Incorporating Geophysics in the Hydrogeological Decision-Making Process

by

EDWARD (NED) ANDREW CLAYTON

B.S.E., Princeton University, 1990

A THESIS SUBMITTED IN PARTIAL FULFILLMENT OF THE REQUIREMENTS FOR THE DEGREE OF MASTER OF APPLIED SCIENCE in

THE FACULTY OF GRADUATE STUDIES

Department of Earth and Ocean Sciences

We accept this thesis as conforming to the required standard

THE UNIVERSITY OF BRITISH COLUMBIA

February 2000

© Edward Andrew Clayton, 2000
In presenting this thesis in partial fulfilment of the requirements for an advanced
degree at the University of British Columbia, I agree that the Library shall make it
freely available for reference and study. I further agree that permission for extensive
copying of this thesis for scholarly purposes may be granted by the head of my
department or by his or her representatives. It is understood that copying or
publication of this thesis for financial gain shall not be allowed without my written
permission.

Department of Earth and Ocean Sciences

The University of British Columbia
Vancouver, Canada

Date April 1, 2000
ABSTRACT

A risk-based economic decision making framework for hydrogeology-related engineering problems was developed, and tested in a real-world case study, for incorporating geophysics in a probabilistic site conceptual model that is directly linked to a decision analysis model. The framework employs a flexible, practical geostatistical methodology that quantitatively accounts for uncertainty associated with measurement inaccuracies and spatial variability. Two types of decision analyses are performed: (1) all types of existing information are integrated in the site conceptual model and the most cost-effective engineering design is determined, accounting for risk costs associated with uncertainty, and (2) the expected future economic worth of different measurements, including geophysics, to the overall problem is evaluated.

Site characterization is often one of the most important and expensive activities in hydrogeology-related engineering problems, having a large impact on the engineering solution and total cost. Thus, this activity should be carried out in the most effective way possible—providing the most worthwhile information at the lowest possible cost. The worth of information is defined by how much the overall economic value of the engineering solution is improved, including reduction in costs associated with the risk of engineering failure. Most conventional characterization techniques involve invasive sampling or testing of the subsurface, sampling only a small fraction of the overall site. Geophysics measurements are non-invasive, have a much larger spatial coverage, can be acquired at a much higher density, and are relatively inexpensive. However, geophysics measurements usually do not measure the properties of interest in these problems, and are often plagued by noise and other uncertainties that limit quantitative usage of the data. These measurements could be a very valuable site characterization technique if they could be quantitatively incorporated in the decision making process, in a way that accounts for their inherent uncertainty.

The Markov-Bayes indicator geostatistics methodology [Alabert, 1987; Zhu and Journel, 1993] provides a versatile and straightforward approach for incorporating geophysics, and any other indirect or direct site characterization measurements of the property of interest, in a probabilistic spatial model of the property. The uncertainty associated with the indirect measurements is determined through a simple calibration between collocated indirect and direct measurements, and accounted for in the probabilistic model using indicator cokriging. The output of the model is either (1) a set of probability of class membership maps for defined interval classes of the property or (2) a set of equally-likely realizations of class membership, or actual values, of the property. The methodology was adapted and expanded to a comprehensive set of routines, optimized for incorporating geophysics data in the uncertainty model and easily linking the model to a decision analysis model.

A real-world case study involving soil contamination remediation was performed to test the integrity and practicality of the developed Markov-Bayes uncertainty framework, and its ability to incorporate actual geophysics measurements. The case study site requires excavation and selective treatment of soil contaminated above specified action levels, with a penalty cost for underclassifying or over-classifying soil contamination. A soil remediation design and data worth analysis decision model, closely linked to the uncertainty framework, was developed for the problem. The expected value optimal remediation design was determined based on the information provided by soil sample data. A sensitivity analysis was also performed—for a range of contamination underclassification unit costs and different real-time sampling alternatives. The optimal remediation designs require excavation of almost the entire site and range in total cost from $4.5 to $9 million Canadian dollars, the higher cost designs corresponding to the scenarios where real-time sampling is not an option and the underclassification costs are substantially higher than when contamination is correctly classified. When excavation and real-time batch sampling, followed by appropriate treatment, is an alternative, it is the optimal action for most of the site—since it eliminates risk costs. A data worth analysis was also performed to evaluate the
worth of additional soil sampling versus geophysics surveys of different data quality levels. The results show that soil sampling—from 20 boreholes evenly spaced across the site—provides negligible worth to the site remediation. The same is true for geophysics, except for the scenario where the geophysics is of very good quality for delineating contamination (greater than 80% probability of correctly identifying the highest contamination levels) and the surveys cover most of the site. Even this scenario provides little value unless the underclassification cost is greater than two times the correct classification cost. Ground penetrating radar (GPR) and frequency domain electromagnetics (FDEM) surveys were acquired as an indirect measurement of hydrocarbon and metals contamination, and the processed data was incorporated in the decision model to evaluate the change in the optimized remediation design. Combining the results from both surveys in the Markov-Bayes uncertainty model produces a significant change in the probability of contamination maps in the region where the surveys were performed. However, the geophysics provides little improvement to the overall design and, consequently, little reduction in total cost. This result is anticipated from the data worth analysis, since the survey covered only a small portion of the overall site.

The case study results illustrate that the developed Markov-Bayes uncertainty framework can be effectively employed in hydrogeology-related problems to (1) evaluate the economic worth of geophysics and (2) incorporate geophysics data in a risk-based decision model. While the geophysics acquired in this study produced little value to the decision making in this particular problem, the integration of the “soft” results from both the GPR and FDEM surveys with “hard” soil sample data had a significant impact on the contamination probability model in the region of the surveys. This suggests that using geophysics within the Markov-Bayes uncertainty framework could provide a promising technique for indirectly measuring hydrogeological properties—providing much greater spatial coverage than conventional sampling techniques.
# TABLE OF CONTENTS

<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>ABSTRACT</td>
<td>II</td>
</tr>
<tr>
<td>TABLE OF CONTENTS</td>
<td>IV</td>
</tr>
<tr>
<td>LIST OF TABLES</td>
<td>XI</td>
</tr>
<tr>
<td>LIST OF FIGURES</td>
<td>XII</td>
</tr>
<tr>
<td>ACKNOWLEDGEMENTS</td>
<td>XXIII</td>
</tr>
<tr>
<td>CHAPTER 1 INTRODUCTION</td>
<td>1</td>
</tr>
<tr>
<td>1.1 Motivation for research</td>
<td>2</td>
</tr>
<tr>
<td>1.2 Research goal</td>
<td>4</td>
</tr>
<tr>
<td>1.3 Research contributions</td>
<td>4</td>
</tr>
<tr>
<td>1.4 Thesis overview</td>
<td>6</td>
</tr>
<tr>
<td>CHAPTER 2 BACKGROUND</td>
<td>9</td>
</tr>
<tr>
<td>2.1 Example hydrogeology problem</td>
<td>9</td>
</tr>
<tr>
<td>2.1.1 Geology of site</td>
<td>10</td>
</tr>
<tr>
<td>2.1.2 Hydrogeology of site</td>
<td>11</td>
</tr>
<tr>
<td>2.1.3 Contamination of site and the resulting implications</td>
<td>11</td>
</tr>
<tr>
<td>2.2 The nature of uncertainty in hydrogeology problems</td>
<td>12</td>
</tr>
<tr>
<td>2.2.1 The relationship between uncertainty and scale</td>
<td>13</td>
</tr>
<tr>
<td>2.2.1.1 Measurement scale</td>
<td>13</td>
</tr>
</tbody>
</table>
## 2.2.1.2 Model scale

2.2.1.3 The relationship between model and measurement scales

2.2.1.4 Network scale

2.2.1.5 Spectral analysis of scale and uncertainty

2.2.1.6 Summary

2.2.2 Uncertainty associated with the measurement process

### 2.3 The decision-making process

2.3.1 Key components of decision-making process for engineering problems

2.3.2 Concept of risk in decision-making

2.3.3 Engineering versus health risk

2.3.4 Concept of data worth

### 2.4 The site characterization process

2.4.1 Objectives

2.4.2 Organization

2.4.2.1 Temporal perspective

2.4.2.2 Problematic perspective

2.4.2.3 Summary

### 2.5 Review of previous work

2.5.1 Geophysics applications in hydrology: general references and examples

2.5.2 Incorporating geophysics in probabilistic framework: measurement uncertainty not quantified

2.5.3 Incorporating geophysics in probabilistic framework: measurement uncertainty quantified

2.5.4 Incorporating geophysics in hydrogeological decision-making framework
# TABLE OF CONTENTS

## CHAPTER 3 GEOPHYSICS IN HYDROGEOLOGY PROBLEMS

3.1 Introduction to geophysics ............................................................... 64

3.2 Applications of geophysics in hydrogeology-related studies .................. 68

3.3 Advantages of geophysics ............................................................... 73

3.4 Difficulties associated with applying geophysics .................................. 76

  3.4.1 The relationship between instrument observations and geophysical properties ........ 78

    3.4.1.1 Physical model – geophysical inversion ....................................... 80

    3.4.1.2 Empirical and statistical models .............................................. 84

  3.4.2 The relationship between geophysical properties and material properties .......... 95

    3.4.2.1 The rock physics problem – conceptual model approach ................. 97

    3.4.2.2 The rock physics problem – statistical approach .......................... 100

    3.4.2.3 The rock physics problem – comparison of approaches ................. 101

  3.4.3 Noise in geophysical measurements ........................................... 103

  3.4.4 Summary .................................................................................... 107

## CHAPTER 4 ACCOUNTING FOR AND MODELING UNCERTAINTY

4.1 Introduction .................................................................................... 111

  4.1.1 Uncertainty analysis approaches ............................................... 111

  4.1.2 Probability: Philosophical schools of thought ............................. 115

4.2 Accounting for uncertainty in spatial random variables ......................... 117

  4.2.1 Derivation of spatial RVs ......................................................... 117

  4.2.2 Classical least squares regression analysis .................................... 120

  4.2.3 Geostatistics .............................................................................. 122

    4.2.3.1 Kriging estimation methods ....................................................... 125
# TABLE OF CONTENTS

4.2.3.2 Geostatistical simulation ........................................................................ 131
4.2.4 Scale ........................................................................................................ 132

4.3 **Probability analysis using the Markov-Bayes indicator estimation methodology** ........ 135

4.3.1 Model inference .......................................................................................... 135
4.3.1.1 Exploratory data analysis ...................................................................... 136
4.3.1.2 Secondary variable calibration model ...................................................... 138
4.3.1.3 Prior expected value .............................................................................. 143
4.3.1.4 Covariance models .............................................................................. 146
4.3.2 Geostatistical estimation ............................................................................ 152
4.3.3 Limitations .................................................................................................. 160

4.4 **Summary** .................................................................................................. 161

## CHAPTER 5 CASE STUDY ............................................................................... 165

5.1 **Introduction** ............................................................................................. 171

5.2 **Background** .............................................................................................. 174

5.2.1 Setting ........................................................................................................ 174
5.2.2 History of site ............................................................................................ 175

5.3 **Site Characterization** ............................................................................... 176

5.3.1 Soil and groundwater sampling and analyses ............................................. 177
5.3.1.1 Borehole logging ................................................................................ 177
5.3.1.2 Soil sampling for chemical analysis ...................................................... 178
5.3.1.3 Groundwater sampling for chemical analysis ...................................... 181
5.3.1.4 Potentiometric head measurements and slug tests ......................... 183
5.3.1.5 Oil-interface probe measurements ...................................................... 184
5.3.2 Geophysics ............................................................................................... 184
<table>
<thead>
<tr>
<th>Section</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>5.3.2.1 Choice of geophysical techniques</td>
<td>185</td>
</tr>
<tr>
<td>5.3.2.2 GPR Surveys</td>
<td>189</td>
</tr>
<tr>
<td>5.3.2.3 FDEM Surveys</td>
<td>194</td>
</tr>
<tr>
<td>5.4 Analysis and Modeling</td>
<td>197</td>
</tr>
<tr>
<td>5.4.1 Risk-Cost-Benefit Decision Model</td>
<td>197</td>
</tr>
<tr>
<td>5.4.2 Geostatistical Model</td>
<td>207</td>
</tr>
<tr>
<td>5.4.2.1 Exploratory data analysis</td>
<td>208</td>
</tr>
<tr>
<td>5.4.2.2 Model inference and validation</td>
<td>217</td>
</tr>
<tr>
<td>5.4.3 Present Day (Prior) Decision Analysis</td>
<td>237</td>
</tr>
<tr>
<td>5.4.3.1 Contamination level probabilities</td>
<td>238</td>
</tr>
<tr>
<td>5.4.3.2 Remediation design</td>
<td>262</td>
</tr>
<tr>
<td>5.4.4 Future Site Characterization Data Worth (Preposterior) Analysis</td>
<td>281</td>
</tr>
<tr>
<td>5.4.4.1 Perfect information</td>
<td>287</td>
</tr>
<tr>
<td>5.4.4.2 Borehole sampling</td>
<td>289</td>
</tr>
<tr>
<td>5.4.4.3 Geophysics</td>
<td>295</td>
</tr>
<tr>
<td>5.4.4.4 Summary</td>
<td>311</td>
</tr>
<tr>
<td>5.4.5 Decision Analysis With Geophysics (Posterior Analysis)</td>
<td>311</td>
</tr>
<tr>
<td>5.4.5.1 FDEM processing</td>
<td>312</td>
</tr>
<tr>
<td>5.4.5.2 GPR processing</td>
<td>318</td>
</tr>
<tr>
<td>5.4.5.3 FDEM calibration</td>
<td>323</td>
</tr>
<tr>
<td>5.4.5.4 GPR calibration</td>
<td>332</td>
</tr>
<tr>
<td>5.4.5.5 Updating contamination probabilities</td>
<td>338</td>
</tr>
<tr>
<td>5.4.5.6 Updating remediation designs</td>
<td>350</td>
</tr>
<tr>
<td>5.4.5.7 Summary</td>
<td>356</td>
</tr>
</tbody>
</table>
**TABLE OF CONTENTS**

CHAPTER 6 CONCLUSIONS AND RECOMMENDATIONS

6.1 Research results

6.1.1 Uncertainty model

6.1.2 Decision model

6.1.3 Case study

6.1.4 Summary

6.2 Recommendations for further research

BIBLIOGRAPHY

APPENDIX A THE GEOPHYSICAL INVERSE PROBLEM

A.1 Background

A.2 Well-posed inverse problems

A.3 Conceptualization of the inverse problem

A.4 Inverse solution methods

A.4.1 Global versus local methods

A.4.2 Direct versus indirect methods

A.4.3 Accounting for uncertainty in the inverse solution

A.4.3.1 Backus-Gilbert approach

A.4.3.2 Minimum relative entropy approach

A.4.3.3 Geostatistical approach
TABLE OF CONTENTS

APPENDIX B THE MARKOV-BAYES GEOSTATISTICAL APPROACH ....... 431

B.1 Introduction ........................................................................................................ 431

B.2 Overview of indicator geostatistics.................................................................. 432

B.2.1 Definition of indicator spatial random variables........................................ 432
B.2.2 Properties of indicators.................................................................................. 432
B.2.3 Indicator coding of data ................................................................................ 433
B.2.4 Indicator kriging ............................................................................................ 436
B.2.5 Indicator simulation ....................................................................................... 440
| Table 5.3.1: Case study site general stratigraphy | 178 |
| Table 5.3.2: Regulatory action levels for soil contamination | 180 |
| Table 5.3.3: Regulatory action levels for groundwater contamination | 182 |
| Table 5.4.1a: Case study decision model parameters—objective and decision questions | 198 |
| Table 5.4.1b: Case study decision model parameters—constraints | 199 |
| Table 5.4.1c: Case study decision model parameters—decision variables | 200 |
| Table 5.4.2: Declustering analysis results for all contaminant levels | 219 |
| Table 5.4.3: Variogram parameters for all contaminant levels | 230 |
| Table 5.4.4: Number of soil samples used as hard and soft data for cokriging estimation | 248 |
| Table 5.4.5: Comparison of site-wide contaminant level expected values of kriging and simulation grid results for 3 conditioning data scenarios: (1) hard sample data, (2) all sample data, and (3) all sample data and soft C prior probabilities | 262 |
| Table 5.4.6: Hypothetical soft data calibration data set where the soft measurement can delineate between hard values of 2 and less than 2 with 80% success, but has no ability to distinguish between values of 1 and 0 | 296 |
| Table 5.4.7: FDEM conductivity cutoffs chosen for calibration of FDEM data to metals and MOG contaminant levels, measured in soil samples | 326 |
| Table 5.4.8: Summary parameters for results of calibration between FDEM conductivity and metals / MOG soil samples | 328 |
| Table 5.4.9: GPR amplitude cutoffs chosen for calibration of GPR data to metals and MOG contaminant levels, measured in soil samples | 335 |
| Table 5.4.10: Summary parameters for results of calibration between GPR relative amplitudes and metals / MOG soil samples | 337 |
| Table 6.1.1: Summarized case study decision analysis remediation design results—for a range of decision model parameters and three information usage scenarios: (1) direct soil sample measurements, (2) direct and co-contaminant soil sample measurements, (3) direct and co-contaminant soil sample measurements plus prior probabilities of contamination exceeding action levels | 376 |
| Table 6.1.2: Summarized case study future data worth decision analysis results—for several contamination underclassification costs and three future site characterization scenarios: (1) soil sample measurements for detecting all contaminants in 21 boreholes spread evenly across the site, (2) geophysics measurements for detecting metals contamination, (3) geophysics measurements for detecting pure phase hydrocarbons contamination. Data worth is represented by reduction in remediation design cost with additional information | 380 |
| Table 6.1.3: Summarized case study decision analysis remediation design results after incorporating both the FDEM and GPR geophysics measurements—for a range of decision model parameters. Data worth is defined by the reduction in the remediation design cost with the additional information provided by geophysics | 382 |
| Table B.1: Indicator coding of data [Alabert, 1987; Deutsch and Journel, 1992] | 434 |
LIST OF FIGURES

Figure 2.1: Schematic map view of example problem site .......................................................... 10

Figure 2.2: Schematic 3-dimensional view of example problem site ......................................... 11

Figure 2.3: Filtering of hypothetical porosity depth section due to the measurement and upscaling processes. Geometric mean boxcar filter is used for measurement filtering—labeled “# m res.” for # meters resolution, corresponding to filter width. Upscaling to 5 meter cell dimension grid performed on 1 meter resolution measurement ........................................... 17

Figure 2.4: Conceptualization of the decision-making process for hydrogeology-related engineering problems .... 27

Figure 2.5: Pyramid approach for defining decision objectives. The apex of the pyramid represents the primary objective, underlain by intermediate objectives that are to be met from bottom to top—culminating with the primary objective ................................................................. 28

Figure 3.1: Spatial averaging associated with a single measurement of a typical surface geophysical method .... 82

Figure 4.1: Hypothetical example of a pdf for hydraulic conductivity (K), on a log scale. The pdf shows that the probability of K being $10^5$ m/s is 0.18 ........................................................................................................ 112

Figure 4.2.2: Example of a hypothetical calibration scattergram. In this example there are two indicator classes. Each point represents a calibration pairing where both the primary and secondary variables were measured at the same location. A1 and A2 (B1 and B2) are the primary (secondary) indicator classes; the line separating the two is the quantity value demarcating the two classes. S1 though S4 are the sum of the calibration pairings within each sector of the scattergram ............................................................................. 140

Figure 4.2.3: Effect of spatial filtering on indicator kriging of hypothetical hard soil chemical concentration measurements. Probability of concentrations below the lower concentration threshold level, assuming the measurements are (a) point measurements and (b) measurements having a circular areal support volume with a radius of 5 units .............................................................. 158

Figure 4.2.4: Effect of spatial filtering on indicator kriging of hypothetical hard soil chemical concentration measurements. Probability of concentrations above the upper concentration threshold level, assuming the measurements are (a) point measurements and (b) measurements having a circular areal support volume with a radius of 5 units ........................................................................................................ 159

Figure 5.1: Example grid of probabilities of contaminant concentration greater than a regulatory threshold level ............................................................................................................. 169

Figure 5.2: Example optimized soil remediation design .................................................................................................................. 169

Figure 5.2.1: Schematic map of case study site .............................................................................. 175

Figure 5.4.1: Schematic case study remediation design ................................................................. 202

Figure 5.4.2a: 3-d cube of case study soil sample metals analysis results (top view) ................. 209

Figure 5.4.2b: 3-d cube of case study soil sample metals analysis results (side view) ............... 210

Figure 5.4.3a: 3-d cube of case study soil sample PAH analysis results (top view) ..................... 211

Figure 5.4.3b: 3-d cube of case study soil sample PAH analysis results (side view) ................. 212

Figure 5.4.4a: 3-d cube of case study soil sample MOG analysis results (top view) ................. 213

Figure 5.4.4b: 3-d cube of case study soil sample MOG analysis results (side view) ............... 214
LIST OF FIGURES

Figure 5.4.5a: Histogram of soil samples analyzed for metals ......................................................... 215
Figure 5.4.5b: Ogive of soil samples analyzed for metals ................................................................. 215
Figure 5.4.6a: Histogram of soil samples analyzed for PAH ............................................................... 216
Figure 5.4.6b: Ogive of soil samples analyzed for PAH ................................................................. 216
Figure 5.4.7a: Histogram of soil samples analyzed for MOG ............................................................ 216
Figure 5.4.7b: Ogive of soil samples analyzed for MOG ................................................................. 217
Figure 5.4.8a: Declustering analysis results for metals threshold level B ........................................ 220
Figure 5.4.8b: Declustering analysis results for metals threshold level C ........................................ 221
Figure 5.4.9a: Declustering analysis results for PAH threshold level B ........................................... 222
Figure 5.4.9b: Declustering analysis results for PAH threshold level C ........................................... 223
Figure 5.4.10: Declustering analysis results for MOG special waste threshold level ....................... 224
Figure 5.4.11a: Horizontal variography for metals level B threshold ............................................. 226
Figure 5.4.11b: Vertical variography for metals level B threshold .................................................... 226
Figure 5.4.11c: Variography for metals level C threshold ............................................................... 227
Figure 5.4.12a: Horizontal variography for PAH level B threshold .................................................. 227
Figure 5.4.12b: Vertical variography for PAH level B threshold ...................................................... 228
Figure 5.4.12c: Variography for PAH level C threshold ............................................................... 228
Figure 5.4.13: Variography for MOG special waste threshold ......................................................... 229
Figure 5.4.14: Calibration of PAH samples to metals – (a) cdf table, (b) misclassification probabilities 232
Figure 5.4.15: Calibration of MOG samples to metals – (a) cdf table, (b) misclassification probabilities 233
Figure 5.4.16: Calibration of metals samples to PAH – (a) cdf table, (b) misclassification probabilities 234
Figure 5.4.17: Calibration of MOG samples to PAH – (a) cdf table, (b) misclassification probabilities 235
Figure 5.4.18: Calibration of metals samples to PAH – (a) cdf table, (b) misclassification probabilities 236
Figure 5.4.18: Calibration of metals samples to PAH – (a) cdf table, (b) misclassification probabilities 237
Figure 5.4.19a: Probability of metals level B soil contamination—from kriging of all samples analyzed for metals. ................................................................. 240
Figure 5.4.19b: Probability of metals level C soil contamination—from kriging of all samples analyzed for metals. ................................................................. 240
Figure 5.4.20a: Probability of metals level B soil contamination—from cokriging of all samples .......... 241
Figure 5.4.20b: Probability of metals level C soil contamination—from cokriging of all samples .......... 241
LIST OF FIGURES

Figure 5.4.21a: Probability of metals level B soil contamination—from cokriging of all samples and soft type C prior probabilities. .................................................. 242

Figure 5.4.21b: Probability of metals level C soil contamination—from cokriging of all samples and soft type C prior probabilities. .................................................. 242

Figure 5.4.22a: Probability of PAH level B soil contamination—from kriging of all samples analyzed for PAH. .243

Figure 5.4.22b: Probability of PAH level C soil contamination—from kriging of all samples analyzed for PAH. .243

Figure 5.4.23a: Probability of PAH level B soil contamination—from cokriging of all samples. ......................... 244

Figure 5.4.23b: Probability of PAH level C soil contamination—from cokriging of all samples. ......................... 244

Figure 5.4.24a: Probability of PAH level B soil contamination—from cokriging of all samples and soft type C prior probabilities. ........................................................................ 245

Figure 5.4.24b: Probability of PAH level C soil contamination—from cokriging of all samples and soft type C prior probabilities. ........................................................................ 245

Figure 5.4.25: Probability of MOG special waste soil contamination—from kriging of all samples analyzed for MOG. ........................................................................ 246

Figure 5.4.26: Probability of MOG special waste soil contamination—from cokriging of all samples. ............... 246

Figure 5.4.27: Probability of MOG special waste soil contamination—from cokriging of all samples and soft type C prior probabilities. ......................................................... 247

Figure 5.4.28a: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—metals level B. ......................... 248

Figure 5.4.28b: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—metals level C. ......................... 249

Figure 5.4.29a: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—PAH level B. ............................ 249

Figure 5.4.29b: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—PAH level C. ............................ 250

Figure 5.4.30: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—MOG SW. ........................................ 250

Figure 5.4.31: Assigned prior probabilities for soft type C data scenario—metals ........................................... 251

Figure 5.4.32: Assigned prior probabilities for soft type C data scenario—PAH ............................................. 252

Figure 5.4.33: Assigned prior probabilities for soft type C data scenario—MOG ........................................... 252

Figure 5.4.34a: Differences greater than 0.1 between probabilities estimated from all sample data and soft C data versus only hard samples—metals level B. ................................ 253

Figure 5.4.34b: Differences greater than 0.02 between probabilities estimated from all sample data and soft C data versus only hard samples—metals level C. ................................ 254

Figure 5.4.35a: Differences greater than 0.1 between probabilities estimated from all sample data and soft C data versus only hard samples—PAH level B. ................................. 254
Figure 5.4.35b: Differences greater than 0.02 between probabilities estimated from all sample data and soft C data versus only hard samples—PAH level C

Figure 5.4.36: Differences greater than 0.02 between probabilities estimated from all sample data and soft C data versus only hard samples—MOG special waste

Figure 5.4.37: Variance of PAH level B probabilities estimated from kriging of soil sample analyzed for PAH

Figure 5.4.38: Variance of PAH level B probabilities estimated from cokriging of all soil sample data

Figure 5.4.39: Variance of PAH level B probabilities estimated from cokriging of all soil sample data and type C soft data prior probabilities

Figure 5.4.40a: Metals level B soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data

Figure 5.4.40b: Metals level C soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data

Figure 5.4.41a: PAH level B soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data

Figure 5.4.41b: PAH level C soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data

Figure 5.4.42: MOG special waste soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data

Figure 5.4.43: PAH level B soil contamination expected value of 25 realizations—simulated using all sample data and soft C prior probabilities as conditioning data

Figure 5.4.44: Venn diagram of hypothetical metals and MOG contaminant level sets

Figure 5.4.45: Optimized total cost of remediation for data scenario where only hard sample data is used in geostatistical estimation

Figure 5.4.46: Optimized total cost of remediation for data scenario where all sample data is used in geostatistical estimation

Figure 5.4.47: Optimized total cost of remediation for data scenario where all sample data and soft C prior probabilities are used in geostatistical estimation

Figure 5.4.48: Value of information (cost savings in remediation design) provided by using soft sample data instead of just hard sample data

Figure 5.4.49: Value of information (cost savings in remediation design) provided by using soft sample data and soft C prior probabilities instead of just hard sample data

Figure 5.4.50a: Optimized remediation design based on hard sample data only—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0

Figure 5.4.50b: Portion of site not requiring contamination treatment for remediation design in Figure 5.4.50a

Figure 5.4.50c: Optimized remediation design based on all sample data—for same scenario in Figure 5.4.50a

Figure 5.4.50d: Portion of site not requiring contamination treatment for remediation design in Figure 5.4.50c
LIST OF FIGURES

Figure 5.4.50e: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.50a. ................................................................. 275

Figure 5.4.50f: Portion of site soil volume not requiring treatment for contamination—remediation design in Figure 5.4.50e. ................................................................. 275

Figure 5.4.51a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is an alternative at $4 / m^3$, (2) ratio of future to present unit remediation cost is 2.0. ...................................................... 276

Figure 5.4.51b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.51a. ...................................................... 276

Figure 5.4.52a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is an alternative at $8 / m^3$, (2) ratio of future to present unit remediation cost is 2.0. ...................................................... 277

Figure 5.4.52b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.52a. ...................................................... 277

Figure 5.4.52c: Portion of site not designated for batch sampling—for remediation design in Figure 5.4.52a. ...................................................... 278

Figure 5.4.53a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.25. ...................................................... 278

Figure 5.4.53b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.53a. ...................................................... 279

Figure 5.4.54a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.125. ...................................................... 279

Figure 5.4.54b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.54a. ...................................................... 280

Figure 5.4.55: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is an alternative at $8 / m^3$, (2) ratio of future to present unit remediation cost is 3.0. ...................................................... 280

Figure 5.4.56: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.0. ...................................................... 281

Figure 5.4.57a: Expected value total remediation cost with additional hypothetical data of different types, including (1) EM geophysics of different data qualities for detecting metals level C, (2) 21 fully-sampled boreholes, and (3) perfect information—for scenario where batch sampling is not an alternative. ...................................................... 283

Figure 5.4.57b: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.57a. ...................................................... 284

Figure 5.4.57c: Expected value total remediation cost with additional hypothetical data—for scenario where $4/m^3$ batch sampling is an alternative and same data types as Figure 5.4.57a. ...................................................... 284

Figure 5.4.57d: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.57c. ...................................................... 285

Figure 5.4.58a: Expected value total remediation cost with additional hypothetical data of different types, including (1) EM geophysics of different data qualities for detecting MOG SW, (2) 21 fully-sampled boreholes, and (3) perfect information—for scenario where batch sampling is not an alternative. ...................................................... 285

Figure 5.4.58b: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.58a. ...................................................... 286
Figure 5.4.58c: Expected value total remediation cost with additional hypothetical data—for scenario where $4/m^3 batch sampling is an alternative and same data types as Figure 5.4.58a

Figure 5.4.58d: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.58c

Figure 5.4.59a: Remediation design based on scenario where there is perfect information about all contaminant levels—for a simulated realization conditioned on all sample data.

Figure 5.4.59b: Remediation design based on scenario where there is perfect information about all contaminant levels—for a simulated realization conditioned on all sample data and soft C prior probabilities.

Figure 5.4.60: Location of 21 hypothetical boreholes and sample depths used for data worth analysis of borehole sampling.

Figure 5.4.61a: Probability of metals level B soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.61b: Probability of metals level C soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.62a: Probability of PAH level B soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.62b: Probability of PAH level C soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.63: Probability of MOG special waste soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.64a: Optimized remediation design based on simulated samples in 21 boreholes and all existing sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0.

Figure 5.4.64b: Optimized remediation design based on simulated samples in 21 boreholes and all existing sample data—for scenario where (1) $4/m^3 batch sampling is an alternative, (2) ratio of future to present unit remediation cost is 2.0.

Figure 5.4.65: Calibration of hypothetical soft data to PAH levels using constructed data set in Table 5.4.6—(a) cdf table, (b) misclassification probabilities.

Figure 5.4.66: Same calibration as Figure 5.4.65, except normalized by the inferred PAH contaminant level expected values—(a) cdf table, (b) misclassification probabilities.

Figure 5.4.67a: Simulation of geophysics soft data outcome where the measurement can perfectly delineate PAH contamination above and below threshold level C, but provides no information about the level B threshold. One (red) corresponds to the geophysics measurement above the calibration soft cutoff, zero (green) below, -1 (blue) outside problem domain.

Figure 5.4.67b: Simulation of geophysics soft data outcome where the measurement can perfectly delineate PAH contamination above and below threshold level B, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a.

Figure 5.4.68a: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level C with 90% success, but provides no information about the level B threshold. Grid values defined same as in Figure 5.4.67a.
LIST OF FIGURES

Figure 5.4.68b: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level B with 90% success, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a. 301

Figure 5.4.69a: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level C with 80% success, but provides no information about the level B threshold. Grid values defined same as in Figure 5.4.67a. 301

Figure 5.4.69b: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level B with 80% success, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a. 302

Figure 5.4.70a: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level C with 60% success, but provides no information about the level B threshold. Grid values defined same as in Figure 5.4.67a. 302

Figure 5.4.70b: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level B with 60% success, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a. 303

Figure 5.4.71a: Probability of PAH level B soil contamination—estimated from cokriging of all sample data and a hypothetical geophysics realization, where the geophysics can perfectly delineate PAH contamination above and below threshold level C, but provides no information about the level B threshold. 304

Figure 5.4.71b: Probability of PAH level C soil contamination—estimated the same way as in Figure 5.4.71a. 305

Figure 5.4.72a: Probability of PAH level B soil contamination—estimated from cokriging of all sample data and a hypothetical geophysics realization, where the measurement can delineate PAH contamination above and below threshold level C with 90% success, but provides no information about the level B threshold. 305

Figure 5.4.72b: Probability of PAH level C soil contamination—estimated the same way as in Figure 5.4.72a. 306

Figure 5.4.73a: Probability of PAH level B soil contamination—estimated from cokriging of all sample data and a hypothetical geophysics realization, where the measurement can delineate PAH contamination above and below threshold level C with 70% success, but provides no information about the level B threshold. 306

Figure 5.4.73b: Probability of PAH level C soil contamination—estimated the same way as in Figure 5.4.73a. 307

Figure 5.4.74a: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics can perfectly delineate metals contamination above and below threshold level C, but provides no information about the level B threshold—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0. 308

Figure 5.4.74b: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics has the same applicability as in 5.4.74a—for scenario where (1) $4/m^3 batch sampling is an alternative, (2) ratio of future to present unit remediation cost is 2.0. 309

Figure 5.4.75a: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics can delineate metals contamination above and below threshold level C with 90% success, but provides no information about the level B threshold—for same scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0. 309

Figure 5.4.75b: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics can delineate metals contamination above and below threshold level C with 70% success, but provides no information about the level B threshold—for same scenario as in Figure 5.4.75a. 310
Figure 5.4.76a: Contour plot of FDEM horizontal dipole, NW-SE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.................................................................314

Figure 5.4.76b: Contour plot of FDEM horizontal dipole, SW-NE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.................................................................314

Figure 5.4.77a: Contour plot of FDEM vertical dipole, NW-SE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.................................................................315

Figure 5.4.77b: Contour plot of FDEM vertical dipole, SW-NE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.................................................................315

Figure 5.4.78a: Contour plot of difference between FDEM horizontal dipole, NW-SE and SW-NE tool-oriented conductivity.................................................................317

Figure 5.4.78b: Contour plot of difference between FDEM vertical dipole, NW-SE and SW-NE tool-oriented conductivity.................................................................317

Figure 5.4.78b: Contour plot of combined FDEM horizontal dipole conductivity grid—constructed by merging low-noise parts of NW-SE and SW-NE oriented grids. 318

Figure 5.4.79a: Processed GPR relative amplitude section from case study site—line 2 of 15 in production survey. .................................................................319

Figure 5.4.79b: Processed GPR relative amplitude section from case study site—line 9 of 15 in production survey. .................................................................320

Figure 5.4.79c: Processed GPR relative amplitude section from case study site—line 14 of 15 in production survey. .................................................................320

Figure 5.4.80: 3-d volumetric representation of processed GPR relative amplitude sections acquired at case study site.................................................................321

Figure 5.4.81a: 3-d, 200 MHz, GPR relative amplitude grid—from processed, transformed, and gridded case study production survey 2-d sections.................................................................322

Figure 5.4.81b: Coarser 3-d, 200 MHz, GPR relative amplitude grid—from processed, transformed, gridded and averaged case study production survey 2-d sections.................................................................322

Figure 5.4.82: Calibration scatterplots of collocated FDEM horizontal dipole conductivity grid values versus soil sample analysis results for: (a) metals contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = greater than level C), (b) MOG contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = between level C and SW, 4 = greater than special waste). Conductivity cutoff chosen for calibration is shown as green line. .................................................................324

Figure 5.4.83: Calibration scatterplots of collocated FDEM vertical dipole conductivity grid values versus soil sample analysis results for: (a) metals contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = greater than level C), (b) MOG contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = between level C and SW, 4 = greater than special waste). Conductivity cutoff chosen for calibration is shown as green line. .................................................................324

Figure 5.4.84: Ogive plots of calibration results for FDEM horizontal dipole conductivity—calibrated to metals sample data for delineating metals contamination above and below level B threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.................................................................329
Figure 5.4.85: Ogive plots of calibration results for FDEM vertical dipole conductivity—calibrated to metals sample data for delineating metals contamination above and below level B threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

Figure 5.4.86: Ogive plots of calibration results for FDEM horizontal dipole conductivity—calibrated to MOG sample data for delineating MOG contamination above and below special waste threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

Figure 5.4.87: Ogive plots of calibration results for FDEM vertical dipole conductivity—calibrated to MOG sample data for delineating MOG contamination above and below special waste threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

Figure 5.4.88: Calibration scatterplots of collocated GPR averaged relative amplitude grid values versus soil sample analysis results for: (a) metals contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = greater than level C), (b) MOG contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = between level C and SW, 4 = greater than special waste). GPR amplitude cutoff chosen for calibration is shown as green line. Cutoff is for absolute value of amplitude in (a).

Figure 5.4.89: Ogive plots of calibration results for GPR absolute value amplitudes—calibrated to metals sample data for delineating metals contamination above and below level B threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

Figure 5.4.90: Ogive plots of calibration results for GPR amplitudes—calibrated to MOG sample data for delineating MOG contamination above and below special waste threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

Figure 5.4.91a: Probability of metals level B soil contamination—estimated from cokriging of all sample data and FDEM horizontal and vertical dipole conductivity.

Figure 5.4.91b: Probability of metals level C soil contamination—estimated from cokriging of all sample data and FDEM horizontal and vertical dipole conductivity.

Figure 5.4.92: Probability of MOG special waste soil contamination—estimated from cokriging of all sample data and FDEM horizontal and vertical dipole conductivity.

Figure 5.4.93a: Probability of metals level B soil contamination—estimated from cokriging of all sample data and GPR absolute value amplitudes.

Figure 5.4.93b: Probability of metals level C soil contamination—estimated from cokriging of all sample data and GPR absolute value amplitudes.

Figure 5.4.94: Probability of MOG special waste soil contamination—estimated from cokriging of all sample data and GPR amplitudes.

Figure 5.4.95a: Probability of metals level B soil contamination—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
Figure 5.4.95b: Probability of metals level C soil contamination—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes...........................................................................344

Figure 5.4.96: Probability of MOG special waste soil contamination—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes...........................................................................344

Figure 5.4.97: Differences greater than 0.1 between metals level B probabilities estimated after and before addition of FDEM and GPR geophysics data to all existing sample data........................................................345

Figure 5.4.97b: Differences greater than 0.02 between metals level C probabilities estimated after and before addition of FDEM and GPR geophysics data to all existing sample data.................................345

Figure 5.4.98: Differences greater than 0.1 between MOG special waste probabilities estimated after and before addition of FDEM and GPR geophysics data to all existing sample data...........................................................................346

Figure 5.4.99a: Probability of metals level B soil contamination across geophysics survey region—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes........346

Figure 5.4.99b: Probability of metals level C soil contamination across geophysics survey region—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes........347

Figure 5.4.100: Probability of MOG special waste soil contamination across geophysics survey region—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes. 347

Figure 5.4.101a: Variance of metals level B soil contamination probabilities—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes. 348

Figure 5.4.101b: Variance of metals level C soil contamination probabilities—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes. 348

Figure 5.4.102: Variance of MOG special waste soil contamination probabilities—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes. 349

Figure 5.4.103a: Probability of contamination above any of the metals, PAH, or MOG regulatory threshold levels—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.................................................................349

Figure 5.4.103b: Probability of contamination above any of the metals, PAH, or MOG regulatory threshold levels—estimated from cokriging of all sample data (no geophysics).................................................................350

Figure 5.4.103a: Optimized remediation design based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0. .................................................................................................351

Figure 5.4.103b: Portion of site soil volume not requiring treatment for contamination—remediation design in Figure 5.4.103a...............................................................................................................................................352

Figure 5.4.104a: Optimized remediation design based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes—for scenario where (1) $4/m$ batch sampling is an alternative, (2) ratio of future to present unit remediation cost is 2.0. .................................................................................................352

Figure 5.4.104b: Optimized remediation design for same scenario as Figure 5.4.104a, except unit cost of batch sampling is $8/m$.................................................................353
LIST OF FIGURES

Figure 5.4.105: Optimized remediation design based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.125.

Figure 5.4.106a: Total cost of remediation for different batch sampling scenarios and across a range of future-to-present remediation cost ratios—based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.

Figure 5.4.106b: Economic value (cost savings in remediation design) of FDEM and GPR geophysics surveys—for same scenario and data types as in Figure 5.4.106a.

Figure 6.1.1: Schematic diagram of different approaches—statistical and physical model—for estimating a subsurface property of interest (hydrocarbon saturation) from indirect geophysics measurements (GPR) that account for uncertainty in the measurement and estimation.

Figure 6.1.2a: Flow diagram of adapted M-B geostatistics approach—see caption next page.

Figure 6.1.2b: Flow diagram of the M-B geostatistics approach adapted and modified for this work, illustrating the process for implementing the methodology—the different modules, required and optional steps, and their sequence. Continued from previous page.

Figure 6.1.3: Flow diagram of decision-making framework developed for the case study, illustrating the process for implementing the different modules of the risk-based decision model. The linkages with the M-B uncertainty model and overall site characterization process are highlighted.

Figure 6.1.4: Total optimized soil remediation design cost derived from decision model, for data scenario where all soil sample data—both direct and soft co-contaminant concentration measurements—are used in geostatistical estimation. The results for three batch sampling scenarios are plotted: (blue) batch sampling is not an alternative, (magenta) batch sampling is an available alternative at a unit cost $4 / cubic meter, and (cyan) batch sampling is an alternative at a unit cost $8 / cubic meter. In addition, the total cost of remediation for the scenario when there is no uncertainty about soil contamination, or Expected Value of Perfect Information ("EVPI") is plotted (red).

Figure A.1: Spatial averaging associated with a typical surface geophysical method.

Figure A.2: Relationship between property distribution and measurement spaces for the forward and inverse problems (after McLaughlin and Townley [1996]).

Figure A.3: Relationship between spatial resolution and measurement spacing in the global inverse problems.
ACKNOWLEDGEMENTS

I am very grateful to Dames and Moore for their approval and assistance in performing the case study; Klohn-Crippen Consultants, Ltd. for gratis use of their EM-31; and the UBC rock physics group for use of their GPR equipment.

Special thanks to all colleagues, friends, and family who provided the extra motivation and encouragement to complete this long overdue thesis. In particular, I am very obliged to my advisor and current and previous manager for their incredible understanding and extra incentives to finish.
CHAPTER 1
INTRODUCTION

The primary goal of this overall study is to assess how well geophysics, used for site characterization, can be rigorously incorporated in the decision-making process for hydrogeological problems. Geophysical measurements, with their large volume of investigation (spatial coverage of measurements), high sampling density, fast acquisition time, and non-invasive nature, potentially provide significant value to hydrogeological problems, but they are inherently indirect or "soft" measurements of the subsurface properties of importance. Thus, there is intrinsic uncertainty associated with the information they provide. This makes these measurements difficult to rigorously incorporate in a risk-based decision-making process and underlies the need for developing a methodology to effectively and efficiently do so. Indeed, this represents the impetus for this work—to develop a comprehensive, yet hopefully practical, decision-making framework that incorporates geophysics-type "soft" measurements into such a process, accounting for the uncertainty in the measurements, in addition to the natural spatial variability in subsurface property distributions.

It was considered essential as part of this development to apply the framework to a comprehensive, non-ideal case study involving a real world problem—where actual geophysics measurements are acquired, analyzed, and incorporated. Such a case study is considered to be, by far, the best way to assess and refine the integrity, applicability, and practicality of the approach. All too often new hydrogeological methodologies are assessed only by applying the methodology to a synthetic problem. While such an exercise is useful for initial algorithmic verification, it omits the all important step of making and assuring the approach is applicable and practical in a real world setting. The case study undertaken for this research is the assessment, acquisition, and incorporation of geophysical measurements into a risk-based decision-making framework for soil environmental remediation at a real estate development. This case study is the focus of the overall research study.
1.1 Motivation for research

Recently it has become increasingly important to formally account for uncertainty in hydrogeology-related problems in order to make more informed, robust decisions. These problems are inherently plagued by a high degree of uncertainty due to their strong dependence on hydrogeological conditions and processes that vary significantly in space and time, and at many different scales. Uncertainty in existing conditions entails even greater uncertainty in future predictions of how the hydrogeological system will behave and, thus, predictions of how proposed engineering systems influenced by hydrogeology will perform. Performance uncertainty translates into the risk of unintended outcomes.

Despite these complexities, there is an increasing need to efficiently allocate limited resources due to the, often, non-revenue generating nature of engineering projects dealing with hydrogeology and, in the environmental remediation arena, the large number of contaminated sites. Efficient allocation of resources requires a cost-effective engineering design process, which can only be accomplished if the many sources and forms of uncertainty are characterized and explicitly accounted for in the analysis.

In addition, decision-makers are increasingly being held accountable for the decisions they make in the, often, adversarial atmosphere surrounding hydrogeology-related problems. This adversarial environment exists because of the regulatory and social context of these problems and the multiple, differing stakeholders involved, as well as the disillusionment of the public about environmental restoration/protection projects. Suspicion accompanying these projects stems from their recent history of large expenditures and unrealized goals. This accountability necessitates documentation of the reasoning behind decisions made in the face of so much uncertainty. In this regard a careful elucidation of uncertainty and risk provides a very useful communication tool for rationalizing decisions.

Finally, progressively more environmental regulations are using health risk to receptors (humans or other forms of life) as an indicator for environmental compliance. In order to reliably assess health risk it is necessary to characterize and propagate uncertainty in any of the linked
Chapter 1. INTRODUCTION

chain of processes along the path—from contaminant release into the environment to contact with potential receptors—that ultimately / potentially influence health risk; this includes hydrogeological processes.

While characterizing uncertainty in hydrogeological engineering problems, and including it in a systematic risk-based decision-making process, helps to increase the reliability of decisions, the only way to make more cost-effective, but still robust, engineering decisions is to reduce the uncertainty in conditions that ultimately influence the decisions and their outcomes. In this respect, the site investigation process plays a very important role by characterizing critical site conditions and, thus, reducing uncertainty in the system. The ultimate goal of site investigations in a decision-making context is to reduce the risk, and associated costs, of unintended outcomes—engineering failure resulting from not accounting for all possibilities, or an overly conservative and, thus, unnecessarily costly decision.

Recently there has been a major interest and effort in using geophysics for site characterization of hydrogeology-related problems. Geophysics provides the advantages of high density, non-invasive, and quick measurements covering large areas, compared to invasive methods such as drilling and direct sampling of subsurface material. The downside of geophysics is that the measurement can have a significant amount of uncertainty associated with it. There has been little effort, however, to assess the overall economic value of geophysics to an engineering project from the perspective of the decision-maker. Does geophysics provide enough new information to a problem to be cost effective? In other words, is the risk reduction (due to the reduction in uncertainty) resulting from the new information outweigh the cost of performing geophysics? Is it best (most cost-effective) to collect geophysics, take borehole samples, or neither? These are the types of questions a decision-maker must confront. To date these issues and ways to answer them have not been addressed in a rigorous risk-based decision-making context.
Chapter 1. INTRODUCTION

1.2 Research goal

The overall objective of this research is to investigate the application of geophysics to hydrogeology-related problems from the perspective of the decision-maker. To accomplish this goal a general framework for incorporating geophysics in the decision-making process has to be developed, so that questions like those posed above can be properly answered. Of utmost importance is to apply the overall process to a real-world case study where a specific hydrogeological problem is addressed and actual geophysical field measurements are collected and incorporated in the decision process. The applicability of the method cannot be effectively demonstrated without carrying out a real-world example due to the unpredictable reliability, site-specific nature, and inherent uncertainties associated with geophysics and other aspects of hydrogeology-related problems. Also, the particular approach is largely untested and, therefore, by taking a deductive approach, the case study can be used as a learning tool to identify problems with the framework and make inferences about what aspects can be applied to hydrogeology-related problems in general. It should be emphasized that the purpose of the case study is to illustrate the methodology; the actual success of geophysics in providing value to the problem is not necessary for meeting this objective.

1.3 Research contributions

The contributions of this research can be grouped into two broad categories—quantitative and qualitative contributions. The quantitative contribution is the development of a rigorous, robust, and generalized methodology for incorporating geophysical measurement data, along with other types of precise and imprecise information, into a risk-based decision-making framework. This methodology accounts for, quantifies, and propagates uncertainty, including measurement uncertainty. Within this framework, a decision-maker can estimate the required accuracy level for geophysics to be worthwhile, compare the economic worth of different
measurement types and strategies, and determine the most cost-effective engineering design based on all existing information.

The qualitative contribution is a recommended approach for applying geophysics to hydrogeology-related problems in a decision-making context—a sequential, iterative, and systematic procedure for determining if, when, why, and how geophysics should be applied for site characterization. Largely inspired and refined as a result of lessons learned from the case study, these general guidelines address the realistic limitations of geophysics. They focus on determining if geophysics has the possibility of providing useful information for a particular problem. If so, which geophysical technique(s) should be used? At what stage of the project should it be performed? How should it be run? (e.g., What type and amount of ground truthing is required?; Should complementary methods be run?; How should the results be interpreted?) Will geophysics be economically viable and worthwhile? How should the results be incorporated in the decision-making process? The recommended approach does not necessarily provide specific answers to these questions for a particular problem, but provides a general, systematic, and realistic way to approach the use of geophysics in hydrogeology-related problems from a risk-based decision-making perspective.
Chapter 1. INTRODUCTION

1.4 Thesis overview

The guiding principle of this research was to develop and test a practical and useable site characterization and decision making approach—applicable to real-world hydrogeology-related problems. This entails that the breadth of the research and the attendant real world case study is quite comprehensive, covering a broad spectrum of disciplines and topics—from geophysics to geostatistics and decision analysis. Consequently, in addition to recording all the results, this thesis attempts to address and explain all the major areas that have important bearing on the work; this has resulted in a very long write-up. The purpose of this section is to provide a “road map” to the thesis, so that readers of different backgrounds and interests can read what they want.

CHAPTER 1: INTRODUCTION (this chapter) summarizes what the thesis is about, what motivated the research, and what the goals are. This chapter should be read by all.

CHAPTER 2: BACKGROUND provides an introduction to and explanation of the important underlying concepts of this work:

1. *Uncertainty and how it affects hydrogeology-related problems* (section 2.2). This section contains a fairly lengthy and in-depth discussion on uncertainty and its relationship to scale, that is not directly applicable to the case study and the results of this research (section 2.2.1). However, the introduction of this section and the discussion on measurement uncertainty (section 2.2.2) should be read.

2. *Decision making, how it is impacted by risk, and how it is accomplished in a formal framework—all in the context of hydrogeology-related problems* (section 2.3). This entire section is important to read if the reader is unfamiliar with the risk-based decision making process and the concepts behind it. Section 2.3.3, a discussion comparing economic risk and health risk, is not needed to understand the results of this work, but is applicable to the philosophical foundation of the case study problem.

3. *The site characterization process, how it can be formally organized, and how it relates to decision making* (section 2.4). This section delves quite heavily into site characterization
theory, basically summarizing, and putting in the context of this research, the work of Baecher [1972]. The introduction to this section, subsection 2.4.1 outlining the objectives of site characterization, and the introduction and summary to section 2.4.2, discussing the organization of site characterization, should probably be read.

In addition, the hypothetical example hydrogeology problem, that is made reference to throughout the thesis to explain different concepts, is introduced (section 2.1) and a review of previous research that is closely related to the subject matter is provided—organized by topic (section 2.5).

CHAPTER 3: GEOPHYSICS IN HYDROGEOLOGY PROBLEMS provides an introduction to geophysics and its application to site characterization in hydrogeology-related problems. The chapter is basically organized to start simple and general, then to become progressively more in-depth and specific to particular aspects related to this work. Section 3.1 is a short overview of geophysics for readers unfamiliar with this type of measurement. Section 3.2 further discusses applications of geophysics in hydrogeology-related problems and section 3.3 outlines the advantages that geophysics provides such problems, using hypothetical examples for illustration. Section 3.4 provides a long discussion on the difficulties associated with applying geophysics to the types of problems of interest and how these can be addressed. It is recommended to read at least the introduction to this section, the introductions to subsections 3.4.1 and 3.4.2, and the summary (subsection 3.4.4); they describe the need and impetus for the uncertainty framework used in this work. The internal subsections of 3.4.1 and 3.4.2 discuss in considerable detail the many different techniques used for dealing with the indirect, non-unique nature of geophysics measurements, including the geostatistical technique used in this work; this is not necessary for understanding the results of this work, since the specific technique used is discussed in detail in CHAPTER 4.

CHAPTER 4: ACCOUNTING FOR AND MODELING UNCERTAINTY provides a comprehensive summary of the different methods of accounting for uncertainty in spatial variables, as well as a detailed discussion of the geostatistics methodology applied in this
work—from a practitioner's perspective. Section 4.1 introduces the concept of uncertainty and how it can be represented and dealt with—recommended for readers unfamiliar with probability theory concepts. Section 4.2 becomes more detailed about the different methods for forming and updating spatial random variables to represent empirical quantity uncertainty. While this section isn't necessary for understanding the results of this work, it does conceptually discuss the general geostatistics approaches used—geostatistics is introduced in section 4.2.3, including the concept behind, and advantages of, the kriging and simulation methods used. Section 4.3 details the workings of the specific geostatistics approach used in this work—the Markov-Bayes indicator random variable approach—and should be read to understand the results.

CHAPTER 5: CASE STUDY describes the soil contamination remediation case study performed in this work—the background of the case study problem, the different tasks undertaken as part of it, and all the representative results of all the key analyses performed. It is the core of this thesis. While undoubtedly a lot of information due to the broad scope of the case study, it should be read in its entirety to understand the full scope of the work accomplished; but, if only if only certain aspects of the work are of interest, the background and the section(s) covering these aspects can be read.

CHAPTER 6: CONCLUSIONS AND RECOMMENDATIONS provides a comprehensive summary of the (1) key findings and contributions of this work (sections 6.1.1 and 6.1.2), including flow diagrams of the uncertainty and decision models that were developed, and (2) the results of the case study (section 6.1.3). In addition, recommendations for further research that builds on this research are presented (section 6.2).

APPENDICES A and B provide detailed, mathematical discussions on the geophysical inverse problem and the Markov-Bayes indicator geostatistics estimation methodology, respectively.
CHAPTER 2
BACKGROUND

This chapter provides a background to the important concepts underlying this research. These concepts include:

- the nature and types of uncertainty inherent to hydrogeology problems,
- the hydrogeological decision-making process and the concept of risk, and
- the site characterization process.

In addition, a comprehensive review of important, closely related previous work is presented at the end of the chapter. To help clarify the explanation of these concepts, and others throughout the thesis, an example hydrogeology problem—akin to the case study problem, but more general—is introduced and referred to repeatedly in this chapter, and throughout the document.

2.1 Example hydrogeology problem

In order to illustrate and provide real-world applicability to the concepts presented throughout this work, a generic, but realistic, example hydrogeology problem is presented. The hypothetical example is made to closely resemble the general components of the case study problem, thus providing a smooth transition into the case study section of the thesis.

The example problem is a waterfront real estate development on a site which has a long history of industrial activity, including operating as a major railyard, fuel storage depot, and commercial ferry terminal. A schematic map view of the site is shown in Figure 2.1. The site contained a nest of underground and above ground fuel storage tanks of different sizes connected by a network of pipelines leading to pump houses and dispensing stations. All of the fuel tanks and most of the pipelines have since been removed. The fuel depot serviced a large railyard with multiple tracks built on cinder and ballast roadbed, a trailer truck yard, and a commercial ferry terminal—all of which are being decommissioned. The site borders a tidal marine inlet. One of
the main concerns facing the developer is the potential for subsurface chemical contamination, resulting from the extensive industrial activity.

![Figure 2.1: Schematic map view of example problem site.](image)

### 2.1.1 Geology of site

The hypothetical site is largely built on manmade fill (up to ten meters thick) overlying natural sediments, and has a generally flat surface topography just above mean sea level (MSL). A schematic three dimensional view of the site is shown in Figure 2.2.

The stratigraphy consists of:

- one to two meters of heterogeneous, mostly gravel-sized surface fill underlain by
- four to eight meters of mostly sandy intermediate fill, interspersed with areas of cobbles, concrete refuse, and wood fragments, which, in turn, is underlain by
- native clay-rich till with interbedded marine sand lenses.

The thickness of the fill layers generally decrease inland from the shoreline and eventually pinch out at the base of a small five meter high bluff and manmade retaining wall.
2.1.2 Hydrogeology of site

The hydrogeology of the hypothetical site is characterized by an unconfined aquifer in the high permeability fill material, underlain by a leaky aquitard associated with the till sediments, that is interspersed with higher permeability zones associated with the sand lenses. The size and interconnectedness of the higher permeability zones in the till is highly variable. The water table averages three meters below the surface. The hydrogeology of the site is greatly complicated by a strong tidal influence which causes large fluctuations in the water table and groundwater flow, in both the magnitude and direction of flow. The area receives a large amount of rainfall, especially in the winter, which acts as a large source of recharge.

2.1.3 Contamination of site and the resulting implications

As mentioned earlier, the hypothetical site has a long history of industrial activity which has lead to subsurface contamination. As a result of leaking fuel tanks, spills, and the release of used mechanical fluid and burned coal (or cinder) there are many scattered ancestral source zones of hydrocarbon and metal contaminants, both above and below the water table. The
contaminants in these source zones can migrate in their initial form and/or migrate with groundwater in other forms. The different types of contamination are schematically depicted in Figure 2.2, where the black plume represents free-phase contaminant and the hatched gray plume aqueous-phase contaminant coming from the free phase. In order to build on the site the developer must meet regulatory requirements for the intended land/water use to insure that the health of future inhabitants is not jeopardized. Thus, the hydrogeology engineering problem at this hypothetical site is to determine:

1. if and where contamination occurs,
2. the future fate and detrimental effects of any contamination, and
3. if, when, and how the contamination should be remediated or contained, if it indeed exists.

2.2 The nature of uncertainty in hydrogeology problems

Due to the complexity of the hydrogeological environment, uncertainty is pervasive and multi-faceted in problems which are influenced by this environment. It is pervasive in that uncertainty exists in one form or another in every real-world hydrogeology problem and it is multi-faceted in that it comes in many different, often overlapping, forms. There are numerous ways to characterize and classify uncertainty. From a practitioner's standpoint, the semantics of how uncertainty is classified is not particularly important, as long as it is recognized. What is important is to identify and understand:

- the different manifestations of uncertainty plaguing a particular problem,
- how these manifestations affect the practitioner's comprehension of the problem, and
- which types of uncertainty can potentially be reduced and which are systemic to the problem.

Once these aspects of the uncertainty are recognized and understood, a robust approach for coping with it in the engineering problem can be developed. A discussion of the types and characteristics of uncertainty in hydrogeology problems is discussed in this section, while approaches for dealing with it are discussed in the following sections.
2.2.1 The relationship between uncertainty and scale

Uncertainty and scale are inextricably linked. Consequently, to properly understand uncertainty, it is crucial to analyze it from the perspective of scale. In an ideal world, reality would be perfectly known at all scales and, thus, uncertainty would not exist. For the hypothetical problem described earlier, this scenario would entail a perfect understanding down to the smallest scale (i.e. the molecular level) of the current partitioning and spatial distribution of contaminants, as well as the spatial distribution of factors controlling groundwater transport. In addition, it would entail a perfect understanding of all the processes controlling mass transport in the subsurface (e.g. multiphase flow, tidal forcing of groundwater, contaminant mass partitioning) at all scales so that the future distribution of contaminant mass across the site could be perfectly predicted. In fact, with such a scenario, the common measurable parameters used to model such processes (e.g. hydraulic conductivity, porosity, dispersivity) would no longer be applicable since they inherently represent averaging across smaller scales of variability. Of course, this extent of knowledge is never achievable. There is a technical and practical limit to the level of detail that can be accounted for in the real-world.

2.2.1.1 Measurement scale

The lower limit of detail (scale of variability) that can be resolved in the real-world is inherently controlled by the measurement process. All measurements inevitably act as filters where the measured parameter represents an average of some variable(s) across the scale of the measurement [Beckie, 1996]. For example, the hydraulic conductivity ($K$) measured from a well core sample using a permeameter test represents an equivalent parameter which accounts for the drag caused by the substrate pore structure on fluid flow, averaged across the volume of the core. This argument applies to measurements of processes which vary in space, as well as time. As a result, the measured parameter will not vary significantly at scales less than the measurement scale. This implies that the measurement process smoothes out small-scale variability. Thus,
there is an intrinsic lower limit to the scale of variability that can be resolved, and this limit is
determined by the smallest measurement scale.

It is important to realize the difference between measurement scale and measurement
volume. The former is defined by the characteristic measurement filter width, itself related to the
spatial frequency spectrum for the measurement (the spectral description of scale is discussed in
the forthcoming Section 2.2.1.5). The measurement volume is the total volume of material
sampled by the measurement. For example, consider a groundwater tracer test performed by
injecting a chemical tracer in one well and measuring the concentration of the tracer chemical
versus time in another well 30 meters away—"downstream" along the mean flow path. The
measurement volume is something of an ellipsoid, with the long axis being the 30 meters along
the flow path between the wells and the shorter axes perpendicular to the flow path in each
direction, with lengths less than 30 meters. The measurement scale, however, could be
considerably smaller if the information from the concentration versus time is used to delineate
heterogeneity in the hydraulic conductivity field between the wells. As illustrated by this
example, the measurement volume and scale are often difficult to quantify.

Strictly, there is no such thing as a point measurement since, in reality, there is an infinite
range of scales of variability, and any measurement inevitably averages out the smaller scales.
The implication of this fact on uncertainty is that there is always some degree of intrinsic
uncertainty—which cannot be reduced—associated with the scales of variability in a variable
that cannot be resolved by measurements. These unresolved scales of variability and their
associated intrinsic uncertainty will be referred to as natural variability and uncertainty due to
natural variability, respectively.

2.2.1.2 Model scale

The model used to represent a system also has an inherent limit on the minimum scale of
variability that can be represented. Ideally, in order to most accurately model the system, the
smallest resolved model scale (e.g. model grid cell size) could be set to the smallest
measurement scale, since this represents the greatest level of detail about a parameter that can be
ascertained through measurements. However, to do so is often technically infeasible or impractical due to our inability to accurately model the small-scale physics and/or computational limitations (i.e. excessive computer effort). For example, to numerically model groundwater flow across a several acre site at the scale of well cores would require millions (if not billions) of grid discretizations. In addition, to accurately characterize the site, the required number of cores would be enormous. Inevitably, simplifications must be made, such as increasing the model scale (e.g. grid cell size), in order to make the problem tractable. Unfortunately, increasing the model scale causes an equivalent increase in the lowest possible scale that can be resolved by the model (the actual resolved scale is dependent on measurement spacing, as discussed in forthcoming Section 2.2.1.4) and, thus, an increase in the level of uncertainty that cannot be reduced (uncertainty due to natural variability).

Conversely, another model scale constraint could be that the model has to accurately represent the engineering design required—setting a minimum level of discretization. For example, if a one meter thick groundwater flow barrier wall is being considered for containing a plume of contaminated groundwater, to accurately model the wall the flow model scale would probably have to be on the order of a meter in and around the modeled wall. If the smallest measurement scale is larger than this, then, even at measurement locations, there would be uncertainty associated with spatial averaging at a scale larger than the model scale.

The larger the model scale the smaller the amount of system behavior explicitly described and resolved by the model and the greater proportion of true behavior which has to be accommodated by the model parameters through some sort of averaging [Beckie, 1996]. For example, as the size of the grid blocks in a groundwater flow model are increased, the scale of the parameter defining hydraulic conductivity ($K$) must also be equally increased, requiring that the $K$ parameter account for more of the physics controlling flow, across a greater volume of, possibly heterogeneous, subsurface material. This increased burden and amount of spatial averaging results in the inability to resolve small-scale features, such as high $K$ channels, which, in turn, means the true physics of flow through these features can no longer be explicitly
modeled. This loss of resolution could entail that the time it takes contaminants in groundwater to travel through the substrate is underestimated by the model due to not accounting for the high $K$ channel. However, for some problems small-scale groundwater flow behavior may be irrelevant, such as regional groundwater studies where only large-scale groundwater flow is of concern; the process model scale can be on the order of kilometers, and still provide the required information.

Model parameters such as $K$ are called phenomenological parameters and are used in conjunction with closure models in an attempt to represent the effect of unresolved physics on resolved physics in the system model [Beckie, et al., 1994]. For example, in fluid transport problems, the effect of unresolved fluid motions upon contaminant movement is often represented using dispersion parameters in a Fickian closure model, that models the unresolved behavior as random motion. The dispersion parameters essentially represent the parameters of a simple probabilistic model describing the random small-scale fluid motion.

### 2.2.1.3 The relationship between model and measurement scales

If the model scale is larger than the measurement scale, then an additional averaging (or filtering) process has to be performed to upscale measurements to the scale of the model. If the parameter being upscaled is process dependent, then the upscaling operation will be context dependent, possibly making the operation complex and imprecise, even with full measurement coverage. Therefore, upscaling represents another potential source of uncertainty that cannot be reduced, even at the model scale.
Chapter 2. BACKGROUND

Figure 2.3: Filtering of hypothetical porosity depth section due to the measurement and upscaling processes. Geometric mean boxcar filter is used for measurement filtering—labeled "# m res." for # meters resolution, corresponding to filter width. Upscaling to 5 meter cell dimension grid performed on 1 meter resolution measurement.

Figure 2.3 illustrates a hypothetical example of the smoothing effects of the measuring process at different measurement scales, and the upscaling process from the measurement scale to a larger model scale, for a porosity 1-d depth section. For this example, the spatial filter associated with these processes is simply an equally-weighted geometric average (a "box car" filter), applied at every 0.1 meter increment, and the parameter being measured or modeled is assumed to be process independent (e.g. porosity). The true variable distribution is highly discontinuous, analogous to a layered stratigraphy where there are large contrasts between the layers and the layer thickness varies significantly. Although small-scale features are evident in the filtered responses to the heterogeneous variable in Figure 2.3, the primary variability is at scales equal to or larger than, approximately, the scale of averaging. The small-scale features and the sharpness of the changes in variability of the averaged parameters is largely an artifact of the sharp boundaries of the filters. The example five meter model scale illustrated in the figure is
generated by averaging the measurement values within each five meter grid cell and assigning that constant value to the cell assuming the measurement scale is 1 meter and the measurements are taken every 0.1 meter. The grid upscaling is seen to average the porosity measurements and the structured discretization masks the true location of bed boundaries.

Usually the fact that variability in parameters, and the associated uncertainty, at scales below the model scale can never be resolved by the model is considered a basic assumption of the model. In other words, it should be understood that a model only models processes at the model scale or larger. The utility of phenomenological parameters and closure models is only an attempt to account for the effect of unresolved behavior on the modeled resolved behavior. In this sense, the variability at scales below the model scale (which can be called natural variability) and the associated uncertainty can be ignored, except in the way it influences resolved scales, since it is understood the model does not represent those scales of the system. For example, a groundwater flow model with grid blocks on the scale of tens of meters cannot model small-scale flow at the size of a well core. If the details of flow at the scale of cores is considered crucial to a particular problem, then the only recourse is to reduce the model scale.

### 2.2.1.4 Network scale

Unfortunately, just because a model has the ability to resolve variability in the behavior of a system at scales equal to or greater than the model scale does not ensure that it will accurately do so. How well the model represents these scales of variability in system behavior depends on how well the variability—in the model parameters corresponding to the same scales—is characterized through measurements. The only way to fully characterize the parameter variability down to the scale of the model is to have full measurement coverage at the spacing of the measurement scale (assuming there is no uncertainty in the measurements and, if upscaling is required, the upscaling process is perfect) [Beckie, 1996]. Beckie [1996] refers to the characteristic spacing between measurements as the network scale, and the inability of the model to accurately resolve scales between the network and model scales as a model closure problem. Therefore, in addition to the uncertainty due to natural variability, there is also uncertainty about
the variability of parameters at scales between the network and model scales, that propagates to uncertainty in the modeled system behavior at those scales. However, unlike the uncertainty due to natural variability, this type of uncertainty can be reduced and, thus, can be referred to as uncertainty due to ignorance.

2.2.1.5 Spectral analysis of scale and uncertainty

Beckie [1996] uses a spectral approach to clearly show the separation between unresolved scales of variability (where there is uncertainty due to natural variability), subgrid scales of variability (where there is uncertainty due to ignorance), and resolved scales (where there is no uncertainty). Spectral techniques transform functions from the space or time domain (e.g. the spatial distribution of \( K \)) to the spatial or temporal frequency domain where the functions plot as an energy spectrum (spectral energy versus frequency of variation). These techniques are valuable for analyzing scale issues because the energy (which is related to the amount of variability) associated with different frequencies or ranges of frequencies (scales) can be directly identified. Nyquist sampling theorems prove that the Nyquist frequency \( f_N = \frac{1}{2A_N} \) is the greatest degree of variability that can be fully described by the Nyquist sampling interval \( \Delta_N \).

Thus, in the frequency-domain, the boundaries between the different scales can be conveniently defined by the appropriate Nyquist frequencies:

- for the boundary between subgrid and resolved scales — \( f_{US} = \frac{1}{2\lambda_M} \), where \( \lambda_M \) is the model scale (e.g. grid block size) or the smallest measurement scale (e.g. the radius of influence of a slug test), whichever is larger.

- for the boundary between unresolved and subgrid scales — \( f_{SR} = \frac{1}{2\lambda_N} \), where \( \lambda_N \) is the network scale (characteristic spacing between the center point of measurements).

These definitions imply that the smallest scale of variability that can possibly be resolved by a model is \( \lambda_M \) and the smallest scale which is actually resolved is \( \lambda_N \).

Beckie [1996] illustrates how the size of the subgrid scale variability section of the frequency spectrum (that represents unresolved variability) can be reduced by taking measurements at a denser spacing across the problem domain, which reduces the size of the
network scale. In essence, this process corresponds to reducing uncertainty due to ignorance. To completely eliminate the closure problem, measurements have to be taken at a spacing equal to the Nyquist interval that corresponds to the unresolved scale - subgrid scale boundary—the larger of the model scale or smallest measurement scale. Taking measurements at closer intervals than this would provide no new information to the model (assuming the measurements are error-free and at a scale equal to or larger than the model scale, i.e. no upscaling is required).

Beckie [1996] also shows that the same reduction of the subgrid scale variability component of the spectrum can be accomplished by increasing the measurement scale (i.e. use a measurement device that measures more of the subsurface), while concurrently increasing the model scale to match the measurement scale. However, it is important to realize that increasing the measurement/model scale will not increase the resolution of the model or reduce the overall uncertainty, since it only moves the spectral boundary distinguishing subgrid and unresolved scales, both of which are unresolved and plagued by uncertainty. It will reduce the model closure problem, though, resulting in a more accurate model representation, albeit the model representation is at a larger scale.

2.2.1.6 Summary

Uncertainty is best understood and characterized from the perspective of scale. Three important scales, common to most problems, are measurement, model, and network scales. Measurement scale corresponds to the smallest scale at which a measuring device measures—also commonly referred to as the measurement resolution. Model scale is the smallest scale of variability in a system that is represented by the model. Network scale characterizes the spacing of measurements across the problem domain. The relationship between these scales determines the true resolving power of a model, the potential resolving power of a model, and the scales of natural variability in a system that can never be resolved by a model. Spectral theory shows that the smallest scale which can be resolved by a model is the characteristic network scale and the best resolution a model can ever achieve, when the problem domain is
Chapter 2. BACKGROUND

fully characterized, is the measurement or model scale, whichever is larger. The implications of these resolution limits on model uncertainty are:

- uncertainty due to natural variability is intrinsic in a model and cannot be reduced,
- uncertainty associated with scales of variability between the model/measurement scale and the network scale can be reduced by taking additional measurements.

The second type of uncertainty can be considered uncertainty due to ignorance since it results from not having enough information—information that is attainable.

There are other types of uncertainty that can afflict system models, including uncertainty resulting from inaccuracies in the model representation of physical processes or uncertainty due to measurement errors. The latter type will be discussed in the following section. A final point is that the uncertainty issues covered in this section apply to the modeling of all types of systems, including economic systems, where model resolution, for example, could correspond to resolving the temporal variability in the unit cost of pumping a well.

2.2.2 Uncertainty associated with the measurement process

The measurement process involves performing a field test (or performing a lab test on a field sample) in order to observe a model parameter, the objective being to reduce uncertainty about the nature of the true model parameter distribution across the site. As shown in the previous section, every measurement inherently acts as a filter, averaging across a certain measurement scale and, thus, filtering out smaller-scale variability. Consequently, every measurement has a certain degree of uncertainty (due to natural variability) associated with it as a result of a lower limit on resolution.

