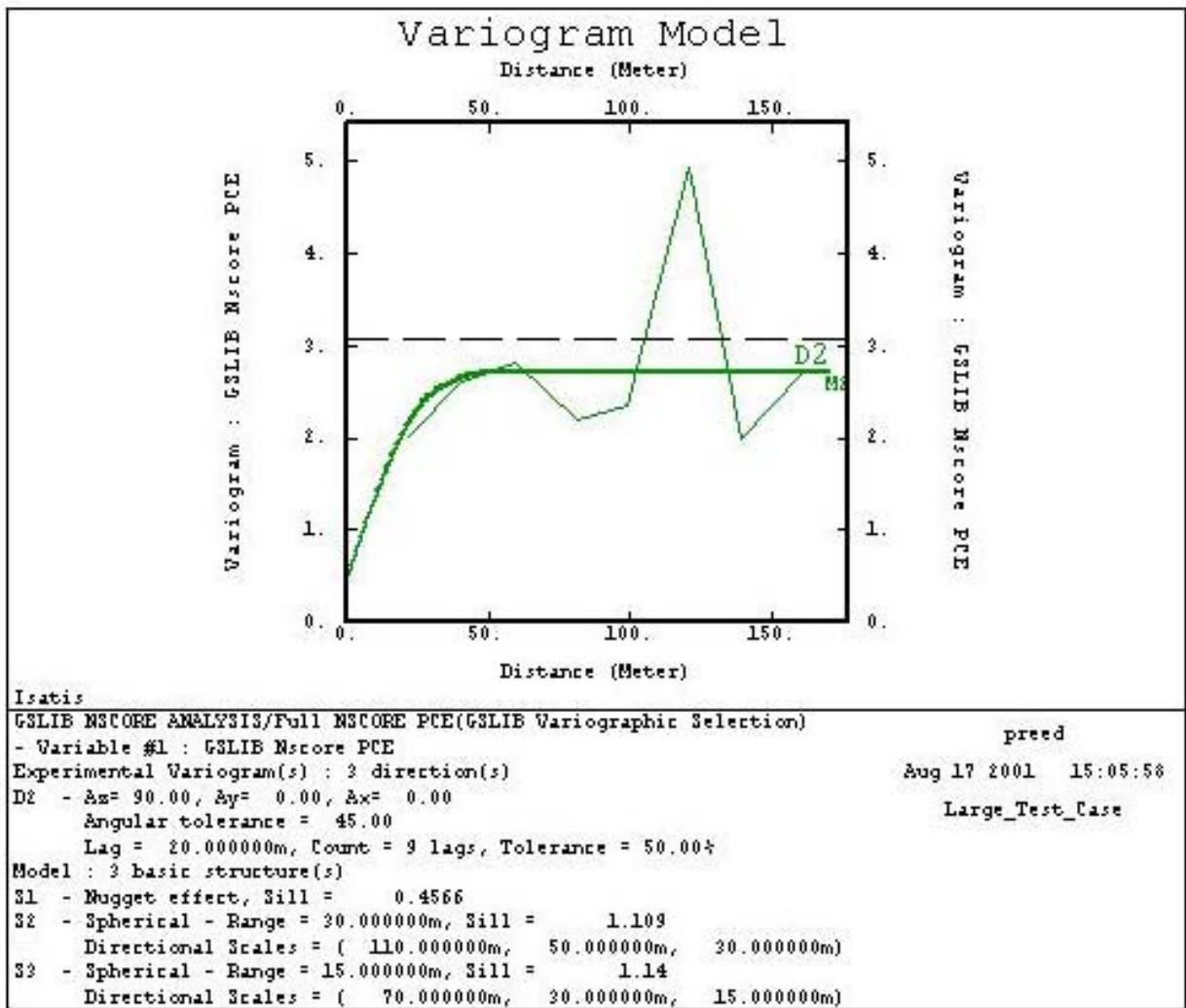


Figure B.5 Large case variogram model for the Multigaussian kriging system in the Y-direction

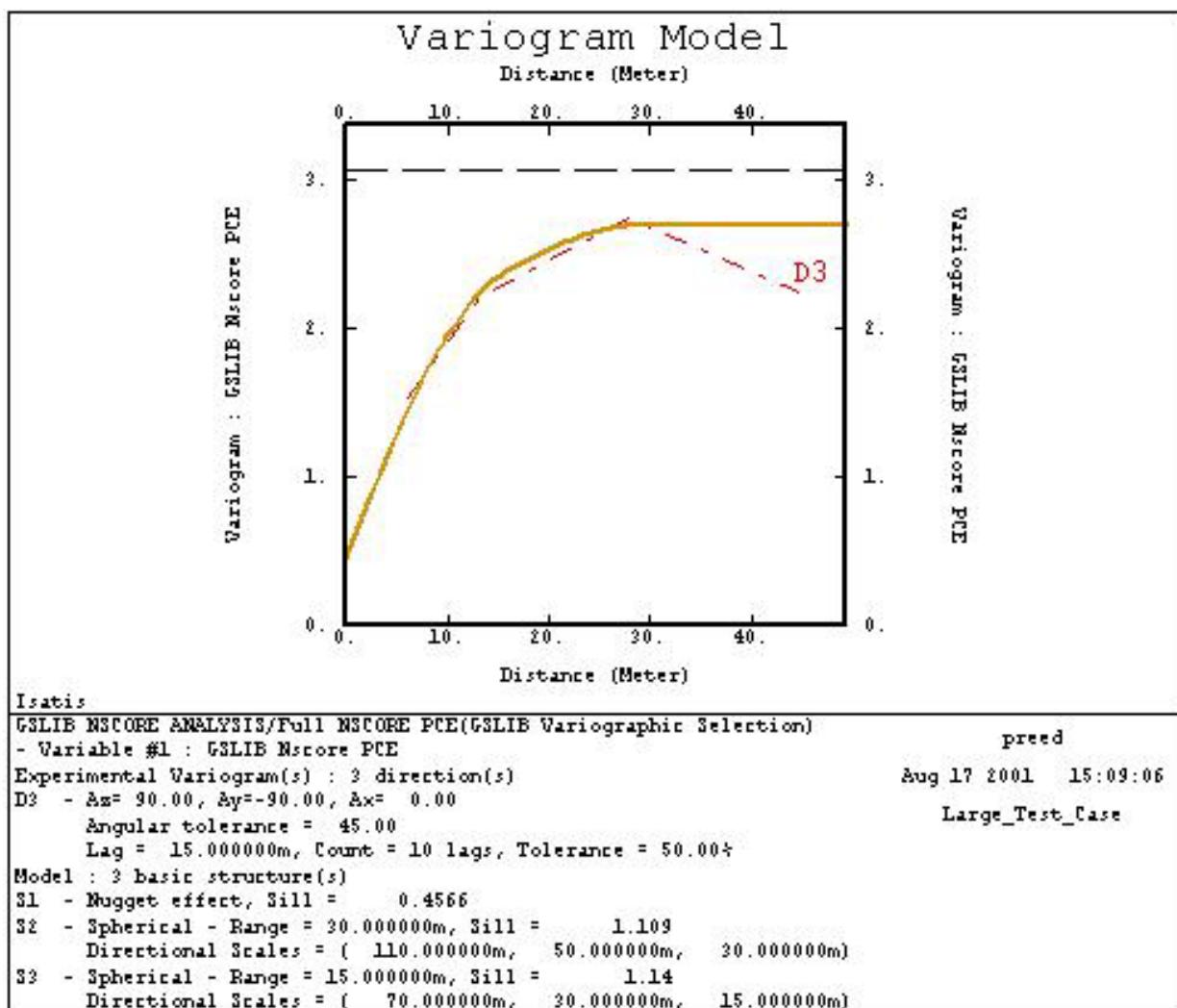


Figure B.6 Large case variogram model for the Multigaussian kriging system in the Z-direction

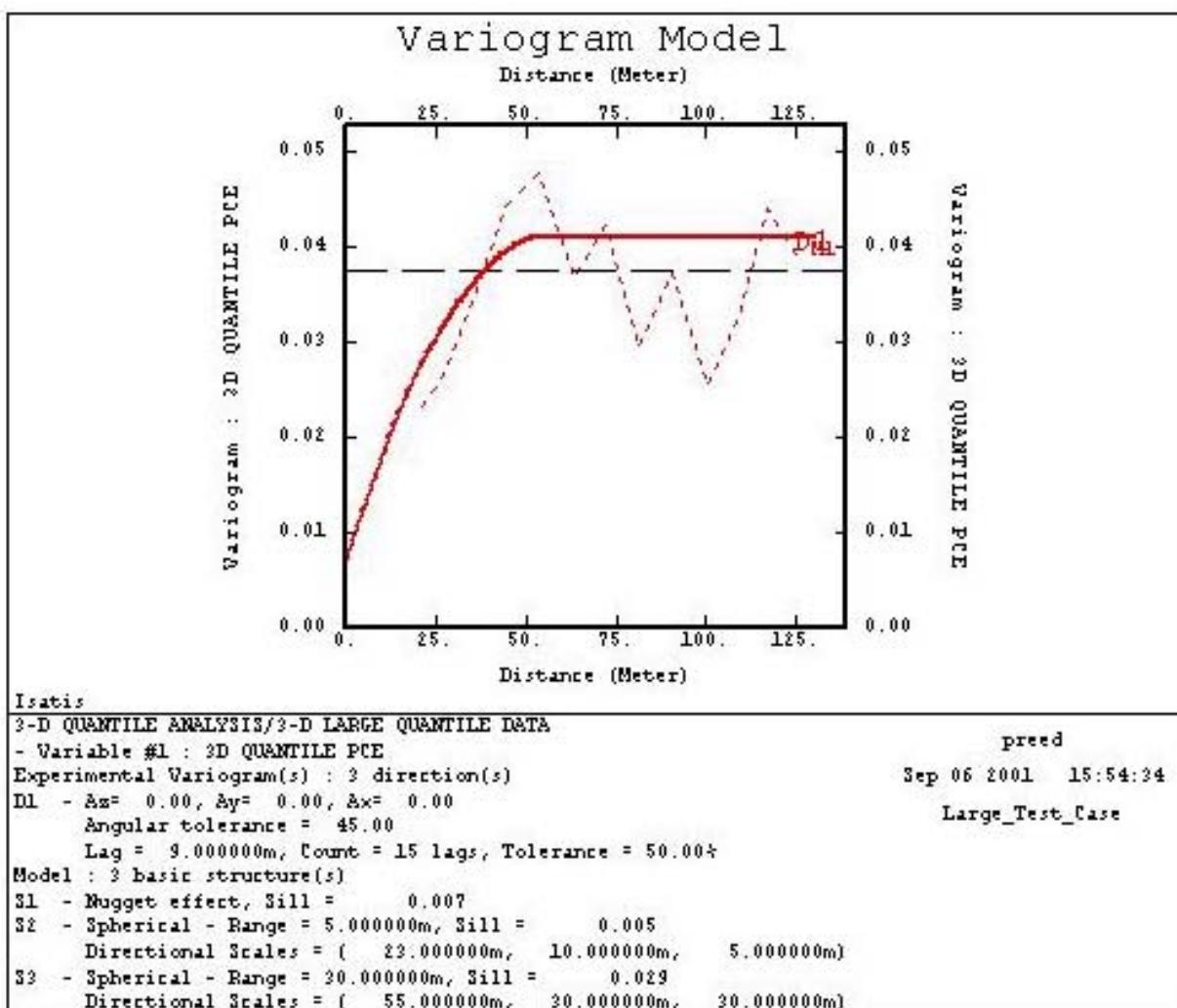


Figure B.7 Large case variogram model for the quantile kriging system in the X-direction

