Optimal Experimental Design Applied to DC Resistivity Problems

by

Darrell A. Coles

B.A. Pure Mathematics, University of Rochester, 1994
M.S. Geosystems, Massachusetts Institute of Technology, 1999

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
in partial fulfillment of the requirements for the degree of
Doctor of Philosophy
at the
MASSACHUSETTS INSTITUTE OF TECHNOLOGY

February 2008

Author ……..

Department of Earth, Atmospheric, and Planetary Sciences
5 December 2007

Certified by ……..

Frank Dale Morgan
Professor of Geophysics
Thesis Supervisor

Accepted by …….

Maria T. Zuber
E.A. Griswold Professor of Geophysics
Head, Department of Earth, Atmospheric and Planetary Sciences

Optimal Experimental Design Applied to DC Resistivity Problems

by

Darrell A. Coles

Submitted to the Department of Earth, Atmospheric, and Planetary Sciences
on 5 December 2007, in partial fulfillment of the
requirements for the Degree of Doctor of Philosophy in
Geophysics

Abstract

The systematic design of experiments to optimally query physical systems through manipulation of the data acquisition strategy is termed optimal experimental design (OED). This dissertation introduces the state-of-the-art in OED theory and presents a new design methodology, which is demonstrated by application to DC resistivity problems. The primary goal is to minimize inversion model errors and uncertainties, where the inversion is approached via nonlinear least squares with L_1 smoothness constraints. An equally important goal is to find ways to expedite experimental design to make it practical for a wider variety of surveying situations than is currently possible.

A fast, sequential ED strategy is introduced that designs surveys accumulatively by an efficient method that maximizes the determinant of the Jacobian matrix. An analysis of electrode geometries for multielectrode data acquisition systems reveals that experiment-space can be usefully decimated by using special subsets of observations, reducing design CPU times. Several techniques for decimating model-space are also considered that reduce design times.

A law of diminishing returns is observed; compact, information-dense designed surveys produce smaller model errors than comparably sized random and standard surveys, but as the number of observations increases the utility of designing surveys diminishes. Hence, the prime advantage of OED is its ability to generate *small*, high-quality surveys whose data are superior for inversion.

Designed experiments are examined in a Monte Carlo framework, compared with standard and random experiments on 1D, 2D and borehole DC resistivity problems in both noiseless and noisy data scenarios and for homogeneous and heterogeneous earth models. Adaptive methods are also investigated, where surveys are specifically tailored to a heterogeneous target in real time or in a two-stage process.

The main contributions this thesis makes to geophysical inverse theory are: 1) a fast method of OED that minimizes a measure of total parameter uncertainty; 2) novel techniques of experiment-space and model-space decimation that expedite design times; 3) new methods of adaptive OED that tailor surveys to specific targets; and 4) though the OED method is demonstrated on geoelectrical problems, it can be applied to any inverse problem where the user controls data acquisition.

Thesis Supervisor: Frank Dale Morgan
Title: Professor of Geophysics

Biography

The author lives in Salem, Massachusetts with his wife, Kristen.

Education

- B.A. Pure Mathematics, University of Rochester. *May 1994*.
- M.S. Geosystems, Massachusetts Institute of Technology. *February2000*.

Employment

- Research Assistant, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology. *August 2002 - present*.
- Teaching Assistant, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology. *August 2002 - present*.

Certificates and Awards

- Excellency in Teaching, Massachusetts Institute of Technology, *Spring 2007*.
- New England Complex Systems Institute: One-Week Intensive Course [in] Complex Physical, Biological, and Social Systems, *January 2003*.

Refereed publications

- **Coles, D.**, et al., 2004, Spatial Decision Analysis of Geothermal Resource Sites in the Qualibou Caldera, Saint Lucia, Lesser Antilles, Geothermics **33 (3)**, 277-308
- Sogade, J., Vichabian, Y., Vandiver, A., Reppert, P.M., **Coles, D.**, Morgan, F.D., 2004, Electromagnetic Cave-to-Surface Mapping System, IEEE Transactions on Geoscience and Remote Sensing, **42 (4)**, 754-763

Submitted Publications

- **Coles, D.** and Morgan, F.D., Optimal Experimental Design Applied to Borehole DC Resistivity, in review Geophysical Journal International

Conference Abstracts, Reports and Presentations

- Morgan, F.D., et al., 2004, *Self-Potential and Gravity Studies for Nevis for Geothermal Exploration*, Renewable Energy in the Americas, Unit for Sustainable Development and the Environment, Organization of American States, Washington D.C.
- Morgan, F.D., et al., 2004, *Analysis for the Sustainability of the Water Resources and Supply of the Vieux Fort Region of Saint Lucia*, Ministry of Planning, Government of Saint Lucia

- Morgan, F.D., et al., 2002, *A Geosystems Sustainability Study of Harrison's Cave, Barbados*: for the Environmental Special Projects Unit, Government of Barbados, Sturges, Barbados
- Morgan, F.D., et al., 2001, *A Geosystems Sustainability Study of Harrison's Cave, Barbados*: for the Environmental Special Projects Unit, Government of Barbados, Sturges, Barbados
- Morgan, F.D., **Coles, D.**, Vichabian, Y., Sogade, J., 2001, *Assessment of the Groundwater Resources at Plantages PortoMari, Curaçao*, MIC n.v. Consulting Engineers, Willemstad, Curaçao
- Morgan, F.D., Sogade, J., Lesmes, D., **Coles, D.**, Vichabian, Y., Scira-Scappuzzo, F., Shi, W., Vandiver, A., Rodi, W., 2003, *Noninvasive Contaminant Site Characterization Using Geophysical Induced Polarization*. Abstracts of Papers of the American Chemical Society 225

Acknowledgements

I owe a profound debt of gratitude to my advisor and good friend, Dr. Frank Dale Morgan. His scientific and private personalities are an inimitable mixture of charm, perspicacity, and nonconformism whose influence has helped me to prosper both at MIT and in my personal life. Particularly, I would like to thank Dale for giving me the freedom to explore and to make – and, hopefully, to eventually correct – my own mistakes, always under supervision but often without direct intervention. It has been a singular and edifying experience to be educated and to educate myself in this way, and I feel stronger and more confident as a scientist and individual for the experience. And, of course, our numerous (mis)adventures in the West Indies have been an incomparable highwater mark in my life, which I will always treasure.