In addition, some measurements do not directly observe model parameters, but instead observe state variables which must be converted to model parameters through some sort of measurement model, leading to the potential for more uncertainty. In particular, this is true when measuring process-dependent phenomenological parameters, such as $K$. For example, the core permeameter test for measuring $K$ involves the observation of the state variables hydraulic head
and specific flux, from which the model parameter $K$ must be inverted by applying Darcy's equation to a one-dimensional, steady-state flow measurement model. The measurement of $K$, in the same material from which the core was extracted, using a slug test could yield different results solely based on the fact that the slug test model assumes radial instead of linear flow (if $K$ in that material is dependent on the flow context, i.e. $K$ is anisotropic). In addition, the results could differ due to the different measurement scales sampled by the two techniques. Measurements such as these represent indirect measurements since they do not directly measure the parameter of interest, the model parameter, but instead measure instrument-dependent state variables, which must somehow be related to the model parameter.

Indirect measurements can contain multiple tiers of indirectness, and complex relationships between the instrument state variables and model parameter. This can especially be true for geophysics, where, not only do the geophysical parameters need to be inverted from the measured instrument-dependent state variable(s) based on a measurement model, but the model parameter of interest must then be somehow related to the geophysical parameters. The use of direct current (DC) vertical electrical surveys (VES) to estimate subsurface saturated porosity will be used as an example to better illustrate this complicated linkage.

The DC VES technique measures the electrical resistivity ($\rho$) of the subsurface by sending a DC electrical current through a region of the subsurface between two electrodes and measuring the resulting electrical potential (voltage) in the subsurface between two other electrodes. A so-called apparent electrical resistivity ($\rho_a$) can be calculated directly from the measured voltage and electrode spacing. Each $\rho_a$ represents some sort of average resistivity across some volume of the subsurface. The electrode spacing can be varied to obtain different depths of investigation. The geophysical information usually desired from a VES is the geoelectrical structure—the spatial distribution of $\rho$ in the subsurface.

Calculation of the geoelectrical structure beneath the survey profile requires simultaneous inversion of the entire set of $\rho_a$ measurements for different electrode spacing using a specified measurement model. The most common measurement models used for electrical profiling are
one dimensional, varying only in the depth direction (i.e. vertically-layered stratigraphy). Even using this simple model, the inversion of VES measurements is non-unique, meaning that many different geoelectrical section solutions are possible from the same data set. The solution that is selected depends on the inversion approach and user judgment. Thus, the calculated geoelectrical section has a significant degree of uncertainty associated with it as a result of the inversion process; this uncertainty can be further exacerbated if there are errors in the $\rho_s$ measurements and/or the true geoelectrical section varies significantly from the one dimensional assumption.

At this stage, the geoelectrical section still has to be converted to a saturated porosity section. This process requires determining the relationship between $\rho$ and porosity, while accounting for the fact that many properties other than porosity affect resistivity (e.g. clay content, pore water resistivity variations). Due to the dependency of $\rho$ on many factors which are usually not well characterized (or not characterized at all), the relationship between porosity and $\rho$ is non-unique, resulting in additional uncertainty in the final porosity estimates. Thus, in the process of calculating porosity from VES measurements, uncertainty has entered into the estimate through:

1. errors in the VES measurements,
2. the scale and assumptions made in the VES measurement model,
3. the non-unique inversion of VES measurements to obtain a geoelectrical section, and
4. the non-unique relationship between $\rho$ and porosity.

As illustrated by the previous geophysics example, many types of error can plague measurements, leading to compounded uncertainty in the estimate of model parameters. The form that these errors take can be classified as either coherent/unbiased or incoherent/biased. Coherency in errors, also called imprecision and non-repeatability, is characterized by variability in the results of measurements taken at the same location (data scatter). The mean of these results eventually converges on the true value after a large number of measurements are taken. This type of data scatter can be considered random error. Incoherence in errors, also called inaccuracy, is
Chapter 2. BACKGROUND

characterized by data scatter as well, but the mean of the data converges on a value which is not the true value. Coherent measurement errors are much easier to cope with than incoherent measurement errors because the associated uncertainty in the model parameter can be accounted for using probabilistic methods. This can not be accomplished with incoherent errors unless the bias (displacement) in the mean from the true value is somehow known. Thus is the importance in distinguishing between these two forms of errors. Incorporating this distinction, measurement errors can be classified based on their source as follows:

- Coherent, random noise due to interference of the measurement (e.g. 60 Hertz noise in a seismic section due to nearby power lines affecting the recording instrument)
- Incoherent, biased noise due to interference of the measurement (e.g. multi-frequency noise in a seismic section due to trucks on a nearby road)
- Measurement scale not equal to model scale (e.g. $K$ measured from a pump test / core samples when interested in $K$ on the order of meters)—can be considered an unbiased error if the downscaling / upscaling approach is unbiased
- Indirect measurement with imperfect correlation between measurement and model parameter uncertainty (e.g. $K$ estimated from seismic velocity)—bias depends on approach for handling uncertainty in the correlation
- Non-unique parameter estimate resulting from inversion of measured state variable(s) (e.g. $K$ parameter field estimated from hydraulic head measurements in five wells)—bias depends on approach for handling uncertainty in the inversion

All these types of measurement error—and many can contaminate a single measurement (as shown in the previous geophysics example)—will result in errors in the estimate of the model parameter, unless the error is accounted for through uncertainty. The uncertainty associated with these error types can be considered uncertainty due to ignorance, since it can be reduced or eliminated by taking measurements with less error. From the perspective of scale, imperfect measurements can result in a loss of resolution of the model parameter variability at some scales.
compared to perfect measurements. Whether a particular scale of variability larger than the network scale can be resolved using imperfect measurements is difficult to assess, but depends on the relationship between the degree of uncertainty in the estimated model parameter and the amount of variability at that scale.

Measurement data is often categorized as either hard or soft data—hard referring to "exact" and soft to "inexact". The categorization can be very important since the two types of data are often treated very differently in a problem. However, this distinction is ambiguous since the terms "exact" and "inexact" are subjective. Every measurement contains some degree of error. Thus, the distinction between hard and soft data should be based on a defined error threshold; measurement errors of a magnitude below the threshold are considered acceptable and errors of a magnitude above unacceptable. Acceptable and unacceptable errors should be defined based on their potential to corrupt the model results to a degree that alters the conclusions made based on them. Performing a model sensitivity analysis can help in choosing the threshold. Although the choice of error threshold will invariably be somewhat subjective, having a strict, quantitative criterion to distinguish between hard and soft data is better than having none at all.

A type of soft measurement that is in a class by itself is engineering or geologic judgment. This type of measurement involves an expert making a subjective, semi-quantitative assessment about a problem parameter, usually a large-scale parameter covering a large portion of the site. For example, in reference to the example problem, hydrofacies classes could be defined based on geological criteria set by a hydrogeologist with experience at similar sites, and an upper and lower bound on $K$ could be assigned to each hydrofacies class. An important component of the assessment is attaching uncertainty bounds to the parameter estimate.

In hydrogeology, the measurement process often consists of performing a field test to measure an intermediate, technique-dependent variable, which must be transformed to obtain the primary parameter of interest—based on some measurement model and inversion approach. There are many places in this multi-stage process where errors can potentially enter in and, potentially, contaminate or create uncertainty in the final parameter estimate. The propensity for
compared to perfect measurements. Whether a particular scale of variability larger than the network scale can be resolved using imperfect measurements is difficult to assess, but depends on the relationship between the degree of uncertainty in the estimated model parameter and the amount of variability at that scale.

Measurement data is often categorized as either hard or soft data—hard referring to "exact" and soft to "inexact". The categorization can be very important since the two types of data are often treated very differently in a problem. However, this distinction is ambiguous since the terms "exact" and "inexact" are subjective. Every measurement contains some degree of error. Thus, the distinction between hard and soft data should be based on a defined error threshold; measurement errors of a magnitude below the threshold are considered acceptable and errors of a magnitude above unacceptable. Acceptable and unacceptable errors should be defined based on their potential to corrupt the model results to a degree that alters the conclusions made based on them. Performing a model sensitivity analysis can help in choosing the threshold. Although the choice of error threshold will invariably be somewhat subjective, having a strict, quantitative criterion to distinguish between hard and soft data is better than having none at all.

A type of soft measurement that is in a class by itself is engineering or geologic judgment. This type of measurement involves an expert making a subjective, semi-quantitative assessment about a problem parameter, usually a large-scale parameter covering a large portion of the site. For example, in reference to the example problem, hydrofacies classes could be defined based on geological criteria set by a hydrogeologist with experience at similar sites, and an upper and lower bound on $K$ could be assigned to each hydrofacies class. An important component of the assessment is attaching uncertainty bounds to the parameter estimate.

In hydrogeology, the measurement process often consists of performing a field test to measure an intermediate, technique-dependent variable, which must be transformed to obtain the primary parameter of interest—based on some measurement model and inversion approach. There are many places in this multi-stage process where errors can potentially enter in and, potentially, contaminate or create uncertainty in the final parameter estimate. The propensity for
these errors to be problematic is often dependent on the problem context. It is important for the practitioner to carefully dissect the measurement process in order to identify potential sources of error and determine if it is necessary to account for uncertainty in the parameter values ascertained from a particular measurement type—thus classifying the data as soft data.

2.3 The decision-making process

The types of decisions that have to be made in engineering problems, especially hydrogeology-related problems, are often vexing, equivocal, and difficult to make as a consequence of the high degree of uncertainty associated with these problems. In the decision-making context, a decision refers to a choice about a future course of action made from a set of possibilities, referred to as decision alternatives. Thus, decision-making requires looking into the future and predicting what decision alternative will be most beneficial / least costly. This process becomes progressively more difficult as uncertainty about factors influencing the decision increases, since the future consequences of choosing a particular decision alternative become less predictable.

Figure 2.4 is a visual conceptualization of the decision-making process for a hydrogeology-related engineering problem. The decision-makers are faced with several important decisions: (1) which engineering design to implement, (2) which site investigation design to follow, and (3) which type(s) of data is most worthwhile to collect (e.g. hard borehole data or soft geophysics data). To make these decisions they must compare the benefits, costs, and risks associated with each alternative, all within a “cloud” of uncertainty. The figure illustrates the types of difficult issues a decision-maker must confront. The following sections will examine these issues and present a general framework for systematically dealing with them in order to ensure that robust decisions are made, even in the face of uncertainty.
2.3.1 Key components of decision-making process for engineering problems

The first step in carrying out a formalized approach to decision-making is to define the key components of the problem, which can be categorized as:

- objectives,
- decision questions,
- constraints, and
- decision variables.

(Note: the terms objective and constraint can be used in many different contexts with different implications. In Section 2.3 these terms will always be used in a decision-making context and, thus, will not be prefixed by the word "decision". However, in other sections of this work, where their usage could be confusing, they will be prefixed with "decision".)

Defining clear objectives before decisions are made is a fundamental, often overlooked, part of the decision-making process, since without objectives there is no context for decision-
making. Indeed, decision-making without objectives is a contradiction in terms; how can a decision be made when the desired outcome of the decision is unknown? The comparison between how well the expected outcomes from different decision alternatives meet the predefined objective(s) is the main criteria for evaluating the attractiveness of an alternative. A problem can have many objectives, although it is important that the objectives do not contradict each other.

An effective way to set objectives is to use a top-down pyramid approach (see Figure 2.5) where the primary problem objective(s) are defined first and represent the apex of the pyramid. Intermediate objectives which will help in the process of achieving the primary objective(s) make up the underlying levels of the pyramid. The structure of the pyramid should be such that intermediate objectives should also assist in reaching other intermediate objectives lying above them, in addition to reaching the primary objective(s). Usually, the different tiers of the pyramid correspond to different stages in the project timeline.

![Figure 2.5: Pyramid approach for defining decision objectives. The apex of the pyramid represents the primary objective, underlain by intermediate objectives that are to be met from bottom to top—culminating with the primary objective.](image)

It is important to stress that the emphasis should be on limiting, not increasing, the complexity in the defined set of objectives in order to simplify and clarify the decision-making process as much as possible. The ideal situation is to have just one primary objective and to be
able to assess the influence of every activity on reaching this objective. Then an activity’s value to the overall project goal can be determined. However, establishing this global linkage of cause and effect between activity and primary objective can be difficult in the early stages of a project when the problem is ill-defined. Before a site is characterized to some degree it may be impossible to determine what the possible design alternatives might be. In fact, it may not even be possible to establish a primary objective before the nature of the problem is identified. This is a situation where an intermediate objective becomes advantageous, defined to reach the level of site description where these obscurities in problem definition are cleared up.

In the example problem, at the early stages of the project, while it may be surmised that there is a potential for contamination, it may be unknown whether there actually is a contamination problem and, if there is one, what type of problem it is (e.g. shallow soil contamination, deep subsurface contamination which is immobile, groundwater contamination entering the inlet). At this stage a generic primary objective could be to implement the most cost-effective engineering design. However, there is so much uncertainty about the nature of the actual problem that the set of possible design alternatives has to account for an unlimited number of scenarios, making it difficult to directly link site activities with the objective (e.g. the influence of taking hydraulic head measurements in existing wells on the design decision). An intermediate objective could be to develop a site conceptual model which describes the general geology, hydrogeology, and sources of contamination (if any) across the site and identifies the parameters that have a strong influence on the design decision. Although this intermediate objective represents a qualitative and subjective criterion (unlike the purely quantitative primary objective), it can be thought of as a compass, used to point the project in the right direction—towards achieving the overall goal.

Once the objective(s) of an engineering problem are defined, the decisions which need to be made in order to achieve them can be set; these will be referred to as the decision questions. The primary decision question facing the decision-maker(s) in engineering problems is which engineering design should be implemented to solve the problem at hand. However, to reach the
point where this decision can be properly made usually requires addressing a set of antecedent questions. These questions could include:

- Which parameters are important to the engineering problem and, thus, need to be characterized?
- Which processes are important and, thus, need to be modeled?
- How complex of a model should be used?
- What site characterization strategy should be implemented?
- How much characterization data is enough?

An important aspect of decision-making is posing the correct questions or, in other words, determining what decisions have to be made.

Constraints, in a decision-making context, represent defined limits imposed on certain aspects of the engineering problem, usually regarding the engineering design performance. Imposing these limits may, in turn, rule out certain potential decision alternatives which are unable to satisfy the constraints. Constraints differ from objectives in that they represent project requirements that do not vary and must be satisfied. In contrast, objectives constitute project goals which are striven for, but not necessarily met, and are usually phrased in terms of maximizing or minimizing some performance metric. Constraints can be made objectives and vice versa, but the two are treated fundamentally different in the problem analysis.

Constraints can be categorized as technical, economic, regulatory, and social. In reference to the example problem, technical constraints could correspond to the maximum excavation slope angle possible before the walls collapse for an engineering design requiring excavation of contaminated soil. Another technical constraint could be the maximum pumping rate in a water well used for containment or extraction of contaminated groundwater—determined based on the well design, pump capacity, and local hydrogeology. An economic constraint could correspond to a limit on the time required for completion of site remediation in order that development of the site can start on time. Regulatory constraints on contamination could be in one of several forms, as defined by the governing regulatory body, including: (1) maximum allowable contaminant
concentrations depending on the intended land use or (2) maximum allowable health risk (e.g. probability of one death in a million over an average lifetime), which would depend on the predicted type and amount of exposure to a contaminant. Another type of regulatory constraint could be a maximum allowable drawdown of the water table in an aquifer used for water supply.

Social constraints are much more abstract and difficult to define, but are included since they are a reality in many problems dealing with the environment and natural resources, which includes hydrogeology problems. A possible social constraint in the example problem is the public perception of living on contaminated ground. In other words, the developer may want to reduce contamination to negligible levels (or eliminate it completely), even though this level of remediation exceeds regulatory requirements, in order to allay the public's fear about the health effects of living on "contaminated" ground. Although difficult to estimate, the economic benefits of abiding by this social constraint may be significant in the long run, if doing so increases the public demand to live there, thus increasing occupancy and allowing higher prices to be charged.

The parameters in a problem which influence the decision(s) are the decision variables. Decision variables include the parameters that define the decision alternatives, and are used as inputs in the model(s) describing system performance, as well as the model output parameters, which are used to differentiate alternatives. In a mathematical context, the former type of parameters are independent variables and the latter type are dependent variables. In the example problem, independent decision variables could be:

- where, how much, and what types of contaminants were released;
- hydrogeological properties affecting flow and transport (e.g. hydraulic conductivity and porosity); and
- unit costs, such as the unit cost of pumping a well, excavating soil, and performing a particular type of treatment.

Dependent decision variables could be:

- the present and future spatial distribution and concentrations of contaminants in different phases (i.e. non-aqueous, aqueous, and sorbed phases of contaminant mass);
Chapter 2. BACKGROUND

- the number, and associated pumping rates, of pumping wells;
- the amount of material excavated, treated; and
- the total cost of remediation for a design alternative, which could be the metric used to compare how well different alternatives meet the objective (if the primary objective is to minimize total cost).

A decision model should be as simple and straightforward as possible while still properly accounting for, through decision variables and system models, all the processes that significantly influence the decision alternatives.

2.3.2 Concept of risk in decision-making

As mentioned earlier, uncertainty complicates the decision-making process by creating unpredictability and, thus, should be accounted for in the process. To account for uncertainty in decision-making it is necessary to understand the relationship between the two. Uncertainty in any of the decision variables in an engineering problem results in the possibility of unintended decision outcomes (decision alternatives failing to meet expectations). There are two types of unintended outcomes resulting from choosing a non-optimal decision alternative:
- not meeting the specified objective(s) and/or constraints (decision failure), and
- meeting all objectives and constraints, but another alternative would have met the objective(s) better.

In the example problem, the first type of outcome could correspond to the existence of soil contamination above the regulatory threshold level (representing a constraint) at certain locations of the site, even after the selected remediation design (representing a decision alternative)—calling for excavation and treatment of soil from identified areas of the site—is implemented. The selected excavation design failed to account for some of the contamination due to underclassification. The second type of outcome would result if uncontaminated soil is inadvertently excavated and treated in an overly-conservative soil remediation plan. The
remediation of clean soil is due to overclassification of contamination and results in unnecessary costs. In this particular example, a single decision alternative can result in both types of outcomes, underclassification in one area of the site and overclassification in another area.

If a problem has no uncertainty then there should be no chance of unintended outcomes, since there is no doubt about the best decision alternative. However, once uncertainty enters into a problem, this chance can become non-negligible. The chance and consequence of an unintended decision outcome will be referred to as risk and risk cost, respectively, in this work. Cost, in this context, represents the difference between the unintended and intended outcome in the value of the dependent decision variable(s) used to distinguish alternatives. It does not have to be a monetary value. In reference to the example problem underclassification scenario described above, the risk cost could correspond to the monetary failure costs (e.g. fines and the cost of delayed remediation) associated with not properly remediating contaminated soil. Thus, as the risks associated with a decision alternative increase (due to heightened uncertainty in the decision variable(s)), the likelihood of having to absorb risk costs increases. Of course, at the time a decision has to be made, it is unknown whether an unintended outcome will actually occur and, thus, whether an associated risk cost will be incurred. However, if the problem is conceptualized as an ensemble of equally-likely descriptions of reality (realizations), based on available information about the decision variables, each realization would have a particular decision alternative outcome and, thus, risk cost (or lack there of) associated with it. Then an expected, or probabilistic risk cost, could be thought of as the average risk cost for all realizations.1

Probabilistic risk cost is the key to risk-based decision-making, which accounts for uncertainty in decision variables. The purpose of formalized decision-making is to identify the decision alternative that best meets the problem objective(s), while satisfying all constraints. In

---
1These definitions for risk are somewhat unique to this work. In most of the work on hydrogeological decision analysis, risk is defined as the expected cost of failure (see Freeze, et al., 1990), which is synonymous with probabilistic risk cost in this work, if failure is taken in a more general sense to mean an unintended outcome. The definitions concerning risk are purposely more descriptive in this work in an attempt to clear up some of the ambiguity associated with the term "risk".
Chapter 2. BACKGROUND

order to differentiate the alternatives based on these criteria, it is necessary to model their future impact on the system, with respect to the objectives and constraints, in terms of standard metrics (dependent decision variables) that can be compared. Probabilistic risk cost represents a statistically robust measure of the influence of uncertainty in decision variables on decision alternative outcomes (i.e. the negative impact of unintended outcomes as compared to their expected outcomes). The measure is in terms of the same dependent decision variables used to gauge the attractiveness of intended alternative outcomes. Thus, probabilistic risk cost can be used in conjunction with the benefits and costs associated with intended outcomes to compare and rank alternatives in a decision-making framework that accounts for risk.

This approach to incorporating risk in decision-making can be illustrated using the example problem. Say that it is determined that the probability of soil contamination above the regulatory threshold level in a certain part of the site is fifty percent. Then the no action alternative would result in a probabilistic risk cost of one-half the dollar cost of failure associated with underclassification of contaminated soil (the product of the probability of failure and the failure cost). The intended costs and benefits for no action are zero. However, for the remediation alternative (excavation and treatment of soil), the probabilistic risk cost is zero since there are no unintended outcomes resulting from underclassification or overclassification of contamination (i.e. the presence of contamination is correctly classified). The intended cost of the remediation alternative is simply the dollar cost of performing the remediation, and the intended benefit is zero (there are no benefits in this problem scenario). Thus, if the primary problem objective is to determine the most cost effective engineering design for site remediation, then the choice of optimal design alternative is determined by which is less expensive: (1) one-half the dollar cost of failure or (2) the cost of soil remediation, corresponding to the no action and remediation alternatives, respectively.
2.3.3 Engineering versus health risk

Risk and the associated terms defined thus far can apply to decision-making problems in general. The specific definition of risk depends on the profession and context it is used in and, thus, should be clearly stated in a problem to avoid ambiguity. The two types of risk that are addressed in this work are (1) engineering risk and (2) health risk (primarily engineering risk, because it applies to the case study). Engineering risk applies to engineering design problems where the risk is the likelihood of a design alternative not performing as expected as a result of uncertainty in independent decision variables that affect design performance (e.g. $K$ of an aquifer, $K$ of a man-made hydraulic barrier wall). The engineering risk cost is either design failure costs due to the design not being robust enough or unnecessary costs due to over-designing, and is usually couched in monetary terms.

Health risk is applied to problems where the health of a living organism or system (e.g. human being or ecosystem) is potentially in jeopardy due to some environmental factor. In the environmental health sciences, risk refers to the likelihood of an individual or population experiencing a detrimental health effect from some harmful substance or condition (e.g. industrial chemicals in groundwater used for drinking). Unlike engineering risk, health risk is not necessarily the likelihood of an unintended decision outcome, but a dependent decision variable used to differentiate decision alternatives. The use of likelihoods to describe the long term health effects of low doses of harmful substances on living organisms instead of trying to identify exactly what will happen to the body stems from the fact that individual organisms can react very differently to the same exposure, simply due to the complexity of biological systems.

Since the health consequences resulting from a particular exposure scenario are dependent on many interlinked, complex processes, it is extremely difficult to predict. Even if the dose of a harmful substance to a particular target area of the body is perfectly known, there can still be significant uncertainty associated with the estimated long term health risk due to limitations in the potency measurement process [Reichard and Evans, 1989]. Potency is the severity of health effects resulting from a specified dose of a certain substance. It is calculated by fitting defined
statistical models to data from controlled laboratory experiments and/or epidemiological studies, usually involving animals exposed to doses much higher than anticipated in the environment of interest [Reichard and Evans, 1989]. In addition, the dose of a harmful substance to different areas of the body for a particular situation is often very difficult to estimate, since it depends on the type and extent of exposure to the substance, which, in turn, depends on the distribution of the substance within the environmental media, which depends on the nature of the substance release and its transport within the environmental media; all of these linked components can add to the overall uncertainty in health risk.

Since health risk is actually a dependent decision variable, usually of great importance to the decision-making process (i.e. it constitutes the measure used in one or more of the objectives and constraints), it is advantageous to take a risk-based approach, where unintended decision outcomes are accounted for, in these decision problems. This is especially true considering the high degree of uncertainty often associated with estimating health risk, as discussed earlier. In this context, an unintended decision outcome is the existence of an unexpected health risk level, either higher or lower than anticipated, after a decision is made; and the risk is the likelihood of one of these outcomes occurring in reality. The risk cost is the consequence of a higher or lower than anticipated health risk, couched in terms of the dependent decision variable that is used to rank decision alternatives in the problem objective(s).

The approach can be more clearly described through an illustrative example. In reference to the example problem, if the objective is to minimize the health risk, the risk cost associated with a soil remediation design is the amount that the actual health risk remaining after remediation exceeds the expected value (in a statistical sense) of the post-remediation health risk. This cost represents an underclassification risk cost. In this scenario there is no overclassification risk since an unexpectedly low health risk decision outcome is a benefit, not a cost, in light of the problem objective. If the problem objective is to minimize the monetary costs of the remediation design while ensuring that the health risk does not exceed the regulatory limit anywhere (i.e. health risk represents a constraint), the risk cost of a soil remediation design is either the
monetary failure cost associated with not meeting the regulatory limit (underclassification) or the unnecessary monetary cost of remediating soil which already has a health risk below the limit (overclassification). The risk described in the last case (where minimizing the design costs is the objective) actually constitutes engineering risk.

In engineering problems where health risk is a concern it is advantageous to account for the uncertainty associated with both health risk and the engineering design in the decision-making process in order to make more robust decisions. There are several ways of accomplishing this depending on how the problem objectives and constraints are defined. As illustrated in the previous example, if the objective is to minimize engineering design costs, health risk can be incorporated as a constraint\(^2\). In setting up a problem as such, engineering risk can result from uncertainties in decision variables affecting health risk as well as design performance. The cost of the performance of an engineered design failing to meet the health risk constraint has to be defined in monetary terms in order to match criteria used in the objective (unless there is no uncertainty in health risk, in which case the question of whether a design alternative satisfies the constraint is definitive). Another way to set up the problem is to minimize health risk as an objective, while setting an upper limit on the engineering design budget as a constraint. With this type of objective, the risk is in terms of health risk and depends on how well the health risk and the design performance can be estimated. There is no uncertainty associated with the engineering design budget constraint since the design cost is unaffected by how much it reduces the health risk in this problem context. The main limitation to incorporating either health risk or engineering design cost as a constraint in the decision-making framework is that, as a constraint, they cannot be optimized, as they could be as an objective.

If both health risk and engineering design cost are included in the problem objective (e.g. minimize health risk and the monetary design cost) then, theoretically, the two decision variables

\(^2\)If health risk is used as a constraint it seems only reasonable that the potency component of health risk is treated as certain and is set by the regulators so that the constraint is rigid and standardized. There could still be plenty of uncertainty associated with the health risk estimate as a result of the uncertainty involved in estimating the dose, which is used in conjunction with potency to calculate health risk.
together can be simultaneously optimized. However, in order to accomplish this using conventional decision analysis or optimization techniques, health risk and design cost must be defined in terms of the same decision variable(s). Since expressing the economic engineering costs in terms of health risk is nonsensical, the only reasonable option is to express health risk in economic terms, which requires assigning an economic value to life. Reichard and Evans [1989] use risk-based decision analysis to identify the optimal environmental control action for a groundwater contamination problem, based on the objective of minimizing the total social cost (which contains the economic cost of the action plus the economic cost associated with the remaining health risk). They account for uncertainty in potency as well as in many of the decision variables used in the dose estimation. They set the assumed value of life at one million dollars for their main analysis, although they also perform a sensitivity analysis of this value. Assigning a value to life raises many political and social questions. There have been some attempts to develop multi-objective optimization techniques, in which the separate objections do not have to be defined in the same terms; however, these techniques are unwieldy to use, require prioritization of the objectives anyway, and, thus, are not discussed in this work.

In this day and age it is becoming more and more common to approach environmental problems from the perspective of health risk. The impetus for using health risk is that, unlike a measure of contaminant concentration, it directly describes the problem of concern—the potentially harmful health effects that can occur from contamination at a particular location for a particular activity. Thus, characterizing health risk enables the focusing of contamination control efforts on dangerous, high health risk problem areas and the disregarding of areas with little health risk; this possibly results in significant monetary savings while still protecting health. Unfortunately, the estimation of health risk is often plagued by a high degree of uncertainty, due to a multifaceted dependency on a chain of many different processes—from contaminant release into the environment to human exposure—that are difficult to accurately model. Thus, to properly incorporate health risk in the decision-making process for environmental control problems, uncertainty in the modeled processes that affect health risk, as well as in the decisions
that are based on health risk (which takes the form of engineering risk), should be accounted for. Although there is no ideal way for accomplishing this while minimizing both engineering design costs and health risk, this section presented several robust options for handling these types of problems in a risk-based decision-making framework.

2.3.4 Concept of data worth

The value of information (VOI) to any problem requiring the making of a decision is determined by the degree to which the information enables a better decision to be made; this is evidenced by how much more favorable the outcome is with the aid of the new information, in comparison to not having it. The likelihood and consequence of undesirable decision outcomes is captured in the risk terms of the decision-making framework, as described earlier. Thus, the true VOI is directly determined by the subsequent reduction in the overall risk costs.

Of course, it is desirable to know the VOI a priori, before an investment is made to collect the information. If it is possible to calculate the probabilistic risk cost before and after simulated collection of information (in which all possible outcomes of the data collection are accounted for) using probabilistic methods, then the expected VOI can be calculated by taking the difference between the a priori and a posteriori probabilistic risk cost. It is important to realize that the expected VOI represents the most likely VOI based on probabilities, not the true VOI, since the true risk cost is unknown beforehand. In the context of engineering problems "information" is usually in the form of data collected from measurements. Thus, data can be expected to have worth from the perspective of a decision-maker only if the estimated reduction in the probabilistic risk cost is greater than the cost of the data collection and analysis. The data worth concept is the basis of the methodology used in this work to explore the usefulness of geophysics data in hydrogeological engineering problems, as applied to the real world case study problem. The mechanics of the methodology—how to calculate data worth—will be explained in

---

3This assumes that unintended outcomes are entirely the result of uncertainty (i.e. if there is no uncertainty the outcome of decisions can be perfectly predicted and, thus, the best alternative can be chosen with certainty).
later chapters. Previous work that has applied the VOI / data worth concept to hydrogeology problems include Reichard and Evans [1989], Freeze et al. [1992], James and Freeze [1993], James and Gorelick [1994], and James et al. [1996].

Just because data reduces uncertainty about site condition(s) does not necessarily imply that the data has worth, since the uncertainty reduction may not translate into a substantial reduction in risk. The amount of risk reduction resulting from parameter uncertainty reduction depends on the decision-making context of the problem. For example, in reference to the example problem, say the primary objective is to prevent contaminated groundwater from entering the inlet at the lowest possible cost, and the two engineering design alternatives being considered are no action or hydraulic containment using a pumping well. If transmissivity ($T$) had already been estimated from one pump test, the worth of collecting $T$ data from a network of slug tests may have little worth, even if the slug tests significantly reduce uncertainty about small-scale variations in $T$. This could occur if, based on the pump test data and delineation of the contaminant plume, it is already evident that the plume will reach the inlet with concentrations above the regulatory limit, within the compliance period. In other words, it is already obvious that containment is the more cost-effective option and the remaining decision, the required pumping rate to ensure plume containment, is better determined from the large-scale pump test measured $T$ value than it would be from the small-scale slug test measured $T$ values anyway. In this case, the lack of worth in taking slug tests would be evident in a data worth analysis since there would be no reduction in the probabilistic risk cost.

2.4 The site characterization process

Characterizing site conditions is a critical and important aspect of hydrogeology-related problems due to the high degree of uncertainty usually associated with hydrogeological processes. This uncertainty has many origins, including simplifications in the models used to describe these processes, but the dominant source is heterogeneity in the hydrogeological environment and our inability to accurately describe it. This unaccounted for heterogeneity can
lead to significant errors in hydrogeological process model predictions (e.g. inaccurate groundwater contaminant travel time, pure-phase contaminant spatial distribution, or hydraulic head depression resulting from pumping) due to errors in the input model parameters (e.g. $K$, porosity, or contaminant source concentrations and extent), even though the physical models themselves may be accurate. Inaccurate model predictions can, in turn, lead to the making of decisions with unfavorable, possibly costly, outcomes (e.g. choosing a no action alternative for groundwater management when groundwater contamination is actually a problem). Hence, there is a need to better characterize the heterogeneous hydrogeological environment in order to enable more accurate modeling of important processes and, thus, better predictions of system behavior; in turn, this enables more robust decisions to be made.

This section examines the site characterization process, emphasizing a systematic, big-picture approach. Although site characterization is important to hydrogeology-related problems, there is a fine line between not having enough information about site conditions to make proper decisions and having superfluous information that provides little or no benefit to decision-making. There is a point where the cost of collecting more data becomes greater than the benefits it provides to the overall problem. In addition, certain types of data are more beneficial than others for improving decision-making. Therefore, in order to pursue a cost-effective site characterization program, it is important to take a systematic approach wherein data collection strategies are justified based on the benefit they provide towards the overall problem objective(s), as represented in the decision-making process.

2.4.1 Objectives

As a result of significant amounts of uncertainty in decision variables, especially in hydrogeology-related problems, the risk in decision-making can be considerable. This leads to the quandary of choosing costly decision alternatives in order to ensure a robust decision outcome or choosing less costly alternatives that have the potential to fail. The only way to reduce risk is to reduce uncertainty, requiring an increase in knowledge about uncertain
independent decision variables (e.g. $K$, soil contamination). In hydrogeology-related engineering problems, this increase in knowledge can only be attained by taking measurements (e.g. slug tests, soil samples) that provide information about these variables. Thus, the main impetus for site characterization can, and should be, directly linked to the decision-making process by making the overall characterization objective to reduce risk [Freeze et al., 1992]. With this underlying objective, the imperative is to determine the most effective and efficient characterization strategy (i.e. type, scale, number, and locations of measurements) for reducing risk; part of this strategy is knowing when to stop sampling, before the cost of additional sampling becomes greater than the benefit provided by risk reduction. This problem is best addressed in a data worth context (as discussed earlier in Section 2.3.4).

Having stated that the general objective of site characterization is to reduce risk, unfortunately, it is not always possible to develop measurement strategies based solely on risk, although every effort should be made to link these strategies to the decision-making process and the project goals. Risk is often ill-defined at the early stages of a project, when there may not be enough information to identify potential design alternatives, or even to define the primary problem objective. Alluding to the example problem, at the beginning of the site investigation it may not be known if and what types of problems (e.g. soil contamination, groundwater contamination, and/or geotechnical problems) might interfere with the real estate development and, thus, it is difficult to determine what environmental control actions may be required. The undefined nature of the problem at this point in time makes it difficult to develop performance criteria for calculating risk. In addition, in environmental problems, decisions about the design of field investigation programs is often largely driven by regulatory requirements [Freeze et al., 1992]. However, it is still important to define objectives, which complement the decision-making process, for all phases of site characterization, and at every stage of an engineering project, so that the site investigation stays focused on the engineering problem at hand.

A common intermediate (albeit somewhat vague) decision objective early in a hydrogeology-related engineering project timeline is to develop a site conceptual model that
describes the general geological and hydrogeological environment, and identifies any anomalies and parameters that may significantly affect the overall problem outcome (see Section 2.3.1). The site conceptual model serves to focus the engineering problem—identify what are the real concerns and their potential solutions. Although criteria for meeting this decision objective are inherently qualitative and subjective, the associated risk can be thought of as the likelihood of not identifying important site conditions which, if accounted for, would alter decisions. In order to reduce this risk it may be rational, depending on the extent of prior knowledge about the site, to initially investigate all types of conditions that might be influential in the problem. Then, the specific site investigation objectives for this phase could be defined as: to determine the general characteristics of the site in terms of the regional geology, regional hydrogeology, present and past human activities, and anomalous areas. These objectives will be met when the problem can be defined in terms of the main concerns, potential engineering alternatives, and a primary decision objective (e.g. to develop the most cost effective engineering design for control of environmental contamination); after this early phase the site investigation can focus on reducing risk. The later phase of site characterization centers on taking measurements of uncertain material properties—in order to better predict engineering design performance and, thus, reduce risk.

Therefore, while basing all site investigation efforts on a universal motive—to reduce risk in the overall solution to the problem—is highly advantageous (since all actions can then be assessed in the context of how they affect the bottom line), it is still often necessary to develop specific objectives for different phases of the site characterization process, especially at early stages of the process when the problem is poorly defined.

2.4.2 Organization

Site characterization is often one of the largest, and most expensive, components of hydrogeology-related engineering problems and usually involves many different stages and types of activities. In order to ensure that a rational and systematic approach to site characterization is
adhered to, it is advantageous to organize the different components of the process in some way, while keeping in mind that all investigation activities should support the stated characterization objective(s). These components can be organized from a temporal or problematic perspective, wherein site investigation activities are grouped based on their sequence in the investigation timeline or how they specifically contribute to the understanding of site conditions, respectively [Baecher, 1972].

2.4.2.1 Temporal perspective

From the temporal perspective the site characterization process is seen as a temporal sequence of activities that start by investigating general characteristics of the site and become progressively more focused and detailed with time. The process is divided into three separate temporal stages: reconnaissance, preliminary exploration, and detailed investigation.

Reconnaissance involves the gathering of mostly qualitative, descriptive information about regional and local geology/hydrogeology, the history of human activity, and possible adverse conditions across the site. The information is ascertained from existing sources (e.g. previous studies, regional geologic/land use maps, air photos, construction records, people's memory, visible site conditions) through literature reviews, interviews, and on-site visual inspections. No actual measurements are taken at this time. Thus, the reconnaissance stage is used to familiarize oneself to the site and learn about large-scale features across the site, by consolidating existing useful information about the site.

Preliminary exploration centers on collecting the first quantitative data, in order to provide a little more detail about subsurface site conditions. The data is used to determine the location and general geometry of large-scale geologic/hydrogeologic features (e.g. geologic stratigraphy, geologic/hydrogeologic structure, depth to the water table), the type and possible location of subsurface anomalies (e.g. fractures, contaminant source zones, buried objects), and the initial estimates of applicable engineering properties. The measurement methods used during this stage are mostly non-invasive (do not require any displacement of subsurface material, e.g. field mapping, surface geophysics) or minimally invasive (e.g. cone penetrometer, hydraulic
punches\textsuperscript{4}, limited borings). The first two temporal stages of site characterization mainly support the development of a site conceptual model, used to refine the engineering problem definition and guide, focus, and optimize future investigation efforts.

Accurate quantitative data on the geometry of important subsurface features and material properties is obtained during the detailed investigation. This requires the taking of large amounts of measurements, usually including intensive emplacement of invasive borings used for in-situ testing (e.g. hydraulic head and conductivity measurements, borehole geophysics) and sampling to obtain accurate data. At this point of the investigation the emphasis is on obtaining better predictions of engineering behavior—in order to reduce risk.

The reasoning behind this type of temporal approach to site characterization is to first obtain information about large-scale geometric relationships before attempting to describe small-scale variability in material properties, since the latter only becomes important when the former is well known; the argument being that accurately characterizing small-scale material property variability in isolated locations has little worth if the large-scale behavior of the system is unknown. Another reason for starting general and becoming progressively more detailed is that the site investigation can be abandoned at any time with minimal wasted effort/cost (due to the prior collection of superfluous information), if it is decided that characterizing site conditions in more detail (i.e. increasing model resolution) is unnecessary or not cost-effective (i.e. the required measurement effort is not justified by the resulting reduction in risk). Alluding to the example problem, if a reconnaissance and preliminary exploration of the site reveals that the clay aquitard underlying the unconfined aquifer is highly discontinuous and underlain by a sand aquifer across the region, then it could probably be assumed with high confidence that any contaminated groundwater that reaches that layer will breach it. Thus, there would be little benefit in undertaking an intensive boring program with the purpose of characterizing in detail the continuity of the clay aquitard. If such an intensive boring program was initiated before the

\textsuperscript{4}Cone penetrometers and hydraulic punches are narrow-radius tubes that are pushed into the subsurface to take various types of measurements.
much more general study of the regional geology/hydrogeology was completed, there would be a large amount of wasted effort and cost.

2.4.2.2 Problematic perspective

Another way of looking at the organization of the site characterization process is from the problematic perspective, wherein the activities are grouped according to the specific problem class they support [Baecher, 1972]. Baecher (1972) identifies five problem classes in site characterization:

1. reconnaissance,
2. pattern recognition,
3. reconstruction,
4. search, and
5. quantification of material properties.

In this context, problem classes represent the types of problems site characterization efforts attempt to solve, grouped on the basis of having similar analysis techniques.

Reconnaissance, which includes the same activities as in the reconnaissance stage of the temporal approach, is used to provide enough information to form initial hypotheses about the subsurface structure of a site.

Pattern recognition represents the classification of objects with similar traits into groups that can be treated alike. It is most commonly used in hydrogeology-related problems to classify geologic substrate into units/formations/facies based on similarities in properties (i.e. patterns), such as grain size distribution, mineral composition, age, degree of fracturing, and degree of consolidation. Pattern recognition reduces the number of degrees of freedom (independent variables) in a problem by consolidating some of them. It is usually a necessary simplification required to make problems more tractable, but it also structures information in a way that is much easier to understand and interpret. The process of pattern recognition is often carried out

\[5\] Baecher (1972) combines pattern recognition and reconstruction and also includes joint surveys as a problem class specific to geotechnical engineering.
subjectively based on human judgment and interpretation (indeed this is almost always the case when geologic logging of boring cores—where a geologist describes in vertical sequence the material extracted from borings—is used to delineate geologic stratigraphy), although there are more rigorous statistical methods (e.g. cluster analysis used to define classes and discriminant analysis used to determine class membership for a set of attributes [Murray, 1994]) that can be employed. Many of these statistical methods are cast in a probabilistic framework, where a probability of class membership is calculated for each set of measured attributes (although the methods usually assume there is no uncertainty in the measurements). The statistical approach has been used to classify petrophysical facies [Murray, 1994], electrofacies [Moline et al., 1995], and hydrofacies [McKenna and Poeter, 1995], based on borehole geophysical log data where the classes are defined by comparing well-core data with the geophysical data (the well-core data included measurements of permeability in the hydrofacies case).

Reconstruction represents the prediction of unobserved features or properties, usually based on observed features or properties, in addition to subjective hypotheses. This process is an inherent requirement for all hydogaology-related problems, since it is impossible to characterize every point on a site and, hence, always necessary to interpolate between and extrapolate beyond measurements, in one way or another. This requirement for reconstruction applies to classes (e.g. geologic and hydrogeologic units), as determined by pattern recognition, as well as material properties (e.g. permeability, porosity, contaminant concentrations). When reconstruction is performed a hypothesis about the spatial structure of the property/feature being estimated is made, regardless of whether or not this fact is known or desired. A robust approach to decision-making should require that the process used to infer the model of spatial structure is rational and makes full use of existing data. Indeed, reconstruction will always require some degree of subjective interpretation. A priority should be to reduce the level of subjectivity in the process by making it as rigorous as possible. (Note: this does not imply that the process should be made as complex as possible. Complexity often hides uncertainty by giving the appearance that more detail can be described than is actually known.)
At this point, the discussion about rigorous approaches to reconstruction will be limited to the distinction between non-probabilistic (deterministic) and probabilistic (stochastic) approaches, leaving a detailed discussion for later chapters and the appendices. Stochastic approaches to reconstruction attempt to account for uncertainty in the estimated values (e.g. geostatistical methods where estimation uncertainty is calculated and accounted for), while deterministic approaches assume the estimates are certain, despite there always being some uncertainty associated with the process (e.g. hand contouring, inverse-square interpolation, geostatistical methods where the estimate represents an expected value and no uncertainty is connected it). Deterministic approaches negate the use of a risk-based approach to decision-making, since uncertainty in decision variables has to be known to calculate risk. In this work a stochastic geostatistical approach is used to perform reconstruction.

Search represents the process of locating anomalous conditions in the subsurface, usually requiring the taking of measurements. Anomalous conditions can be defined as large-scale heterogeneity or irregularities in any property or condition considered important to the problem. Examples include discontinuous geologic units (e.g. sand lenses, pinchouts, nonconformities), aquitard discontinuities, buried objects, contaminant source zones, and contaminant plumes in groundwater. The objective of the search process is to locate targeted anomalies as effectively, while staying within a limited budget. Obviously, if there is no limit on the number of measurements that can be made (corresponding to an unlimited characterization budget), the search strategy with the greatest chance of success (i.e. 100% chance) could require an enormous amount of measurements (depending on the measurement scale, target size, and site size). (Note: this is a situation where geophysics could potentially be very valuable since geophysical measurement coverage can be very extensive and densely spaced. The potential uses of geophysics will be explored in detail in the next chapter.) Thus, there has to be a compromise between measurement effort and search chance of success.

Search theory is well established for locating targets of simple geometric shapes by using different grid patterns of point measurements [Freeze et al., 1992]. The theory is based on simple
geometrical relationships and provides probabilities of hits and misses for different geometric configurations assuming the target(s) are equally likely to exist anywhere on the measurement grid and have no spatial correlation (i.e. the presence or absence of the target at a location provides no information about the presence or absence of the target in the zone surrounding the first location). Indicator geostatistical methods can account for target spatial correlation, but then it is no longer possible to calculate the probabilities of hitting and missing the target(s) before any measurements are taken, since these probabilities will depend on the results of nearby measurements (although the probabilities for the no correlation case would represent a "worst case" bound with respect to search integrity).

The search approach espoused in this work is to compare the data worth between different measurement search strategies (i.e. measurement programs with different types, numbers, and/or locations of measurements) and choose the strategy that has the most data worth. This approach assumes that risk and probabilistic risk cost can be defined and calculated in quantitative terms, which is not true before the problem is defined and potential engineering alternatives are identified. In the early stages of a project, when the decision objective is to develop a site conceptual model, a subjective decision has to be made about how important it is to determine if a certain type of anomaly is present and, thus, how much money should be allocated to searching for it and what likelihood of search success should be aimed for—these two criteria being complementary, as explained in the above discussion. The thought process used to make this decision can still incorporate risk, if the risk cost of not identifying a certain type of anomalous site condition is made to represent the consequences on the overall project.

Alluding to the example problem, if, at the beginning of the site investigation, it is unsure whether subsurface contamination will be a concern at all, a high priority would probably be placed on searching for any subsurface anomalies that could be contamination source zones, since the consequences (risk cost) of completely ignoring subsurface contamination would be very costly. However, since at this time the only concern is whether there is any contamination, the search process could be focused, at least initially, on the areas determined during the
reconnaissance stage to have the highest potential for contamination (e.g. where fuel storage tanks once were located).

Once the problem becomes defined, and risk can be quantitatively calculated based on the primary decision objective, a rigorous data worth analysis can be performed using search theory (see Freeze et al. [1992] for example) or indicator geostatistics (see James and Freeze [1993] for example) to determine the best measurement search strategy. Referring to the example problem again, this type of rigorous search approach would be initiated for locating contaminant source zones as soon as the first evidence of subsurface contamination was discovered. In this work, a data worth methodology based on a form of indicator geostatistics is used to perform the search function, as applied in the case study.

The final site characterization problem class is quantification of material properties, which can only be accomplished by taking measurements. This function is usually performed when precise values of material properties (or model parameters) are needed for inclusion in models used to predict hydrogeologic system and engineering design performance. In a decision-making context, the material properties being estimated represent independent decision variables, used as input to models whose outputs (dependent decision variables) are, in turn, used to compare the benefits and costs associated with the different engineering design alternatives. Since the need for precise estimates of material properties usually does not originate until after the engineering problem is defined, the best approach, from a decision-making context, for identifying effective measurement strategies for obtaining these estimates is a risk-based data worth approach [Freeze et al., 1992; James and Freeze, 1993; James and Gorelick, 1994; James et al., 1996].

The requirement of precise, accurate quantitative model parameter estimates places heavy reliance on the accuracy of the measurement process, which, unfortunately, can be riddled with many types of errors (see Section 2.2.2 of this work). Therefore, before a robust decision can be made about the most effective way (i.e. the most effective measurement strategy) to estimate model parameters, an error analysis should be performed on proposed measurement types to identify all sources of measurement error. This will help determine if the cumulative amount of
error in the estimate is significant enough that it needs to be accounted for by attaching uncertainty to the measurement estimate (i.e. determine whether the measurement estimate is hard or soft data). Another important consideration is how the measurement scale compares to the model scale of interest, how the measurement estimate can be rescaled to match the model scale if the two scales are indeed different, and the associated uncertainty that results from this rescaling process (see Section 2.2.2). The assessment of a measurement type's integrity is necessary for ensuring that a valid data worth analysis of proposed measurement strategies based on that measurement type can be performed. Otherwise, the data worth analysis may over or under estimate the proposed measurement strategy’s expected worth to the overall project goals (the latter being less troubling than the former, since it represents a conservative estimate).

Unfortunately, it is not always possible to accurately assess the degree of error that will afflict a measurement before the measurement is actually taken (and sometimes can be difficult to determine even after the measurement is taken). How well the quality (degree of accuracy) of a measurement can be predicted beforehand can be referred to as the measurement’s reliability. Geophysics is especially vulnerable to having low measurement reliability, since geophysics survey results are often highly site-specific. One way of accounting for low reliability is to identify the possible range in error magnitudes that can plague a measurement and perform the data worth analysis over the corresponding range of measurement uncertainties, resulting in a range of expected data worth (EDW) values for that measurement strategy. In essence, this process represents performing a data worth sensitivity analysis on the measurement quality level. Then the decision about which measurement strategy to implement for estimating the model parameter(s) of concern can be made by comparing the range of EDW values associated with the different measurement strategy alternatives, and ranking the alternatives based on some subjective criteria (e.g. the EDW associated with lowest possible quality level for a measurement type). Even if this decision is equivocal due to low measurement reliability, this type of data worth analysis is still useful for determining the minimum quality level required for a measurement type to be worthwhile—in terms of the overall project goal(s). This data worth
approach is one of the primary topics explored in this work. Details about the approach, and a methodology for applying the approach to real-world problems, will be elaborated upon in later chapters.

2.4.2.3 Summary

In order to better understand and ensure that a systematic and rational approach is followed in the site characterization process, it is useful to organize the process into stages or components. Organization of the site characterization process can be viewed from a temporal or problematic perspective. From a temporal perspective, the process can be divided into separate stages—which should be carried out sequentially in time—on the basis of (1) the site characterization objectives at that time in the investigation, (2) the level of detail about the site that is desired at that time (which is related to the objectives), and (3) the types of information gathering methods that are employed to obtain that level of detail. From a problematic perspective, the site characterization process can be divided into problem classes based on the particular type of problem being addressed and the specific types of analytical techniques which are required in order to solve them. The problematic perspective details the mechanics of site characterization and is particularly useful when applying a probabilistic approach to the process. The two perspectives for organizing site characterization should not be considered mutually exclusive approaches, but instead, just different ways of viewing the process that can be used in conjunction (e.g. search is usually performed during the preliminary exploration and detailed investigation temporal stages) to better understand and organize the multi-faceted process [Baecher, 1972].

2.5 Review of previous work

There has been a significant amount of previous work on using geophysics for hydrogeologic site investigations, and incorporating soft data, including geophysics, in a probabilistic framework of site characterization. However, there has been limited work done
investigating the use of geophysics in a rigorous probabilistic risk-based decision-making framework. This section provides a summary of some of the existing work on these subjects that is relevant to this research, categorized by topics that address particular aspects of this work.

2.5.1 Geophysics applications in hydrology: general references and examples

There have been several very useful reference guides published, for geophysicists and non-geophysicists alike, on the use of geophysics in near-surface problems (e.g. Ward, 1990; EPA, 1993; Goldman and Neubauer, 1994) and numerous documented case studies on the use of geophysics in qualitative, deterministic (non-probabilistic) site investigation applications for a wide variety of geophysical techniques (e.g. Brewster et al., 1995; Olhoeft, 1986; Goldman and Neubauer, 1994).

2.5.2 Incorporating geophysics in probabilistic framework: measurement uncertainty not quantified

There also has been a significant amount of work done attempting to incorporate geophysical measurements in a stochastic (probabilistic) framework for reducing uncertainty in site conditions. Lahm et al. [1995] use borehole geophysical logs to precisely predict hydraulic conductivity in a deep regional aquifer by performing linear regression between direct hydraulic conductivity measurements and geophysics-measured porosity. Both sets of measurements are then used to update a geostatistical model of the hydraulic conductivity structure through a kriging type of interpolation.

Hyndman et al. [1994, 1996] use a "split inversion method", a fairly involved co-inversion process, to estimate large-scale hydraulic conductivity zonation in a shallow aquifer. This is accomplished by coupling the subsurface seismic velocity structure, derived from cross-well seismic tomography data, with tracer concentration data.
Bortoli et al. [1994] use seismic reflection data and borehole geophysical logs from an oil reservoir, the latter used as direct control data, to generate an optimized acoustic impedance image, which, in turn, can be transformed into a petrophysical image through a calibration (the last step they don't cover). Their method is to use a Monte Carlo exhaustive updating approach, where forward seismic modeling is performed on a set of equi-probable acoustic impedance realizations. The realizations are generated using geostatistical simulation based only on the well data. Then the realization that best matches the actual seismic data (based on a set of defined criteria) is chosen as being closest to reality.

In both the techniques of Bortoli et al. [1994] and Hyndman et al. [1994, 1996] uncertainty is not explicitly accounted for, as evidenced by the fact that the final optimized estimates don't have any measure of uncertainty associated with them.

2.5.3 Incorporating geophysics in probabilistic framework: measurement uncertainty quantified

Attempts also have been made to explicitly quantify and propagate geophysical measurement uncertainty for both oil industry and hydrogeology applications. Wolf et al. [1994] use a geostatistical simulation method based on kriging with external drift to estimate and assess the accuracy of the spatially varying thickness of an oil sand, using seismic reflection amplitudes as an imperfect guide. The specifics of the methodology is not described.

Thadani [1994] uses a statistical pattern recognition technique to calibrate seismic reflection amplitude data with geological well data from an oil reservoir. He then uses an indicator geostatistical simulation technique to generate indicator representations of porosity classes, which are constrained to a single value at well locations and to upper and lower bounds away from the wells where seismic data exists.

Moline et al. [1995] use the concept of regionalized classification to classify a suite of geophysical borehole logs from a heterogeneous deep aquifer into electrofacies, based on collocated hydrologic core data. Probabilities of electrofacies class membership are assigned to
well locations containing only geophysical logs, based on electrofacies characteristics, using a statistical process known as discriminant analysis. Then all the data is used in an indicator kriging algorithm to estimate the 3-d distribution of electrofacies occurrence. Finally, permeability and porosity are estimated at each 3-d grid location by probability weighting of the statistical relationship between electrofacies classes and the hydrogeologic parameters.

A multiple linear regression (MLR) methodology is used by Lesch et al. [1995] to predict the spatial distribution of soil salinity, in the form of point or probability estimates, from electromagnetic conductivity measurements. Direct measurements of soil salinity are used to calibrate the electromagnetic data. They also present an algorithm for selecting the optimal locations for MLR calibration sites to minimize the number of sites required.

Rubin et al. [19921 combine exhaustive seismic velocity data with sparsely sampled hydraulic head and hydraulic conductivity data to estimate a single hydraulic conductivity value at every point in the field. The estimation process they use links geostatistical inversion, geostatistical estimation, and a semi-empirical petrophysical relationship between seismic velocity, permeability, and effective pressure.

Copty et al. [19931 improve on the above methodology by taking a rigorous Bayesian approach, wherein:

- all probability density functions (pdf) are fully defined,
- Bayes' theorem is used to update pdfs to conditional pdfs based on new data,
- errors in seismic velocity are permitted (in the form of upper and lower bounds), and
- the final hydraulic conductivity estimate is in the form of an entire pdf along with a measure of estimation uncertainty.

However, their framework still requires a petrophysical relationship between seismic velocity, permeability, and pressure—which is non-unique and highly dependent on the local setting. Also, with their approach the hydraulic conductivity estimate is only conditioned on the seismic velocity value located at the estimation point. The influence of velocity data at neighboring points due to spatial correlation is ignored.
Copty and Rubin [1995] further generalize the methodology for incorporating seismic velocity data in the estimation of hydraulic conductivity ($K$). They employ a statistical calibration between coinciding lithology and seismic velocity log data from wells to compute the probabilistic relationship between the two. The calibration results are, in turn, used to calculate the probability of occurrence for each lithology at every point where there is seismic reflection velocity data. The calculation is based on the velocity pdf at the point of estimation, which they calculate using small perturbation analysis, thus quantifying the uncertainty associated with the seismic inversion process.

They use the lithology probabilities of occurrence along with the direct lithology data to compute indicator covariances for each lithology based on the approach of Alabert [1987], taking advantage of the much larger number of covariance pairs. The covariance models inferred from this process are used in a simple kriging system to condition the lithology probabilities of occurrence at each seismic velocity data point on the direct lithology measurements from well logs.

The final step they perform is updating $K$ at each seismic data point, based on the lithology conditional probabilities and neighboring direct measurements of $K$, by kriging and probability weighting. This process involves, first, conditioning the prior $K$ pdf for each lithology at each point on only $K$ data from the same lithology using a geostatistical updating algorithm, such as kriging. Second, at each estimation point, the updated $K$ pdf for each lithology is weighted by the previously calculated probability of occurrence for that lithology. The weighted $K$ pdfs for all lithologies are summed to construct a single $K$ pdf, conditioned on all information, at that point. This approach assumes that the defined lithologies correspond to distinct and somewhat different $K$ populations and the $K$ field is spatially uncorrelated across lithologies.

Copty and Rubin [1995] is one of the first approaches to use surface geophysics to predict an indirect property (lithology) that is loosely related to the property of interest ($K$), but much more strongly related to the geophysics measurement than the property of interest is related to the geophysics. The idea is to use a two step approach. First use geophysics and other direct
measurements to identify the heterogeneous spatial arrangement of an indirect, secondary property, taking advantage of the high spatial resolution and coverage of geophysics. Then estimate the property of interest using the secondary property information, accounting for uncertainty in the relationship between the two. However, the methodology still ignores seismic velocity spatial correlation and can, thus, only calculate results at locations with seismic or well data.

_Cassiani and Medina, Jr. [1997]_ present a straightforward statistical approach for reducing hydraulic conductivity (K) uncertainty with formation electrical resistivity (ρ) measurements, and illustrate the methodology in a real-world case study carried out in full. In their approach they calibrate an electrical formation factor \( F = \rho / \rho_w \) where \( \rho_w \) is the pore water resistivity), used instead of ρ to reduce the effects of \( \rho_w \) variations on the measurement of ρ, with K using linear regression analysis. Then they invoke a simplified co-kriging procedure to estimate K and its associated uncertainty at every point in the field, conditioned on K and F measurements. The regression analysis and co-kriging procedure are linked through spatial covariance. The F covariance and F-K cross-covariance models are calculated from the K covariance model through a linear scaling based on the regression model parameters. Inference of all random field and regression model parameters (six parameters in total, if a stationary and isotropic Gaussian random field, exponential covariance, and linear regression model are assumed, as the authors did) is performed simultaneously based on all K and F data using the robust and objective maximum likelihood method.

In their case study, _Cassiani and Medina, Jr. [1997]_ perform vertical electrical soundings measuring apparent ρ at a contaminated landfill site. The sounding curves have to be inverted in order to obtain a prediction of the ρ depth section, used in the calibration and K estimation process. This inversion process is inherently non-unique and, although the authors constrain the solution to match well log and depth to bedrock measurements, they do not explicitly account for uncertainty in the interpreted ρ depth section. However, this uncertainty, along with every other
types of uncertainty plaguing the p data, is implicitly accounted for through the regression analysis.

In summary, Cassiani and Medina, Jr. [1997] present a general, robust, and computationally feasible approach for using geophysics to reduce the uncertainty in the spatial distribution of a hydrogeological property, while accounting for geophysical measurement uncertainty. McKenna and Poeter [1995] integrate multiple types of data including borehole geophysics logs, cross-hole seismic tomography measurements, hydraulic test data, hydraulic head observations, and geologic knowledge, in a process they call "data fusion", to generate a probabilistic description of the heterogeneous hydraulic conductivity structure in a shallow aquifer. An electrofacies classification, similar to the approach taken by Moline et al. [1992], based on a suite of geophysical borehole logs (with greatest emphasis on the sonic velocity log) and core samples is employed to identify electrofacies groupings. These groupings are then examined in terms of their similarities and combined if they are found to be similar based on geology, discriminant analysis, or air permeameter measurements performed on core samples. Discriminant analysis is used to determine, for all locations with borehole geophysics data, probabilities of class membership into the redefined groupings, now referred to as hydrofacies.

Seismic velocity data calculated from the inversion of cross-borehole seismic tomography surveys are used as imprecise estimates of hydrofacies type, based on a calibration with velocities measured from borehole seismic logs. The authors make the assumption that the hydrofacies can be perfectly distinguished based on seismic velocity thresholds, even though the range of velocity values within each hydrofacies—measured from sonic logs—overlap significantly. Realizations of the hydrofacies distribution within the aquifer, conditioned on borehole geophysics data and seismic velocity from cross-borehole tomography, are geostatistically simulated using the Markov-Bayes (M-B) method [Alabert, 1987]. With the M-B method the soft auto-covariances and hard-soft cross-covariances are calculated from the hard covariance by multiplying it by a scaling factor, which is determined from the calibration.
The final step in the process is to use inverse groundwater flow modeling to eliminate hydrofacies realizations that are inconsistent with hydrogeologic data. The hydrofacies distribution within each realization act as zones in the inverse flow model. For each zone the model determines a single hydraulic conductivity (\( K \)) value, that is optimized on the differences between measured and predicted hydraulic head values (the head residuals) being as small as possible. If the results of the inverse modeling do not meet a set of defined criteria, that realization is eliminated from the set of realistic realizations. The criteria for culling out realizations are inherently subjective—based on the acceptable magnitude of head residuals and hydrogeological judgment. The desired effect of the realization elimination process is a reduction in uncertainty regarding the true nature of \( K \) heterogeneity in the aquifer.

The innovativeness of the approach by McKenna and Poeter [1995] is the ability to effectively combine disparate types of direct and indirect, precise and imprecise, data in a probabilistic description of site conditions, resulting in enhanced uncertainty reduction and a better understanding of subsurface heterogeneity.

2.5.4 Incorporating geophysics in hydrogeological decision-making framework

Previous work approaching the application of geophysics in near-surface problems from a risk-based decision-making perspective is limited. Parks and Bentley [1996] touch upon the potential "worth" of ground conductivity surface geophysical measurements at oil and gas production flarepit sites that potential groundwater salinity contamination; they assess the comparative reduction in uncertainty versus cost, comparing the geophysics with drilling, sampling, and chemically analyzing groundwater. They use full ordinary co-kriging of total dissolved solids (\( TDS \)) data from wells and ground conductivity (\( C \)) data from an electromagnetic (EM) survey to estimate the most likely value and estimation variance of \( TDS \) at every point on a two-dimensional grid surrounding a flarepit area. This particular geostatistical estimation algorithm is considered to be a statistically robust method for incorporating measurements of a secondary variable (e.g. \( C \)) into the spatial estimation of the primary variable
(e.g. TDS). The main drawback of the algorithm is that, in addition to the need to infer the underlying probability distribution and auto-covariance model for the primary model, inference of the secondary variable auto-covariance and primary-secondary cross-covariance models is also required. Reliably inferring these models requires a large amount of data and is computationally intensive, especially when there is more than one secondary variable. The authors assume a normal probability model (required for proper estimates of estimation variances) and spherical covariance models, and determine the parameters defining these models based on statistical analysis of the data.

The data worth of EM measurements relative to drilling additional wells and taking groundwater samples for this application is demonstrated by comparing the cost of performing an EM survey to the cost of drilling enough additional wells to reduce the uncertainty in TDS concentrations across the site to the same level as that resulting from the EM survey. The overall uncertainty associated with each site characterization strategy is defined as the average of all estimation point variances calculated from kriging. Since the estimation point variance is independent of the estimated point value, the uncertainty reduction resulting from drilling can be estimated by sequentially adding fictitious boreholes. The location chosen for the fictitious borehole is the point of maximum point variance in the study area and a new estimate of the spatial distribution of variance is made by kriging. The process is repeated until the average variance is equal to or less than that calculated from the EM data. Based on the approximation that the total cost of one well is equal to the cost of an EM survey over the entire area, the authors found that it costs about ten times more to reach the same amount of uncertainty reduction by emplacing additional wells compared to performing the geophysics.

In summary, Parks and Bentley [1996] use a rigorous co-kriging algorithm, that can account for and propagate most types of uncertainty, to incorporate geophysical measurements into an example site investigation, and illustrate a simple method for comparing the economic data worth of different sampling strategies based on the kriging variance. Although they address their problem from a probabilistic decision-making framework, in doing so, they take a very
simplified approach wherein economic risk is assumed to be directly related to uncertainty in TDS. A more realistic approach is to base data worth on the economic benefits to the decision process resulting from risk reduction, which in turn is related to uncertainty reduction. In addition, uncertainty in TDS is assumed to be perfectly defined by the kriging variance, regardless of the values of surrounding data.

James and Freeze [1993] assess the worth of additional data from wells (in different configurations) versus a seismic reflection survey covering the entire study area for determining aquitard continuity in a contaminated groundwater management problem. The data worth is based on the economic benefits the data provides to the engineering design process through risk reduction. The analysis is performed within a risk-based decision-making framework. The framework links together an indicator geostatistical model, a contaminant transport model, and an economic risk-cost-benefit model.

The geostatistical model is used to create equally-likely realizations of the aquitard continuity based on all types of precise and imprecise data, using a two step approach. First the parameters defining the probability model (e.g. mean, spatial correlation length) are updated based on new data using a variant of Bayes' equation and classical statistics. Then a Monte Carlo geostatistical simulator based on the Markov-Bayes algorithm [Alabert, 1987] is invoked to generate the multiple realizations conditioned on all data. The data consists of aquitard thickness picks, made from geological and geophysical logs (the former considered imprecise and the latter precise) at existing wells, and picks resulting from the drilling of additional hypothetical boreholes. The information extracted from the data is simplified to a question of aquitard existence—either the aquitard exists or doesn't at a particular location; the actual thickness is considered irrelevant. A set of realizations has to be simulated for all possible values of the hypothetical data in order to carry out the data worth analysis.

The contaminant transport model is used to model the groundwater transport of contaminants from source areas for each geostatistical realization, determining whether or not
contamination breaches the aquitard and pollutes the underlying aquifer within a regulatory time horizon. The former represents failure, resulting in considerable extra cost.

Then the economic risk-cost-benefit model is used to compare the expected total cost (the average of the total cost for all realizations) of each possible contaminant management design alternative, for each sample outcome. The design with the lowest expected cost is selected and the same analysis is performed for all sample outcomes. The expected value of the design costs taken over all sample outcomes is calculated by taking the average of expected costs of the designs selected previously, where the expected costs are weighted by the probability of that sample outcome occurring. The final value represents the expected cost of management when that particular sampling strategy is implemented. The expected cost of management for different sampling strategies can then be compared with each other, and with the expected cost for no additional sampling, to estimate the relative and absolute economic worth of the sampling proposals.

While James and Freeze [1993] follow a very rigorous, robust, and general approach for assessing the economic worth of different patterns of additional boreholes, their approach for analyzing the worth of a surface geophysical survey is much more simplistic—reliant on assumptions and problem-specific. By requiring that the geophysical survey must cover the entire area where a window in the aquitard would cause a failure that alters the choice of best design alternative, the updating problem simplifies to a direct application of Bayes' equation. In their approach, the geophysical survey has only two possible outcomes—a failure-inducing window exists or does not exist, both of which are not necessarily accurate predictions, due to the imprecision of the geophysical measurement. The probabilities required for this data worth assessment—the probabilities of contaminant control failure for the two possible geophysics outcomes—are calculated using Bayes' equation; required inputs to the equation are the probability of failure prior to performing (hypothetical) geophysics and the two likelihood probabilities of (1) sampling failure given failure and (2) no failure actually occurring in reality.
The two likelihood probabilities characterize the imprecision of the geophysics survey and have to be inferred through a calibration with collocated precise data or subjective expert opinion.

In summary, James and Freeze [1993] effectively, and efficiently, assess the worth of performing a seismic reflection survey, used to identify aquitard windows detrimental to contaminant management, by calculating the worth of the survey over a range of precision levels (by varying the likelihood probabilities; in this way the level of precision required for the survey to be worthwhile is estimated. However, while their approach is appropriate for their specific problem, the methodology would not work for many other types of problems due to its large assumptions. These assumptions include a problem risk defined by a single criteria for failure or no failure, with only one possible cause of failure, and the assumption that geophysics has the ability to distinguish between the two. For example, their methodology will be insufficient if

- several failure mechanisms exist, such as high hydraulic conductivity channels in the upper and lower aquifers, in addition to aquitard windows in the aquitard that separates them, or
- failure is dependent on a combination of related factors, such as the size and/or arrangement of separate windows, or
- several levels of failure exist.

Another limiting assumption is that the geophysics survey must cover most of the site, an often difficult task to accomplish, due to on-site physical impediments and the presence of sources of noise to the measurement.

The work of James and Freeze [1993] represents a very comprehensive framework for assessing and comparing the economic worth of different types of data in a risk-based decision-making context. Their general approach is closely followed in this research, but much more emphasis is placed on rigorously incorporating geophysical measurements in the framework in a general way, applicable for a variety of problem types and geophysical techniques.
This chapter examines the application of geophysics to hydrogeology-related problems, with the main emphasis on its use in site characterization. The chapter begins by presenting a brief introduction to geophysics. Next, the specific types of applications that geophysics can be, and in many cases have been, used for is described. The advantages geophysics can provide in these types of applications, especially in comparison to other measurement types, are then discussed, as well as the inherent disadvantages and difficulties associated with using geophysics. In the discussion on the difficulties associated with applying geophysics, special attention is given to the indirect nature of geophysical measurements and ways to cope with this. Appendix A provides a detailed and thorough review of the geophysical inverse problem—the process of deriving subsurface properties of interest from geophysical measurements—including methods of solving it, and methods of accounting for the uncertainty associated with it.

The primary purpose of this chapter is to provide the reader, who may not be familiar with geophysical methods, an introduction to geophysics and its use in the context of near-surface problems—a foundation for the remainder of this work, which integrates geophysics and near-surface site characterization in a rigorous and cohesive risk-based decision-making framework.

### 3.1 Introduction to geophysics

Geophysics applies to a class of measurements wherein some type of energy field emanating through/from a medium of interest (e.g. the subsurface) is measured by an instrument. The measured energy can be either a potential energy field, such as gravitational/electrical potential energy, or a kinetic energy field, such as impinging atomic/subatomic particles (radiation), and can vary on different time scales. The energy field measured by a geophysical recording instrument (receiver) is either naturally-occurring (e.g. the earth's gravitational, magnetic fields) or artificially generated by an external, manmade source (e.g. electrical current,
electromagnetic field generators). The energy source (transmitter) and receiver can be located separately or together, the latter only possible if the source is manmade. Depending on the energy field type and source-receiver configuration, the source energy is modified in some way as it travels through the carrying medium to the receiver.

The extent to which the energy field is modified between the source and receiver employed in geophysics is dependent on the physical properties of the field itself (e.g. the inverse-squared decay in gravitational potential strength with increasing distance), as well as the value and spatial distribution of certain material properties along the energy field's travel path. Therefore, geophysics is potentially useful for subsurface characterization because characteristics of the recorded energy field are dependent on geophysical material properties of the transmitting medium, which may be of interest themselves or through some relationship with another property. For example, with electrical/electromagnetic geophysical methods, the magnitude of the electrical potential (or voltage) is dependent on electrical conductivity/resistivity of the transmitting medium. The electrical conductivity/resistivity structure of the subsurface may be of interest itself (e.g. for characterizing total dissolved solids in the groundwater) or through its relationship with another property (e.g. porosity).

Geophysical measurements are remote-sensing measurements in that information about the property of interest is obtained by sensing a transmitted signal without the instrument being in direct contact with the material being sampled. Limiting ourselves to solid-earth geophysics, where the medium of interest is the subsurface, surface and borehole geophysics represent methods where the measuring device that records the incoming signal is located above (or on) the ground surface which overlies the medium of interest and within a borehole penetrating the medium, respectively. Borehole geophysics are still remote-sensing measurements, even though the borehole is in contact with the subsurface, since the measurements obtain information about the substrate beyond the borehole wall. The fact that surface geophysics requires no major disruption of the subsurface—removal of material or significant penetration into the subsurface)—makes these measurement methods non-invasive. Of course, this is not the case for
borehole geophysics, since a borehole has to be emplaced to apply these methods. However, it can be argued that borehole geophysics is less invasive than direct sampling methods (e.g. well-cores), since in order to sample the same volume of subsurface as borehole geophysical methods direct sampling would require the removal of a much larger amount of material.

Most geophysical techniques do not directly observe the parameter of interest, but instead observe some other state variable from which the parameter of interest must be somehow derived; in this way geophysical measurements are indirect measurements. There are two sources of indirectness in geophysics:

1. spatial filtering, where the measured signal response represents some sort of average across a volume of substrate that is at a larger scale than the scale of interest, and
2. measuring a property that is not the property of interest, but is somehow related to it.

(See Section 2.2.2 for further elaboration on the measurement process.) Depending on the final information desired for a particular application, the second source of indirectness often has several layers, wherein (1) the geophysical parameters must be derived from the observed state variables and (2) the material property of interest must be estimated from the derived geophysical parameters.

As an example of these types of indirectness, consider ground-penetrating radar (GPR). GPR measures reflected electromagnetic energy versus time resulting from the interaction between a surface-transmitted source signal and variations in subsurface electromagnetic properties. The response signal represents a spatial average of subsurface conditions since the source signal does not have an infinitesimally small bandwidth and, thus, acts as a filter. Ideally, it would be desirable (although not necessary) to resolve variations in the electromagnetic properties at a scale smaller than the source signal bandwidth. In addition, the response signal is measured versus time when, in actuality, the response versus depth is the desirable result. In order to transform the recorded section from time to depth it is necessary to ascertain the electromagnetic velocity structure of the subsurface, which is related to the change in travel-time versus offset between source and receiver—a quantity that can be measured. A third layer of
indirectness exists if the true parameter of interest is, say, hydraulic conductivity, that is somehow related to electromagnetic velocity through a complex, non-independent relationship.