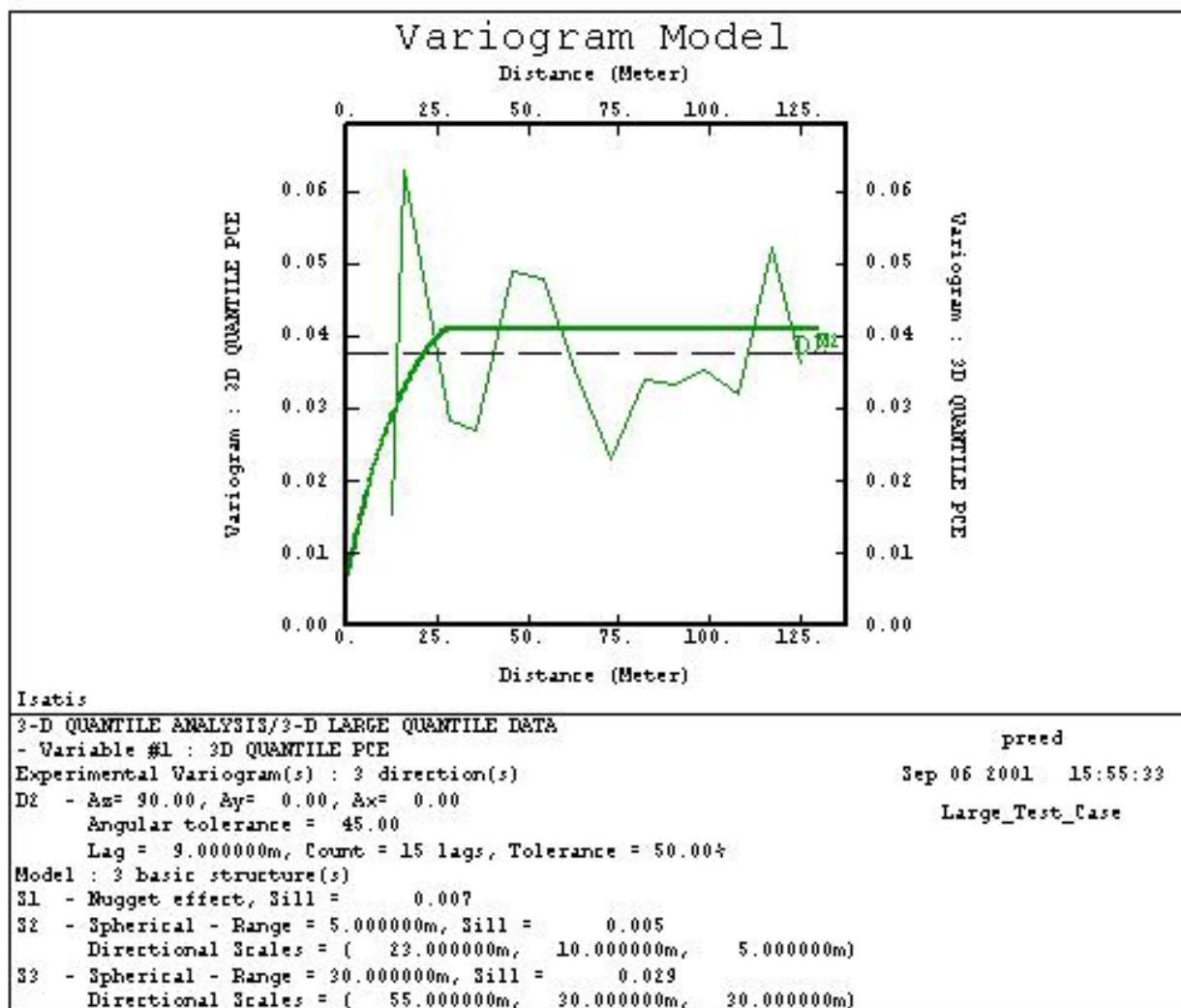


Figure B.8 Large case variogram model for the quantile kriging system in the Y-direction

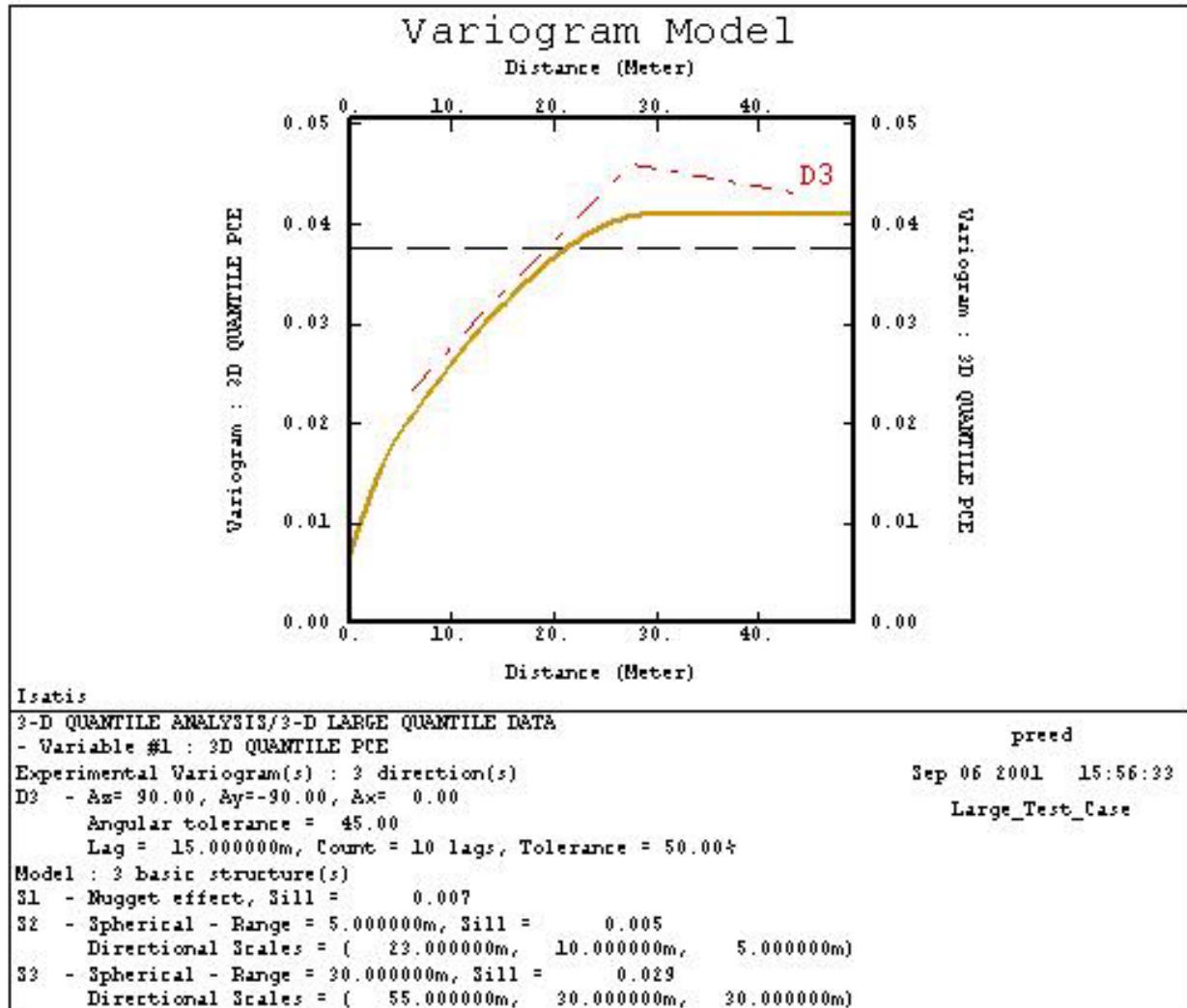


Figure B.9 Large case variogram model for the quantile kriging system in the Z-direction

preed - Aug 13 2001 13:55:19

---

---

| Structure Identification |

---

---

Data File Information:

Directory = Snapshot at Yr 25  
File = Large Data Set  
Selection = Variographic Selection

Seed File Information:

Directory = Snapshot at Yr 25  
File = Large Data Set  
Selection =  $0 < \text{PCE} < 40$   
Type = POINT (124 points)

Model Name = IRF Model-Moving Neigh

Neighborhood Name = Moving Omnidirectional Neighborhood - MOVING

Model : Covariance part

---

Number of variables = 1  
Number of basic structures = 4

**S1 : Nugget effect\***

**S2 : Order-1 G.C. - Scale = 114.136788m\***

S3 : Spline G.C. - Scale = 114.136788m - C-H2 = 1.0

S4 : Order-3 G.C. - Scale = 114.136788m - C-H2 = 1.

Covariance Identification Step

---

Rank	Score	Structure 1	Structure 2	Structure 3	Structure 4
1	1.0131	73.136	93.888	0	0
2	1.0131	71.207	96.422	1.0307	0
3	1.0131	72.522	94.645	0	0.21992
4	1.0253	141.01	0	0	0
5	0.95154	0	204.2	14.046	0
6	0.94059	0	206.06	0	4.8068
7	0.93058	0	209.67	0	0
8	3.7128	0	0	0	183.84
9	23.826	0	0	135.43	0

Successfully processed = 65

Written to the disk = 0

Warning: the following structures have not been saved

as their sills were almost zero for all variables:

- Discarded structure : Spline G.C. - Scale = 114.136788m - C-H2 = 1.0

- Discarded structure : Order-3 G.C. - Scale = 114.136788m - C-H2 = 1.

**\*Selected structures are in bold above**

The Model Parameter File (IRF Model-Moving Neigh) has been updated

Figure B.10 Large case Isatis output specifying the generalized covariance model for the intrinsic kriging system

preed - Aug 13 2001 13:55:19

---

---

Structure Identification	
--------------------------	--

---

---

Data File Information:

Directory = Snapshot at Yr 25  
File = Large Data Set  
Selection = Variographic Selection

Seed File Information:

Directory = Snapshot at Yr 25  
File = Large Data Set  
Selection =  $0 < \text{PCE} < 40$   
Type = POINT (124 points)

Model Name = IRF Model-Moving Neigh

Neighborhood Name = Moving Omnidirectional Neighborhood - MOVING

Drift Identification Step

---

Results are based on 62 measures

Trial	Mean Rank	Mean of Error	Variance of Error	Drift Trial
9	<b>6.6</b>	<b>-2.0297</b>	<b>156.13</b>	<b>1 z z2 (selected trend function)</b>
14	6.7	-3.1745	127.26	1 x y z
13	6.7	-2.7627	127.63	1 y z
12	6.9	-2.6174	131.85	1 x z
8	7	-2.3635	133.98	1 z
1	7.1	-2.4211	142.96	No Drift
5	7.1	-2.7578	138.96	1 y
2	7.2	-2.6817	143.33	1 x
6	7.3	-1.4244	163.45	1 y y2
11	7.4	-3.2242	139.74	1 x y
10	7.6	-1.9905	175.83	1 z z2 z3
3	7.7	-2.4304	164.06	1 x x2
7	8.2	-0.75455	183.43	1 y y2 y3
4	8.3	-2.1522	181.63	1 x x2 x3
15	8.6	-1.8936	221.77	1 x y x2 xy y2
16	9.5	-1.1353	313.36	1 x y x2 xy y2 x3 x2y xy2 y3

The Model Parameter File (IRF Model-Moving Neigh) has been updated

Figure B.11 Large case Isatis output specifying the trend function for the intrinsic kriging system

## B.2 Medium Test Case Results

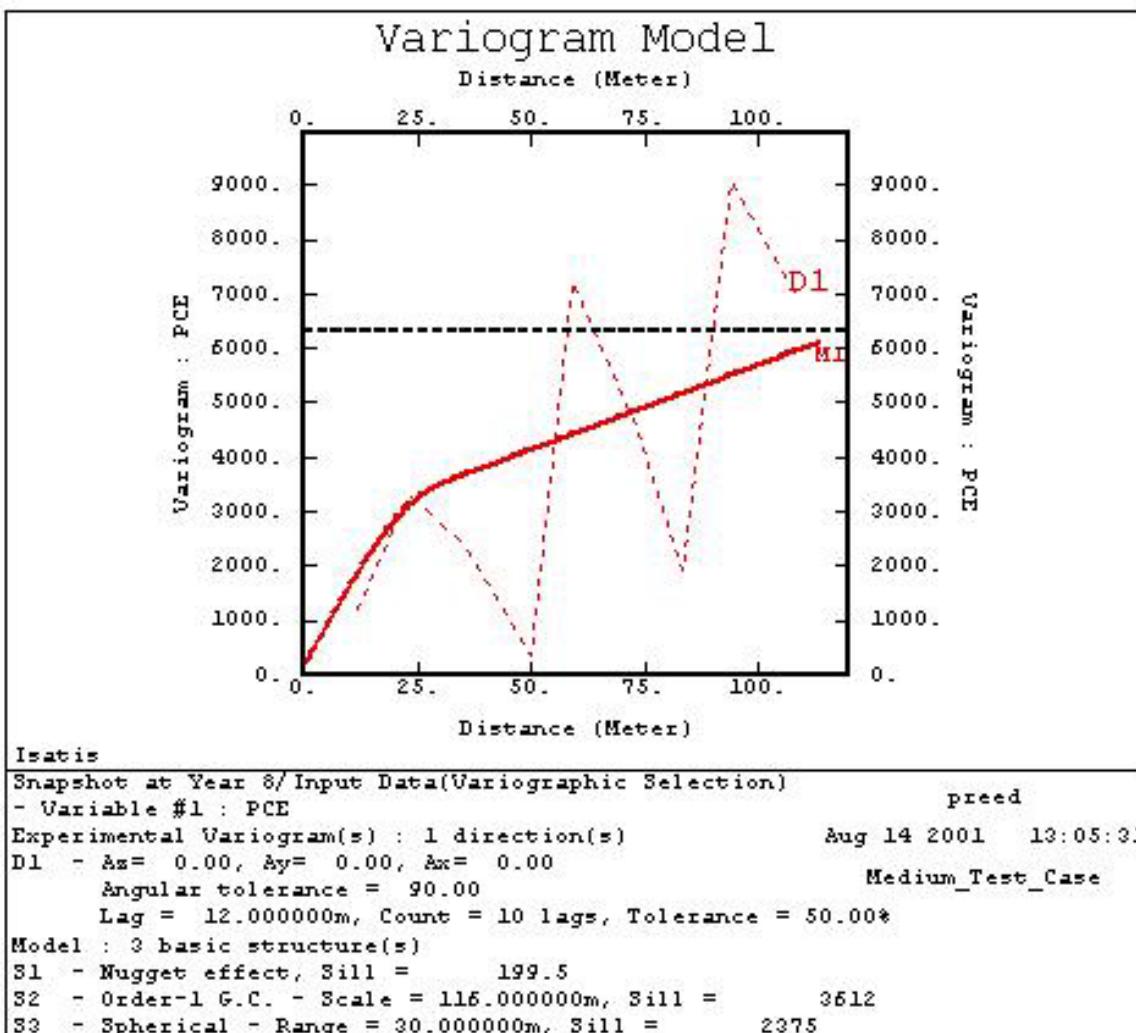


Figure B.12 Medium case isotropic variogram model for the ordinary kriging system