Next, with hat in hand, I humbly thank my wife, Kristen, who has patiently persevered with me in my *nine* years at MIT. During these years she has never failed to encourage me in my frustration, console me in my failure, and endure me in those rare moments of triumph when I simply could not shut up about some breakthrough. I could not dream of a better partner to have shared this experience with, and I cannot begin to express how deeply comforting and necessary her presence has been to me. I thank God for granting me the wisdom to choose a spouse who is *not* a scientist! She is the yin to my yang.

Of course, I would like to extend thanks to my thesis committee, Professor Dale Morgan, Professor Nafi Toksöz, Professor Tom Herring, Professor Stéphane Rondenay, and Dr. Michael Prange. I appreciate your time, effort and interest in my work.

Around campus I have had the pleasure of much lively discussion, both scientific and otherwise, with many students, faculty and postdocs. I would like to thank Burke

Minsley, Victoria Briggs, Bob Fleming, Samantha Grandi, Yervant Vichabian, John Sogade, Bill Rodi, Felix Hermann, Jonathan Kane, Xander Campmann, Jonathan Ajo-Franklin, the numerous Data and Models and Geosystems students I have TA'd, Aileen McLeod and many others for the happy times and memories that I will take away from MIT.

Outside of MIT, I am grateful for all the friends that have surrounded me and made my time in and out of metro Boston so thoroughly enjoyable. To Dio and Caroline Deltoro, thanks for all the great times seeing Waitiki and staying up too late dancing. To Dio Deltoro and Jeff Tanger, thanks for letting me bring the noise in Kobayashi. To Paul and Shari Lyden, thanks for being the crazy fun-loving folks you are and for sharing that with Kristen and me; I wish we had met sooner! To Dale and Connie Gephart for being the best landlords in the world. To Alex Mason for all the great conversations and for being such a good friend to Kristen. Thanks to all my hometown friends, Kyle, Marnie, Nate, Scott, Bill, Mike, Wes and Silas, for staying in touch and getting up to all kinds of trouble with me through the years! You have been as instrumental as anyone in helping me maintain my center!

Finally, I would like to thank my parents in my hometown of Fair Haven, NY as well as in Salem, MA. Moms and Dads, you are my rock and pillar, the people I admire and look up to most. It is thanks to you that I have become the man I am today and that I have been lucky enough to marry an extraordinary woman who completes me. I hope to always deserve your respect and love.

This work is dedicated to my sons Nathan Gibson Coles and Samuel Theron Coles and to their grandfather, Ronald Theron Coles. May you three be getting into untold mischief together somewhere beyond the stars.

Contents

List of Figures

List of Tables

List of Flowcharts

Chapter 1

Introduction

1.1 Motivation

In exploration geophysics, it is not uncommon that little or no information is available on an area of interest prior to surveying. The traditional approach is to deploy a standardized or *ad hoc* survey to query the earth, and it is taken for granted that such surveys are sufficient to query arbitrary targets. Contextually, 'sufficient' means that the acquisition strategy is adequate to produce a data set whose inversion is robust in terms of *modeling accuracy*, which depends strongly on the conditioning of the inverse problem, and/or in terms of *model uncertainty*, which depends strongly on the noise in the data and the degree to which it influences the inversion model. However, there is no *a priori* reason to assume that a standardized survey shall produce a data set whose inversion produces either particularly accurate models or particularly small modeling uncertainties. In this sense, standardized geophysical surveys are arbitrary and generally suboptimal, for the data they produce are not tailored for the unique spatial distribution of material properties at any particular target site. Standardized surveys might produce data that are often adequate, but they generally do not produce data that are optimal for particular targets.

At first glance, geophysical survey *optimality* can be qualitatively defined as the degree to which the data created by a survey 'illuminate' the target region. For example, a survey might be termed optimal if its inverted data produce the most accurate

model of the material property of interest when compared to models created by other surveys. Alternately, it may be more natural to define optimality by a survey whose inverted data are maximally impervious to the effects of data noise. If one intends to tailor a geophysical survey to an individual target site, a critical first step is to clearly define what survey optimality shall mean in that context.

The deliberate creation of geophysical surveys tailored for specific settings is termed *optimal experimental design* (OED). Tangentially, the terms *survey* and *experiment* shall hereafter be used interchangeably. Optimal experimental design is distinguished principally by the fact that it treats design as a *computational problem*. Rather than taking an empirical or *ad hoc* approach to the design of geophysical experiments, *optimal* experimental design differentiates itself in that some sort of design objective is mathematically posed and then solved in a computational, optimization framework. There are numerous ways of implementing OED depending on the goals of the geophysicist and the setting in which it is applied. Nonetheless, at the end of the day, all varieties of optimal experimental design have the same prime objective: to create a compact set of observations that minimizes expense (in terms of time and money) while producing superior data quality.

Where geophysical data acquisition and analysis are concerned, there are arguably two distinct classes of experimental design: *Adaptive* and *Specific OED*. Adaptive OED, which might also be called *Unspecific OED*, assumes that a functional relationship between data and model is known but that information on the particular target region is *unspecified*. This is why it is adaptive; the experiment can be refined or *adapted* as information on the target is gathered. Specific OED also assumes that a functional relationship between data and model is known, but in contrast to Adaptive OED, information on the particular target region is already known or *specified*. In this light, adaptive OED is really just a generalization of Specific OED, because specific models are hypothetically refined through time, allowing the designer to iteratively adapt the experiment to be better tailored to the site where it is employed. On the face of things, Specific OED is debatably the easier of the experimental design frameworks because the designer can focus on optimizing a survey for a single,

known model. Adaptive OED is more complicated because the geophysicist must establish design objectives and solver algorithms for a problem where model information does not initially exist or is being refined in real time.