As a result of their indirect nature, geophysical measurements usually do not directly provide information about the material property of interest; this information must somehow be extracted from the observed state variables. Hence, in order to obtain useful information from the measurements, it is first necessary to understand the relationship between the observed state variable and the property of interest. As discussed above, the indirectness associated with geophysics is often multi-tiered and, thus, it is often (but not always) necessary to define the relationship between known and desired variables at each level of indirectness. Once these relationships are defined there is still the problem of using them in such a way that the greatest degree of valid information about the property of interest can be ascertained.

It is important to realize that these relationships between the measured geophysical state variable and the property of interest are rarely one-to-one and/or independent (i.e. the two are rarely perfectly correlated and the geophysical observations are often dependent on other properties, in addition to the one of interest). Hence, there are usually uncertainties associated with geophysical estimates of material property values as a result of the indirect nature of geophysical measurements.

Another concern associated with the indirect nature of geophysics is how sensitive the measurement is to variations in the property of interest. As this sensitivity decreases and/or uncertainty associated with the relationship between the observed state variable and property of interest increases, the viability of the geophysical technique for estimating the parameter of interest decreases. This occurs because the signal-to-noise ratio—the degree of variation in the response signal originating from actual variations in the parameter of interest compared to the degree of variation due to other sources—decreases.
3.2 Applications of geophysics in hydrogeology-related studies

There are many possible applications for geophysics in near-surface problems, but the actual effectiveness of using geophysics for these applications is highly technique-, problem-, and site-dependent. In the most general sense, geophysics can be used to either:

- delineate subsurface geometry (e.g. anomalous conditions, large-scale geologic/hydrogeologic structure) or
- estimate subsurface material properties (e.g. porosity, hydraulic conductivity, chemical concentration).

These two types of geophysical applications require different levels of information to be extracted from the measurements. In order to delineate subsurface geometry, the change in subsurface properties producing the geometric feature must be recognizable in the geophysical response, implying that the geophysical technique is sensitive to contrasts in those properties and it can resolve the feature. This application does not require an exact quantitative estimate and, thus, only relative changes in the geophysical response are important, making the data interpretation more qualitative than quantitative. In contrast, estimating material properties requires that the final estimate generated from the geophysics be in precise quantitative terms. This requirement entails that the geophysical measurements must be sensitive to changes in the material property being estimated and that the relationship between the measured state variable and the property of interest is known with reasonable precision and accuracy. Hence, using geophysics to estimate material properties is a much more demanding application than delineating subsurface geometry, since the final measurement output must be in exact, quantitative terms—requiring a much higher level of inference from the data. This does not imply the geophysical estimate has to be perfect, though, since uncertainty can be accounted for in quantitative terms.

Investigating applications of geophysics from the organized site characterization perspective presented in Section 2.4, geophysics could potentially be useful during the preliminary exploration and detailed investigation temporal stages. Geophysics is more likely to
be useful (i.e. has a higher likelihood of success) during the preliminary exploration stage, since it is often plagued by measurement uncertainties and the emphasis during that stage is on delineating large-scale features, as opposed to collecting quantitative information about detailed, small-scale features and material properties—the emphasis of the detailed investigation. For this reason, most of the real-world applications of geophysics to near-surface problems are associated with the preliminary exploration stage of site characterization. This is not to say that geophysics cannot be used for quantitative applications—in fact it is used for such an application in the real-world case study of this work—just that there is much greater difficulty in doing so compared to qualitative applications; this a consequence of the characteristic indirect and imprecise nature of geophysical measurements.

From the problematic site characterization perspective, geophysics has the potential to be very useful for performing search and pattern recognition functions due to the typically high density and areally-extensive spatial coverage of geophysical measurements, and the fact that precise quantitative estimates are not required to successfully carry out these functions. The high spatial coverage of geophysics significantly increases the likelihood of locating targets and reduces the need for reconstruction of large-scale features, in comparison to the limited coverage provided by direct-sampling methods (e.g. borehole coring).

Specific geophysical applications in hydrogeology-related problems can be grouped into five categories:

1. geologic characterization,
2. hydrogeologic characterization,
3. contaminant delineation,
4. locating buried objects, and
5. measuring time-varying processes.

A very common geophysical application is delineating large-scale geologic structure such as stratigraphy and discontinuities (e.g. fault displacements, pinch-outs). Pattern recognition is inherent in this process since in order to delineate structure the contrasts in material properties
defining the structure must be identified. This requirement is usually addressed in a non-rigorous manner by defining a geologic classification scheme (or using a pre-existing one) based on certain material properties (e.g. grain size, lithology) and identifying the relationship (if any exists) between geophysical response and the geologic class differentiation. Another much more demanding geologic characterization application of geophysics is quantifying small-scale variations in geologic material properties (e.g. mineralogy, porosity, density).

Since geology usually varies much more in the vertical than horizontal direction (e.g. uniform vertical layering of stratigraphy), geophysical methods with good vertical resolution are usually more appropriate for geologic characterization. For highly discontinuous geology horizontal resolution may be important as well. Other considerations in determining which geophysical techniques are potentially useful for geologic characterization are depth of investigation and, of course, how sensitive the geophysical response is to the changes in geology that are of interest, higher sensitivity being more desirable.

As with geologic characterization, geophysical applications for hydrogeologic characterization can be separated into delineating large-scale features, usually boundary conditions in a hydrogeological system, and quantifying small-scale variations in hydrogeologic material properties, the latter application being considerably more difficult. Common geophysical applications are determining the depth to the water table and the depth/thickness of aquitard/aquifer layers. Vertical resolution and sensitivity to properties whose contrasts define these hydrogeologic boundaries are the main concerns in determining the applicability of geophysical techniques for these uses. Another more difficult application is identifying fracture zones and the fracture orientation within these zones. Identifying individual fractures can be considered a small-scale application and only high resolution borehole geophysical techniques would have much chance of success. Other small-scale hydrogeologic properties that geophysics might help to quantify are porosity, pore water content, pore water chemistry (e.g. total dissolved solids), and hydraulic conductivity, although these applications require high accuracy and precision in geophysical estimates for success. Another hydrogeologic application for
geophysics, in a class by itself, is predicting the direction (and possibly magnitude) of subsurface water movement (e.g. locating discharge zones).

Geophysical uses for contaminant delineation include identifying large-scale anomalies caused by contaminants (e.g. source zones, plume geometry), as well as obtaining quantitative estimates of contaminant concentration, the latter application being considerably more difficult. Contaminant delineation could be classified as a component of hydrogeologic characterization, but is considered separately here due to its importance in hydrogeology-related problems.

Probably the most common geophysics application in near-surface problems is locating buried objects. The types of buried objects that have been located using geophysics include utilities, pipelines, and buried waste (which can be a source of subsurface contamination). Geophysics is particularly well suited to this application since geophysics is usually effective at identifying anomalies in the subsurface caused by large and sharp contrasts in physical properties—as these buried objects often cause—that influence the energy field being measured. The actual value of material properties is often difficult to quantify using geophysics, but this is not required for simply locating buried objects. This task can be accomplished if the buried object causes a detectable anomaly in the geophysical response and the anomaly can be recognized as being associated with the buried object of interest.

Geophysics can be used to measure time-varying processes that vary on relatively short time scales since the acquisition of geophysical data is usually very fast. An example of this type of application is using geophysics to monitor leakage of hazardous liquids from an underground storage tank. Geophysics is well-suited to this type of application since, as mentioned repeatedly, it is much more effective in identifying relative changes in subsurface conditions—commonly referred to as “measurement precision”—than obtaining accurate quantitative estimates of properties—commonly referred to as “measurement accuracy”. Temporal changes in subsurface conditions can be identified by comparing background geophysical measurements with new, up-to-date measurements, highlighting the differences in the geophysical response. Such an approach was taken by Brewster et al. [1995] to monitor the migration of dense non-aqueous
phase liquid (DNAPL) through the subsurface over a period of time using ground penetrating radar.

An important supplementary application of geophysics, which supports all the previously described site characterization operations (geologic/hydrogeologic characterization, contamination delineation, and locating buried objects), is to guide and optimize the placement of more exact, invasive measurements (e.g. borehole measurements). This is potentially a very valuable application since geophysics often provides relatively inexpensive, easily acquired, and spatially-extensive data, but the measurements tend to be unreliable and inexact and, thus, incapable of providing precise quantitative parameter estimates. In contrast, more precise measurements that can provide quantitative estimates are usually much more expensive, difficult to obtain, and spatially-limited in comparison to geophysical measurements. Thus, in a two-stage process, geophysics can be used first to identify potential areas of concern or anomalies in the subsurface across a site and, hence, direct the placement of hard measurements to priority areas, which are used in the second stage to obtain quantitative estimates of important material properties.

Finally, a more general and recently extolled geophysics application is to aid in the development of a probabilistic site conceptual model by helping to infer the form(s) of the spatial uncertainty model(s) used in the conceptual model. Spatial uncertainty models account for uncertainty in the spatial distribution of a parameter by treating the parameter as a probabilistic random variable, constrained to a finite range of possible values based on an assumed and/or inferred structure. The structural constraints are usually applied in the form of a probability and spatial correlation model (e.g. lognormal probability model and exponential spatial correlation model) for the variable, that are defined by a limited number of structural parameters (e.g. mean, variance, correlation length). These models and their structural parameters need to be inferred from existing information and knowledge at the time. (The topic of spatial probability models and methods for inferring their structure is covered in CHAPTER 4.)
Geophysics can significantly aid in the model inference process by providing a large number of densely-spaced and spatially-extensive data. Data play an important role in the process by providing a means to validate/invalidate hypothesized models, and data with a wide range of separation distances between data locations—as is often the case with geophysics data—is especially useful for gaining incite into the nature of spatial correlation across a wide range of scales. The main difficulty in using geophysics for this purpose is accounting for the uncertainty associated with the data obtained from indirect geophysical measurements.

Several researchers have recently used geophysics to aid in the process of inferring a spatial uncertainty model. Rae and Knight [1995] use ground penetrating radar (GPR) data to deduce the correlation model for horizontal continuity of geologic beds in a glacio-fluvial sedimentary environment, assuming that the GPR response is indicative of changes in geologic bedding. Copty and Rubin [1995] use soft estimates of lithology—derived from seismic reflection data—along with hard lithology data from well logs, to compute a weighted, significantly-improved (as compared to just using hard data) estimate of the lithology indicator covariance function, using the Markov-Bayes model [Alabert, 1987]. Thus, the seismic geophysics enables the inference of a more accurate lithology spatial-correlation model. James and Freeze [1993] use soft estimates of aquitard presence (whether a particular aquitard layer is present or not at a location) from geological well-logs, hard aquitard presence data from borehole geophysical well-logs, and prior geologic knowledge about the continuity of the aquitard in the area to estimate the indicator mean of aquitard continuity (i.e. the probability that a window exists in the aquitard) across a study area. They employ a Bayesian estimation approach. In all these studies geophysics greatly improves the ability to confidently estimate probability model structural parameters.

3.3 Advantages of geophysics

The advantages provided by geophysics in subsurface investigations stem from the spatial measurement coverage and density that can be obtained from geophysics, as well as the non-
invasive nature of the measurements. Whereas the amount of subsurface sampled by direct borehole sampling techniques (e.g. well-coring) is necessarily restricted by the narrow circumference and, thus, small sampling volume of boreholes, coupled with an economic constraint on the number of boreholes that can be emplaced, surface geophysical techniques can exhaustively sample the subsurface underlying a large portion of a site in a comparatively short period of time, within economic constraints. Surface geophysics has the ability to accomplish this because the sampling technique is based on a remote sensing and, thus, non-invasive measurement, that enables rapid acquisition and, thus, extensive measurement coverage at a high density. As an illustration of this attribute, consider the example problem, where the approximate areal extent of the site is approximately 16,000 meters squared (200 m² x 80 m²). Assuming the entire site is accessible and amenable to ground-penetrating radar (GPR) measurements, it is conceivable that the entire site could be sampled by a GPR survey on a one meter by one meter areal grid spacing to a depth of up to ten meters, with a vertical sampling density on the order of tens of centimeters, all within the time period of a week. This results in some 1.6 million data points. In contrast, it would probably not be feasible (due to economic and time constraints) to drill/dig more than a hundred boreholes/test pits across the site; this in itself would, undoubtedly, take much longer than a week.

Although borehole geophysical methods are invasive—they require the existence of a borehole—and, hence, cannot provide the extensive spatial coverage of surface geophysics, they can still provide significantly better sampling coverage than direct sampling methods. This is because they have the ability to investigate tens of centimeters (sometimes more) beyond the borehole wall with a vertical sampling density on the order of tens of centimeters.

In order to understand the overall advantages that geophysics can provide to a problem, it is important to investigate how geophysics can influence, through increased measurement coverage, the decision-making process. From a problematic site characterization perspective (see Section 2.4.2.2 for a description of this perspective), better coverage increases the chance of identifying anomalous subsurface conditions when performing the search function, and reduces
the need for performing reconstruction (inference of conditions/properties beyond measurement locations, commonly referred to as spatial estimation) since a significantly higher percentage of the subsurface is sampled. Referring to the GPR example above, it is easy to see that the extensive site coverage provided by GPR would greatly increase the chances of identifying subsurface contaminant source zones (e.g. pure phase contaminant pools, buried drums), in comparison to the limited coverage provided by borings, assuming GPR can effectively recognize these zones. In addition, the high density of the GPR measurements would reduce the need for reconstructing unmeasured parts of source zones.

The advantageous influence of geophysics on the search and reconstruction site characterization operations translates into a reduction in the uncertainty due to ignorance associated with a problem. By increasing measurement coverage, geophysics reduces the characteristic network scale (separation between measurements—see Section 2.2.1), thus increasing model resolution (by reducing the smallest resolvable scale of variability in the property of interest). Of course, the amount of uncertainty reduction (if any) depends on the accuracy and precision of the geophysics data. Depending on the decision-making context of the problem and the impact of geophysics on uncertainty, this uncertainty reduction can reduce the risk of unintended decision outcomes—improving decision-making by enabling better, more cost-effective decisions to be made and, thus, benefiting the overall project.

Additional benefits provided by geophysics to hydrogeology-related problems include:

• minimization of the potential health risk posed to site workers responsible for taking measurements, by eliminating the need to remove subsurface materials (e.g. the health risk associated with exposure to subsurface contaminants),

• elimination of the potential for alteration of the natural hydrogeological system due to subsurface intrusion (e.g. introducing a pathway along a borehole between two aquifers), and

• fast acquisition times, enabling its use in applications where the measurement of processes which vary on relatively short time scales is important (e.g. monitoring leakage from underground storage tanks).
3.4 Difficulties associated with applying geophysics

There are two major factors that complicate the application of geophysics to subsurface characterization problems: (1) the multiple layers of indirectness in geophysical measurements and (2) the susceptibility of the measurements to noise and low signal-to-noise ratios as a result of adverse site conditions. In addition to complicating the process of estimating the property of interest from geophysical measurements, these factors can potentially contaminate the estimates with significant errors, making geophysics an inherently "soft" measurement. This section discusses the nature, causes, and consequences of these complicating factors and summarizes general approaches for overcoming the difficulties they cause, highlighting the approaches taken in this work and stressing the importance of accounting for uncertainty in the geophysics measurement process.

The source of indirectness in the geophysical method is multifaceted. In the most general sense, indirectness results from observing a state variable which is not directly of interest itself, but is somehow related, possibly through a complex linkage, to the actual property of interest (see 2.2.2 for a general description of the measurement process). With geophysics the relationship between the measured state variable and the parameter of interest can be especially circuitous. Geophysics usually measures a technique- and instrument-dependent state variable (e.g. "apparent" electrical resistivity in direct-current resistivity techniques) that is somehow related to a subsurface geophysical property (e.g. the "true" subsurface electrical resistivity) through some measurement (spatial-filtering) process. However, the application interest is often a material property (e.g. porosity), not the geophysical property that is often only indirectly related to the material property. In other words, the physical property that the geophysical measurement directly responds to is not necessarily of interest in a problem, but is somehow related to the material property that is of interest.

The types of indirectness associated with a particular geophysical measurement, its consequences on a specific application, and ways to deal with it vary significantly with the
Chapter 3. GEOPHYSICS IN HYDROGEOLOGY PROBLEMS

geophysical technique being employed and the application it is being used for. For example, if a surface electromagnetic method is being used to locate buried metallic waste, it may be possible to base the delineation strictly on an interpretation of the anomalies seen in the raw measurements of "apparent" electrical conductivity (\(\rho_a\)). However, if the application is to estimate subsurface hydraulic conductivity \((K)\), the apparent electrical conductivity measurements would have to somehow be related to \(K\) at specific locations in the subsurface. The linkage between \(\rho_a\) and \(K\) is multi-tiered: \(\rho_a\) is directly dependent on the subsurface true electrical conductivity (geoelectrical) structure and the geoelectrical structure is (possibly) indirectly related to \(K\). Therefore, the \(K\) distribution has to be back-calculated from \(\rho_a\), either directly in one step or in two separate steps:

1. first determine the geoelectrical structure based on \(\rho_a\).
2. then derive the \(K\) distribution from the geoelectrical structure).

As alluded to earlier at the beginning of this chapter, in order to estimate the property of interest from the observed geophysical state variable, it is first necessary to define the relationship between the two using an assumed physical, empirical, or statistical model. Before defining this relationship, though, it is critical to understand the nature of the dependencies between the key variables—the observed state variables, pertinent geophysical properties, and material properties of interest. The observed state variable(s) represents the geophysical instrument response to a particular geophysical property (or properties). It is important to understand that the geophysical measurement is not responding to (and, thus, is not directly related to) a material property (properties), but instead a geophysical property (properties)—that is potentially related to a material property (properties) indirectly. Ideally, one would want to combine the two separate dependencies and define how the material property of interest is directly related to the geophysical instrument response. However, depending on the nature of the instrument response for a particular geophysical technique and the priorities associated with the application the technique is being used for, it is not always appropriate or possible to combine the two linked dependencies, as will be explained below.
The two above-mentioned dependencies will now be examined in greater detail, with the purpose of investigating how they can be defined and used to ascertain information about the property of interest from the geophysical measurement. To apply geophysics to engineering problems in a risk-based decision-making framework—as advocated in this work—it is necessary to quantitatively account for uncertainty in the information provided by geophysics; this uncertainty can be significant as a result of the often high degree of indirectness in geophysical measurements. Hence, an emphasis and high importance is placed on the ability of the methods discussed in this section to account for uncertainty in the property value estimation process. First, in Section 3.4.1, the relationship between the geophysical instrument response and the geophysical property will be considered, followed by the dependence between the geophysical and material properties (Section 3.4.2). Methods that can combine the two linked dependencies into one—relating the geophysical instrument response directly to the material property of interest—and the implications of such a simplification are highlighted in the discussion.

3.4.1 The relationship between instrument observations and geophysical properties

Usually a geophysical instrument does not directly measure a geophysical property, but instead another variable that is a function of the geophysical property. For example, the direct current (DC) resistivity technique measures voltage, which is proportional to resistivity; the proportionality constant is a function of input current and electrode spacing. In addition, the instrument response is inherently a spatial average of the property being measured. This is especially true for surface geophysics since it is a remote-sensing measurement; measurements of the subsurface are made from a remote location on the surface, requiring the response signal to travel through a large volume of subsurface material. Therefore, in mathematical terms, the geophysical observable(s) represents a dependent variable(s) that has some functional dependence on the geophysical property (or properties), which represents an independent
variable(s); the dependency is determined by the instrument-dependent measurement process (see Section 2.2.1.3 for a discussion of the measurement process).

Defining the dependency of the instrument response (observed state variable(s)) on the pertinent geophysical property (or properties) requires modeling the measurement process in some way. Ideally, this would be accomplished by developing from first principles a forward physical model—a mathematical representation wherein the dependent variable is defined in terms of the independent variable—of how the energy field associated with the particular geophysical technique is transmitted, modified by the medium through which it travels, and recorded by the instrument. Then, since the unknown of interest (the geophysical property) is the independent variable and the known (observed state variable) is the dependent variable in the model formulation, the process of estimating spatially-distributed values of the pertinent geophysical property from measurements is a mathematical inverse problem. The use of a physical model to describe the relationship between the geophysical property and instrument response is ideal, in that it exactly replicates the measurement process and, thus, is potentially more informative than the other alternatives—empirical and statistical models—that attempt to imitate the true physical relationship using statistics and, thus, possibly neglect certain characteristics. However, if the physical model does not accurately represent the measurement process, potentially significant errors can be introduced in the estimation of the geophysical property distribution anyway.

The following subsections discuss the three types of models—physical, empirical, and statistical models—that can be used to describe the dependency between the observed geophysical state variable and pertinent geophysical property. The emphasis is on how these models are used to estimate geophysical property values, their ability to account for uncertainty in the estimation process, their advantages, and their limitations. First the geophysical inversion problem—the problem that must be addressed when using a physical model—is briefly discussed, followed by a discussion of empirical and statistical model approaches. A much more in-depth discussion of the geophysical inverse problem is provided in Appendix A.
3.4.1.1 Physical model – geophysical inversion

In order to incorporate a physical model in the geophysical property estimation process it is necessary to solve an inverse problem; this, at least, requires the ability to find a tractable and valid solution to the forward problem (calculating a hypothetical instrument response from a given geophysical property distribution). There are two principle objectives in geophysical inversion: (1) to transform measurements from the observed state variable to the geophysical property and (2) to reverse the effects of spatial filtering (or, in mathematical nomenclature, “measurement deconvolution”) in order to obtain higher resolution information, than the measurement support scale provides, about the geophysical property’s spatial structure. The priorities placed on these objectives for a particular application should be used to determine how the solution of the associated geophysical inverse problem is approached.

In discussing how the above objectives are achieved in geophysical inversion it is important to distinguish between local inversion, where each measurement is treated separately in the inverse problem, and global inversion, where all measurements are treated together. With local inversion only one value of the geophysical property is estimated within the locale of the single measurement being inverted and, thus, the scale of the estimate is determined by the support scale of the measurement; no attempt is made to deconvolve the spatial filtering associated with the measurement process. In contrast, with global inversion the geophysical property distribution across a spatial domain containing all the measurements is estimated simultaneously, taking advantage of the fact that the spacing between measurement locations may be smaller than the measurement process support scale. This approach enables a possibly improved resolution of the geophysical property spatial structure to a scale as small as the network scale (the characteristic spacing between measurements—see Section 2.2.1.4) and, thus, addresses the second objective mentioned above.

\[1\text{In fact, according to sampling theory, it is impossible to reconcile scales smaller than the support scale of the measurement if only one data value is used in the inverse problem [Beckie, 1996].}\]
Therefore, in terms of gleaning the most information as possible about the geophysical property distribution from measurement inversion, approaching the inverse problem from the global (inversion of all measurements at once) as opposed to local (inversion of each measurement separately) perspective is advantageous (assuming the network scale is smaller than the measurement scale). However, if the model scale of the overall engineering problem—the scale of parameterization used for solving the engineering problem at hand (e.g. the size of blocks used in a groundwater flow and transport model of contaminant migration from a proposed landfill)—as determined based on the economic/technical limitations and the decision-making context of the overall problem (see Section 2.2.1.2), is equal to or larger than the measurement scale, then there is probably no added advantage to performing global inversion. For this situation, local inversion of each measurement separately can be used to obtain a description of the smallest scale of spatial variability required in the engineering problem (the model scale)—the advantage of local over global inversion being that the solution is often much easier to obtain.

Geophysical inversion can be complicated by many factors. As a result of the remote-sensing nature of geophysics, the measured energy field/pulse possibly travels through, and is therefore potentially influenced by, a large volume of the subsurface; it can be very difficult (or impossible) to decipher the exact circumstances (i.e. spatial distribution of geophysical properties) producing a measured response. This circumstance, common with surface geophysical methods, is illustrated in Figure 3.1, where the source energy travels through a large volume of subsurface material before reaching the receiver to provide a single measurement. If the goal of the geophysical survey is to estimate the geophysical property distribution at a scale of variation smaller than the distance between the source and receiver (i.e. the measurement scale), then, in order to provide this resolution, measurements have to be collected at a spacing less than or equal to the smallest scale of interest and global inversion must be performed. (According to Nyquist sampling theory, in order to reconstruct the variability of a function at a certain scale two or more points per scale length, or period, need to be sampled.) However,
measurements cannot be taken at different depths (unless the surface geophysical technique has the capability of varying the depth of investigation).

From intuition, one can see that, if information is desired about the geophysical property spatial variability at a scale smaller than the network scale, the inverse problem does not have a unique solution—unless further constraints are added—due to the fact that there are more unknowns than measured data. In a mathematical context this situation is referred to as the inverse problem being underdetermined; the consequence is nonuniqueness in the solution. The only way that a solution can be obtained for an underdetermined problem is by adding additional problem constraints (which represent "knowns"). This is usually accomplished by establishing criteria (or norms) that the solution must meet. It is critical to understand that, even though a unique solution to the inverse problem may be obtainable by adding constraints, the solution is highly dependent on how the norm(s) are defined. This translates as uncertainty in the inverse solution.

Most of the geophysical inverse problems encountered in practice when using surface geophysics are underdetermined due to the inherently high degree of spatial filtering (i.e. large measurement support volume) associated with "observing" into the subsurface as deep as possible from the surface. Borehole geophysics tend to be less affected by the underdetermined problem than surface geophysics, as a result of having a shallower depth of investigation, a more
focused measurement, and a much higher measurement sampling density; this translates into a much lower degree of spatial filtering and a much greater number of data.

Other factors that can complicate the geophysical inversion problem are (1) the inability to calculate a direct inverse solution and (2) instability in the solution. If the forward operator (the mathematical expression that maps the geophysical property to the observed state variable) is a nonlinear functional, it may not be possible to directly identify the inverse operator (the mapping in the opposite direction). Consequently, an indirect iterative search approach must be used to search for the inverse solution. If the forward operator is highly nonlinear (e.g. nonconvex), it may be very difficult to locate the global solution using this approach; instead, local minima/maxima in the solution function are inadvertently located, resulting in, possibly highly, inaccurate estimates of the geophysical property distribution. Instability of the inverse solution is characterized by the unfavorable situation where small differences in measurement results cause large changes in the estimated geophysical property distribution obtained through inversion (chaotic behavior); consequently, minor amounts of noise in geophysical measurements can lead to significant errors in the inverse solution. The implications of both these complicating factors on the geophysical inverse process is additional uncertainty in the estimated geophysical property distribution.

The inverse problem in geophysics has most commonly been solved using deterministic, as opposed to stochastic, approaches; therefore, uncertainty is not rigorously accounted for in the property distribution estimate. Uncertainty in the inverse solution is usually only qualitatively addressed through "exploration" of the property distribution space—examining the effects on the estimate of solving the problem using different norms, constraints, and/or solution techniques—and appraisal of the resolving power and average property value estimate at particular locations. However, much more rigorous stochastic approaches to inversion have been attempted in geophysics and other fields; most noteworthy to the author are the stochastic approaches taken in hydrogeology to solve inverse problems, such as estimating the hydraulic conductivity distribution from hydraulic head measurements (taken when the groundwater
system is stressed) and/or solute concentration measurements (taken during a tracer test). Some of the more powerful and noteworthy methods that attempt to rigorously address and quantify uncertainty in the inverse problem include Backus and Gilbert appraisal and inference approaches (reviewed by Oldenburg [1984] and applied in a hydrogeology context by Vasco et al. [1997]), minimum relative entropy inversion (applied in a hydrogeology context by Woodbury and Ulrych [1996]), and the geostatistical inversion approach (outlined in a hydrogeological context by Kitanidis and Vomvoris [1983]). An overview of these methods, highlighting their advantages and limitations, is provided in Appendix A. At this point, suffice it to say that the methods that have been developed to rigorously address uncertainty in the inverse problem tend to be highly involved, both in a theoretical and computational context, due to the mathematically complex nature of the problem.

3.4.1.2 Empirical and statistical models

The physical complexity of many geophysical methods prohibits the ability to develop and solve tractable physical forward models of instrument response. In this case, an empirical or statistical model describing the dependence between the observed state variable and geophysical property—that, inherently, does not rigorously account for the physical theory underlying the geophysical instrument response—must be resorted to. Such models rely on the comparison between laboratory and/or field instrument observations (indirect measurements) and reality—defined by accurate, precise, direct (hard) measurements of the geophysical property of interest—to develop a quantitative description of the relationship between the observed state variable and geophysical property. In this work, a distinction is made between empirical and statistical models, although the two types of models are very similar, to account for different methodologies used for developing them. With empirical models deterministic mathematical expressions of the instrument response to highly-controlled, well-known conditions of the geophysical property (e.g. constructed laboratory models) are developed. In contrast, with statistical models a quantitative and, usually, probabilistic relationship defining the dependence between the instrument response and geophysical property is developed based on prior
knowledge and a statistical analysis of a large number of indirect (instrument response) and direct field measurements, at least some of which are collocated.

A major limitation of empirical and statistical models of the measurement process is that, unlike physical models, they cannot rigorously account and correct for spatial filtering; they are only capable of performing the task of transforming the observed state variable from measurements to the property of interest. Consequently, the use of these models for addressing the geophysics problem is only appropriate if the spatial support scales (the volume of subsurface material that predominantly influences the measurement) of the direct and indirect "calibration" measurements used to develop the models are similar, or regularization techniques are used to make them so. As a result of this limitation, and the purely statistical nature of the empirical and statistical model approaches, if either of the two approaches has to be taken for developing both the mathematical relationships between (1) the observed state variable and geophysical property and (2) the geophysical property and material property of interest, there is no advantage in treating the two dependencies individually; such a treatment requires a two-step estimation process to ascertain the material property of interest from the geophysical measurements. Instead, the geophysical instrument response can be directly related to the material property of interest in the empirical or statistical model, significantly reducing the required computational effort. If this scenario exists for a particular problem, then the usage of "geophysical property" in the following discussion of empirical and statistical approaches to modeling the geophysical instrument response should be replaced with "material property of interest".

3.4.1.2.1 Empirical model

An example of an empirical model is the taking of a set of neutron count rate measurements with a borehole geophysics neutron-porosity instrument in carefully constructed borehole models with a range of known hydrogen index values (HI—the volumetric fraction of hydrogen in the formation, which is typically directly related to water content in the formation). Then, a curve (representing a mathematical expression) is fit to the results of measured neutron count rates versus known HI using some form of regression analysis; the mathematical
representation of the fitted curve represents an empirical model relating HI to neutron-porosity instrument response.

It is also possible to develop an empirical instrument response model for borehole neutron-porosity instruments by mathematically modeling the response of the particular instrument to known conditions for a range of constant HI formations, and substituting the modeled response results for the measured results in the above formulation. In this case, a forward physical model of instrument response exists, but only for stationary and homogeneous conditions (i.e. the tool response cannot be practically modeled for normal borehole logging conditions and operations). A combined physical-empirical model that quantitatively describes the dependence between the geophysical property and observed state variable can be developed. It is important to realize that this physical-empirical model still does not account for spatial filtering.

Once the empirical neutron-porosity instrument response model is defined using one of the two methods mentioned above, neutron count rate measurements can be directly converted to HI values by evaluating the mathematical model at the appropriate count rate value. It is important to note that use of an empirical model to estimate geophysical property values from geophysical observations is necessarily a local, as opposed to global, estimation procedure; estimation can only be performed at measurement locations.

3.4.1.2.2 Statistical models

An example of a statistical model of geophysical instrument response is the statistical "calibration" of electromagnetic (EM) "apparent" conductivity readings in the field to collocated (and separately located for certain statistical approaches) direct measurements of soil conductivity (e.g. laboratory analysis of core sample or conductivity readings from a small handheld electronic probe that is inserted into the soil). Soil conductivity may be of particular interest in quantifying soil salinity since the two are directly related; often "apparent" conductivity can be directly related to soil salinity in the statistical model [Lesch et al., 1995]. To analyze the field data and develop statistical "calibration" models in such scenarios, two different general approaches have been used: (1) regression analysis and (2) a geostatistical linear estimation
procedure known as cokriging; the two approaches have many similarities and are, in fact, theoretically equivalent in certain situations [Lesch et al., 1995]. These methods are summarized next; more detailed descriptions are provided in Appendices A and B.

In very simple terms, regression analysis fits a simple mathematical function to the calibration scatterplot of collocated indirect (e.g. the geophysical instrument response) versus direct (e.g. accurate, precise measurements of the geophysical property of concern) measurements. This is accomplished by minimizing the sum of the squared residuals between estimates of the direct measurement variable—predicted by evaluating the regression function for indirect measurement values—and the direct measurements (i.e. a minimum least squares criterion). When linear regression (the approach most commonly taken) is implemented, a linear function—a straight line—is fit to the calibration data. Therefore, a regression model can be used to provide minimum residual variance estimates of the geophysical property of concern at locations on the field site where measurements taken using the corresponding geophysical device exist; the estimates are based on the assumption that the dependence between the instrument response and geophysical property can be approximated by a linear function (if linear regression is used) or other simple function (if other types of regression are used). In addition, the regression variance can be computed, providing a measure of uncertainty in the regression model. Then, the entire probability distribution function (pdf) describing the uncertainty in the regression model estimates can be characterized, providing a complete description of estimate uncertainty, if the following assumptions are valid: (1) the residuals are assumed to be independent and normally distributed about the fitted mean function, (2) the residual variance is assumed to be homoscedastic (equal for all values of the observed state variable), and (3) the regression model statistics are the same across the site—referred to as the assumption of statistical stationarity.

The geostatistical cokriging methodology uses a probabilistic approach, wherein the pertinent primary and secondary variables—the geophysical property of concern and the observed state variable, respectively, in the current discussion—are treated as spatial random
variables, to estimate the property value of concern and an associated measure of uncertainty in the estimate. The estimate uncertainty is represented by a variance or an entire pdf, depending on the specific cokriging approach and the chosen/assumed form of the random variables. The estimates are made at the locations of the indirect measurement (geophysical sensor response), as well as all locations in the vicinity of measurements, if desired. The approach, as applied to the geophysics problem here, entails a two step process:

1. Definition of the prior global geophysical property spatial random variable—the probabilistic model of the property before it is locally updated on the measurements. This requires inference of statistical spatial moments for the study area, based on analysis of the direct and indirect measurements of the geophysical property of concern—
   - auto-covariances\(^2\) for both the geophysical property (e.g. electrical conductivity) and observed state variable (e.g. voltage measured with the DC resistivity method),
   - the cross-covariance\(^3\) between the geophysical property and observed state variable, and
   - the expected value of the geophysical property (for certain cokriging methods).

2. Local updating of the geophysical property random variable at, and in the vicinity, of the instrument state variable observations (geophysical instrument response), based on these observations, as well as the direct measurements of the property. This is the actual cokriging operation, and the result is an estimation of the expected value of the geophysical property—with a measure of uncertainty. It relies on the inferred statistical spatial moments.

The following discussion provides an introduction on the use of cokriging for geophysics. A much more in-depth discussion of the theory and mechanics of geostatistics, focusing on the particular Markov-Bayes cokriging approach [Zhu and Journel, 1993] applied in this research, is presented in Section 4.3 (covering concepts and specific recommended procedures for the applied practitioner) and Appendix B (a mathematical description of the approach).

---

\(^2\) Auto-covariance, in this context, is a measure of the correlation of an empirical quantity from one location to a nearby location, for a range of different separations.

\(^3\) Cross-covariance is a measure of the correlation of an empirical quantity with another, different, quantity at varying separation distances, including at the same location.
Chapter 3. GEOPHYSICS IN HYDROGEOLOGY PROBLEMS

Elaborating on the general two-step process for cokriging, there are three key components required for the cokriging approach:

1. direct measurements of the geophysical property where the instrument state variable readings were made—required for inference of the spatial moment models—in addition, of course, to the instrument readings themselves—required for the cokriging estimation;

2. models of the first and second statistical spatial moments of the pertinent geophysical property and the instrument state variable(s)—inferred from, and, thus, dependent on, field measurements of the variables at the problem site (as well as any other type of information about the statistical moments, such as expert knowledge); and

3. the estimator, the form of which is independent of specific problem characteristics, such as measurement types and values.

Cokriging is identical to the more familiar kriging, except that the estimator is expanded to take advantage of the information that auxiliary (indirect) measurements, in addition to direct measurements, provide about the spatial distribution of the variable of interest.

The fundamental component for successful estimation of geophysical properties from the measured instrument response using the cokriging approach is information that can be used to infer the required statistical spatial moment models—most notably a sufficiently comprehensive data set of the direct and indirect (instrument response) measurements of the geophysical property. The inference process requires having a sufficiently large number of both indirect and direct measurements at many different separations, so that informative sample statistics can be constructed. Important sample statistics include:

- the direct geophysical property measurement sample mean for inferring the geophysical property expected value and
- experimental auto- and cross-covariograms/semivariograms calculated from direct and indirect measurements.

Preferably, the measurement data for such sample statistics is from the field site being investigated, but data from other similar field sites can be used.
The cokriging estimator is constructed to provide the minimum variance, unbiased, linear estimate of the variable of concern, at all locations desired, by taking a linear combination (weighted sum) of all direct and indirect measurements. The weights are computed based on unbiasedness (the expected value of the estimation error must be zero) and minimum variance (the variance of the estimation error is as small as possible) constraints [ASCE Task Committee on Geostatistical Techniques in Geohydrology, 1990]. The exact form of the estimator depends on the specific cokriging approach employed (e.g. ordinary, simple, or universal cokriging). The primary calculation required by the estimator is the computation of the measurement weights; this is a linear algebra problem—the solution of simultaneous linear equations. The system of equations used to compute the weights contain terms with the pertinent statistical moment models.

Therefore, to use the cokriging estimator it is necessary to infer, beforehand, the required statistical moment models. Ultimately, it is necessary to select mathematical models to represent the required statistical moments, based on the direct and indirect measurement data set and any other useful source of information, including expert judgment. Selecting a mathematical model entails choosing a functional form and adjusting a limited number of parameters defining the specific shape of the function. This is usually accomplished by fitting the function to useful sample statistics, adding expert judgment to arrive at a final form. Alternatively, there is a much more objective inference method—the maximum likelihood method [Kitanidis and Lane, 1985]; it works by identifying the parameter values that maximize the likelihood of the observed measurement values. Critical to the statistical moment inference process is validation of the selected mathematical models to ensure that they are consistent with available measurement data from the field site being investigated; appropriate validation techniques are discussed by Kitanidis [1991].

In the cokriging estimation operation the estimator uses the measurement data, constrained by the inferred statistical moment models, to locally reduce the uncertainty in the global geophysical property random variable field across the area where the instrument state variable
readings were made. The influence of the measurements at the locations where they are taken, as well as locations nearby, is accounted for in a probabilistic manner. If the measurement is an indirect (i.e. soft) measurement of the geophysical property, as is the case for the geophysical instrument response, the pdf of the geophysical property random variable at that location will potentially be narrowed—the amount of uncertainty reduction dependent on the correlation between the instrument response and geophysical property. The random variable can also be updated at locations near the instrument readings, wherein the uncertainty will be to a lesser degree—dependent on the spatial cross-covariance between the instrument response and geophysical property. With most cokriging approaches it is possible to reconstruct the entire random variable conditional pdf at any location in the vicinity of measurements; although, in some cases this requires random simulation of many equally-likely realizations, from which a probability histogram can be computed. The simulation operation incorporates the cokriging procedure (see Section 4.2.3.2).

Particularly important, and difficult, in cokriging is the inference of the cross-covariance model between the primary (e.g. geophysical property) and secondary (e.g. instrument state variable) variables; this model defines the correlation between the two as a function of spatial separation. Conventional statistical moment inference methods attempt to directly estimate the required covariances based on experimental auto- and cross-covariogram/semivariograms, computed from the direct and indirect measurement data set. Then the inferred covariance models are used in an extended, computationally-intensive cokriging estimator. Such an inference process requires having a large number of both indirect and direct measurements at many different separations, preferably including collocated indirect and direct measurements (a requirement for some cokriging estimators), so that informative experimental variograms can be constructed. In addition, when all the auto- and cross-covariance models are treated as separate and independent models the estimation process can be very computationally intensive, since, for the cokriging estimation, covariances need to be computed and stored for all measurement-measurement and measurement-estimation point pairings.
Alternative methods that indirectly calculate the secondary variable statistical second moments (auto-covariance and cross-covariance models) have been developed to cope with these inference and computational difficulties with conventional cokriging. In the method of auxiliary variables, the cross-covariance is calculated from a simple relationship between the auto-covariances of the primary, secondary, and a third auxiliary variable; the auxiliary variable is defined to be a simple function of the primary and secondary variables and to have an relatively easily estimated auto-covariance (e.g. the auto-covariance of the auxiliary variable is the sum of the auto-covariances of the primary and secondary variables and their cross-covariance) [Kitanidis and Vomvoris, 1983]

Other methods estimate the secondary variable auto-covariance and primary-secondary variable cross-covariance models by rescaling the primary variable auto-covariance model; the rescaling is based on a calibration model relating the secondary and primary variables. The calibration model is typically derived by performing some sort of statistical analysis of collocated direct and indirect measurements. Cassiani and Medina [1997] define a calibration model based on the results of a linear regression analysis of the calibration data. The necessary secondary variable auto- and cross-covariance models are determined through a linear rescaling of the primary variable auto-covariance using parameters from the regression model. The Markov-Bayes method treats both the primary and secondary variables as a set of binary indicator random variables (IRV), having values of either 0 or 1—defined by threshold values or class membership. The IRV classification enables the derivation of the secondary variable indicator auto-covariance and the indicator cross-covariance directly from the primary variable indicator auto-covariances, based on only a few basic assumptions [Alabert, 1987]. Alabert [1987] shows that the desired indicator covariance functions at each indicator threshold level, or class, is simply the primary variable indicator auto-covariance function for the matching indicator level/class multiplied by a rescaling factor. The rescaling factor is computed from likelihood probabilities between the defined threshold levels/classes of the secondary and primary variable random variables—obtained through a simple statistical calibration of
collocated indirect and direct measurement data. A slightly modified version of the Markov-Bayes method [Zhu and Journel, 1993] is the approach used in the case study of this work, selected for its generality—there are no restrictions on the acceptable form of random variable pdfs and multiple types of indirect/secondary data can be efficiently incorporated—and simplicity—the relationship between the material property of interest and the geophysical instrument response is derived through a simple statistical calibration. The Markov-Bayes method will be discussed in much greater detail in CHAPTER 4.

3.4.1.2.3 Comparison of cokriging to other empirical and statistical approaches

The geostatistical cokriging approach to modeling the geophysical measurement response, and estimating the geophysical or material property distribution from geophysical observations, has many conceptual advantages over regression and empirical approaches; the main tradeoff is computational efficiency. Both statistical and empirical approaches have conceptual limitations, albeit computational advantages, as compared to a robust physical model approach. Unlike regression and empirical approaches, cokriging is an exact interpolator—the estimator will always converge on the value of exact direct measurements at those measurement locations. In addition, cokriging can produce estimates anywhere within the survey area, accounting for the influence of any nearby direct and indirect measurements as determined by the spatial correlation and cross-correlation of the primary and secondary variables. Regression and empirical models permit estimation only at measurement locations, and the estimate is independent of any neighboring measurement values. Other advantages of the cokriging approach are:

- the ability to always estimate the entire primary variable pdf, either directly for certain techniques/situations or indirectly through random simulation, and
- the ability to incorporate prior information other than the localized measurement values in the estimation process (e.g. prior information about the statistical mean of the primary variable across the site).

The main limitation of the cokriging approach to other statistical and empirical methods is the need to infer auto- and cross-covariance models and incorporate them in the estimator—often
a difficult and computationally intensive process. However, as discussed earlier, methods have been developed to simplify and accelerate this process.

3.4.1.2.4 Comparison of empirical and statistical models to physical models

Both empirical and statistical model approaches to estimating geophysical property values from geophysical observations are conceptually inferior to robust physical model approaches in the following ways:

- they are inherently less informative about the relationship between observations and property values, since physical details about the dependency are lost and
- they cannot increase estimation resolution beyond that of the geophysical measurement support scale by deconvolving the effects of spatial filtering; a physical model can potentially accomplish this, if the spacing between measurements is smaller than the support scale.

The latter limitation can be crudely overcome using certain cokriging estimation techniques that account for specified spatial averaging in measurements, but the statistical calibration between indirect and direct geophysical property measurements (i.e. auto- and cross-covariance estimation) still requires that the two types of measurements have similar support volumes.

Another limitation of both empirical and statistical approaches compared to physical model approaches is that they require a large number of "calibration" measurements to accurately model the observed state variable-geophysical property dependency. Sometimes analogous calibration results from other similar field sites can be included (or used exclusively) in the statistical analysis to help overcome this limitation.

A major advantage of empirical and statistical approaches over physical model approaches to estimating geophysical property values from geophysical observations is their conceptual/theoretical simplicity and computational efficiency; this is achieved by eliminating the need to perform a potentially unstable and complex formal mathematical inversion. This advantage is especially important with geophysics, since the physical underpinnings of many geophysical techniques is highly complex, and often not well understood. Although methods have been developed to quantify the uncertainty in geophysical inverse problem solutions, they
are typically mathematically and theoretically complex, and often reliant on many simplifying assumptions. In contrast, geostatistical cokriging techniques can rigorously and elegantly account for uncertainty in the estimates; indeed, these techniques have been purposefully developed to deal with estimation problems in a probabilistic framework. For the above-mentioned reasons, a cokriging approach to the geophysics estimation problem—specifically, a version of the Markov-Bayes approach [Zhu and Journel, 1993]—is implemented in the case study of this work.

3.4.2 The relationship between geophysical properties and material properties

Often there is one more step required after the geophysical property values—that the measuring instrument is sensitive to—have been estimated for the useful application of geophysical measurements in hydrogeology-related problems—the estimation of the hydrogeological material property of interest from the geophysical property values. There are applications where this additional step is unnecessary. For example, in reference to the example problem, if an electromagnetic (EM) technique that responds to variations in electrical conductivity is used to locate buried steel fuel pipes, then the characterization of subsurface electrical conductivity ($\rho$) is all that is required, since steel has a much higher $\rho$ than natural earth materials. Electrical conductivity, itself, is one of the more sensitive property differentiating the steel pipes from the surrounding earth materials. However, if the application was to use the EM technique to characterize hydraulic conductivity ($K$), an extra estimation step would be required to estimate $K$ values from $\rho$ values, since there is not a direct, simple relationship between the two property types. This extra step—estimating hydrogeological material properties from geophysical properties—can be a particularly vexing problem due to the highly complex dependency of geophysical properties on the subsurface environment. Indeed, the field of rock physics (or petrophysics) was created primarily to address this problem—to determine theoretical and/or laboratory models relating substrate properties and geophysical properties; this will be referred to as the *rock physics problem* in this work.
As with the dependency between geophysical observations and the pertinent geophysical properties that influence them, in order to estimate the material property of interest from the geophysical property, it is necessary to define the relationship between the two using a physical, empirical, or statistical model (or combination of model types). This requires determination and an understanding of the key controls affecting what the values of the geophysical property are in the material of concern (e.g. porous sediments). An important consideration in this process is scale:

• the scale of interest with respect to the geophysical property,
• the corresponding scales of variability in the key subsurface controls, and
• the relationship between the two.

If the subsurface consisted of a single component (e.g. a nonporous rock with a simple composition), or even a few components arranged in a simple geometry (e.g. sandstone with spherical quartz grains and pores completely filled with water, whose constituents are in chemical equilibrium), the process of understanding the key controls and defining the relationship would be relatively simple. It would only be necessary to determine (1) the value of the geophysical property for the individual substrate component(s) and (2) how the values change when the individual component(s) are varied. Even in this simple scenario an assumption is being made that the variations in the components occur at a scale larger than the scale that the geophysical property can be resolved; otherwise spatial filtering has to be accounted for.

Unfortunately, in the real-world the subsurface is intrinsically much more complex—composed of multiple fluid and solid components, arranged in many different geometry, that potentially interact chemically with each other, all of which can potentially influence the value of geophysical properties. To complicate matters further, the geophysical measurement, itself, can sometimes significantly alter conditions in the subsurface, since certain methods emit or induce energy fields in the subsurface (e.g. seismic techniques send acoustic pressure waves through the subsurface that can cause pore fluids to move).

There are two principal approaches to the rock physics problem:
Chapter 3. GEOPHYSICS IN HYDROGEOLOGY PROBLEMS

1. Develop a simplified conceptual model of the subsurface that can quantitatively account for and describe the salient features in the subsurface—those that are most influential on the value of the geophysical property—using basic parameters of the model. Then isolate the dependence of the geophysical property on each problem parameter; and

2. Concede that the subsurface is too complex to accurately model and, instead, perform a statistical calibration, based on lab or field measurements, to define the relationship between material properties of interest and the geophysical property (or observed state variable).

Both approaches have their merits and advantages, as well as their weaknesses and limitations.

3.4.2.1 The rock physics problem – conceptual model approach

The first approach requires the completion of three major steps: (1) development of a subsurface conceptual model wherein important subsurface features are quantitatively represented through defined model parameters that can be varied, (2) determination of how the model parameters influence the geophysical property of concern, and (3) definition of the relationship between the model parameters and the hydrogeological material property of interest, so that the geophysical property can be related directly to the material property. The third step is sometimes not required if the model parameter and material property of interest are one and the same (e.g. porosity).

A subsurface conceptual model should describe, at least, the solid component (matrix) and void (pore) space of subsurface earth material, and, ideally, the interfaces between the two, since most geophysical properties are highly sensitive to these factors. The types of subsurface features that are typically considered important (i.e. influential on the behavior of geophysics) in rock physics problems are, in order of increasing level of difficulty to incorporate in the problem:

- identity of components (both solid and fluid),
- properties of individual components (e.g. electrical conductivity, density),
- volume fractions of components,
- geometry of components (internal arrangement), and
• interaction between components (e.g. how water interacts with solid matrix to form surface tension or an electrical coating).

To quantitatively incorporate these features in a subsurface conceptual model, they have to be represented by defined model parameters (e.g. porosity to describe void space, surface area to describe pore space geometry). As the number of subsurface features represented is increased, it becomes increasingly more difficult to properly represent them in the model, and predict the geophysical behavior in the model; however, the resulting model of geophysical property dependency is more comprehensive and realistic—a tradeoff. It is also important to realize that the model's description of subsurface material can only be as complicated as the ability to characterize the material in the field, if the model of geophysical property dependency is to be practical. For these reasons, macroscopic (or bulk) model parameters, defined at scales much larger than the pore-scale, are usually used to describe the key controls on geophysical properties (e.g. the ratio of surface area to volume of pore space to describe pore space geometry), instead of parameters that attempt to explicitly describe small-scale features (e.g. the dimensions and shape of matrix pores).

There are two principal approaches for determining the influence of subsurface conceptual model parameters on geophysical property values:

• A theoretical approach, wherein a mathematical relation between the two variables is derived, using petrophysical principles applied to the conceptual model; and

• An empirical approach, wherein a laboratory experiment is built to replicate the conceptual model. Conditions in the laboratory setup are varied so as to represent changes in the model parameter(s) defining subsurface features, and the effects of the changes on the geophysical property of concern are measured using a laboratory instrument.

Often hybrid approaches, incorporating both theoretical and empirical components, are implemented; for example, a petrophysical relation is first developed, after which a laboratory study is performed to verify and fine-tune the relation.
Chapter 3. GEOPHYSICS IN HYDROGEOLOGY PROBLEMS

Mixture formulas, that define bulk, or effective, geophysical properties as volumetric averages of the geophysical property values of individual components in the modeled substrate, are used as the framework for many theoretical approaches. The averaging method can range from geometric to harmonic as end members, or any combination in-between. An example of a mixture formula is the complex refractive index model (CRIM). CRIM is a simple mixing law that expresses the bulk dielectric constant (which can be derived from certain types of ground-penetrating radar surveys) as a linear combination of the square root of the dielectric constants of air, water, sand, and clay weighted by their volumetric fractions, all of which represent conceptual model parameters [Wharton et al., 1980]. Under laboratory conditions the CRIM model was found to fit the observed dielectric behavior of a constructed unconsolidated sand-clay mixture [Knoll and Knight, 1994; Knoll et al., 1995]. Using such a relation, the influence of each model parameter on the bulk dielectric constant can be determined. However, if the material property of interest in an investigation is intrinsic permeability \( k \), instead of one of the conceptual model parameters, then a relation must be found between \( k \) and the model parameters. Knoll et al. [1995] used the CRIM mixing law with a series of additional petrophysical models to produce relations between dielectric constant, saturation, and intrinsic permeability. The petrophysical models required to establish the linkage between the dielectric constant and these hydrogeological properties are (1) a simple microgeometrical porosity model for sand-clay mixtures (2) a relation for \( k \), derived from the Kozeny-Carman relation, and (3) a definition of surface area per solid volume—in terms of the air, water, sand, and clay volumetric fractions, as well as the grain particle radii of clay and sand (see Hubbard et al. [1997] for a review of the derivations). In this example, the relations defining the influence of subsurface conceptual model parameters on the hydrogeological material property of interest were derived entirely from theoretical relations (microgeometrical models); however, all the theoretical relations were verified in the laboratory (albeit not in the field, at least at the time of this work).

The above example could have been carried out similarly using an empirical approach wherein the curves defining the relations are determined by (1) constructing laboratory models of
air-water-sand-clay mixtures, (2) altering conditions in the laboratory models in a controlled manner, so as to vary conceptual model parameters by known amounts, and (3) observing the effects of the changes on the bulk dielectric constant—measured with a laboratory radar instrument. The empirical relations are derived by fitting a mathematical function to a plot of the parameter of interest versus the different influencing model parameters.

Once the functional relation between the geophysical property and material property of interest are determined using a theoretical and/or empirical approach, field estimates of the material property can be computed from measured geophysical property value distributions by applying the relation. This estimation procedure actually requires a mathematical inversion, since the material property and geophysical property variables are independent and dependent variables, respectively, in the rock physics problem—a “forward model” (e.g. the dielectric constant is derived as a function of petrophysical properties in the above example). The forward model relations are usually relatively simple algebraic expressions that can be inverted without much difficulty through simple mathematical rearrangement, or using graphical techniques. However, the relations are often (1) non-unique (e.g. the seismic velocity-permeability relation in Copty et al. [1993]) and (2) dependent on many subsurface characteristics—represented in the relation by the conceptual model parameters (e.g. dielectric constant-permeability relation in Hubbard et al. [1997]). The consequence of this is, respectively, (1) inherent uncertainty in the material property estimates and (2) the need to characterize many different subsurface features.

### 3.4.2.2 The rock physics problem – statistical approach

The statistical approach to the rock physics problem involves performing a statistical calibration between field observations of the geophysical property of concern and the material property of interest—to estimate the correlation relationship between the two properties. The approach is essentially identical to the statistical approach to defining the relationship between the observed state variable (geophysical instrument response) and the geophysical properties that influence it, discussed earlier in Section 3.4.1. The relationship is defined through either a regression or cross-covariance model, depending on whether a regression analysis or
geostatistical cokriging methodology, respectively, is used. Either way, uncertainty in the relationship can be accounted for. The integrity of the statistical approach is contingent on having a large enough calibration data set of geophysical and material property measurements, with a range of values representative of the area being characterized; the measurements should have similar support scales (volumes of investigation). The measurements of the two property types have to be collocated for the regression and many cokriging techniques. After the statistical model(s) relating the material property and geophysical property is defined from the calibration data sets, field estimates of the material property are made from the geophysical property measurements using a regression or cokriging estimation technique. A detailed discussion of the use of statistical regression and cokriging methodologies for (1) defining the relationship between two different variable types and (2) estimating values of one from the other was provided earlier in this chapter in the discussion on the geophysical property-instrument response relationship (Section 3.4.1) and, thus, will not be repeated here.

3.4.2.3 The rock physics problem – comparison of approaches

In general terms, the subsurface conceptual model approach to the rock physics problem—based on theoretical or laboratory-derived empirical relations—can provide a more informative and detailed description of the relationship between the geophysical property and material property than the statistical approach. However, the statistical approach is usually more universally applicable and straightforward than the conceptual model approach, since it is not reliant on accurately characterizing ancillary subsurface features, and less reliant on other assumptions, to define the relationship; and it can also account for uncertainty in material property estimates. If the subsurface conceptual model accurately depicts the conditions being investigated, the derived theoretical/empirical relations provide a very accurate model of the true geophysical property-material property dependency. Then, if all the conceptual model parameters—except the material property being investigated—can be accurately characterized throughout the problem site, accurate estimates of the material property can be made from measurements of the geophysical property (assuming the measurements are accurate).
However, these requirements are rarely met in the real-world, except at the most ideal field sites. The implications of not meeting these requirements are (1) an inaccurate model of the dependency relationship, due to incorrect assumptions in the subsurface model, and/or (2) inaccuracies in the assigned values of conceptual model parameters, due to an inability to characterize the site properties that define the true values of these parameters—both leading to potentially significant errors in the material property estimates. Unlike the statistical approach, the conceptual model approach cannot conveniently account for uncertainty in quantitative terms.

As more complicated subsurface conceptual models—that more accurately depict real-world conditions—are being developed and used to investigate geophysical behavior, the more complicated the picture of geophysical property dependency on conditions in the subsurface is becoming. The difficulty in developing and implementing accurate and practical theoretical and/or empirical models for this increasingly complicated dependency, that can be applied to real-world problems, is becoming more apparent. For example, through rock physics research it is becoming apparent how important the characteristics of solid-fluid interfaces are on the geophysical response (i.e. the value taken on by geophysical properties); properties of these interfaces controlling geophysical response can be process-dependent (e.g. water wetting properties in sub-saturated porous material are dependent on whether the pore water was most recently drained or emplaced), and are very difficult to independently characterize. The number and variety of subsurface conditions that influence geophysical properties is often very large, and these conditions can vary significantly from site to site. This makes it difficult to use the conceptual model approach for real-world problems, since it is difficult to account for all the property interdependencies, and especially difficult to properly characterize the pertinent subsurface conditions, at a real-world field site. Even though the conceptual model approach may not be ideal for real-world applications, its importance should not be understated; it provides a basis for understanding the key controls on geophysical behavior in the subsurface environment that, in the very least, can help to qualitatively predict what types of geophysical
property-material property relationships can be expected for certain types of subsurface conditions.

The statistical approach to the rock physics problem is much more universal and conceptually straightforward to apply, and, therefore, more practical for real-world applications, than the conceptual model approach. The methodology does not require an understanding of the complex nature and types of factors influencing the geophysical property in order to define a probabilistic relationship between the geophysical and material properties of interest. Site-specific conditions are inherently accounted for in the defined statistical model of the relationship—if the calibration data is acquired from the actual field site being investigated (or one that has very similar conditions). In addition, the statistical approach is much more adept at accounting for uncertainty in the material property estimates—a high priority when using a risk-based decision-making process, as advocated in this work. For these reasons, in the case study of this work a geostatistical cokriging approach is used to quantitatively model the relationship between geophysical and material properties, and to estimate, in an uncertainty framework, the spatial distribution of the material property of interest from measurements of the geophysical property.

3.4.3 Noise in geophysical measurements

In addition to the problems created by the indirectness of geophysical measurements, the application of geophysics to subsurface characterization can also be hindered by noise in the measurement signal—defined as measurement response resulting from site conditions/properties that are not of interest. This difficulty is exacerbated by the propensity for problematic noise to be highly site-specific, and, thus, difficulty to predict. Although the S/N ratio—the proportion of the measured signal resulting from "true" subsurface conditions of interest—is difficult to quantify in actual geophysical measurements, it represents a useful metric for conceptualizing the extent to which noise corrupts the measured geophysical response. The S/N ratio can cause difficulties for interpreting geophysical measurements in two ways:
• if the noise in the geophysical response is recognizable, but it masks or significantly interferes with the component of the response resulting from actual subsurface conditions of interest—making it difficult or impossible to reconstruct the true signal from the noisy measurement;

• if the noise is unrecognizable and the S/N ratio is low enough, or the noise is of the form (e.g. biased noise), to cause significant errors in the geophysical interpretation.

Whether or not these difficulties are problematic to a particular application depends on the (1) magnitude of the geophysical interpretation errors, (2) how these errors corrupt the estimation of the material property of interest, and (3) the estimation accuracy, precision, and resolution required by the application. The S/N ratio can become problematically low for a particular application—enough to make a particular geophysical technique unsuitable for use in certain, or all, regions of a site—if site conditions are not conducive to performing that technique. Site conditions can induce this unfavorable situation by causing low levels of signal transmission energy and/or high levels of noise in the measurement.

Low signal transmission occurs when the strength of the energy emitted by the geophysical source is significantly reduced by the time it reaches the receiver where, as a consequence, the recorded geophysical response is very weak. The causes of this reduction in signal strength can be categorized as either (1) poor coupling between the source/receiver and the earth medium—the signal transmission between the source/receiver and subsurface material is poor—or (2) high signal attenuation within the subsurface along the travel path between source and receiver. For example, when applying seismic methods to near-surface problems it can be very difficult to achieve good coupling between acoustic sources/receivers and the ground, as well as maintaining sufficient signal strength as the acoustic wave travels between source and receiver, if the shallow subsurface consists of unconsolidated, large-grained materials (e.g. loose sands and gravels). The acoustic coupling and transmission problems stem from the difficulty in effectively transmitting a pressure pulse into such loose material and the inability of the pressure
wave to effectively propagate through the material, respectively\textsuperscript{4}. Insufficient signal transmission presents problems for many geophysical methods; indeed, one of the main objectives, and difficulties, in developing practical geophysical techniques is overcoming the sources leading to these problems. However, many of the limiting factors are inherent to the physics associated with the technique itself and, thus, there are always certain site conditions for which specific types of geophysics cannot perform effectively. For example, GPR has poor signal transmission through and, correspondingly, very high signal attenuation within, high clay environments, resulting in very low S/N ratios in the measured response from subsurface materials within and below the highest clay layer.

The noise that afflicts geophysical measurements has many origins and forms. Noise can be considered any contribution to the measured geophysical response that is not accounted for by the measurement model and, hence, represents interference to the signal of interest, possibly corrupting the estimate obtained from the measurement. Noise is caused by conditions that significantly influence the geophysical measurement, but are not of any interest in the study; it can be either coherent (unbiased) or incoherent (biased) in character (see Section 2.2.2 for an explanation of the different forms of measurement error). The main contributors to geophysical noise are cultural (man-made) sources, both above and below the ground surface, although natural conditions can sometimes induce measurement noise as well. Another source of errors in geophysical measurements, which can be classified as cultural noise, is the intrinsic imprecision of the measuring instrument, usually resulting from electronic limitations.

Cultural examples of geophysical noise sources include the influence of:

- steel-link fences on GPR/electromagnetic techniques—

  $\Rightarrow$ a high amplitude reflection in the recorded wavetrain at a time proportional to the distance between the fence and receiver that can: mask/interfere with true subsurface reflections (GPR)

---

\textsuperscript{4}Think of the difference between hitting a sand pile and hitting concrete with a sledge hammer; the hammer produces a much greater sound and far-reaching vibration when hitting the concrete.
Chapter 3. GEOPHYSICS IN HYDROGEOLOGY PROBLEMS

⇒ erroneous values of the measured electromagnetic potential near the fence that vary greatly in space, depending on the exact location and orientation of the instrument (electromagnetic techniques);

- power-lines on electrical/electromagnetic techniques—skewed values of the measured electrical/electromagnetic field near the powerlines; and

- jet aircraft traffic on seismic techniques performed near an airport—low frequency (usually below 100 Hertz) acoustic energy in the recorded acoustic amplitude versus time section.

An example of a natural source of geophysical noise is the influence of buried boulders on surface seismic reflection techniques—numerous diffractions in the recorded acoustic amplitude-time section that interfere with signal energy caused by reflections.

The propensity for noise and/or signal attenuation to be problematic when using geophysics for subsurface characterization is dependent on many factors, and is highly site-specific—varying greatly from site to site, or within the area of a single site. Among these factors are (1) the particular geophysical technique being used (e.g. GPR usually performs well/poorly in low/high electrical conductivity subsurface environments, whereas the performance of electromagnetic methods is usually the opposite) and (2) the particular application requirements (e.g. accurately estimating porosity requires a much higher measurement accuracy and precision than that required for locating buried waste dumps).

Ultimately, the degree and severity of noise and signal attenuation affecting geophysical measurements depends on the specific subsurface and surface, cultural and natural conditions at the site being investigated. Because of the high sensitivity of geophysical performance (as determined by noise and signal strength) to site-specific conditions, the success of a geophysical technique can be highly variable, and sometimes difficult to predict beforehand; this translates into geophysics generally having low reliability, compared to other characterization techniques that are not sensitive to these conditions (e.g. chemical analysis of physical soil samples).

All geophysical measurements contain some degree of noise. The form and severity of this noise, how well it can be recognized and removed from the true signal, as well as the S/N ratio,
determine the impairment it will cause to the application of geophysics at a site. A worst case scenario is a significant proportion of the geophysical measurements covering large areas of the site to having low S/N ratios, contaminated with high amplitude, biased noise that is not recognized by the practitioner and, thus, is not removed from the data prior to analysis and interpretation of the data. This scenario would lead to significant errors in the final results. Errors in the measurement state variables propagate through the process of estimating the material property of interest (accounting for the indirect nature of the measurement)—resulting in an amplification of unaccounted for estimation errors of the material property. These errors can, and should be, accounted for by attaching uncertainty to the geophysical measurements and propagating this uncertainty through the different stages of the estimation process, leading to an uncertainty assessment of the final result; this is especially important if a rigorous, risk-based approach to decision-making is to be taken—as in the case study of this work.

3.4.4 Summary

The difficulty facing the practitioner using geophysics to characterize the subsurface is vexing: how to extract information about the subsurface property/condition of interest—often some hydrogeological material property (e.g. hydraulic conductivity)—from, quite possibly noisy, technique-dependent geophysical measurements of some energy field value, which is only indirectly related to the property/condition of interest. There are two levels of dependencies in the relationship between geophysical measurements and material properties: (1) the dependency of the geophysical observations on the spatial distribution of the geophysical property (properties) that the measurement technique is directly responding to and (2) the dependency of the geophysical property (properties) on the subsurface material property of interest. The first dependency is influenced by (1) the spatial filtering of the geophysical measurement process—the measurement is responding to a geophysical property, but the response is averaged over a large volume of subsurface material—and (2) internal electronics of the geophysical measuring device—the device is actually measuring an instrument-dependent electrical
parameter that is sensitive and directly related to changes in the geophysical property distribution. The second dependency is often an indirect relationship; the bulk geophysical property—that which the geophysical measurement actually responds to—is directly influenced by specific conditions in the subsurface, including pore-scale attributes (e.g. the surface area of pores, the nature of the pore fluid-solid matrix interface) that are very difficult to characterize in the field. The conditions affecting the bulk geophysical property do not necessarily include the material property of interest; in this case the relationship between the material property and those conditions has to be determined as well.

All these dependencies need to be modeled in order to make predictions about the material property of interest from geophysical measurements. There are two broad approaches for modeling these dependencies and estimating the property of interest from geophysical observations: (1) a theoretical or empirical approach, wherein mathematical relations are developed based on physical principles or highly controlled laboratory measurements, respectively, to represent the dependency and (2) a statistical approach, wherein quantitative statistical models are developed based on field measurements to represent the dependency. Comparing the pros and cons of the two approaches, the first approach provides a more informative and detailed model of the dependency and, therefore, is capable of providing more accurate estimates of the property of interest from geophysical measurements; however, it is much more theoretically and mathematically-involved and usually places more reliance on assumptions. The second approach is usually much more straightforward (both conceptually and mathematically), accounts for site-specific conditions and their effects on geophysical response, and can estimate the property of interest at locations without measurements—using inferred spatial correlation models (if a cokriging technique is employed). If a statistical approach is used for both dependencies, the material property of interest can be related directly to the geophysical observable, instead of modeling the two dependencies separately; this is the approach taken in the case study.
As a result of the multiple levels of indirectness between the observed state variable—measured by a geophysical device—and the subsurface property/condition of interest, and the propensity for noise in geophysical measurements, uncertainty is inherent in the predictions acquired from these measurements. Since quantitatively accounting for uncertainty is crucial for implementing a risk-based decision-making process, as advocated in this work, the geophysical measurement analysis techniques considered in this work must have this capability. The statistical approach is better equipped than most theoretical/empirical approaches for quantifying uncertainty in the material property estimates obtained from geophysical observations, since the statistical techniques typically employed in such an approach are purposely designed to work within a probabilistic framework.

A statistical approach—geostatistical cokriging (see CHAPTER 4)—is taken in the case study of this work to estimate the material property of interest from geophysical measurements. The approach was selected for its

- generality—ability to handle many different geophysical measurement types, material property types, and site conditions,
- conceptual and mathematical simplicity—ability to compute material property estimates anywhere on the field site, and
- rigorous treatment of uncertainty.
CHAPTER 4
ACCOUNTING FOR AND MODELING UNCERTAINTY

This chapter discusses how uncertainty can be quantitatively accounted for and modeled in hydrogeology-related problems, with an emphasis on problems incorporating geophysical measurements and, particularly, on the methodology employed in this work to incorporate such measurements in a probabilistic framework. As explained in CHAPTER 3, there is often a high degree of uncertainty associated with the rigorous incorporation of geophysics in these problems. A probability analysis approach for representing uncertainty, wherein a probability distribution function (pdf) is used to characterize the uncertainty associated with a quantity, is followed in this work; the specific method of choice is a variant of the Markov-Bayes indicator geostatistical algorithm [Zhu and Journel, 1993]. In this method, uncertain spatial quantities (e.g. subsurface concentrations of chemical contaminants) are characterized by their probabilities of being within defined class intervals, and the information provided by measurements is determined based on an inferred correlation between the measurements and the quantity—both at coinciding and spatially separated locations. The uncertainty associated with indirect measurements of a quantity (e.g. measurements of subsurface electrical conductivity used to locate subsurface contaminants) is based on a statistical calibration between collocated indirect and direct measurements of the quantity.

The main purpose of this chapter is to provide an overview of this method—its basis, implied assumptions, usage, and limitations—particularly as applied to the case study in this work. Emphasis is placed on how the method is applied in practice, as illustrated through examples from the case study. Theoretical and mathematical details of this geostatistical method are provided in Appendix B. In order for the reader to understand the basis and attractiveness of the approach for the types of problems addressed in this work, an introduction to uncertainty analysis, in the context of these types of problems, is provided before the details of the approach are explained.
4.1 Introduction

Uncertainty is a concern in hydrogeology-related problems—as well as any problem in which critical decisions about future actions have to be made—because it results in the risk of unintended decision outcomes. Therefore, to account for the consequences of risk in decision-making, uncertainty must be identified and quantified in the problem. Uncertainty can afflict many, if not most, of the components of a decision problem—technical (scientific and/or engineering), economic, and political quantities, as well as the functional form of technical, economic, and political models. The approach for dealing with uncertainty depends on the problem component it is associated with, the source and type of uncertainty, and the overall context of the particular problem being addressed. Morgan and Henrion [1990] provide a very thorough discussion on the most appropriate uncertainty analysis approaches to use for the different components of a problem, which they explicitly define.

4.1.1 Uncertainty analysis approaches

The two principal approaches for treating and analyzing uncertainty are (1) probability analysis and (2) sensitivity/parametric analysis. Probability analysis characterizes uncertainty in quantities by representing each quantity as a random variable\(^1\) (RV) with a probability distribution function (pdf), which identifies the likelihood of the quantity taking on any particular value (see Figure 4.1 below).

\(^1\)A term used in predictive statistics that simply refers to a variable that has a component of random variability in it, the nature of which is defined by a pdf. Usually the RV is split into two additive components: (1) a deterministic component which does not have any random variability and (2) a stochastic component with random variability and a mean of zero.
The primary task in probability analysis is the derivation of an appropriate and representative pdf for the quantity of concern. Once a pdf is defined for the quantity (e.g. hydraulic conductivity pdf), the pdf can be used as input in system models that are influenced by the quantity to determine the uncertainty in system output (e.g. the cost pdf associated with an alternative in a decision model). The input pdf can be incorporated either (1) directly, through quantity value probabilities, or (2) indirectly as a set of equally-likely realizations of the quantity value, such that the histogram of the quantity value for the set of realizations approaches the pdf; the method is determined by the system model requirements.

In a sensitivity analysis approach, the quantity of concern (e.g. the failure cost associated with an engineering alternative designed to contain groundwater contamination that leaks) is discretely varied across a selected range of values. The effect of these variations on other important problem parameters (e.g. expected cost of containment) is characterized by running the system model(s) that define(s) the relationship between the quantity and parameter(s) of concern for multiple scenarios, each with a different value of the quantity from within the range of variation. The end result of a sensitivity analysis is a set of output parameter values calculated for a predefined set of input quantity values—providing insight into how the output is influenced
Probability analysis takes a much more rigorous approach to accounting for and representing uncertainty; however, it is only appropriate for empirical quantities (e.g. contaminant concentration, hydraulic conductivity), not decision variables (e.g. treatment type and cost for a particular soil remediation alternative) nor value parameters (e.g. cost of failure, "value of life", discount rate) [Morgan and Henrion, 1990]. Most of the following discussion deals with applying probability analysis to empirical quantities, which in the context of hydrogeology-related problems are usually hydrogeological material properties (e.g. porosity, hydraulic conductivity) distributed across the problem site.