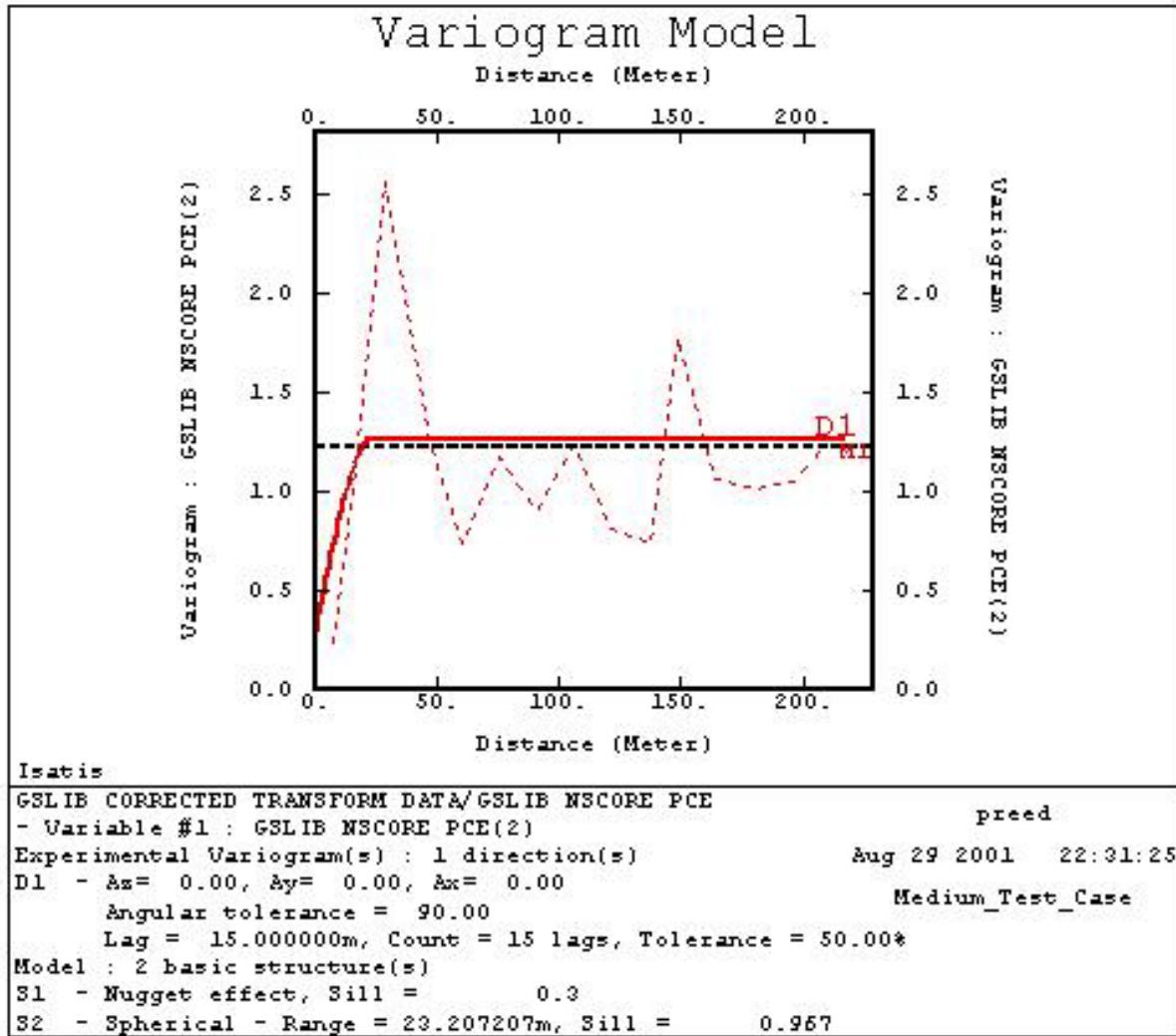


Figure B.13 Medium case isotropic variogram model for the Multigaussian kriging system

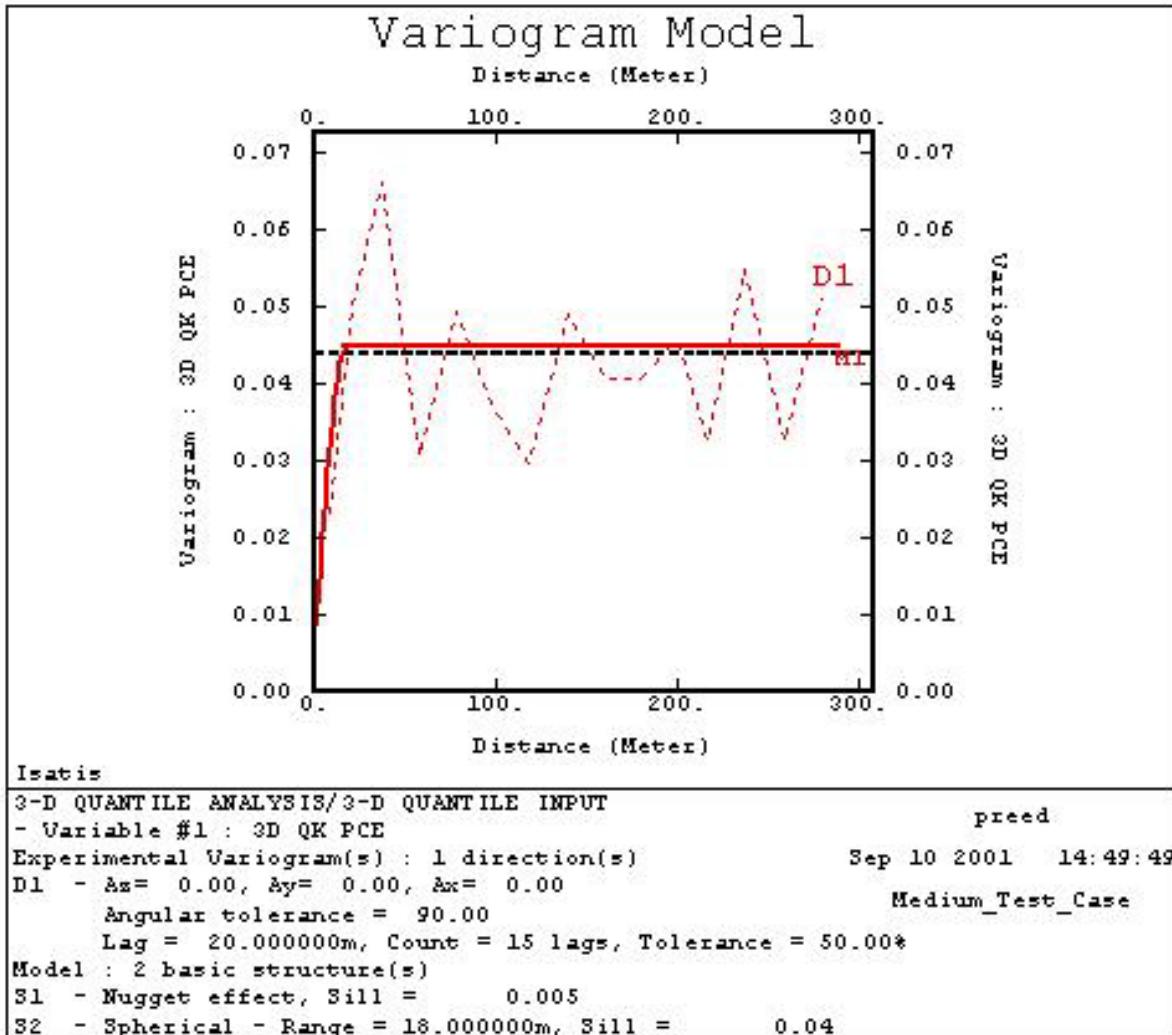


Figure B.14 Medium case isotropic variogram model for the quantile kriging system

preed - Aug 23 2001 08:29:52

=====| Structure Identification |=====

**Data File Information:**

Directory = Snapshot at Year 8  
File = Input Data  
Selection = Variographic Selection

**Seed File Information:**

Directory = Snapshot at Year 8  
File = Input Data  
Selection = Variographic Selection  
Type = POINT (58 points)

Model Name = IRF MODEL FOR VAR SEL (MOVING)

Neighborhood Name = 3-D MOVING NEIGHBORHOOD - MOVING

**Model : Covariance part**

Number of variables = 1

Number of basic structures = 4

S1 : Nugget effect\*

**S2 : Order-1 G.C. - Scale = 76.954438m**

**S3 : Spline G.C. - Scale = 76.954438m - C-H2 = 1.0**

**S4 : Order-3 G.C. - Scale = 76.954438m - C-H2 = 1.**

**Covariance Identification Step**

Rank	Score	Structure 1	Structure 2	Structure 3	Structure 4
1	<b>0.86579</b>	16633	0	0	0
2	<b>0.61067</b>	0	20362	0	0

Successfully processed = 52

Written to the disk = 0

**Warning: the following structures have not been saved**

as their sills were almost zero for all variables:

- Discarded structure : Order-1 G.C. - Scale = 76.954438m
- Discarded structure : Spline G.C. - Scale = 76.954438m - C-H2 = 1.0
- Discarded structure : Order-3 G.C. - Scale = 76.954438m - C-H2 = 1.

\*Selected structures are in bold above

The Model Parameter File (IRF MODEL FOR VAR SEL (MOVING)) has been updated

Figure B.15 Medium case Isatis output specifying the generalized covariance model for the intrinsic kriging system

preed - Aug 23 2001 08:29:52

Structure Identification	

Data File Information:

Directory = Snapshot at Year 8  
File = Input Data  
Selection = Variographic Selection

Seed File Information:

Directory = Snapshot at Year 8  
File = Input Data  
Selection = Variographic Selection  
Type = POINT (58 points)

Model Name = IRF MODEL FOR VAR SEL (MOVING)  
Neighborhood Name = 3-D MOVING NEIGHBORHOOD - MOVING

Drift Identification Step

Results are based on 32 measures

Trial	Mean Rank	Mean of Error	Variance of Error	Drift Trial
13	<b>5.7</b>	<b>-6.6525</b>	<b>20131</b>	<b>1 y z (selected trend function)</b>
5	6.1	0.78829	20090	1 y
7	6.7	1.438	26734	1 y y2 y3
14	6.8	-9.4693	23459	1 x y z
8	6.8	-4.4978	20824	1 z
11	7.1	1.2472	22715	1 x y
1	7.1	4.6045	19764	No Drift
16	7.3	-24.6	1.0549e+005	1 x y x2 xy y2 x3 x2y xy2 y3
12	7.4	-7.6167	23608	1 x z
2	7.8	2.6357	22335	1 x
3	8.1	-3.4158	20599	1 x x2
4	8.3	-8.7268	22629	1 x x2 x3
10	8.3	-0.30538	22255	1 z z2 z3
9	8.4	6.5105	20113	1 z z2
6	8.6	-1.712	26751	1 y y2
15	9.6	17.862	56174	1 x y x2 xy y2

The Model Parameter File (IRF MODEL FOR VAR SEL (MOVING)) has been updated

Figure B.16 Medium case Isatis output specifying the trend function for the intrinsic kriging system