There are two important distinctions worth noting here. First, there is a difference between *active* and *passive* experimental design problems. 'Active' refers to *active remote sensing* techniques, where the observer controls both when/where observations are made and when/where energy is injected into the system. 'Passive' refers to *passive remote sensing* techniques, where the observer controls when/where observations are made but not when/where energy is *injected* into the system. DC resistivity is an example of active remote sensing, where the observer injects electrical current into the system at known positions and times and observes how the intervening medium reacts to the excitation. Earthquake location is an example of passive remote sensing, where the observer waits for, or depends on, an event (an excitation source) to make observations, and which frequently also entails trying to *locate* the excitation source.

The second distinction is between what might be termed *functional* and *non-functional* experimental design. 'Functional' experimental design refers to situations where there is a known mathematical/physical relationship between data and model, whereas 'non-functional' ED refers to the case where no such relationship is known. The former case applies for standard inverse problems, which assume as given a mathematical relationship between data and model, while the latter case typically applies where no deterministic mathematical relationship is posited but where a statistical relationship is at least assumed.

To appreciate the difference between an optimal and a non-optimal experiment, let us consider an amusing, albeit unrealistic, lunar physics problem. There is an astronaut on the moon who wants to be a professional baseball pitcher. This astronaut has brought a baseball to the moon to conduct an experiment where he will throw the ball as hard as he can while his companion records the trajectory of the ball using time-lapse photography. The pitching astronaut wants to determine (1) the moon's gravitational acceleration, (2) his release velocity and (3) how tall he is. Unfortunately, his companion is so tickled by the experiment that she cannot help laughing while taking

measurements and is jiggling the camera, creating errors in observation both in the horizontal and vertical axes. For the sake of exposition, let us assume that the time-lapse camera can be programmed to take snapshots at any instant desired, with infinite precision. The equations of state governing this simple problem are

$$\begin{aligned} x(t) &= v_0 t + \varepsilon_x(t) \\ y(t) &= -\frac{1}{2} g t^2 + h + \varepsilon_y(t) \end{aligned} \tag{1.1}$$

where $x(t)$ and $y(t)$ are the x and y positions of the ball recorded at time t, v_0 is the release velocity, g is the gravitational acceleration on the moon, h is the height of the pitching astronaut, and $\varepsilon_x(t)$ and $\varepsilon_y(t)$ are the errors in observation created by the laughing astronaut at time t. An example of what the time-lapse photo might look like is shown in Figure 1-1.

The problem is to invert the noisy x and y positions to derive the most accurate possible estimates of g, v_0 and h. The astronauts have decided to conduct the experiment two different ways. The first experiment records the ball's position at 50 equispaced time intervals, from the moment of release to the moment of landing, and the second experiment records the ball's position 50 times according to a very particular sampling schedule (Sampling Schedules 1 and 2, respectively; see Figure 1-2). To reiterate, let us assume that the camera can be programmed to take snapshots with infinite precision at any instant in time, even if these instants are separated by an infinitesimal increment of time.

To get an idea which of the experiments is superior for accurately estimating the inversion parameters, we performed a Monte Carlo investigation wherein each experiment was simulated 1000 times with different realizations of observation noise, and the data were inverted for each instance. This simulates the expected range of modeling errors that would be caused by the laughing astronaut. Figure 1-3 shows cross-plots of the parameter estimates for the Monte Carlo simulations of the two experiments and histograms of their model % rms errors. The scatter plots clearly indicate that the expected ranges of percent error in the parameter estimates are smaller with respect to Sampling Schedule 2, especially for estimates of lunar gravitation, g.

The histogram in Figure 1-3 further validates that Sampling Schedule 2 is better than Schedule 1. The histogram shows the model % rms errors for all three parameters together, and it clearly shows that Sampling Schedule 1 has a higher probability of producing large model errors compared to Schedule 2. Hence, the astronauts should use Sampling Schedule 2 to get the best estimates of g, h and v_0, so long as the laughing astronaut cannot get control of herself! For the curious reader, the true values were g = 1.6 m/s^2, v_0 = 35.8 m/s and h = 1.8 m. The estimated values were g = 1.6 ± 0.08 m/s^2, v_0 = 35.8 ± 0.03 m/s and h = 1.8 ± 0.04 m for Sampling Schedule 1 and g = 1.6 ± 0.05 m/s^2, v_0 = 35.8 ± 0.03 m/s, and h = 1.8 ± 0.03 for Sampling Schedule 2. These estimates further demonstrate that Schedule 2 produced less model uncertainty than did Schedule 1.

Naturally, the reader might wonder where Sampling Schedule 2 came from for this astronomical sports problem. That question is at the heart of this research. This question is addressed in this thesis in the limited framework of geoelectrical problems, but *the reader should appreciate that the ideas presented here are applicable to any remote sensing technique*.

1.2 Context

As the title states, geophysical optimal experimental design in this thesis is approached in the context of geoelectrical problems. To provide a general reference for survey design, a brief introduction to the instrumentation and procedures by which resistivity data are collected and inverted is provided. Please note that only 1D and 2D problems are examined in this research.

Figure 1-4 shows a cartoon representation of a resistivity setup for 2D surface resistivity. Instrumentation and procedures for 1D resistivity are similar to the 2D case shown here, and what differences exist will be discussed. A number of survey electrodes are arrayed collinearly along the ground, most often at equispaced intervals, and are connected to a *resistivity meter* via a *multiplexor*. In nearly all resistivity applications, four electrodes are used to make a field observation; two as a *current* or

source dipole, across which a known current is injected into the ground; and two as a *potential* or *receiver dipole*, across which the earth's response to the injected current is measured as an electrical potential. The resistivity meter is the primary instrument in field exercises. This instrument contains a program/schedule that specifies the combinations of *4-electrode or quadrupole* observations comprising a survey or experiment. The resistivity meter communicates this program to the multiplexor, which is like a sophisticated switchboard that accepts and executes observation commands from the resistivity meter. Following these instructions, the multiplexor opens and closes relays to the electrodes for each source/receiver combination. The current dipole is excited from the *power supply*, which forms a closed electrical loop with the multiplexor, and the values of injected current and observed potential pass from the multiplexor back to the resistivity meter. The *master computer* has one or more roles. It can be used to control the resistivity meter, communicating survey instructions, and/or it inverts the resistivity data via an inversion algorithm. Geoelectrical inversions are typically carried out on piecewise constant, mesh-grid models of the earth, like the one shown in Figure 1-4. In view of the work carried out in this thesis, the master computer may also be tasked with optimizing geoelectrical surveys in the field and automatically passing these survey instructions to the resistivity meter for execution.