Probability analysis relies on a probabilistic perspective of the real-world. A simple way to conceptualize uncertainty about the spatial distribution of an empirical quantity of interest across a problem site due to limited understanding of the subsurface is to think of the actual spatial distribution as a single realization—the true one—from a set/ensemble of (infinitely) many equally-likely realizations of the property distribution. The pdf of the empirical quantity at any location in the problem spatial domain, then, corresponds to the histogram (frequency distribution) of all the realization values at that location. In predictive statistics jargon, the empirical quantity at each separate location represents an RV, and the set of all the RVs across the spatial domain (i.e. the ensemble of realizations) represents a random field/function (RF).

To model the uncertainty in an empirical quantity across the region of interest using a probabilistic approach—requiring determination of the pdf of the quantity at all locations of interest—the influence of different types of information on the quantity-value uncertainty across the region has to be modeled and, thus, reasonably well understood. At the beginning of an investigation, when little is known about the property (empirical quantity) of interest, the variation in the property values at different locations within a single realization, and at the same location between different realizations, in the ensemble of acceptable realizations is very
large—corresponding to a very wide, non-informative property value pdf (e.g. a uniform pdf with widely separated upper and lower bounds) at all locations.

When information about the empirical quantity at a particular location is obtained, all realizations that do not satisfy the information are eliminated from the ensemble, thus narrowing the width of the pdf at that location. If the information is perfect (e.g. direct measurements with no significant errors) the pdf at that location reduces to a single value—all the remaining realizations are identical at that location; in addition the pdf at nearby locations may be reduced if the quantity is correlated in space—the value of the quantity at distances close to each other tend to be similar. If the information is imperfect (e.g. indirect measurements with/without error, direct measurements with significant errors) the width of the pdf at the location where the information was acquired will be narrowed to some degree, but not reduced to a single value. The amount of uncertainty reduction (narrowing of the pdf) depends on the degree of correlation between the soft measurement and the quantity; this uncertainty reduction corresponds to a reduced variance between realization values at that location. Additionally, the uncertainty in the quantity values at locations surrounding the soft measurement might be reduced, depending on the degree of correlation between the measurement and the quantity of interest and the degree of spatial correlation between quantity values. As more information about the spatial distribution of the empirical quantity is obtained, the ensemble of acceptable realizations is effectively reduced to only those realizations that satisfy all the information.

Therefore, the primary tasks in probability analysis, as applied to hydrogeology-related problems, are (1) the derivation of RV pdfs at all locations of interest across the problem spatial domain (or the ensemble of acceptable realizations) and (2) the updating of these pdfs (or the culling of realizations from the ensemble) as more information about the quantity is acquired at any location in the area of interest. The distinction between these two tasks is really not necessary, since deriving the initial RV pdfs of the RF can be thought of as updating the RF from a state of complete uncertainty. Probability analysis is inherently an iterative process wherein the RF is updated whenever new information about the quantity of interest becomes available.
Hence, a critical aspect of any probabilistic model of uncertainty is a justifiable and robust updating methodology.

### 4.1.2 Probability: Philosophical schools of thought

Before examining the different approaches in probability analysis for characterizing and updating uncertainty in RVs, it is important to highlight the different philosophical theories about probability, since they determine what types of approaches are appropriate. There are two general schools of thought about the definition and treatment of probability: (1) the frequentist school and (2) the Bayesian/subjectivist school. The frequentist school of thought defines probability as the *frequency* an event occurs in a long sequence of trials. Based on this definition, only actual measurements, samples, or observations—not personal judgment—can be used in a statistical analysis to derive pdfs, which should, thus, actually be referred to as frequency distributions. This restriction implies that an assessment of uncertainty cannot be made until a significant amount of data has been collected—a long sequence of trials is performed. Updating of the RF statistics is accomplished by redoing the statistical analysis using an expanded data set containing any new data, in addition to previously acquired data. The Bayesian school of thought defines probability as the *degree of belief* that an event occurs—an inherently subjective assessment of uncertainty. The derivation of pdfs is achieved using all available information, including expert opinion, in addition to measurements, samples, and observations. Updating of the RF statistics when new information becomes available is performed using Bayes' equation for conditional probability, which theoretically accounts for the uncertainty associated with different types of information.

The approach used in this work generally adheres to the Bayesian school of thought, although it is the author's opinion that the philosophical differences between the two schools of thought are not that important in practice, at least in the context of the types of problems addressed in this work. When measurements are sparse (e.g. at the beginning of an investigation) it is still necessary to rigorously account for uncertainty in a risk-based decision-making process,
requiring the use of subjective information (e.g. expert judgment) to infer pdfs of the quantity of interest—an impossibility from a frequentist standpoint. Even when numerous measurements do exist, all geostatistical probability analysis approaches, including the approach used in the case study of this work, require the inference of some aspect of a probability/statistical model (e.g. the covariance function) and, thus, rely on some degree of personal judgment, at least so far as selecting the form of the mathematical function used to represent the model.

A Bayesian approach can also account for uncertainty in the parameter of a selected statistical model for the quantity of interest (e.g. the mean of a Gaussian pdf for some quantity) by treating the parameter as a random variable\(^2\); then the *compound* or *Bayesian* pdf (or certain statistical moments of this pdf) for the quantity of interest, which accounts for the inherent uncertainty in the value of the quantity as well as the uncertainty in parameters defining the quantity value pdf model, can be derived using Bayesian conditioning \([Benjamin and Cornell, 1970; Kitanidis, 1986\]). Treating parameters of a statistical model as random variables themselves is strictly forbidden from a frequentist perspective.

Although the two approaches of uncertainty assessment have fundamentally different philosophical foundations, the descriptions of uncertainty provided by the two approaches for a particular problem should converge as more and more measurements of the quantity of interest are taken; the information provided by the measurements eventually eclipses any less reliable subjective information in the Bayesian approach. Hence, as more measurements become available, or there is very little in the way of subjective prior information (often referred to as a "diffuse prior" pdf \([Kitanidis, 1986\]), the distinction between the two approaches, as applied in practice, becomes blurry. Regardless of which approach or particular method is taken to characterize the RV of an empirical quantity, it is important to realize that the resulting

\(^2\)The use of the classifier "Bayesian framework" in predictive statistics has been a source of considerable confusion—at least for this author. Many authors (e.g. Kitanidis [1986], Morgan and Henrion [1990]) refer to a Bayesian framework as one in which the parameters of an RV are treated as RV's themselves (i.e. given pdfs), and updated when new information becomes available using Bayes' theorem. In fact, any approach that uses Bayes' theorem to update existing RV pdf's can be considered a Bayesian approach, even if only the RV of the empirical quantity of interest is updated (i.e. the uncertainty in the true values of the parameters defining the functional form of the empirical quantity pdf is not treated rigorously).
description of uncertainty about the true value of the quantity—at locations in which direct, hard measurements are not available—is always just an estimate; the uncertainty can never be fully characterized unless no uncertainty exists at all.

4.2 **Accounting for uncertainty in spatial random variables**

This section provides a background on how spatial RVs are used to represent uncertainty in empirical quantities, and how these RVs are updated as new information becomes available, in some commonly employed methods—directly applicable to hydrogeology-related problems. The specific method used in the case study of this work is described in the following section (Section 4.3).

4.2.1 **Derivation of spatial RVs**

There are two principal types of information about a material property of interest that can be used to derive estimates of an RV, including its pdf: (1) measurements and (2) expert judgment (e.g. information about a quantity value based on experience from similar sites). Measurements are generally objective (although this is not always the case, as when the measurement results are interpreted—e.g. interpreting lithological stratigraphy from ground penetrating radar) and, thus, can be incorporated in both frequentist and Bayesian approaches of characterizing probability. Expert judgment is subjective and, hence, can only be used in Bayesian approaches. Ideally, one would like the derivation of pdfs to be completely objective, but this is usually not feasible due to the inherent sparcity of measurements relative to the volume of material that the investigator wants to characterize in a typical problem. Indeed, before any measurements are taken (e.g. at the beginning of an investigation) expert judgment has to be relied on entirely to estimate uncertainty in the quantity value.

For the spatially distributed RVs of interest in hydrogeology-related problems (e.g. a material property distribution across a problem site), most of the information about an unsampled
value comes from observations—measurement values—at nearby locations; this is a result of the propensity for material properties to be spatially correlated within the subsurface. In geostatistics—a branch of statistics that is concerned with phenomena that fluctuate in space, and sometimes in time as well—this spatial correlation characteristic of certain RVs is used to provide additional information about the attribute of interest at locations without measurements by using multivariate statistics (statistics associated with joint probability distributions of two or more RVs at different locations), in addition to univariate statistics (statistics associated with the RV pdf at a single location).

If measurements are used to derive the pdfs of a spatially distributed property, then, obviously, measurements from more than one location have to be pooled together in order to produce meaningful statistics about the property value uncertainty across the site (e.g. histograms, variograms). Otherwise, no variability in the property value would be detected since measurements at a single location result in, or converge on (if there are sampling errors), a single property value from a single realization—that representing the true property value spatial distribution. The required pooling together of measurement data across a region to analyze the property value uncertainty automatically presumes that the pdf, or any other estimated statistics used to model uncertainty (e.g. spatial correlation) in the property value spatial distribution, is the same across that region—an implied condition referred to as "stationarity" in geostatistics. In addition, the statistics associated with the measurements from a single realization (that being reality) are used to model the uncertainty in the material property across all realizations; this is an acceptance of the "ergodic theorem", which states that statistics from a single realization can be used to infer pdfs and other statistics for the full ensemble of realizations.

Therefore, there exists an incongruity between the need to use measurements from different locations across a region in order to glean information about a property value pdf—invoking the assumption of stationarity across that region—and the desire to characterize the local uncertainty in the property value at scales much smaller than the region encompassing all the measurements (e.g. at the scale of an individual measurement). Indeed, using the statistics
ascertained from analyzing a group of spatially separated measurements to describe the local uncertainty in a property value at all locations across a study area, individual "hard" measurements (i.e. measurements with no uncertainty) would not be honored at their locations and the property value pdfs at those locations would exhibit false uncertainty—despite the presence of hard measurements. To overcome this limitation a two step approach of estimating the location-dependent pdfs of a property value is used in geostatistics:

1. First, stationary statistics of the RF are inferred based on an analysis of measurements and/or subjective judgment and assigned to all locations of interest across a predefined region (usually containing the measurements),

2. Followed by local updating of the stationary statistics at specific locations of interest by conditioning on measurement values—in essence, transforming the RF statistics to nonstationary.

Classical statistical regression analysis differs from this geostatistical approach in that it uses a one step approach, wherein the deterministic component of the RV (i.e. the mean) at individual locations is locally updated, but the stochastic portion of the RV (i.e. the description of uncertainty) remains stationary.

In predictive statistics—wherein, based on existed information, the estimated value of an unknown RV is predicted and, equally important in the context of this work, the uncertainty associated with the estimate is quantified—a variety of different methods have been developed for predicting/updating unknown RVs. Prediction and updating are synonymous in the context of this work, since characterizing the entire RV—not just predicting a single value—is required for probabilistic analysis. These methods can be classified based on whether they use (1) a strictly classical statistics (frequentist) approach, (2) a Bayesian approach that simplifies to a classical statistics approach if there is no subjective information, or (3) a full Bayesian approach that even treats the parameters of empirical quantity RVs as RVs themselves—which are updated when new information is obtained. The methods can also be differentiated based on the specific
Chapter 4. ACCOUNTING FOR AND MODELING UNCERTAINTY

1 technique employed for updating—particularly its mechanics, implied assumptions, and approximations.

4.2.2 Classical least squares regression analysis

Classical least squares regression analysis is an example of a predictive statistics method that uses a strictly frequentist approach to RV estimation. The method is designed to estimate values of an empirical quantity from indirect measurements—measurements of another quantity (a secondary variable) that are somehow related to the quantity of interest (the primary variable). Regression analysis requires that a statistical calibration be performed between collocated direct and indirect measurements of the empirical quantity of interest; conceptually, this can be thought of as constructing a graph—a scatterplot or scattergram in statistics jargon—of the secondary variable value versus the primary variable value for each measurement pair. A simple, usually linear, mathematical function is fit to the data pairs such that the expected value of the regression errors squared—the regression error variance—is a minimum. Estimates of the primary variable value at a particular location are obtained by evaluating the fitted function for the primary variable from the secondary measurement value at that location. Hence, estimates can only be produced at locations containing indirect measurements, since regression analysis does not account for spatial correlation between neighboring RVs—the RVs are assumed to be independent.

The regression estimate actually represents a prediction of the conditional expected value of the primary RV from an observation of the secondary RV—the indirect measurement. An inherent assumption is that the relationship between the conditional expected value and the secondary variable value has the prescribed functional form of the particular regression technique being implemented (e.g. linear). If the true relationship between the two varies significantly from the assumed functional form, for example it is not linear when assumed so, then the estimate does not represent the expected value of the conditional pdf, but the best estimate that minimizes the error variance in, for example, a linear least squares sense. The only
measure of uncertainty in the estimated value obtained from regression analysis is the error variance—calculated for the entire calibration set and, thus, assumed to be stationary and independent of the actual value of the secondary variable across the region containing all the indirect measurements. If all these conditions and assumptions hold and the conditional pdf of the primary RV is modeled as normal with stationary variance across the study area, the pdf for each prediction location can be fully characterized from the estimated expected values and error variance. Otherwise, the shape of the pdf cannot be characterized.

If the conditional pdf of the empirical quantity RVs can be fully characterized, a simple extension of the regression analysis is the simulation of equally-likely realizations of the RF using Monte Carlo modeling (i.e. generating random numbers that conform to a probability distribution). Then the realizations can be used as input to dependent system models, enabling characterization of the uncertainty in system model outputs resulting from uncertainty in the empirical quantity values; this can be acquired through a statistical analysis of system output values for all realizations (e.g. constructing histograms). This Monte Carlo simulation is very straightforward, since the regression model assumes the errors in the empirical quantity value estimates at different locations are independent of each other (i.e. spatially uncorrelated).

Regression analysis takes a purely frequentist approach to RV estimation; it provides no way of incorporating subjective information about the empirical quantity being estimated. As more measurements are taken, the RV estimates are updated by simply re-performing the regression analysis with the new measurement results included.

In summary, classical least squares regression analysis is a methodology for estimating the expected value of an empirical quantity of interest at locations where measurements of a secondary variable exist. The methodology has many limitations if used for probability analysis. Estimates cannot be made at locations where there are no measurements; any spatial correlation characteristics of the quantity of interest cannot be accounted for to provide additional information. The only measure of uncertainty is a stationary single regression error variance,
which cannot be used to construct a pdf of the RV at an estimation location unless many assumptions are satisfied.

4.2.3 Geostatistics

Geostatistics provides a wide range of RV estimation techniques that, in general, have more capabilities and greater universal applicability than regression analysis methods. Depending on the geostatistical estimation method employed, any of the three general philosophical approaches to uncertainty characterization can be followed:

- a strictly frequentist approach,
- a Bayesian approach that does not rigorously account for uncertainty in the parameters of the functional model used to represent the pdf of the empirical quantity, or
- a full Bayesian approach that even treats the functional model parameters as RVs with pdfs that are updated with the addition of new information.

All geostatistical estimators, similar to regression analysis, are linear weighted least squares estimators that minimize the mean squared error (or error variance) by varying the weights in a linear combination of all collocated, as well as neighboring, data; this is accomplished while satisfying the constraint that the expected value of the estimation error must be zero. The error variance is defined from previously determined statistical influence (or covariance) functions representing the correlation between the data and the particular RV being estimated. Unlike regression analysis, the covariance functions account for spatial correlation, in addition to the correlation between different data types.

As mentioned earlier, geostatistics takes a two step approach to estimating the initial RF model for a spatially-distributed property value, elaborated on here:

1. The larger-scale statistical process model is first estimated using classical statistical analysis techniques applied to existing measurement data or using Bayesian updating of an initial prior estimate of the statistical process model based on the measurement data (the latter approach is only possible for certain classes of geostatistical estimation methods),
2. Then the RF model resulting from the previous step is locally conditioned on measurements using a geostatistical estimator, as mentioned above.

If a full Bayesian statistical approach is being followed (i.e. the parameters of the RV models are treated as RVs), the form of the conditional geostatistical estimator will be intrinsically different than those used for the other two approaches; the full conditional compound pdf, which accounts for the uncertainty in the empirical quantity RV, as well as the uncertainty in the RVs of parameters defining the empirical quantity RV, has to be estimated (see [Kitanidis, 1986]).

As new measurement data becomes available, the updating of the RF can proceed along one of two paths, depending on the geostatistical approach and the practitioner's confidence in the statistical process model that was previously inferred:

- a two-step approach, similar to that used to estimate the initial RF model or
- a one-step approach that incorporates the updated nonstationary, local statistics from the last updating iteration.

If a strictly frequentist approach is being followed, the two-step updating is identical to the initial RF model estimation process, except that the new data is included in the analysis. If a Bayesian approach is being followed, in the first step of the two-step updating procedure the stationary statistical process model (from the previous updating iteration) is updated based on the new measurements using Bayes' theorem; the local conditional updating step is identical to that used in the initial RF model estimation. The same applies to the full Bayesian approach, except that the RV models of the parameters defining the stationary empirical quantity RV have to be updated in the first step as well.

The one-step approach to updating the RF model estimate locally conditions the individual RVs on neighboring new data, wherein the estimation error and error variance used in the geostatistical estimator is defined based on the nonstationary statistics resulting from the most recent updating iteration (see Harvey and Gorelick [1995] for a hydrogeology-related example of this approach). The procedure is the same for frequentist, Bayesian, and full Bayesian
approaches, and requires calculation and storage of the updated RF nonstationary covariances for all locations of interest across the study area after each updating iteration.

The geostatistical estimation process for conditionally updating an empirical quantity RV at a particular location from an RF, based on neighboring data, is the hallmark algorithm of geostatistics; it is often referred to as "kriging", if only direct measurements of the quantity are considered, or "cokriging", if both direct and indirect measurements are considered. Kriging is a very powerful, useful, and versatile probabilistic estimator. There are many forms and variations of kriging—based on different RF models and requiring different types of statistical information as inputs to the estimation process. Similar to regression analysis, all kriging methods actually provide a minimum variance, unbiased single value estimate from an empirical quantity RV based on a weighted linear combination of neighboring data; the estimate is not necessarily related to any particular statistic of the conditional RV pdf, which is the primary interest in probability analysis. However, there are two characteristic properties of kriging that enable definition of the entire conditional pdf of each RV in the empirical quantity RF for certain cases, depending on the functional model used to represent the RV pdfs or the type of kriging implemented [Deutsch and Journel, 1992]. These cases are described in the next two paragraphs.

If the functional form of the RV pdfs is modeled as multivariate normal (Gaussian), or can be mathematically transformed into that functional form, then the updated conditional pdfs always retain normality and the kriging estimate and estimation error variance (the latter easily calculated during the kriging procedure) equate to the expected value and the variance, respectively, of the conditional pdf at a particular location; these two statistical parameters are all that are needed to fully describe the conditional pdf.

If a binary indicator transform of the empirical quantity RV is performed, resulting in a set of indicator RVs for each location, then the kriging algorithm applied to each indicator RV provides the least squares estimate of the conditional cumulative distribution function (cdf)

---

3 The indicator RV is set to one if the original RV is less than a prescribed empirical quantity value and set to zero otherwise. Indicator RV's are defined for a series of prescribed empirical quantity threshold levels.

4 Mathematically, the cdf represents the integrated pdf.
value for that threshold level—the probability that the empirical quantity is less than the indicator RV threshold value. Therefore, by combining the kriging estimates for each indicator RV at a location, a discrete version of the conditional cdf can be directly estimated; the level of discretization is controlled by the number of prescribed threshold levels.

All kriging methods require, at the very least, a model of the bivariate primary variable covariance—the statistical second moment—for every pairing of locations within the problem spatial domain—defining the spatial correlation between any two primary variable RVs (i.e. RVs at different locations). This covariance model is required for computing the error variance in the estimation process. Most kriging algorithms limit their analyses to the information provided by the first two statistical moments. However, some recently developed geostatistical methods attempt to incorporate higher order statistical moments, based on multivariate pdfs, in the estimation process [Journel, 1993; Guardiano and Srivastava, 1993]; these methods have limited practicality in practice since the inference of multivariate covariance models is extremely difficult.

4.2.3.1 Kriging estimation methods

4.2.3.1.1 Ordinary kriging

One of the most robust and popular kriging algorithms is ordinary kriging, which does not require a prior model of the expected value of the RVs as input in the estimation process. The algorithm internally computes a location-dependent estimate of the mean based on the data used for a particular RV estimation—allowing for a nonstationary prior RF mean [Deutsch and Journel, 1992]. The unbiasedness constraint in the kriging system is satisfied by ensuring the sum of the weights computed for the data used in the estimation is one. Ordinary kriging is usually only used in a frequentist framework, since the means of the RVs and, thus, prior subjective descriptions of their pdfs cannot be specified; although, theoretically, subjective RF covariance models could be specified and updated in a Bayesian sense with the addition of new information.
The ordinary kriging estimation methodology can be easily extended to account for covariate data—measurements of a secondary variable that is indirectly related to the primary variable of interest—by including this data in the weighted linear combination of neighboring data used to generate the empirical quantity RV estimate. For this extension, referred to as ordinary cokriging, the prerequisite information that is used to define the error variance includes (1) bivariate covariance models (auto-covariances) for each type of secondary variable as well as (2) all the combinations of bivariate covariance models (cross-covariances) between the different types of variables, secondary and primary. The required inference of so many covariance models can be very difficult in practice when there is a limited amount of data; indeed, this is the major limitation of ordinary cokriging.

4.2.3.1.2 Simple kriging

Simple kriging is identical to ordinary kriging except that the algorithm requires prior specification of the means for all of the RVs in the RF being estimated. Since the means can be selected or computed independently of the algorithm, the use of simple kriging for local RV conditioning is very conducive to a Bayesian probability analysis approach. Following such an approach, initially subjective information about the empirical quantity would be used to estimate the RV means at informed locations. This initial expected value model can be defined as (1) stationary across the entire RF domain (i.e. a constant mean), (2) stationary across subregions of the domain, or (3) nonstationary. If measurement data is initially available it can be incorporated in the unconditional prior estimate of the mean by using Bayes' theorem to update the subjective mean model—requiring the imposition of stationarity across the regions in which data are averaged; this is the most statistically robust way to infer the mean if both subjective information and measurement data is available. Simple kriging is then used to locally update the empirical quantity RVs at every location of interest by conditioning on any neighboring measurement data. Additional updating on new measurements is accomplished using either a two-step or one-step geostatistical updating methodology, as described earlier. If a full Bayesian approach is being
followed the form of the simple kriging estimator will be different in order to estimate the compound pdfs of the empirical quantity RVs.

Identical to ordinary kriging, the extension of simple kriging to simple cokriging (i.e. including covariate data in the kriging estimator) is simple and straightforward; however, as explained for ordinary cokriging, the full form of simple cokriging requires the inference of many first and second statistical moment models.

4.2.3.1.3 Universal kriging

Universal kriging (also referred to as kriging with an external drift) is a more general and comprehensive kriging method than simple or ordinary kriging that can rigorously account for a variable mean model of the RF (e.g. a linear trend in the mean). This method of kriging requires inference of a generalized covariance function model⁵, as well as specification of the functional form (e.g. a second order polynomial with unknown coefficients) of the RV mean model. Ordinary kriging is a special case of universal kriging where the functional form of the mean is a constant. Since the mean model represents the deterministic part of the RV, specifying a more detailed and complicated functional form of the mean model implies that more information is known about the spatial variability of the property of interest—placing less reliance on the stochastic part of the RV to describe the spatial structure [Kitanidis, 1992]. Placing a heavier reliance on the deterministic portion of the RV is appropriate and advantageous for certain applications (e.g. the geostatistical approach to the groundwater inverse problem—see Appendix A, Section A.4.3.3), in that more information is provided about the spatial structure. In the case study of this work the spatial structure of the properties of interest are highly uncertain, making the use of a model with a variable mean inappropriate; consequently, universal kriging will not be discussed any further. Useful references on universal kriging and generalized covariance functions include Kitanidis [1993] and Deutsch and Journel [1992].

⁵A simplified version of the customary covariance function where redundant parts, resulting from the mean (deterministic) model being more informative, are neglected [Kitanidis, 1993].
4.2.3.1.4 Indicator kriging

Indicator kriging, which simply corresponds to the kriging of indicator RVs (IRVs) using any of the previously mentioned methods of kriging, is a very versatile and flexible form of kriging—very conducive to probability analysis applications since the kriging estimate represents an actual probability value. It can also rigorously incorporate a wide range of "soft" information types—indirect measurements, subjective beliefs, measurements with errors—if the Markov-Bayes methodology [Alabert, 1987; Zhu and Journel, 1993] is employed. For these reasons indicator kriging is the geostatistical estimation method of choice for this work. IRVs are binary RVs that can only take on the values of zero or one. They can be used to represent:

- categorical variables, such as geologic rock types, in which case the IRV is defined to take on a value of one if the variable is in a certain category (e.g. geologic rock type) and zero otherwise, or
- continuous variables, in which case the IRV is defined to be one if the variable is less than or equal to a threshold value, greater than a threshold value, or within an interval—the choice of criteria depending on the application—and zero otherwise.

A very attractive property of IRVs is that the expected value of an IRV is the probability that the IRV is equal to one. Therefore, since the kriging estimate is an expected value estimate, indicator kriging directly provides the least squares estimate of the conditional probability that a certain event (or condition) exists (or will occur)—the event being defined by criteria fixed in the definition of the IRV. The estimate is based on neighboring measurement observations and prior statistical model(s) of certain aspects of the IRV (e.g. indicator covariance model, indicator mean model). As mentioned earlier, a discrete form of the conditional pdf or cdf can be constructed from the indicator kriging estimates for a series of IRVs defined over a range of threshold levels or categories, each IRV corresponding to a single level/category. This conditional probability estimation is accomplished without any restrictions placed on acceptable functional forms of the empirical quantity RV pdf model, as required for kriging of a continuous RV (see above); the mathematical form of the model does not have to be specified since the pdf is effectively
discretized. This property of indicator kriging is extremely attractive for probability analysis, wherein the objective is to characterize the uncertainty associated with an empirical quantity value using probabilities.

When the two step approach of estimating the RF model for a spatially-distributed property value is applied to IRVs, the larger-scale statistical process model expected value—required for simple kriging—is estimated/inferred in the first step, corresponding to the overall probability of the empirical quantity of concern falling within a defined range of values or category. Thus, uncertainty is inherently represented by the inferred IRV expected value; a discrete representation of the RF global pdf/cdf is provided by combining the inferred expected values for all the IRVs. Then, employing simple indicator kriging (SIK), these global probabilities of occurrence are directly updated at a particular location from the information provided by nearby observations—effectively updating the discrete pdf/cdf. Thus, empirical quantity uncertainty and the uncertainty reduction resulting from improved information are implicitly handled by employing IRVs and indicator kriging. In fact, if the probability of occurrence is only conditioned on one observation, it can be shown that SIK simplifies to an exact statement of Bayes' relation for conditional probability [Alabert, 1987]; for more than one observation SIK has been shown to be a close approximation of the Bayesian conditional probability [Solow, 1986].

Many different types of both perfect (hard) and imperfect (soft) information about an empirical quantity can be encoded in the IRV framework, and incorporated in indicator kriging. Hard information at a particular location (e.g. direct observation of the quantity with no measurement errors) is simply encoded as either ones or zeros in the IRVs at that location, corresponding to whether the value of the empirical quantity falls within or outside, respectively, the range of values defined in the IRV. In addition, to use the hard information in indicator kriging—to update the probabilities of occurrence at uninformed locations—requires specification of an indicator covariance model for each indicator category/cutoff level.
Chapter 4. ACCOUNTING FOR AND MODELING UNCERTAINTY

There are several different ways of encoding soft information, depending on the type of information and the specific indicator kriging algorithm implemented. If the quantity value at a location is only known to be within a certain range of values, the IRVs defined above and below the upper and lower bounds of the known interval of occurrence at that location are assigned zeroes and ones, respectively; the IRVs defined within the known interval remain uninformed (i.e. they are not assigned a value). The same indicator covariance model specified for the hard data can be used to update probability estimates at locations nearby the interval data, but only for indicator cutoffs outside of the known interval of occurrence in the data; probabilities corresponding to cutoffs within the known interval are unaffected by the data.

If the source of information is indirect data, the IRVs at the data locations are assigned either ones and zeroes or values between zero and one, depending on the specific algorithm used (see Section 4.3.1.2 below). The covariance model for the secondary variable—the variable measured by the indirect measurement—and the secondary-primary variable cross-covariance model have to be specified for each indicator cutoff/category, in addition to the primary variable covariance, to perform indicator cokriging; these models are used to update the primary variable probabilities of occurrence based on the indirect measurements. Alternatively, methods have been developed for deriving these additional covariance models from the primary variable covariance model—based on a calibration between collocated measurements of the secondary and primary variable [Alabert, 1987; Zhu and Journel, 1993]; such methods considerably reduce the burden placed on model inference. These methods are explained in the following section (Section 4.3) on the Markov-Bayes geostatistical estimation approach.

Finally, qualitative information (e.g. expert judgment) can be incorporated as a prior pdf model for all locations across the region being informed—represented by assigning values between zero and one at each category/cutoff level for all the affected IRVs. These probabilities should be treated as expected values that are updated based on neighboring hard and/or soft data in the indicator kriging procedure [Alabert, 1987].
Another attribute of indicator kriging is the ability to account for spatial correlations that vary depending on the value of the variable, since each IRV cutoff level is treated separately. This property enables much better modeling of the spatial distribution and variability of extreme values of an empirical quantity than, for example, Gaussian methods, which only model the spatial correlation at the center of mass of the RV. Extreme values are often of great concern in engineering problems, since they are often the values that result in engineering failure.

4.2.3.2 Geostatistical simulation

The map generated from kriging is strictly an expected value estimate of an empirical quantity RF. As a result, it is typically a very "smoothly" varying depiction of the spatial distribution. While the kriging estimate can be very useful for determining the nature of the conditional pdf of the quantity of concern, it is incorrect to assume that the map of kriged values is a valid representation, or realization, of what the true property value spatial distribution might look like. Such a realization is very important in hydrogeology-type engineering problems, because it is required as input to predictive models—models of the natural and engineered systems (e.g. a realization of hydraulic conductivity is required for input into groundwater flow and contaminant transport models). If equally probable realizations of the quantity value RF can be generated, and used to producing equally probable system model responses, then uncertainty in the model outputs—resulting from the uncertainty in the quantity value—can be analyzed from the variability in the output results. This is the concept behind, and the attraction of, geostatistical simulation—the simulation of any number of equally probable realizations of a spatially distributed empirical quantity.

Many methodologies have been developed for performing geostatistical simulation. Some of the more popular and well accepted simulation approaches are based on sequentially applying a Monte Carlo random number generator for each simulation value. Since this is the approach taken in this study, it will be the only simulation methodology discussed here. This fairly straightforward approach can be described as a set of iterative steps:

1. A random node (or cell) is chosen for simulation.
2. The conditional cdf at the node is derived by calculating the kriging estimate conditioned on neighboring measurements and previously simulated nodes.

3. A random number between zero and one is generated.

4. The random number is taken as a cdf probability. The quantity value corresponding to this number in the cdf generated in step 2 is taken as the realization value at this node.

5. Steps 1-4 are repeated until all nodes have been simulated producing a single realization.

6. Steps 1-5 are repeated until the desired number of realizations has been generated.

The two working algorithms of this sequential approach are (1) Monte Carlo random number generation and (2) conditional kriging.

The only requirement for using the sequential Monte Carlo simulation approach is that a conditional cdf can be generated from kriging. However, this requirement limits the geostatistical simulation to empirical quantities that have Gaussian pdfs or a pdf that can be transformed to Gaussian; alternatively, an indicator transform can be applied to the quantity and a form of indicator kriging implemented in the simulation procedure. The simulation approach implemented in this work is sequential indicator simulation, using the Zhu and Journel [1993] M-B indicator kriging algorithm. This approach enables the incorporation of indirect measurements (e.g. geophysics).

4.2.4 Scale

Before discussing the specifics of the geostatistical estimation methodology used in this work (Markov-Bayes indicator cokriging), it is important to consider the issue of scale and its implications on probability analysis, especially as concerns geostatistical approaches. The scale at which a probability analysis is performed can significantly affect the results of the probability analysis, as well as the engineering analysis. Hence, a critical question that should be addressed before developing a probabilistic model for the important empirical quantities in a hydrogeology-related engineering problem is, “what is the subsurface scale of concern in the problem?”. There are several points to consider when addressing this question, including (1) what subsurface scales
have significant influence on the engineering problem, (2) what scales can be measured, and (3)
what scales can be adequately modeled. Often there has to be a tradeoff between the scales
arising from these different, sometimes conflicting, perspectives.

For example, in reference to the hypothetical problem introduced at the start of CHAPTER
2, if substrate matrix contamination is the primary concern at the site prior to development, and
excavation and treatment of subsurface matrix material is the only remediation alternative for
highly contaminated soil (Note: this is actually the situation for the case study used in this study),
the smallest engineering spatial scale of concern could be determined by the volume of material
extracted in a single scoop of a backhoe; it is impractical to extract smaller volumes. The
primary engineering decision then—whether or not to excavate a backhoe volume from a
particular location—is determined by whether or not that volume of subsurface material is
considered contaminated enough to warrant remediation.

However, the measure of contaminant concentration, whether it is volumetric or mass
derived, is inherently scale dependent since substrate contamination is heterogeneous by nature;
should the concentration be measured across the entire volume/mass of the single backhoe scoop
or at a smaller scale, stipulating excavation if the maximum contaminant concentration in the
material is above a preset threshold level? Of course the smallest scale of contaminant
characterization is limited by the smallest measurement scale. For this particular problem, the
choice of the scale of characterization to use for making decisions (which is referred to as the
decision scale in this work) is mostly a public health and regulatory (i.e. political) issue. A
reasonable choice for the decision scale in this case would seem to be the smallest sample scale
used to obtain accurate and efficient contaminant concentrations. Thus, a particular batch of
material, equal in size to the backhoe scoop volume, would be designated for excavation removal
if the maximum concentration measured in samples taken from it is above the preset threshold
level.

Geostatistical estimation methods have been developed to estimate an RV at a larger or
smaller scale than the measurement scale, and include block kriging or averaging small scale
stochastic simulations [Deutsch and Journel, 1992] and incorporation of spatial filtering [Hachich and Vanmarcke, 1983], respectively; however, these methods should be used very carefully since the meaning and usefulness of their results is dependent on many factors specific to the particular problem being addressed.

Point kriging produces a point estimate of an RV at any selected point location, conditioned on neighboring measurement data that, likewise, are effectively considered to be point measurements in the algorithm. Block kriging, on the other hand, can estimate an upscaled RV value, representing an average of the point estimates of the RV across a defined region, by effectively doing just that—taking an average of kriged point estimates; this can be accomplished by calculating a point-to-block covariance, instead of a point-to-point covariance, between the data and estimation locations for use in the kriging equations [Deutsch and Journel, 1992]. However, if block kriging is used with IRVs the resulting estimate is not an estimate of the probability of the average RV meeting the indicator criteria (i.e. falling below a threshold value), but an estimate of the proportion of point values within the defined region that meet the indicator criteria. If an estimate of the probability of the average RV meeting the indicator criteria is desired, equally probable realizations of the RV small-scale distribution within the region in which an average is desired can be generated using stochastic simulation techniques, after which the simulated point values are averaged for each realization; the proportion of averaged values across all realizations meeting the indicator criteria represents an estimate of the probability desired (i.e. the conditional expected value of the large-scale IRV) [Deutsch and Journel, 1992].

If the measurement scale is larger than the scale of interest for geostatistical estimation the larger volume of influence and lower resolution of a measurement can be accounted for in the conditional updating of the RV by incorporating a spatial filtering function associated with the measurement in the kriging equations. A spatial filtering function is a mathematical description of the influence of small-scale parcels of the subsurface on the larger-scale measurement. Accounting for the measurement volume of investigation in this manner requires that the spatial
filtering function be defined; it is often difficult to characterize this function, especially since the function can vary depending on the heterogeneity and magnitude of the spatial quantity being measured.

4.3 Probability analysis using the Markov-Bayes indicator estimation methodology

This section describes how the Markov-Bayes (M-B) indicator geostatistical estimation methodology can be used, and is used in this work, to perform probability analysis. The steps required to accomplish this are illustrated using the hypothetical example problem. In addition, the problems and limitations of the methodology are highlighted, and the means in which they are addressed and attempted to be overcome in this work are detailed.

4.3.1 Model inference

The first step in applying any geostatistical estimation method to a characterization problem is the inference of all statistical spatial moment models required for implementation of the method; prior to and in conjunction with this step is the choice of model types and assumptions (e.g. stationarity across the site) to use in the problem. In order to implement M-B estimation algorithms it is necessary to infer spatial covariance models—primary and secondary variable spatial covariances, as well as cross covariances—and the primary variable expected value model; the latter is not required if an ordinary kriging/simulation algorithm is being used. A critical step in the M-B model inference process, required to incorporate secondary variable information, is the development of a secondary to primary variable calibration model. This is usually accomplished by performing an experimental calibration of secondary variable (soft) measurement data with collocated primary variable (hard) data.

Model inference can be separated into two general steps: (1) choosing overall spatial RV functional models for the quantity of concern (e.g. stationary SRV with an isotropic spherical covariance to represent primary variable spatial correlation) and (2) inferring the variables (or
parameters) within the models (e.g. the mean of the stationary SRV and the variance and correlation length of the spherical covariogram). The first step is best accomplished by closely interrogating all available information about the site under investigation—commonly referred to as exploratory data analysis. The second step is accomplished through more rigorous statistical analysis of data about the quantity of concern across the site.

As discussed earlier in this chapter, there are two philosophical approaches for inferring a population’s statistical model parameters as information becomes available: the frequentist and Bayesian updating approaches. These approaches differ in that, following the frequentist approach, statistical parameters are estimated using all measurements every time the calculations are performed (e.g. when more measurements become available the calculation is performed again using all previous and new measurements together). Usually only direct, “hard” measurements are used to calculate the estimates, but algorithms for incorporating indirect, “soft” measurements and subjective information along with “hard” measurements—within a Markov-Bayes IRV framework—have been developed, and will be discussed later in this section. Following the Bayesian approach, statistical parameter estimates are updated by applying Bayes’ theorem, whereby only new information, along with the previous estimate value(s), are required in the updating. The Bayesian approach is very amenable to incorporating subjective information in the parameter inference, especially qualitative expert opinion (e.g. geological intuition of the continuity of a clay aquitard) at the beginning of an investigation, before a substantial amount of measurements are taken. As more and more measurements are taken, parameter estimates inferred using the Bayesian approach should converge with those inferred using the frequentist approach. The following discussion will focus on the frequentist approach, since this is the approach used in this study.

4.3.1.1 Exploratory data analysis

Before any statistical moment models can be inferred, the choice must be made about which geostatistical estimation methodology to apply to the problem and what models and assumptions are most appropriate for the chosen estimation method. An exploratory analysis of
existing relevant data from the site should be made to help in these decisions. The objective of this exercise is to glean as much information as possible about the overall magnitude, variability, and spatial distribution of the property of interest across the problem site. This is accomplished through the use of informative plots and the calculation of summary statistics.

A three dimensional (3-d) plot of all the measurement data collected across the problem site is a valuable tool for qualitatively mapping spatial trends in the quantity of interest across the site. The visual information provided by such plots can be used to determine (1) if the entire site can be approximated as stationary, or (2) if there are separate, distinct populations in different areas of the site that might need to be treated separately in the geostatistical analysis, or (3) if the entire site should be treated as nonstationary.

Another family of plots that are useful for examining the univariate statistical distribution of the overall population is the histogram, or frequency distribution—the “experimental” estimate of the pdf—and variants of the histogram (e.g. the ogive—the experimental estimate of the cdf). The shape of the histogram largely determines which geostatistical estimation (kriging) methodology is most appropriate for the property of interest. In fact, as mentioned earlier in this chapter, non-indicator kriging methods only provide true expected value and conditional pdf/cdf estimates when the quantity being estimated has a normal (Gaussian) distribution; obtaining the best possible estimate of the quantity pdf is critical for determining whether these methods, which include non-indicator simulation algorithms since they rely on kriging, can be used. If it is determined that the quantity of interest does not have a normal distribution, either (1) a transform has to be applied to the quantity in order to force it to retain a normal distribution (e.g. log normal transform\(^6\)), after which it is valid to apply a non-indicator geostatistical method to the transformed quantity, or (2) an indicator geostatistical estimation method needs to be used for the quantity.

\(^6\) A log normal transform involves taking the logarithm of all the data values (the histogram of the transformed data should approximate a normal distribution) and incorporating them in the non-indicator estimation method of choice to estimate the logarithm of the quantity at all locations of interest. The estimated values can then be back-transformed to the original quantity.
If an indicator geostatistical estimation method is used (as in this study), there are no restrictions on the shape of the quantity value pdf. In this case, the histogram and ogive are useful for determining what indicator cutoffs should be used and how the tails of the cdf should be modeled, the latter only necessary if continuous values are to be estimated.

The ogive can be used as the input prior cdf for an indicator geostatistical estimation algorithm, wherein the ogive frequency values, defined as a percentage of the overall sample number, represent the expected value of the measured quantity being less than or equal to the threshold value immediately to the right in the ogive. However, as will be discussed below, it is often best to try to correct for spatial sampling bias—clustering of measurements in zones with similar quantity values—when estimating statistical process means; simple, equal data weighting averages (e.g. the arithmetic average) do not account for this bias.

Other plots that are helpful in exploratory data analysis include the box and whisker plot, which summarize many important statistical parameters of a data set in a single intuitive plot.

4.3.1.2 Secondary variable calibration model

In order to take advantage of information about the quantity of interest provided by measurements of a secondary variable using the M-B estimation approach (which is one of primary reasons for using this approach), a secondary to primary variable calibration model must be developed; this calibration model is required for M-B cokriging. The most objective way to infer such a model is to perform a calibration between collocated measurements of the secondary and primary variable. Ideally, the calibration data set is (1) obtained from the actual field site and (2) sufficiently large enough, with calibration pairs for the full range of primary variable values, to provide an accurate representation of the relationship between the secondary and primary variable. Depending on the M-B methodology being used, (1) Alabert [1987] or (2) Zhu and Journel [1993], the parameters required in the calibration model and, thus, that need to be calculated from the calibration, are, respectively, (1) a quality factor or (2) a quality factor and probabilities of the primary variable being within defined interval categories when the secondary variable falls within a certain interval. These parameters will be clarified later.
Chapter 4. ACCOUNTING FOR AND MODELING UNCERTAINTY

The calibration procedure for the M-B approach is, first, to generate a scatterplot, or scattergram, of collocated secondary versus primary variable measurement values (see Figure 4.2.2 for an example of such a scattergram). Indicator thresholds are applied to each variable (or axis on the scattergram) and the number of points (calibration pairings) within each primary-secondary variable combination of indicator intervals is tallied. From these summations within the different sectors of the scattergram, the conditional probabilities of the primary/secondary variable falling between a range of values, given that the secondary/primary variable at the same location falls within a range of values, is derived (see Figure 4.2.2). From these conditional probabilities the secondary variable classification probabilities P1 and P2 can be derived, which are used to derive the secondary variable quality factor mentioned above. P1 and P2 are, respectively: (1) the probability that the secondary variable is within a secondary variable indicator interval when the primary variable is within a certain primary indicator interval, (2) the probability that the secondary variable is within the same secondary variable indicator interval when the primary variable does not fall within the primary indicator interval noted above. P1 and P2 are calculated for each primary variable indicator threshold value. The secondary variable quality factor, one for each threshold, is simply the difference between P1 and P2. This quality factor is used as multiplicative factor with the primary variable auto-covariance model to derive (1) the primary-secondary variable cross-covariance model (which defines the spatial correlation between the secondary and primary variable) and (2) the secondary variable auto-covariance model. All these models are required to perform cokriging.
Chapter 4. ACCOUNTING FOR AND MODELING UNCERTAINTY

Figure 4.2.2: Example of a hypothetical calibration scattergram. In this example there are two indicator classes. Each point represents a calibration pairing where both the primary and secondary variables were measured at the same location. $A_1$ and $A_2$ ($B_1$ and $B_2$) are the primary (secondary) indicator classes; the line separating the two is the quantity value demarcating the two classes. $S_1$ though $S_4$ are the sum of the calibration pairings within each sector of the scattergram.

The derivation of the classification probabilities is different for the Alabert [1987] and Zhu and Journel [1993] M-B approaches; this is a consequence of the latter approach permitting a different number of indicator threshold cutoffs (or indicator classes when a discrete RV is being considered) for the primary and secondary variable, unlike the former where the number has to be the same.

In the Alabert [1987] M-B approach, the classification probabilities are calculated directly from the scattergram by comparing the number of calibration pairings within different sectors of the scattergram. Referring to Figure 4.2.2, $P_1$—the probability of measuring a secondary variable value within the secondary indicator interval $B_1$, given that the primary variable is within the primary indicator class $A_1$, or in mathematical notation $P[B_1 \mid A_1]$—is simply the number of calibration pairings in sector $S_1$ of the scattergram divided by the sum of the pairings in sectors $S_1$ and $S_3$ (or $P[B_1 \mid A_1] = S_1 / (S_1 + S_3)$). Similarly, $P_2$—the probability of measuring a secondary variable value within the secondary indicator interval $B_1$, given that the primary
variable is not within the primary indicator class $A_1$ (which, in the case of only two indicator classes, infers that the primary variable is within class $A_2$), or $P[B_1 | A_2]$—is simply the number of calibration pairings in sector $S_2$ of the scattergram divided by the sum of the pairings in sectors $S_2$ and $S_4$ (or $P[B_1 | A_1] = S_2 / (S_2 + S_4)$). Since there is only one indicator threshold value in this example, only one set of these classification probabilities has to be determined. If more indicator thresholds are used, similar calculations would be performed to determine $P_1$ and $P_2$ for these additional indicator thresholds.

The Zhu and Journel [1993] M-B approach requires a more involved analysis for calculating $P_1$ and $P_2$ for each primary variable indicator class; they cannot be calculated directly from ratios of the sectors in the scattergram, since the number of primary variable indicator classes can be different than the number of secondary variable ones. First, the conditional probability of the primary variable being within the primary class $A_1$ given the secondary variable is within $B_1$ (which will be referred to as $Y(i=1; j=1)$, where $i$ and $j$ correspond to primary and secondary indicator classes, respectively) is calculated from the scattergram (or $P[A_1 | B_1] = S_1 / (S_1 + S_2)$ as defined in Figure 4.2.2); the conditional probability $Y(i;j)$ for all the primary-secondary indicator threshold combinations are calculated—only two in this case. These conditional probabilities represent “prior” information about the primary variable at locations where secondary variable information exists. This “soft” information, in the form of conditional probabilities between 0 and 1, is used directly as conditioning information in the indicator cokriging estimate of the primary variable at coincident and nearby locations; $P_1$ and $P_2$ are also required for this estimate.

In the Zhu and Journel [1993] M-B approach, $P_1$ and $P_2$ are calculated by taking the expected value of $Y(i;j)$ given that the primary variable is within a primary indicator class and not within that class, respectively; these are calculated for all primary indicator thresholds. For the calibration example illustrated in Figures 4.2.2:

$$P_1 = E\{Y(1; j) | B_1\} = (Y(1; 1) \cdot S_1 + Y(1; 2) \cdot S_3) / (S_1 + S_3)$$
$$P_2 = E\{Y(1; j) | B_2\} = (Y(1; 1) \cdot S_2 + Y(1; 2) \cdot S_4) / (S_2 + S_4)$$
Since in this example only one primary indicator threshold is used, only one set of classification probabilities is required. Unlike the Alabert [1987] M-B approach, any number of secondary variable indicator thresholds can be used, regardless of the number of primary indicator thresholds.

In summary, the calibration model required by both M-B approaches to perform cokriging can be derived directly from a scattergram of collocated primary versus secondary measurement values. The methodology for deriving the quality factors and conditional probabilities from the scattergram, which are used to define the information that the secondary variable measurements provide about the primary variable in the calibration model, is different for the Alabert [1987] and Zhu and Journel [1993] approaches. The Zhu and Journel [1993] approach directly uses, as soft data indicator values in the indicator kriging algorithm, the conditional probabilities of the primary variable being within a range of values (defined by the chosen indicator thresholds) given that the secondary variable falls within defined ranges, in addition to the secondary-primary variable correlation quality factors (P1 and P2) used in the Alabert [1987] approach; soft data indicator values are always either 0 or 1 in the Alabert [1987] approach, depending on whether measured secondary variable is above or below the defined indicator threshold. Using this additional information in the Zhu and Journel [1993] approach provides the flexibility of having a different number of primary and secondary thresholds/classes; this flexibility enables customization and optimization of the calibration model to glean the most information from indirect measurements based on the (1) character of the relationship between the secondary and primary variable and (2) the completeness of the calibration data set.

It should be emphasized that the validity of the M-B calibration model is highly dependent on the integrity of the calibration data set. The calibration data set should be representative of the primary and secondary variable RF distributions across the study area. There should be enough calibration pairings to ensure that the quality factors calculated from the scattergram are not mistakenly good due to an insufficient, small statistical pool.
Exactly how the secondary variable auto-covariance and primary-secondary variable cross-
covariance models are derived from the above mentioned conditional probabilities and quality
factors, in combination with the primary variable auto-covariance model, is discussed later in
this chapter on covariance model inference (Section 4.3.1.4); a detailed mathematical discussion
is also provided in Appendix B.

4.3.1.3 Prior expected value

Geostatistical estimation methods that rely on simple kriging require the inference of the
quantity value overall population expected value(s) prior to estimation. A very common way of
estimating the expected value is to take the average (usually the arithmetic average) of all the
hard (primary variable) data; every data value is given the same weight in such an estimate. If
there is no subjective information (e.g. expert opinion based on similar sites) about the mean and
the measurements of the quantity were sampled randomly across the site, this hard data average
is as valid an estimate of the expected value as can be ascertained. However, measurements are
typically not sampled randomly, instead being sampled as biased clusters in “areas of interest”.
This sample clustering especially tends to be true for contaminant characterization in
environmental problems, wherein contaminant “hot spots” are densely sampled compared to
“less interesting” clean zones. While such a sampling scheme is useful for delineating individual
plumes, an equally weighted average of the samples undoubtedly results in a biased estimate of
the overall population expected value.

A simple, versatile, and effective method of correcting for skewed estimates of the prior
expected value resulting from spatially biased data sampling is to calculate a declustered mean
using a technique known as declustering analysis [Deutsch and Journel, 1992]; this technique is
used in this study. Declustering analysis employs an iterative process of calculating weighted
data averages, wherein the data weights are determined by overlaying a grid with uniform cell
sizes on the problem area; the weight applied to each datum is inversely proportional to the
number of data in the cell where the datum resides. The weighted data average calculation is
repeated for a range of cell sizes; the smallest or largest average, depending on whether the
measurement sampling was biased towards areas of high or low values, respectively, is taken as the declustered mean for the data set. For example, if measurements of substrate contaminant concentrations were clustered in areas of high concentrations (which is a typical scenario), the lowest weighted average should be chosen as the declustered mean; if an indicator approach is being used, the average probability corresponding to the lowest probability of contamination (e.g. the highest probability of being below a threshold, if a cdf representation is being used) should be chosen. This declustering methodology is a heuristic approach and, thus, can be sensitive to the range of cell sizes and the grid placement on the problem domain used in the calculation; care should be taken to compute the weighted average for enough combinations of these parameters such that the optimum, or the near-optimum, declustered mean is intercepted.

Subjective information, such as expert judgment about the mean, can be incorporated with hard measurements to estimate the statistical process mean using Bayesian updating; however, such updating is only practical if certain assumptions are met. If the quantity of interest is represented using a continuous (non-indicator) RV, the RV must be Gaussian for practical implementation of Bayesian updating [Kitanidis, 1986]. James and Freeze [1993] present a methodology for Bayesian updating of the mean of an indicator RV that incorporates indirect “soft” measurements and expert judgment, in addition to direct “hard” measurements, using the Alabert [1987] M-B framework. However, the algorithms used in this methodology assume the soft data form an independent data set from the hard data, potentially causing artificially inflated weighting of the soft data if the two data types are strongly spatially correlated [James and Freeze, 1993]. While this IRV Bayesian updating approach was not used in this study due to its relative complexity and ad hoc, not yet fully proven and developed status, such a Bayesian method could be very versatile and useful for rigorously incorporating soft data in the statistical model inference process.

Certain types of “soft” data can also be incorporated in the estimation of indicator means using a frequentist approach within the MB indicator RV framework. This can be accomplished by taking a weighted average of the hard data indicator mean and the mean of “corrected” soft
data indicator values, the means calculated independently (e.g. using declustering analysis). The weights are a function of the number of hard versus soft data (and sum to one) and the soft indicator "correction", applied to each soft data indicator value before calculating the soft mean, is a function of the quality of the soft data—defined by the quality factors that are determined from the secondary variable calibration [Alabert, 1987]. A methodology for estimating the mean of an IRV using indirect (secondary variable) and direct (primary variable) measurements is derived in Alabert [1987] and applied in a real world study in Copty and Rubin [1995]. This methodology is only strictly valid for the Alabert [1987] M-B approach.

In the Zhu and Journel [1993] approach the expected value of the soft indicator data is actually a direct estimate, itself, of the primary variable IRV mean. The quality of the indirect measurement is implicitly accounted for in the soft indicator values—derived from all secondary variable measurements in the calibration (see above). As the information about the primary variable provided by the indirect measurements becomes less and less (lower quality or lower correlation), the soft indicator values approach the hard indicator means; thus, their averages approach the hard indicator means.

A logical variation, then, of the Alabert [1987] indicator mean estimation methodology applicable to the Zhu and Journel [1993] M-B approach is to calculate the weighted average of the hard and, this time, raw (uncorrected) soft indicator means, since the quality is already accounted for; again, the weights are made proportional to the relative number of data for each measurement type.

In summary, the estimation of indicator RV expected values is analogous to the estimation of continuous RV expected values, except that "soft" information can be much more easily incorporated in the IRV expected value estimation. With both RV types, care should be taken to account and correct for non-random sample bias in the data set used for the estimation, which

---

7 This convergence will only occur if the primary variable measurements in the calibration data set (used to calculate secondary variable indicators) is representative of the primary variable indicator means inferred for the overall stationary population (i.e. the equally weighted means of those data values is identical to the inferred means). This is a limitation of the Zhu and Journal [1993] approach, and is discussed in the "Limitations" section (4.3.3) of this chapter.
can be successfully accomplished using declustering analysis. This is especially true for engineering problems that use, and are sensitive to, the expected values of IRV classes—which represent probabilities of the quantity of concern being within defined intervals—in a risk-based decision-making process. For example, the case study in this work centers on determining the worth of additional data, and choosing remediation alternatives, for the cleanup of contaminated soil, based on the probability—IRV expected value—of soil being contaminated above the regulatory threshold. The risk-based decision-making framework for this problem is very sensitive (i.e. costs fluctuate significantly) to small changes in the overall expected values of soil contaminant concentration threshold IRVs (probabilities of being above the regulatory threshold levels).

4.3.1.4 Covariance models

The inference of spatial covariance models should be an iterative three step process: (1) calculating the experimental covariance from measurement data, (2) fitting a mathematical model to the experimental covariance, and (3) validating the covariance model [Kitanidis, 1991]. When using an indicator random variable geostatistical approach, such as the M-B approach, a covariance model is needed for each indicator class or cutoff, each one requiring repetition of this inference process. In addition, when implementing a cokriging geostatistical estimation approach a secondary variable covariance and primary-secondary variable cross-covariance is required for each secondary variable used, at each indicator cutoff! The Markov-Bayes indicator RV approach eliminates the need for inferring all these additional covariance models by deriving them from the primary variable indicator covariance model—scaled using soft-hard data calibration parameters.

The objective of the inference process is to choose a representative, relatively simple covariance equation that conforms with all available information about the quantity distribution across the study site, as well as provides the best prediction of data values when the geostatistical kriging methodology of choice is invoked. While certain parts of this process can be automated and made objective (e.g. the fitting of covariance equations to data by varying certain parameters
in the equations can be automated through mathematical optimization), certain aspects of the process require user input and the overall process should be iterative in nature—testing and tweaking the models over and over again to improve their overall conformance with the data.

The covariance model inference process is heavily reliant on available data; the uniqueness and integrity of an inferred covariance model depends on the number and distribution of data that can be used in the inference procedure. While for most geostatistical methods only "hard", direct measurements of the primary variable can be used to infer the covariance model for that variable, Alabert [1987] presents a methodology for using information provided by "soft", indirect measurements as well, in an M-B indicator RV framework. Similar to the Alabert [1987] methodology for incorporating soft data in the estimate of the mean, the primary variable covariance is calculated by taking a weighted sum of the primary variable indicator auto-covariance, primary-secondary variable indicator cross-covariance, and secondary variable indicator auto-covariance; "correction" factors, derived from the calibration misclassification probabilities, are applied to the latter two covariances. This methodology can be very valuable for improving the integrity of the covariance model inference when there is a lot more indirect than direct measurement data; Copty and Rubin [1995] use the methodology to greatly improve the estimate of a lithology indicator covariance model by incorporating information provided by seismic data having full site coverage, supplementing the limited spatial information provided by core logs in a few boreholes. To the author's knowledge, a similar methodology has not been applied to the Zhu and Journel [1993] M-B approach; such a methodology may not be possible since indirect measurements are represented by cumulative probabilities, instead of binary indicators.

4.3.1.4.1 Computation of experimental covariance

The first step in covariance model inference is to calculate the experimental covariance from measurement data. This entails computing the covariance between all data pairs and binning the results by user defined separation distance intervals (or lags) [Deutsch and Journel, 1992]. The results are plotted on a graph of lag distance on the horizontal axis versus a
covariance measure on the vertical axis. In practice, the analytical tool used to evaluate the
spatial covariance of a variable is usually the variogram, which can be thought of as the
“inverse” of covariance—the variogram values are low (have low variance) between quantity
values at small separations and high (have high variance) at large separations. It is important to
realize that the results of this computation are highly dependent on the parameters (e.g. lag,
search bandwidths) chosen for the computation. Another important point about experimental
variograms is that it is often very difficult to delineate meaningful variogram shapes from them,
as a result of the data not providing enough information about the RVs spatial structure. This is
where the ability to incorporate the, usually, much more voluminous soft data in the
experimental covariogram/variogram calculation can be very advantageous.

Copty and Rubin [1995] illustrate the benefits of the multi-data type experimental indicator
covariance estimation methodology, put forth by Alabert [1987], for a synthetic hydrogeology
problem. They use surface seismic reflection survey data to provide “soft” information about
lithology, based on an Alabert [1987] M-B type of calibration of collocated borehole geophysical
sonic logs and geologist’s logs of lithology. Seismic velocities derived from a hypothetical 2-d
seismic profile survey are then transformed to soft lithology indicators, providing lithology
information coverage across the entire section. They calculate separate hard and soft data
experimental indicator auto-covariances, as well as the hard-soft cross-covariance, from the hard
geologist’s well log indicators and the soft surface seismic indicators. All three covariances are
combined into one estimate of the hard indicator experimental covariance by taking a weighted
average of the unmodified hard data covariance, scaled soft covariance, and scaled hard-soft
cross-covariance; the scaling factors are derived from the soft data calibration and the
weights—applied to each of the three covariances—are an optimized function of (1) the number
of soft versus hard data and (2) the quality of the soft data [Copty and Rubin, 1995]. For their
synthetic test case, the hard data alone provides only one point on the horizontal experimental
covariogram—due to all the data being in a few vertical boreholes—while the combined
covariogram estimate, incorporating spatial information provided by the soft seismic data, defines points across all horizontal and vertical spatial lags.

McKenna and Poeter [1995] illustrate a methodology for incorporating qualitative information in the form of probabilities (e.g. expert judgment) in the primary variable indicator covariance estimate.

Regardless of whether soft data is employed in the computation, when the M-B is being used one experimental covariogram/variogram should be computed for each primary variable indicator threshold level (or class).

### 4.3.1.4.2 Choice of mathematical covariance model

Experimental versions of the covariance provide a starting point for defining potential mathematical models of the covariance. The experimental versions represent discrete representations of what the true covariance may look like; continuous mathematical equations can be fit to the points. The model equations can be fit either manually through visual inspection or automatically using regression analysis, both of which can be accomplished very easily using interactive computer programs. It is generally prudent to use mathematical models that are as simple as possible with as few parameters (unknowns to be fit) as possible. The models should not be expected to cross each point on the experimental covariance, which, it must be remembered, is only a visual approximation of the true covariance. A model is fit to each of the experimental covariograms/variograms representing each indicator level.

### 4.3.1.4.3 Validation of covariance model

Once a preliminary mathematical model of the covariance is chosen it should be validated by evaluating the residuals between collocated predicted values of the variable from the model and observed values. The predicted values are calculated by invoking the kriging methodology chosen for the study with the preliminary covariance model. The variable is estimated separately, and in succession, at each location containing an observation using the chosen kriging technique, conditioned on some subset of the observations that doesn’t contain the observation at the
estimation location; the nature of the observation subset is dependent on the validation procedure. At each observation point the kriging residual is calculated by taking the difference between the observed and estimated values, and usually normalizing the difference in some way.

A rigorous method for performing the very important covariance model validation step, used in the case study of this work, is the method of orthonormal residuals [Kitanidis, 1991]. Without getting into details, the method proceeds as follows:

1. the observations / measurements are ordered from one to the total number of measurements \( n \), emulating the order in which the measurements were taken (usually the actual order does not have much affect the analysis);
2. the value of the second observation is predicted from the first observation using conditional kriging;
3. the value of the third observation is predicted from the first and second observations using conditional kriging;
4. the values of the remaining observations are predicted in succession using the same approach;
5. the orthonormal residuals for each of the \( n-1 \) predicted observation values—defined such that they are uncorrelated to each other and normalized to one—are calculated by subtracting the predicted from the measured value, and dividing the result by the kriging estimation variance at that location;
6. the orthonormal residuals are evaluated against a set of defined criteria (see Kitanidis [1991] for examples of statistically derived criteria) to determine the validity of the chosen covariance model, as well as its comparative goodness to other models tested;
7. the covariance model is adjusted, typically by changing a few of the parameters in the mathematical equation defining the model (e.g. the correlation length and directional anisotropy);
8. steps 1-7 are repeated until the near optimal model—defined as the model that best meets the evaluation criteria—is converged on.
Kitanidis [1991] presents a number of tests for evaluation of model goodness of fit. The two primary ones use the $Q_1$ and $Q_2$ statistics, defined as:

$$Q_1 = \frac{1}{n-1} \sum_{k=1}^{n} \varepsilon_k,$$

$$Q_2 = \frac{1}{n-1} \sum_{k=1}^{n} \varepsilon_k^2,$$

where $n$ is the total number of observations and $\varepsilon_k$ is the orthonormal residual for the $k$th observation. The model should be rejected if these statistics meet the following criteria:

$$|Q_1| \geq \frac{2}{\sqrt{n-1}},$$

$$|Q_2 - 1| \geq \frac{2.8}{\sqrt{n-1}}.$$

The orthonormal residuals validation methodology is very general; it can be applied to any continuous or indicator RV geostatistical estimation process, as long as direct hard measurements of the primary variable exist. While an iterative trial and error approach was used in the case study of this work, the methodology can also be automated by employing an iterative optimization routine—linear or nonlinear, depending on the mathematical nature of the particular validation problem [Kitanidis, 1991].

### 4.3.1.4.4 Derivation of secondary variable covariance model

As noted above, with the M-B indicator RV approach, once the inference of the indicator covariance model for each of the primary variable indicator levels is completed, the secondary variable indicator covariance and primary-secondary variable indicator cross-covariance models are derived simply through a scaling of the primary variable indicator covariance model. The scaling factors are calculated from the secondary variable misclassification probabilities:

$$\{\text{secondary variable auto-covariance}\} = B^2 \cdot \{\text{primary variable auto-covariance}\}$$

$$\{\text{primary-secondary variable cross-covariance}\} = B \cdot \{\text{primary variable auto-covariance}\}$$

where $B = P1 - P2$ (see above Section 4.3.1.2 for the definitions and derivations of the misclassification probabilities $P1$ and $P2$); “indicator” has been dropped as a prefix to
Chapter 4. ACCOUNTING FOR AND MODELING UNCERTAINTY

"covariance" for brevity. The secondary variable auto-covariance has a discontinuity at a lag distance (separation between two measurement locations) of zero, which corresponds to the secondary variable variance. This variance is calculated differently for the Zhu and Journel [1993] and Alabert [1987] M-B approaches due to differences in the "soft" indicator RVs between the two approaches. In the Zhu and Journel [1993] approach, the "soft" IRV variance is equal to

\[ \text{soft IRV variance} = B \cdot \text{primary variable auto-covariance}. \]  \hspace{1cm} (4.7)

In the Alabert [1987] approach it is calculated from the soft IRV mean using the formula for a binary RV variance:

\[ \text{soft IRV variance} = \text{soft IRV mean} \cdot (1 - \text{soft IRV mean}), \]  \hspace{1cm} (4.8)

where the soft IRV mean is calculated from the hard IRV mean:

\[ \text{soft IRV mean} = P1 \cdot \text{hard IRV mean} + P2 \cdot (1 - \text{hard IRV mean}). \]  \hspace{1cm} (4.9)

4.3.2 Geostatistical estimation

Once the required "global" spatial random variable models are inferred, the chosen geostatistical estimation algorithm can be implemented to estimate local values of the random variable at any location in the study area. As discussed earlier, there are two general types of geostatistical estimation:

- kriging, wherein the expected value of the RV is estimated at each location, and
- simulation, wherein equally likely realizations of the spatial random field are generated; a value of the RV is estimated at each location, for each realization.

Both techniques account for the RVs spatial correlation through the inferred covariance model(s), and can be performed with continuous or discrete (indicator) RVs. As discussed earlier in this chapter, the crux of the primary simulation approach—sequential simulation—is kriging. Kriging uses data from the study area to locally condition the expected value of the RV, by accounting for the spatial correlation influence of the data on neighboring locations; estimates at hard data locations always preserve the measurement values at those locations. Simulation can be
performed in both unconditional and conditional modes, without and with conditioning data, respectively. Both modes preserve the spatial correlation structure of the SRF, and the global mean if simple kriging is employed; in addition, hard data values are preserved with conditional simulation.

This section focuses on the mechanics of implementing the Markov-Bayes indicator kriging and simulation methodologies, both the Alabert [1987] and Zhu and Journel [1993] approaches. The general theory behind geostatistical estimation was discussed earlier in this chapter (Section 4.2.3); Appendix B provides a detailed mathematical discussion of the M-B indicator RV geostatistical estimation approach.

In order to incorporate the information provided by observations in the probabilistic representation of an empirical quantity across a spatial area / volume, the information has to be coded in a way that the M-B kriging / simulation methodologies can appropriately handle. The indicator coding is predicated on the practitioner's definition of indicator classes for discrete RVs (e.g. lithofacies types), or indicator threshold cutoffs for continuous RVs (e.g. hydraulic conductivity). As mentioned earlier, hard data (error free direct measurements of the empirical quantity) are simply coded as ones or zeroes, corresponding to 100% and 0% probability, respectively, of being within an indicator class. Soft data are coded in different ways depending on the type of soft data and the M-B approach. Three types of soft data can be handled, encompassing most types of imperfect information:

1. measurements of a secondary variable that is indirectly related to the primary variable (e.g. measuring electrical conductivity and relating it to water chemistry);
2. bounds on the value of the primary variable (e.g. from geological experience knowing that the hydraulic conductivity has to be between 0.1 and 0.5 meters/day); and
3. information in the form of a probability distribution (e.g. from work in a similar geological environment, a geologist estimates there is a 40% chance the lithology is sand, 40% chance it is silt, and 20% chance it is clay).
The classifications numbered 1, 2, and 3 above are given the names Type A, Type B, and Type C soft data, respectively [Alabert, 1987].

As discussed earlier in this chapter in the section on secondary variable calibration (Section 4.3.1.2), Type A soft data is coded as either zeroes and ones (Alabert [1987] approach) or values between zero and one, inclusive (Zhu and Journel [1993] approach); both methods are based on the results of the collocated hard-soft data calibration and have misclassification probabilities associated with them.

Type B soft data is coded the same way hard data is—zeroes and ones, depending on indicator class membership—except that the data is ignored outside the bounds where it is defined. For example, if the soft information about hydraulic conductivity is defined as being somewhere between 0.1 and 0.5 meters/day, then for all indicator thresholds below 0.1 m/day indicator values are coded as zeros; for all indicator thresholds above 0.5 m/day they are coded as ones. Indicator values are left unassigned for indicator thresholds falling between 0.1 and 0.5 m/day, where there is no information about the hydraulic conductivity.

Type C soft data are coded as probability values (which, by definition, are between zero and one, inclusive) for each indicator class—basically representing the soft information as a known discrete probability distribution, with the cutoffs corresponding to the indicator class thresholds.

The inferred “global” RV models and the coded information about the empirical quantity at the study site are used in making the kriging estimate of the quantity at specified locations, through the solution of a system of kriging equations. These linear regression equations are formulated based on two estimation constraints:

1. the estimation variance is minimized, in a least-squared error sense, and
2. the estimation is unbiased, in that the expected value of the estimated values is equal to the expected value of the RV for simple kriging, and the weights in each kriging equation sum to one for ordinary kriging.
The simple end result of applying these statistical constraints is an estimator that is a linear combination of properly coded neighboring data, plus an unbiasedness constraint; the weights associated with each datum are calculated from a system of normal equations—a matrix equation—relating the covariances between the estimation and data locations (see Appendix B for more detail). In the indicator RV case this process corresponds to the updating of the local pdf/cdf for a quantity of interest, based on information from neighboring local pdf/cdfs [Deutsch and Journel, 1992].

Since the M-B approach uses a cokriging methodology to incorporate multiple data types, auto-covariances for each data type, and cross-covariances between all different data types, are required in the normal equations. Fortunately, all the covariances are derived from the single primary variable auto-covariance model using primary-secondary variable calibration parameters (see above sections).

While primary variable measurements and type A and B soft data are incorporated in the weighted linear combination part of the kriging estimator, type C data are treated as a local expected value that is conditioned on neighboring data [Alabert, 1987]. It should be realized that using such type C data invalidates any assumption of first order stationarity.

Essentially the same sequential simulation approach as used for continuous Gaussian SRVs can be applied for M-B indicator simulation of indicator SRVs. (See Section 4.2.3.2 for a short discussion of sequential simulation.) Instead of determining a single empirical quantity value from a random number applied to a continuous Gaussian ccdf distribution—generated by conditional kriging of the prior cdf from neighboring measurements and previously simulated nodes—the random number is applied to a discrete ccdf—generated by M-B conditional kriging—to determine an indicator class membership. If realizations of the actual empirical quantity value distribution is desired, instead of indicator classes, a continuous ccdf can be constituted from the discrete ccdf (making assumptions about the shape of the probability distribution) and a single empirical quantity value determined from the random number. For the case study of this work (see CHAPTER 5), realizations of whether soil contaminant
concentrations are above threshold levels are required—not realizations of actual concentration values.

The above description of the M-B indicator kriging/simulation approaches represent estimation of the conditional cdf (ccdf) of an IRV at defined point locations. In what is commonly referred to as block kriging, kriging can also be used to directly estimate the ccdf of the proportional number of point locations in a defined volume, or block, that have values within an indicator threshold class. It is important to realize that this ccdf is different than the ccdf of the entire block volume property value being within an indicator class; as discussed earlier, that ccdf can be computed by (1) simulating multiple realizations of the continuous RV at a much finer scale, (2) upscaling the resulting values in each realization to the larger volume of interest, and (3) computing cdf histograms, with discretizations chosen to match the indicator thresholds, at each volume location using all the upscaled realizations [Deutsch and Journel, 1992]. In comparison to the estimated ccdf values from point kriging, block kriging ccdf estimates have less spatial variance.

One problem with standard kriging and simulation is that it does not account for different sampling volumes of the measurements used for estimation. Massman and Freeze [1989] document a methodology to account for different sampling volumes in the estimation by modifying the covariance matrices used in the kriging calculation, based on a defined measurement spatial filter matrix; the approach is applied to kriging/simulation of a continuous RV. An attempt has been made to apply the same approach to the M-B indicator kriging algorithm; a spatial filter is associated with each measurement point and incorporated in the covariance computation for that measurement in the M-B indicator kriging equations. A spatial filter can be associated with a hard or soft Type A measurement.

Figures 4.2.3 and 4.2.4 are 2-d examples of the effect of including a hypothetical measurement support volume in the M-B indicator kriging for the concentration of a chemical in soil—at a lower and higher concentration threshold level, respectively. The kriging estimation map is shown for the cases when the concentration data are assumed to be obtained from a point
measurement (Figures 4.2.3a and 4.2.4a) and a measurement with a circular support volume with a radius of 5 units (Figures 4.2.3b and 4.2.4b); the spatial filter in the latter case is equally weighted across the 5 unit radius. The effect of the spatial filter is much more obvious at the upper concentration threshold (Figure 4.2.4), where the correlation length of the indicator RV is much less (5.5 units compared to 16 units for the lower threshold). When the measurements have the circular support volume associated with them they have a larger spatial influence on nearby kriging estimates, but the change in estimated probabilities of concentrations being above the upper threshold is smaller at locations near the center of the measurement; at the measurement location where a concentration above the threshold was observed the probability drops to less than 0.8, compared to 1.0 for the point measurement. This result is expected because the measurement sampled an “average” concentration above the threshold, but the individual soil parcels—at the smaller estimation scale—that contributed to the high concentration are unknown. The results for the lower concentration threshold exhibit the same features (Figure 4.2.3)—the estimated probabilities of concentrations above the threshold level are no longer 0.0 or 1.0 at the measurement locations when the spatial filtering is applied, but the area of influence is greater.