### B.3 Small Test Case Results

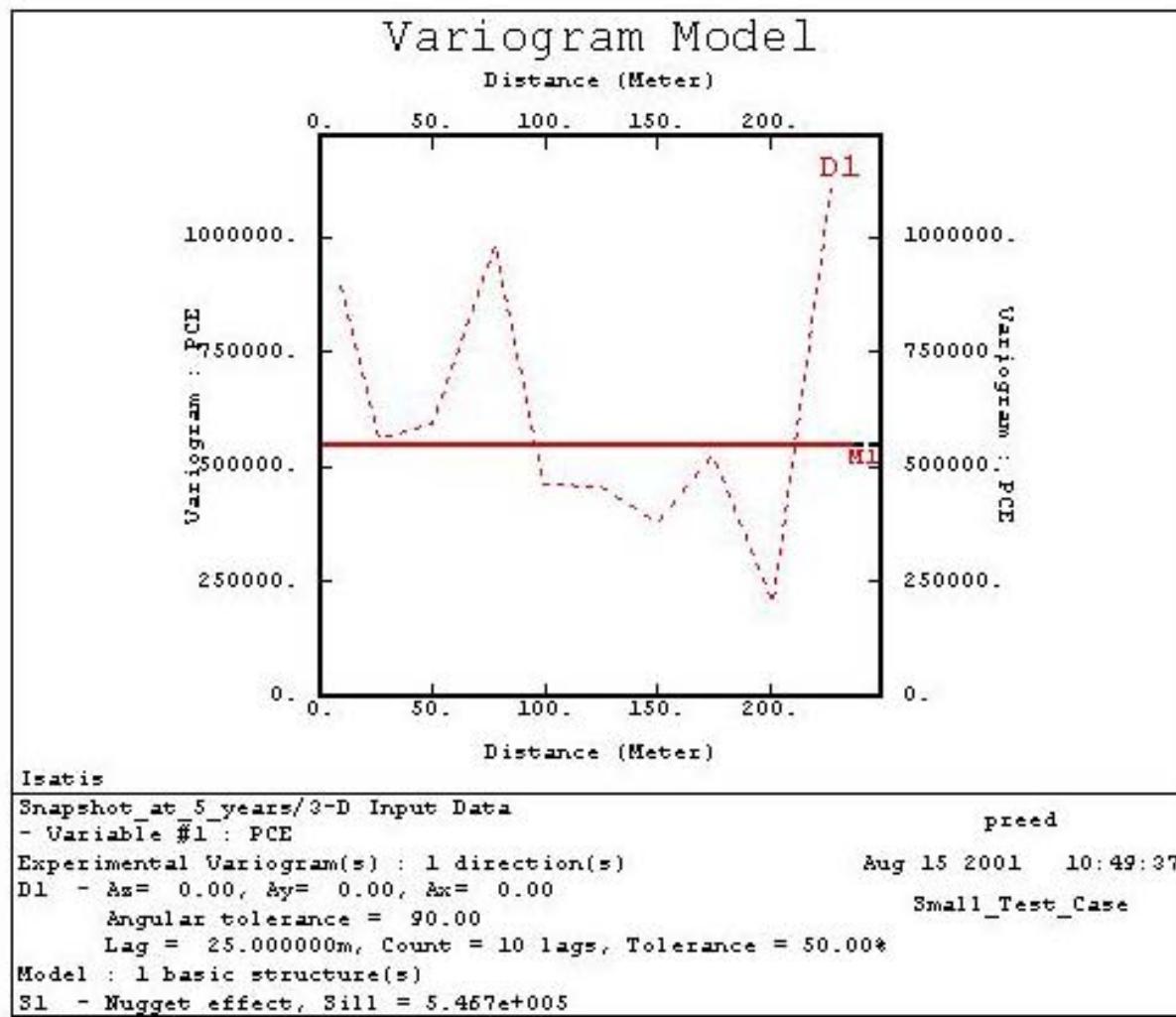


Figure B.17 Small case pure nugget variogram model for the ordinary kriging system

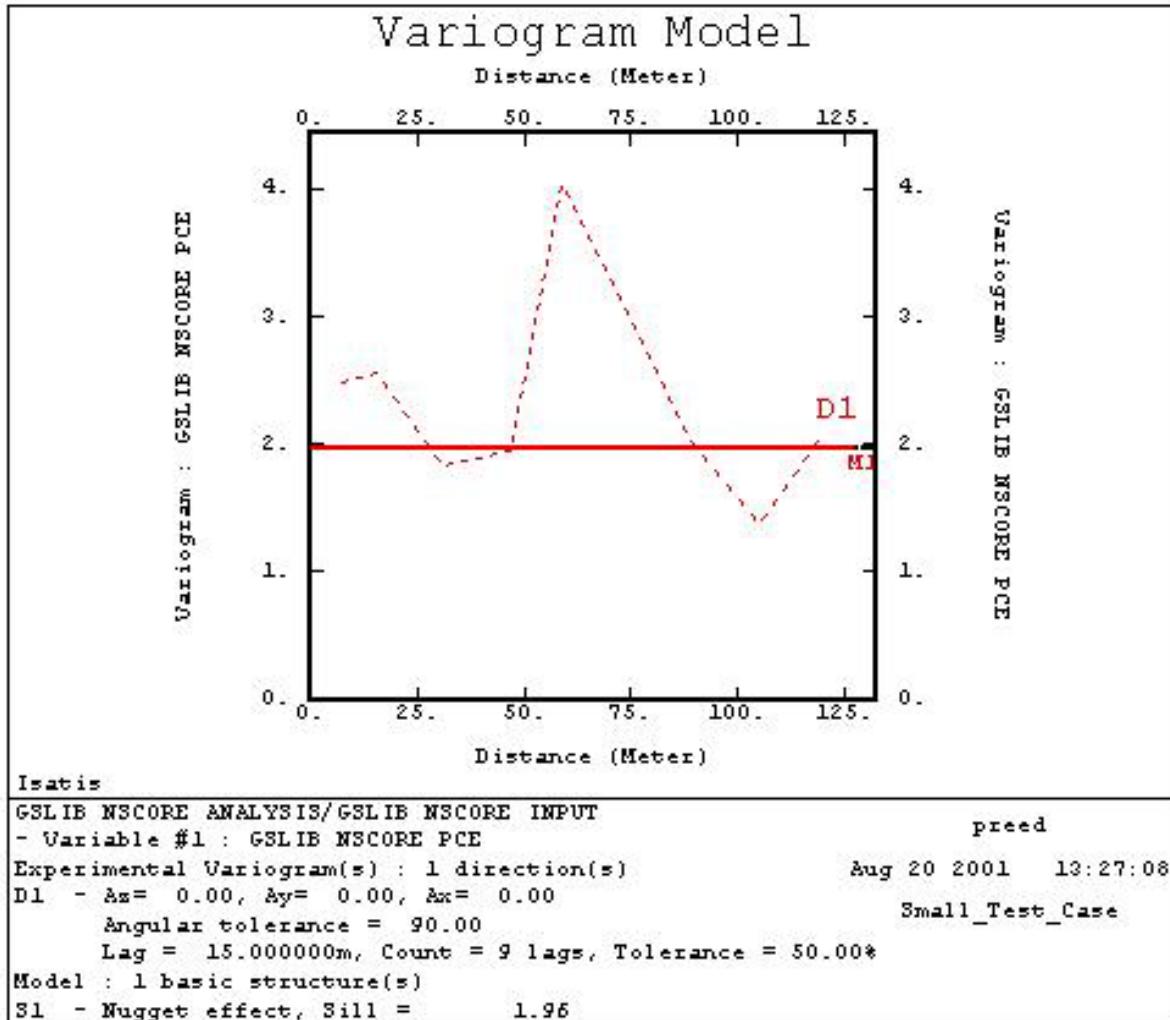


Figure B.18 Small case pure nugget variogram model for the Multigaussian kriging system

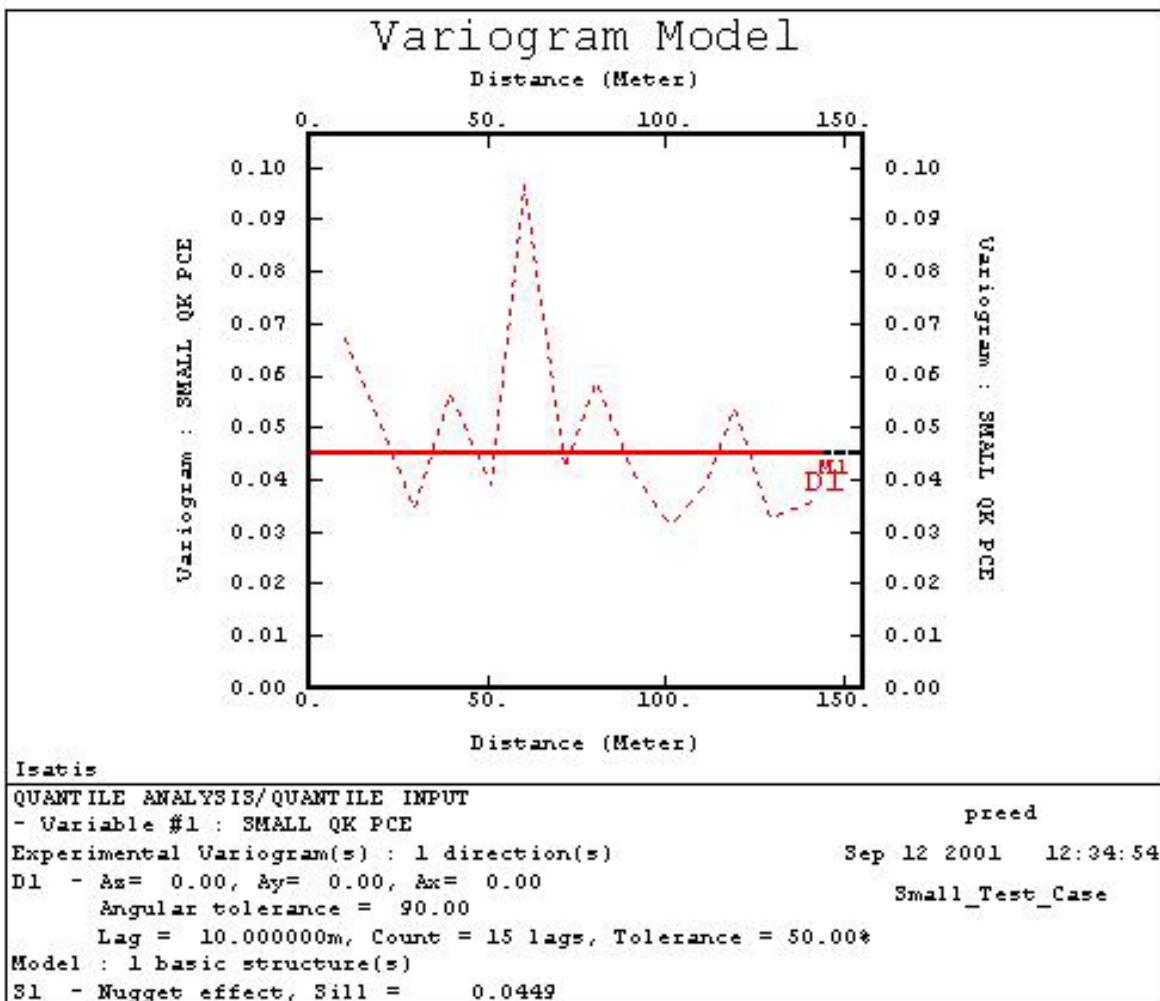


Figure B.19 Small case pure nugget variogram model for the quantile kriging system

preed - Aug 15 2001 11:34:59

---

---

| Structure Identification |

---

---

Data File Information:

Directory = Snapshot\_at\_5\_years  
File = 3-D Input Data

Seed File Information:

Directory = Snapshot\_at\_5\_years  
File = 3-D Input Data  
Type = POINT (26 points)

Model Name = 3-D IRF-k Model

Neighborhood Name = Unique - UNIQUE

Model : Covariance part

---

---

Number of variables = 1  
Number of basic structures = 4

**S1 : Nugget effect\***

S2 : Order-1 G.C. - Scale = 34.443554m

S3 : Spline G.C. - Scale = 34.443554m - C-H2 = 1.0

**S4 : Order-3 G.C. - Scale = 34.443554m - C-H2 = 1\*.**

Covariance Identification Step

---

---

Rank	Score	Structure 1	Structure 2	Structure 3	Structure 4
1	1.0181	5.0006e+005	0	0	97.554
2	0.97794	5.5e+005	0	0	0
3	0.95732	5.3788e+005	0	307.35	0
4	1.5984	0	9.5147e+005	32879	0
5	1.6848	0	9.0504e+005	0	11295
6	2.4531	0	6.5103e+005	0	0
7	110.37	0	0	26654	0
8	356.37	0	0	0	18327

Successfully processed = 26

Written to the disk = 0

Warning: the following structures have not been saved

as their sills were almost zero for all variables:

- Discarded structure : Order-1 G.C. - Scale = 34.443554m
- Discarded structure : Spline G.C. - Scale = 34.443554m - C-H2 = 1.0

**\*Selected structures are in bold above**

The Model Parameter File (3-D IRF-k Model) has been updated

Figure B.20 Small case Isatis output specifying the generalized covariance model for the intrinsic kriging system

preed - Aug 15 2001 11:34:59

---

---

Structure Identification	
--------------------------	--

---

---

Data File Information:

Directory = Snapshot\_at\_5\_years  
File = 3-D Input Data

Seed File Information:

Directory = Snapshot\_at\_5\_years  
File = 3-D Input Data  
Type = POINT (26 points)

Model Name = 3-D IRF-k Model

Neighborhood Name = Unique - UNIQUE

Drift Identification Step

---

---

Results are based on 26 measures

Trial	Mean Rank	Mean of Error	Variance of Error	Drift Trial
9	<b>6.7</b>	<b>11.359</b>	<b>5.9871e+005</b>	<b>1 z z2 (selected trend function)</b>
6	6.8	-9.8004	6.0758e+005	1 y y2
2	6.8	0.54892	6.2438e+005	1 x
13	6.8	22.44	6.4027e+005	1 y z
1	6.8	-4.3726e-014	5.9133e+005	No Drift
8	6.8	8.5282	6.1734e+005	1 z
12	7	7.1991	6.5188e+005	1 x z
11	7.3	15.081	6.5562e+005	1 x y
5	7.3	13.833	6.1646e+005	1 y
14	7.4	21.814	6.8002e+005	1 x y z
10	7.4	117.89	1.1388e+006	1 z z2 z3
7	7.6	-18.58	6.6107e+005	1 y y2 y3
15	8	6.3196	8.0575e+005	1 x y x2 xy y2
3	8.4	8.6376	6.5827e+005	1 x x2
16	9	137.4	1.0734e+006	1 x y x2 xy y2 x3 x2y xy2 y3
4	9.8	12.31	7.4213e+005	1 x x2 x3