The field setup described above is what is known as an *automated multielectrode system*. Once the electrodes and cables have been laid out and the resistivity meter (or the master computer) has been programmed with a geoelectrical survey or experiment (a suite of quadrupole observations), the field operator simply hits the 'execute' button and the resistivity data set is automatically collected. For the 20-electrode setup in Figure 1-4, it might take one hour to lay out the electrodes and cables. A survey of several hundred to as many as a thousand observations would typically be collected with this number of electrodes, and depending on the sophistication of the multiplexor, the field data might take anywhere from a half hour to two hours to collect. Because the survey electrodes must be pre-placed to take advantage of automated data collection, there are a finite number of possible quadrupole combinations. The com-

binatorial character of these surveys presents interesting challenges for geoelectrical experimental design as will be seen in later chapters.

Setup and data collection for 1D resistivity is slightly different. First, one-dimensional resistivity assumes the earth can be modeled by a series of horizontal layers, rather than by a piecewise-constant mesh. Second, while it is necessary to pre-place the electrodes for automated 2D data collection, 1D resistivity actually goes faster without automation, so electrodes can be placed anywhere along the survey line. As will be seen later on, this opens up 1D resistivity experimental design to an infinitude of possible observations, which poses its own unique challenges.

1.3 Background

Despite nearly five decades since survey optimization began to appear in the literature, experimental design is still an emerging field in computational geophysics. Improvements in fast, cheap computing have helped geophysical data inversion to blossom over the past several decades, and modern computational power is now making geophysical experimental design a serious possibility. There are two reasons it has taken OED so long to develop. First, geophysical inverse theory/practice ostensibly needed to mature first (10,000 papers have been published on data inversion over the past five decades while only 100 have been published on experimental design over the same interval (Maurer & Boerner, 1998a). It could not have been readily obvious to inversion practitioners the importance of the experiment until geophysical inversion had itself fully matured; only then would it be natural to inquire whether *smarter* data could be collected. Second, OED is generally more computationally expensive than data inversion. The high computational overhead of OED derives from the fact that its search space (experimental space) is frequently discrete, combinatoric and compatible only with so-called exhaustive or global search algorithms like the genetic algorithm, simulated annealing, et cetera (e.g. Curtis, 1999b).

Research in geophysical OED has increased in the past decade, with contributions coming primarily from the independent and collaborative works of Maurer (1998a;

1998b; 2000), Curtis (1999a; 1999b; 2000; 2004) and their associates (Stummer *et al.*, 2002; van den Berg *et al.*, 2003; Stummer *et al.*, 2004), with a few other authors having published during this time as well (Forsell & Ljung, 2000; Furman *et al.*, 2004; Narayanan *et al.*, 2004; Wilkinson *et al.*, 2006a; Wilkinson *et al.*, 2006b; Furman *et al.*, 2007).

Geoscientists have considered optimal experimental design for such problems as cross-borehole seismic tomography (Curtis, 1999a), microseismic monitoring surveys (Curtis *et al.*, 2004), oceanographic acoustic tomography (Barth & Wunsch, 1990), 1D electromagnetic soundings (Maurer & Boerner, 1998a; Maurer *et al.*, 2000), 2D resistivity (Stummer *et al.*, 2002; Stummer *et al.*, 2004; Wilkinson *et al.*, 2006a; Wilkinson *et al.*, 2006b), seismic amplitude versus offset (van den Berg *et al.*, 2003), and oceanographic bathymetry inversion (Narayanan *et al.*, 2004). These works use a variety of objective functions to establish a mathematical definition of experiment optimality, but all are alike in that the basic formulation of the OED problem, as will be seen in the next chapter.

Recent work in geoscientific experimental design (see any of the aforementioned citations) has been based on the linear approximation of (predominately) nonlinear model functions, whose elements have been manipulated in an optimization framework to give rise to optimal data acquisition strategies. Aside from linearization, the defining characteristic of many of these efforts has been their use of so-called *global* or *stochastic search* methods to optimize data surveys. Examples of such optimization techniques include simulated annealing and genetic algorithms, both of which have been employed in experimental design exercises (e.g., Barth & Wunsch, 1990; Curtis, 1999a, respectively). It is well known that global search strategies are well suited to finding global minima, even in the presence of multiple local minima, which is one of the reasons these methods have been so widely used in experimental design research (Curtis & Maurer, 2000). However, though they can be guaranteed to find global minima, global search strategies are among the least efficient optimization techniques (acknowledging that sometimes they are among the only strategies for finding global minima for certain kinds of problems, such as combinatorial problems

(Du & Pardalos, 1998)). Design computation times on the order of hours or days have been reported in the literature (e.g., Curtis & Snieder, 1997; Wilkinson *et al.*, 2006a) for modestly-sized inverse problems.

Because of the mathematical nature of optimal experimental design, a discussion of the background of OED methodologies is forestalled until the next chapter, where it provides an instructive foundation for a proposed, novel method of OED.

In some situations, the computational expense of optimal experimental design, implemented via global search methods, may be comparatively low, such that the user is willing to accept the cost/benefit ratio, where *cost* equates with the computation time and *benefit* equates with the degree to which data quality can be improved. However, there are plenty of practical situations where the cost/benefit ratio is prohibitively high. For example, suppose a team of geophysicists is given a week to assess a geothermal region with a set of 20 electromagnetic soundings distributed over tens of square kilometers of unknown subsurface. Once some prelimnary soundings have been performed and subsurface images produced, the team may wish to adapt their EM surveys, based on that information, to reduce inversion model uncertainties. However, an OED method that takes hours or days per survey would clearly be impractical.