While this measurement spatial filter implementation in the M-B indicator kriging methodology appears to work effectively, it will not be addressed in any detail because it is not required for this study. However, this extension is considered a potentially very useful extension to the M-B kriging approach. It expands upon the existing capability of rigorously accounting for the uncertainty associated with indirect measurements at the same scale of the estimation scale to also rigorously handle the uncertainty associated with the measurement support scale being larger than the estimation scale; a measurement support scale smaller than the estimation scale can be addressed using block kriging—also built into the M-B methodology as an option in this work.
Figure 4.2.3: Effect of spatial filtering on indicator kriging of hypothetical hard soil chemical concentration measurements. Probability of concentrations above the lower concentration threshold level, assuming the measurements are (a) point measurements and (b) measurements having a circular areal support volume with a radius of 5 units.
Figure 4.2.4: Effect of spatial filtering on indicator kriging of hypothetical hard soil chemical concentration measurements. Probability of concentrations above the upper concentration threshold level, assuming the measurements are (a) point measurements and (b) measurements having a circular areal support volume with a radius of 5 units.
4.3.3 Limitations

A problem has been found with the Zhu and Journel [1993] approach to conditional indicator kriging with soft data, when the mean of the hard measurements used for calibration of soft measurements is different than the overall global mean used for geostatistical estimation. Even when the quality of the soft data is very poor (when there is little or no correlation between the secondary and primary measurement), the soft data can have a significant, albeit obviously erroneous, influence on the estimation. This bias is due to the soft data prior cdf (denoted as “$Y$” in section 4.3.1.2 on M-B soft data calibration) being significantly different from the global mean of the IRV; this results in a significant adjustment to the prior expected value cdf at the estimation location in the Zhu and Journel [1993] formulation of the simple kriging equation. The reason that the $Y$ cdf is significantly different than the global cdf can be traced to the way the $Y$ cdf values are calculated in the secondary variable calibration; specifically, the indicator threshold expected values of the primary variable measurements used as calibration pairs with collocated secondary measurements is different than the global indicator means, causing a bias in the $Y$ values calculated from these data pairs. A methodology has been developed in this work to alleviate this bias problem—normalizing the distribution of the primary variable measurement values within each indicator class to mimic the distribution of the overall global population of the IRV, prior to calculating the $Y$ cdf values for each indicator class.

The Alabert [1987] approach does not appear to have this “mean” biasing problem when conditioning on soft data. This is most likely due to the different way that this approach accounts for soft Type A data quality in the kriging equations. Type A soft indicator data can only take on values of 1 or 0; all of the soft data uncertainty is accounted for in the $B$ calibration factor, used to derive the soft indicator auto-covariance and soft-hard indicator cross-covariance. Unlike in the Zhu and Journel approach, in the Alabert [1987] approach the secondary-primary variable calibration—used to estimate the $B$ calibration factor—is unaffected by the indicator class distribution of the primary variable calibration data. (See above Sections 4.3.1.2 and 4.3.1.4.4, and the mathematical discussion in Appendix B, for an explanation of the $B$ calibration factor.)
Even when the calibration normalization procedure developed in this work is implemented in the Zhu and Journel [1993] methodology, the kriging estimates generated from the M-B indicator kriging of soft Type A data still appear to become biased to the soft data—instead of preserving the global prior mean—when a large number of conditioning soft data are used. This occurs even if the soft data is of very poor quality—represented by a very low $B$ factor. From intuition, the kriging estimate should change very little when soft data from far outside the data’s spatial correlation length is included the estimation; in practice, in this study the estimate has been shown to change quite substantially as more and more data outside this range are included. The average of the estimated conditional point cdf values become progressively biased from the input global prior mean as more and more soft data are present and they are permitted as soft conditioning data in the kriging estimation, even if they are outside the correlation length of the data. The exact cause of, and remedy for, this bias problem has not been identified by the author, but it is postulated that it could be associated with the bias problem in the $Y$ cdf calibration methodology—discussed above. While this bias in the estimation mean is generally small, it can be very important in problems where engineering risk costs are sensitive to probabilities of occurrence; this condition is true for the case study in this work, where the probabilistic cost associated with contaminated soil excavation is very high.

### 4.4 Summary

Probability distribution functions (or cumulative distribution functions) are the ideal and most general representation of the uncertainty associated with an empirical quantity. However, in order to employ them rigorously and effectively within a risk-based decision-making framework, they must be (1) defined for the quantity of concern based on all available information, (2) represented in a way that can be incorporated in the mathematical decision framework, and (3) updated as additional information is obtained. For hydrogeology-related problems, the material properties of concern are usually spatially-varying and, thus, a pdf is required at every location across the study area.
Chapter 4. ACCOUNTING FOR AND MODELING UNCERTAINTY

Geostatistics provides an effective way for representing and manipulating spatially-varying variables, such as those common to hydrogeology-related problems (e.g. hydraulic conductivity), that are naturally occurring and distributed across a large area/volume—and, thus, are inherently uncertain. Such variables are mathematically represented as spatial random variables (SRV), containing deterministic and stochastic components, that rigorously account for uncertainty and take advantage of the information provided by the spatial correlation such variables have. While there are many geostatistical approaches, only in a few circumstances can uncertainty be represented and handled in the “full” form of a pdf. The only case in which the full form of the pdf can be represented by a continuous SRV and accounted for after geostatistical estimation is if the pdf is assumed to have a Gaussian normal functional form; in this case the pdf can be represented by and reconstructed from a few parameters—mean and variance. However, representing spatial properties as binary indicator spatial random variables (ISRV)—taking on values of 1 or 0, depending on if the property is within or not within a defined interval of values—enables a discrete representation of the variable’s pdf, with no restrictions on the functional form; the expected value of an ISRV is the probability of the variable falling within the interval of values defined by the ISRV. How well the continuous pdf is represented by the discretization depends on how the ISRV cutoffs are defined for the continuous variable. As with continuous SRVs, ISRVs are locally updated through geostatistical estimation, which updates the discrete pdf/cdf to a conditional pdf/cdf.

Another advantage of using ISRVs instead of SRVs is that the indicator kriging estimation directly results in a probability—the conditional probability of the variable being within the interval defined for the ISRV. The most consistent, robust, and universal probability updating method is Bayesian updating—employing Bayes’ conditional probability equation. Indicator kriging is an approximation of Bayesian updating [Solow, 1986].

The pdf for an SRV at a particular location is determined in two steps when employing a geostatistical approach:
1. The global prior (a) stationary pdf/cdf, (b) spatial correlation, and (c) correlation between different variables (the latter only if some of the data corresponds to measurement of a secondary variable) are inferred based on the integration of all available information. These inferred statistical models are generally spatially invariant and do not account for the local information next to each particular location.

2. The local conditional pdf/cdf for each individual SRV at particular locations is estimated using geostatistical kriging/cokriging or simulation of multiple realizations, by updating the global pdf based on all nearby information that affects the SRV at that location.

Geostatistics enables the incorporation of direct observations of the variable of interest in the estimation of the SRV pdf, as well as indirect observations of a secondary variable that is somehow related to the variable of interest. However, to incorporate indirect observations in the SRV estimation requires the inference of many covariance models—a computationally intensive and difficult process requiring a large and representative data set of primary and secondary variable measurements.

The Markov-Bayes ISRV approach [Alabert, 1987; Zhu and Journel, 1993] is a very flexible and practical indicator geostatistics approach that has the advantages of ISRV representation of variables—no restrictions on the functional form of the pdf, direct conditional probability estimation—as well as the ability to efficiently incorporate observations of secondary variables in the random variable representation of the variable of interest. The difficulty in inferring all the covariance models when incorporating secondary variable observations is overcome by deriving all secondary variable covariances and secondary-primary variable cross-covariances from the primary variable covariance, based on a simple calibration between collocated primary and secondary variable observations. In addition, subjective information about the variable of interest in the form of prior probabilities, such as expert judgment, can be incorporated in the cokriging estimator for updating ISRVs.

Thus, the Markov-Bayes indicator geostatistics methodology provides a very good match to the spatial uncertainty model needs of this work—the ability to incorporate geophysics, an
inherently indirect measurement technique, in the probabilistic representation of the variable of interest and, for the case study, the ability to directly estimate probabilities of threshold exceedance for input in the soil contamination decision model.
CHAPTER 5
CASE STUDY

This chapter discusses the real-world case study undertaken for this work—a risk-based decision analysis of the site investigation for an active environmental remediation site, focusing on how it can be optimized, particularly through the use of geophysics, to make the entire remediation more cost-effective. The site is a real estate development on a recently deactivated railyard where soil contamination has been found. Since development of the site has to meet environmental compliance for its intended future use, either (1) contaminated portions of the site have to be remediated or (2) the development has to be changed so that the future activities in those areas (or lack thereof) are acceptable for the type and level of contamination there. Clearly these decisions are largely dependent on what is known about the site, particularly what contamination exists, where it exists, and what will happen to it in the future. To address this decision-making process in the case study, the site is analyzed from three perspectives using the risk-based decision-making approach developed in this work:

- **Present day scenario** (also commonly referred to as “prior” analysis) - Using the existing soil sample data set, a geostatistical model of soil contamination is developed using the modified Zhu and Journel [1993] Markov-Bayes approach discussed throughout this work. Using the geostatistical model and site specific decision analysis model, an optimal soil excavation and treatment program is designed, based on a risk-cost-benefit decision making methodology.

- **Hypothetical future scenario** (also commonly referred to as “pre-posterior” analysis) - The worth of future, additional site characterization data—more soil samples or surface geophysics data of different qualities—is investigated by applying the decision-making model with M-B geostatistical simulation [Zhu and Journel, 1993] of hypothetical characterization programs. This analysis is performed with respect to the information provided by the present day volume of data.

- **Future scenario** (also commonly referred to as “posterior” analysis) - Actual surface geophysical measurements (electromagnetics and ground penetrating radar) acquired as part
of this study, are incorporated into the geostatistical model and the remediation design is re-optimized.

When the case study was initiated, the actual site characterization program was already at an advanced stage and soil remediation activities had started in some areas. A total of some 350 soil samples had been extracted; 266 were chemically analyzed for hydrocarbon contamination and 33 for heavy metals. The results from these analyses were used by the developer at the site to build a conceptual model of where different types and levels of soil contamination were known or interpreted to exist. This model was to be used for planning where soil excavation should be performed and how the excavated soil should be treated. It was also planned that routine soil batch samples would be taken and chemically analyzed during excavation to monitor the existence and extent of contamination. If necessary, the excavation and remediation could be adjusted based on the new information.

The case study consisted of two primary types of activities: (1) taking geophysical measurements on the site and analyzing the results and (2) developing and running geostatistical and decision analysis models for soil remediation of the site, based on the existing soil sample data, as well as trying to incorporate the geophysical measurements.

Ground penetrating radar (GPR) and electromagnetic (EM) surveys were performed in an accessible area of the site. These surveys had to be performed in an asphalt parking lot surrounded by steel chain link fences and powerlines. The site subsurface possibly contained some pipelines and many possible forms of junk (e.g. plastic and steel) within the heterogeneous fill material used to create the synthetic land in the harbor. The subsurface is also heavily influenced by sea tides—especially near the shoreline—which cause the water table to rise and fall, flushing sea water in and out of the soil pore space as it does. These conditions pose difficulties for both these types of geophysics, but are regarded as making the study more realistic.
Chapter 5. CASE STUDY

Normally a three step process for applying geophysics to hydrogeological problems is recommended in this work:

1. First perform a data worth analysis to determine whether, in general, potential geophysical techniques can provide added value to the problem and what “quality” level is required to do so,

2. If step one indicates geophysics may have value, run a small reconnaissance survey over a region of the site where the measurements can be ground-truthed to determine the “quality” level of the chosen geophysical techniques, and

3. Perform a full-scale production survey(s) with the chosen geophysical technique(s) across as much of the site as possible.

However, due to time constraints of this work, this process could not be followed for the case study; only a very small reconnaissance survey and analysis could be performed and the final GPR and EM production surveys were taken only over a small portion of the site. The resulting measurements are processed; rectified to the soil sample locations on site; presented as 2-D contour maps (EM) and 3-D cut-away pixel grids (GPR) to enhance interpretation; and calibrated to soil analysis results, using the modified M-B calibration approach discussed in CHAPTER 5, for incorporation in the future scenario analysis.

For the modeling component of the case study an indicator geostatistical model of soil contamination across the site is developed using the existing soil sample chemical analysis database. Indicator thresholds are set at the different regulatory chemical concentration action levels for the contaminants of concern. Development of the geostatistical model entails inferring and validating overall indicator means and variograms for each threshold level of each contaminant, using the approach discussed in Chapter 5. Many of soil samples were not analyzed for all the contaminants. In order to glean as much information from the data as possible, the Markov-Bayes co-kriging approach [Zhu and Journel, 1993] is employed to take advantage of correlations between the presence of different contaminants (e.g. hydrocarbon concentration measurements used to estimate heavy metals indicator threshold values).
A risk-based soil remediation decision model is developed for the site. This model estimates optimal—meaning most cost-effective—remediation engineering designs and their respective component and total costs based on the contamination spatial uncertainty models; soil excavation, batch sampling, and treatment are accounted for. Unit costs for the above activities have to be assumed, but are based on cost estimates given by the developer. The model takes into account the following risks: (1) not remediating contamination during the remediation phase of development, thus having to deal with it during the actual building phase or after, (2) unnecessarily remediating soil for contamination that does not actually exist. The solution is calculated using a heuristic optimization approach. The model is designed to perform decision analyses for all three scenarios described above.

Three dimensional grids of the probability of contamination above the regulatory action level and equally-probable contaminant spatial distribution realizations (depictions of possible reality) are generated for each contaminant (and action level, for the former); this is accomplished by invoking the modified M-B conditional cokriging and simulation approaches (discussed in CHAPTER 4), respectively, with all the sample data and inferred spatial structure models as inputs. Figure 5.1 is an example of one of the estimated probability grids—the probability of a particular type of hydrocarbons having concentrations above a defined regulatory threshold. The probability grids are used in the decision model to estimate the optimal remediation design (see Figure 5.2) and its expected cost based on all existing information. In addition, the expected value of perfect information (EVPI) is calculated from the simulated realizations. The analyses are performed for a range of cost values (referred to as future remediation costs) associated with not remediating contaminated soil during the remediation phase of the project. Figure 5.2 is an example of an optimal remediation design for a particular set of assumed failure costs. The optimized total remediation cost ranges from 4 – 9 million Canadian dollars. These analyses comprise the present day scenario.
Figure 5.1: Example grid of probabilities of contaminant concentration greater than a regulatory threshold level.

Figure 5.2: Example optimized soil remediation design.
The hypothetical future scenario of acquiring additional site characterization data—from geophysical measurements of different data qualities or soil sampling and chemical analysis—is simulated so as to estimate the worth of the different site investigation scenarios. This is accomplished with a combination of conditional geostatistical M-B simulation and cokriging, wherein: (1) multiple, equally-probable realizations of the spatial extent of contamination across the site are generated with simulation using the same information employed in the present day scenario simulations, (2) the results of the hypothetical characterization programs are determined for each realization, and (3) the hypothetical data is used as conditioning data, in addition to existing real data, to generate, with cokriging, sets of contamination probability grids for each hypothetical data realization. Using the decision model, the optimal remediation design and its corresponding cost of remediation is estimated for each contamination probability map. Since, theoretically, the results of each of these decision model analyses is equally-probable, the expected value of the remediation cost if the proposed characterization programs are implemented can be calculated by taking a simple average of the results. The expected data worth of the different programs is derived by subtracting the present day (prior) optimal remediation cost from the above-mentioned expected value remediation costs. Then the comparative worth of the different characterization programs can be assessed. The results of the data worth analysis indicate that analyzed soil samples from 20 additional boreholes spread across the site provide little data worth; geophysics only provides a significant cost savings if the data quality for delineating contamination is very good and the measurement coverage is site-wide.

The final modeling component of the case study is to incorporate the results of the production EM and GPR geophysical surveys in the geostatistical model of the site, which is then used in the decision analysis model to update the optimal remediation design and its associated costs. Incorporation of the geophysical measurements in the geostatistical model requires identification and quantification of correlations between categorized levels of certain attributes of the geophysical measurements and threshold levels of hydrocarbon and/or metals
soil contamination. This is accomplished through a calibration exercise, where those geophysical results that are sensitive to the subsurface at locations where soil samples were taken are compared with those same soil sample results—using the procedure discussed in Chapter 5. After processing the geophysics data, it is successfully calibrated with soil sample data, indicating several reasonable correlations with certain types of soil contamination. The processed geophysics data, along with the calibration results, are successfully incorporated in the M-B indicator cokriging algorithm [Zhu and Journel, 1993] to generate updated contaminant level probabilities, subsequently used in the decision model to update the optimal remediation design.

The change in the total remediation cost from incorporation of the new geophysics information is negligible, despite a considerable effect on the probabilities—albeit over a small proportion of the site.

5.1 Introduction

The real world example chosen to illustrate the risk-based decision making framework investigated in this study is the incorporation of existing site investigation data, and evaluation of future site investigation programs, for contaminated soil remediation design at an active engineering site. The objective of the case study is twofold: (1) to apply the framework espoused in this research to a real world problem, and (2) to follow through the process, from start to finish, of applying geophysics to a hydrogeological engineering problem. A real world case study, especially one that is not set up or chosen to provide ideal results, provides the opportunity to test, recognize limitations and pitfalls in, and improve upon the methodology being proposed. Inevitably, a real world case study, especially one involving the acquisition and analysis of actual field measurements, will encounter and expose more difficulties than a synthetic case study. This is especially applicable to this study since the results and success of geophysical measurements, in general, are very site specific; one of the main research objectives of this work is to develop a flexible, applicable framework for evaluating and utilizing geophysics in hydrogeology problems. Thus, the case study was chosen to represent a typical hydrogeology problem at a site.
where geophysical measurements could be acquired, but the problem and site are not necessarily conducive to the application of geophysics.

In order to assure an effective case study a set of criteria was developed. The following is a list of the requirements for the case study:

1. A decision has to be made concerning the design of a hydrogeological engineering system (e.g. which alternative to implement for a contaminated site—no action, active hydraulic containment, or barrier walls).
2. The decision is sensitive to hydrogeological parameter(s) characterizing the site (e.g. contaminant concentration, aquitard continuity, or hydraulic conductivity).
3. The critical parameter(s) represent a target which can potentially be detected by geophysics (i.e. geophysical measurements are sensitive to the type of contrasts in the parameter(s) expected).
4. The equipment for appropriate geophysical methods are available and can be run or, next best, a valid geophysics data set already exists.

The case study chosen is a major commercial and residential development on an industrial site, located along the shoreline of a tidal marine inlet. The site is predominantly underlain by fill and contains subsurface contamination which, if exceeding regulatory limits, must be remediated in order to meet environmental compliance for the land’s intended use. The specific section of the site used for the case study contains hydrocarbon and heavy metals soil contamination which must be remediated if above regulatory thresholds. The chosen method of remediation is excavation and treatment of contaminated soils. The design of the excavation depends on the spatial distribution of the contaminants. Thus, the site meets the first two criteria for an effective case study since a decision has to be made about the excavation design and this decision is sensitive to the spatial distribution of contaminant concentrations. Numerous soil and groundwater samples from boreholes have been taken and chemically analyzed in a lab for
contaminant concentrations, but the subsurface spatial distribution of contaminants is still uncertain due to the limited volume of investigation of soil samples.

Geophysics provides a possible alternative for better characterizing the contaminant distribution and, thus, improving the excavation design and overall remediation. However, the case study site conditions and problem requirements make the application of geophysics very difficult. The near subsurface is characterized by heterogeneous fill, possible pipelines, and a tidally-influenced water table, while the surface is punctuated by numerous steel fences, parking lots, rail tracks, powerlines, and buildings. Also, the use of geophysics, in general, for delineating hydrocarbon contamination is not a well established application. However, as mentioned above, these difficulties can actually be beneficial to the research objectives of this study. That being said, soil samples from the site have contained high saturations of pure phase hydrocarbons, which ground penetrating radar (GPR) and frequency-domain electromagnetics (FDEM) can potentially detect. In addition, FDEM can potentially detect high levels of metals contamination. Thus, the site meets the last two criteria for a case study listed above: (1) there is the potential that GPR and FDEM are be sensitive to hydrocarbon and metal contamination—critical parameters affecting remediation design and (2) GPR and FDEM equipment were available for use.

To evaluate the geophysics potential at the site a geophysics data worth analysis is performed and geophysical surveys using GPR and FDEM are taken in an attempt to delineate subsurface hydrocarbon and metals contamination. Before starting the geophysics data worth analysis a prior risk-cost-benefit analysis using existing borehole soil sample data is performed to estimate the optimal prior remediation design. The maximum characterization budget is also determined at this time by executing an expected value of perfect information (EVPI) analysis. Data from the geophysical surveys are included in the remediation design process through a risk-cost-benefit posterior analysis.

It should be noted that the framework and guidelines developed through this research were not necessarily followed in the case study (e.g. the field work was performed before any risk-
cost-benefit analyses) due to limited time and resources, but also due to the fact that the lessons learned from the case study were instrumental in actually developing the recommended guidelines.

5.2 Background

5.2.1 Setting

The case study site is a recently deactivated railyard and commercial ferry dock, which included a fuel storage and marine fueling station in earlier times, and is now the site of a major real estate development. All of the fuel storage tanks have since been removed (see Figure 5.2.1) and the rail tracks and ferry dock were being removed at the time of the geophysics field campaign. Most of the tank farm area (approximately 250 by 100 meters) is now paved with parking lots (many of which are 24 hour parking lots) and surrounded by chain link steel fences. There are also some buildings and overhead powerlines in the tank farm area. The site is located along the shoreline of a tidal marine inlet and is largely built on fill (up to 14 meters thick) overlying natural sediments. The topography of the site is generally flat with elevations below four meters above mean sea level (MSL). As a result of the proximity to the sea and the absence of topography there is a strong tidal influence on the groundwater. The water table depth fluctuates significantly and the groundwater flow actually changes direction in some places on the site as the tide changes.
5.2.2 History of site

The case study site has a history of more than a century of industrial activity. In the late 1800s a railyard was built along the harbor waterfront. In the early 1900s the shoreline was extended further out into the harbor by filling in intertidal and subtidal areas. At this time a petroleum products bulk plant (storage area) and marine fueling station were built on the fill. This facility grew to include six above ground fuel tanks (the largest of which could hold close to two million gallons of fuel), a pumphouse, a wharf for fueling ships, fuel unloaders to supply the tank farm from rail cars, and a network of underground and above ground pipelines. The fuel tanks were predominantly filled with bunker fuel, but some tanks were also filled with diesel fuel and distillate gas. Also during this time the rail yard was extended over filled land and freight transfer docks and sheds were built. In the 1970s the fueling facility was demolished and fuel tanks removed.

At present the railyard is being decommissioned and the whole site is being built into a commercial and residential development. To proceed with the development the property owner has to insure that the site meets government environmental standards. To do this the property owner has hired an environmental consultant to manage environmental aspects of the
development. At the time of this study the environmental consultant had developed soil management plans for different parts of the overall site based on soil and groundwater sampling data; construction and any associated required remediation had been initiated on certain parts of the overall site. No construction or excavation had started in the tank farm area and adjacent railyard (Phases 2 and 3 of the project) prior to the geophysics field campaign and, thus, this section of the site was the focus of the case study. Later references to the "overall site" correspond to this area of the site (or phases of the project), while references to the "tank farm" area correspond to the area where the former fuel storage and supply facilities existed.

5.3 Site Characterization

Site characterization data at the case study site consists of two major components: "prior" data collected by the environmental consultant for the site and "posterior" geophysical data collected afterwards as part of this research. This sequence of data collection is not necessarily the most effective order of investigation, but is a consequence of the advanced stage of the engineering project prior to being chosen for this study. As mentioned earlier in Section 3.3, geophysics often can be used most effectively in the early stages of a site investigation to help focus and confine the collection of more expensive, smaller-scale "hard" data (e.g. borehole core samples). The site characterization program completed by the environmental consultant included research into historical land use at the site and the emplacement of boreholes and test pits for taking physical samples or making direct in-situ measurements of soil and groundwater. Results from the historical land use study were discussed above and will not be discussed in this section except to say that the contribution of this information is to identify what the likely sources of contamination were (and possibly still are), what types of contamination to expect (e.g. contaminated soil and/or groundwater), and where the most likely areas of contamination are. The soil and groundwater sampling program and geophysical field campaign are discussed below, with emphasis on objectives, strategies, and procedures. Also, some of the measurement
results are summarized and interpreted in a descriptive sense, but the main discussion of results is covered in the analysis section.

5.3.1 Soil and groundwater sampling and analyses

Soil and groundwater sampling from borings, test pits, and completed boreholes is the primary source of information for site characterization used by the environmental consultant. The environmental consultant's objectives for the sampling programs were to characterize the subsurface stratigraphy, soil conditions, and hydrogeology of the site and, thus, aid in the design and regulatory approval of the soil management plan for remediation of the property. The design of the sampling programs (i.e. where and what to test for) was based on the results of the historical land use study and earlier measurements. Borehole logging of cores and cuttings were used to characterize the stratigraphy and observable soil conditions (e.g. grain size, color, odor). Laboratory chemical analysis of representative soil and groundwater samples were used to characterize soil and groundwater contaminant concentrations. Potentiometric head measurements and slug tests in observation wells were used to estimate the groundwater flow regime. Oil-interface probes were used in observation wells to measure the existence and thickness of light non-aqueous phase liquids (LNAPLs) floating on the water table.

5.3.1.1 Borehole logging

A total of some 140 borings and test pits were emplaced by the environmental consultant at sites scattered across the overall site, but mostly concentrated in the tank farm area. Interpreted cross sections of stratigraphy, based on geologic logging of cores and cuttings at various locations across the overall site, were constructed. The cross sections reveal that there are three distinct stratigraphic layers: surficial fill, intermediate fill, and native material. Most of the coal cinder encountered was along rail right-of-ways. The composition and maximum depth ranges of these layers—reported as meters relative to mean sea level (MSL)—is summarized in Table 5.3.1:
<table>
<thead>
<tr>
<th>Layer type (top to bottom)</th>
<th>Composition</th>
<th>Maximum depth range (meters relative to MSL)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Surficial fill</td>
<td>Heterogeneous granular material, predominantly gravel-sized with some fines.</td>
<td>2 to 3.5</td>
</tr>
<tr>
<td></td>
<td>Coal cinder in some areas, predominantly in railyard areas.</td>
<td>Thickness &lt; 0.5, average 0.25</td>
</tr>
<tr>
<td>Intermediate fill</td>
<td>Mostly sand and silty sand with localized areas of cobbles, concrete refuse, wood fragments, and other manmade junk.</td>
<td>-6.0 to 2.0</td>
</tr>
<tr>
<td>Native material</td>
<td>Glacial till or marine sand.</td>
<td>Top found between -6.0 to -3.0</td>
</tr>
</tbody>
</table>

Table 5.3.1: Case study site general stratigraphy.

5.3.1.2 Soil sampling for chemical analysis

The borings and test pits were used to extract a total of some 870 soil samples across the overall site and 350 soil samples within the immediate tank farm area for laboratory chemical analysis. Soil sample sizes ranged from 0.1 to 2.5 meters in length with a mean of about 0.5 meters, and diameters close to in the range of 0.05-0.25 meters. The samples were analyzed for:

- polycyclic aromatic hydrocarbons (PAHs),
- monocyclic aromatic hydrocarbons (MAHs),
- mineral oil and grease (MOG),
- total petroleum hydrocarbons (TPHs),
- light aromatic hydrocarbons (LAHs), and/or
• metals.

Each of these chemical analyses consisted of one concentration measurement in milligrams per kilogram (ppm) except PAH (for which 14 separate chemical species were measured), MAH (5 chemical species were measured), and metals (14 metal elements were measured). The results are classified according to required action levels based on a regulatory framework consisting of four contaminant concentration criteria, as shown in Table 5.3.2.

The criteria for site remediation are based on the contaminant levels of PAH, MAH, metals, and MOG. The remaining types of chemical analyses, TPH and LAH, were used as gross investigative parameters.

Results from the chemical analyses of soil samples suggest that hydrocarbon contamination predominantly occurs in the tank farm area. The tank farm area was the only place where samples contained special waste hydrocarbons (12 samples had greater than 3% MOG, which is defined as special waste and is the only classification of MOG that has to be remediated) and hydrocarbons above Level C (10 samples with PAH). MAH above Level B was only found in this area as well, although only in one sample and only benzene. In fact, nowhere else on the overall site contained soil samples with hydrocarbon contamination at levels which had to be remediated.
### Table 5.3.2: Regulatory action levels for soil contamination

<table>
<thead>
<tr>
<th>Action level</th>
<th>Criteria and required actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Approximate achievable analytical detection limits for organics and natural background concentrations for inorganics and metals. Soils with concentrations above this level and less than Level B are considered slightly contaminated, but not requiring remediation.</td>
</tr>
<tr>
<td>B</td>
<td>Remediation criterion for residential, recreational, or agricultural land use. Soils with concentrations above this level must be remediated to concentrations at or below this level for this type of land use.</td>
</tr>
<tr>
<td>C</td>
<td>Remediation criterion for exclusive commercial or industrial land use. Soils with concentrations above this level are considered significantly contaminated and must be remediated to concentrations at or below this level for this type of land use.</td>
</tr>
<tr>
<td>Special waste</td>
<td>A separate classification for certain contaminants when their concentrations exceed certain thresholds. Soils which fall into this classification must be managed, transported, and stored in a special way according to stringent regulations.</td>
</tr>
</tbody>
</table>

From initial observations, the sample results suggest that PAH contamination in the tank farm area is quite extensive—32 of 60 test locations across this area contain PAH contamination exceeding Level B, at a maximum depth of 9 meters below the ground surface (or approximately 6 meters below MSL). One would expect from scientific intuition that most pure phase (undissolved) hydrocarbon contamination is above the water table or in the zone of water table fluctuation (the subsurface depth range over which the water table varies due to tidal influence). The types of hydrocarbons documented as potential source material from the historical study, and those which have been found in the subsurface from sampling thus far, have densities less
than that of water in their non-aqueous phase (i.e. they are LNAPLs) and would, thus, be expected to float on top of the water table. Possible hypotheses for LNAPL contamination occurring below the maximum water table depth are that at the time the hydrocarbon source was active LNAPL loading caused the water table to depress even further due to buoyancy forces or that the chemical analyses were actually measuring aqueous phase (dissolved) hydrocarbon contamination. The later could only be true if the hydrocarbon concentrations were below the solubility limit for that hydrocarbon type.

Metals contamination above required action levels was found predominantly along existing or former rail tracks. Most of the soil samples which contained metals contamination in these areas consisted of either coal cinder or rock ballast. The cinder identified thus far is predominantly located in the near subsurface, above the water table, and in the vicinity of rail tracks. It appears that the cinder was used as rail track bedding or at least dumped near the tracks. Three boreholes, all in one localized cluster, contained samples with metals contamination above Level B in the tank farm area.

5.3.1.3 Groundwater sampling for chemical analysis

Groundwater samples were extracted for laboratory chemical analysis from 16 observation wells across the overall site, and eight wells within the immediate tank farm area, between the years of 1990 and 1992. The samples were analyzed for chemical concentrations of MOG, TPH, MAH, PAH, and dissolved metals. Similar to the soil analyses, the results are classified into action levels based on government regulatory criteria as shown in Table 5.3.3.
Chapter 5. CASE STUDY

<table>
<thead>
<tr>
<th>Action level</th>
<th>Criteria and required actions</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>Natural background concentrations or achievable analytical detection limits for organic, inorganic, and metallic compounds. For groundwater with concentrations above this level and less than Level Bds a detailed investigation is required, but not remediation.</td>
</tr>
<tr>
<td>Bdw</td>
<td>Remediation criterion for drinking water. Groundwater with concentrations exceeding this level must be remediated to concentrations at or below this level if the water is intended for human consumption.</td>
</tr>
<tr>
<td>Bds</td>
<td>De minimus criterion for water-based discharges to protect aquatic life. Groundwater with concentrations exceeding this level requires further investigation to assess the impact of the contamination on aquatic life and determine appropriate remediation, if necessary. Groundwater with concentrations at or below this criterion does not require remediation if the receiving water is not used for human consumption.</td>
</tr>
</tbody>
</table>

Table 5.3.3: Regulatory action levels for groundwater contamination

Since the groundwater from the site will not be used for human consumption, the only criterion of concern is Level Bds. For this site there are at least two sets of standards or limits that apply to this criterion for some chemicals, depending on where the water is being discharged.

The results from most of the groundwater chemical analyses were below at least the less stringent standard for Level Bds, including all that were tested for MAHs and dissolved metals (with the exception of one which contained barium). Some of the samples contained MOH, PAHs, and/or MAHs above one or both of the standards used for Level Bds; most, but not all, of these samples are from the tank farm area. Most of the samples which exceeded the criterion are from sites located near, but not necessarily overlapping, sites where soil samples showed...
hydrocarbon contamination. The only MOH samples which exceeded the less stringent Level Bds standard were those taken early in the sampling campaign when sediment was not removed, suggesting that the elevated concentrations could be attributed to contamination adsorbed to the soil. Later samples with sediments removed from the same wells met discharge standards. Thus, in the tank farm area only two observation wells had groundwater samples which did not meet relevant criteria; some of these samples exceeded limits for benzene and PAHs. The environmental consultant postulates that, once the benzene and PAH source is removed during soil remediation, groundwater contamination should be eliminated. Thus, groundwater contamination is not an immediate concern, but will be carefully monitored during and after soil remediation.

5.3.1.4 Potentiometric head measurements and slug tests

Potentiometric head measurements were made in all the observation wells to determine the site groundwater head gradient and, thus, flow direction and order of magnitude velocity. Most of the observation wells are screened between about one and six meters below ground surface. The gradient was found to be very sensitive to and dependent on the tide—so much so that the gradient changes direction at many locations of the site. To further analyze the tidal influence on groundwater flow, continuous water level measurements were collected in seven observation wells and compared to the inlet water levels. Water levels within a single observation well fluctuate by as much as 3.5 meters with fluctuations between 1.5 and 2.4 meters in the wells that are located in the tank farm area. As expected, the water level fluctuations in the wells follow the sinusoidal tidal fluctuations in the inlet, having the same wavelength, but different phase due to a delayed response and reduced amplitude. The phase delay and amplitude reduction increases with increasing distance from the shoreline. The environmental consultant estimated that groundwater velocities in the tidally influenced area, which includes the tank farm area, is on the order of meters per day due to the large hydraulic gradient created by the tide. These high velocities and changing flow directions indicate that large quantities of water move in and out of this region on a regular basis.
5.3.1.5 Oil-interface probe measurements

Oil-interface probe measurements were taken in all the observation wells at various times and none of the reported measurements indicate the existence of LNAPL pools floating on the water table. However, apparent floating LNAPL was encountered in two observation wells while attempting to take water table depth measurements during the geophysics campaign. From scientific intuition it would seem that the likelihood of LNAPL pools and high LNAPL pore volume saturations could be significantly diminished, although not necessarily precluded, at this site due to extensive tidal flushing and constant movement of the water table. These characteristic site conditions could also accelerate the dissipation of soil contaminants.

5.3.2 Geophysics

The acquisition and analysis of pertinent geophysical surveys was attempted at the case study site in order to: (1) provide geophysical data for a posterior analysis, (2) to help focus on the best and most realistic methodology for evaluating data worth of geophysical measurements, and (3) to reveal the difficulties inherent to real life field conditions that can be encountered when attempting to implement geophysics in hydrogeology problems. A logical, structured procedure for acquiring geophysical data, as recommended in this work, was attempted, but several deviations from this methodology had to be made due to research limitations of time and resources. Since geophysical surveys were going to be attempted regardless of the expected data worth of geophysics for the engineering problem, and were indeed collected before any data worth analysis was performed, the data acquisition methodology consisted of:

1. identifying targets that potentially can be delineated using geophysics,
2. identifying geophysical techniques that can detect the targets (and were also available for use),
3. performing limited reconnaissance surveys over known targets,
4. analyzing, in a gross sense, the results from the reconnaissance runs to determine whether 
production geophysical surveys would be worthwhile (e.g. are the measurements so noisy as 
to be useless), and 

5. performing production surveys if the answer to (4) was yes.

5.3.2.1 Choice of geophysical techniques

Potential targets for geophysics were identified as non-aqueous phase hydrocarbon 
(NAPL) and coal cinder, both of which are critical to the engineering problem. Thus, appropriate 
geophysical techniques would have to detect a difference in response between regions in the 
subsurface containing NAPL or cinder and uncontaminated regions.

For identifying geophysical techniques which could potentially respond to the targets it is 
important to: (1) determine what small-scale subsurface physical properties (e.g. pore fluid 
electrical conductivity) the targets affect, (2) how these changes in small-scale physical 
properties influence the bulk physical properties of the subsurface (e.g. overall electrical 
conductivity of a unit volume of the subsurface accounting for the contribution of all constituents 
and phases in the grain matrix and pores), and (3) which geophysical methods are sensitive to the 
magnitude of contrasts expected in these bulk properties.

Past research in the form of laboratory or field studies for similar applications can be very 
useful in this quest, although it is important to realize that every site is unique. There has been a 
fair amount of research performed on delineating NAPLs in the subsurface using geophysics and 
there has been some success using GPR and electrical conductivity measuring devices [Olhoeft, 
1986; Brewster, et al., 1995], mostly for monitoring temporal changes in NAPL saturation at a 
single location, although these applications are not well established. Since NAPLs have low 
electrical conductivity and low dielectric constant relative to water, there is a possibility that the 
bulk electrical conductivity and bulk dielectric constant decrease when NAPLs displace water in 
the subsurface; the amount of change depends on many factors including the water saturation, 
water chemistry, NAPL saturation, and soil matrix mineral composition [Olhoeft, 1986; Brewster 
et al., 1995]. As discussed in Section 5.3.1, soil samples showed that hydrocarbon contamination
occurs both in the unsaturated and saturated zones, depending on the position of the water table at any one time. The matrix material was interpreted to be mostly sand or gravel fill, with very little in the way of fine materials across the zone of potential contamination. Sands and gravels generally have low matrix electrical conductivity and dielectric constant while clays (which are very fine grained) have high matrix conductivity and constant.

For these site conditions one would expect the degree of change in bulk electrical conductivity and bulk dielectric constant due to the existence of LNAPLs to be different in the unsaturated and saturated zones and to be highly dependent on the LNAPL pore volume saturation. Since non-clay matrix materials and NAPLs both have similarly low dielectric constants, Olhoeft [1986] argues that the presence of NAPL in the pores instead of water results in a much more uniform bulk dielectric constant. This property would translate into fewer and lower amplitude reflections across the region of NAPL contamination for a GPR reflection survey, since radar reflections are caused by, and proportional to, magnitude of contrasts in dielectric constant [Olhoeft, 1986]. However, for this type of GPR response to occur and be discernable, the NAPL must displace a significant portion of original pore water volume (i.e. there must be a high LNAPL saturation) and cover a large enough volume of the subsurface to be resolvable from the GPR image. For a significant change in bulk electrical conductivity these conditions would also be required, since a large decrease in the value of this physical property would not be expected if most pores and pore throats were still crossed by conductive water. As mentioned in Section 5.3.1.5 above, this scenario is unlikely at this site.

Another possible GPR response to the presence of NAPL in the subsurface is a reflection off the water-NAPL pool interface as a result of the large contrast in dielectric constant between the two. This characteristic response was used to image the migration of a controlled dense non-aqueous phase liquid (DNAPL) release into a saturated sand [Brewster et al., 1995]. However, identifying GPR reflections caused by NAPLs in the subsurface is much more difficult when there is no uncontaminated baseline profile for comparison. NAPL reflections have to be distinguishable from other reflections caused by other factors such as stratigraphic layering,
heterogeneity, and natural changes in water saturation. In the unsaturated zone (probably better referred to as the partially-saturated zone since there is always at least residual water saturation) a significant bulk dielectric constant contrast occurs only if the NAPL displaces a significant proportion of the water content in the pores. NAPL displacing air in the pores does not cause a significant change in bulk dielectric constant since both air and NAPL have low dielectric constants. Thus, a measurable GPR response of this type is expected to be much more possible in the saturated versus the unsaturated zone, unless high saturation pools of NAPL exist in the unsaturated zone.

The effect of coal cinder on subsurface properties is not as obvious as that of NAPLs and there does not appear to be much past research and/or case studies on specifically detecting cinder using geophysics. Chemical analysis of cinder samples indicate the cinder contains a fairly high concentration of metals compared to the average non-cinder material, but also contains significant residual carbon. Thus, the cinder matrix may have higher electrical conductivity than the average, mostly sand, non-cinder matrix, but not necessarily. The cinder material also could have a different porosity and, thus, different bulk electrical conductivity and bulk dielectric constant than surrounding non-cinder material—if the difference in porosity translated to a difference in water content.

Another critical consideration in identifying potentially useful geophysical methods is how suitable the site conditions are to operating the equipment successfully. A particular geophysical method may be very sensitive to the target of interest, but if the measurement signal is not able to reach the target or the signal is overwhelmed by noise then the method is useless. These types of problems are not always identifiable before taking actual field measurements, but often certain basic criteria can be established.

Another tool for evaluating the sensitivity of a geophysical measurement type to the types of contrasts in physical properties expected at a site is geophysical forward modeling. (This technique is discussed in more mathematical detail in 0.) Geophysical forward modeling consists of calculating the geophysical response to a known subsurface distribution of physical
parameters. The idea is to replicate a geophysical survey before taking any measurements. This process requires two models: a model to represent the subsurface distribution of the physical property (or combination of properties) which the geophysical method is sensitive to and a physical model, in the form of mathematical equations, to represent the physics of the measurement. Unfortunately, there is usually considerable uncertainty in the physical property model, considering that it represents the unknown we’re trying to measure. In addition, the model representing the physics, usually in the form of computer algorithms, is often necessarily simplified (e.g. approximating the subsurface as one dimensional) in order to be tractable and acceptably efficient on standard computer systems. Also, it is usually impossible to account for site specific "noise" which might contaminate the measurements. Despite these limitations, if a useable physical model exists for the geophysical method being evaluated, forward modeling can be beneficial for making a first estimate as to whether the method has any chance of success for the types of conditions expected at the site. Forward modeling was not used in this study due to the complexity, cost, and availability of forward models for GPR and FDEM, which were deemed too high relative to the potential benefits gained from performing such an analysis.

Probably the best method of assessing whether a geophysical technique can detect a target is to actually take test measurements on-site using the technique; these are referred to as reconnaissance surveys in this work. In the very least, the effects of noise on the measurements can be assessed. Ideally, the measurements are made in a well-characterized area that has regions where the target is known to exist and others where it is known not to exist. In this way the type of discernable anomaly the target causes on the geophysical response, if any, can be identified directly, accounting for site specific conditions. Limited GPR and FDEM reconnaissance surveys were run in the tank farm area and in a part of the site known to have cinder in the near subsurface. The main purpose of these surveys was to assess the effects of cultural noise on the measurements. While the measurement responses of the two techniques showed significant sensitivity to surface metallic objects—fences and cars—producing high amplitude events in the surveys, away from these objects the signal-to-noise ratio seemed much better—the
measurements being sensitive to actual subsurface properties. Thus, the reconnaissance surveys indicated both GPR and FDEM techniques should not be deemed ineffective due to cultural noise, but care should be taken to avoid measurements near the fences and any cars.

Based on the conclusions from the feasibility analysis for applying geophysics at the case study site, GPR and FDEM were assessed to have some potential of success for delineating NAPL. Coal cinder was dropped as a potential target for geophysics because of limited time for this work and the need to prioritize tasks; GPR and FDEM were deemed to have a higher likelihood of being able to detect hydrocarbons than cinder due to the poorly understood GPR and FDEM response to cinder and the cinder's location as thin layers in the very shallow subsurface—difficult to resolve with both techniques. Delineation of locations with metals contamination could also be attempted with the same measurements taken for hydrocarbons, since both contamination types exist in the tank farm area.

5.3.2.2 GPR Surveys

This section describes the GPR measurement technique that is used for this study, providing a brief overview on the physics of the measurement and a description of the field acquisition campaign undertaken for this study. The detailed analysis of the production survey results is discussed in Section 5.4.5.

5.3.2.2.1 Measurement overview

GPR measures reflected electromagnetic energy versus time resulting from the interaction between a surface transmitter signal and variations in subsurface electromagnetic properties. From the ground surface a radar transmitter antenna emits an electromagnetic wave into the subsurface that travels at certain velocities, loses energy, and reflects and refracts off subsurface boundaries—the nature of these transformations and interactions depending on the electromagnetic properties of the subsurface. A receiver antenna measures reflected/refracted energy—in the form of electromagnetic waves—that returns to the ground surface at the receiver location. The measurement and technique is very analogous to seismic reflection surveys, using
electromagnetic instead of acoustic energy that are sensitive to electromagnetic instead of acoustic properties. The electromagnetic properties that have the greatest influence on the transformations and interactions that the GPR transmitter signal undergoes are dielectric constant and electrical conductivity—the former primarily determining the signal velocity and reflections/refractions and the latter determining the signal attenuation.

The measured GPR response is a signal wavetrain (or trace) of amplitude versus time at each measurement surface location, forming a 2-d cross section of wavetrains when collected at regular increments along a line or a 3-d volumetric section when collected on a surface grid. The amplitude events (of both negative and positive magnitude) in each trace correspond to radar reflections from material interfaces where contrasts in dielectric constant occur. The usual operating mode is to place the transmitter and receiver antenna close to each other (one meter separation was used for this study) so that the measured return signal can be assumed to be reflecting off subsurface horizons directly below the midpoint between the two antenna. The small receiver-transmitter spacing is kept constant as the array is moved along a line or grid of measurement locations. However, since electromagnetic waves propagate spherically in 3-d space, the receiver measures reflections from all directions, including those from non-horizon dielectric constant variations (e.g. buried drums, steel fences on the surface) to the sides of the transmitter-receiver array—if these reflection events arrive at the receiver within the measurement time window and with measurable amplitudes. Such reflections are considered unwanted noise.

The resolution of the GPR measurement is a function of the wavelength of the transmitter signal; an accepted guideline from wave theory is that objects as small as one quarter of this wavelength can be resolved. Wavelength is a function of signal frequency and the electromagnetic wave velocity of the subsurface: \( \text{wavelength} = \text{velocity} \times \text{frequency} \) (e.g. the wavelength of a 100 megahertz antenna system in a medium with a velocity of 0.05 meters per nanosecond is 50 centimeters). Depth of investigation (maximum reflector depth that provides a coherent, recognizable amplitude event in the measured receiver trace) decreases with increasing
signal frequency, but is also highly dependent on the types of subsurface materials. A 100 megahertz (MHz) signal can have a depth of investigation greater than 10 meters in a clean sand while that for 200 MHz signal might be only 6 meters. Thus, there is definitive tradeoff between resolution and depth of investigation that should be optimized for the problem at hand. Both GPR frequencies can have a depth of investigation less than one meter in clay-rich material. In general, the GPR measurement cannot penetrate below horizons with significant clay content amounts, making the measurement ineffective and unusable in high clay environments—another important consideration when determining the feasibility of GPR. Most standard GPR systems have nominal frequencies of 100, 200, or 500 Hertz.

GPR data can be processed in much the same way as seismic reflection data. Typically, the data is initially filtered to reduce noise contributions to the measurement, helping to focus on the data of interest. Measured amplitudes can be reported as: (1) absolute amplitudes across an entire section, wherein amplitudes are boosted only to account for wave spreading losses (wave energy decreases as the inverse radius from the source squared) or (2) relative amplitudes across a predefined time window, wherein amplitudes are normalized to maximum and minimum levels within the window. Since GPR provides an amplitude versus time measurement, the time axis needs to be converted to depth in order to be applicable in the real world for site characterization. The two most common time-to-depth conversion methods used in practice are:

- comparison to groundtruth borehole data where a prevalent horizon of known depth from the groundtruth data can be compared to the associate GPR reflection. Brewster et al. [1995] use this approach to estimate an average velocity which they use to convert all GPR measurements throughout their small subsurface test volume.

- performing a common depth point survey where measurements are made at different receiver-transmitter spacing around a stationary midpoint. By plotting the results as a time versus separation distance section, the shape of the hyperbolae demarcated by reflections from common reflection horizons can be analyzed to estimate layer velocities. This is analogous to normal moveout processing for seismic reflection surveys.
Chapter 5. CASE STUDY

Often this is the extent of processing carried out for standard GPR surveys. However, many of the routines used in seismic reflection processing can be applied to GPR processing (e.g. migration to convert time to depth), depending on the level of sophistication desired, and dollar budget available. It is possible to convert a GPR profile into a dielectric constant section if the electromagnetic property distribution is well known [Knoll and Knight, 1994].

The output from standard GPR processing is a series of depth trace sections displaying the distribution of GPR amplitudes located in depth and horizontal spacing (see Figure 5.4.79 for an example). If a grid of measurements are collected, a 3-d volume can be displayed (see Figure 5.4.80 for an example). In standard practice an interpretation is made from these displays about what subsurface features are causing the GPR reflections. The information derived from the interpretation is then incorporated in the site conceptual model in one form or another. Typically this is a very subjective process. In this study, a more rigorous, quantitative approach for incorporating GPR information is attempted by calibrating the interpreted results to groundtruth data using the Markov-Bayes approach described in this work.

The GPR equipment used for the case study was a PulseEKKO™ system with interchangeable 100 and 200 megahertz (MHz) antenna, controlled from a laptop computer.

5.3.2.2.2 Reconnaissance surveys

Reconnaissance GPR surveys were collected on a limited scale in the tank farm and cinder sections of the site, mainly to determine whether reasonable signal-to-noise ratios could be attained across the regions and depths of interest. A few survey lines were collected in both sections of the site. Station measurements were made every 0.5 meters along these lines. The tank farm lines were acquired near a steel chain-link fence, the effect of which is clearly visible as a diagonal reflector—starting near the surface where the measurement station is adjacent to the fence and extending downwards in time/depth as the stations move away from the fence. The same effect is produced by nearby parked vehicles. Based on these results, it is obvious that steel objects on the surface produce unwanted noise in the GPR measurements, and measurements
should be made at least 5 meters from such objects. Otherwise, it was apparent that measurements of true subsurface horizon reflections could be obtained.

The survey lines from the cinder area are less conclusive and promising about the applicability of GPR for delineating cinder material. High amplitude, near surface reflections occur in the GPR sections, and are continuous across the section. This is the region where the cinder resides. The survey lines cross from an area where it is known there is no cinder into one where it is known there is cinder. The GPR response in the upper part of the section is essentially the same across these areas, indicating there is very little hope in delineating a unique signal response from the cinder. The biggest problem appears to be that the strong direct arrival from the transmitter to the receiver occurs in the same time frame as any reflections from the near surface cinder would be arriving, thus masking any useful signal.

Based on the reconnaissance surveys, GPR was judged to have potential for delineating pure phase hydrocarbons in the tank farm area, but very low likelihood for providing any information about cinder distribution. Thus, it was decided to perform GPR production phase surveys only in the tank farm area.

5.3.2.2.3 Production surveys

Twenty-five GPR production survey lines were acquired in the tank farm area. The lines are in a dense grid bounded by steel chain-link fences on two sides (at the time of acquisition) and the harbor on one side. The surface material at the time of acquisition was asphalt pavement. Spacing between lines is one meter and spacing between measurement stations along the lines 0.5 meters for most lines and 0.25 for the rest. The production survey centered on collecting 200 MHz frequency data, although 100 MHz measurements were taken along a number of the lines as well. The follow-up analysis was only performed on the 200 MHz data since it was considered to have a greater possibility of utility for contaminant delineation, based on its higher resolution. The higher resolution of the 200 MHz data, but greater depth of investigation of the 100 MHz data, is evident from the two sections. The processing, analysis, and results of the GPR production surveys are described in Section 5.4.5.
5.3.2.3 FDEM Surveys

This section describes the FDEM measurement technique that is used for this study, providing a brief overview on the physics of measurement and a description of the field acquisition campaign undertaken for this study. The detailed analysis of the production survey results is discussed in Section 5.4.5.

5.3.2.3.1 Measurement overview

FDEM measures electrical conductivity of a region surrounding the probe by powering an alternating current (AC) of a certain frequency through a transmitter coil(s), inducing eddy currents in the measurement region. The induced AC eddy currents, in turn, induce an electrical response in a receiver coil(s) that is calibrated to measure apparent electrical conductivity in the region of measurement. The measurement is largely unfocused, although slightly different depths of investigation can be measured by rotating the coils 90 degrees between horizontal and vertical coil orientations (dipoles). The instrument used for this case study, the Geonics, Limited EM31™, has an approximately three meter long tube with an electronics control box in the center. The transmitter and receiver induction coils are at the ends of the tube. The volume of investigation of the EM31™ is something like an ellipsoid, with axes of 4 meters (parallel to long axis of tool) by 3 m (perpendicular to long axis) by 4 m (depth) for the horizontal dipole measurement, and about the same dimensions, except 6 m depth, for the vertical dipole. Since air is highly non-conductive and the tool is sensitive to conductivity, not resistivity, the half of the ellipsoidal volume of investigation above ground contributes a negligible amount to the overall measured signal. However, this changes dramatically if there are conductive objects on the surface. The measurement is severely influenced by such objects (e.g. steel fences, cars), resulting in measurement noise that masks the subsurface signal of interest, possibly reducing the signal-to-noise ratio significantly.

The value reported by the EM31 measurement is internally calibrated to be conductivity in millisiemens (mS). It is often referred to as “apparent conductivity” because the value is not the
conductivity of a particular object or lithological unit, but some sort of filtered average of the conductivity of all the soil, water, and surface objects within the volume of investigation; the value is dependent on the measurement filter (see Section 2.2.1.1 for a description of the inherent filtering process associated with any measurement). With this viewpoint, the “true conductivity” of subsurface materials of interest is only obtained after deconvolving the filtering process of the raw measurement response, requiring knowledge of the filtering process and an inversion algorithm. While rigorous inversion approaches can be applied to FDEM measurements, usually no processing beyond simple gridding of the apparent conductivity data is applied to interpolate/extrapolate the data to the model scale grid; this is especially true for straightforward shallow-reading, single frequency FDEM techniques, such as the EM31.

5.3.2.3.2 Reconnaissance surveys

Reconnaissance FDEM surveys were collected in the same areas as the GPR reconnaissance surveys—locations in the tank farm and cinder sections of the site, mainly to determine whether reasonable signal-to-noise ratios could be attained across the regions and depths of interest. Station measurements were made every one meter along these lines—four separate measurements at each station:

1. horizontal dipole with the long axis of the tool oriented parallel to the survey line,
2. horizontal dipole with the long axis oriented perpendicular to the line,
3. vertical dipole with the long axis oriented parallel, and
4. vertical dipole with the long axis oriented perpendicular.

As with the GPR, the effect of the steel chain-link fences in the tank farm lines are clearly visible in the measurements made at stations near the fences. These measurements have conductivity anomalies, either very high or very low, compared to the measurements further away from the fences. A telltale of an anomalous EM31 result is when there are large differences in the conductivity values measured at the two different orientations; the conductivity results near the fences clearly exhibit this difference. Similar anomalies, but with much lower magnitudes, are produced by nearby parked vehicles; it appears that the longer, more continuous
the electrically conductive surface object, the worse the measurement is hindered. Besides the anomalies near large surface steel objects, the EM31 tank farm results appear to be well-behaved, with reasonable conductivity values—very similar at the two tool orientations—that vary smoothly in space—expected for the fairly large measurement volume of investigation. Based on these results, it is obvious that steel objects on the surface produce unwanted noise in the FDEM measurements, and measurements should be made at least 2 meters from such objects. Otherwise, it was apparent that measurements of true subsurface conductivity could be obtained.

As with the GPR results, the survey lines from the cinder area are less conclusive and promising about the applicability of FDEM for delineating cinder material. There is no detectable change in the measured conductivity between the area in the survey lines where it is known there is cinder and those where it is known there is not. The cinder typically occurs as a thin layer in the top one meter or less and, thus, constitutes only a small portion of the overall EM31 measurement volume; this could lead to the difficulty in resolving its presence.

Based on the reconnaissance surveys, FDEM was judged to have potential for delineating metals and pure phase hydrocarbons in the tank farm area, but very low likelihood for providing any information about cinder distribution. Thus, it was decided to perform FDEM production phase surveys only in the tank farm area.

5.3.2.3.3 Production surveys

Fifteen FDEM production survey lines were acquired in the tank farm area using the EM31. The lines are in a one-by-one meter grid bounded in the direction of the surveys by a road and the harbor. Within the survey grid are two steel chain-link fences at approximately right angles to each other. Measurements were made in the direct vicinity of the fences; how to deal with the known noise in these measurements was left for the processing. The surface material at the time of acquisition was mostly asphalt pavement with some grass or topsoil in some areas. The same four separate measurements as in the reconnaissance surveys were made at each station. The processing, analysis, and results of the FDEM production surveys are described in Section 5.4.5.
5.4 Analysis and Modeling

5.4.1 Risk-Cost-Benefit Decision Model

The section describes the key components of the decision model developed for the case study and how the model is used to perform important decision-making tasks—choosing site remediation designs and evaluating the value of different site characterization programs—using a risk-cost-benefit decision analysis approach.

The first step in performing a risk-cost-benefit decision analysis for the remediation phase of the case study real estate development project is to develop a decision model for performing economic analyses, wherein the economic outcome of all possible decision alternatives can be assessed. The decision model should have the ability to automatically determine the most cost effective decision alternative for the following scenarios of knowledge about the site:

- Current, present-day information
- Perfect information
- Information provided by possible future site characterization programs, in addition to existing information.

First, the decision-making parameters for the engineering problem of concern (remediation of the site prior to construction) are defined. These are presented in Table 5.4.1.
Chapter 5. CASE STUDY

<table>
<thead>
<tr>
<th>Objective</th>
<th>Implement most cost-effective remediation design—in context of overall project.</th>
</tr>
</thead>
<tbody>
<tr>
<td>Decision</td>
<td>Which engineering design should be implemented to remediate site?</td>
</tr>
<tr>
<td></td>
<td>• Excavation, treatment, and disposal of soil for level C / special waste, level B, or no contamination</td>
</tr>
<tr>
<td></td>
<td>• Excavation and batch sampling and testing of soil; treatment and disposal as warranted by test results</td>
</tr>
<tr>
<td></td>
<td>• Confirmatory sampling and testing of soil on sides of excavation hole; further excavation, treatment, and disposal of sampled volume—as warranted by test results</td>
</tr>
<tr>
<td>Questions and Their Alternatives</td>
<td>Should further site characterization be carried out and, if so, what technique(s) should be employed and where?</td>
</tr>
<tr>
<td></td>
<td>• Borehole sampling (how many boreholes?)</td>
</tr>
<tr>
<td></td>
<td>• Geophysics – GPR and/or FDEM (what data quality level is required?)</td>
</tr>
</tbody>
</table>

Table 5.4.1a: Case study decision model parameters—objective and decision questions
## Constraints

- **Technical**
  - minimum volume of soil that can be excavated at one time – cube of 1 x 1 x 1.5 meters (approximate volume of standard backhoe shovel)
  - excavation wall slope angle – 1 meter benches required for step down
- **Economic**
  - maximum allowable time for remediation before site development starts – time not considered
- **Regulatory**
  - must remediate soil above thresholds set for land use at that location
- **Social**
  - limit amount of remediation required during development phase – fuzzy constraint so perform analyses with range of values for cost of remediation during development phase

<table>
<thead>
<tr>
<th>Constraints</th>
<th>Technical</th>
<th>Economic</th>
<th>Regulatory</th>
<th>Social</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>• minimum volume of soil that can be excavated at one time – cube of 1 x 1 x 1.5 meters (approximate volume of standard backhoe shovel)</td>
<td>• maximum allowable time for remediation before site development starts – time not considered</td>
<td>• must remediate soil above thresholds set for land use at that location</td>
<td>• limit amount of remediation required during development phase – fuzzy constraint so perform analyses with range of values for cost of remediation during development phase</td>
</tr>
</tbody>
</table>

**Table 5.4.1b: Case study decision model parameters—constraints**
### Table 5.4.1c: Case study decision model parameters—decision variables

<table>
<thead>
<tr>
<th>Decision Variables</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Independent</strong></td>
<td>• contaminant concentrations from soil samples</td>
</tr>
<tr>
<td></td>
<td>• unit costs during <em>remediation</em> phase of project</td>
</tr>
<tr>
<td></td>
<td>⇒ excavation - $4 per cubic meter</td>
</tr>
<tr>
<td></td>
<td>⇒ treatment and disposal for different contaminants:</td>
</tr>
<tr>
<td></td>
<td>Level B - $19 per cubic meter</td>
</tr>
<tr>
<td></td>
<td>Level C and special waste - $135 per cubic meter</td>
</tr>
<tr>
<td></td>
<td>⇒ batch sampling - $4 or $8 per cubic meter</td>
</tr>
<tr>
<td></td>
<td>⇒ confirmatory sampling - $1 per cubic meter</td>
</tr>
<tr>
<td><strong>Dependent</strong></td>
<td>• probability of different types of contamination across site</td>
</tr>
<tr>
<td></td>
<td>• soil partitioning - leave in place, excavation, treatment and disposal for different contaminants, batch sampling, confirmatory sampling</td>
</tr>
<tr>
<td></td>
<td>• costs - remedial option costs for individual soil volumes, total cost of remediation for different design alternatives</td>
</tr>
</tbody>
</table>

These decision parameters need to be represented mathematically in the decision model, in the form of an objective function. The objective function addresses the problem objective and,
thus, should reflect the economic cost of remediation in the overall project. This overall cost can be split into the sum of two costs:

1. remediation costs during the remediation phase of the project and
2. the cost of remedial activities that are required after the remediation phase is completed, due to contamination that is bypassed during remediation.

The costs include soil excavation, sample testing of excavated soil, and treatment and disposal of excavated soil. The second cost above is a consequence of the uncertainty about where soil contamination occurs on the site; it represents the risk cost. These costs are a function of the decision variables. The objective function is a mathematical representation of these costs that can be evaluated for different remediation alternatives. The goal is to minimize this objective function, thus identifying the most cost effective overall remediation design for a particular scenario (see above).

A prevailing assumption in the decision model is: if there is soil contamination above level B or C (special waste for MOG) regulatory thresholds, it has to be excavated and treated at some point in time. If it is missed during the remediation it will be uncovered at a later time, most likely during the development phase when building foundations are dug, at which time it will have to be dealt with, probably at an elevated cost.

To set up the decision model the site is divided into a 3-d grid of blocks, having dimensions of the minimum volume of soil that can be efficiently excavated by a backhoe, taken as 1 x 1 meter sides x 1.5 meter depth. This volume represents the decision scale for the problem and, thus, should also be the model scale. The spatial domain chosen for the model is a 220 by 100 meter area that encompasses the tank farm zone of the site. The domain is bordered, and partially penetrated, on one side by the harbor; areas with water are excluded from the analysis. The model extends to a depth of nine meters, which is greater than the maximum depth of measured contamination and below the inferred maximum depth of the underlying clay aquitard. This designated analysis domain is divided into a 3-d grid of the unit model/decision scale blocks.
Chapter 5. CASE STUDY

The easiest way to determine and describe the mathematical representation of the costs for different alternatives is to look at a schematic representation of possible soil remediation designs, as shown in Figure 5.4.1. For each soil block there are six possible remediation design alternatives, as shown in the diagram. Boreholes have been drilled in some locations, providing soil samples that have been chemically analyzed to determine contaminant concentrations. The distribution of contamination reaching level B and C regulatory levels are depicted as solid blue and red blobs; these are unknown to the decision maker, except at the locations intersected by boreholes. Suppose that remediation designs for many of the blocks have already been determined (shaded blocks). Now the most cost effective alternative for the block labeled “x0”, needs to be determined. The costs of all possible alternatives need to be computed and compared with each other. The block could have contaminated soil reaching level B or C, or not reaching either of these levels (the latter being reality, unbeknownst to us).

![Figure 5.4.1: Schematic case study remediation design](image)

**Remediation design alternatives:**
- **excavate, treat, and dispose for level B**
- **excavate, treat, and dispose for level C**
- **excavate and batch test; treat accordingly**
- **take confirmatory sample (only if exposed)**
- **excavate only to reach blocks below (no treatment)**
- **no action**

**Figure 5.4.1:** Schematic case study remediation design
The cost of remediating the block for level C contamination is:

\[ C_{\text{rem}}^{(C)}(x_0) = C_{\text{exc}} V_{\text{exc}} + C_{\text{dis}}^{(C)} V_b \quad (5.1a) \]

\[ V_{\text{exc}} = V_b n_{\text{exc}}(x_0) \quad (5.1b) \]

where \( C_{\text{exc}} \) is the unit cost of excavation, \( V_{\text{exc}} \) is the total volume of soil requiring excavation (including soil above), \( C_{\text{dis}}^{(C)} \) is the unit cost of treatment and disposal for level C contamination, \( V_b \) is the block volume (1.5 cubic meters), and \( n_{\text{exc}}(x_0) \) is the number of blocks requiring excavation (including blocks above that have to be removed to reach block \( x_0 \)).

The cost of remediating the block for level B contamination is:

\[ C_{\text{rem}}^{(B)}(x_0) = C_{\text{exc}} V_{\text{exc}} + C_{\text{dis}}^{(B)} V_b + P_u^{(C)}(x_0)(C_{\text{dis},u}^{(C)} - C_{\text{dis}}^{(B)}) V_b \quad (5.2) \]

where \( P_u^{(C)}(x_0) \) is the probability of underclassifying the soil contamination in block \( x_0 \)—where it actually has contamination reaching level C. \((C_{\text{dis},u}^{(C)} - C_{\text{dis}}^{(B)})\) represents the additional unit treatment and disposal cost that will be incurred if the block contains level C soil. Thus, the third term on the right hand side is a probabilistic risk cost. The level C disposal cost in this case, \( C_{\text{dis},u}^{(C)} \), represents the elevated cost of having to dispose of the contaminated soil at a future time as a consequence of underclassifying the contamination during the remediation phase.

The cost of taking no action—leaving the soil intact—is:

\[ C_{\text{na}}(x_0) = P_u^{(C)}(x_0)(C_{\text{exc},u} V_{\text{exc}} + C_{\text{dis},u}^{(C)} V_b) + P_u^{(B \leq C)}(x_0)(C_{\text{exc},u} V_{\text{exc}} + C_{\text{dis},u}^{(B)} V_b) \quad (5.3) \]

where \( P_u^{(B \leq C)}(x_0) \) is the underclassification probability of soil contamination reaching level B, but not C, in block \( x_0 \). This cost is entirely a risk cost; no costs are incurred during the remediation phase, but there is a risk that elevated remediation costs will occur at a future time. All unit costs on the right hand side of the equation are elevated costs associated with having to perform remedial activities after the remediation phase of the project.

The cost of excavating the block, performing a batch sample test, and treating and disposing of the soil accordingly is:

\[ C_{\text{bt}}(x_0) = C_{\text{exc}} V_{\text{exc}} + C_{\text{bt}} V_b + P^{(C)}(x_0) C_{\text{dis}}^{(C)} V_b + P^{(B \leq C)}(x_0) C_{\text{dis}}^{(B)} V_b + (1 - P^{(C)}(x_0) - P^{(B \leq C)}(x_0)) C_{\text{dis}}^{(B)} \quad (5.4) \]
Chapter 5. CASE STUDY

where $C_{bs}$ is the unit cost of performing a batch sample test. The second and third terms on the right hand side are the probabilistic costs of treating and disposing of the soil for level C and B contamination, respectively. The last term is the probabilistic risk cost of having to dispose of or reuse the soil if it contains no contamination. This cost is a risk cost because the uncontaminated soil could have been left intact; now that it has been excavated it has to be dealt with.

The cost of performing a confirmatory sample test on the block (assuming it is exposed on the side of an excavation), and excavating, treating, and disposing of the soil accordingly, is:

$$C_{cs}(x_0) = C_{cs} + P^{(C)}(x_0)(C_{dis}^{(C)} V_b + C_{exc} V_{exc}) + P^{(B+C)}(x_0)(C_{dis}^{(B)} V_b + C_{exc} V_{exc})$$

(5.5)

where $C_{cs}$ is the unit cost of performing a confirmatory sample test. The second and third terms on the right hand side are the probabilistic costs of excavating, treating, and disposing of the soil for level C and B contamination, respectively.

An underlying condition of this model is that, if soil is treated and disposed of for contamination of a certain severity level, no additional action is required for any contamination at less severe levels in that soil. (While true for this problem, if this was not the case, the model could be adjusted to allow different actions depending on the combination of contamination types by adding more design alternative equations and adjusting the equations to account for these additional options.) This exclusivity condition is handled in the model by prioritizing the regulatory action levels across all three contaminants together. For this problem, all level C and special waste contamination are dealt with in the same way, regardless of contaminant type, requiring more extensive (and costly) treatment and disposal than level B contamination. The actions required for Level B contamination are the same for all contaminants, as well.

All variables on the right hand side of the above equations are independent decision variables (known or assumed), except the probabilities of contamination at different regulatory levels. These need to be estimated from the geostatistical model. As will be discussed in the next section, probabilities of contamination can be directly estimated using the M-B indicator geostatistical methodology described in this work. As denoted above, the probabilities in the decision model are actually mutually exclusive probabilities—the probability of a certain type of
contamination at a certain regulatory level and no contamination at any of the higher priority levels—due to the remedial action exclusivity condition (if both level B and C contamination, treat for C only). This requires a conversion from the contaminant-independent probabilities output of the geostatistical estimation model to conditional probabilities, described later in Section 5.4.3.1.

The costs for the different possible remediation design alternatives represent solutions of the risk-cost objective function. Once computed, the least costly remediation design for block $x_0$ can be identified by simply comparing the costs. Ideally, then, the same computation can be performed for all blocks across the site, the final result being the overall most-effective remediation design for the site. However, the problem is greatly complicated by the fact that the action taken for an individual block of soil affects the remediation alternative decision for nearby blocks. This occurs because the excavation “envelope” required for level B and C remediation, and batch sampling of a block, influences the costs for nearby blocks within this envelope. The consequence of this complication is that the decision analysis cannot be made on a block by block local basis if the truly most cost effective solution is to be obtained. It has to be performed on a global basis where the influence of an action at an individual block on every other block it can affect is accounted for in the choice of optimal design alternative for that block. This quickly becomes an intractable mathematical problem for a grid of any considerable size, as with this case study, due to the enormous amount of possible outcomes to assess.

To tackle the intractable global nature of the decision analysis problem a heuristic iteration technique is developed to identify the global remediation design solution. The technique can be summarized as follows:

1. Starting from the layer of blocks on surface and progressing downwards layer by layer to the bottom depth, the blocks that have the same optimal remediation design regardless of the number of blocks in its excavation envelope that are excavated are identified. Since the number of possible excavation combinations become very large with greater depth, the configuration of combinations is randomly limited to a certain chosen number.
Chapter 5. CASE STUDY

2. The same process is repeated, but the results from the first iteration are accounted for in choosing possible excavation scenarios, effectively reducing the number of scenario combinations.

3. The iterations are continued until a remediation design solution has been obtained for every block. If blocks without a unique solution remain after all those with unique solutions have been accounted for, or a maximum number of iterations has been reached, the remaining blocks are looped through and the optimal remediation design for each is determined, assuming the most current existing excavation scenario. The order of block-to-block calculation loop is chosen randomly.

While there is no guarantee this technique finds the optimal design solution, the hope is that it at least finds a “near-optimal” solution, at a reasonable computational cost.

The decision model is designed for three modes of operation—to determine the near-optimal overall site remediation design for the three information scenarios described earlier: (1) present-day knowledge about the site, (2) perfect information, and (3) additional information provided by hypothetical future site characterization. The first mode, referred to as prior or posterior analysis (depending on the stage of site characterization), simply incorporates the probability of contamination site-wide distributions for metals, PAH, and MOG, determined from existing site data, in the decision analysis process described above.

The second decision analysis mode, referred to as expected value of perfect information (EVPI) analysis, uses a more simplified decision model that doesn’t have any risk cost terms. The type of, or lack thereof, contamination in each block is assumed known from prefect information about the site, provided in contaminant distribution indicator geostatistical simulations. Calculating the optimal remediation design simply entails computing the known cost of remediation for each block, starting on the surface layer and looping through all blocks, layer by layer, down to the bottom layer. EVPI is the difference between the calculated total prior remediation cost and that calculated assuming perfect information about site contamination.
The third decision analysis mode, referred to as preposterior, data worth, or value of information (VOI) analysis, uses an iterative version of the prior analysis to determine the site near optimal remediation design for each in a series of site probability of contamination maps. The probability maps reflect the added information provided by simulated hypothetical site characterization results. After the remediation designs are computed for all simulated scenarios, the expected value of the site remediation cost with the additional characterization information is computed by taking the simple average of the equally probable individual design costs.