The Model Parameter File (3-D IRF-k Model) has been updated

Figure B.21 Small case Isatis output specifying the trend function for the intrinsic kriging system



## APPENDIX C. CROSS VALIDATION

### C.1 Large Test Case Results

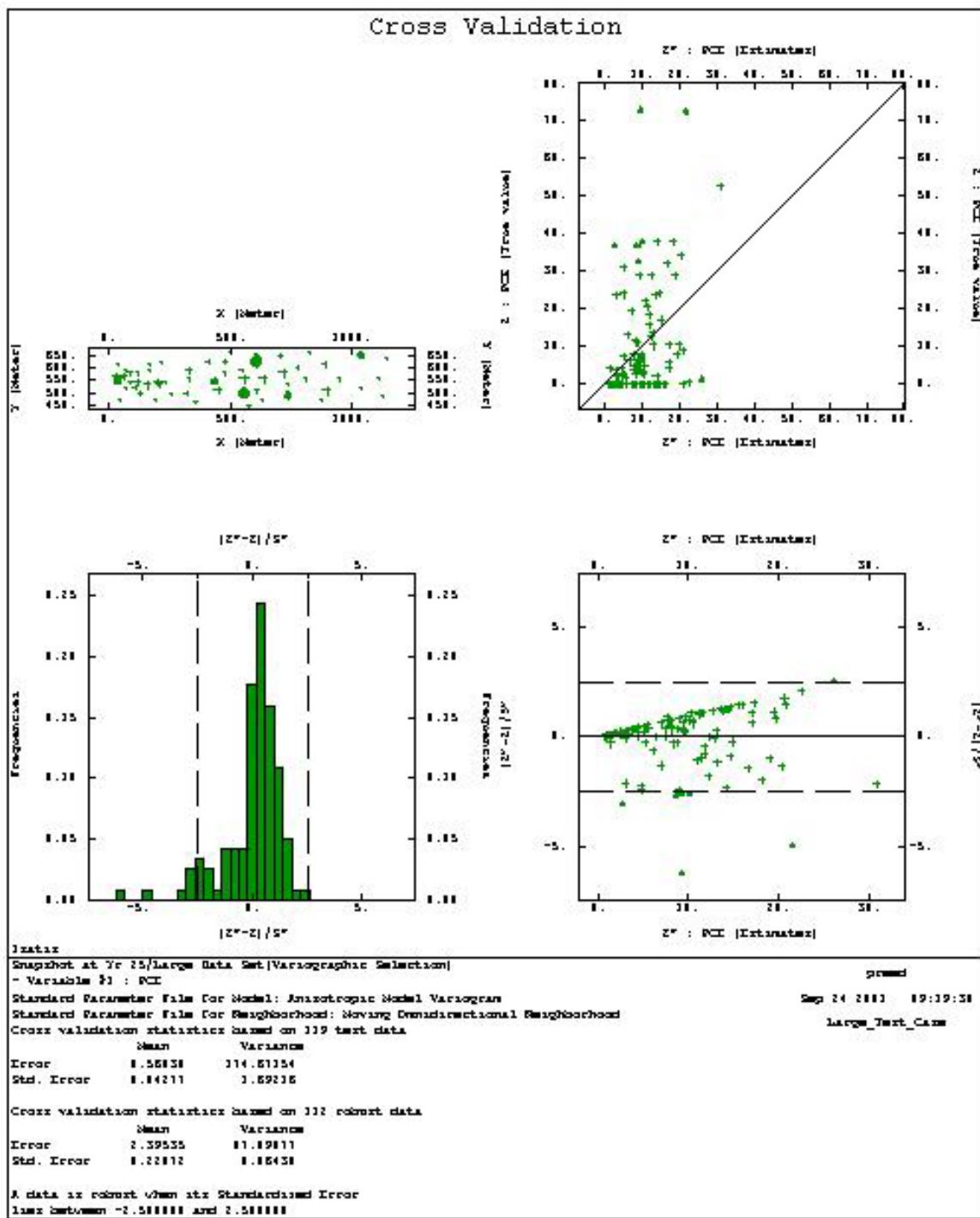


Figure C.1 Cross validation analysis of the anisotropic variogram model used for ordinary kriging in the large case

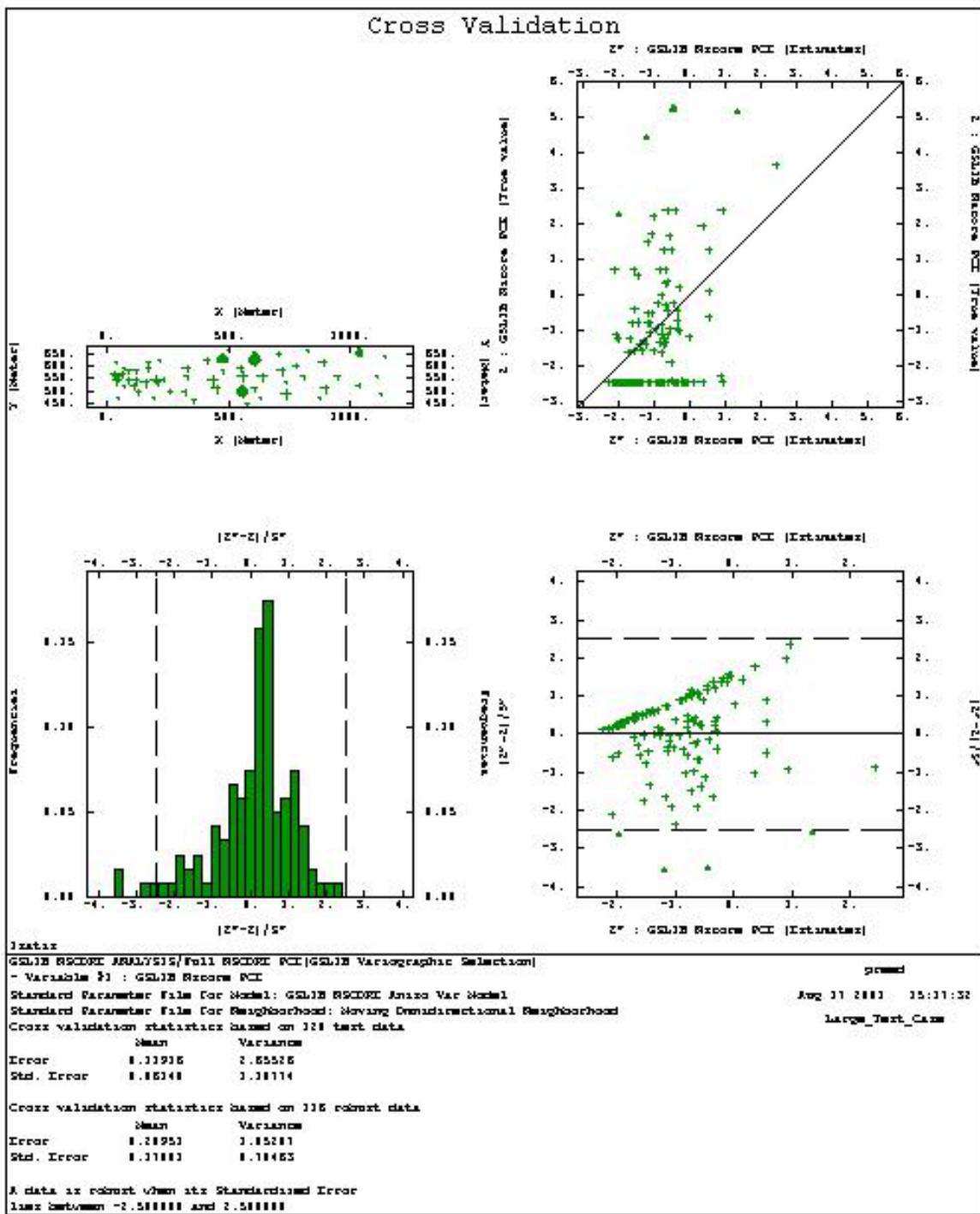


Figure C.2 Cross validation analysis of the anisotropic variogram model used for Multigaussian kriging in the large case

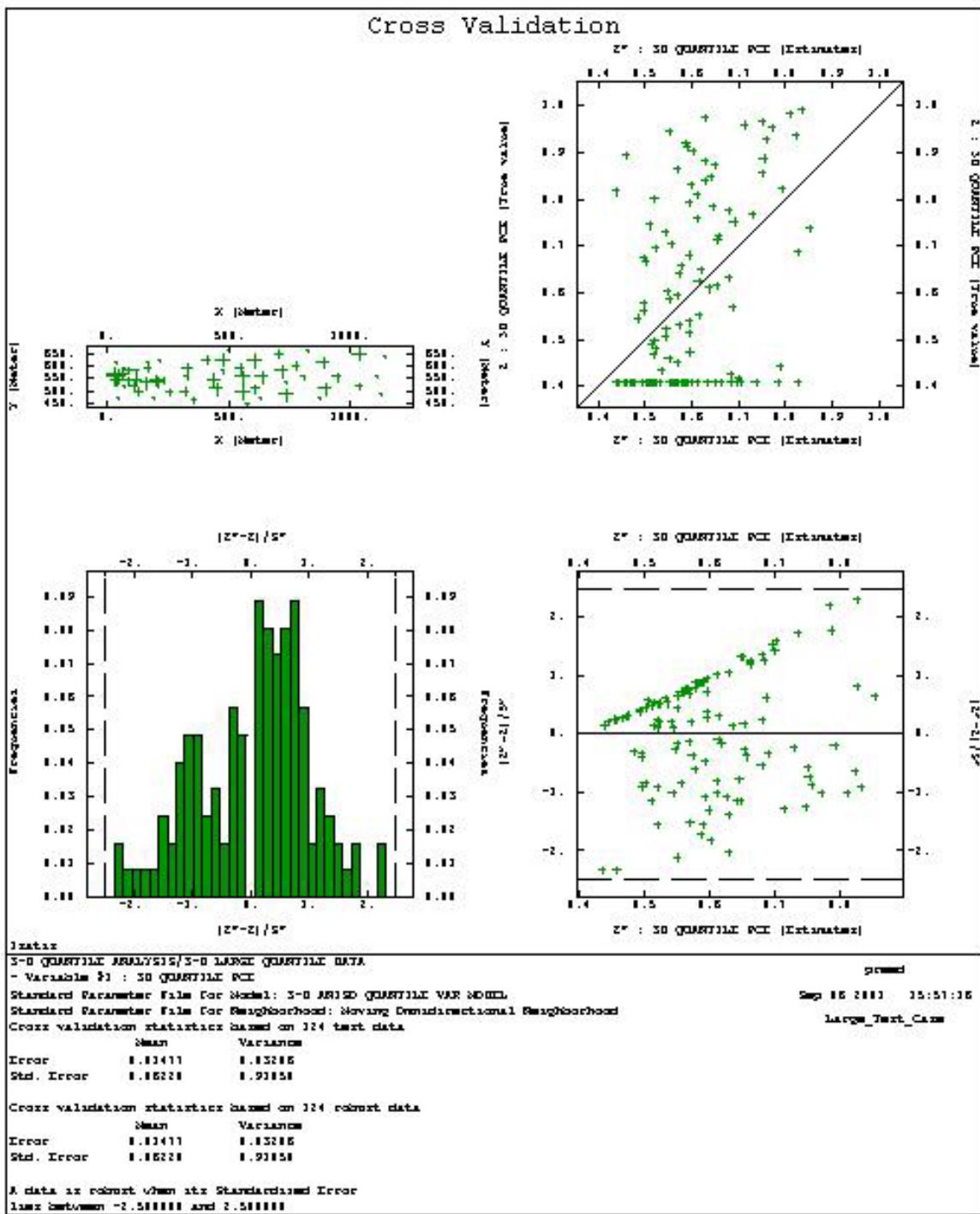


Figure C.3 Cross validation analysis of the anisotropic variogram model used for quantile kriging in the large case