The question naturally arises: *Can experimental design be formulated so that the inefficiency of global search methods can be avoided? Or, put another way, can OED be streamlined to make it practical for a wide variety of remote sensing settings?*

1.4 Thesis Plan

The primary objective of this thesis is to develop a method for systematically designing geoelectrical surveys that produce *smart data*, data that have superior characteristics for inverse problems. To be clear then, the geophysical experimental design exercises examined herein are *functional*, meaning that constitutive equations governing the relationship between model and data are given, and they pertain to *active* remote sensing techniques, because the observer controls both the sources and receivers. The

systematic design of experiments is treated as an optimization problem, solvable by any number of optimization algorithms. The concept of smart data is fully developed in Chapter 2, where a mathematical objective function is proposed whose minimization gives rise to optimal experiments. Chapter 2 also discusses several historic objective functions that have previously been published, in an effort to clarify the meaning of survey optimality. This chapter concludes with the introduction of several novel experimental design optimization algorithms. Chapters 3 – 5, which are annotated below, examine the design methodology in the context of three geoelectrical scenarios of increasing difficulty, in the sense that the number of model parameters increases for each scenario. This necessitates various tricks and workarounds that attempt to expedite the design procedure, which becomes slower commensurate with an increase in the number of model parameters for which experiments are designed.

Chapter 3 applies the theory developed in Chapter 2 to 1D surface resistivity problems, particularly in the presence of noisy data. A distinction between continuous and discrete experimental search spaces is discussed, which has ramifications for the types of solver algorithms one can apply to the design problem. Designed surveys are compared with standard Schlumberger and Wenner surveys in a Monte Carlo framework to produce a statistically meaningful purview of their relative inversion performances. Both *unspecific* (homogeneous earth assumption) and *adaptive* OED approaches are examined. In addition, Chapter 3 examines the issue of data noise created by electrode misplacement and comes up with a surprising discovery.

Chapter 4 examines the novel OED method in the context of single-borehole DC resistivity. As in Chapter 3, designed surveys are compared with standard and random surveys in a Monte Carlo setting. Both unspecific and adaptive experimental designs are considered, and especially 'real-time' adaptive OED is taken up and examined at length. Chapter 4 works with discrete experimental design spaces, which are beneficially reduced in size through a novel examination and application of the properties of electrical resistivity tomography (ERT) quadrupoles. Both noiseless and noisy data scenarios are investigated.

Chapter 5 investigates the OED method in the context of 2D surface resistivity. The number of model parameters in this chapter is the largest of the three chapters, and special attention is paid to expediting the design algorithm through several novel model-parameter reduction schemes. Again, designed surveys are compared with standard and random surveys in a Monte Carlo setting, and both unspecific and adaptive OED techniques are considered.

Chapter 6 summarizes the results, insights and conclusions of this thesis. Following the concluding chapter are four Appendices. Appendix A introduces the Differential Evolution algorithm, which is used to search the continuous experimental search spaces in Chapter 3. Appendix B provides an alternate analytic derivation explaining empirical results in Chapter 3 concerning electrode misplacement errors. Appendices C and D tabulate the Pseudosection and ERL surveys, respectively, which are used extensively throughout Chapters 4 and 5.

1.5 Contributions

Below is a list of significant contributions this thesis adds to the fields of optimal experimental design and geoelectrical theory in general.

- A novel, computationally inexpensive basic method for optimizing data acquisition strategies, which is applicable not only to geoelectrical problems, but also to any active source remote sensing technique that relies on data inversion.
- Several novel techniques of adaptive experimental design.
- Identified a law of diminishing returns for optimal experimental design, showing that OED is best used to create *compact*, *smart* data sets.
- Several computational expedients that can usefully be applied to any kind of optimal experimental design problem.
- An analysis of how electrode misplacement errors affect data noise, and thereby modeling errors, for 1D resistivity.

- An analysis of electrode geometries for fixed-electrode surveys that identifies a significant cause of resolution limitations in the inverse problem.
- An analysis of electrode geometries for fixed-electrode surveys showing that all quadrupole observations can be classified as one of three types. Further analysis of these three configurations demonstrates that one of the most widely used geoelectrical survey, the Pseudosection survey, produces markedly inferior data quality.

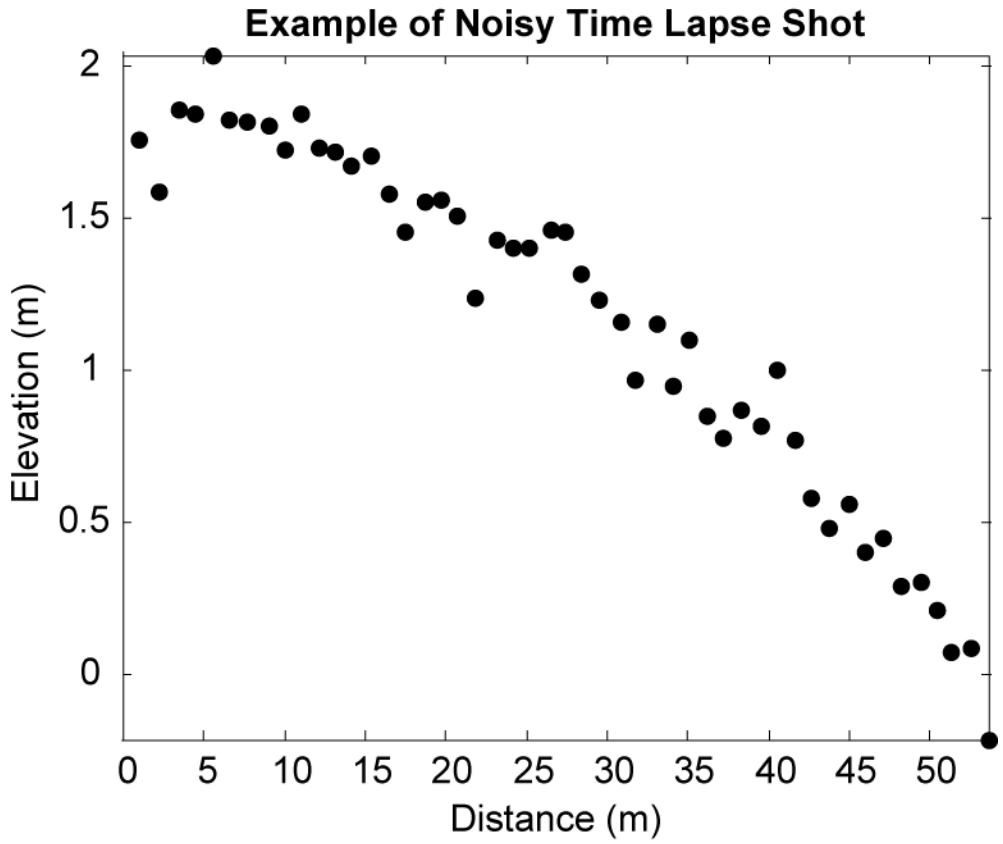


Figure 1-1 Example of a noisy time-lapse shot.