5.4.2 Geostatistical Model

This section describes the development of the geostatistical model used for predicting the spatial distribution of soil contamination across the case study site. Estimates of the probability of PAH, metals, and MOG contamination above regulatory threshold levels at all the grid blocks in the decision model are required for execution of the model. They are the only independent decision variables that are empirical quantities—underlying the importance of the geostatistical model to the decision-making in this problem. In addition, simulation of equally probable realizations of the different contaminant distributions across the model domain are required for the data worth analysis. As mentioned throughout this work, the goal is to implement the Markov-Bayes indicator geostatistics methodology [Zhu and Journel, 1993] to compute these probability maps and realizations, so as to enable the local updating of expected value and simulation estimates of contamination levels from indirect information, especially geophysics. Implementing this methodology requires the inference of a stationary indicator geostatistical spatial structure model for each contaminant, at each regulatory threshold level; these models are necessary for indicator kriging and simulation, which locally update the contamination level random variable at each grid block. The following two sections discuss: (1) the preliminary analysis of existing soil sample data for constraining the pool of possible geostatistical models and (2) the inference and validation of the stationary expected value and spatial correlation models for each indicator random variable, incorporating all available information.
Chapter 5. CASE STUDY

5.4.2.1 Exploratory data analysis

The first step in developing the geostatistical model for the study site is to analyze any data that currently exists, primarily the soil sample chemical analysis results. The purpose of this initial analysis is to become familiar with the important characteristics of the data, including patterns of regularity, the distribution of data values independent of location, and the spatial variability of the data. Since the timing of the case study occurred in the late stages of site characterization for the site, a significant amount of data was available for the this exploratory analysis.

With the large amount of data from across the study area, a good way to visualize the general existence and pattern of contamination is to display the soil sample results in 3-d space. The metals, PAH, and MOG sample results are shown in Figures 5.4.2, 5.4.3, and 5.4.4, respectively, from different viewing perspectives; the symbols identifying the sample locations are color coded according to classification of the results relative to their respective contaminant regulatory threshold levels. Evident from these displays is: (1) the spatial clustering of contaminated samples in two zones, (2) the close spacing of contaminated and uncontaminated samples, especially in the vertical direction, and (3) the high concentration of samples taken in the identified contaminated zones. The sampling was able to identify several contaminated zones, where a majority of the sampling was then focused. While these zones are quite large features, the sample results are not consistent within the zones, suggesting there is significant heterogeneity in the contaminant distribution—especially for metals and MOG, and in the vertical direction. While many of the samples were analyzed for more than one contaminant type, many were not, and the number of samples analyzed for metals is much less than PAH and MOG. Because of this fact, gleaning information about one contaminant type from the others could be quite useful; this can be (and is) accomplished with the M-B cokriging approach utilized for geostatistical estimation.
Figure 5.4.2a: 3-d cube of case study soil sample metals analysis results (top view)
Figure 5.4.2b: 3-d cube of case study soil sample metals analysis results (side view)
Figure 5.4.3a: 3-d cube of case study soil sample PAH analysis results (top view)
Figure 5.4.3b: 3-d cube of case study soil sample PAH analysis results (side view)
Figure 5.4.4a: 3-d cube of case study soil sample MOG analysis results (top view)
Figure 5.4.4b: 3-d cube of case study soil sample MOG analysis results (side view)

While these plots seem to indicate contamination occurring in spatial clusters, since the sample spacing across the site is non-uniform—mostly concentrated in the identified areas with contamination—it was decided to assume the a general stationary spatial correlation model, instead of non-stationary (e.g. trend), model; not enough information was available from other areas of the site.

Histogram and ogive plots provide information about the spatially-independent distribution of the contaminants across the site. Figures 5.4.5, 5.4.6, and 5.4.7 are histogram and ogive plots of the metals, PAH, and MOG sample analysis results, respectively, classified according to the regulatory threshold levels. The percentage of samples with contaminant concentrations above
the highest regulatory level is low—6% and 10% for metals and PAH level C, respectively, and
6% for MOG special waste. While the frequency of samples with metals level B classification is
also relatively low (18%), occurrence of samples with PAH level B contamination is surprisingly
high (54%). Since all but one of the contaminant classifications have very low frequencies of
occurrence, it can be anticipated that the decision model will be sensitive to small percentage
changes in the inferred expected values for these classifications. This concern, as well as the fact
that the soil samples are highly spatially biased, emphasis the importance of performing a careful
and robust inference of the stationary expected values, including a data declustering analysis.

Figure 5.4.5a: Histogram of soil samples analyzed for metals

Figure 5.4.5b: Ogive of soil samples analyzed for metals
Figure 5.4.6a: Histogram of soil samples analyzed for PAH

Figure 5.4.6b: Ogive of soil samples analyzed for PAH

Figure 5.4.7a: Histogram of soil samples analyzed for MOG
5.4.2.2 Model inference and validation

The geostatistical model chosen for the case study problem is a set of three indicator spatial random fields, one for each of the contaminant types (metals, PAH, and MOG), that are assumed to be stationary with respect to the global expected value and spatial correlation function. Each random field consists of indicator random variables (IRVs) for the regulatory level classifications that apply to the respective contaminant; these random variables are binary—either having a value of one if contaminant concentration is within the class or zero if not—and can vary spatially. The expected value of the IRVs is the probability of soil contamination within the bounds of the regulatory level class, at a particular location. The stationarity assumption, in essence, means that the global expected value and spatial correlation function of each IRV do not vary across the study area; the expected value is a constant while the spatial correlation function can take on any form, but does not vary from location to location. These global models are locally updated in geostatistical kriging and simulation estimation.

This geostatistical model necessitates the inference of three separate model components: (1) the global IRV expected value for each contamination class, representing the overall probability of contamination at that respective level, (2) the global IRV spatial covariance for each contamination class, and (3) the spatially-independent cross correlation between different contamination classes.
Chapter 5. CASE STUDY

contaminant IRVs. The third component is associated with using the information about concentration levels for one contaminant type to provide indirect information about another contaminant—soft data cokriging—and is specific to the M-B indicator cokriging approach.

Inference of the IRV global expected value for each contamination classification is accomplished through declustering analysis. The declustering analysis routine from the GeoStatistical Library (GSLIB) [Deutsch and Journel, 1992] was modified to permit the inclusion of “soft”, secondary data—other contaminant concentration levels—in the estimation of the primary variable mean (see Section 4.3.1.3). Using soft data in this calculation, however, creates a contradiction, because the secondary data cdf values used in the calculation are themselves a function of the primary variable mean. To account for this contradiction an iterative process is employed:

1. First the declustered mean estimate from hard data only is calculated.
2. This hard mean is used in the soft data calibration for normalization.
3. The hard-soft declustered mean estimate is calculated using the soft data information.
4. The hard-soft mean is used in the soft data calibration once more to compute the final calibration parameters.

The results of the declustering analysis for metals, PAH, and MOG are shown in Figures 5.4.8, 5.4.9, and 5.4.10, respectively. The change in the calculated mean (vertical axis)—from hard data only, soft data of each of the two other contaminants only, and the combined hard-soft mean—for a range of declustering cell sizes (horizontal axis) is plotted. Only the results of the final iteration of the analysis are shown. The cell size with the highest mean value, the apex of the curve, is chosen as the optimal declustering cell size, and the associated mean the optimal spatially unbiased mean. This mean corresponds to the highest probability that the respective contaminant concentration is below the said regulatory threshold level (lowest probability of contamination). That we are choosing a declustered mean that relates to lower probability of contamination makes sense from the spatial distribution of the samples—clustered in identified
contamination "hotspots" leading to biasing of the number of samples having high contaminant concentrations.

The final declustered means chosen as the inferred expected values for each contaminant classification are tabulated in Table 5.4.2.

<table>
<thead>
<tr>
<th>Contaminant type</th>
<th>Regulatory threshold</th>
<th>Inferred expected value of contamination below threshold level</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metals</td>
<td>C</td>
<td>0.931</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.974</td>
</tr>
<tr>
<td>PAH</td>
<td>C</td>
<td>0.628</td>
</tr>
<tr>
<td></td>
<td>B</td>
<td>0.950</td>
</tr>
<tr>
<td>MOG</td>
<td>Special waste</td>
<td>0.965</td>
</tr>
</tbody>
</table>

Table 5.4.2: Declustering analysis results for all contaminant levels
Figure 5.4.8a: Declustering analysis results for metals threshold level B
Figure 5.4.8b: Declustering analysis results for metals threshold level C
Figure 5.4.9a: Declustering analysis results for PAH threshold level B
Figure 5.4.9b: Declustering analysis results for PAH threshold level C
The global IRV spatial covariance models for each contamination class are inferred using the three step approach described in Section 4.3.1.4:

1. Calculate indicator experimental variograms using all sample data results for that contaminant type.
2. Fit variogram mathematical function models to the experimental variograms.
3. Validate the models using the orthonormal residuals methodology [Kitanidis, 1991]. If the models do not suffice, alter them and perform the validation again. Iterate this procedure until an acceptable model fit is identified.

While covariograms are the spatial correlation model type used in the M-B indicator geostatistical estimation algorithms employed for this study, variograms are much more easier to calculate from data. Variogram mathematical models are easily converted to covariance models.

Plots of the experimental variograms and their corresponding fitted and validated models for metals, PAH, and MOG regulatory classifications are shown Figures 5.4.11, 5.4.12, and 5.4.13, respectively. Immediately evident from these plots is that there are very few points in the experimental variograms and, in many cases, the variogram amplitudes (denoted gamma) vary significantly between adjacent points; and these experimental variograms are the best ones that could be generated. This makes it very difficult to visually fit a mathematical model to the variograms. The inability to calculate well-behaved experimental variograms stems from the fact that, while the overall sample pool for analysis is fairly large, the number of samples above the indicator threshold levels is relatively small. The consequence is that there are very few data pairs (or none) with zero-one or one-one indicator value combinations (zero and one corresponding to contaminant concentration above and below the threshold level, respectively) at the different lag distances; most of the data pairs are one-one combinations. This condition accentuates the importance of performing the iterative model validation step.
Chapter 5. CASE STUDY

Figure 5.4.11a: Horizontal variography for metals level B threshold.

Figure 5.4.11b: Vertical variography for metals level B threshold.
Figure 5.4.11c: Variography for metals level C threshold.

Figure 5.4.12a: Horizontal variography for PAH level B threshold.
Figure 5.4.12b: Vertical variography for PAH level B threshold.

Figure 5.4.12c: Variography for PAH level C threshold.
While choosing appropriate variogram models is very difficult in this case due to the limited information provided by the experimental variograms, some parameters, or parameter constraints, defining the general features of the models can be set, and the parameters that should be allowed to vary identified. As Kitanidis [1991] recommends, it is always better to start with a simple model, especially if little information is available. A spherical variogram model type seems to fit the data as good as any other. The correlation length, or range, in the horizontal and vertical directions, reflected as the slope of the variogram—how fast gamma values reach the asymptotic gamma with increasing lag distance, is seen to vary significantly for many contamination classes; an example is the difference between the PAH Level B vertical (green crosses) and horizontal (red circles) indicator experimental variograms in Figure 5.4.12. Thus, vertical anisotropy is set as a model parameter that is allowed to vary. Since the variance of an IRV is a function of the expected value only (\( C_i(0,z_c) = F(z_c)(1 - F(z_c)) \)), where \( F(z_c) \) is the expected value for cutoff \( z_c \), the “C” variogram model parameter was set to the variance calculated from the inferred expected value for that contaminant class, and not allowed to vary.
The nugget for all variogram models was set to zero; this seemed to result in good model validation results and there is not enough information to suggest otherwise.

The inferred spherical variogram model parameters for each contaminant classification are tabulated in Table 5.4.3, along with the orthonormal residual analysis results, defined by the criteria and actual final values for the \( Q_1 \) and \( Q_2 \) goodness of fit statistics (see Section 4.3.1.4.3 and [Kitanidis, 1991] for an explanation and definition of these statistics). The inference process was accomplished by curve fitting the spherical mathematical function to the experimental variogram, validation checking of the function, and adjustment of the range and vertical anisotropy—performed in an iterative manner until a valid model, with the lowest possible \( Q_1 \) and \( Q_2 \) values, was identified. This process was performed manually; it could be set up so that the parameters are optimized mathematically, where the optimization criteria are defined by the orthonormal residual statistics [Kitanidis, 1991].

![Table 5.4.3: Variogram parameters for all contaminant levels](image)

The inference of the secondary to primary variable correlation models is accomplished using the M-B calibration procedure described in Section 4.3.1.2, in conjunction with the expected value inference (see above). First calibration data sets are constructed, comprised of all
collocated primary-secondary variable soil sample data pairs. The calibration is performed with these data sets, wherein the primary and secondary variable thresholds are set as the regulatory threshold concentrations for each contaminant. The outputs from the calibration are:

- a cdf probability table of the probabilities of the primary variable having a contaminant concentration below each threshold that applies to that contaminant, given the secondary variable belongs to a certain contaminant class,
- the P1 and P2 misclassification probabilities for each primary variable threshold level.

See Section 4.3.1.2 for a more detailed description of these outputs and how they are computed.

Two calibrations are performed for each contaminant type, one for each of the other contaminants, to determine the correlation between it and the others. The two calibration cdf tables for metals, PAH, and MOG, and the corresponding misclassification probabilities, normalized by their inferred contaminant level expected values, are displayed as ogives in Figures 5.4.14-18.
Figure 5.4.14: Calibration of PAH samples to metals – (a) cdf table, (b) misclassification probabilities.
Figure 5.4.15: Calibration of MOG samples to metals – (a) cdf table, (b) misclassification probabilities.
**Figure 5.4.16**: Calibration of metals samples to PAH — (a) cdf table, (b) misclassification probabilities.
Figure 5.4.17: Calibration of MOG samples to PAH – (a) cdf table, (b) misclassification probabilities.
Figure 5.4.18: Calibration of metals samples to MOG –
(a) cdf table, (b) misclassification probabilities.
5.4.3 Present Day (Prior) Decision Analysis

This section describes the decision analysis calculations performed using the developed decision and geostatistical models described above, with present day soil sample data only. The primary analysis performed with the information provided by this data is the determination of the near-optimal remediation design, and its corresponding cost, for a range of post-remediation phase activity unit costs and different batch sampling options and unit costs.
5.4.3.1 Contamination level probabilities

The soil contamination distribution inputs for these calculations—probability of contamination above regulatory thresholds and simulated contamination classifications at each block in the decision model for (1) and (2), respectively—are derived from the global geostatistical model developed using the soil sample data, locally updated and conditioned on surrounding hard and soft data using kriging (for probabilities) and simulation (for contamination realizations). These calculations are also performed using (1) hard data only and (2) additional subjective-type C soft data (see Section 4.3.2 for a description of the different types of soft data), for comparison.

First, probability of soil contamination within each of the regulatory classes at each of the 113,400 grid cells (grid dimensions of 220 x 100 x 3), minus cells that are blanked because they are in the harbor, are computed for metals, PAH, and MOG using the M-B simple indicator kriging methodology [Zhu and Journel, 1993], modified for this work. The kriging is performed using the existing soil sample data in three different ways:

1. Only the soil sample data analyzed for the contaminant being estimated is used as conditioning data.
2. In addition, the other two contaminants are treated as soft conditioning data for the contaminant being estimated.
3. In addition, expert opinion about the probability of contamination in certain parts of the site are included, treated as type C data.

Since the size of soil samples analyzed for contamination varies from 0.2 to 3 meters in length and generally less than 25 centimeters in width, and the grid block size is 1 x 1 x 1.5 meters, there is a scale discrepancy issue. It was decided that if a sample of any of these sizes was contaminated, the block containing the sample has to be treated as being contaminated, since the block size is the smallest volume of material that can be excavated. Thus, samples that cross grid block boundaries are divided into separate samples within each of the blocks; the sample location is taken as the midpoint of the sample length within each block. The kriging algorithm was modified to ensure that the estimated probability of contamination for a block is never lower...
than the highest contamination classification of the samples contained within that block (e.g. if a sample in the corner of a grid block has level C metals contamination and the kriging estimate at the middle of the block is 0.9 for level C, the estimate is boosted to 1.0).

The probability of contamination grids for each contamination class, computed using the three conditional kriging scenarios described above, are shown in Figures 5.4.19-27.
Figure 5.4.19a: Probability of metals level B soil contamination—from kriging of all samples analyzed for metals.

Figure 5.4.19b: Probability of metals level C soil contamination—from kriging of all samples analyzed for metals.
Figure 5.4.20a: Probability of metals level B soil contamination—from cokriging of all samples.

Figure 5.4.20b: Probability of metals level C soil contamination—from cokriging of all samples.
Figure 5.4.21a: Probability of metals level B soil contamination—from cokriging of all samples and soft type C prior probabilities.

Figure 5.4.21b: Probability of metals level C soil contamination—from cokriging of all samples and soft type C prior probabilities.
Figure 5.4.22a: Probability of PAH level B soil contamination—from kriging of all samples analyzed for PAH.

Figure 5.4.22b: Probability of PAH level C soil contamination—from kriging of all samples analyzed for PAH.
Figure 5.4.23a: Probability of PAH level B soil contamination—from cokriging of all samples.

Figure 5.4.23b: Probability of PAH level C soil contamination—from cokriging of all samples.
Figure 5.4.24a: Probability of PAH level B soil contamination—from cokriging of all samples and soft type C prior probabilities.

Figure 5.4.24b: Probability of PAH level C soil contamination—from cokriging of all samples and soft type C prior probabilities.
Figure 5.4.25: Probability of MOG special waste soil contamination—from kriging of all samples analyzed for MOG.

Figure 5.4.26: Probability of MOG special waste soil contamination—from cokriging of all samples.
Figure 5.4.27: Probability of MOG special waste soil contamination—from cokriging of all samples and soft type C prior probabilities.

While the impact of using the data from the other two contaminants as soft data (Figures 5.4.20, 23, and 26) is not readily evident from looking at the comparison with the hard data only results (Figures 5.4.19, 22, and 25), it is highlighted by computing a difference grid between the two (Figures 5.4.28, 29, and 30). The soft soil sample data provides information in areas where there is no hard data information. The number of additional data points provided by the soft sample data for each contaminant being estimated is shown in Table 5.4.4. The increase in the number of data available for local updating of the IRV is especially prevalent for metals (see red numbers in Table 5.4.4).
### Table 5.4.4: Number of soil samples used as hard and soft data for cokriging estimation.

<table>
<thead>
<tr>
<th>Hard data:</th>
<th>Metals</th>
<th>PAH</th>
<th>MOG</th>
</tr>
</thead>
<tbody>
<tr>
<td># of samples:</td>
<td>33</td>
<td>94</td>
<td>172</td>
</tr>
</tbody>
</table>

| Soft data:     | Metals - # of soft samples | NA | 21 | 16 |
|                | PAH - # of soft samples    | 92 | NA | NA |
|                | MOG - # of soft samples    | 156| 91 | 81 |

**Figure 5.4.28a:** Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—metals level B.
Figure 5.4.28b: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—metals level C.

Figure 5.4.29a: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—PAH level B.
Figure 5.4.29b: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—PAH level C.

Figure 5.4.30: Difference between probabilities estimated from all sample data versus only samples analyzed for contaminant being estimated—MOG SW.
The soft type C data consists of assigned prior local probabilities of soil contamination greater than the regulatory threshold levels—a cdf or pdf assigned to particular locations/areas, replacing the global expected values at those locations. These local probability values are treated just like the global expected value in the geostatistical estimation (see Alabert [1987]). This type of soft information is usually derived from expert judgment. For this study a hypothetical soft C data set is constructed, reflecting a hypothetical expert’s opinion that the likelihood of any type of soil contamination at the side boundaries of the site is negligible, as well as at any depth below 8 meters for all contaminants except MOG. In addition, it is of the (hypothetical) opinion that the whole western half of the site has a very low likelihood of metals contamination. All these zones are assigned zero soft C probabilities of contamination for all classification levels of the respective contaminant(s). Figures 5.4.31, 32, and 33 show the assigned zones in the decision model grid for the three contaminant types—metals, PAH, and MOG, respectively.

Figure 5.4.31: Assigned prior probabilities for soft type C data scenario—metals.
Figure 5.4.32: Assigned prior probabilities for soft type C data scenario—PAH.

Figure 5.4.33: Assigned prior probabilities for soft type C data scenario—MOG.
Chapter 5. CASE STUDY

The effect of the soft C data on the estimated probability of contamination grids is very evident in Figures 5.4.21, 24, and 27; the difference between these probability grids and the hard data only ones is shown in Figures 5.4.34-36. The IRV expected values at the locations where the prior probabilities are assigned retain those probabilities, unless there is other nearby hard or soft type A or B data to locally update them with; the same is true of the global expected values at locations distant from conditioning data. The overall expected values for the model change after estimation because of biasing from the assigned prior probabilities. Indeed, by applying a spatially non-uniform prior mean to the model, the assumption of stationarity in the expected value is broken, and the model becomes non-stationary. As would be expected, the overall mean of the estimated expected value grid is different than the prior inferred global expected value. These values, for the five different contamination classifications, are itemized in Table 5.4.5, located later in this section.

Figure 5.4.34a: Differences greater than 0.1 between probabilities estimated from all sample data and soft C data versus only hard samples—metals level B.
Figure 5.4.34b: Differences greater than 0.02 between probabilities estimated from all sample data and soft C data versus only hard samples—metals level C.

Figure 5.4.35a: Differences greater than 0.1 between probabilities estimated from all sample data and soft C data versus only hard samples—PAH level B.
Figure 5.4.35b: Differences greater than 0.02 between probabilities estimated from all sample data and soft C data versus only hard samples—PAH level C.

Figure 5.4.36: Differences greater than 0.02 between probabilities estimated from all sample data and soft C data versus only hard samples—MOG special waste.
Another way to compare these kriging results is to look at the difference in their variances. As mentioned earlier, the variance of an IRV is a simple function of the expected value of the IRV, and ranges between zero (when the expected value is one or zero) and 0.25 (when the expected value is 0.5). The set of PAH level B probability of contamination variances for the three results, where the variance is calculated at each grid node, are shown in Figures 5.4.37, 38, and 39. In general, the variance at soft data locations is reduced compared to the hard data only results. The soft data can also increase the variance, though, if the global expected value of contamination is very low (or high) and the soft data indicates the likelihood (or unlikelihood) of contamination, depending on the quality of the soft data.

Figure 5.4.37: Variance of PAH level B probabilities estimated from kriging of soil sample analyzed for PAH.
Figure 5.4.38: Variance of PAH level B probabilities estimated from cokriging of all soil sample data.

Figure 5.4.39: Variance of PAH level B probabilities estimated from cokriging of all soil sample data and type C soft data prior probabilities.
Indicator realizations of the contaminant class IRVs are also generated with the Markov-Bayes conditional simulation algorithm [Zhu and Journel, 1993] for the three data sets; these are required for the data worth analysis components of the decision model. Twenty-five realization grids are generated for each contaminant, wherein each grid cell has a binary number representing which regulatory contamination class that cell belongs in. The expected value of the suite of realizations should converge on the expected value field generated from kriging as the number of realizations increases. The soft sample data case realization expected value grids for all the five contamination classes, computed by taking the simple average of the 25 realization values at each grid cell, are shown in Figures 5.4.40-42. The PAH level B realization expected value for the soft C data case is also displayed in Figure 5.4.43. While the realization expected value fields are not nearly as smooth as the kriging-generated ones, the general features and major anomalies in the field are captured by the realizations.

Figure 5.4.40a: Metals level B soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data.
Figure 5.4.40b: Metals level C soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data.

Figure 5.4.41a: PAH level B soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data.
Figure 5.4.41b: PAH level C soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data.

Figure 5.4.42: MOG special waste soil contamination expected value of 25 realizations—simulated using all sample data as conditioning data.
Figure 5.4.43: PAH level B soil contamination expected value of 25 realizations—simulated using all sample data and soft C prior probabilities as conditioning data.

The average probabilities of contamination within the regulatory classes across the entire problem domain, for each of the three data scenarios (denoted hard, soft A for the soft sample data, and soft C) and for both geostatistical estimation modes (kriging and simulation), are itemized in Table 5.4.5. The comparison is very good for both kriging and simulation, except for the kriging performed with soft type C prior probabilities, wherein the computed averages are less than the global expected values; as discussed above, this is expected since the assigned prior probabilities are less than the global expected values.
### Table 5.4.5: Comparison of site-wide contaminant level expected values of kriging and simulation grid results for 3 conditioning data scenarios: (1) hard sample data, (2) all sample data, and (3) all sample data and soft C prior probabilities.

<table>
<thead>
<tr>
<th>Contaminant type and regulatory threshold level</th>
<th>Expected value of contamination below threshold level</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>Inferred From kriged grids From simulated realizations</td>
</tr>
<tr>
<td></td>
<td>Hard Soft A Soft C Hard Soft A Soft C</td>
</tr>
<tr>
<td>Metals B</td>
<td>0.931 0.931 0.931 0.979 0.931 0.932 0.978</td>
</tr>
<tr>
<td>Metals C</td>
<td>0.974 0.974 0.974 0.987 0.974 0.974 0.985</td>
</tr>
<tr>
<td>PAH B</td>
<td>0.628 0.622 0.658 0.745 0.621 0.648 0.717</td>
</tr>
<tr>
<td>PAH C</td>
<td>0.950 0.949 0.950 0.962 0.949 0.949 0.956</td>
</tr>
<tr>
<td>MOG SW</td>
<td>0.965 0.966 0.966 0.972 0.965 0.965 0.971</td>
</tr>
</tbody>
</table>

#### 5.4.3.2 Remediation design

Once probability of contamination grids for all the applicable regulatory levels have been estimated from existing site knowledge (primarily soil sample data), the prior decision analysis can be performed—determining the near-optimal remediation design and its associated costs, based on defined unit costs for the pertinent remediation activities. As discussed earlier in this chapter, the decision model requires that the contamination levels be prioritized and mutually exclusive contamination probabilities determined. A very useful way of understanding these probabilities and how they can be calculated is to use Venn diagrams. Figure 5.4.44 is a Venn diagram showing the probability space for three contaminant classifications—metals level B, metals level C, and MOG special waste—that are depicted as circles. The area of the entire box represents the probability of any possible outcome, that being one. The proportion of the total box area enclosed by a circle is the overall probability of that contamination classification, regardless of whether there is also other types of contamination, while the proportional area
enclosed by the intersection of circles is the probability of contamination in all the classes represented by those circles.

![Figure 5.4.44: Venn diagram of hypothetical metals and MOG contaminant level sets.](image)

Assuming metals level C has highest priority, MOG special waste next highest, and metals level B the lowest, we need the following mutually exclusive probabilities:

- Probability of metals C contamination, or mathematically represented as: \( P[\text{Met}_C] \)
- Probability of MOG special waste contamination, but not metals C: \( P[\text{MOG}_{sw} \notin \text{Met}_C] \)
- Probability of metals B contamination, but not metals C or MOG special waste: 
  \[ P[\text{Met}_B \notin (\text{Met}_C \cup \text{MOG}_{sw})] \]

\( P[\text{Met}_C] \) is simply the area of the red circle, proportional to the area of the box. 
\( P[\text{MOG}_{sw} \notin \text{Met}_C] \) is the portion of the blue circle area that is not intersected by the red circle, proportional to the box area. It can be represented mathematically as:

\[
P[\text{MOG}_{sw} \notin \text{Met}_C] = P[\text{MOG}_{sw}] - P[\text{MOG}_{sw} \cap \text{Met}_C] = P[\text{MOG}_{sw}] - P[\text{Met}_C](P[\text{MOG}_{sw} | \text{Met}_C])
\]

(5.6)

The last term in the expression is derived from the probability of the intersection between the contaminant classes using Bayes' Theorem. If MOG contamination is independent of metals contamination, \( P[\text{MOG}_{sw} | \text{Met}_C] = P[\text{MOG}_{sw}] \) and the equation simplifies to:

\[
P[\text{MOG}_{sw} \notin \text{Met}_C] = P[\text{MOG}_{sw}](1 - P[\text{Met}_C])
\]

(5.7)
Finally, $P[\text{Met}_B \not\in (\text{Met}_C \cup \text{MOG}_{sw})]$ is the portion of the green circle area that is not intersected by the red or blue circles. It can be represented mathematically as:

$$P[\text{Met}_B \not\in (\text{Met}_C \cup \text{MOG}_{sw})] = P[\text{Met}_B] - \{P[\text{MOG}_{sw}](P[\text{Met}_B | \text{MOG}_{sw}]) - P[\text{Met}_B \cap \text{Met}_C \cap \text{MOG}_{sw}] \} \quad (5.8)$$

The last intersection term can be reformulated in many ways, including:

$$P[\text{Met}_B \cap \text{Met}_C \cap \text{MOG}_{sw}] = P[\text{Met}_B \cap \text{Met}_C \cap \text{MOG}_{sw}]$$

In this case, equation (5.8) can be greatly simplified because, by definition, level B and C metals contamination cannot occur simultaneously, zeroing out all terms of the equation containing an intersection between these two contamination classes. Equation (5.8) then simplifies to:

$$P[\text{Met}_B \not\in (\text{Met}_C \cup \text{MOG}_{sw})] = P[\text{Met}_B] - P[\text{MOG}_{sw}](P[\text{Met}_B | \text{MOG}_{sw}]) \quad (5.9)$$

The decision model is formulated to calculate the mutually exclusive contamination level probabilities at each grid cell using equations like those above, and can incorporate (1) user defined conditional probabilities, as a constant or defined locally at the grid scale, between contamination classes of different contaminants (conditional probabilities between classes of the same contaminant are always zero) or (2) assume contamination classes are independent from each other. To keep the decision model simple, independence between contamination classes was assumed for this study.

In addition to estimating the probabilities in several different ways, other model parameters are parametrically varied in the decision model to account for their uncertainty. Other than the probabilities, the decision variables with the greatest uncertainty are the future unit remediation activity costs—the costs for performing remediation after completion of the remediation phase of development, due to the underclassification of contamination. These costs are very important to the decision model since they determine the magnitude of the risk costs. To account for this uncertainty the ratio of future to present costs is treated as a parametric variable and the decision analysis is performed for a discrete range of these ratios between 1.0 to 3.0.

Another decision variable that is varied is the unit cost for batch sampling, and whether batch sampling is considered an alternative at all. The unit cost estimate for batch sampling that
was provided ($4 / cubic meter) seems low, especially considering the same cost was given for excavation. Thus, the decision analysis was also performed with a batch sampling cost of $8 / cubic meter. A scenario wherein batch sampling is not an alternative is also evaluated with the decision model.

The only other parameter that is varied in the decision analysis is the constraint on the maximum excavation slope angle, or bench width between excavation layers. The primary analysis is performed assuming a one meter (one block) wide bench is required between excavation layers (e.g. if a block has to be excavated some depth below the surface, then eight blocks have to be excavated the layer above to reach it). Some of the analyses are also performed with the assumption that vertical excavation is permissible (e.g. if one block has to be excavated at depth, only the blocks directly above it have to be excavated to reach it).

The total cost of remediation of the near optimal remediation design results of the prior decision analysis for the three batch sampling options (no batch sampling, $4, and $8 / cubic meter batch sampling cost), across a range of future to present remediation cost ratios, are displayed graphically in Figures 5.4.45, 46, and 47 for hard data only, soft sample data in addition, and soft type C data in addition cases, respectively. The cost of remediation is seen to be substantial—in the range of $4-9 million. For the no batch sampling scenario, the remediation cost increases linearly with increasing future cost of remediation, except when the future cost is close to the present day cost, wherein the total cost increases more rapidly. For the scenario where batch sampling is an alternative, the total remediation cost increases from future-to-present cost ratios of 1.0 to 1.5, wherein it reaches a plateau for increasing ratios. The double batch sampling cost scenario has the same pattern as the batch sampling case, but reaches a plateau that is about $0.5 million higher. The double batch remediation cost values are the same as the no batch case for ratios of 1.0-1.5; the batch sampling costs diverge from the no batch costs at much lower ratios.

Also plotted in these figures is the estimated total remediation cost if perfect information about contamination exists (denoted EVPI). It is the same for the all batch sampling cases and all
future remediation cost scenarios because, if perfect information is available, there is no need for sampling and there are no risk costs. These values are computed from the 25 contamination level realizations for each contaminant by determining the remediation design and cost for each realization and taking the average across all 25 to estimate the expected value cost. The EVPI remediation costs should be the same as the total remediation cost calculated from the prior analysis for a future-to-present remediation cost ratio of 1.0, where all batch sampling scenarios converge.

The following is an explanation of why these costs should be the same. When the future-present remediation cost ratio is 1.0 in the prior analysis, almost all the required remedial actions are delayed until the future and, thus, treated as a risk cost. This is because in the future, as the site is being excavated for development, it is assumed that, if there is contamination, it will be directly identified and remediated. Thus, there is no risk of misclassification, as there is if the soil is remediated in the present, during the remediation phase, making the delay in remedial action cheaper (when the future cost is no more expensive). As explained above in the description of the decision model, the risk cost is simply the probability of contamination multiplied by the cost of future remediation for all contamination classes—the expected value of the future cost. Since, in this case, all remedial activities are being delayed to a future time when the extent of contamination becomes known, the total cost of remediation calculated using the prior analysis corresponds to the expected value cost when there is perfect information about contamination across the site.

Thus, a useful way of evaluating the validity of the overall decision and geostatistical models is to compare the total remediation cost results from the EVPI analysis and prior analysis for a future-present remediation cost of 1.0; these costs should be the same for all batch sampling options. As illustrated in Figures 5.4.45-47, the EVPI and prior 1.0 ratio costs match very closely for the hard data only and soft sample data scenarios; however, there is a discrepancy for the soft C data scenario. The discrepancy is attributed to a problem in some of the PAH contaminant simulated realizations, wherein average probabilities of contamination are considerably higher.
than the other realizations and the average expected values of the kriging results; this is highlighted in the comparison between the IRV spatial field expected values computed from simulation and kriging (see Table 5.4.5).

Figure 5.4.45: Optimized total cost of remediation for data scenario where only hard sample data is used in geostatistical estimation.
Figure 5.4.46: Optimized total cost of remediation for data scenario where all sample data is used in geostatistical estimation.

Figure 5.4.47: Optimized total cost of remediation for data scenario where all sample data and soft C prior probabilities are used in geostatistical estimation.
Chapter 5. CASE STUDY

The comparison between the different input data scenarios shows that the hard data only and soft sample data cases have very similar total costs for all three batch sampling scenarios, and across all future-present cost ratios. The difference between the hard data only and soft data remediation costs are plotted in Figure 5.4.48; these values represent the value of the information provided by the soft sample data. These values average about $70,000 and do not reach above $120,000. While these values are not substantial, especially when considered in the context of the total remediation costs, it should be remembered the extra sample data already exists, so the cost incurred to glean this extra information is negligible.

The difference in total costs between the hard data only and soft C data scenarios (the value of the soft C information) is very substantial—$1.9-2.9 million—as illustrated in Figure 5.4.49. As elucidated above in the comparison between the probability grid estimates (Table 5.4.5), the overall mean probability of contamination drops significantly when the soft C data is incorporated in the kriging estimation, and the variance also drops considerably. These are probably good indicators that the total remediation cost of the optimized design will be considerably lower, as it is. While these results suggest this type of information can be very valuable, they also illustrate how misleading it can be if based on false presumptions. This type of information—subjective expert opinion—should be treated very carefully in decision making, since there is usually no hard data to verify its accuracy.
Figure 5.4.48: Value of information (cost savings in remediation design) provided by using soft sample data instead of just hard sample data.

Figure 5.4.49: Value of information (cost savings in remediation design) provided by using soft sample data and soft C prior probabilities instead of just hard sample data.
Instructive examples of the optimized remediation designs for the different scenarios are provided in Figures 5.40.50-56. The remediation actions are color-coded:

- Blue is no action (NA)
- Light blue is confirmatory sampling (CS)
- Light green is excavation and batch sampling (BS)
- Darker green is excavation only (E)
- Light orange is excavation, treatment and disposal for level B contamination (TB)
- Red is excavation, treatment and disposal for level C contamination (TC).

Figure 5.4.50a is the near optimal design for the scenario where (1) hard data only is employed, (2) batch testing is not an alternative, and (3) the cost of remedial activities in the future are twice what they are at the present. Almost every block of soil on the site has to be excavated, most requiring level B treatment and disposal. Figure 5.4.50b displays only the blocks that do not have to be treated. Most of these are in zones near the waterfront and in the areas where most of the soil samples were taken from.

Figure 5.4.50c is the design for the scenario having the same parameters as Figure 5.4.50a, but where soft sample data is included in the analysis. The area on the bottom layer where confirmatory sampling is chosen has expanded and the number of blocks not requiring treatment is slightly greater than the hard data only scenario, as shown in Figure 5.4.50d, where the blocks requiring treatment are made invisible.

The same scenario, except soft C data is included, is displayed in Figure 5.4.50e. The results are significantly different than the other two cases in that many fewer blocks have to be treated for level B contamination, especially on the boundaries of the site; this is a logical result since prior probabilities of contamination are set to zero on the boundaries. Even in the center of the site, a much larger number of blocks do not require treatment, only excavation; this is clearly visible in Figure 5.4.50f.

The remediation design for the scenario where batch sampling at $4 per cubic meter is an alternative and the future remediation costs are still twice as costly as those today is displayed in
Figure 5.4.51 for the soft sample data and soft C data cases. By far the predominant action is excavation, batch sampling, and treatment accordingly. The soft C data case has more outright treatment, excavation only, and no excavation blocks, replacing batch sampling blocks.

The soft sample data and soft type C data results for the scenario where batch sampling is $8 per cubic meter, with the same future-to-present cost ratio are exhibited in Figures 5.4.52a and 5.4.52b, respectively. With the higher batch sampling cost, fewer blocks are designated for batch sampling, but it is still the predominant action. Many of the blocks are to be excavated only, especially in the soft C data case. All the blocks that are not designated for batch sampling in the soft sample data design are displayed in Figure 5.4.52c.

Figure 5.4.53 shows the no batch sampling, soft sample data and soft C data cases where unit future remedial action costs are 1.25 times greater than present ones. A much larger potion of the site does not require outright level B treatment, in comparison to the 2.0 future-to-present cost scenario, instead requiring excavation only or no excavation at all (confirmatory sampling or no action). This is especially true for the soft C data case, where some of the surface now doesn’t require excavation.

The same results for a future-to-present remediation unit cost ratio of only 1.125 are presented in Figure 5.4.54. Even with this small ratio most of the site still requires remediation activities today in the soft sample data design, although most of the work is excavation to reach widely scattered blocks that are to be treated for level B contamination. In contrast, only a few hot spots require action in the soft C case, the rest of the site left alone—no action required.

The soft sample data scenario wherein batch sampling is an alternative and the future remediation activity unit cost is three times greater than present day costs is identical to the same scenario wherein future-to-present cost ratio is 2.0, as illustrated in Figure 5.4.55.

Finally, Figure 5.4.56 shows the soft sample data, no batch sampling scenario where the unit future remediation costs are the same as present day costs. Most of the site is left alone, except for a small zone in the western part of the site. This result is expected, as explained above.
Figure 5.4.50a: Optimized remediation design based on hard sample data only—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0.

Figure 5.4.50b: Portion of site not requiring contamination treatment for remediation design in Figure 5.4.50a.
Figure 5.4.50c: Optimized remediation design based on all sample data—for same scenario in Figure 5.4.50a.

Figure 5.4.50d: Portion of site not requiring contamination treatment for remediation design in Figure 5.4.50c.
Figure 5.4.50e: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.50a.

Figure 5.4.50f: Portion of site soil volume not requiring treatment for contamination—remediation design in Figure 5.4.50e.
Figure 5.4.51a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is an alternative at $4 / m^3$, (2) ratio of future to present unit remediation cost is 2.0.

Figure 5.4.51b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.51a.
Figure 5.4.52a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is an alternative at $8 / \text{m}^3$, (2) ratio of future to present unit remediation cost is 2.0.

Figure 5.4.52b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.52a.
Figure 5.4.52c: Portion of site not designated for batch sampling—for remediation design in Figure 5.4.52a.

Figure 5.4.53a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.25.
Figure 5.4.53b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.53a.

Figure 5.4.54a: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.125.
Figure 5.4.54b: Optimized remediation design based on all sample data and soft C prior probabilities—for same scenario in Figure 5.4.54a.

Figure 5.4.55: Optimized remediation design based on all sample data—for scenario where (1) batch sampling is an alternative at $8 / m^3$, (2) ratio of future to present unit remediation cost is 3.0.
5.4.4 Future Site Characterization Data Worth (Preposterior) Analysis

Once the prior decision analysis for the case study problem is completed, determining the near optimal soil remediation design and its associated costs from existing data (soil sample chemical analysis), the next step in the analysis is to determine the economic value, or data worth, of additional site characterization information that can be feasibly acquired. The caliber of interest is an estimate of the cost savings that can be ascertained by collecting this additional information, the savings incurred by developing a more cost-effective remediation design as a result of having better information about the site. The focus in this study is to evaluate the worth of soft data, particularly geophysics, of different data quality levels. The case study centers on applying EM and GPR to the site, so the data worth analysis is geared towards evaluating the possible applicability of these non-invasive geophysics techniques, in comparison to the hard data provided by invasive soil sampling.
Chapter 5. CASE STUDY

The data worth analysis is performed for three types of information:

- Perfect information
- Borehole sample data
- Geophysics data of different data qualities.

The purpose of the worth of perfect knowledge about the site is to determine the absolute maximum budget that should be spent on site characterization; spending more money than this would cost more than could possibly be saved by reducing the remediation design cost from the extra knowledge. This perfect information worth also provides a normalized scale for evaluating the worth of real information, how it compares to the maximum amount of cost reduction possible. The purpose of the worth of additional borehole sampling and worth of geophysics is to assess whether proposed site investigation programs using these techniques will be worthwhile in the overall economic framework of the project.

It is important to understand that these calculations are only estimates of the cost savings provided by the additional information, based on the existing geostatistical model for the site. The geostatistical model accounts for uncertainty about site conditions and, thus, the calculations are expected value estimates. Thus, they are commonly referred to as the expected value of perfect information (EVPI) and the expected value of sample information (EVSI), the latter referring to any type of information that is provides anything less than perfect information about the site (e.g. the borehole samples, geophysics measurements). If the assumptions in the geostatistical model are inaccurate, then EVPI and EVSI are not even accurate expected value estimates. This being said, these analyses still provide very useful insights about the potential value of site characterization techniques, and a means for assessing the relative value of different techniques, applied at different site locations, all within a rigorous uncertainty framework. Within this framework, the expected value of the cost savings is only one of many statistical metrics that can be analyzed; theoretically, the whole probability distribution of the predicted cost savings for the whole range of possible site realizations can be constructed.

For all information types, the data worth is evaluated by:
Chapter 5. CASE STUDY

1. simulating a set of measurement outcomes, as they relate to the primary variable of interest—contamination level for metals, PAH, and/or MOG;

2. updating the contamination probability fields based on the information provided by each outcome, for all outcomes, by performing geostatistical kriging with the outcome data as conditioning data, in addition to the existing soil sample data;

3. determining the near-optimal remediation design and associated total cost for each set of updated contaminant probability fields for all outcomes; and

4. estimating the expected value of the total remediation cost with the additional information.

The difference between the existing information (prior) total remediation cost and the expected value cost with the new information is the expected value, or data worth, of the new information.

Figures 5.4.57a-d and 5.4.58a-d are plots summarizing the results of the data worth analyses for the different data types. They are explained in the following sections, which describe the analysis for each of the data types.

**Figure 5.4.57a:** Expected value total remediation cost with additional hypothetical data of different types, including (1) EM geophysics of different data qualities for detecting metals level C, (2) 21 fully-sampled boreholes, and (3) perfect information—for scenario where batch sampling is not an alternative.
Figure 5.4.57b: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.57a.

Figure 5.4.57c: Expected value total remediation cost with additional hypothetical data—for scenario where $4/m³ batch sampling is an alternative and same data types as Figure 5.4.57a.
Chapter 5. CASE STUDY

**Figure 5.4.57d:** Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.57c.

**Figure 5.4.58a:** Expected value total remediation cost with additional hypothetical data of different types, including (1) EM geophysics of different data qualities for detecting MOG SW, (2) 21 fully-sampled boreholes, and (3) perfect information—for scenario where batch sampling is not an alternative.
Figure 5.4.58b: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.58a.

Figure 5.4.58c: Expected value total remediation cost with additional hypothetical data—for scenario where \$4/m^3\ batch sampling is an alternative and same data types as Figure 5.4.58a.
Chapter 5. CASE STUDY

5.4.4.1 Perfect information

As briefly discussed in Section 5.4.3, the worth of having perfect information about the distribution and contamination level of all contaminants across the site (EVPI) is estimated by (1) generating a ensemble of 25 contamination level realizations for the three contaminants with the M-B IRV simulation routine, (2) calculating the total cost for remediating each of the site realizations (there is no uncertainty in this calculation because the contamination levels of every block are known), and (3) computing the expected value of the total remediation cost by taking the simple average of the costs for all realizations. In the EVPI case, the step for updating the probability fields based on the sample outcomes is skipped because, with perfect information, the realizations replace the probability fields. An example of the remediation design for one of the perfect information realizations, for the scenario where existing soft sample data is incorporated and the scenario where soft type C data is also included, is shown in Figure 5.4.59.

Figure 5.4.58d: Economic value (cost savings in remediation design) of additional hypothetical data—for same scenario and data types as in Figure 5.4.58c.
Figure 5.4.59a: Remediation design based on scenario where there is perfect information about all contaminant levels—for a simulated realization conditioned on all sample data.

Figure 5.4.59b: Remediation design based on scenario where there is perfect information about all contaminant levels—for a simulated realization conditioned on all sample data and soft C prior probabilities.
5.4.4.2 Borehole sampling

The approach used for calculating the EVSI of borehole samples uses the ensemble of 25 contamination level realizations to provide a set of sample outcomes at proposed borehole locations for each realization. Only the contaminant realization values at the borehole locations are retained; the rest of the grid is left unused. Probability maps for each contaminant are generated with M-B IRV kriging—25 sets, each conditioned on a different borehole sample outcome. A decision analysis is performed for each of these contaminant level probability sets to determine the near optimal remediation design and total remediation cost, from which the expected value total remediation cost is computed. Since the sample outcomes are generated from equally-likely realizations, the expected value can be computed by simply taking the simple average of the results.

A set of 21 borehole locations was chosen for the hard soil sample data EVSI analysis; their locations within the decision model grid are displayed in Figure 5.4.60. The borehole locations were chosen so as to be distributed across the site, but not in zones where extensive soil samples already exist. The boreholes are designed to be fully-penetrating to a depth of nine meters below the surface. The borehole sampling data worth analysis is performed for the scenarios where (1) batch sampling is not an alternative, (2) batch sampling is an alternative, at a unit cost of $4 per cubic meter, and (3) batch sampling is an alternative, at a unit cost of $8 per cubic meter; all existing hard and soft sample data are used as conditioning data, with the hypothetical new borehole sample data.
The influence of the borehole sampling on the kriged probability maps is illustrated in Figures 5.4.61-63, example probability maps from one borehole sampling outcome for the 5 contamination levels (3 contaminants). The contamination level probability that the borehole sampling has the greatest influence on across the site is PAH level B, since this contaminant level IRV has the longest spatial correlation length. In general, though, the 21 borehole program does not appear to provide much information about the overall probability of contamination, due to limited overall volume of material sampled by the boreholes and the generally short correlation length of the contaminant level IRVs.

The no batch sampling and $4 / cubic meter batch sampling scenario optimized remediation designs, determined for one of the borehole sampling outcomes, at a future-to-present unit remediation cost ratios of 2.0, is exhibited in Figure 5.4.64. While the no batch sampling remediation design is considerably different than the prior analysis design (Figure 5.4.50a), in that significantly less of the site is remediated for level B contamination. However,
essentially the entire site still is slated for excavation. There is very little from the prior analysis remediation design in the batch sampling case, where almost all the blocks are to be batch sampled.

The expected value total remediation cost for the simulated borehole sampling program, across a range of future-to-present unit remediation cost ratios, is graphically displayed in Figures 5.4.57b and 5.4.57c for no batch sampling and $4 / cubic meter batch sampling, respectively. These costs are so close to the prior analysis remediation total cost, the two are hard to distinguish; the addition of the hard sample data from the 21 borehole layout has a negligible impact on the overall remediation design cost. The borehole sampling EVSI, computed by subtracting the borehole sampling expected value total remediation cost from the prior total cost, is plotted in Figures 5.4.57b and -57d for the no batch sampling and $4 / cubic meter batch sampling scenarios, respectively, and a range of future-present cost ratios from 1 to 3. The EVSI is negligible for all scenarios, the highest being about $200,000 for the scenario where batch sampling is not an option and the future remediation unit costs are twice those for the present.

Figure 5.4.61a: Probability of metals level B soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.
Figure 5.4.61b: Probability of metals level C soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.62a: Probability of PAH level B soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.
Figure 5.4.62b: Probability of PAH level C soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.

Figure 5.4.63: Probability of MOG special waste soil contamination—estimated from cokriging of hypothetical sample outcomes in 21 boreholes and existing hard and soft samples.
**Figure 5.4.64a:** Optimized remediation design based on simulated samples in 21 boreholes and all existing sample data—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0.

**Figure 5.4.64b:** Optimized remediation design based on simulated samples in 21 boreholes and all existing sample data—for scenario where (1) $4/m^3$ batch sampling is an alternative, (2) ratio of future to present unit remediation cost is 2.0.
5.4.4.3 Geophysics

As with the perfect information and borehole sampling analyses, the data worth analysis for geophysics uses 25 contaminant level realizations for each contaminant as a starting point. Unlike the borehole sampling approach, the realizations need to be transformed to reflect the uncertainty in the geophysics indirect measurement of contamination. The quality of the geophysics is defined by the M-B soft calibration, more specifically the cdf table and misclassification probabilities calculated from the calibration. A hypothetical calibration needs to be performed to generate these calibration probabilities. In addition, realistic geophysics outcomes for the contaminant level realizations need to be constructed. The approach used in this study to simulate geophysics results that can be used as conditioning data for kriging is to: (1) build a hypothetical geophysics calibration data set and perform a calibration to ascertain the cdf and misclassification probabilities, (2) transform the contaminant realizations into simulated geophysics result realizations using the calibration results, and (3) incorporate the simulated geophysics realizations as soft data in conditional kriging to update the applicable contaminant level probabilities, using the calibration results calculated in step (1).

Building a hypothetical geophysics calibration data set is accomplished by, first, deciding on a geophysics data quality level for each of the regulatory levels of the contaminant the geophysics provides information about. The quality level is defined as the percentage of geophysics measurements that can correctly identify the true contaminant level. For example, if 10 geophysics measurements are made in soil with PAH contamination above level C, 8 correctly identify the contamination above level C; the geophysics has a 80% success rate in identifying levels C contamination. The hypothetical calibration data set for this example could consist of any number of hard-soft data pairs where half of the hard data are above level C, half below. Of the hard level C data, 80% of them have soft data of one value (say 1), the rest of another value (say 0). Of the hard data below level C, 20% have soft data of value 1 and 80% have soft data of value 0. The total number of data pairs used for the hypothetical calibration data set is irrelevant, as long as the relative number of different pairings is in the correct proportion to
yield the proper percentages. To replicate the inability of the geophysics to provide any information about one of the contaminant levels, the calibration data pairs for that level can be proportioned such that the soft data results are split 50% for one of the soft data values (e.g. 1) and 50% for the other (e.g. 0). To clarify, a possible hypothetical calibration data set for the above example is provided in Table 5.4.6.

<table>
<thead>
<tr>
<th>Hard data</th>
<th>Soft data</th>
<th>Hard data</th>
<th>Soft data</th>
<th>Hard data</th>
<th>Soft data</th>
</tr>
</thead>
<tbody>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>1</td>
<td>1</td>
<td>1</td>
<td>2</td>
<td>0</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
<tr>
<td>0</td>
<td>0</td>
<td>1</td>
<td>0</td>
<td>2</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.4.6: Hypothetical soft data calibration data set where the soft measurement can delineate between hard values of 2 and less than 2 with 80% success, but has no ability to distinguish between values of 1 and 0.

Once the hypothetical calibration data set is built, a M-B soft data calibration can be performed. The calibration results for the above example data set, a cdf probability table and misclassification probabilities, are displayed as 3-d bar plots in Figure 5.4.65. The cdf probabilities represent the probability PAH contamination is less than the said level, given the said soft geophysics data; one minus the cdf table values is the probability of PAH contamination greater than or equal to the said level. The results show there is about a 70% likelihood that PAH contamination is above the level C threshold if the soft data is in class level one. Conversely,
there is about a 90% likelihood PAH is less than level C if the soft data is in class level zero. These likelihoods differ somewhat from the those set for the soft data due to the influence of data at PAH levels lower than level C; the \( B \) calibration correlation parameter is essentially the same whether PAH level B is included in the calibration or not (see Section 4.3.1.2 for a description of this parameter). Figure 5.4.66 displays ogive plots for the same calibration, except the probabilities have been normalized by the stationary global expected values for PAH contamination at the two regulatory cutoffs, as required for properly incorporating the soft data in the M-B conditional kriging algorithm (see Section 4.3.3 for an explanation).

![Figure 5.4.65: Calibration of hypothetical soft data to PAH levels using constructed data set in Table 5.4.6—(a) cdf table, (b) misclassification probabilities](image-url)
The approach used for transforming the contaminant level realizations into simulated geophysics results is to choose a geophysics soft data outcome, for each contaminant level realization value, according to the calibration probability distribution. This is accomplished by generating a random number between 0 and 1 and choosing the geophysics outcome that corresponds to where the number falls with respect to the calibration cdf for that realization contaminant level. (For this operation, the calibration cdf used should NOT be normalized for the stationary contaminant level expected values.) Using the example described above, if the PAH realization contamination level is C and the random number generated is 0.85, then a soft data value of 1 will be assigned to that location; if the random number is less than 0.8, a soft data value of 0 will be assigned. Examples of simulated geophysics soft data for PAH across the site
are displayed in Figures 5.4.67-70. Figure 5.4.67 show the cases where: (1) the geophysics can identify PAH contamination at level C perfectly, but cannot provide any information about contamination at level B or below, except that it is below level C, (2) it can identify contamination above level B perfectly, but cannot distinguish between levels B and C, respectively. These simulations exactly replicate the PAH realizations at the respective levels, since the data quality is perfect. Figures 5.4.68-70 show the same cases, except where the soft data quality levels drop to 90%, 80%, and 60% likelihood of accurately distinguishing the said contamination levels, respectively. All these geophysics simulations are generated from the same PAH contaminant level realization, enabling the deterioration in the ability of the geophysics to accurately distinguish contamination (or, in particular, to falsely distinguish contamination) as the data quality decreases to be easily discerned.

Figure 5.4.67a: Simulation of geophysics soft data outcome where the measurement can perfectly delineate PAH contamination above and below threshold level C, but provides no information about the level B threshold. One (red) corresponds to the geophysics measurement above the calibration soft cutoff, zero (green) below, zero (green) below, -1 (blue) outside problem domain.
Figure 5.4.67b: Simulation of geophysics soft data outcome where the measurement can perfectly delineate PAH contamination above and below threshold level B, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a.

Figure 5.4.68a: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level C with 90% success, but provides no information about the level B threshold. Grid values defined same as in Figure 5.4.67a.
Figure 5.4.68b: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level B with 90% success, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a.

Figure 5.4.69a: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level C with 80% success, but provides no information about the level B threshold. Grid values defined same as in Figure 5.4.67a.
Figure 5.4.69b: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level B with 80% success, but provides no information about the level C threshold. Grid values defined same as in Figure 5.4.67a.

Figure 5.4.70a: Simulation of geophysics soft data outcome where the measurement can delineate PAH contamination above and below threshold level C with 60% success, but provides no information about the level B threshold. Grid values defined same as in Figure 5.4.67a.
Once the 25 simulated geophysics realizations are generated, the realizations are used to conditionally update the contamination level probability fields for the contaminant the geophysics is sensitive to. Examples of updated PAH contamination level probabilities for geophysics with data qualities of 100%, 90%, and 70% likelihood of distinguishing level C contamination are exhibited in Figures 5.4.71, -72, and -73, respectively. At the highest data quality (perfect identification of level C contamination), level C probabilities are either 1.0 or 0.0, except below six meters, considered to be beyond the depth of investigation of the geophysics measurement; the geophysics from above still influences the probabilities below 6 meters, due to spatial correlation. While the geophysics provides no direct information about level B contamination, where the geophysics indicates there is a higher probability of level C contamination, there is a lower probability of level B contamination; in Figure 5.4.71b, blocks having zero probability of level B contamination, corresponding to 100% probability of level C contamination, are scattered across the site. As the data quality decreases, the proportion of the
problem domain indicated by geophysics to have level C contamination increases, but there is progressively less effect on the level C probabilities in these zones.

**Figure 5.4.71a:** Probability of PAH level B soil contamination—estimated from cokriging of all sample data and a hypothetical geophysics realization, where the geophysics can perfectly delineate PAH contamination above and below threshold level C, but provides no information about the level B threshold.
**Chapter 5. CASE STUDY**

Figure 5.4.71b: Probability of PAH level C soil contamination—estimated the same way as in Figure 5.4.71a.

Figure 5.4.72a: Probability of PAH level B soil contamination—estimated from cokriging of all sample data and a hypothetical geophysics realization, where the measurement can delineate PAH contamination above and below threshold level C with 90% success, but provides no information about the level B threshold.
Figure 5.4.72b: Probability of PAH level C soil contamination—estimated the same way as in Figure 5.4.72a.

Figure 5.4.73a: Probability of PAH level B soil contamination—estimated from cokriging of all sample data and a hypothetical geophysics realization, where the measurement can delineate PAH contamination above and below threshold level C with 70% success, but provides no information about the level B threshold.
Examples of near-optimal remediation designs, determined for scenarios of geophysics with different data qualities for measuring level C metals contamination, are displayed in Figures 5.4.74 and -75. Incorporating geophysics with a 100% data quality for correctly identifying level C metals contamination changes the remediation design considerably for the no batch sampling case, by significantly reducing the amount of soil that has to be treated for level B contamination (compare Figure 5.4.74a to Figure 5.4.50c), but still requiring excavation of most of the site. The batch sampling case for the same data quality does not change much; most of the soil is still slated for excavation and batch testing (compare Figure 5.4.74b to Figure 5.4.51c). Reducing the data quality just slightly to a 90% likelihood of correctly identify level C metals contamination results in a substantially different remediation design than the 100% data quality scenario, for the no batch sampling case (Figure 5.4.75a); as with the prior, no geophysics design, most of the site once again requires treatment for level B contamination. The 70% data quality, no batch
sampling scenario (Figure 5.4.75b) is very similar to the 90% scenario, except even more of the site requires level B treatment instead of excavation only.

Figure 5.4.74a: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics can perfectly delineate metals contamination above and below threshold level C, but provides no information about the level B threshold—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0.
Figure 5.4.74b: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics has the same applicability as in 5.4.74a—for scenario where (1) $4/m^3$ batch sampling is an alternative, (2) ratio of future to present unit remediation cost is 2.0.

Figure 5.4.75a: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics can delineate metals contamination above and below threshold level C with 90% success, but provides no information about the level B threshold—for same scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0.
Figure 5.4.75b: Optimized remediation design based on all existing sample data and a simulated geophysics realization, where the geophysics can delineate metals contamination above and below threshold level C with 70% success, but provides no information about the level B threshold—for same scenario as in Figure 5.4.75a.

The overall results of the geophysics data worth analysis are plotted in Figures 5.4.57a-d and 5.4.58a-d, at the beginning of section 5.4.3. Geophysics with poor to excellent data qualities for delineating metals level C and MOG special waste (SW) were investigated, for the two scenarios where batch sampling is not an alternative (Figures 5.4.57a,b and 5.4.58a,b) and where it is with a unit cost of $4 per cubic meter (Figures 5.4.57c,d and 5.4.58c,d). The results are reported in two ways: (1) the expected total cost of remediation, in comparison with the borehole sampling and perfect information expected total remediation cost, as well as the prior total cost (Figures 5.4.57a,c and 5.4.58a,c); (2) the expected value (worth) of information (EVSI), in comparison with those from perfect information and borehole sampling (Figures 5.4.57b,d and 5.4.58b,d). These results DO NOT take into account the data acquisition and analysis costs; these costs should be subtracted from the reported engineering EVSI to obtain the overall EVSI. In these plots, the qualitative naming of geophysics data quality as “excellent” or “very good”, “good”, “medium”, and “poor” quality correspond to the quantitative data qualities of 100%,...
90%, 80%, and 60% likelihood of identifying the respective contamination levels, respectively. While the plots are labeled as being for EM (electromagnetics) geophysics, they apply equally to GPR of the same data quality levels for detecting metals and MOG.

The results show that only at a 100% data quality level does geophysics that is sensitive to metals level C or MOG SW have any considerably worth in reducing overall remediation costs. In general, geophysics is more worthwhile if it can detect the MOG SW than the metals level C; its highest data worth for detecting MOG is $850,000 (no batch testing) and $500,000 (batch testing), whereas for metals it is $650,000 (no batch testing) and $0 (batch testing). While the highest data worth occurs at the highest future-to-present unit remediation cost ratio (3.0), there is only a small difference for a ratio of 2.0; at a ratio of 1.0 the worth is zero, as is the EVPI.

With a 90% data quality level, the geophysics for detecting MOG SW does have a data worth of $500,000 in the no batch sampling case, but for all other scenarios and data quality levels the data worth is negligible.

5.4.4.4 Summary

In summary, the data worth analysis indicates that neither realistic additional borehole sampling (on the order of 10s of holes) nor geophysics site characterization programs will provide any significant economic worth (cost savings) to the overall project, especially if batch sampling is an alternative. For geophysics that can delineate metals level C or MOG special waste contamination to be worthwhile, the data quality has to be very good—unrealistically good, considering the difficult site conditions for EM and GPR and the difficulty in applying geophysics successfully for the desired targets, even when site conditions are ideal.

5.4.5 Decision Analysis With Geophysics (Posterior Analysis)

The frequency-domain electromagnetics (FDEM) and the 200 megahertz ground penetrating radar (GPR) data collected in the production surveys are incorporated in the decision analysis through a number of sequential steps:
1. The raw data is processed, rectified to the case study decision model grid, and transformed into a format that can be input into the geostatistical and decision models.

2. The processed data that is collocated with available hard sample data is calibrated for detecting metals and MOG contamination above level B and special waste level, respectively.

3. The fully processed data is incorporated in the geostatistical model through conditional kriging to produce updated probability fields for metals and MOG contaminant levels.

4. The updated probabilities are incorporated in the decision model to determine an updated near-optimal remediation design for a number of future remediation activity costs and batch sampling scenarios.

These steps and their results are discussed in the following sections.

5.4.5.1 FDEM processing

In the FDEM production surveys four measurements were made at each station—horizontal and vertical dipoles with the long axis of the tool oriented in the southwest-northeast (parallel to the survey line) and northwest-southeast (perpendicular to the survey line) directions. Horizontal and vertical dipoles refer to the transmitter-receiver coil orientation, changed by tilting the short axis of the tool 90 degrees. The two dipoles have different depths of investigation; the vertical dipole is deeper. The southwest-northeast and northwest-southeast measurements should be similar; the difference between these measurements is a good quality control.

The measurement values reported by the Geonics. Ltd. EM-31™ tool in the field are fully calibrated to be conductivity in millisiemens per meter (mS/m). However, processing is still required to eliminate noisy, misleading measurements and to transform the electrical conductivity data into a form that can be directly input into the decision model. The processing stream applied to the FDEM data consists of:

1. rectifying the conductivity data from the survey coordinates to the decision grid coordinates,
2. generating a grid from the data that exactly overlaps with the decision grid,
3. identifying zones/points where the measurements are degraded by unwanted noise and, thus, unusable, and

4. editing the data to eliminate noisy data and create final horizontal and vertical dipole grid data sets for analysis.

Rectifying the measurement station locations to the decision model coordinate system simply requires performing a transform on the station coordinates. After the transformation the data locations are compatible with all the data and results from the decision analysis, but do not necessarily fall on grid locations. Since the volume of investigation of the EM-31™ is quite large (something like a 4 meter by 2 meter by 4-6 meters depth ellipsoid) relative to the decision scale (1 x 1 x 1.5 meters), it was decided no information would be lost by interpolating the data to the decision model grid; this enables much easier soft data calibration and geostatistical estimation. Gridding of the four rectified data sets is accomplished with a simple inverse-distance algorithm. Contour plots of the rectified conductivity grids for the four conductivity measurements, blanked away from measurement locations, are shown in Figures 5.4.76 and -77. Immediately obvious are large (positive and negative) magnitude conductivity anomalies that differ between the southwest-northeast and northwest-southeast oriented measurements of the same dipole, the anomalies having orientations primarily parallel and perpendicular to the survey lines; these are indicative of problems with some of the data.
Figure 5.4.76a: Contour plot of FDEM horizontal dipole, NW-SE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.

Figure 5.4.76b: Contour plot of FDEM horizontal dipole, SW-NE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.
**Figure 5.4.77a:** Contour plot of FDEM vertical dipole, NW-SE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.

**Figure 5.4.77b:** Contour plot of FDEM vertical dipole, SW-NE tool-oriented conductivity, rectified and gridded to decision model grid coordinates.
The next step is to identify problematic data that is not an accurate representation of soil conductivity. One useful way of recognizing problem data is to look at the difference between southwest-northeast and northwest-southeast measurements of a particular dipole at the same location. This is accomplished by computing a difference grid between the southwest-northeast and northwest-southeast conductivity grids for each dipole. Contour maps of these difference grids, blanked away from measurement locations, are displayed in Figure 5.4.78. While large areas of both difference grids exhibit little difference between the two orientations (green zones), there are a number of zones with differences of greater than 20 mS/m absolute value. These zones correspond with the location of two chain-link fences, one positioned parallel to the survey lines in the northwest portion of data area of the grid, the other positioned roughly perpendicular to the lines in the southwest portion. It is expected that the measurements would be severely compromised in the vicinity of large metal objects. The impact appears to be more severe when the tool is oriented parallel to the fence and when the measurement is made with horizontal dipoles.

To eliminate a majority of the noisy data from the horizontal dipole results, the northwest-southeast and southwest-northeast grids are combined, taking the northwest-southeast grid data located greater than 40 meters in the northing direction and the southwest-northeast data below 40 meters. This combined grid is the final, fully-processed horizontal dipole conductivity grid—to be used in the geostatistical analysis; it is displayed as a contour map in Figure 5.4.79. Since the southwest-northeast vertical dipole data (Figure 5.4.77b) appears much better behaved than the northwest-southeast data (Figure 5.4.77a), the southwest-northeast grid is used as the final processed vertical dipole conductivity grid—to be used in the geostatistical analysis.
Chapter 5. CASE STUDY

Figure 5.4.78a: Contour plot of difference between FDEM horizontal dipole, NW-SE and SW-NE tool-oriented conductivity.

Figure 5.4.78b: Contour plot of difference between FDEM vertical dipole, NW-SE and SW-NE tool-oriented conductivity.
Figure 5.4.78b: Contour plot of combined FDEM horizontal dipole conductivity grid—constructed by merging low-noise parts of NW-SE and SW-NE oriented grids.

5.4.5.2 GPR processing

The raw 200 megahertz GPR data from each production survey line required processing before the data could be analyzed. This processing consisted of applying to the raw time series data:

- amplitude enhancement with increasing time to account for spherical wave losses
- amplitude normalization within a time window to reduce the effect of other types of energy losses
- temporal frequency-domain filtering to reduce low and high frequency noise
- spatial wavenumber-domain filtering to reduce the influence of high dip events—produced from radar reflections off the chain-link fences.

The data also has to be transformed from time to depth. In lieu of more concrete information, an average GPR velocity of 0.1 meter per nanosecond, appropriate for the lithology at the site, is
assumed for this transformation. A common midpoint (CMP) survey was attempted to determine the velocity, but, for some unknown reason, the data was corrupted.

Several examples of processed 2-d GPR sections are displayed in Figure 5.4.79. The sections were also fit into a 3-d volumetric grid for 3-d visualization, as displayed in Figure 5.4.80. Below about 4.5 meters depth there are no coherent reflections, suggesting this depth is about the maximum depth of penetration for the GPR at this site. Strong reflections occur in the upper part of the section, especially within the top 3 meters. One of these reflections is probably the reflection off the top of the water table, but it is difficult to discern, especially since the water table is known to vary considerably, depending on the level of the sea tide in the harbor. There are also zones of low amplitudes (green areas) that are of notable interest because one of the GPR responses to the presence of pure phase hydrocarbon is thought to be a dampening of amplitudes, due to the lowering of the dielectric constant [Olhoeft, 1986].

![Processed GPR relative amplitude section from case study site—line 2 of 15 in production survey.](image)

**Figure 5.4.79a:** Processed GPR relative amplitude section from case study site—line 2 of 15 in production survey.
Figure 5.4.79b: Processed GPR relative amplitude section from case study site—line 9 of 15 in production survey.

Figure 5.4.79c: Processed GPR relative amplitude section from case study site—line 14 of 15 in production survey.
Figure 5.4.80: 3-d volumetric representation of processed GPR relative amplitude sections acquired at case study site.

The processed GPR sections are rectified to the decision model grid coordinate system by applying a transform to the measurement station locations. To interpolate the rectified data to actual horizontal grid nodes ordinary kriging is used with a short range to retain the high resolution of the measurement. This is performed in two ways: (1) so as to retain the 0.04 meter depth sampling of the GPR and (2) increasing the scale to match closer the sample data and decision model scales closer by averaging across 0.5 meter depth increments, enabling easier calibration with hard sample data. The results of the gridding operations of (1) and (2) are displayed in Figure 5.4.81. While the GPR resolution is lost with the averaging, and associated smoothing, in the 0.5 meter depth increment grid, this enables an assessment of whether there is a correlation between contaminated soil and the average GPR amplitude for a block. These two GPR amplitude grids, particularly the coarser one, are the final processed GPR data for use in the geostatistical analysis.
Figure 5.4.81a: 3-d, 200 MHz, GPR relative amplitude grid—from processed, transformed, and gridded case study production survey 2-d sections.

Figure 5.4.81b: Coarser 3-d, 200 MHz, GPR relative amplitude grid—from processed, transformed, gridded and averaged case study production survey 2-d sections.
5.4.5.3 FDEM calibration

Both the horizontal and vertical dipole conductivity data are calibrated with collocated soil sample data located at depths less than 4 meters and 7 meters from the surface, respectively. At the locations where the FDEM grid overlaps sample data, all the sample analysis results within the 1 x 1 meter by 4 or 7 meter depth block are paired with the single conductivity value for that grid cell as separate hard-soft calibration pairs. In this way, the larger volume of investigation of the FDEM measurement is accounted for in the calibration. The largest viable horizontal dipole calibration data sets consist of 15, 8, and 5 data pairs for metals, MOG, and PAH contaminants, respectively; vertical dipole data sets contain 23, 18, and 7 data pairs for metals, MOG, and PAH contaminants, respectively. Because there are so few available collocated data pairs for PAH, it is not considered a valid calibration candidate.

The next step is to make a scatterplot of the collocated hard soil sample data versus the soft conductivity values to determine the optimal conductivity cutoff(s)—those that provide the most effective delineation of the contaminant level(s)—for the four vertical/horizontal dipole, metals/MOG combinations. The horizontal and vertical dipole scatterplots are shown in Figures 5.4.82 and 5.4.83, respectively. In these figures, sample classifications of 1, 2, and 3 correspond to soil contamination below level B, below level C, and above level C, respectively, for the metals scatterplot; classifications of 1, 2, 3, and 4 correspond to contamination below level B, below level C, below special waste, and above special waste, respectively, for the MOG scatterplot. For all these combinations, higher conductivity is chosen as an indicator of contamination—metals and MOG contamination greater than level B and special waste thresholds, respectively. The soft conductivity cutoff values chosen for the FDEM calibration are displayed as green horizontal lines in the scatterplots and are itemized in Table 5.4.7.
Figure 5.4.82: Calibration scatterplots of collocated FDEM horizontal dipole conductivity grid values versus soil sample analysis results for: (a) metals contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = greater than level C), (b) MOG contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = between level C and SW, 4 = greater than special waste). Conductivity cutoff chosen for calibration is shown as green line.
Figure 5.4.83: Calibration scatterplots of collocated FDEM vertical dipole conductivity grid values versus soil sample analysis results for: (a) metals contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = greater than level C), (b) MOG contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = between level C and SW, 4 = greater than special waste). Conductivity cutoff chosen for calibration is shown as green line.
Table 5.4.7: FDEM conductivity cutoffs chosen for calibration of FDEM data to metals and MOG contaminant levels, measured in soil samples.

With the FDEM soft data conductivity cutoffs chosen, the calibration is performed to calculate a soft data cdf probability table and misclassification probabilities. As recommended in this work, a sensitivity analysis is performed on the calibration results by adding a data pair for every possible combination of hard-soft data outcomes and performing the calibration for each combination. Following a conservative approach, the result with the worst calibration is taken as the final calibration—to be used in the geostatistical analysis. The final calibration results, after the sensitivity analysis, are plotted as ogives in Figures 5.4.84-87 for the four calibrations. Each calibration result is presented in two ways:

- Normalized to the case where there are an equal number of hard samples in each of the primary variable contamination level classes. This manipulation permits the classification of FDEM data quality as the likelihood of identifying contamination above/below the respective contaminant level from the calibration results. The data quality determined in this way can be related back to the preposterior data worth analysis data qualities.

- Normalized for the inferred primary variable contamination threshold expected values. These are the calibration results required for the M-B geostatistical estimation approach used in this study.