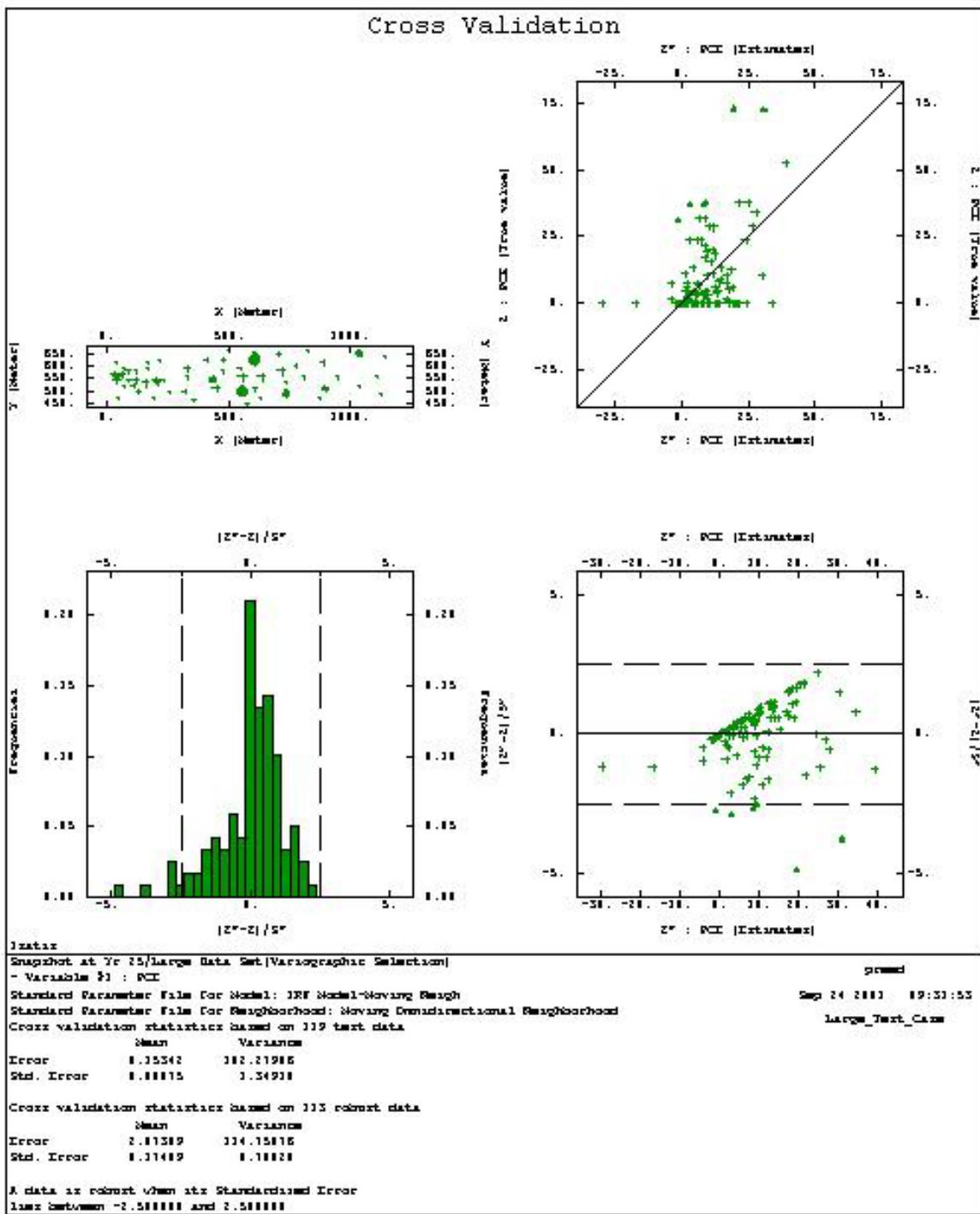


Figure C.4 Cross validation analysis of the isotropic variogram model used for intrinsic kriging in the large case

## C.2 Medium Test Case Results

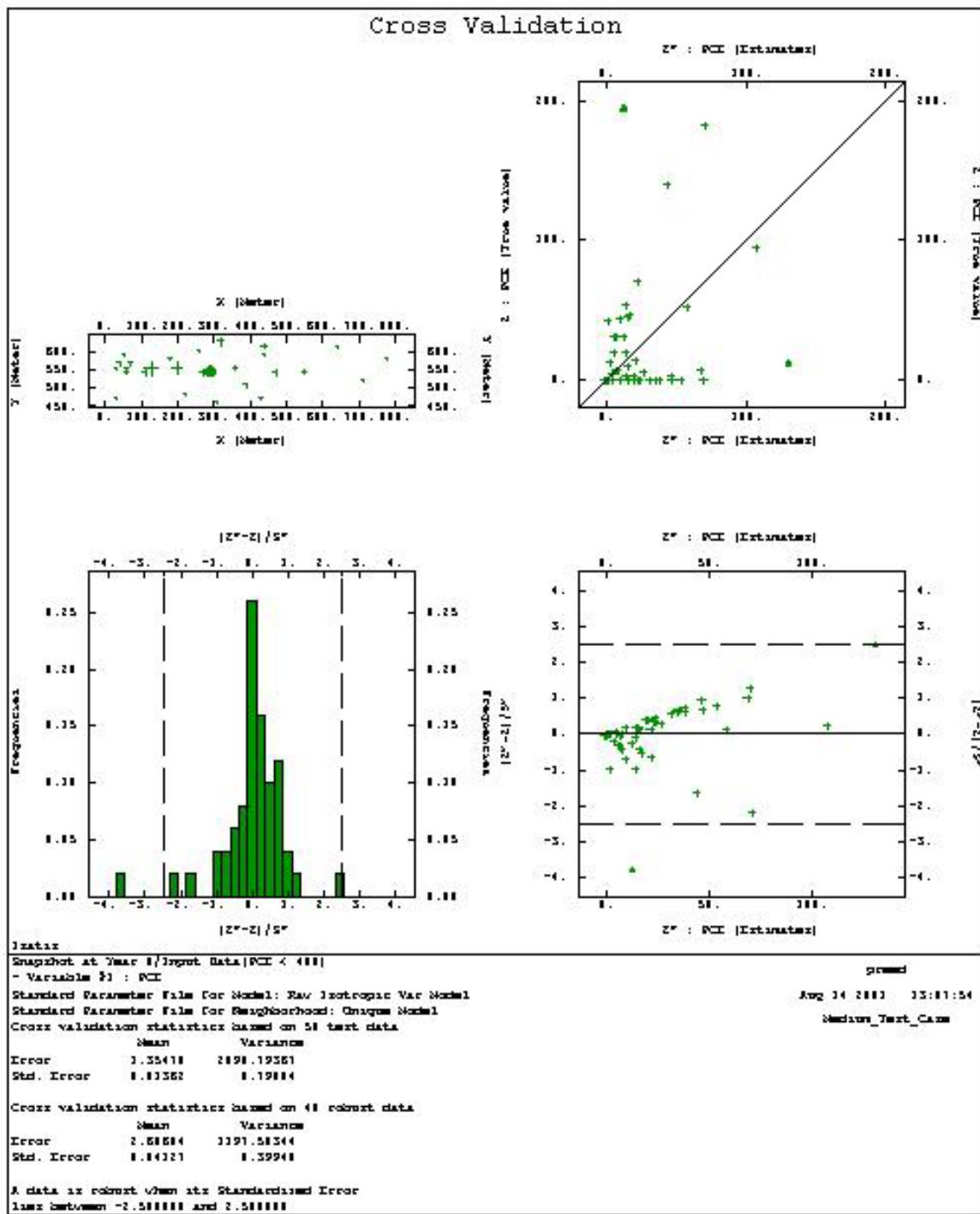


Figure C.5 Cross validation analysis of the isotropic variogram model used for ordinary kriging in the medium case

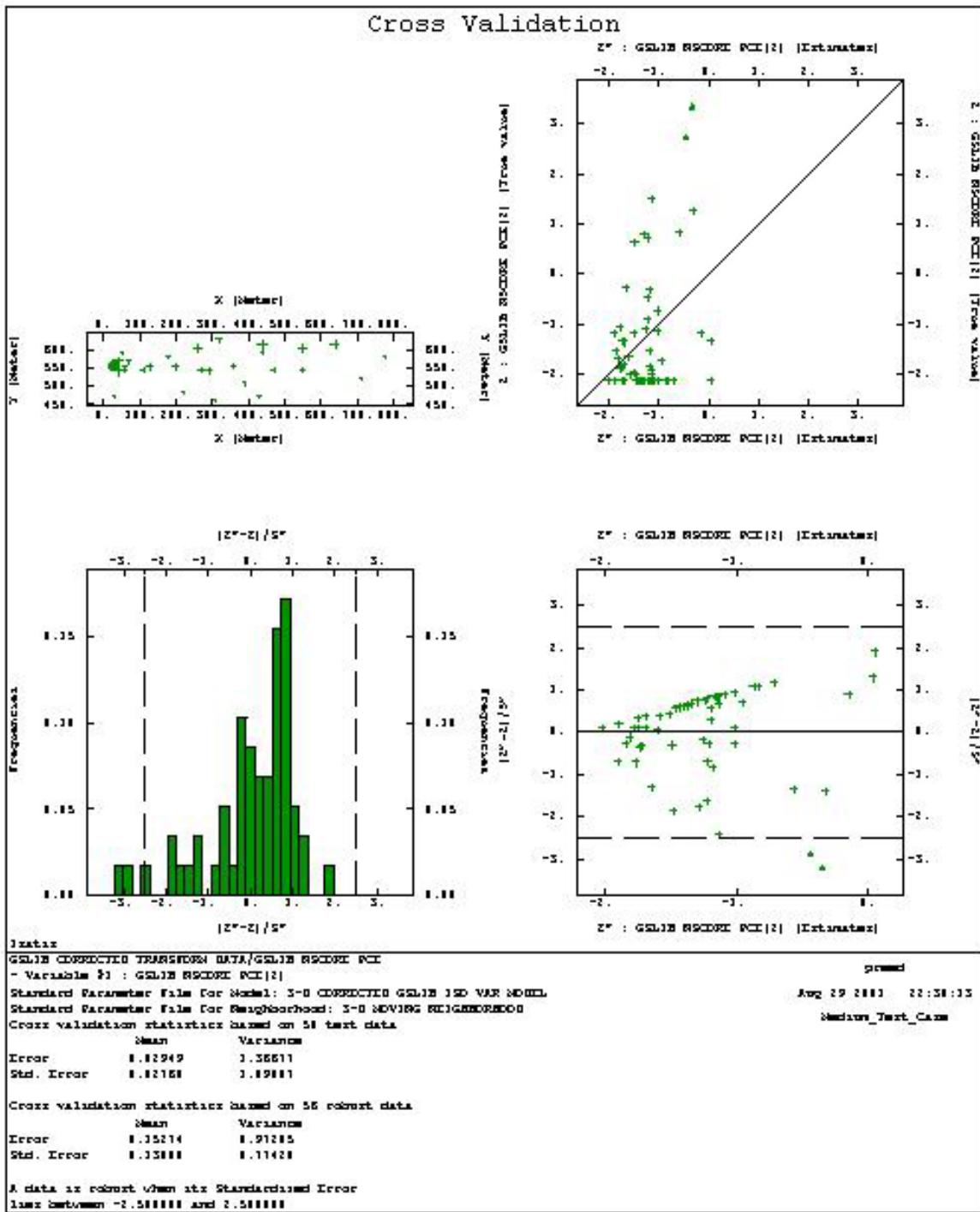


Figure C.6 Cross validation analysis of the isotropic variogram model used for Multigaussian kriging in the medium case

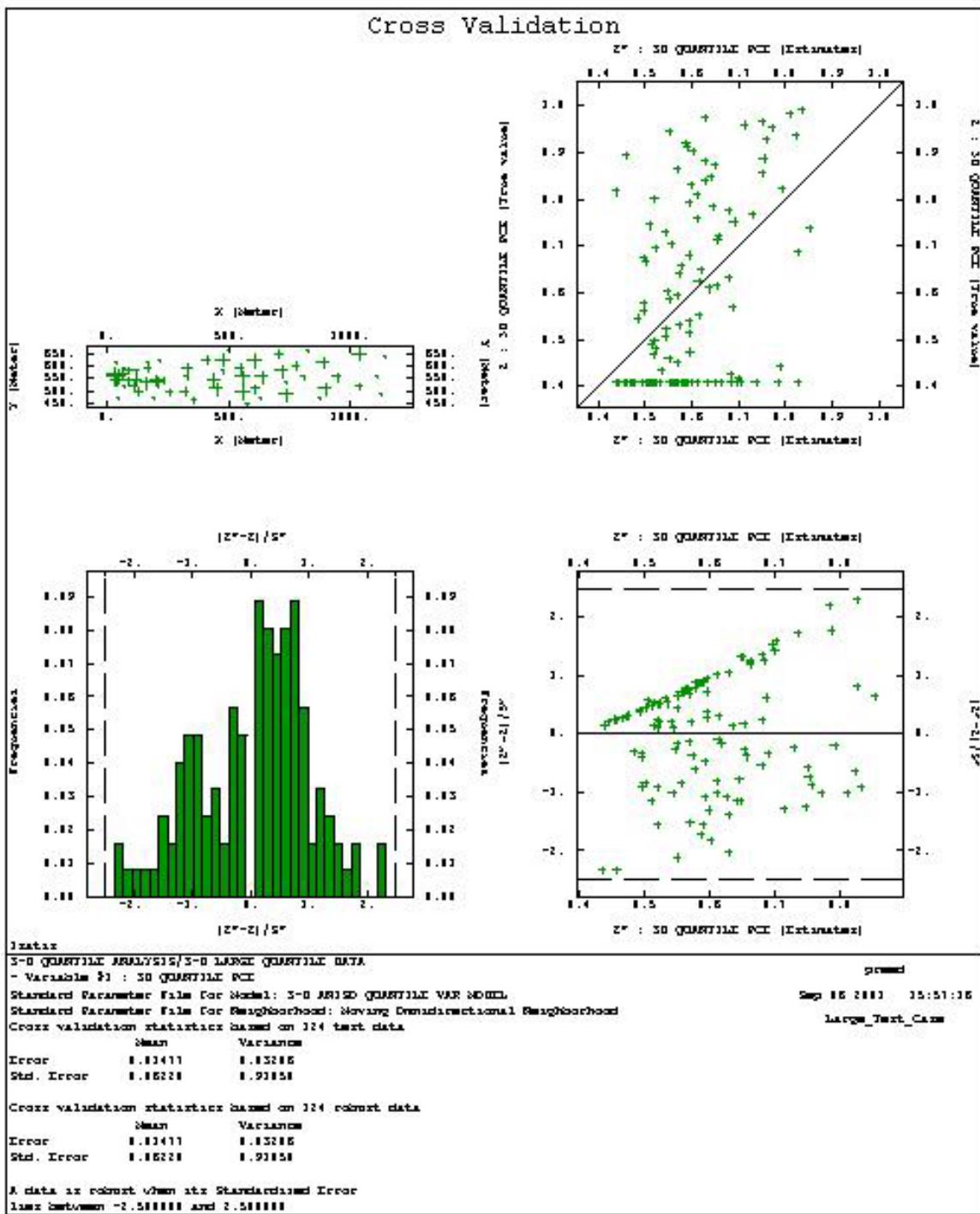


Figure C.7 Cross validation analysis of the isotropic variogram model used for quantile kriging in the medium case