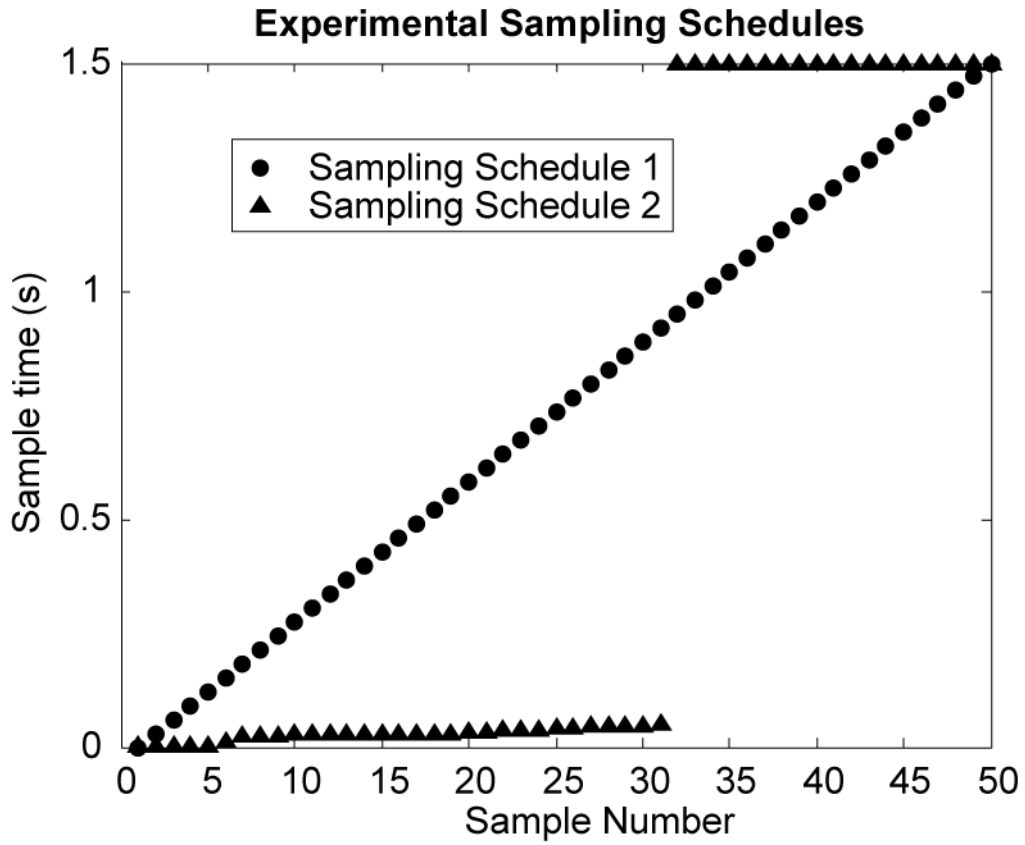


Figure 1-2 Two sampling schedules (experiments) the astronauts can use to observe the ball trajectory. Schedule 1 uses equispaced time intervals from the ball's release to its landing. Schedule 2 uses an 'optimized' experiment where most of the sample snapshots are taken right after release and right before landing. Noisy data from these two experiments are ultimately used to estimate the lunar gravitational acceleration, g, the release velocity of the pitching astronaut, v_0, and height of the pitching astronaut, h.