The data quality, as defined above, can be easily determined from the calibration cdf table computed from the equally-proportioned hard data case; one minus the cdf value for FDEM data above the defined conductivity cutoff is the likelihood that conductivity greater than the cutoff correspond to soil contamination above the respective contaminant level. Conversely, the cdf
value for FDEM data below the cutoff is the likelihood that conductivity less than the cutoff correspond to contamination below the respective level. Another parameter for gauging soft data quality based on the calibration is the $B$ factor, calculated from the equally-proportioned hard data case, which is equal to the difference between misclassification probabilities $P1$ and $P2$, (see Section 4.3.1.2 for a description of these parameters). It varies between $-1$ and $1$, negative being an anti-correlation between soft and hard data; the closer the absolute value of $B$ is to one, the better the correlation/"data quality", the closer to zero the worse it is. The $B$ parameter is probably a better overall metric of soft data quality (for measuring the primary variable), because it combines classification and misclassification probabilities into a single, comprehensive "degree of correlation" number. However, in the author's opinion, it is less intuitive than the "likelihood of identification" description of soft data quality, especially when trying to simulate, and describe, soft data measurements of different data qualities, as was done in the geophysics data worth analysis.

A summary, using likelihoods of identification and $B$ values, of the ability of the horizontal and vertical dipole FDEM measurements to delineate metals and MOG soil contamination above level B and special waste regulatory thresholds, respectively, based on the M-B soft data calibration results, is presented in Table 5.4.8. The calibrations suggest both horizontal and vertical dipole measurements are much better at distinguishing whether or not there is MOG special waste contamination, versus metals level C contamination. In the context of the geophysics data worth analysis, the FDEM data quality for this MOG application would be qualitatively ranked as between "good" and "excellent" (nearer to "excellent") for the vertical dipole, and "medium" for the horizontal dipole. The vertical dipole is considerably better than the horizontal for both the MOG and metals application. Even for the metals level C delineation the vertical dipole shows some success ("medium" quality), whereas the horizontal appears borderline useless.

Of course, the best way to evaluate how well the FDEM measurements reduce the uncertainty about the subsurface distribution of these contaminants is examine the metals and
Chapter 5. CASE STUDY

MOG contaminant level probability maps after conditional kriging with this additional data; ultimately, the economic value of the FDEM measurements is determined by examining the overall effect on the optimized remediation design. These analyses are enumerated later in this chapter.

<table>
<thead>
<tr>
<th>FDEM Measurement</th>
<th>Contaminant level</th>
<th>Likelihood of correctly identifying contamination</th>
<th>Likelihood of correctly identifying no contamination</th>
<th>B correlation parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Horizontal dipole</td>
<td>Metals &gt; B</td>
<td>59%</td>
<td>64%</td>
<td>0.05</td>
</tr>
<tr>
<td></td>
<td>MOG &gt; SW</td>
<td>75%</td>
<td>100%</td>
<td>0.50</td>
</tr>
<tr>
<td>Vertical dipole</td>
<td>Metals &gt; B</td>
<td>75%</td>
<td>65%</td>
<td>0.16</td>
</tr>
<tr>
<td></td>
<td>MOG &gt; SW</td>
<td>94%</td>
<td>100%</td>
<td>0.89</td>
</tr>
</tbody>
</table>

Table 5.4.8: Summary parameters for results of calibration between FDEM conductivity and metals / MOG soil samples.
Figure 5.4.84: Ogive plots of calibration results for FDEM horizontal dipole conductivity—calibrated to metals sample data for delineating metals contamination above and below level B threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.
Figure 5.4.85: Ogive plots of calibration results for FDEM vertical dipole conductivity—calibrated to metals sample data for delineating metals contamination above and below level B threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.
Figure 5.4.86: Ogive plots of calibration results for FDEM horizontal dipole conductivity—calibrated to MOG sample data for delineating MOG contamination above and below special waste threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.
Chapter 5. CASE STUDY

Figure 5.4.87: Ogive plots of calibration results for FDEM vertical dipole conductivity—calibrated to MOG sample data for delineating MOG contamination above and below special waste threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

5.4.5.4 GPR calibration

The calibration of the processed GPR data is performed in much the same way as the FDEM calibration. Calibration data sets of hard metals, MOG, and PAH sample data that are collocated with the averaged GPR amplitude grid are built. At each sample location that intersects the grid, all sample depth intervals that are within the top 4.5 meters of the subsurface are paired with the GPR amplitude of the 0.5 by 0.5 by 0.5 meter grid cell the samples fall within. The largest viable calibration data sets consist of 17, 19, and 7 data pairs for metals,
MOG, and PAH contaminants, respectively. As with the FDEM data set, because there are so few available collocated data pairs for PAH, it is not considered a valid calibration candidate.

Figures 5.4.88 and 5.4.89 display scatterplots of the GPR versus metals and GPR versus MOG calibration data sets, respectively (hard soil sample data plotted versus the collocated soft GPR amplitude values)—used to determine the GPR amplitude cutoff(s) that provide the most effective delineation of the metals and MOG contaminant level(s). In these figures, sample classifications of 1, 2, and 3 correspond to soil contamination below level B, below level C, and above level C, respectively, for the metals scatterplot; classifications of 1, 2, 3, and 4 correspond to contamination below level B, below level C, below special waste, and above special waste, respectively, for the MOG scatterplot. In the metals scatterplot, GPR amplitudes are also plotted as absolute values. A cutoff of 0.7 for absolute value GPR amplitudes seems to provide the most effective delineation between metals contamination below and above level B, for absolute value GPR values below and above 0.7, respectively. For the MOG data set, a GPR amplitude cutoff of -0.7 seems to provide the most effective delineation between MOG contamination below and above the special waste threshold, for GPR values above and below -0.7, respectively. These cutoff values are displayed as green horizontal lines in the scatterplots, and are itemized in Table 5.4.9.
Figure 5.4.88: Calibration scatterplots of collocated GPR averaged relative amplitude grid values versus soil sample analysis results for: (a) metals contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = greater than level C), (b) MOG contaminant levels (1 = less than level B threshold, 2 = between level B and C, 3 = between level C and SW, 4 = greater than special waste). GPR amplitude cutoff chosen for calibration is shown as green line. Cutoff is for absolute value of amplitude in (a).
Chapter 5. CASE STUDY

The metals and MOG calibration data sets are used with the defined GPR cutoffs to perform M-B calibrations of how well the GPR distinguishes between contaminant levels above and below metals level B and MOG special waste regulatory limits. As part of each calibration, a sensitivity analysis is performed on the calibration results by adding a data pair for every possible combination of hard-soft data outcomes, performing the calibration for each combination, and retaining the result from the worst calibration. Ogive plots of the cdf probability table and misclassification probabilities from the final calibration results for the metals and MOG data sets are shown in Figures 5.4.89 and 5.4.90, respectively. Two probability normalization cases for each data set are displayed, where the results are normalized to: (1) equally-proportioned number of hard data in each primary variable contaminant class and (2) the inferred contaminant level expected value. The likelihoods of identification and B values for using GPR to delineate metals and MOG soil contamination above level B and special waste regulatory thresholds, based on the M-B soft data calibration results, is presented in Table 5.4.10.

The calibration results suggest the GPR measurements are slightly better at distinguishing whether or not there is metals level C contamination, versus MOG special waste contamination. For both applications, though, the “quality” of the GPR data is low—qualitatively ranked as less than “medium” using the definitions applied in the geophysics data worth analysis.

<table>
<thead>
<tr>
<th>GPR amplitude</th>
<th>&gt; Metals level B</th>
<th>&gt; MOG special waste</th>
</tr>
</thead>
<tbody>
<tr>
<td>Absolute value</td>
<td>&gt; 0.7</td>
<td>&lt; -0.7</td>
</tr>
</tbody>
</table>

Table 5.4.9: GPR amplitude cutoffs chosen for calibration of GPR data to metals and MOG contaminant levels, measured in soil samples.
Figure 5.4.89: Ogive plots of calibration results for GPR absolute value amplitudes—calibrated to metals sample data for delineating metals contamination above and below level B threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.
Figure 5.4.90: Ogive plots of calibration results for GPR amplitudes—calibrated to MOG sample data for delineating MOG contamination above and below special waste threshold: (a) cdf table, normalized to hard data being equally proportioned in classes, (b) misclassification probabilities for same normalization; (c) cdf table, normalized to inferred contaminant level expected values, (d) misclassification probabilities for same normalization.

<table>
<thead>
<tr>
<th>Contaminant</th>
<th>Likelihood of correctly identifying contamination</th>
<th>Likelihood of correctly identifying no contamination</th>
<th>B correlation parameter value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Metals &gt; B</td>
<td>72%</td>
<td>82%</td>
<td>0.28</td>
</tr>
<tr>
<td>MOG &gt; SW</td>
<td>73%</td>
<td>62%</td>
<td>0.11</td>
</tr>
</tbody>
</table>

Table 5.4.10: Summary parameters for results of calibration between GPR relative amplitudes and metals / MOG soil samples.
5.4.5.5 Updating contamination probabilities

Metals and MOG contaminant level probabilities are updated based on the new FDEM and GPR data by conditional M-B indicator kriging with the processed data sets and their calibration probability tables. Both horizontal and vertical dipole FDEM grid data are used, the former applied to the top two decision model grid layers (maximum depth of 3 meters) and the latter to the top four layers (maximum depth of 6 meters). The GPR data used in the conditional kriging is obtained from the averaged GPR data grid, to a depth of 4.5 meters below the surface; data below this depth is considered noise. The contaminant level probability fields for metals and MOG estimated from conditional kriging of both horizontal and vertical dipole FDEM data, in addition to all existing hard and soft sample data, are displayed in Figures 5.4.91 and -92. The same fields, estimated with the GPR data, are displayed in Figures 5.4.93 and -94.

The FDEM data delineates several large new zones with higher probability (up to 50%) of metals level B contamination, in the vicinity of where sample data measured the only metals contamination of concern. Outside these zones, in the region where there is FDEM data, the probability of level B contamination is reduced. While the metals level C contamination probability field is not directly conditioned on the FDEM data, where the probability of level B contamination is reduced because of the FDEM data (and, correspondingly, the probability of metals concentrations below level B increased), the probability of level C contamination is reduced as well, since the overall probability of contamination at level B and above has been reduced. The impact of the FDEM data on the probability of MOG special waste contamination field is similar to the metals level B results. New isolated zones of higher probabilities are delineated in the vicinity of where MOG special waste was identified with soil samples; the remaining zones within the region where FDEM data exists have lower probabilities.

The impact of incorporating the GPR data into the conditional kriging for estimating metals contaminant level probabilities is a patchwork of widely distributed, highly variable zones within the GPR survey region, where there are elevated, or reduced, level B probabilities (see Figure 5.4.93a). These zones are much more discontinuous and pervasive across the entire survey area than those indicated by the FDEM; the higher variability can be explained by the
large difference in the spatial resolution of the FDEM and GPR, the GPR providing a much more
detailed picture. As with the FDEM results, the metals level C probabilities are reduced in the
zones where level B probabilities are reduced (Figure 5.4.93b). The MOG special waste IRV
kriging results have a similar type of pattern of anomalous probabilities as the metals level B
results, except the anomalies have a different spatial distribution and the change in probability
values from before-geophysics probabilities are not as large (Figure 5.4.94).

Figure 5.4.91a: Probability of metals level B soil contamination—estimated from
cokriging of all sample data and FDEM horizontal and vertical dipole conductivity.
**Figure 5.4.91b**: Probability of metals level C soil contamination—estimated from cokriging of all sample data and FDEM horizontal and vertical dipole conductivity.

**Figure 5.4.92**: Probability of MOG special waste soil contamination—estimated from cokriging of all sample data and FDEM horizontal and vertical dipole conductivity.
Chapter 5. CASE STUDY

Figure 5.4.93a: Probability of metals level B soil contamination—estimated from cokriging of all sample data and GPR absolute value amplitudes.

Figure 5.4.93b: Probability of metals level C soil contamination—estimated from cokriging of all sample data and GPR absolute value amplitudes.
Figure 5.4.94: Probability of MOG special waste soil contamination—estimated from cokriging of all sample data and GPR amplitudes.

While the previous probability estimation from the FDEM and GPR data treated separately is valuable for evaluating the impact of each measurement independently, the goal from the decision-making framework is to integrate all available information into a single description of contamination across the site, with uncertainty quantified. Towards this end, both the FDEM and GPR data, along with all hard and soft sample data, are incorporated in a single M-B conditional indicator kriging estimation of metals and MOG contaminant level probabilities.

Figures 5.4.95a, -95b, and -96 display the entire site probability fields for metals level B, metals level C, and MOG special waste contaminant levels, respectively, generated from the integrated data kriging. The combination of FDEM and GPR data reduces the degree of increased probability anomalies delineated by the GPR for metals level C and MOG special waste, especially to the northeast of the grid. The higher probability anomalies, in the vicinity of where contamination was detected by sampling, are consolidated and more clearly delineated.

To highlight changes, Figures 5.4.97a, -97b, and -98 show the difference in the estimated grids between the new and pre-geophysics results for metals level B, level C, and MOG special
waste, respectively, where only difference grid values greater than 0.1, 0.02, and 0.1 absolute value, respectively, are displayed.

Zoomed-in views of the portion of grid where the geophysics surveys were acquired are displayed in Figures 5.4.99 and -100.

The variances of the combined geophysics kriging results are exhibited in Figures 5.4.101a, -101b, and 102, showing increases in the IRV variance in the delineated contamination “hot zones” and a decrease everywhere else where geophysics was collected; this is what would be expected from the probability field observations.

Finally, the grid of overall, exclusive probability of contamination, at any of the contaminant levels of concern, is shown in Figure 5.4.103a, for comparison with the pre-geophysics counterpart (Figure 5.4.103b). There is a noticeable increase in the contamination probability in the metals and MOG contaminated zone delineated by the geophysics.

**Figure 5.4.95a**: Probability of metals level B soil contamination—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
Figure 5.4.95b: Probability of metals level C soil contamination—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.

Figure 5.4.96: Probability of MOG special waste soil contamination—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
Figure 5.4.97: Differences greater than 0.1 between metals level B probabilities estimated after and before addition of FDEM and GPR geophysics data to all existing sample data.

Figure 5.4.97b: Differences greater than 0.02 between metals level C probabilities estimated after and before addition of FDEM and GPR geophysics data to all existing sample data.
Chapter 5. CASE STUDY

Figure 5.4.98: Differences greater than 0.1 between MOG special waste probabilities estimated after and before addition of FDEM and GPR geophysics data to all existing sample data.

Figure 5.4.99a: Probability of metals level B soil contamination across geophysics survey region—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
Figure 5.4.99b: Probability of metals level C soil contamination across geophysics survey region—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.

Figure 5.4.100: Probability of MOG special waste soil contamination across geophysics survey region—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
**Figure 5.4.101a:** Variance of metals level B soil contamination probabilities—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.

**Figure 5.4.101b:** Variance of metals level C soil contamination probabilities—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
Chapter 5. CASE STUDY

Figure 5.4.102: Variance of MOG special waste soil contamination probabilities—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.

Figure 5.4.103a: Probability of contamination above any of the metals, PAH, or MOG regulatory threshold levels—estimated from cokriging of all sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.
5.4.5.6 Updating remediation designs

The decision analysis is performed again with the geophysics updated contaminant level probability fields to investigate the resulting changes in the optimized remediation design, for a range of future unit remediation activity costs and batch sampling scenarios. Figures 5.4.103-105 are updated remediation designs for several scenarios. The contaminated zone delineated by the geophysics is now allocated for level C / special waste treatment in the no batch sampling, 2.0 future-to-present unite remediation cost ratio scenario (see Figure 5.4.103a). More of the soil is designated for excavation only in eastern part of the site now, also, for this scenario, as illustrated in Figure 5.4.103b—the blocks not requiring treatment. Little change is evident in the $4 per cubic meter batch sampling scenario for the same future-to-present unit remediation cost ratio (Figure 5.4.104a). More blocks in the eastern section of the site are to be treated or excavated now, rather than batch sampled, in the $8 batch sampling, same future costs scenario (Figure 5.4.104b). Similarly, treatment and excavation zones are more clearly delineated in this
section of the site for the no batch sampling scenario where the future-to-present unit remediation cost ratio is only 1.125 (Figure 5.4.105).

**Figure 5.4.103a:** Optimized remediation design based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 2.0.
Chapter 5. CASE STUDY

Figure 5.4.103b: Portion of site soil volume not requiring treatment for contamination—remediation design in Figure 5.4.103a.

Figure 5.4.104a: Optimized remediation design based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes—for scenario where (1) $4/m^3 batch sampling is an alternative, (2) ratio of future to present unit remediation cost is 2.0.
Figure 5.4.104b: Optimized remediation design for same scenario as Figure 5.4.104a, except unit cost of batch sampling is $8/m³.

Figure 5.4.105: Optimized remediation design based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes—for scenario where (1) batch sampling is not an alternative, (2) ratio of future to present unit remediation cost is 1.125.
The total remediation cost for the new remediation designs, updated with the geophysics information, are plotted for the different batch sampling scenarios across a range of future-to-present unit remediation cost ratios in Figure 5.4.106a. In addition, the difference between the pre-geophysics total remediation cost and the post-geophysics cost—the economic value, or cost savings, provided by the FDEM and GPR geophysics surveys—for the different scenarios is plotted in Figure 5.4.106b. As can be seen from this plot, the value of information is quite insignificant, having a maximum value of $45,000 at a future-to-present unit remediation cost ratio of 3.0, if batch sampling is not an alternative. If batch sampling at a unit cost of $4 is an alternative, the cost savings is negligible, while if the unit cost is $8, the geophysics provides a maximum cost savings of $22,000, at a 1.5 future-to-present cost ratio. Interestingly, at the latter batch sampling cost, the value of the geophysics falls back to near zero with an increasing cost ratio; it is possible this behavior is simply due to modeling errors. All these value of information / cost savings numbers do not account for the acquisition, processing, and analysis costs of the geophysics, which will reduce the net cost savings.
Figure 5.4.106a: Total cost of remediation for different batch sampling scenarios and across a range of future-to-present remediation cost ratios—based on all existing sample data, FDEM horizontal and vertical dipole conductivity, and GPR amplitudes.

Figure 5.4.106b: Economic value (cost savings in remediation design) of FDEM and GPR geophysics surveys—for same scenario and data types as in Figure 5.4.106a.
The FDEM and GPR geophysics survey measurements, collected over a small area in the eastern part of the case study site, are successfully incorporated in the decision model by performing a simple statistical calibration between collocated geophysics and soil sample data. The geophysics data, with the calibration results, is used in the geostatistical kriging estimation model to update the metals and MOG contaminant level probabilities across the site. The change in these probabilities within the region where the geophysics surveys were acquired is fairly significant after the updating. However, the overall impact on the decision analysis results—cost of the optimized remediation design—is insignificant. This is not surprising, and is in fact expected based on the data worth analysis, which indicated that only geophysics of very good quality, applied across the site, provides a significant economic value (cost savings) to the overall remediation. While some of the acquired geophysics indicated good quality for some contaminant delineation applications from the calibration, the areal coverage of the site from the geophysics surveys is very small.
CHAPTER 6
CONCLUSIONS AND RECOMMENDATIONS

This chapter summarizes the important findings from the study presented in this work—the adaptation of a risk-based decision-making framework that rigorously accounts for measurement uncertainty, and application of the framework to a real-world case study wherein geophysics is used as a site characterization technique. The research results are summarized, highlighting the modeling and analysis framework that was developed, the conclusions and lessons learned from the case study, and important contributions that can be applied in practice to other projects. The chapter finishes with a section providing recommendations for further research on topics addressed by, or related to, this work.

6.1 Research results

The work completed for this study consisted of:

- developing a generalized methodology for directly handling—in the form of probabilities—the many types of uncertainty associated with hydrogeology-related problems, especially in regards to the application of geophysical measurements for site characterization in such problems;
- incorporating the methodology in a risk-based decision-making framework, wherein robust engineering decisions can be made that account for all the inherent risks of such problems; and
- applying the framework to a real-world case study—soil remediation of a real estate development site, wherein geophysical measurements were acquired and incorporated in the site characterization process.

Figure 6.1.1 schematically illustrates the main concept explored in this research—how to incorporate geophysics measurements in the quantitative characterization of a subsurface property, wherein the many uncertainties associated with the geophysics (indirectness,
measurement errors, nonuniqueness, etc.), as well as the natural spatial variability of the property, are accounted for. The schematic hydrogeology problem is a site with subsurface contamination wherein two types of site characterization techniques are employed to help delineate the pure phase hydrocarbons contamination:

- Ground Penetrating Radar (GPR) that non-invasively measures radar reflection versus propagation time profiles from the surface; and
- Physical samples of subsurface material that are analyzed for contaminant concentrations—extracted by emplacing invasive boreholes.

Two conceptually different approaches for incorporating the GPR measurements to estimate hydrocarbon saturation are depicted:

- a statistical approach that accounts for the correlation and uncertainty between a GPR response attribute and hydrocarbon saturation levels based on a calibration of collocated GPR and borehole sample measurements (Magenta arrows show the key steps for applying this approach in Figure 6.1.1.); and
- a physical model approach that first uses mathematical inversion to convert the GPR amplitude versus time results to a dielectric constant versus depth section, based on an electromagnetic wave physical model, and then relates dielectric constant to hydrocarbon saturation, based on a rock physics model (Blue arrows show the key steps for applying this approach in Figure 6.1.1.).

A combination of the two approaches is also shown—inversion of the GPR section to obtain dielectric constant, which is then calibrated against the borehole samples for usage in the statistical approach.

The statistical calibration uses a scattergram plotting the GPR attribute value versus hydrocarbon saturation for the collocated data set to construct histograms of the GPR attribute distribution, from within defined intervals, versus hydrocarbon saturation levels. These histograms represent the uncertainty associated with indirectly determining hydrocarbon
saturation from GPR in a simple, straightforward way—using probabilities—that inherently accounts for all the different forms of uncertainty the GPR measurements can have.

Accounting for uncertainty in the physical model approach is more elusive and difficult. The inversion of the GPR amplitude section to dielectric constant is typically a non-unique problem, wherein a number of different dielectric constant versus depth combinations provide acceptable solutions to the inversion; however, representing and accounting for this uncertainty is a complicated task. The rock physics model relating dielectric constant to hydrocarbon saturation can be designed to account for non-uniqueness in the relation—in particular, the fact that dielectric constant is a function of many other subsurface properties that vary across the site. Using the mathematical functions describing this relation for the range of anticipated variability in other unmeasured functional dependencies, histograms of hydrocarbon saturation levels for the dielectric constant distribution in defined intervals can be calculated—providing a workable representation of uncertainty, as in the statistical approach. In lieu of performing the rock physics model step, the derived dielectric constants can be statistically calibrated with the borehole sampling results to produce the same types of histograms.

The advantage of the physical model approach is that the derived relation between the GPR measurements and the property of concern—subsurface hydrocarbon saturation—should be more constrained than that for the statistical approach; this reduces the amount of uncertainty associated with estimating the property. However, accounting for the uncertainty in the estimates is mathematically complicated and difficult. In contrast, the statistical approach very cleanly accounts for uncertainty in the estimation of hydrocarbon saturation from any attribute, raw or derived, resulting from the GPR measurement; if the calibration is based on a large enough and representative data set, then all forms of uncertainty in the measurement should be correctly accounted for.

Because of its simplicity of use, robustness, generality, and ease of linkage to a risk-based decision making model, a statistical approach is used for the uncertainty model in this research, as described below.
Figure 6.1.1: Schematic diagram of different approaches—statistical and physical model—for estimating a subsurface property of interest (hydrocarbon saturation) from indirect geophysics measurements (GPR) that account for uncertainty in the measurement and estimation.

6.1.1 Uncertainty model

The Markov-Bayes (M-B) indicator geostatistical cokriging and simulation approach of spatial variable estimation [Zhu and Journel, 1993] was chosen for dealing with uncertainty, because of its generality, versatility, and straightforward representation of uncertainty; uncertainty is directly represented and updated as probabilities, and both uncertainty associated with spatial variability and measurement/observation uncertainty can be represented and
accounted for in a straightforward manner. The methodology was adapted and generalized so that:

- indirect geophysical measurements are readily incorporated in the processing stream;
- all the steps required for robust geostatistical estimation are implemented in compatible modules and can be performed in a practical, efficient manner;
- many options are available for application to a variety of problems, including using either the Alabert [1987] or Zhu and Journel [1993] M-B approach, performing point or block estimation, applying the methodology to continuous or categorical variables; and
- the algorithms, and their input and output, are optimized for linkage to a broad range of risk-based decision models.

Variables are represented in the M-B geostatistical approach by one or more indicator random variables (IRVs), defined as zero or one depending on whether the variable falls within a defined interval or class. Serendipitously, the expected value of an IRV, including the kriging estimate, is the probability that the variable falls within the interval/class; the set of IRV expected values (kriging estimates) for a series of defined cutoffs/classes of a variable represents a discrete probability distribution. Uncertainty due to spatial variability is accounted for by a spatial covariance model for each IRV; uncertainty due to measurement/observation error or indirectness (i.e. not directly measuring the primary variable of interest) is accounted for by a pre-defined probability or cross-covariance model for each IRV. The cross-covariance models for indirect measurements are derived from (1) the covariance model for the variable of interest and (2) a simple calibration between collocated indirect ("soft") measurements and direct ("hard") measurements of the variable; this eliminates the need for complicated, intensive methods of deriving cross-covariance models, such as mathematical inversion or computation of experimental cross-covariograms—requiring computationally intensive algorithms and very large combined direct-indirect measurement data sets, respectively.

Figure 6.1.2 is a flow diagram illustrating the steps, and options, for implementing the adapted M-B geostatistics approach developed in this work. The first step is to define indicator
threshold cutoffs or categories for the variable of interest (number (1) in the flow diagram); these
define the IRVs used for representing uncertainty and should be determined based on the specific
needs for the problem at hand (e.g. soil contamination regulatory action levels were chosen as
the cutoffs for the case study in this work, since required soil remediation actions are determined
from these thresholds). All information about the variable, including indirect measurements such
as geophysics, is represented as IRV values according to the defined IRV classifications (number
(2) in the flow diagram).

The information is used to infer prior "global" statistical models for each IRV—the first
step in the two step process of updating IRVs based on all existing information about the variable
of interest (number (3) in the flow diagram). This step entails inferring field-wide models for the
mean, covariance, and correlation model that relates the secondary variable measured by an
indirect measurement to the primary variable of interest—the latter only required if indirect
measurements are being used. The global means (number (3a) in the flow diagram) are computed
using a declustering algorithm published in *Deutsch and Journel* [1992], modified for indicator
RVs, including the ability to incorporate indirect measurements in the computation. The
secondary-primary variable correlation model (number (3b) in the flow diagram) is calculated
from a data set of collocated direct-indirect measurement pairs using the M-B calibration routine
from *Deutsch and Journel* [1992]—modified to include options for performing (1) an IRV mean-
normalized calibration and (2) a calibration for either the *Alabert* [1987] or *Zhu and Journel*
[1993] M-B approaches. IRV covariance model inference (number (3c) in the flow diagram) is
carried out in a three step process:

1. Experimental semi-variograms are calculated for each IRV class.

2. A chosen semi-variogram functional model is fit to each experimental result by adjusting
   several key parameters in the model. Both this operation and the one above are performed
   using the variogram modules within the UNCERT suite of software [Wingle et al, 1999].

3. The fitted models are validated against the available data for the study area using an
   orthonormal residuals validation routine [Kitanidis, 1991] developed for this work—wherein,
for each IRV class, expected values are estimated from kriging at each data location and compared with the data value at that location. The software routine was written by modifying the cross validation and “jacknifing” code in Deutsch and Journel [1992]. Iterative runs of the validation routine are performed, and the parameters of the functional model are manually varied, to minimize a set of normatives based on the orthonormal residuals.

Together, these prior global models comprise the field-wide spatial uncertainty model, which contains little information about the spatial structure of the variable of interest on a local scale.

The second step in the two step updating process is to update the uncertainty model on a local scale using M-B conditional estimation routines (number (4) in the flow diagram) with all the available information about the variable of interest—transformed to IRV values. The types of “conditioning” information that can be easily incorporated in the estimation is very broad, including information that is uncertain, such as indirect measurements of and personal judgment about the variable of interest; this is one of the strongest attributes of the M-B estimation approach. The estimation is either in the form of an IRV expected value (corresponding to the probability of occurrence within the defined IRV class) or a simulated realization at designated locations in the region of interest, comprising two separate routines—an M-B indicator cokriging routine (number (4a) in the flow diagram) and stochastic simulation routine (number (4b)), respectively. The M-B simple kriging and simulation computer programs in Deutsch and Journel [1992] were used as the foundation for these routines; these programs have been modified and generalized to be more conducive to wide-ranging decision-making problems and applications, and include the following additions:

- Data input and estimation output are amenable to linkage within a decision-making framework and model (e.g. output of multiple probability distribution files, with file names that are sequentially numbered, based on the input of multiple realizations of indirect measurement results).
Chapter 6. CONCLUSIONS AND RECOMMENDATIONS

- "Soft" information (i.e. information containing uncertainty) in the form of probabilities of occurrence within IRV indicator classes (e.g. from expert judgment) can be incorporated in the estimation, at the level of prior means—as proposed by Alabert [1987].
- Multiple indicator kriging and simulation options are provided—for continuous or categorical variables, simple or ordinary kriging, point or block kriging.
Chapter 6. CONCLUSIONS AND RECOMMENDATIONS

1. **Definition of IRVs for variable of interest.** Requires choosing:
   - threshold cutoffs (for continuous variables)
   - categories (for categorical variables)

2. **Collection and compilation of information/data about variable of interest across study site:**
   - Direct measurements of variable
   - Indirect measurements of variable (measurements of secondary variable that is related to primary variable)
   - Expert inference/judgment about variable (in form of probabilities)

3. **Inference of Global Geostatistical Models**

   - **If using direct observations of variable of interest**
     - **Global Mean:** Derive indicator global means for each IRV cutoff/category from direct measurements using declustering analysis approach.
     - **Global Covariance:** Calculate experimental indicator semi-variograms from direct measurements for each IRV.
     - Choose semi-variogram function models (e.g. spherical) for each IRV cutoff/category and fit to experimental IRV semi-variograms by adjusting parameters (e.g. range).
     - Iterate between these operations to obtain optimum fit between models and data.
     - Validate semi-variogram mathematical models against existing data using the orthonormal residuals method [Kitanidis, 1991].

   - **If using direct observations and indirect measurements of variable of interest**
     - **Indirect Measurement to Primary Variable Correlation Model:**
       - Calculate indirect measurement miss-classification probabilities for each IRV cutoff/category. Accomplished by performing normalized M-B calibration of indirect measurements from scattergram between collocated direct and indirect data pairs (different algorithms for Alabert, Zhu-Journel approaches).

   - **If using Zhu-Journel approach iterate between these calculations since calibration is dependent on global mean**

3b. **Iterate between these operations to obtain optimum fit between models and data.**

   - Use inferred global models in M-B geostatistical estimation (see next page).

---

Figure 6.1.2a: Flow diagram of adapted M-B geostatistics approach—see caption next page.
Use inferior global models and all available data about the variable of interest as input to M-B geostatistical estimation (continued from previous page).

| Chapter 6. CONCLUSIONS AND RECOMMENDATIONS | 366 |

Conditional M-B indicator estimation of variable of interest at defined points/blocks across area/volume of interest.

**Types of information that can be used to guide estimation:**

- "Hard" data IRVs - derived from direct measurements/observations of variable of interest
- "Soft" data IRVs - derived from indirect measurements of variable of interest
- "Soft" data prior IRV expected values - inferred probabilities about the variable
- Data falling within an interval of values - represented in the form of "hard" data IRVs, wherein IRVs falling within the interval are not assigned any values

**Indicator kriging estimation of IRV expected values, as defined by the IRVs for the variable of interest:**

- Bayesian updating of prior global IRV models (mean and covariance) at each estimation location - based on all information about the variable of interest in the region surrounding the estimation location.
- Estimate is equivalent to probability of variable being within defined indicator class - below/above defined threshold cutoff or within defined interval/category, depending on IRV definition.
- Results in a discrete pdf/cdf at each estimation location, conditioned on all neighboring information, when estimates for the different IRV cutoffs/categories are combined.

**Simulation of variable spatial distribution realizations OPTIONS:**

- Simulate realizations of indicator values (zeros or ones) for each IRV cutoff/category
- Simulate realizations of the variable values (if the variable is a continuous variable)

**Estimation Options:**

- Estimation at designated point locations - at the scale of a discrete point
- Estimation of the fraction of point values in a defined area/volume falling within the IRV indicator classes
- Accounting for the difference in uncertainty reduction for different measurement support scales [prototype algorithm that needs to be further tested and evaluated]

**Perform process again as more information about variable of interest is obtained.**

**Figure 6.1.2b:** Flow diagram of the M-B geostatistics approach adapted and modified for this work, illustrating the process for implementing the methodology—the different modules, required and optional steps, and their sequence. Continued from previous page.
6.1.2 Decision model

A risk-based decision-making framework was developed around the M-B IRV-based uncertainty model; specifically, a risk-based soil excavation and treatment decision model was developed—applicable for the soil remediation case study. The objective of the decision model is to determine the optimal soil excavation and treatment design for a site containing contaminated soil by minimizing the unnecessary costs associated with soil contamination underclassification and overclassification (i.e. choosing a remediation design that does not account for contamination that is actually present or is over-designed for the level of contamination, respectively). The decisions are based on contaminant concentration action levels—concentrations above these levels require certain actions, depending on the planned usage of the land. Incorporation of the M-B method enables the direct utilization of cokriged probability maps—probabilities of soil contamination above the action levels—to determine how a parcel of soil should be remediated, if at all; the decision is based on the comparative costs for the different alternatives, including the risk cost of not remediating the soil when it actually contains contamination above the limits. In contrast, if a continuous RV geostatistical approach is used, simulation of a large number of contaminant concentration distribution realizations is required, from which probabilities of contaminant concentrations above the action levels can be determined; thus, using the M-B indicator kriging approach is considerably more computationally efficient, as well as more versatile with its ability to incorporate “soft” information in the probability estimates.

The soil excavation and treatment decision model consists of a set of three modules—computer programs—to perform key analyses in the decision-making process:

1. Determine the expected value most cost effective excavation and treatment design, based on all existing information. As discussed above, this analysis is performed using the cokriged probability of contamination maps as input from the M-B uncertainty model.
2. Determine the expected value of perfect information in terms of cost savings—to set a maximum site characterization budget. This analysis is performed using simulated
realizations of contaminant concentration action levels across the study site—whether the contaminant is above or below the defined action levels—as input from the M-B uncertainty model.

3. Determine the expected value of possible future site characterization programs in terms of cost savings—to evaluate the comparative worth of performing different types of site exploration. This is accomplished using kriged probability of contamination maps, conditioned on equally-likely realizations of hypothetical site characterization results, as input from the M-B uncertainty model. The site characterization realizations are generated from M-B contaminant action level simulations, culled to account for the coverage of the hypothetical site characterization program, and transformed into “soft” data realizations using hypothetical M-B calibration probabilities—if indirect measurements are being simulated.

The computer programs for these analyses were developed from square one, without the benefit of existing computer programs to build upon.

Figure 6.1.3 is a flow diagram that schematically summarizes the process for implementing the different modules of the risk-based decision model within a decision-making framework—highlighting their linkages with the M-B uncertainty model and overall site characterization process. After all existing hard and soft information about a contamination remediation site has been analyzed and incorporated in the M-B uncertainty model, the “near-optimal” excavation and treatment design can be determined using contamination probability grid(s) from the uncertainty model (number (1) in the above itemized list of analyses and in the flow diagram). Contamination level realizations of the site—generated by conditional simulation in the M-B uncertainty model with the existing information—can be used to estimate what the maximum site contamination characterization budget should be by calculating the difference between the cost of the “near-optimal” remediation design and the expected value cost of the remediation designs for all the realizations (number (2) above and in the flow diagram). To estimate the overall site remediation cost savings from executing additional site contamination
characterization work, the expected value cost of remediation designs for a series of equally-probable contamination probability grids resulting from the hypothetical characterization program is subtracted from the cost of the "near-optimal" remediation design (number (3) above and in the flow diagram); the hypothetical results account for all the existing hard and soft information, as well, through a combination of the M-B conditional simulation and cokriging. The "data worth" of any number of potential site characterization programs can be evaluated—as defined by the practitioner, in terms of measurement coverage and quality of data (i.e. how accurate the contamination action level measurement is)—including indirect measurements, such as geophysics. Based on the results from these analyses decisions about the future direction of the project are made (as depicted in number (4) of the flow diagram):

- whether to perform more site characterization work to further delineate contamination or to move on to site remediation using the existing information and the corresponding "near-optimal" remediation design;
- if the decision to acquire more data is made, what characterization program should be implemented.

The decision model used to determine the "optimal" excavation design—for all three of the analyses—is built around the determination of the most cost effective alternative for a discrete block of subsurface material, based on the contaminant concentration levels in that block. Uncertainty in the subsurface contaminant concentrations in the block is accounted for in this decision through risk costs—defined as the underclassification cost of not treating the material for a contamination action level (this cost is treated as a "known" input to the model), multiplied by the probability that contamination above the action level actually exists. If each block could be analyzed on its own, without influence on other blocks, and vice-versa, this decision model would be very simple and straightforward—a direct comparison of the costs for the different alternatives, performed for every block across the site. However, since material from above has to be excavated to reach a block at depth, the remediation decision made for an individual block affects the excavation cost component of the decision made for other blocks, and vice versa; the
decision is inexorably a global decision, not a local one, if several depth layers of blocks are
included in the decision model. While this decision-making problem is still a decision analysis
problem—not an optimization one—in that there are a discrete number of decision alternatives,
instead of a continuous function defining the alternatives, the number of possible alternatives
when there are many depth layers and areal blocks is very large; the set of possible alternatives
includes all the possible combinations of the alternatives in each individual block across the
entire site! Since evaluation of all possible alternatives is intractable, a heuristic approach is used
to evaluate a representative subset of the alternatives, with options for how large the subset is.
While such an approach does not ensure that the optimal solution (most cost effective
remediation design) will be attained, the goal is to obtain a "near-optimal" solution. All the
results from the case study portion of this study appear to indicate that the model is consistent in
obtaining solutions; in general, the prior (existing data) and preposterior (hypothetical future
data) decision analysis results from the case study are consistent with each other and coincide
with the independent expected value of perfect information analysis results (see CHAPTER 5).

Model parameters that can be varied in the overall decision model include the number of
depth layers of blocks, the size of the blocks, and the width of the bench required for excavation
(e.g. at each successive depth level in an excavation a bench width of one block size is required
to prevent wall collapse). The decision model is also set up to account for conditional
probabilities between different contaminants. Conditional probabilities can significantly
influence the decision analyses if there is a significant correlation between the presence/absence
of different contaminants or the problem dictates that the different contaminant action levels
need to treated as mutually exclusive; the latter is the case if, when soil contains more than one
contaminant breaching action levels, the soil is only treated as prescribed by the most severe
action level that is breached. Accounting for these conditional probabilities requires a conversion
from the contaminant-independent probabilities output of the geostatistical estimation model to
conditional probabilities; this is accomplished either in the form of user defined conditional
probabilities between contamination classes of different contaminants, as a constant or defined locally at the grid scale, or assuming contamination classes are independent from each other.
Chapter 6. CONCLUSIONS AND RECOMMENDATIONS

Carry out site characterization program to gain information about subsurface contamination across site.

M-B Uncertainty Model

1. M-B IRV cokriging with all existing hard and soft information on contamination — to estimate probability of contamination above action levels, for each contaminant action level at each grid block.

2. M-B IRV conditional simulation with all existing hard and soft information on contamination — to estimate realizations of contamination above action levels, for each contaminant action level at each grid block.

3. M-B IRV cokriging with all existing hard and soft information on contamination, plus hypothetical measurement results — to estimate probability of contamination above action levels when additional information is available, for multiple possible measurement outcomes.

4. Creation of hypothetical contamination measurement results from culled contamination realizations, transformed to "soft" data if indirect measurements are being evaluated — generated using M-B IRV conditional simulation with all existing information.

Decision Model

1. Determine "near-optimal" excavation and treatment remediation design based on all available information about the site. Required input:
   - Excavation block size, grid size, and excavation bench width
   - K riged probability of contamination above each action level for each contaminant, in each grid block
   - Unit costs of alternatives — excavation, different types of treatment. Includes underclassification costs (costs for not taking appropriate actions).

2. Determine maximum budget for future site characterization — Expected Value of Perfect Information (EVPI). Required input:
   - Simulated realizations of contaminant distribution across site.
   - Unit costs.

3. Determine economic value of possible future site characterization programs — Expected Value of Sample Information (EVSI) or "data worth". Required input:
   - K riged probability maps of contamination above each action level for each contaminant — for an ensemble of equally-likely site characterization outcomes for the program being evaluated.
   - Unit costs.

4. Make decision on:
   - Whether to continue site characterization
   - Which site characterization program to pursue

   Yes → Move on to site remediation
   No

Figure 6.1.3: Flow diagram of decision-making framework developed for the case study, illustrating the process for implementing the different modules of the risk-based decision model. The linkages with the M-B uncertainty model and overall site characterization process are highlighted.
6.1.3 Case study

The focus of this research was to undertake a real-world case study wherein the risk-based decision-making framework—the combined M-B uncertainty model and excavation-treatment decision analysis model—is applied to the site characterization and engineering design process, which includes the incorporation of geophysical measurements. The real-world engineering problem is soil remediation of a real estate development site, wherein soil has to be excavated, treated, and/or disposed of in a prescribed way if it contains contaminant concentrations above regulatory action levels. The primary decision is what action should be taken with the soil in different places across the site, from six possible alternatives (ordered from most costly to least):

1. excavate, treat, and dispose of soil for the highest contamination classification;
2. excavate, treat, and dispose of soil for the lower contamination classification;
3. excavate, batch test sample for contamination, and treat and dispose of accordingly;
4. perform confirmatory sampling on small-sized samples on the wall/floor of an existing excavation, and treat and dispose of accordingly;
5. excavate to reach contaminated soil beneath; and
6. leave soil intact.

Important secondary decisions are (1) whether to perform further site characterization and, if yes, (2) how to proceed with site characterization.

An underlying assumption of the analysis is that all soil with contamination above regulatory action levels has to be treated in the way required by that action level—whether at the time of planned site remediation or at some future time, when there is a cost penalty for not taking action earlier. In other words, if the contamination in soil is mistakenly underclassified there is no chance that the contamination will go unnoticed and the land containing that soil allowed to be used in a way that is unacceptable for the actual contamination level. Therefore, the risk costs in the problem are associated with misclassifying soil contamination levels—either underclassifying contamination and not abiding by a stringent enough action level or
overclassifying contamination and over designing the remediation. The sampling alternatives (numbers (3) and (4) above) eliminate the risk of misclassifying the contamination level of the soil that is sampled.

Existing information about soil contamination in the study area consists of a large number of soil samples, analyzed for contaminant concentrations. As part of the case study, Ground Penetrating Radar (GPR) and Frequency Domain ElectroMagnetics (FDEM) geophysics surveys were performed across a section of the site. The complete set of decision analyses, including evaluation of the worth of collecting GPR and FDEM geophysics measurements, are performed using contaminant distribution uncertainty models built from all the soil sample results. The same analyses, minus the data worth evaluation, are carried out after incorporation of the acquired geophysics results in the uncertainty models. Approximate unit costs for the remediation alternatives were provided by the real estate developer; however, estimates of underclassification costs—the costs of not remediating soil at the action level required for the level of contamination it contains—could not be ascertained.

The decision analysis using all previously existing soil sample data to determine the near-optimal soil remediation design is performed a number of times for different “soft” data scenarios and different values of key model parameters—the cost increase resulting from underclassification of contamination and the cost of batch sampling—to evaluate the sensitivity of the analysis results to these changes. The direct measurements of contaminant concentration from the soil samples are used to infer prior global IRV mean and covariance models for each contaminant regulatory action level that, in turn, are used in conditional indicator kriging with the same data to generate the contamination action level probability maps required for the decision model. Since the case study investigates three contaminant types of concern and the sample data for each type are not necessarily coincident, additional information about each contaminant can be ascertained from the sample data of the other two contaminants—by using the options in the M-B uncertainty model designed for incorporating indirect measurements in the inference of the global IRV mean and the conditional cokriging/simulation algorithm. In
addition, a soft "type C" data scenario is evaluated, wherein prior probabilities of contamination above action level thresholds for certain sections of the site—representing hypothetical, but not unrealistic, "expert judgment"—are incorporated in the IRV conditional updating step.

Table 6.1.1 summarizes the remediation design results obtained by executing the decision analysis model for a range of the decision model parameters and information usage scenarios described above. Immediately evident from the results is that the estimated total cost of remediation is substantial—$4 to $9 million Canadian—for the full range of costs associated with contamination underclassification. If large volume soil batch sampling, testing, and the corresponding remedial action (if required) is an available option, it is the preferred alternative for most of the site, regardless of underclassification costs (see "Batch sampling - % of total site" column in Table 6.1.1); the overall remediation cost is considerably less than when batch sampling is not an option, especially at higher underclassification costs. This is true for the batch sampling cost provided, as well as for a cost of twice that amount. If batch sampling is not an option, soil excavation, treatment and disposal for the lower action level, is the predominant alternative across the whole site (see Table 6.1.1). Except when the underclassification costs are nearly the same as present day remediation costs, all the remediation designs designate almost the whole site for excavation—near 90%.

Using co-contaminants as indirect measurements in the M-B geostatistical estimation provides little improvement in the soil remediation design. On the contrary, incorporating prior probabilities derived from "expert judgment" results in a significant remediation cost reduction, that can be attributed to a reduction in the overall average of the prior IRV means (probabilities of contamination above the action levels) across the whole site, compared to if the means are inferred from the soil sample data alone.
### Table 6.1.1: Summarized case study decision analysis remediation design results— for a range of decision model parameters and three information usage scenarios: (1) direct soil sample measurements, (2) direct and co-contaminant soil sample measurements, (3) direct and co-contaminant soil sample measurements plus prior probabilities of contamination exceeding action levels.
The sensitivity of the remediation design results to changes in the underclassification costs follows the same trend for all three data usage scenarios—hard soil sample data, hard plus soft sample data, and hard plus soft sample data plus type C soft data (prior probabilities). As shown in Figure 6.1.4, without batch sampling as an option the total remediation cost increases sharply, but with a decreasing slope, from the underclassification-to-correct classification remediation cost ratio of 1.0 to 1.25, at which point the total cost increases linearly as the ratio increases. When batch sampling is an option, the total remediation cost increases in a similar fashion as the no batch sampling scenario until a classification cost ratio of 1.25, but then reaches an asymptote, whose amplitude is positively correlated to the cost of batch sampling (see Figure 6.1.4).

**Figure 6.1.4:** Total optimized soil remediation design cost derived from decision model, for data scenario where all soil sample data—both direct and soft co-contaminant concentration measurements—are used in geostatistical estimation. The results for three batch sampling scenarios are plotted: (blue) batch sampling is not an alternative, (magenta) batch sampling is an available alternative at a unit cost $4 / cubic meter, and (cyan) batch sampling is an alternative at a unit cost $8 / cubic meter. In addition, the total cost of remediation for the scenario when there is no uncertainty about soil contamination, or Expected Value of Perfect Information (“EVPI”) is plotted (red).
Based on the uncertainty and decision model information provided by the existing soil sample data—direct and co-contaminant concentration measurements—an evaluation of the economic value of additional site characterization is performed. First, the Expected Value of Perfect Information (EVPI)—which represents an expected value estimate of what the absolute maximum site characterization budget should be—is determined from an ensemble of 25 realizations for each contamination level IRV. The realizations are generated from the M-B uncertainty model simulation program, using all direct and co-contaminant soil sample data as conditioning data. The expected value total remediation cost for the case of perfect knowledge is Canadian $5.2 million—calculated in the decision model by determining the optimal remediation design for each set of contamination level realizations and taking the average of the computed total remediation cost; EVPI is calculated from this cost value by subtracting it from the total remediation cost of the near-optimal remediation design derived using the actual existing data, as described in the previous paragraphs. The calculated EVPI varies from zero—for the scenario wherein the contamination underclassification costs are the same as the correct classification costs—to Canadian $3.8 million—for the scenario wherein batch sampling is not an option and underclassification costs are three times those when the contamination is correctly classified (see Figure 6.1.4).

Next, the expected value of different possible site characterization programs—direct contamination measurements from borehole soil samples and indirect measurements from GPR and FDEM geophysics of different data qualities—are determined from an ensemble of 25 cokriged contamination level probability grids for each contamination level IRV. The probability grids are generated from the M-B uncertainty model cokriging program—using as conditioning data 25 hypothetical measurement outcome grids and all direct and co-contaminant soil sample data; the measurement outcomes are built from simulated contamination level realizations. The expected value of the site characterization scenarios, often referred to as the Expected Value of Sample Information (EVSI), is calculated by:
Chapter 6. CONCLUSIONS AND RECOMMENDATIONS

1. determining the near-optimal remediation design for each of the 25 sets of conditional contamination probability grids,
2. taking the average of the 25 total remediation design costs for each design, and
3. subtracting the average remediation design cost from the cost of the near-optimal remediation design derived using the actual existing data.

Table 6.1.2 summarizes EVSI analysis results for two, very different, possible site characterization programs: (1) soil samples from 21 fully-penetrating boreholes distributed evenly across the site, analyzed for all three contaminant types, (2) geophysics used to indirectly delineate metals or pure phase petroleum contamination, across a range of soft data quality levels. Both characterization scenarios were analyzed for a range of contamination underclassification costs and different batch sampling scenarios. The results, presented as the cost savings in the remediation design with the new information, indicate that the borehole sampling provides negligible value—probably because the actual volume of the site sampled is very small. The same is true for the geophysics scenario, unless the data quality level (degree of correlation between the measured geophysical parameters and the actual contamination levels) is excellent and the underclassification remediation cost is at least two times higher than the correct classification cost; the maximum EVSI is $831,000 for geophysics that can detect with excellent quality the highest contamination level for metals or pure phase petroleum, if batch sampling is not an available option (see “Cost savings...Excellent quality..” column in Table 6.1.2). Thus, the EVSI analysis indicates that it is very unlikely that a future borehole sampling (on the order of tens of wells) or geophysics site characterization program would be of economic value to the soil remediation project.
<table>
<thead>
<tr>
<th>Future site characterization scenario (In addition to all existing direct and co-contaminant soil sample measurements)</th>
<th>Batch sampling an option? If yes.</th>
<th>Ratio of cost of underclassification to cost of remediation</th>
<th>Cost savings in rem. design (Can $ x1000)</th>
<th>Cost savings in rem. design Good quality measurements (Can $ x1000)</th>
<th>Cost savings in rem. design Excellent quality measurements (Can $ x1000)</th>
</tr>
</thead>
<tbody>
<tr>
<td>21 fully-penetrating boreholes, sampled across entire depth for all contaminants</td>
<td>No</td>
<td>2.0</td>
<td>N/A</td>
<td>N/A</td>
<td>$206</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0</td>
<td>N/A</td>
<td>N/A</td>
<td>$51</td>
</tr>
<tr>
<td></td>
<td>Yes ($4)</td>
<td>2.0</td>
<td>N/A</td>
<td>N/A</td>
<td>$0</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0</td>
<td>N/A</td>
<td>N/A</td>
<td>$0</td>
</tr>
<tr>
<td>Geophysics for detecting metals contamination above highest action level</td>
<td>No</td>
<td>2.0</td>
<td>$182</td>
<td>$212</td>
<td>$714</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0</td>
<td>$0</td>
<td>$4</td>
<td>$619</td>
</tr>
<tr>
<td></td>
<td>Yes ($4)</td>
<td>2.0</td>
<td>$5</td>
<td>$2</td>
<td>$10</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0</td>
<td>$0</td>
<td>$0</td>
<td>$15</td>
</tr>
<tr>
<td>Geophysics for detecting pure phase hydrocarbon contamination above highest action level</td>
<td>No</td>
<td>2.0</td>
<td>$173</td>
<td>$304</td>
<td>$763</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0</td>
<td>$4.1</td>
<td>$140</td>
<td>$831</td>
</tr>
<tr>
<td></td>
<td>Yes ($4)</td>
<td>2.0</td>
<td>$7</td>
<td>$66</td>
<td>$394</td>
</tr>
<tr>
<td></td>
<td></td>
<td>3.0</td>
<td>$0</td>
<td>$63</td>
<td>$483</td>
</tr>
</tbody>
</table>

Table 6.1.2: Summarized case study future data worth decision analysis results—for several contamination underclassification costs and three future site characterization scenarios: (1) soil sample measurements for detecting all contaminants in 21 boreholes spread evenly across the site, (2) geophysics measurements for detecting metals contamination, (3) geophysics measurements for detecting pure phase hydrocarbons contamination. Data worth is represented by reduction in remediation design cost with additional information.
Despite the unpromising conclusions of the data worth analysis, the data obtained from GPR and FDEM geophysics surveys—performed over a relatively small section of the site, as part of this study—are incorporated in the remediation design decision framework. After the data was acquired it was processed and transformed into a form and scale that could be directly compared with existing collocated soil sample data in the M-B calibration module of the uncertainty model. Based on the calibration, it was determined that both the GPR and FDEM measurements were best correlated with metals and pure phase hydrocarbons contamination; the cutoffs on GPR amplitude and FDEM conductivity that produce the best correlations with the action levels of these contaminants are chosen. The processed geophysics measurements and calibration parameters are used as conditioning data, along with all the existing soil sample data, in the M-B cokriging routine. The resulting contamination probability grids are employed in the decision model to determine near-optimal remediation designs for the cases of GPR only, FDEM only, and both GPR and FDEM combined. Table 6.1.3 shows key characteristics of the near optimal remediation design determined with the new information provided by both types of geophysics, including the overall remediation cost reduction in comparison with the information provided by the soil sample data alone; results are presented for several batch sampling scenarios and several contamination underclassification remediation costs. The cost reduction is very minor, as is anticipated from the data worth analysis; the maximum cost reduction is $44,000—for the scenario when both GPR and FDEM data are incorporated, batch sampling is not an option, and the contamination underclassification costs are three times the correct classification costs.
### Table 6.1.3: Summarized case study decision analysis remediation design results after incorporating both the FDEM and GPR geophysics measurements—for a range of decision model parameters. Data worth is defined by the reduction in the remediation design cost with the additional information provided by geophysics.

<table>
<thead>
<tr>
<th>Geophysics data used in M-B conditional cokriging (In addition to all soil sample measurement data)</th>
<th>Batch sampling an option?</th>
<th>Cost savings in rem. design resulting from geophysics (Can x $1000)</th>
<th>Highest contamination level % of total site</th>
<th>Lower contamination level % of total site</th>
<th>Batch sampling cost % of total site</th>
<th>Confirmatory sampling cost % of total site</th>
<th>Excavation only cost % of total site</th>
<th>No action cost % of total site</th>
</tr>
</thead>
<tbody>
<tr>
<td>Combined FDEM and GPR geophysics for detecting metals and pure phase hydrocarbons contamination levels</td>
<td>No</td>
<td>2.0</td>
<td>$16.8</td>
<td>0.3%</td>
<td>79%</td>
<td>N/A</td>
<td>5%</td>
<td>6%</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>$44.4</td>
<td>0.6%</td>
<td>71%</td>
<td>N/A</td>
<td>0.4%</td>
<td>18%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Yes ($4)</td>
<td>2.0</td>
<td>$0.6</td>
<td>.02%</td>
<td>71%</td>
<td>15%</td>
<td>4%</td>
<td>11%</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>$0.07</td>
<td>.02%</td>
<td>.01%</td>
<td>72%</td>
<td>14%</td>
<td>3%</td>
<td>10%</td>
</tr>
<tr>
<td></td>
<td>Yes ($8)</td>
<td>2.0</td>
<td>$11.2</td>
<td>.02%</td>
<td>2%</td>
<td>61%</td>
<td>14%</td>
<td>12%</td>
</tr>
<tr>
<td></td>
<td>3.0</td>
<td>$1.6</td>
<td>.02%</td>
<td>.01%</td>
<td>72%</td>
<td>15%</td>
<td>3%</td>
<td>10%</td>
</tr>
</tbody>
</table>

**Table 6.1.3:**

6.1.4 Summary

A case study of a soil remediation site was performed to test, evaluate, and fine tune a risk-based decision-making framework that can incorporate site characterization information containing uncertainty, including geophysics measurements and prior probabilities based on personal judgment. The framework consists of an uncertainty model, developed based on the Markov-Bayes indicator geostatistics approach [Zhu and Journel, 1993; Alabert, 1987], linked to a decision model that determines the “near-optimal” soil remediation design and the expected value economic worth of different possible future site characterization programs. The different geostatistical and decision analyses performed for the case study produce consistent and veritable results—as best as can be assessed from the information available at the time of the case study.
The soil remediation design obtained from the decision model requires costly excavation of almost the entire site, regardless of whether batch sampling is an option or not; only if the risk costs are very low (the ratio of contamination underclassification to correct classification remediation costs less than 1.25), or for the case where prior contamination action level probabilities—representing expert judgment—replace the global IRV mean across certain sections of the site, is the degree of excavation and treatment significantly reduced. The expected data worth of future site characterization contamination measurements is negligible, unless the measurements cover most of the site and have little uncertainty associated with them; even so, further site characterization is only justified if the risk costs are high (underclassification to correct classification remediation costs greater than 2.0) and batch sampling is not an option. These conclusions are obtained by using (1) the information on soil contamination provided by an objective statistical analysis of all the soil sample results and (2) the unit costs for the different soil remediation alternatives provided by the site developer. When the acquired geophysics measurements are included in decision framework, the remediation design and its total cost are little changed.

Other key findings—resulting from the case study—about the M-B uncertainty and decision analysis methodologies employed in this research are itemized below:

- *The sensitivity of the decision analysis results to the uncertainty model global “stationary” parameters*. In the case study small changes in the means inferred for the contaminant level IRVs produced significant variations in the total cost of the decision analysis near-optimal remediation design; this occurs predominantly because of changes in the probabilistic risk costs—resulting from changes in the average probability of contamination above the action levels, which correspond to the changes in the means. This sensitivity—higher than any other uncertainty model parameter—highlights the importance of using an robust, objective, and repeatable methodology for inferring the IRV means; a declustering analysis approach, that accounts for spatial biasing in measurement sampling, is recommended.
Chapter 6. CONCLUSIONS AND RECOMMENDATIONS

• The sensitivity of the uncertainty model cokriging and simulation estimation results to the methodology used in the indirect measurement data calibration. It was found that the M-B calibration [Deutsch and Journel, 1992] results—soft data conditional probabilities and misclassification probabilities—are very sensitive to the IRV means of the hard data used for in the calibration data set; in turn, the M-B geostatistical estimation results are very sensitive to these calibration parameters. A methodology was developed to renormalize the calibration data set so that the segregation of hard data containing values within the IRV classes replicates the inferred IRV means; this methodology was used in the case study analyses and seems to produce legitimate, consistent estimation results.

• The sensitivity of the uncertainty model estimation results to the number of indirect “soft” data used as local conditioning data. While it is expected that the cokriged IRV estimate will change as more soft data are used as conditioning data within the correlation range of the IRV, the influence of the soft data should reach a limit at some point as more and more soft data—outside the correlation range—are included in the estimate. It was found in the case study that there is no upper limit on the number of soft data that influence the IRV estimate, even if the data quality is very low and the data being included in the cokriging is outside the correlation range. Conceptually, there appears to be a problem with the M-B cokriging estimation when indirect data are used; this problem may emanate from the soft data calibration IRV mean inconsistency discussed above, even when the calibration is normalized using the method developed here. One way to handle this problem is to limit the number of soft conditioning data that can be used in the cokriging estimation, but the number to choose as a maximum is subjective.

• The difficulty in simulating hypothetical indirect measurement results for the data worth analysis component of the decision framework. Generating realistic, equally-probable hypothetical results, that preserve the global mean, covariance, and simulated soft data calibration models, was found to be a complicated task. Simulated realizations of the contamination level IRVs are used as a starting point of equally-likely representations of
Chapter 6. CONCLUSIONS AND RECOMMENDATIONS

reality, but how to “soften” these realizations to mimic an indirect measurement of the contaminant levels, that conforms to M-B indirect data representation, is a conundrum. In this work the approach is to generate a random number at each hypothetical data location and use it to identify a transformed indirect data value, based on the M-B calibration cdf for the simulated realization hard IRV value at that location. While this approach surely accounts for the univariate uncertainty in the soft data, it is unsure whether the resulting hypothetical indirect measurement realization conforms with the bivariate covariance model that the indirect measurement IRVs should have according to the Zhu and Journel [1993] M-B method. The case study data worth analysis results obtained by using this hypothetical indirect measurement simulation approach do appear to be consistent and realistic, though.

- **The need to use a consistent scale for all components of the framework.** In the case study it became apparent how much the spatial scale of different constituents of the decision model can vary from each other. The important scales for the case study are the measurement, engineering, and model scale; the volume of these scales vary from tens of cubic centimeters for soil sample measurements, about a cubic meter for the smallest engineering scale (the volume of soil a backhoe shovel can excavate), to about ten cubic meters for FDEM measurements. It is necessary to reconcile these scales to a common scale in order to perform an integrated decision analysis, either by upscaling or downscaling; since the engineering scale controls the decision-making and it cannot be changed, all the other scales need to be transformed to a volume of one cubic meter. A 3-d model grid with cells of this size was defined for the site and the soil sample data was partitioned/combined to fall into the grid blocks as single data points. The evenly spaced FDEM data, acquired on a one meter spacing, was interpolated to the model grid nodes using a simple inverse-distance algorithm; the high resolution processed GPR data (normalized amplitudes) was upscaled to the grid scale using simple unweighted averaging. The spatially-transformed geophysics data could then be appropriately calibrated with the equivalently-scaled soil sample data. While the M-B geostatistical estimation routine developed in this work can make cokriging/simulation
estimates at a larger scale than the conditioning data with block kriging, and possibly make estimates at a smaller scale than the data using spatially-filtered covariances (this is a prototype addition to the M-B indicator methodology), it is necessary that indirect measurement data and the direct data that are used to calibrate it in the M-B geostatistical model have the same support scale.

The case study results suggest that the M-B uncertainty model can be used effectively and efficiently for risk-based decision-making problems in which probabilities are used directly in the decision model, such as remediation of contaminated soils. The model also provides a quantitative, yet relatively simple, way for rigorously incorporating geophysics measurements, as well as any number of other types of indirect or direct measurements and prior probabilities based on personal judgment, in the decision model. Other types of hydrogeology-related problems the uncertainty framework could be applied to include delineating aquitard continuity (e.g. holes in an aquitard that permit inter-aquifer vertical flow) and determining aquifer-aquitard geometry in general (e.g. sand and clay lenses/layers); these problems can be defined in terms of indicator random variables.

While not required for this study, the M-B approach also can be used for problems in which continuous variables need to be evaluated in the decision model; continuous probability distributions (cokriging) or values of the variable (simulation) are reconstructed from a series of estimated expected values or values of IRVs, respectively. For example, realizations of hydraulic conductivity ($K$) within an aquifer can be estimated by M-B conditional simulation—using borehole hydraulic test measurements, core permeameter tests, and/or geophysics calibrated to the test measurements—then used in a flow simulation model to predict groundwater flow and its influence on an engineering design in the future; there are no restrictions on the acceptable form of the $K$ probability distribution, but a sufficient number of $K$ IRV cutoffs have to be defined, and each IRV mean and covariance model inferred, to accurately reconstruct the continuous distribution.
Thus, the M-B decision framework developed in this work provides a flexible and comprehensive method for incorporating and integrating many different types of information about a spatial variable, including indirect measurements, in a probabilistic uncertainty model that directly represents and accounts for the variable uncertainty; the uncertainty model is designed for direct linkage to a risk-based decision-making model.

6.2 Recommendations for further research

While a working version of the Zhu and Journel Markov-Bayes methodology was developed and applied in this work, further investigation and refinement is needed to address some problems that were found with the indirect measurement calibration and cokriging estimation with transformed indirect measurement data (see section 6.1.4 above for an explanation of these problems). As part of this investigation it would be very worthwhile to compare cokriging/simulation estimates generated using the Zhu and Journel [1993] methodology with corresponding estimates generated with the Alabert [1987] methodology—for a comprehensive direct and indirect measurement data set. All the computer programs written for this work have the option of using either methodology. These two methodologies are based on the same Markov-Bayes indicator cokriging theory, but have fundamental differences in the way they define the soft indicator random variables (IRVs) used to represent the secondary variable—the variable measured by indirect measurements—and how they incorporate these IRVs in the cokriging estimator; this includes using different calibration approaches for producing the probabilities that are used in the soft IRV definition. The Zhu and Journel [1993] approach was used in this work because it provides more flexibility in representing and handling indirect measurement data, permitting a different number of secondary and primary variable cutoffs/classifications. The two methodologies should produce comparable estimation results when the same primary and secondary variable cutoffs are applied to the same conditioning data.

The methodology for generating ensembles of equally-likely hypothetical indirect measurement outcomes, required for the data worth decision analysis, should be further
investigated. While it is thought the approach developed in this work is simple and fairly robust, it needs to be further evaluated and verified—in particular as related to its preservation of the inferred global statistical models for a problem, as discussed in section 6.1.4 above.

Several of the global statistical model inference routines of the M-B uncertainty framework developed for this work could be fine-tuned to be more comprehensive and efficient. An approach for including "expert judgment" local prior probabilities in the inference of the global IRV mean would make the existing mean inference routine complete, since it already incorporates declustered direct and indirect measurements. A Bayesian updating approach, that directly employs a version of Bayes' equation for conditional probability, along the lines of James and Freeze [1993], would be recommended, since it is the most complete and theoretically robust way of updating probabilities. The covariance model orthonormal residuals validation routine could be improved by automating the covariance parameter optimization procedure, using a parameter optimization algorithm; as is, this has to be performed in a manual, iterative process.

While not a focus of this research, a methodology of accounting for different indirect and direct measurement sampling support volumes in the M-B indicator cokriging and simulation algorithms was implemented in the computer routines. The methodology uses spatial filters defined for particular measurement types to transform the point covariance for the measured variable to a measurement-scale covariance. A limited number of tests were performed with the routine and qualitatively the results make sense; however, the methodology needs to undergo considerably more testing, especially since, to the author's knowledge, this approach has not been applied to indicator random variables before. The value of the methodology is that it enables the incorporation of measurements of any scale in the powerful M-B geostatistical uncertainty model.

The next step in expanding and testing the usefulness of the developed M-B uncertainty framework for hydrogeology-related decision-making problems is to apply it to continuous variables, such as hydraulic conductivity. The type of problem recommended is a real-world case
study wherein hydraulic conductivity is an important decision variable in the engineering problem. Geophysics (e.g. GPR) is evaluated and used as a site characterization technique, in combination with hydrologic testing, for estimating hydraulic conductivity in the relevant aquifer of concern; the M-B geostatistical methodology is employed for integrating the geophysics and testing measurements in a hydraulic conductivity IRV uncertainty model. Realizations of hydraulic conductivity as a continuous variable can be simulated from the IRV uncertainty model and used as input in a groundwater flow model. The decision-making context of the problem would presumably be sensitive to groundwater flow and how it affects the performance of an engineering design. The output of the groundwater flow model is used as input in an engineering design performance model, linked to a decision model wherein the expected value most cost effective engineering design is determined by performing a risk-cost-benefit decision analysis. In addition, a data worth analysis of different site characterization programs, including different geophysics techniques, can be performed—using the methodology developed in this work. Successful completion of such a case study would ascertain a valuable and flexible technique for incorporating spatially-extensive indirect measurements, most notably geophysics, in the characterization and probabilistic representation of hydrogeological properties—intended for usage in a risk-based decision making framework.

It was found in the case study that the most attractive, and cost effective, soil remediation options were the ones that involved sampling as the remediation proceeded, enabling the design to be changed in "real-time". The decision-making framework developed in this work is not designed for "real-time" iterative updating of the uncertainty model as new information is provided during the remediation; this would allow the remediation design to be updated continuously. Obviously, such an ideal approach requires very efficient updating; the methodology as developed requires a laborious two step approach: updating of the global IRV means and covariances based on the new plus all the previous information, followed by conditional local estimation with all data. Bayesian updating of prior global statistical models would improve the efficiency of the routine (see [James and Freeze, 1993] for an example of
IRV mean Bayesian updating), as would the approach used by Harvey and Gorelick [1995], wherein the covariance matrix is saved after each kriging update and used in the next iteration when more data is available. Investigating the modification of the linked uncertainty-decision framework developed in this work to enable iterative updating of the decision model is another area for further research.
BIBLIOGRAPHY


APPENDIX A
THE GEOPHYSICAL INVERSE PROBLEM

Geophysical measurements are remote-sensing measurements wherein an energy field, originating from either a natural or artificial source, travels through, and is potentially modified by, the subsurface and is recorded at a remote location either above the ground surface (surface geophysics) or in a borehole (borehole geophysics). The specific objective of the geophysical method is to determine what values and spatial distribution of geophysical properties (e.g. bulk electrical resistivity) that cause the modification in the measured energy field between source and receiver; this is often referred to as the geophysical inverse problem. After the geophysical property spatial distribution is characterized, the hydrogeological engineering objective is to determine how the geophysical property values relate to hydrogeological material properties of interest, which is often referred to as the rock physics problem. The purpose of this appendix is to examine the geophysical inverse problem in significantly more depth than covered in the main text.

A.1 Background

Since all geophysical measurements represent observations of the physical response of an energy field/pulse to a system and the interest is not in the measured response, but instead the cause of the response, in order to infer information about the system (i.e. subsurface properties) an inverse problem must be solved. In other words, geophysical measurements (the observations) are dependent on the distribution of subsurface geophysical properties (the information of interest) through some physical relationship. The physical relationship can be described (or at least conceptualized) using a forward mathematical model. The relationship can be defined in very general mathematical terms as [McLaughlin and Townley, 1996]:

\[ d = F(p) + \varepsilon \]  

(A.1)
where $d$ is the set of measurement data, $p$ is the geophysical property distribution, $\varepsilon$ is the set of measurement errors due to noise associated with each datum, and $F$ is a forward operator representing the mathematical model that maps $p$ to $d$. There are some geophysical methods wherein the physical processes are very complex and not well understood and, as a result, no accurate and/or practical mathematical models exist; this nullifies the ability to solve the inverse problem using a physical model approach, leaving empirical or statistical model approaches as the only option.

The geophysical inverse problem centers on identifying an inverse operator $G$ that maps the set of measurement data to an estimate of the geophysical property distribution:

$$p^* = G[F(p) + \varepsilon]$$

(A.2)

It is important to distinguish between the case wherein each measurement is treated separately in the inverse problem, which will be referred to as local inversion, and the case wherein all measurements are treated together, referred to as global inversion. With local inversion $d$ contains only one value and inversion is performed separately using a single measurement at a time to estimate the property distribution only within the locale of measurement. In contrast, with global inversion $d$ contains the set of all measurement values and inversion is performed using all measurements at once to estimate the property distribution across a spatial domain containing all the measurements. There are distinct advantages to treating the inversion problem as a global inversion (as is usually done in the geophysics discipline) instead of local inversion; most notably it provides an ability to possibly attain a higher resolution estimate of the geophysical property spatial variability, although the inversion can be significantly more complicated. This topic will be further addressed later in this section.

The inverse problem (i.e. identifying $G$) is encountered in many disciplines other than geophysics, including hydrology (e.g. estimating hydraulic conductivity distribution from hydraulic head measurements), signal processing, astronomy, medicine, and applied mathematics—wherever mathematical models are used to represent reality as defined by observations. Special note is made of the similarities between the groundwater and geophysical
inverse problem since this work concentrates on hydrogeology-related problems; indeed, much of the following discussion on the geophysical inverse problem is based on work directed towards the groundwater version of the inverse problem. The purpose of this discussion is not to provide a detailed, comprehensive, or theoretical description of the broad, highly developed field of inverse theory, but instead to highlight the nature of the geophysical inverse problem, the difficulties it presents, and the general approaches that have been developed for solving it, emphasizing probabilistic approaches which attempt to account for uncertainty in the parameter estimates.

Geophysical inversion (i.e. identifying $G$) can be complicated by many factors. As a result of the remote-sensing nature of geophysics, the measured energy field/pulse possibly travels through, and is therefore potentially influenced by, a large volume of the subsurface, the consequence being that it can be very difficult (or impossible) to decipher the exact circumstances (i.e. spatial distribution of geophysical properties) causing a measured response. This circumstance, which is common with surface geophysical methods, is shown in Figure A.1 (and was illustrated earlier in Figure 3.1), where the source energy travels through a large volume of subsurface material before reaching the receiver to provide a single measurement. If the goal of the measurement is to estimate the geophysical property distribution at a scale smaller than the distance between the source and receiver (i.e. the measurement scale) then, from intuition, one can see that this problem does not have a unique solution, unless further constraints are added, due to the fact that there are more unknowns than measured data. (According to Nyquist sampling theory, in order to reconstruct the variability of a function at a certain scale two or more points per scale length, or period, need to be sampled.) In a mathematical context this situation is referred to as the inverse problem being underdetermined. Most of the geophysical inverse problems encountered in practice when using surface geophysics are underdetermined due to the inherently high degree of spatial filtering (i.e. large measurement support volume) associated with "looking" into the subsurface as deep as possible from the surface. Borehole geophysics tend to be less affected by the underdetermined problem than
surface geophysics as a result of having a shallower depth of investigation, a more focused measurement, and a much higher measurement sampling density, translating into a much lower degree of spatial filtering and a much greater number of data.

Other factors that can complicate the geophysical inversion problem are the inability to calculate a direct inverse solution and instability in the solution. If $F$, the forward operator, is a nonlinear functional it may not be possible to directly identify $G$. Consequently, an indirect iterative search approach must be used to search for the inverse solution. If $F$ is highly nonlinear (e.g. nonconvex) it may be very difficult to locate the global solution using this approach, instead inadvertently locating local minima/maxima in the solution function, resulting in, possibly highly, inaccurate estimates of the geophysical property distribution. Instability of the inverse solution is characterized by the unfavorable situation where small differences in measurement results cause large changes in the estimated geophysical property distribution obtained through inversion; the consequence is that minor amounts of noise in geophysical measurements can lead to significant errors in the inverse solution.