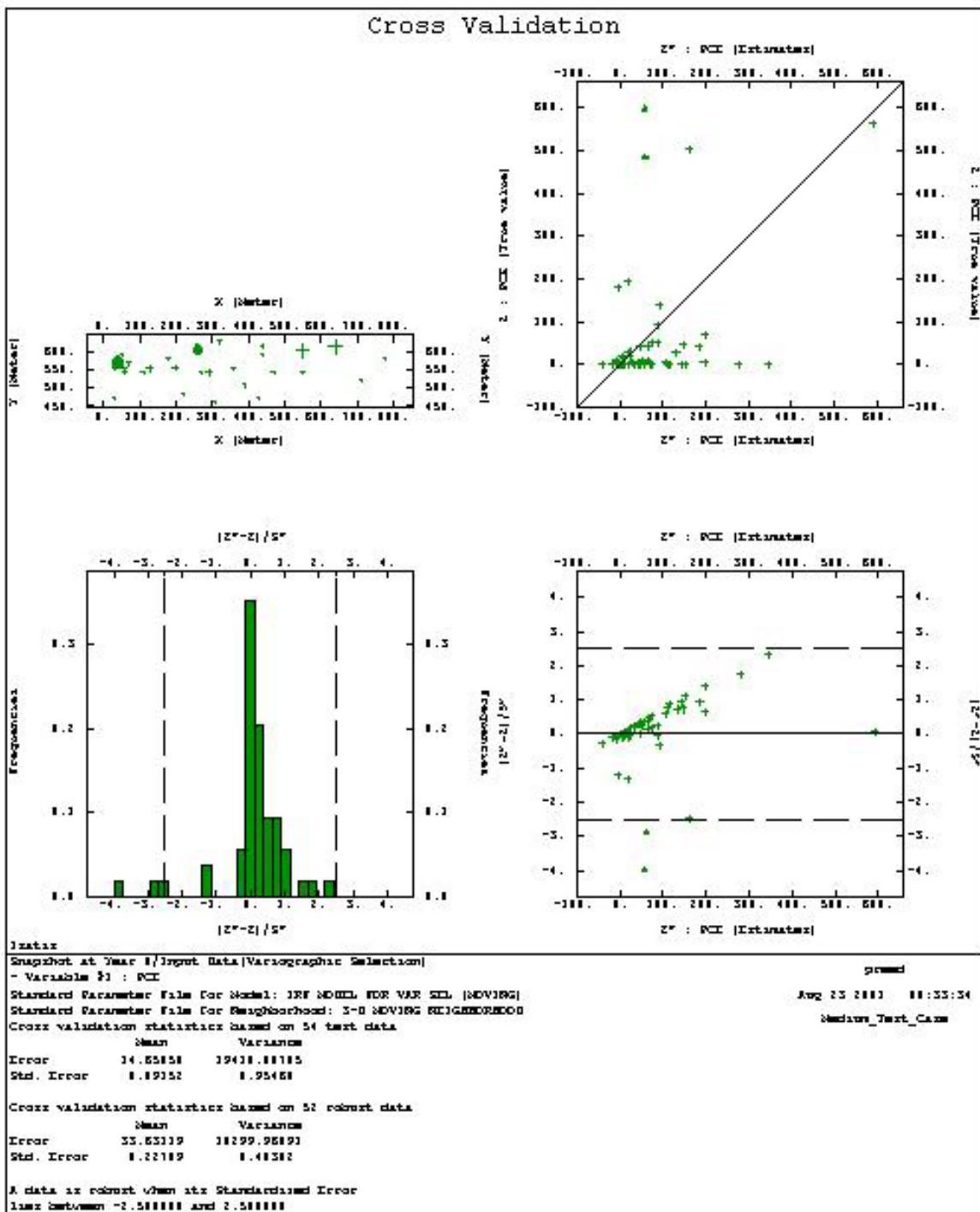


Figure C.8 Cross validation analysis of the nugget variogram model used for intrinsic kriging in the medium case

### C.3 Small Test Case Results

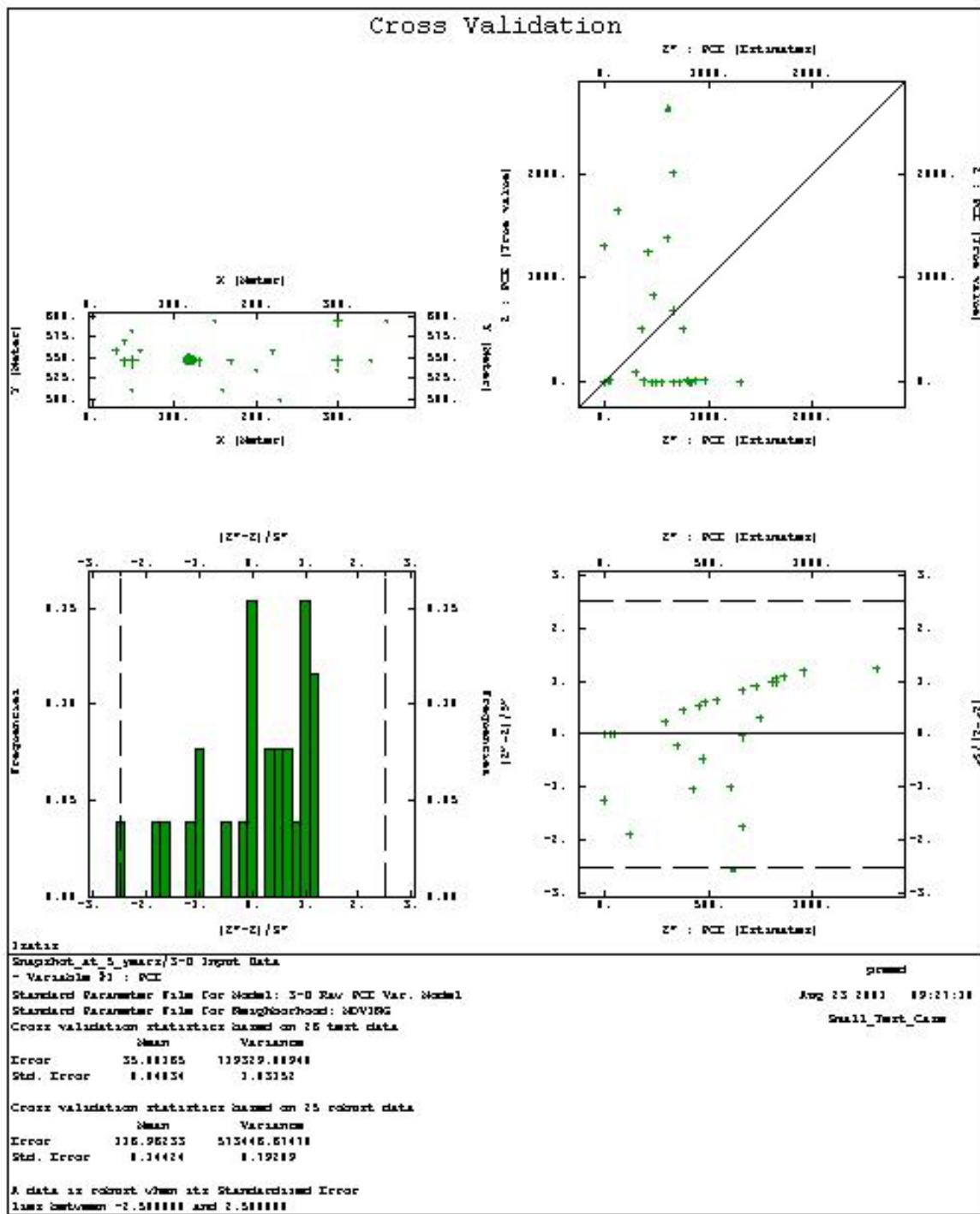


Figure C.9 Cross validation analysis of the nugget variogram model used for ordinary kriging in the small case

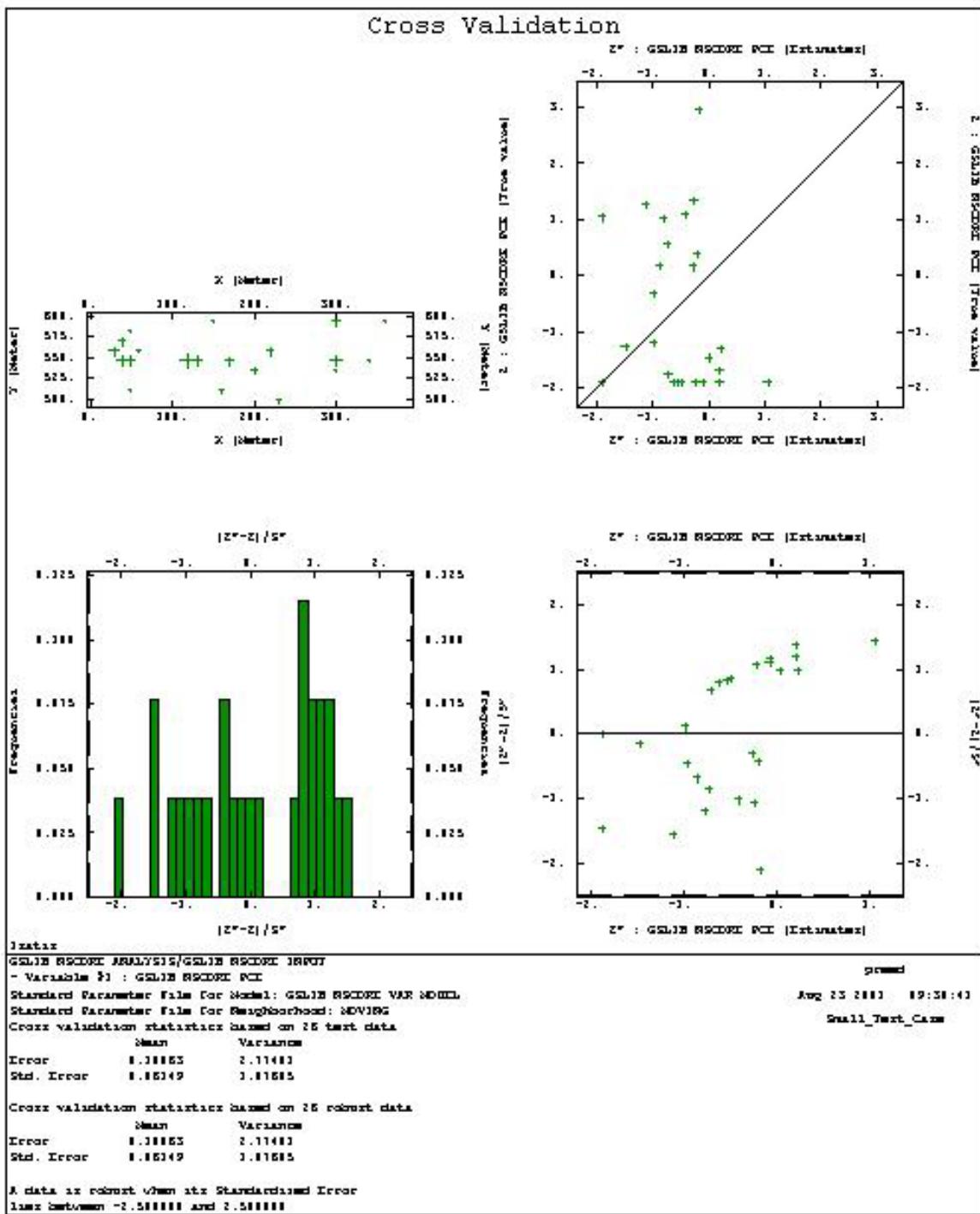


Figure C.10 Cross validation analysis of the nugget variogram model used for Multigaussian kriging in the small case