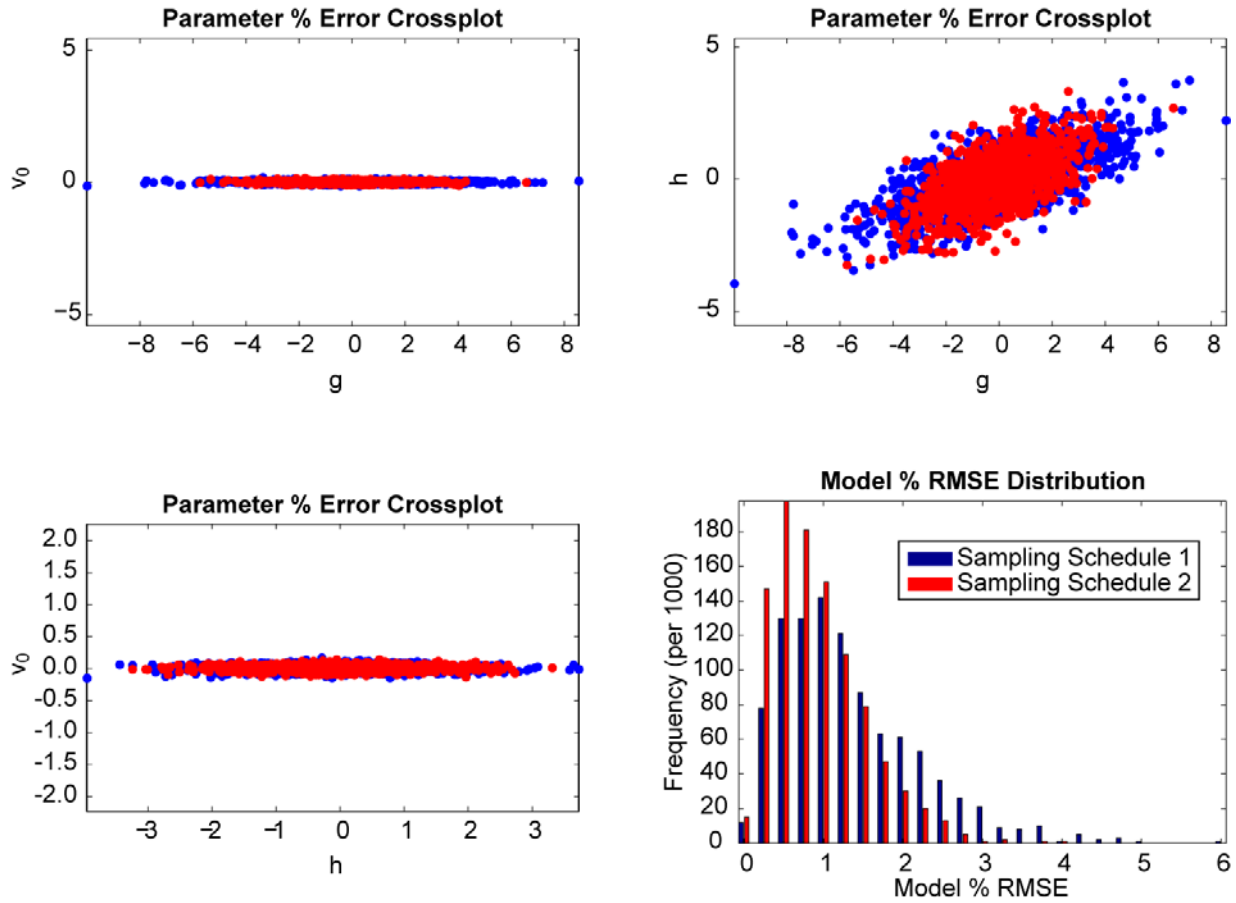


Figure 1-3 Inversion results of Monte Carlo simulations for the two sampling schedules in the presence of observational errors. The scatter cross-plots show the percent errors between the true parameters and those derived by inversion. The histogram shows the distribution of model % rms errors for the three model parameters. For the curious reader, the true values of the model parameters were g = 1.6 m/s^2, v_0 = 35.8 m/s and h =1.8 m.

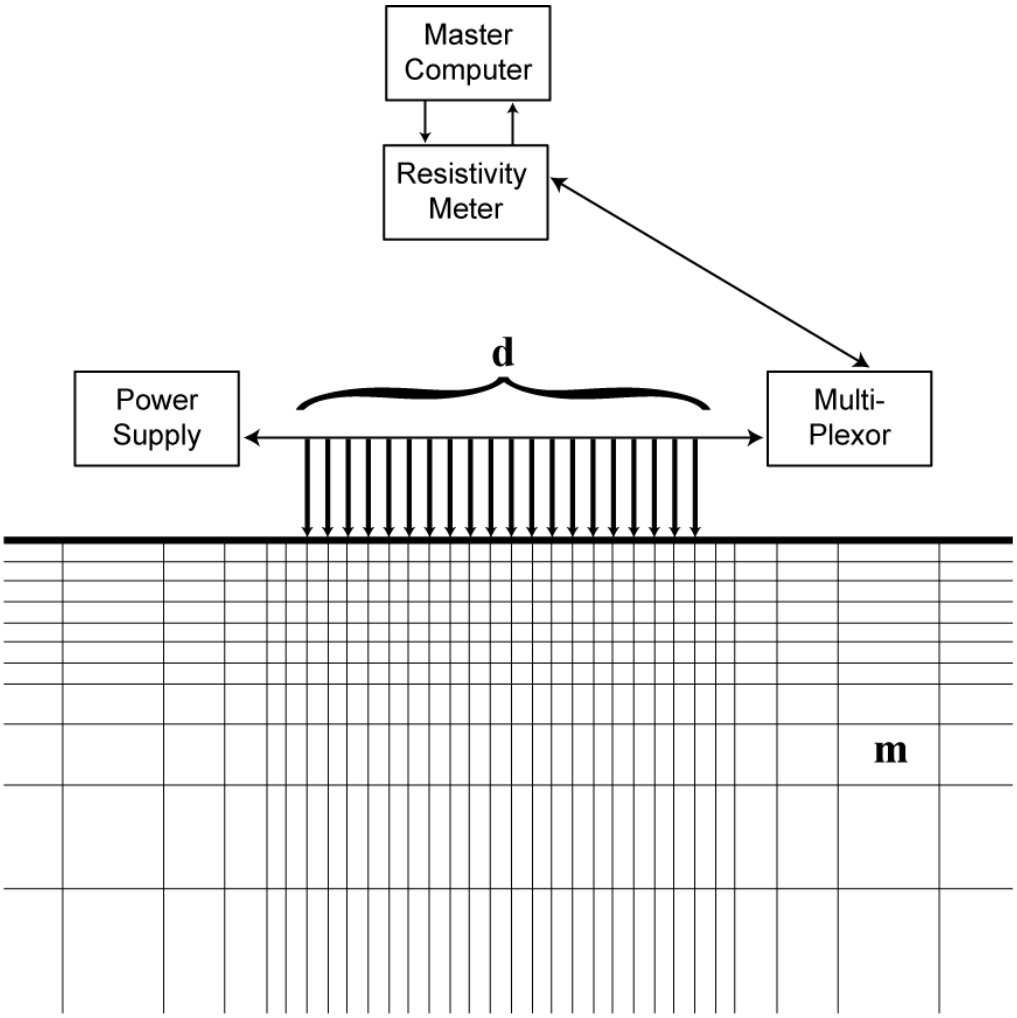


Figure 1-4 Cartoon of a typical setup for an automated, multielectrode 2D resistivity system. Because data are automatically collected, the electrodes must be pre-placed and remain in position during the survey. The *resistivity meter* contains a user-supplied progam that describes the resistivity survey, which is communicated to the *multiplexor*. The multiplexor is like a sophisticated switchboard that accepts commands from the resistivity meter on which quadrupoles to use for each datum in the experiment. The *power supply* forms a closed electrical loop with the multiplexor and supplies the source electrodes with current. A *master computer* is sometimes used to remotely control the resistivity meter, making the resistivity meter a slave. The master computer also inverts the resistivity data and, in the future, may be responsible for designing optimal experiments.