### A.2 Well-posed inverse problems

Inverse problems that are afflicted by solution non-uniqueness and/or solution instability are referred to as being ill-posed and, in order to produce meaningful inverse solutions, the
problem formulation has to be modified to make the problem well-posed (i.e. not beset by either of these complications) [Yeh, 1986; McLaughlin and Townley, 1996]. There are three inverse problem components that can be modified in an attempt to replace an ill-posed inverse problem with a well-posed one: (1) the measurements, including imposed constraints on the values of hypothetical measurements at unmeasured locations (e.g. treat the measured dependent variable as a random variable with assumed probability distribution and spatial covariance); (2) the forward operator (i.e. the mathematical model representing the physical relationship between the geophysical property distribution and the measurements); and (3) the representation of the geophysical property spatial distribution estimate, including imposed constraints (e.g. represent the distribution as a continuous second-order polynomial function or a discrete number of blocks of prescribed sizes with bounds placed on the property values that can be assigned to each block) [McLaughlin and Townley, 1996].

If no prior structure is imposed on the spatial variability of the geophysical property of interest then the corresponding geophysical inverse problem is always underdetermined—no matter what the other circumstances of the problem are—since there are an infinite number of unknowns (associated with property variability on an infinitesimally small scale) to solve for; hence, the number of degrees of freedom in the property spatial variability has to be to reduce. There are numerous ways of imposing structure on the spatial variability in order to reduce the degrees of freedom and, thus, make the inverse problem determined; the choice depends on the requirements and constraints of the project at hand. To start with, the problem spatial domain has to be bounded (e.g. assume the volume of subsurface influencing an electrical conductivity measurement made using an electromagnetic technique is 4 by 4 meters horizontally by 4 meters vertically with the measuring device located in the center of the top surface of the block).

In the broadest sense, the spatial variability of a property can be mathematically described either discretely or as a continuous, smooth function. In fact, these two descriptions of spatial variability can be considered special cases of a more general description where the spatial variability is represented by a weighted sum of (usually linearly-independent) “basis functions,”
which are pre-designated (usually very simple) functions [McLaughlin and Townley, 1996]. By limiting the number of basis functions and weights to a finite number (which represents "parameterization" of the problem, as often referred to in the literature), the number of degrees of freedom in the inverse problem is correspondingly limited. One way to make the inverse problem determined, then, is to reduce the number of "unknowns" (parameters) in the geophysical property spatial variability description (e.g. reduce the number of basis functions) so that there are less unknowns than "knowns" (i.e. informative data).

Another obvious way of confronting an underdetermined inverse problem is to take more, properly located, measurements—hence increasing the number of knowns. The number and placement of measurements required to make a problem determined depends on the parameterization of the problem domain—indeed determinacy is co-dependent on the interplay between parameterization and measurement-taking—as well as the degree of error in the measurements.

An alternative to taking enough measurements to provide useful information about all the unknown parameters is to interpolate between measurements (i.e. estimate the dependent observed state variable at unmeasured locations). The danger with this approach is that the dependent state variable estimates may contain significant errors, leading to potentially significant errors in the inverse solution estimates of the geophysical property distribution. This danger can be somewhat alleviated by accounting for uncertainty in the state variable estimates and propagating this uncertainty through to the final inverse solution (e.g. treat the state variable as a random variable), although major assumptions about the unmeasured state variables are still required (e.g. the shape of the probability distribution, the mean, the spatial covariance).

Another way of imposing constraints on the geophysical property distribution estimate and, therefore, potentially helping to make the inverse solution unique and stable (i.e. well-posed) is to use a performance criterion against which "good" parameter estimates are determined [McLaughlin and Townley, 1996]. The definition of a "good" parameter estimate is usually based on some type of prior knowledge and is inherently subjective. Common examples
include upper and/or lower bounds on parameter estimates and desired traits in the overall parameter model such as "smoothness", "overall low magnitude", or being "close" to a prior estimate.

Performance criteria are typically implemented in the inversion process by minimizing a norm (mathematical representation of the performance index) that penalizes parameter models based on their overall deviation from "good" parameter estimates. The norm for the smoothness criterion would characterize the overall amplitude of spatial variations in the parameter model. The overall low magnitude criterion norm would characterize the overall magnitude of the parameter model; while the prior estimate norm would characterize the overall misfit between the estimated parameter model and the prior parameter model. The norm for the bound(s) criterion would be different than the others in that it would probably be constructed to flag and eliminate any parameter model solution that contains a property value outside the bound(s). In addition, all these norms would contain a term characterizing the overall misfit between the measurement model state variables \((F(p) + \epsilon)\), predicted from the forward geophysical model, and the observed state variables (true measurement values). (More details about implementing norms in the inverse problem are provided later in this subsection where inverse solution methods are discussed.)

Except for the bound(s) criterion, applying these types of performance criteria to the geophysical inverse problem should result in a unique solution, regardless of the parameterization and available data in the problem. The bound(s) criterion reduces the number of acceptable solutions to those which fall within the bound(s), but does not necessarily result in only one solution and, thus, represents more of a probabilistic constraint on property value distribution. In essence, the performance criteria impose structure on the spatial variability of the geophysical property distribution and, in doing so, they can help to stabilize the inverse solution by insuring that the solution remains "close" to a prior estimate (where "close" is determined by the definition of the norm). Therefore, they can help to make an inverse problem well-posed. However, it must be stressed that the parameter model predicted from this process is purely an
estimate that might be considerably different from the true parameter model, especially if the number of parameters being estimated greatly outnumbers the number of data.

Sometimes an ill-posed inverse problem can be made well-posed by modifying the forward operator (i.e. the mathematical model relating the measured state variables to the parameters of interest). McLaughlin and Townley [1996] provide a simple example of how a one-dimensional steady-state saturated groundwater flow problem becomes well-posed by changing the downstream head boundary from specified head to specified flux. The problem assumes hydraulic head measurements are available (or estimated) everywhere between the two boundaries, the head is specified at the upstream boundary, and the flux is specified at the downstream boundary (highly unlikely in reality, but appropriate for a simple illustration).

A.3 Conceptualization of the inverse problem

![Figure A.2: Relationship between property distribution and measurement spaces for the forward and inverse problems (after McLaughlin and Townley [1996])](image)

At this point in the discussion on the geophysical inverse problem it is useful to conceptualize the problem in terms of mappings between mathematical spaces, as is illustrated in Figure A.2. Let $P$ represent a function space containing all possible functional descriptions of the
geophysical property distribution, \( p \) (as used in equation (A.1))—the property distribution space. Let \( P_{\text{ad}} \) represent a function space containing all admissible estimates of the geophysical property distribution—the admissible property distribution space. \( P_{\text{ad}} \) is a sub-space of \( P \), made "smaller" by putting constraints on the type or form of functions that can be used to describe the property distribution. Constraints are applied by imposing structure on the spatial variability of the property distribution through parameterization (i.e. limiting the functional description to functions of a certain form) or setting performance criteria for the estimate of the property distribution to meet. For example, if \( P \) is parameterized such that the weighted sum of \( N \) defined basis functions is used to describe the spatial variability of the property distribution (as suggested earlier in this subsection), \( P_{\text{ad}} \) would span all possible functions that can be constructed from taking different combinations of basis function weights. Let \( D \) represent the \( M \)-dimensional vector space containing all possible results of \( M \) measurements—the measurement space. Both \( d \) and \( e \), as used in equation (A.1) lie in \( D \).

The forward operator, \( F \), maps the true parameter model, \( p \), from \( P \) to a vector, \( d \), in \( D \) containing \( M \) measurements with noise (see Figure A.2). This is a one-to-one (unique) mapping. In geophysics (or any discipline coping with the inverse problem) \( p \) is unknown, but of interest, while \( d \) is known; hence, the goal is to perform the above mapping in the opposite direction. The inverse operator, \( G \), obtained by solving the inverse problem, attempts to accomplish this goal by mapping \( d \) to \( p^* \), an estimate of \( p \) which lies within \( P_{\text{ad}} \). There is no guarantee that \( p^* \) is "close" to \( p \) or that \( p \) lies within \( P_{\text{ad}} \), even if the inverse problem is well-posed. The reason for this incongruity between reality and the inverse estimate is that often some of the constraints used for defining \( P_{\text{ad}} \) are based on prior, subjective judgments which are not necessarily accurate in reality, or they represent simplifications imposed to make the problem more tractable which are not necessarily physically justified (e.g. parameterization of the spatial variability in the property distribution being inverted).

The nature of the inverse mapping depends on how \( P_{\text{ad}} \) is defined. If the property spatial variability is parameterized such that the number of unknown parameters is equal to the number
of measurements, then the inverse problem becomes determined and $P^{(1)}_{ad}$ (as shown in Figure A.2) is defined by the parameterization constraints. If the data contains no errors and $F$ is invertible then a unique inverse mapping should exist between $d$ and $p^{*}_{(1)}$, where $G$ is the inverse of $F$. Usually $p^{*}$ does not coincide with $p$ since parameterization imposes structure on the spatial variability of the geophysical property. However, $p^{*}$ is often some sort of spatial average of $p$ (i.e. the spatial variability of the property distribution described at larger scale), which, depending on the application, may actually be desirable. Often, however, geophysical measurements do contain noise (as depicted by the inner circle surrounding $d$ in Figure A.2, which bounds the region in $D$ where the unknown values of the noise-free measurements lie); this results in the inability to directly invert $F$ and, consequently, the need to indirectly solve the inverse problem by minimizing a norm that penalizes the misfit between forward model predictions of measurement values and the true measurement data. Conceptually, this norm represents another constraint, which is reflected in the function space representation of the problem as a reduction in $P^{(1)}_{ad}$.

A second approach to the same inverse problem could be to not impose any parameterization of the property distribution, but instead identify a unique inverse solution by minimizing a norm that imposes a smoothness performance criterion and penalizes data misfit. Taking this second approach would result in a very different definition of $P_{ad}$ ($P^{(2)}_{ad}$ in Figure A.2), spanning a different region of $P$ (a region characterized by smooth functions) than $P^{(1)}_{ad}$, as illustrated in Figure A.2. The ultimate consequence of taking these different approaches to the inverse problem is that the geophysical property distribution estimates ($p^{*}_{(1)}$ and $p^{*}_{(2)}$ in Figure A.2 for the first and second approaches, respectively) may be very different.

### A.4 Inverse solution methods

Numerous methods have been developed for setting up and solving the inverse problem. In the most general sense, inverse solution methods can be classified as direct or indirect, deterministic (non-probabilistic) or stochastic (probabilistic), and local or global (as briefly
introduced at the beginning of this section). There are two main goals in inversion: to transform the measured state variable to the variable of interest and to decompose (or deconvolve in mathematical/geophysical parlance) the spatial filtering effect of the measurement in order to obtain higher resolution information about spatial structure. The priorities placed on these goals for a particular application should be used to determine how the solution of an inverse problem is approached. If accurate estimates of the inverted property are of much greater priority than resolving, in detail, the property's spatial variability in a problem (e.g. inverting hydraulic conductivity hydraulic head measurements in wells during a pumping test for use in a regional flow model), then an inverse solution method that can provide accurate large-scale estimates of the property distribution should be implemented. On the other hand if imaging spatial heterogeneity or small-scale subsurface anomalies is of much greater importance than obtaining accurate property value estimates for a particular application (e.g. using magnetic methods to image small buried objects such as steel drums), then an inverse solution method that prioritizes high resolution of spatial variability over accuracy in property distribution estimates should be implemented.

A.4.1 Global versus local methods

As mentioned earlier, local inversion treats each measurement separately while global inversion incorporates all the measurements at once to estimate the property distribution. The smallest scale of spatial variability in the property being inverted that can be resolved through local inversion is the measurement scale, as characterized by the spatial filtering associated with a particular measurement technique [Beckie, 1996]. This is evident from the fact that there is only one data value to inform the measurement model and, thus, there can be only one degree of freedom (unknown) in the model (i.e. one parameter describing the geophysical property distribution) in order to obtain a unique estimate of the property distribution. Although a unique inverse solution can be obtained with more than one parameter, or even if no parameterization is imposed, by minimizing a norm, the description of spatial variability in the property distribution
estimated from the solution is by no means unique or precise. In fact, an entirely different description of property spatial variability can be obtained by simply changing the norm (e.g. changing from a smoothness norm to an overall low amplitude norm could significantly increase the degree of spatial variability).

In contrast to local inversion, when using global inversion it is possible to attain a higher resolution of the property distribution's spatial variability than that associated with the individual measurement scale of a technique—if the separation between measurements is smaller than the measurement scale. The smallest scale of spatial variability that can possibly be resolved through global inversion is the network scale (the characteristic spacing between measurements—see Section 2.2.1.4). This conclusion can be deduced from a mathematical perspective by considering that, in order for the inverse problem to be determined and, thus, have the potential of having a unique solution (without applying subjective constraints), the description of the property distribution has to be parameterized to a number of parameters equal to or less than the number of measurements in the problem domain. If this parameterization is imposed by splitting the problem spatial domain into discrete blocks of constant property value, containing one measurement per block as illustrated in Figure A.3, the smallest resolvable scale of property spatial variability is easily visualized as the size of the smallest block.

![Figure A.3: Relationship between spatial resolution and measurement spacing in the global inverse problems](image-url)
Therefore, in terms of gleaning the most information about the property distribution as possible from measurement inversion, approaching the inverse problem from the global (inversion of all measurements at once), as opposed to local (inversion of each measurement separately), perspective is advantageous if the network scale is smaller than the measurement scale. However, if the model scale of the overall engineering problem—which represents the scale of parameterization used for solving the engineering problem at hand (e.g. the size of blocks used in a groundwater flow and transport model of contaminant migration from a proposed landfill), as determined based on the economic/technical limitations and the decision-making context of the overall problem (see Section 2.2.1.3)—is equal to or larger than the measurement scale, then there may be no added advantage to performing global inversion. For this situation local inversion of each measurement separately can be used to obtain a description of the smallest scale of spatial variability required in the engineering problem (i.e. the model scale)—the advantage of local over global inversion being that the solution is often much easier to ascertain.

A.4.2 Direct versus indirect methods

Direct and indirect inverse solution methods (as defined in this work) have very different mathematical approaches in that direct methods identify $G$ by calculating a mathematical inverse of $F$ (or a variant of $F$), whereas indirect methods do not calculate a mathematical inverse of $F$, instead using an estimator to optimize the forward model fit to measurements by iteratively varying the property distribution estimate. The main advantage of the direct inversion method is the ability to always identify the optimal solution to a problem setup. This implies that the inverse problem must be posed in such a way that the inverse functional (either $G$ or a regularized variant of $G$) exists and is mathematically tractable (i.e. can be computed)—usually only true for linear functionals. An example of a regularized variant of the inverse functional is one that minimizes the least squares error between the forward solution and measurements with
noise—see Yeh [1986] and Oldenburg [1984] for examples of regularized direct inversion in a groundwater and geophysical context, respectively. It is important to realize that the property distribution estimate obtained from direct inversion is not necessarily the best estimate (i.e. the obtainable estimate that is closest to reality); the integrity of the estimate depends on the structure and assumptions made in the inverse problem setup (e.g. type of parameterization, type and form of performance criteria), as well as the stability of the inverse solution.

Instability—wherein small fluctuations in measurement values or the choice of mathematical method for performing direct inversion can cause significant changes in the inverse solution—is a common problem with inverse problems, especially when measurements are contaminated with noise (as is typical in geophysics). The direct inversion approach can be particularly vulnerable to instability due to limits on the degree of regularization that can be implemented in the inverse problem (e.g. it is difficult to apply constraints on the acceptable values of particular property estimates, such as upper and lower bounds) and the inherent need to perform mathematical operations that are very sensitive to numerical errors (e.g. division by small numbers). The numerical instability problem can be easily illustrated for direct inverse problems when the relationship between the property distribution and observations is expressed as a matrix equation and direct inversion constitutes inverting a matrix (which is true of most direct inversion problems)—a problem representation amenable to computer solution. If the matrix being inverted is nearly singular (as indicated by very small singular values, as determined from singular value decomposition, or very small diagonal pivots after the matrix is transformed to an upper diagonal matrix), the matrix inversion process will be highly susceptible to numerical errors and the inverse solution possibly sensitive to small measurement errors.

Indirect inversion represents an optimization process in which there is significant flexibility in how the optimal property distribution estimate is chosen and significant generality in the types of inverse problems that can be handled. Mathematical optimization represents the process of identifying the optimal value of a variable based on certain defined criteria, and possibly subject to constraints. The criteria are usually posed as satisfying an overall objective by
minimizing/maximizing a single performance index (or norm) or a combination of norms that are given different importance (by using weights). The function incorporating all performance criteria is referred to as the objective function or estimator. The solution to an optimization problem can be highly dependent on the estimator used, particularly the norm(s) incorporated in the estimator and the weights given to different norms.

In context of the inverse problem, there is always a norm defining the misfit between forward model predictions of the observed state variable and the actual observations, referred to as the residual (e.g. the least squares error norm, which minimizes the sum of all residuals after they are squared). In addition, regularization norms (e.g. a measure of the deviations of estimates from a prior model of the property distribution) are also often included in the optimization to make the inverse problem more stable. It is important to realize that even the choice of measurement misfit norm (e.g. the sum of the absolute value of the residuals as opposed to the least squares error) can result in very different inverse solutions; the appropriateness of one norm over others is often subjective and difficult to physically justify.

Once the estimator is constructed the indirect inverse problem becomes one of solving an optimization problem; the choice of solution method is multitudinous, although somewhat limited by the characteristics of the problem (e.g. linearity/nonlinearity of forward operator, convexity/nonconvexity of estimator). If the forward operator in the inverse problem is linear then the optimization problem can be solved analytically in one iteration (e.g. using linear programming techniques) with 100% chance of finding the optimal solution, if it exists. However, in the real-world the dependency of measured state variables on property values is usually nonlinear—requiring that, either, the inverse problem is linearized based on certain approximations, or a nonlinear, iterative optimization solution method is used to search for the optimal solution of the indirect inverse problem. Nonlinear optimization solution methods are much more computationally-intensive than linear ones due to their iterative nature and, unlike linear methods or the direct inversion approach, have the potential of not finding the optimal
(albeit not necessarily best) inverse solution if the estimator function surface is nonconvex (i.e. contains local minima in addition to a global minimum).

Both linear and nonlinear optimization approaches can be hindered by instability of the inverse problem, which manifests itself in the optimization problem as a flat estimator function surface (i.e. the estimator is relatively insensitive to changes in the property distribution estimate); but with indirect inversion, unlike with direct inversion, a variety of regularization conditions (based on prior knowledge, scientific judgment, and/or problem requirements) can be imposed to make the problem more stable. In addition, indirect inversion using nonlinear optimizers is the only way to solve nonlinear inverse problems without making approximations to the governing relationship between property values and observations.

A.4.3 Accounting for uncertainty in the inverse solution

In order to apply geophysics to engineering problems in a risk-based decision-making framework (as is advocated in this work), uncertainty in the property value estimates obtained by solving the geophysical inverse problem must be quantitatively accounted for. However, the inverse problem in geophysics is most commonly solved using deterministic, as opposed to stochastic, approaches; therefore, uncertainty is not rigorously accounted for in the property distribution estimate. Uncertainty in the inverse solution is usually only qualitatively addressed through "exploration" of the property distribution space (i.e. examining the effects on the estimate of solving the problem using different norms, constraints, and/or solution techniques) and appraisal of the resolving power and average property value estimate at particular locations. Much more rigorous stochastic approaches to inversion have been attempted in other fields; most noteworthy to the author are the stochastic approaches taken in hydrogeology to solve inverse problems such as estimating the hydraulic conductivity distribution from hydraulic head measurements—taken when the groundwater system is stressed—and/or solute concentration measurements—taken during a tracer test. The following subsections provide an overview of the different types of approaches and specific methods that have been developed to address
uncertainty in the inverse problem, highlighting their advantages and limitations. Extra emphasis is given to the geostatistical approach since it is the approach used in this work.

**A.4.3.1 Backus-Gilbert approach**

*Oldenburg* [1984] recommends a three step approach to the geophysical inverse problem based on the methods developed by Backus and Gilbert [1967, 1968, 1970]: model construction, appraisal, and inference. In model construction a number of diverse property distribution models are constructed using different inverse solution methods based on different norms and types of constraints. Only the features present in all constructed models can be considered likely to exist in reality. If the models differ significantly a best guess estimate should be constructed based on the type of model most desired (e.g. smallest, flattest, or highly discontinuous), using the most appropriate norm and including as much additional information as possible.

*Oldenburg* [1984] then recommends using Backus-Gilbert appraisal to assess the average property distribution model value at particular locations in the problem domain, representing an average of all possible inverse problem property value estimates at that location, and the standard deviation of the computed average; uncertainty is associated with the average if the measurements contain error. In addition, Backus-Gilbert appraisal can be used to calculate the spatial averaging function associated with the above estimation of average property distribution model values—of the spatial filtering required to obtain that estimate—for a certain standard deviation of the average model estimate at that point, providing useful insight about the resolving power of the model. There is a tradeoff between the resolution of the model point average estimate and the accuracy of that estimate, even when the measurements are error-free; albeit, there is a lower limit on attainable resolution as determined by the network scale (see Section 2.2.1.4). If the measurements are error free, the computed average model value is unique, although not necessarily well resolved (as identified by the shape of the averaging function at that location). The methodology incorporated in the appraisal approach is set up to identify the inverse estimation averaging function that produces the most accurate average property distribution model value at a point. It is not devised in a way that is convenient for calculating
the uncertainty associated with a certain specific resolution (as characterized by an averaging function). In addition, only simple statistical models of measurement error can be efficiently incorporated in the appraisal.

If the resolving power of the experiment is found to be poor—as is common for many surface geophysics experiments—and, thus, it is difficult to glean much information from the computed average property distribution model values attainable from inversion, Oldenburg [1984] suggests using a so-called inference approach to obtain more information. The upper and lower bounds on the average model property values for different characteristic scales of resolution (based on the width of a boxcar filter) and data accuracy can be computed using the inference approach. Unlike the appraisal approach, in the inference approach the estimation averaging function (which characterizes the resolution of the property distribution model) is exactly specified, from which quantitative bounds on the uncertainty in the resulting average model property value estimate at a point are calculated. In addition, the approach can incorporate physical constraints on acceptable inverse estimate values (e.g. positivity, upper and lower bounds on acceptable property distribution estimates) as well; this helps to better constrain the inverse solution, as indicated by a reduction in the distance between the computed upper and lower bounds on the average model values.

The emphasis of the geophysical inverse problem approach taken by Oldenburg [1984] is to identify the salient features of the property distribution that can be ascertained from the geophysical inversion while recognizing the fact that other features of the distribution, apparent in certain inverse models, can not be properly resolved through the inversion. Oldenburg [1984] generally ascribes to a non-parametric statement of the inverse problem (or if parameterization is required the problem is kept well underdetermined by using a sufficiently fine partitioning) and limits his examples to linear inverse problems, but suggests that the methodologies can be extended to nonlinear problems—albeit with some computational and theoretical limitations (especially as relates to computing accurate estimates of average model values).
Vasco et al. [1997] use an approach similar to the Backus-Gilbert appraisal approach outlined by Oldenburg [1984] to quantitatively examine the trade-off between spatial resolution and uncertainty in model parameter estimates for the groundwater inverse problem of estimating hydraulic conductivity \( K \) from hydraulic head and tracer concentration measurements collected during pressure and tracer tests, respectively. They parameterize the inverse problem by representing the \( K \) distribution as a finely-spaced grid of equally-sized cells, each cell having a single \( K \) value. Since a fine grid spacing is used the inverse problem is inherently underdetermined (i.e. there are more unknowns than data). They use a two step indirect inverse estimation process wherein, first, nonlinear optimization is used to minimize the sum of two weighted least squares norms defining data misfit and parameter model roughness (determined from finite difference approximations of the spatial derivatives of \( K \)) in order to obtain an initial estimate of the \( K \) parameter structure. In the second step the initial \( K \) structure estimate is used as a background model from which small deviations are made in order to fine-tune the model and, more importantly, assess the model resolution and parameter uncertainty.

In order to estimate parameter uncertainty and resolution and the inherent trade-off between the two, either a Monte Carlo approach must be used or the problem has to be linearized by making approximations to the nonlinear physical model; the latter approach requiring the calculation of a sensitivity matrix relating a change in each \( K \) parameter to a corresponding change in a measurement observation (the approach taken by Vasco et al. [1997]). The metrics used for assessing spatial resolution and model parameter uncertainty of inverse solution estimates are the so-called resolution matrix and the model parameter covariance matrix, which are estimated from a generalized inverse that incorporates the sensitivity matrix. In order to assess the trade-off between \( K \) distribution model resolution and uncertainty a generalized inverse that minimizes the weighted sum of overall data misfit, resolution spread (ascertained from the resolution matrix), and parameter covariance size (ascertained from the parameter covariance matrix) norms is calculated, from which a trade-off curve is constructed by varying the weights (referred to as the trade-off parameters) associated with the resolution and
covariance norms. The authors (Vasco et al. [1997]) only consider the trade-off between fitting the data and optimizing the resolution, although after an optimal point on the trade-off curve is identified, they examine parameter covariance to determine if it is reasonable; if not, another value of the trade-off parameter can be tried.

The inverse assessment approach employed by Vasco et al. [1997] is especially useful in that it provides a quantitative estimate of not only the uncertainty associated with individual parameters defining the property distribution, as represented by the parameter covariance matrix, but also the inverse model resolution and the trade-off between the two. The method has the added advantage of not placing any restrictions on the statistical structure of parameter uncertainty (e.g. normal probability distribution with exponential spatial covariance) since the uncertainty is represented by a covariance matrix. The major limitations of the approach are that the mathematical formalism requires the relationship between the property distribution of interest and observations to be linear, necessitating linear approximations of the forward operator if it is nonlinear, and that the methodology is computationally intensive, requiring the computation of very large matrices.

A.4.3.2 Minimum relative entropy approach

Minimum relative entropy inversion methods can be used to compute the posterior pdf and expected value of the parameterized property distribution for any parameter in the inverse model based on prior information, measurements, and physical constraints (as defined by the forward physical model) [Woodbury and Ulrych, 1996]. The approach is usually applied to highly parameterized and, thus, underdetermined (i.e. more parameters than observations—resulting in solution nonuniqueness) inverse problems. The parameters used to describe the property spatial distribution are treated as random variables. The essence of the approach is an optimization problem: the minimization of the entropy of the posterior pdf of the property distribution model relative to a prior pdf of the model based on expected value constraints. The expected value constraints are: (1) the expected value of the posterior pdf multiplied by the forward physical model operator evaluated for each separate property distribution model parameter is equal to the
expected value of the observed value of that parameter and (2) the integration of the posterior pdf over all parameters is equal to one [Woodbury and Ulrych, 1996]. Therefore information that needs to be supplied to implement this method are a prior pdf of the parameterized property distribution model (e.g. upper and lower bounds on acceptable seismic velocities, representing a uniform distribution, for grid blocks in a vertical 2-d block model of seismic velocity spatial distribution), measurements of the observed state variable (e.g. seismic travel times between transmitters and receivers in different boreholes), and a parameterized forward operator (e.g. seismic tomography forward operator).

The formulation developed by Woodbury and Ulrych [1996] allows the exact form of prior pdf to be an unknown; only upper and lower bounds for all the property distribution model parameters and available estimates of specific model parameters, based on information (e.g. direct measurements of the property value of interest, expert judgment) obtained prior to inversion, are required. The prior model parameter estimates are assumed to be the expected value of a pdf which is chosen so that it has minimum relative entropy to a uniform pdf defined by the supplied upper and lower bounds.

The main limitation of the approach presented by Woodbury and Ulrych [1996] is the required mathematical form of the forward physical model operator; it has to be linear and must be represented as a Fredholm integral equation of the first kind (a functional inner product). Most forward physical models for geophysical problems are nonlinear.

The outstanding attribute of minimum relative entropy inversion is its ability to fully characterize parameter uncertainty associated with solution of the inverse problem—yielding exact expressions for the posterior pdf of the property distribution model [Woodbury and Ulrych, 1996]. In addition, prior information about the property distribution can be incorporated through a prior pdf; a noninformative prior pdf (i.e. a uniform distribution with wide bounds) can be used if no prior information exists.
Another way to quantitatively account for uncertainty in the inverse problem solution is to use a geostatistical approach where the property distribution of interest is treated probabilistically in terms of spatial random variables and estimates of the property distribution and their associated uncertainty are derived from statistical/probabilistic models—an approach that has been widely applied to the groundwater inverse problem, but that has seen little application in the geophysical inverse problem. In essence, geostatistics refers to the application of random fields to describe the spatial variability of some property in the subsurface, usually accounting for major spatial trends (the statistical first spatial moment) and spatial correlation (the statistical second spatial moment) of the property (i.e. the tendency of subsurface property values to be more similar when they are closer together). In geostatistics the unknown property distribution is represented by a probability distribution function (pdf) at every location in the problem domain and a measure of spatial correlation (e.g. auto-covariance) between the property value at each location and those at all other locations, the entirety of which represents the property distribution random field. The property distribution random field can be thought of as an ensemble of equally-likely realizations of the property spatial distribution, of which one of the realizations is the true distribution. (Note: CHAPTER 4 and APPENDIX B provide a much more detailed discussion on geostatistics and the geostatistical estimation process. The description presented here is only designed to provide a context for understanding the geostatistical approach to the inverse problem.)

The shape of the pdf and spatial correlation relationship at each location is dependent on prior assumptions about the random field structure, the value of measurements, and the correlation relationship between measurements and property distribution values (as represented by cross-covariance, which is integrally different than the auto-covariance between property values at different locations) through a statistical estimator (e.g. conditional cokriging). If a measurement is exact and direct (i.e. hard), the property value pdf at that location will collapse to a single value with 100% probability of occurrence (i.e. no uncertainty) and the property uncertainty in the region surrounding the measurement will be reduced (the amount of reduction
depending on the auto-covariance). However, if a measurement is inexact and indirect (i.e. soft), as is the case in the inverse problem, the pdf at that location will potentially be narrowed, but will most likely retain some uncertainty (the amount of uncertainty reduction depending on the cross-covariance) while the surrounding uncertainty in the property distribution may be reduced, but not to the same degree as with a hard measurement.

The geostatistical estimation process, including using a geostatistical approach for estimating the inverse problem solution (e.g. conditional cokriging), is (at least) a two step procedure. First the initial random field model representing the overall uncertainty and spatial structure of the property distribution must be defined, after which the random field is locally updated based on existing measurement data using a geostatistical estimator. The first requirement entails selecting a random field model, as represented by mathematical functions (e.g. a normal property value pdf whose mean varies with spatial location according to a defined trend and exponential property value spatial auto-covariance), and inferring the statistical parameters, or coefficients in the mathematical functions, associated with that model (e.g. property value mean, variance, and spatial correlation length for the above example random field model)—using as much available information as possible, including measurements and expert knowledge. The random field model usually describes the first two statistical spatial moments of the random variable (which represents the property distribution for the inverse problem)—the mean and covariance—ignoring higher order statistical moments. An intrinsic assumption made in the initial spatial random field model selection process is that the mathematical functions used to describe the model apply across a certain region (usually the entirety) of the problem domain (although the mathematical function, itself, may vary with location). This inference process is inherently subjective, based on a modeler's understanding of the phenomenon as well as data, although the selected model(s) should be properly validated to assure that it is consistent with the available information [Kitanidis, 1991].

As an example of selecting an initial random field model for a subsurface property spatial distribution, reference is made to the hypothetical problem introduced at the beginning of
CHAPTER 2. Based on the statistical analysis (i.e. computation of experimental statistics such as histograms, arithmetic means, and covariograms/semivariograms—see CHAPTER 4) of $K$ data obtained by performing permeameter tests on core samples and prior experience at similar sites, it might be found that a lognormal probability distribution model with constant mean across the site and isotropic (i.e. no directional dependence) exponential covariance model can successfully describe the spatial variability and correlation, respectively, of $K$ in the unconfined aquifer. In this case the statistical parameters that have to be inferred in order to define the shape of the mathematical functions representing the random field model are the mean, variance, and correlation length of $K$.

The computational process in the initial random field model inference process that is specific, and potentially most troublesome, to the application of geostatistical estimation to the inverse problem is the determination of the cross-covariance between the observed state variable (measurements) and property distribution and the observed state variable auto-covariance; these covariances are directly dependent on the physical relationship between the observed state variable and the property of interest for a particular problem. It is potentially possible (depending on the specifics of the inverse problem) to estimate these covariances using a purely experimental statistics approach—as typically done when estimating the initial property value auto-covariance (see above)—by computing experimental cross/auto-covariograms (or semivariograms) from measurement data and fitting mathematical models to them. However, that approach neglects very useful information that is already known about the specific relationship between the dependent (observed state variable) and independent (property distribution) variables through the forward problem. Since this physical relationship is mathematically described through the forward operator, theoretically the desired covariances should be derivable from the forward mathematical model. There are two approaches for estimating the cross-covariance from the forward model: Monte Carlo simulation and first-order (or small perturbation) approximation.
With Monte Carlo simulation a set of equally-likely realizations of the property distribution are generated using a geostatistical simulator (see CHAPTER 4 for a discussion of Monte Carlo geostatistical simulation) based on an inferred initial random field model of the property distribution (see above) and conditioned on direct property value measurements (if they exist). For each realization a forward model prediction of the observed state variable across the problem domain is simulated. A statistical analysis is performed on all the forward model simulations to estimate the cross-covariance between the property distribution and observed state variable, observed state variable auto-covariance, and observed state variable mean (if required) using simple approximation equations such as:

\[
D_m(x) = \frac{1}{N} \sum_{i=1}^{N} d_{m(i)}(x) \tag{A.3}
\]

\[
C_{D_mD_m}(x, x') = \frac{1}{N} \sum_{i=1}^{N} (d_{m(i)}(x) - D_m(x))(d_{m(i)}(x') - D_m(x')) \tag{A.4}
\]

\[
C_{D_mP}(x, x') = \frac{1}{N} \sum_{i=1}^{N} (d_{m(i)}(x) - D_m(x))(p_{(i)}(x') - P(x')) \tag{A.5}
\]

where \( p_{(i)} \) and \( d_{m(i)} \) are the property value and simulated observed state variable, respectively, for realization \( i \) at location \( x \); \( P(x) \) and \( D_m(x) \) are the mean function values for the property distribution and observed state variable, respectively, at location \( x \); \( C_{D_mD_m}(x, x') \) is the auto-covariance between observed state variables at locations \( x \) and \( x' \); \( C_{D_mP}(x, x') \) is the cross-covariance between the observed state variable at location \( x \) and the property distribution value at location \( x' \); and \( N \) is the number of realizations.

The primary advantage of the Monte Carlo approach is that, as long as the forward problem can be modeled, there are no limitations as to the mathematical form of the forward operator (e.g. nonlinear functionals) required to proceed with the approach. The only required approximations and assumptions are those associated with solving the forward problem (e.g. discretization of the problem domain, defined boundary conditions) and the finite number of simulations used in the covariance estimation (i.e. to be truly representative of the spatial variability the number of simulations should be infinite). The main limitations of this approach are that typically only direct property value measurements, excluding the usually more abundant
indirect measurements, are used to infer the initial property distribution random field model and the approach requires intensive computational effort to generate a large enough number of Monte Carlo simulations to compute robust statistics.

In the first-order approximation approach the cross-covariance between property values and observations is directly calculated from the forward operator and the inferred initial random field model of the property distribution (see above) by linearizing a stochastic representation of the forward operator and then deriving the second moments. The idea is to calculate the second moments of a dependent variable (e.g. measurement observations) from the first two moments of an independent variable (e.g. geophysical property distribution) based on a functional relationship between the two, where both variables are treated as random variables. The principal steps required to carry out this calculation are (in addition to inferring the initial random field model of the property distribution described earlier): (1) representation of the observed state variable and property distribution as random variables; (2) estimation of the transformation of the governing mathematical equation(s) (forward operator) relating the two variables to account for the fact that the variables are now random variables, making necessary approximations in order to facilitate solution of the equation(s) in the next step; (3) solution of the forward equation(s) for the dependent (observed state) variable—using approximations, if necessary; and (4) calculation of the desired statistical second moments (cross-covariance and observed state variable auto-covariance) from the forward solution, which is greatly simplified if the relation between dependent and independent variable in the solution is linear.

A common and simple way to represent random variables is as a sum of a deterministic and stochastic components: \( Y(x) = F(x) + f(x) \), where \( Y(x) \) is the random variable, \( F(x) \) is the mean function (the deterministic component), and \( f(x) \) is the fluctuation of the variable about the mean function (the stochastic component). Both the components (and thus the random variable) are functions of space. For the inverse problem two random variables have to be defined—one for the property distribution and one for the observed state variable. If the fluctuations about the mean are assumed to be "small" compared to the mean term, the
fluctuation (or first-order) terms can be dropped after substitution of the random variable representations into the forward model equation. This approximation leaves a deterministic equation with only mean value (or zero-order) terms; the equation can be solved for the mean observed state variable function based on the inferred initial estimate of the mean property value function using the usual solution methods for solving the deterministic forward equation (e.g. finite differences or finite elements if the forward equation is in the form of a differential equation).

Once the mean observed state variable function is calculated, it can be considered a known, along with the mean property distribution function, in the random variable representation of the forward model where, this time, the first-order terms are retained. The objective at this point in the process is to manipulate the stochastic forward equation such that the first-order measurement state variable fluctuation is an explicit dependent variable; this corresponds to moving all terms containing those fluctuations to the left-hand side of the equation, somehow combining them into one term, and moving all other terms to the right-hand side—essentially an open form solution of the stochastic forward equation for the state variable fluctuation. The purpose of posing the stochastic forward equation in this form is that the mathematical expressions for the second moments can be easily derived by taking the expected value of both sides multiplied by the fluctuation in the property value (for the cross-covariance between the observed state variable and property distribution) or the fluctuation in the observed state variable (for the observed state variable auto-covariance):

\[ E[d_m(x)p(x')] = E[F_s(D_m(x), P(x), p(x))p(x')] \quad (A.6) \]

\[ E[d_m(x)d_m(x')] = E[F_s(D_m(x), P(x), p(x))F_s(D_m(x'), P(x'), p(x'))] \quad (A.7) \]

where \( E[*] \) represents the expectation operator (\( E[ab] \) being then the second statistical moment, or cross-covariance, between \( a \) and \( b \); \( D_m(x) \) and \( P(x) \) are the mean functions for the observed state variable and property distribution, respectively; \( d_m(x) \) and \( p(x) \) are the fluctuations about the mean functions of the observed state variable and property value, respectively; \( F_s \) is the stochastic forward operator that maps \( D_m(x) \), \( P(x) \), and \( p(x) \) to \( d_m(x) \); and \( x' \) represents a
different location than \( x \). Equation (A.6) is the expression for the cross-covariance between the observed state variable and property distribution variable and equation (A.7) is the expression for the auto-covariance of the observed state variable.

Unfortunately, the expectations on the right-hand side of equations (A.6) and (A.7) are unknown. The goal is to manipulate the right-hand side so that it only contains knowns; this requires that the only expectation on the right hand is \( E[p(x)p(x')] \) (the property distribution auto-covariance estimated in the initial inference process), which entails that the forward operator has to be linear in order to simplify the right-hand side expectation. For many problems this simplification requires linearization of a nonlinear forward operator—often accomplished using numerical approaches that discretize the spatial domain. After this simplification the expressions for the second statistical moments become:

\[
E[d_m(x)p(x')] = F'(D_m(x), P(x), E[p(x)p(x')]) \tag{A.8}
\]

\[
E[d_m(x)d_m(x')] = F''(D_m(x), P(x), D_m(x'), P(x'), E[p(x)p(x')]) \tag{A.9}
\]

where \( F' \) is a functional representing the expected value of a linearized version of \( F \) multiplied by \( p(x) \) and \( F'' \) is a functional representing the expected value of the same linearized version of \( F \) multiplied by itself (evaluated at a different location). Now the problem has been reduced to evaluating these modified forward operator functionals, either analytically or numerically, to obtain estimates of the desired auto- and cross-covariances.

The main advantages of the first-order approximation approach over the Monte-Carlo approach for estimating the observed state variable statistical second moments directly from the forward operator are that it is much more computationally efficient and both indirect and direct property value measurements can potentially be used to infer the initial property distribution random field model—taking advantage of formulations already required for the dependent variable second moments estimation. Elaborating on the latter advantage, the initial spatial statistical parameter estimates of a predesignated mathematical model describing the property distribution random field can be objectively selected such that they maximize the likelihood of both direct and indirect property value measurements using the maximum likelihood method.
[Kitanidis and Lane, 1985]. The likelihood is represented through a joint likelihood function (pdf) for all measurement values, usually selected as a Gaussian pdf (even if the selected structure of the property distribution random field is non-Gaussian), wherein the statistical moments of the observed state variable are related to the property distribution moments (the ones being estimated) using the formulations developed in the first order approximation. See Harvey and Gorelick [1995] for an example of using the maximum likelihood method in this manner to estimate the statistical spatial moments of the initial property distribution random field model.

The primary limitation of the first-order approximation approach is the many approximations (e.g. linearization, dropping of terms in the stochastic forward equation) needed to calculate the observed state variable statistical second moments, resulting in potentially significant errors in the estimates of these moments. In addition, the calculated covariances can be very large and cumbersome to deal with—requiring significant computer memory—if the problem domain is large and/or highly parameterized.

It is important to realize that the covariances estimated by both the Monte Carlo and first-order approximation approaches for estimating the observed state variable statistical second moments from the forward operator can be (and often are) non-stationary; this means that the estimated correlation between the observed state variable, or the observed state variable and property value, at two locations can be location-dependent. These methods take a global approach to the covariance estimation process in that the influence of every part of the problem spatial domain on the observed state variable is accounted for by solving the global boundary value forward problem, but estimate the desired covariances for all pairs of localized locations in the domain separately. In essence, these methods estimate covariance by perturbing the property value from its mean value at every location in the problem domain, one at a time, and calculating the change (through the forward model) in the predicted observed state variable at all possible measurement locations resulting from the perturbation—thus providing a sensitivity measure. The covariances are determined directly from these sensitivity measures; the result is location-dependent covariances with no stationarity restrictions and, thus, the addition of localized
information about the observation-property value relationship to the geostatistical estimation process. This additional information potentially reduces uncertainty in the probabilistic inversion process as compared to using strictly stationary (i.e. location-independent) observed state variable auto- and cross-covariances.

As mentioned at the beginning of this subsection on calculating the observed state variable second statistical moments, it is also potentially possible to estimate these moments strictly from measurements using an experimental statistics approach—thus ignoring the mathematical relationship between the property distribution and measurement observations. The experimental statistics approach consists of three principal steps for estimating each required covariance (that is the cross-covariance between the observed state variable and the property distribution and observed state variable auto-covariance):

1. use simple statistical formulas to calculate an experimental auto/cross-covariogram (or semivariogram) from measurement data pairs at a variety of different separations (indirect-direct pairs for the cross-covariance and indirect-indirect pairs for the auto-covariance);
2. select a mathematical model (e.g. exponential covariance function) and set of parameter values defining the model (e.g. sill, nugget, correlation length) to best fit the experimental covariogram (usually accomplished using graphical fitting); and
3. validate that the model accurately represents the spatial variation of the data.

All three steps are appropriate for calculating the measurement state variable auto-covariance, but the cross-covariance relation can often be complex (e.g. unsymmetrical, having both positive and negative values) and, thus, difficult to fit graphical. It is not always the case that this relation has a complex form, as shown in the case study by Parks and Bentley [1996] where the calculated experimental cross-covariance between the natural log of total dissolved solids in the groundwater and the natural log of electrical ground conductivity as measured by an electromagnetic instrument is closely fit by a simple spherical model. In cases where the cross-covariance relation is complex, validation techniques, such as the method of orthonormal
residuals [Kitanidis, 1991], must be mostly relied upon to check and help select an appropriate postulated functional model to represent the cross-covariance relation.

As discussed in CHAPTER 3 (Section 3.4.1.2.2), alternative cross-covariance calculation methods have also been developed to cope with complexity in the dependent-independent variable spatial relation based on inferring the representative cross-covariance model indirectly from known or easily derived auto-covariance models. These methods include the method of auxiliary variables (see Myers [1982] for a specific application) and methods that estimate the dependent variable covariance and dependent-independent variable cross-covariance models solely from the inferred independent variable auto-covariance model and some sort of calibration model relating the two variables. Elaborating on the latter set of methods, the calibration model characterizes the relationship between the dependent and independent variable at the same location and is derived by performing a statistical analysis of collocated direct and indirect measurements of the independent (property value) and dependent (observed state variable) variables, respectively. Ideally the calibration data set is from the site being investigated, but sometimes it may be necessary to use calibration data from an earlier study in a neighboring area or in an area with similar site conditions.

Cassiani and Medina [1997] define a calibration model based on the results of a statistical linear regression between the measurements of the two variables. The desired covariance models are simple functions of the independent variable auto-covariance model, the regression slope, and the regression variance—the functional relationship determined by taking the expected value of the regression equation squared. They use a maximum likelihood procedure to objectively estimate all the statistical parameters that define the covariance and linear regression models at once. This method relies on the approximation that the relationship between the mean of the dependent and independent variables is close to linear since linear regression is used to describe the relation.

The Markov-Bayes method treats both the independent and dependent variables as a set of indicator random variables defined by threshold values or class membership, which enables the
Appendix A. THE GEOPHYSICAL INVERSE PROBLEM

derivation of the dependent variable indicator auto-covariances and the indicator cross-
covariances directly from the independent variable indicator auto-covariances based on only a
few basic assumptions [Alabert, 1987]. From the general assumption that if a direct (or hard)
measurement already exists at a location, a collocated indirect (or soft) measurement has no
influence on the value of the independent variable (property value) at nearby points (i.e. the hard
measurement prevails over the soft), Alabert [1987] shows that the desired indicator covariance
functions at each indicator threshold level or class is simply the independent variable indicator
auto-covariance function for the matching indicator level/class multiplied by a rescaling factor.
The rescaling factor is derived from likelihood probabilities between the dependent and
independent variable random variables obtained through a calibration between collocated direct
and indirect data. An advantage of this method is that there are no restrictions on the acceptable
form of the pdf (e.g. normal) describing the random variability of the independent variable due to
the use of indicator random variables, whose expected values collectively represent a discrete
approximation of a continuous cumulative pdf (the degree of discretization defined by the
number of indicator threshold levels). The Markov-Bayes method is discussed in much greater
detail in CHAPTER 4 and APPENDIX B.

The form of the covariances estimated using the experimental statistics approach to
estimating the dependent-independent variable cross-covariance and dependent variable auto-
covariance is necessarily stationary across the spatial domain encompassing all the
measurements used to calculate them since the invoked statistical formulas compute an average
covariance/variogram value for all measurement pairings at similar separations (and similar
orientations, if desired). Indeed, according to probability theory, taking the average covariance
value is required since the observations represent single realizations of a random variable (the
random variable being the observed state variable), which, by definition, has a degree of natural
random variability. This randomness in the observations leads to variability in the
experimentally-derived covariance estimates, which can only be smoothed out to converge on
the true covariance by averaging over many measurement pairings—thus treating the overall
covariance relationship as spatially invariant (i.e. stationary). As a consequence, localized spatial variations in the physical relationship between the measurement state variable and property value are filtered out and, thus, can not be accounted for when applying the experimental statistics approach to estimate the measurement state variable second moments in the inverse problem.

Since experimental statistics calculations require the determination of averages across many measurements, it is necessary to have a sufficiently large number of both direct and indirect measurements at many different separations—preferably including collocated indirect and direct measurements (a requirement for some geostatistical estimation techniques, such as the Markov-Bayes method)—in order to construct informative experimental covariograms from which covariance models can be inferred with a reasonable degree of confidence.

In essence, all experimental statistics methods for calculating cross-covariances perform a calibration between indirect and direct measurement values to determine the average degree of correlation between the two; consequently, the use of this approach for addressing the inverse problem is only appropriate if the spatial support scales (i.e. characteristic volume of subsurface material that has the most influence on the measurement) of the direct and indirect measurements are similar or regularization is performed to make them so. For example, in a seismic tomography survey a seismic signal is transmitted at different depths in one well and the travel-times required for the signal to reach an array of receivers in another well are measured. Since the tomographic travel-time measurement is influenced by the seismic velocity distribution along the entire length of the seismic signal path between source and receiver, the corresponding measurement support scale is approximately equal to the distance between wells. Therefore, calculating an experimental cross-covariogram between tomographic travel-time and seismic velocity based on tomographic travel-time measurements and direct seismic velocity measurements obtained from geophysical well logs (which typically have a support scale on the order of tens of centimeters) is not very insightful, and possibly misleading, due to the discrepancy between the measurement scales. Experimental statistics approaches do not rigorously address the spatial filtering aspect of the inverse problem.
Both the forward model (i.e. Monte Carlo and first-order approximation methods) and experimental statistics approaches to estimating the observed state variable statistical second moments can account for measurement errors. Experimental statistics methods intrinsically account for indirect measurement noise in the experimental covariance calculation process. Measurement errors are not intrinsically accounted for in the covariances estimated by forward operator methods, but can be incorporating by adding error covariance terms to the observed state variable auto-covariance before local conditioning on measurements is performed.

Once the initial random field model is defined—including the observed state variable and property value auto-covariances, as well as their cross-covariance and possibly their mean functions (depending on the estimation procedure)—the next step in the geostatistical approach to the inverse problem is to locally update the random field model based on existing direct and indirect measurements in a geostatistical estimation procedure. The selection of an initial random field model, in essence, corresponds to choosing an ensemble of solutions to the inverse problem (i.e. property value spatial distributions) on the basis of spatial structure considerations; one of these solutions, it is assumed, represents reality. However, at that first stage, the ensemble contains realizations that do not honor the measurements since the initial random field model only provides information about the intrinsic variability, without regard to the specific value of measurements at particular locations. The purpose of conditional estimation, then, is to eliminate from the initial ensemble those realizations that do not honor the measurements—to reduce the size of the property distribution space to conform with measurement observations.

The conditional estimation procedure is usually performed using some form of cokriging, which provides the minimum variance unbiased estimate of the property value distribution if the observed state variable is linearly related to the property value [Harvey and Gorelick, 1995]. If the above assumption does not hold true for a particular problem cokriging provides the best estimate which is a linear function of the measurements, although a better nonlinear estimator might exist [Harvey and Gorelick, 1995]. The property value estimates derived from cokriging can be taken as the best "average" solution to the inverse problem themselves, or a geostatistical
simulation routine, combining cokriging and Monte Carlo random modeling, can be implemented to generate equally-likely property distribution realizations that satisfy the inverse problem. Most geostatistical simulation algorithms require that the pdf of the initial property value random field be fully described—usually in the form of a normal pdf, or some other generic pdf that can be mathematically transformed to a normal pdf. An exception to this requirement is indicator-based simulation methods (e.g. the Markov-Bayes method) for which the form of the pdf does not have to be prescribed.

One of the primary reasons and advantages for using a geostatistical approach to the inverse problem is to rigorously account for uncertainty in the inverse solution. If the initial property distribution random field model is multi-Gaussian (or can be transformed into a multi-Gaussian model) the cokriged property value estimate at each location represents the expected value of a location-dependent Gaussian pdf (or can be transformed into the expected value, if a Gaussian transform was performed on the original pdf model) and the estimation variance, which can be easily calculated in the cokriging process, represents the standard deviation of the pdf; these two statistical parameters are all that are needed to fully describe the conditional Gaussian pdf. The geostatistical simulation approach, requiring the same multigaussian assumption, is advantageous in that the resulting property distribution realizations more accurately represent the expected real-world spatial structure than expected value or "best" average maps and they can be used as input to other system models; the response pdf for these system models can be estimated by constructing histograms from the system response to each of the realizations in the set of equally-likely realizations. Indicator cokriging techniques enable the conditional location-dependent pdfs of the property value estimates to be reconstructed in a discretized form—in addition to the same simulation capabilities mentioned above—regardless of the mathematical form of the initial property distribution random field model. Indeed, the geostatistical approach to inversion provides one of the most versatile and robust methods of accounting for uncertainty in the inverse problem solution—while enabling the estimation of the "best" average property distribution solution to the inverse problem, as well.
APPENDIX B

THE MARKOV-BAYES GEOSTATISTICAL APPROACH

B.1 Introduction

This appendix describes the mathematical and statistical basis of the framework used to represent and account for uncertainty in this study; the basis of this framework is the Markov-Bayes geostatistical approach [Zhu and Journel, 1993; Alabert, 1987]. The general statistical approach is to:

1. represent variables of interest as indicator spatial random variables (SRVs),
2. to condition these SRVs on both precise (hard) and imprecise (soft) measurements, and
3. to calculate expected value estimates of the SRVs or simulate multiple realizations of the SRVs in the area of interest—using a indicator cokriging or Monte Carlo sequential indicator simulation algorithm based on the Markov-Bayes model of accounting for soft data.

A probabilistic, as opposed to fully deterministic, representation of the variables of interest is implemented to account for the typically complex spatial variability in the distribution of hydrogeologic variables in geologic media. This variability is the result of the inherent heterogeneity of geologic strata and/or the complexity of the physical and chemical processes and interactions controlling the distribution of the variable of interest. The probabilistic representation also accounts for measurement uncertainty. With this approach the true, unknown, spatial distribution of the variable can be thought of as a single realization amongst an ensemble of equally possible realizations.

Defining the hydrogeologic variables as indicator SRVs (ISRVs) provides the flexibility to handle direct and indirect types of data with different reliabilities. Since the expected value of an ISRV translates directly into a probability of occurrence below/above a threshold value, ISRVs are very amenable to problems dealing with threshold levels, such as regulatory action levels for subsurface contaminants. Also, any probability density function (pdf) can be discretized into an
ISRV, although the degree to which the pdf shape is preserved depends on the amount of
discretization.

In general, a Bayesian approach to updating of statistical parameter estimates (e.g. mean,
variance, and spatial correlation length) is invoked. This approach entails updating existing
(prior) estimates of statistical parameters to new (posterior) estimates when new information
becomes available by applying Bayes' Theorem. Bayesian updating permits the incorporation of
subjective prior estimates of statistical parameters (such as geological intuition), unlike the
classical statistics approach which only uses actual measurements in updating [Freeze et al.,
1990]. If there are no subjective prior estimates these two approaches should essentially overlap.

B.2 Overview of indicator geostatistics

B.2.1 Definition of indicator spatial random variables

If Z(x) is a SRV with outcomes z(x) then a binary function can be defined:

\[ i(x, z_c) = 1 \quad \text{if} \quad z(x) \leq z_c \]
\[ i(x, z_c) = 0 \quad \text{if} \quad z(x) > z_c, \quad \text{for continuous SRV Z(x) and cutoff } z_c, \text{or:} \quad (B.1) \]
\[ i(x, z_c) = 1 \quad \text{if} \quad z_c = 0 \]
\[ i(x, z_c) = 0 \quad \text{if} \quad z_c \neq 0, \quad \text{for discrete SRV Z(x) and value } z_c. \quad (B.2) \]

From here on in indicators will be defined only for continuous SRVs in this section, but all
the equations and definitions apply for categorical indicators as well by just replacing the
inequalities with equalities and non-equalities as shown above. Indicators can be thought of as
posterior cumulative probabilities on Z(x) given z(x):

\[ i(x, z_c) = P[Z(x) \leq z_c \mid z(x)] = 0,1, \quad (B.3) \]

where P[A|B] represents the conditional probability of event A given event B.

B.2.2 Properties of indicators

The first order moment of I(x,z_c):

\[ \int i(x, z_c) \, d(x) = \int P[Z(x) \leq z_c \mid z(x)] \, d(x) = 0,1, \quad (B.3) \]
Appendix B. THE MARKOV-BAYES GEOSTATISTICAL APPROACH

\[ F(x,z_c) = E[I(x,z_c)] = 1 \times P[z(x) \leq z_c] + 0 \times P[z(x) > z_c] = P[z(x) \leq z_c], \quad (B.4) \]

where \( E\{A\} \) represents the expected value of SRV \( A \).

The indicator mean is simply the cumulative distribution function (cdf) of \( Z(x) \). If \( Z(x) \) and \( I(x,z_c) \) are first order stationary then \( F(z_c) = F(x,z_c) \), the mean is no longer a function of space.

The second order indicator moments:

*non-centered covariance:*

\[ E[I(x_i,z_c)I(x_j,z_c)] = 1 \times P[i(x_i,z_c) = 1, i(x_j,z_c) = 1] + 0 \times P[i(x_i,z_c) = 1, i(x_j,z_c) = 0] \]
\[ + 0 \times P[i(x_i,z_c) = 0, i(x_j,z_c) = 1] + 0 \times P[i(x_i,z_c) = 0, i(x_j,z_c) = 0] \]
\[ = P[z(x_i) \leq z_c, z(x_j) \leq z_c], \quad (B.5) \]

which represents the joint bivariate cdf of \( Z(x_i) \) and \( Z(x_j) \).

*centered covariance:*

\[ C(x_i,x_j,z_c) = P[z(x_i) \leq z_c, z(x_j) \leq z_c] - F(x_i,z_c)F(x_j,z_c) \]
\[ = P[z(x_i) \leq z_c, z(x_j) \leq z_c] - F(z_c)^2 \quad (B.6) \]

If \( Z(x) \) and \( I(x,z_c) \) are second order stationary the covariance depends only on the magnitude and direction of the vector \( h = x_i - x_j \) and the centered covariance becomes:

\[ C(h,z_c) = P[z(x) \leq z_c, z(x + h) \leq z_c] - F^2(z_c) \]
\[ = P[z(x) \leq z_c, z(x + h) \leq z_c] - F(z_c)^2 \quad (B.7) \]

Since \( I(x,z_c) \) is binary the variance is only a function of the mean:

\[ C(0,z_c) = F(z_c)(1 - F(z_c)) \]
\[ = F(z_c)(1 - F(z_c)) \quad (B.8) \]

B.2.3 Indicator coding of data

One of the advantages of ISRVs is the flexibility to handle many different types of hard and soft data. The codification of different data types is illustrated in the table below:
Appendix B. THE MARKOV-BAYES GEOSTATISTICAL APPROACH

<table>
<thead>
<tr>
<th>Data type</th>
<th>Coding of (i(x,z_k))</th>
<th>Remarks</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hard</td>
<td>0 or 1</td>
<td>(z(x)) perfectly known</td>
</tr>
<tr>
<td>Soft type A</td>
<td>[0,1] with quality index</td>
<td>Quality index computed from misclassification probabilities (see below).</td>
</tr>
</tbody>
</table>
| Soft type B   | • 0 if \(z(x) \leq a\)  
               • undefined if \(z(x) \in (a,b]\)  
               • 1 if \(z(x) > b\)            | Data provides hard inequality constraints: \(z(x) \in (a,b]\). |
| Soft type C   | \(F(x,z_c) \in [0,1]\)  | Prior cdf - treat independently from indicator data values as a local prior indicator mean. |

Table B.1: Indicator coding of data [Alabert, 1987; Deutsch and Journel, 1992]

Type (c) soft data represents knowledge about the form of the pdf for an SRV over a certain portion of the domain prior to measurements being made (e.g. knowledge from similar studies performed in neighboring areas). Alabert [1987] recommends treating this knowledge separately from other data types; at the same level as the mean. Otherwise if this knowledge is treated as indicator data the pdf would be perfectly preserved at the locations it is assigned, unaffected by nearby hard and soft data during indicator kriging. This is highly undesirable since it entails that the initial uncertain prior pdf is never updated. Thus, Alabert treats this knowledge as a "local" prior indicator mean which acts to re-scale the posterior distribution estimated through kriging (see later section on indicator kriging).

Data from indirect measurements, such as geophysical measurements used to glean information about hydrogeological properties, or imprecise direct measurements are best represented by type (a) soft data. The measurements are transformed to indicator values of the primary variable, if necessary, and assigned misclassification probabilities. The misclassification probabilities are defined as:

\[
p_1(x,z_c) = P[I_1(x,z_c) = 1 | I(x,z_c) = 1], \tag{B.9}
\]

\[
p_2(x,z_c) = P[I_2(x,z_c) = 1 | I(x,z_c) = 0]. \tag{B.10}
\]
Appendix B. THE MARKOV-BAYES GEOSTATISTICAL APPROACH

The probability \( p_1(x,z_c) \) represents the probability that the soft sample indicates \( z(x) \leq z_c \) given that this is indeed true and \( p_2(x,z_c) \) is the probability that the soft sample indicates \( z(x) \leq z_c \) even though in reality \( z(x) > z_c \). These misclassification probabilities are used to calculate a quality index:

\[
B(x,z_c) = p_1(x,z_c) - p_2(x,z_c).
\] (B.11)

If \( B(x,z_c) = 1 \), the soft data is perfectly correlated with the primary variable at this cutoff and is, thus, treated as hard data. If \( B(x,z_c) = -1 \), the soft data is perfectly negatively correlated with the primary variable (i.e. when \( i(x,z_c) = 1 \), \( i_s(x,z_c) = 0 \) and vice-versa). If \( B(x,z_c) = 0 \), the soft data has no correlation to the primary variable and, thus, provides no new information. The quality index is in turn used to compute the soft-soft covariances and soft-hard cross-covariances for indicator kriging and simulation if the Markov-Bayes approach is used (see below). The transformation of the secondary variable to a secondary indicator data series and the calculation of the misclassification probabilities is ideally performed using a calibration scattergram.

A calibration scattergram is constructed by plotting the values of soft measurements versus the values of direct hard measurements at collocated sampling locations. The soft measurements can be imprecise direct measurements or indirect measurements. Alabert [1987] requires that type (a) soft data be given indicator values of 0's and 1's strictly while Deutsch and Journel [1992] permit values between 0 and 1 based on the calibration. Alabert's requirements restrict the measurements to being of the same type (both direct measurements) unless a transformation of secondary measurements to primary values is performed before the calibration. The method used by Deutsch and Journel permits the direct calibration of secondary measurements with hard primary measurements and, thus, is more flexible. The method of Alabert also doesn't appear to account for the variance in primary variable values for a measured secondary value, or in terms of ISRV probabilities: \( P[I(x,z_c) = 0 | I_s(x,z_c) = 1] \). For an explanation of how the calibration scattergram is used to calculate the soft indicator prior probability cdf see Deutsch and Journel [1992, p.87,88]). Deutsch and Journel also provide a computer routine for performing the calibration and calculating \( p_1(x,z_c) \) and \( p_2(x,z_c) \). Whenever a calibration is performed the
integrity of or confidence in the calibration should somehow be assessed. No formalized or quantitative method has been developed for performing this assessment. It would seem reasonable to assess the integrity of calibration based on the number of collocated calibration pairs within each and all indicator classes as well as how well the calibration agrees with existing scientific models and previously performed studies. A confidence factor could be assigned and used to re-scale the quality measure for the soft data.

If no calibration sample is available for type (a) soft data then \( p_1(x,z_c) \) and \( p_2(x,z_c) \) must be estimated through models or expert human judgement. This approach is necessarily more qualitative than using a calibration sample. See Alabert [1987, p.39, Appendix A] for a discussion of using human judgement to estimate \( p_1(x,z_c) \) and \( p_2(x,z_c) \).

### B.2.4 Indicator kriging

Indicator kriging has the same basis as kriging with continuous SRVs, both of which constitute a minimum error variance estimation of unknown values based on a linear combination of known values (data). Indicator kriging is very useful in that it directly estimates (in a least squares sense) the conditional cumulative distribution function (ccdf) of an SRV at a cutoff since:

\[
i'(x,z_c) = E[I(x,z_c) | (n)]^* = P'[Z(x) \leq z_c | (n)],
\]

where \( A^* \) represents the kriging estimate of quantity \( A \) and \( (n) \) represents the conditioning information available. Thus, indicator kriging comes down to estimating the ccdf of \( Z(x) \) at cutoff \( z_c \) as a linear combination of available indicator information which minimizes the estimation variance.

Each kriged estimate could be conditioned on indicator data from all the cutoffs, but normally only data from the cutoff being estimated are used based on the argument that the unknown indicator value is more correlated with data from the same cutoff than from other cutoffs [Journel and Alabert, 1989]. If the indicator cross-correlations between cutoffs is
considered significant a more demanding formalism using indicator principal components can be used [Deutsch and Journel, 1992].

The linear combination of more than one data value could include terms which are products of indicator data, but this would require the use of higher order statistics which can rarely be determined [Journel and Alabert, 1989]. Thus, the expansion is normally limited to terms which are a function of only one indicator. This approximation results in the indicator kriging estimate of the ccdf being a linear approximation of Bayes Theorem for more than one datum [Alabert, 1987]. However, the errors caused by this approximation have been shown to be small, especially if the tails of the ccdf (i.e. less than .01 and greater than .99) are not of concern [Solow, 1986].

To accommodate both soft and hard data the linear expansion is simply expanded to include a linear combination of both data types. Then the kriging system becomes a cokriging system of hard and soft indicator data. This cokriging system is only required for type (a) soft data.

The two types of indicator kriging addressed in this study are simple and ordinary soft indicator cokriging. The simple cokriging system of equations is:

\[
\hat{y}(x_0, z_c) = \mathbb{P}[I(x_0, z_c) = 1| i(x_j, z_c), j = 1, N_h; i_t(x_k, z_c), k = 1, N_s] = a_0 + \sum_{j=1}^{N_h} a_j i(x_j, z_c) + \sum_{k=1}^{N_s} b_k i_t(x_k, z_c)
\]

where

\[
a_0 = F(x_0, z_c) - \sum_{j=1}^{N_h} F(x_j, z_c) a_j - \sum_{k=1}^{N_s} F(x_k, z_c) b_k
\]  

(B.13)

The cokriging weights \(a_j, b_k\) are calculated by solving the matrix equations:

\[
\sum_{j=1}^{N_h} a_j C_i(x_j - x_l, z_c) + \sum_{k=1}^{N_s} b_k C_{it}(x_k - x_l, z_c) = C_{il}(x_0 - x_l, z_c); l = 1, N_h
\]

\[
\sum_{j=1}^{N_h} a_j C_{it}(x_j - x_m, z_c) + \sum_{k=1}^{N_s} b_k C_{it}(x_k - x_m, z_c) = C_{it}(x_0 - x_m, z_c); m = 1, N_s
\]  

(B.14)

where \(C_i(x_j - x_l, z_c)\) is the covariance at the cutoff \(z_c\) between two hard data, one at point \(x_j\) the other at point \(x_l\); \(C_{it}(x_k - x_l, z_c)\) is the covariance at cutoff \(z_c\) between a soft datum at point \(x_k\) and a hard datum at point \(x_l\); and \(C_{it}(x_k - x_m, z_c)\) is the covariance at cutoff \(z_c\) between...
two soft data, one at point $x_k$ and the other at point $x_m$. A more simple, compact matrix representation is:

$$
\begin{bmatrix}
C_{s1} & C_{s2} & \ldots & C_{s \cdot m} \\
C_{s1} & C_{s2} & \ldots & C_{s \cdot m} \\
\end{bmatrix}
\begin{bmatrix}
a \\
b \\
\end{bmatrix}
= 
\begin{bmatrix}
C_f(x_0) \\
C_{u_s}(x_0) \\
\end{bmatrix}
$$

More than one type of soft data can be utilized in the cokriging system. This would change the formulation by adding an additional set of kriging weights, summation term, and equation for each additional soft data type. More simply in terms of notation, all the soft data types could be amalgamated into the soft data terms in the above system of equations by numbering the soft data types sequentially into distinct groups (i.e. $1, \ldots, n_s, n_s+1, \ldots, n_s+n_s', \ldots, n_s+n_s'+\ldots+n_s = N_s$; where $n_s$ is the number of soft data of type $j$ and $m$ is the number of soft data types). The soft covariance matrix and soft-hard cross-covariance matrices would then be augmented matrices:

$$
C_s = 
\begin{bmatrix}
C_{s1} & \ldots & C_{s1} \cdot C_{s \cdot m} \\
\ldots & \ldots & \ldots \\
C_{s1} \cdot C_{s \cdot m} & \ldots & C_{s1} \cdot C_{s \cdot m} \\
\end{bmatrix}, 
C_{u_s} = 
\begin{bmatrix}
C_{u_s1} & \ldots & C_{u_s \cdot m} \\
\ldots & \ldots & \ldots \\
C_{u_s \cdot m} & \ldots & C_{u_s \cdot m} \\
\end{bmatrix}, 
\text{and } 
\begin{bmatrix}
C_{t1} \\
\vdots \\
C_{tn} \\
\end{bmatrix}
$$

As written above simple indicator cokriging requires prior knowledge of $N_h + N_s + 1$ means as well as $N_h + 1$ by $(N_h + 1)$ hard covariances, $(N_s + 1)$ by $(N_h + 1)$ hard-soft cross-covariances, and $(N_s + 1)$ by $(N_s + 1)$ soft covariances for each estimate. However, under the common assumption of second order stationarity the prior mean is considered constant for the entire random field and the covariance is assumed to be a function of only separation and direction, not location.

According to Alabert [1987] type (c) data would enter the kriging system as a "local" prior indicator mean at any of the locations $x_0, x_j, x_k$ where the data is present in the equation for $a_0$. At locations where there is no type (c) data the global mean is used (assuming stationarity). Adjusting the mean locally would seem to be in conflict with the assumption of stationarity, although once the random field is updated through kriging the local expected values of the SRV become location dependent (i.e. the random field becomes nonstationary) anyway. Another way
Appendix B. THE MARKOV-BAYES GEOSTATISTICAL APPROACH

439

to approach the incorporation of this type of data would be to separate the domain into sub-domains based on areas where the type (c) data pdfs are the same to create separate stationary random fields. Then this data could be treated as subjective prior stationary indicator means for each indicator class. These prior process means (process referring to the overall statistical process as opposed to a particular point) could be updated with the taking of measurements within the respective sub-domains using Bayesian updating equations [James and Freeze, 1993]. Then soft simple indicator kriging could be applied to each sub-domain independently using the updated process means to attain a posterior ccdf at each location. This approach for incorporating type (c) data is more statistically sound than Alabert's [1987] approach, but possibly also significantly more computationally expensive due to the creation of separate stationary random fields and the need to update the process statistics within each of them.

The system of equations for soft ordinary indicator cokriging is the same as those for simple indicator cokriging except that the $a_0$ term is dropped, a Lagrange parameter is added to the left hand side of the covariance matrix equations, and the additional constraint:

$$\sum_{j=1}^{N_s} a_j + \sum_{k=1}^{N_s} b_k = 1.0$$

For ordinary kriging an estimate of the global mean is not required. Since the mean is not employed in ordinary kriging type (c) soft data can not be accounted for following the approach of Alabert [1987].

Under the Markov-Bayes model [Alabert, 1987; Deutsch and Journel, 1992] the soft covariance and hard-soft cross-covariance is calculated directly from the hard covariance and the quality indexes of the soft data using the relationships:

$$C_{ij}(x, x_k, z_c) = B(x_k, z_c)C_{ij}(x - x_k, z_c),$$

and

$$C_{ij}(x, x_k, x_m, z_c) = B(x_k, z_c)B(x_m, z_c)C_{ij}(x - x_m, z_c),$$

if $B(x_k, z_c) \neq B(x_m, z_c)$ or $|x_k - x_m| > 0$  \hspace{1cm} (B.17a)

$$C_{ij}(x, z_c) = B(x_k, z_c)C_{ij}(x - x_k, z_c),$$

if $B(x_k, z_c) = B(x_m, z_c)$ and $x_k - x_m = 0$  \hspace{1cm} (B.17b)
These relationships rely on the assumption that if hard and type (a) soft data are collocated the soft data provides negligible information about the value of the primary SRV at any point compared to the hard data and is, thus, ignored. The relationship for the soft covariance also relies on the assumption that the soft data from a location where hard and soft data are collocated provides negligible information about the value of the secondary SRV at any point. See Alabert [1987] for the full derivation of the soft simple indicator cokriging equations.

**B.2.5 Indicator simulation**

Indicator simulation is used to generate multiple, equally probable realizations of an ISRV, each of which is conditioned by existing data at their locations and reproduces first and second order process statistics, within ergodic fluctuations. Each realization consists of a spatial map of 0's and 1's for each indicator class.

The indicator simulation methodology used in this work is the Markov-Bayes sequential indicator simulation approach [Alabert, 1987]. This methodology is based on the Markov-Bayes cokriging estimator combined with a Monte Carlo random number generator:
1. An initial estimation point is chosen at random and each ISRV expected value probability is estimated based on cokriging with all nearby information, including previously simulated values—resulting in a ccdf at the point.
2. Another random number between 0 and 1 is generated and, depending on where this number intersects the ccdf, a value of the SRV is identified; this value represents the simulated realization value at this point.
3. Continuing on a random path through the points to be estimated, another point is chosen and steps 1 and 2 are repeated.
4. Steps 1 through 3 are repeated until all points have been estimated, resulting in a single complete realization.
5. Steps 1 through 4 are repeated, using a different random path through the estimation points, until the total number of desired realizations have been generated.
Necessary properties of a good simulation algorithm are reproduction of the first and second order process statistics, and the honoring of data at their locations, accounting for data quality. The reproduction of process statistics should be close for each separate realization and, more importantly, almost exact for the expected value of these parameters from a large number of realizations. Hard data should be exactly honored at their locations. The ISRV simulation algorithm used in this study, the Markov-Bayes sequential indicator simulator, has been shown to fulfill these properties [Alabert, 1987].