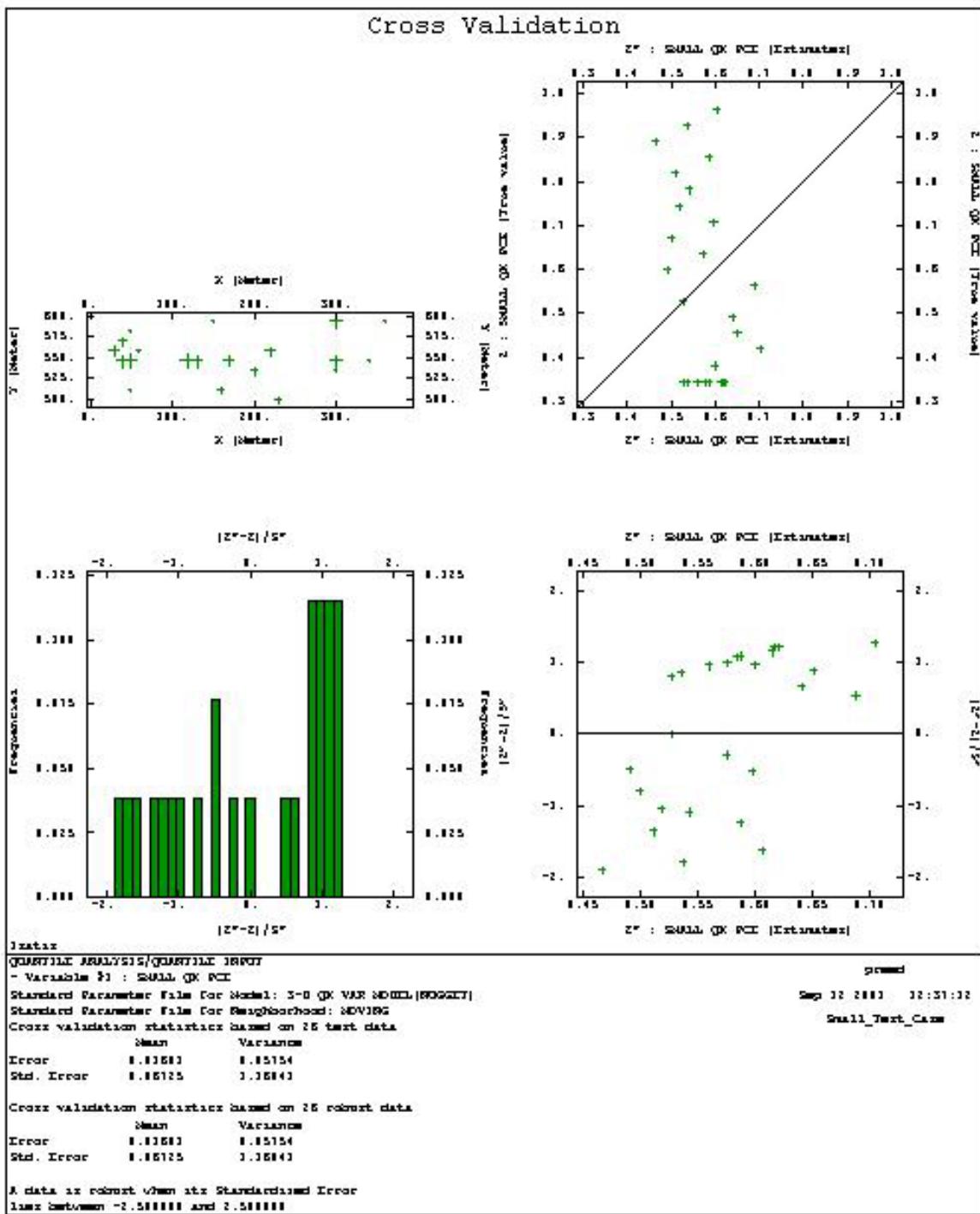


Figure C.11 Cross validation analysis of the nugget variogram model used for quantile kriging in the small case

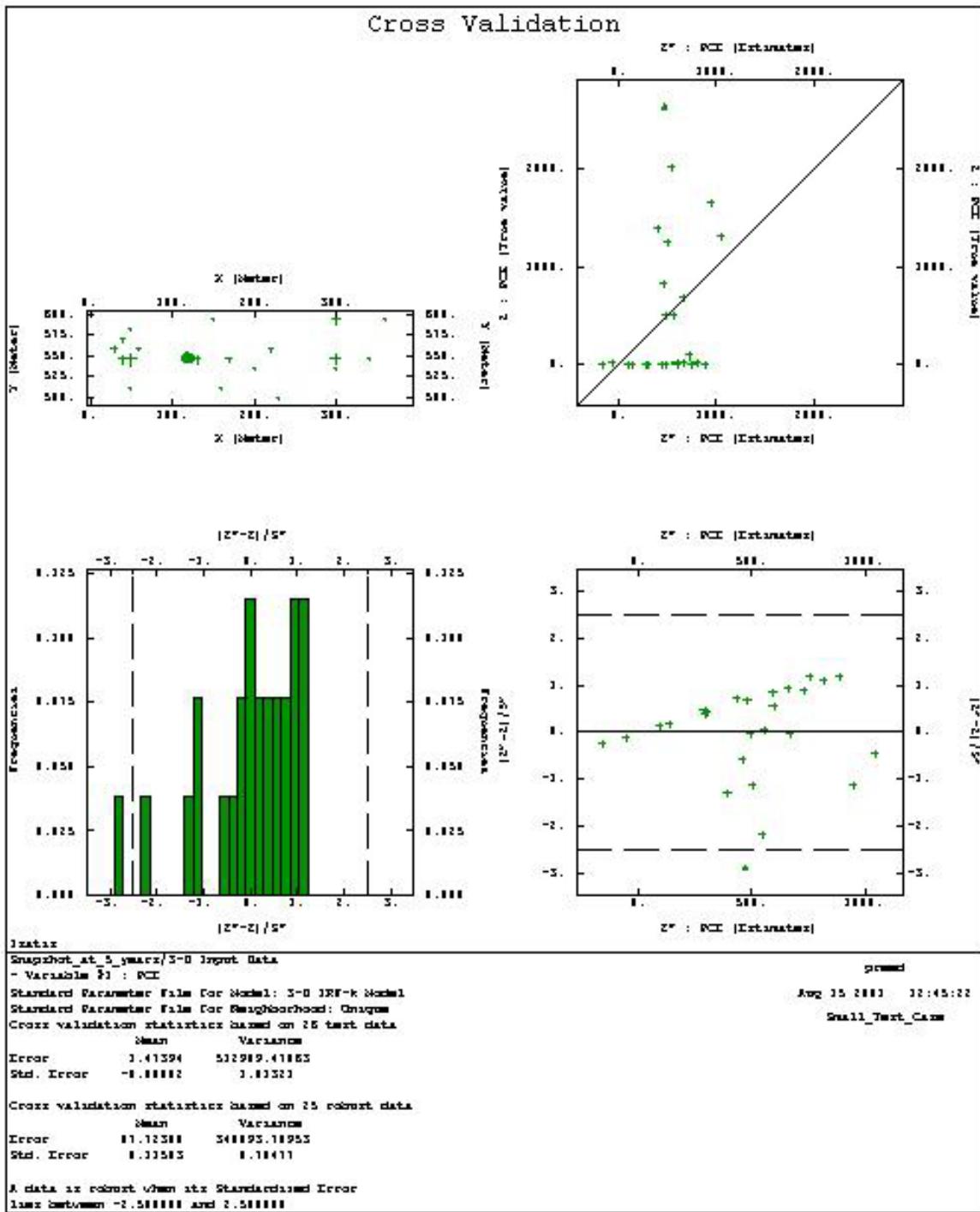


Figure C.12 Cross validation analysis of the isotropic variogram model used for intrinsic kriging in the small case



## APPENDIX D. FREQUENCY ANALYSIS WITH ZEROS

The highly heterogeneous nature of real-world groundwater contamination combined with the need to delimit the plumes often results in datasets that have multiple zero measures. These zero measures complicate the traditional implementation of frequency analysis used to attain the empirical cumulative distribution function (cdf) for contaminant concentrations. In this thesis, the empirical cdf is vital for quantile kriging discussed in Chapters 5 and 6. Guidance for dealing with multiple zero measures can be taken directly from the field of hydrology where zero measures are also prevalent. *Haan* (1977) shows how total probability theory can be used to address this issue. First, let us define the following quantities:

$C$  = random variable representing contaminant concentrations;

$S_n$  = set of nonzero contaminant data;

$S_o$  = set of zero data;

$c_i$  = the  $i^{th}$  contaminant measure;

$Pr(c_i \in S_n)$  = probability of observing a nonzero measure;

$Pr(c_i \in S_o)$  = probability of observing a zero measure;

$F(C \leq c_i)$  = empirical cdf for contaminant concentrations;

$Pr(C \leq c_i | c_i \in S_n)$  = probability of the  $i^{th}$  contaminant observation given the sample is a member of the set of nonzero data,  $c_i \in S_n$ ;

$Pr(C \leq c_i | c_i \in S_o)$  = probability of the  $i^{th}$  contaminant observation given the sample is a member of the set of zero data,  $c_i \in S_o$ ;

$i$  = the index for the contaminant data measures;

$N$  = total number of contaminant data;

$n$  = total number of nonzero contaminant data;

$r_i$  = the rank assigned to the  $i^{th}$  contaminant data measure.

Given these quantities, equation (D.1) utilizes total probability theory to divide the contaminant data set into 2 populations (1) the nonzero data and (2) the zero data when computing the empirical cdf,  $F(C \leq c_i)$ .

$$F(C \leq c_i) = Pr(C \leq c_i | c_i \in S_n) Pr(c_i \in S_n) + Pr(C \leq c_i | c_i \in S_o) Pr(c_i \in S_o) \quad (\text{D.1})$$

Given a total  $N$  contaminant measures of which  $n$  are nonzero measures, equations (D.2) and (D.3) illustrate how to compute the probabilities of attaining nonzero and zero measurements, respectively.

$$Pr(c_i \in S_n) = \frac{n}{N} \quad (\text{D.2})$$

$$Pr(c_i \in S_o) = \frac{N-n}{N} \quad (\text{D.3})$$

Equation (D.4) shows, as expected, the probability that the  $j^{th}$  observation is less than or equal to zero given it has been taken from the set of zero data is equal to 1.

$$Pr(C \leq c_i | c_i \in S_o) = \frac{N-n}{N-n} \rightarrow 1 \quad (\text{D.4})$$

Equation (D.5) is the result of substituting equations (D.2) - (D.4) into equation (D.1).

$$F(C \leq c_i) = Pr(C \leq c_i | c_i \in S_n) \frac{n}{N} + \frac{N-n}{N} \quad (\text{D.5})$$

The final quantity that remains to be specified in equation (D.5) is the conditional probability of attaining the  $i^{th}$  contaminant measure given that the measure is a member of the nonzero data set,  $Pr(C \leq c_i | c_i \in S_n)$ . This quantity is computed by ranking all of the data in ascending order (i.e., from the least to the greatest). When ranking the data all of the zero measures are assigned a rank of 0 (since they have been segregated into a different statistical population) and the nonzero ranks range from 1 to  $n$ . Equation (D.6) specifies how to compute  $Pr(C \leq c_i | c_i \in S_n)$ .

$$Pr(C \leq c_i \mid c_i \in S_n) = \frac{r_i}{n+1} \quad (\text{D.6})$$

Finally, substituting equation (D.6) into equation (D.5) yields equation (D.7), which is an appropriate measure of the empirical cdf in the presence of multiple zero measures.

$$F(C \leq c_i) = \frac{r_i}{n+1} \frac{n}{N} + \frac{N-n}{N} \quad (\text{D.7})$$

**APPENDIX D. FREQUENCY ANALYSIS WITH ZEROS .....210**

## **VITA**

Patrick Michael Reed was awarded the University of Missouri's Curators Scholarship to attend the University of Missouri at Rolla (UMR) in 1993. He graduated from UMR in May 1997, Summa Cum Laude with a Bachelor of Science in Geological Engineering. While at UMR, he also worked for the Missouri Department of Natural Resources Division of Geology and Land Survey helping to develop a database cataloguing hazardous waste sites throughout the state of Missouri. In August 1997, Patrick was admitted into the Civil and Environmental Engineering's Masters Program at the University of Illinois at Urbana-Champaign. His initial year of study was funded by a University of Illinois Fellowship from 1997 to 1998. His Masters research investigated cost effective sampling strategies for contaminated groundwater sites undergoing long term monitoring. He graduated with his Master of Science in August 1999 and was subsequently accepted into the Ph.D. program in Civil and Environmental Engineering at the University of Illinois at Urbana-Champaign. His doctoral studies were funded by the United States Environmental Protection Agency's Science to Achieve Results (STAR) Fellowship from 1999 to 2002. In February 2002, Patrick was honored with the Ross J. Martin Award for Outstanding Research Achievement by a Graduate Student within the College of Engineering at the University of Illinois at Urbana-Champaign for his research into groundwater monitoring design, multiobjective genetic algorithms, and spatial interpolation methods. In August 2002, he will join the faculty of the Civil and Environmental Engineering Department at the Pennsylvania State University as an Assistant Professor of Hydroinformatics.