Chapter 2

Theory

2.1 Basics

Optimal experimental design can be qualitatively viewed as a systematic method to maximize the connection between the data collected and the model from which they derive. Whenever there is a known mathematical relationship between *data space* and *model space*, experimental design is clearly contingent upon that relationship. The relationship between geophysical data and model is expressed by some forward operator g, which maps a model m to data d, via

$$d = g(m). \tag{2.1}$$

g may either be linear or nonlinear with respect to m, though in the context of geophysical problems, it is almost always nonlinear. For example, for geoelectrical problems, Poisson's equation states that

$$\nabla \cdot \mathbf{J} = -\nabla \cdot (\sigma \nabla \phi) = \delta(\mathbf{x}), \tag{2.2}$$

where $\mathbf{J}$ is the current density, σ is the spatial distribution of electrical conductivity (the reciprocal of resistivity, ρ), and ϕ is the electrical potential field arising from a point current source at position $\mathbf{x}$. In this context, the data, d, correspond to the potential field, ϕ, and the model, m, corresponds to the electrical conductivity distribution, σ (or its reciprocal, ρ). Hence, g is a nonlinear function that operates on the conductivity distribution (model) to produce data (electrical potentials) such that both

data and model satisfy the nonlinear partial differential equation in (2.2). For the purposes of the following theory, it shall be necessary to assume that the forward operator is at least once differentiable with respect to the model values; that is, $\partial g / \partial m$ must exist.

Equation (2.1) is incomplete insomuch as it does not explicitly account for the experiment that is used to establish the connection between model and data. It is more accurate to explicitly state this mapping as

$$d = g(m;\omega), \tag{2.3}$$

where ω is a data station (see Figure 2-1.a), which can be thought of as a point in *experiment space*. Equation (2.3) explicitly expresses a mapping from two spaces (model and experimental space) to one space (data space), but this is still not yet a fruitful way to think about optimal experimental design. Rather, experimental space is a large, possibly infinite, set of mappings, where each mapping associates with a single experiment (Figure 2-1.b) that links data and model space in a unique way. Some mappings are two-way (invertible), some are one-way (noninvertible), and each has a different 'quality' or 'fitness', specifying how strongly it connects data space and model space together.

Viewed in this simple, qualitative manner, the object of the experiment designer is to find an experiment from the set of all possible experiments that produces the highest 'quality' mapping between model and data. Preferably, this mapping should be two-way or what is known in real analysis as an *injective* or *one-to-one mapping* (Bartle & Sherbert, 2000). Briefly, denoting data space *D* and model space *M*, an *injective mapping* between *M* and *D* requires that every model in *M* maps to a unique data set in *D*, bearing in mind that the mapping is actualized by the particular experiment used (see Figure 2-1). *Injectivity* ensures uniqueness in the forward and inverse mappings between *M* and *D*, and it thereby also ensures that the modeling operator, *g*, is invertible, as alluded to above. This follows because $g(m;\omega)$ is invertible with respect to *m* if and only if $g^{-1}\left(g(m;\omega)\right) = m$ for all *m*. If $m \in M$ and $g(m;\omega) \in D$, and if every *m* in model space maps to a unique $g(m;\omega)$ in data space, then the *in-*

verse mapping, from D to M, logically must return $g(m;\omega)$ to the point in model space from which it originated. Hence, if g is injective then $g^{-1}\left(g(m;\omega)\right) = m$ for all m, satisfying the definition of invertibility. The question is: what has this to do with geophysical inverse problems?

The answer to that question is straightforward: we should like for designed experiments to give rise to well-posed inverse problems that have unique solutions. Therefore, we seek an experiment, ω, that ensures the invertibility of Equation (2.3) such that for any model, m, we can solve for a unique d. To do this, the experimental design problem must be cast in the mathematical framework of the inverse problem. The 'quality' of an experiment must depend in part on the 'strength' of the forward mapping it produces from model to data space but *especially* on the 'strength' of the *inverse* mapping it produces from data to model space. Consequently, experimental 'quality' must be conflated with the inverse problem. Later on, it will become necessary to consider the possibility that an inverse problem can never be well posed, and this necessarily affects the definition of experiment 'quality'. For now, however, it is important to lay the basic foundation for optimal experimental design, and focusing on problems that can be well determined is a good starting point.

Equations (2.1) and (2.3) are nonlinear continuous expressions. To estimate the model that gives rise to some observed data, Equation (2.3) must be inverted. To do so for geophysical problems, the earth is traditionally discretized into semi-infinite discrete layers (1-dimensional models) or into regular/irregular meshes (2- and 3-dimensional models) within which earth material properties are treated as constant. An analogous mathematical relationship is then established between the discretized earth model and the constitutive equations that govern the physical phenomenon of interest (electromagnetic, elastic, gravitational, etc.). From the discretized representation derive data vectors, **d**, and discretized-model vectors, **m**. A linear approximation of g is generated by first-order Taylor expansion (e.g., Lanczos, 1961), giving rise, in matrix-vector notation, to

$$\Delta\mathbf{d} = \mathbf{G}\Delta\mathbf{m}\,, \tag{2.4}$$

where **G** is the sensitivity (Jacobian) matrix, comprising the partials of $g(\mathbf{m}, \omega_i)$ with respect to the discretized model parameters m_j:

$$G_{ij} \equiv \frac{\partial g(\mathbf{m}; \omega_i)}{\partial m_j}. \tag{2.5}$$

For Equation (2.4) to be a tractable geophysical problem in the inverse sense, it is necessary to model the earth with a finite number of discrete parameters such that piecewise-constant models approximate the earth. The following development treats earth models as being piecewise-constant approximations to the true earth model, and particularly, the mathematical formulation of the optimal experimental design problem is cast in terms of piecewise constant models.

Because of its widespread use throughout the remainder of this and subsequent chapters, it is necessary to formally define a *data sensitivity kernel* (hereafter labeled simply *sensitivity kernel*), which is the vector of partials of g with respect to all model parameters for a single data observation:

$$\mathbf{g} \equiv \frac{\partial g(\mathbf{m}; \omega)}{\partial \mathbf{m}}, \tag{2.6}$$

where ω is a single observation or data station. Clearly, the rows of **G** comprise these sensitivity kernels:

$$\mathbf{G} = \begin{bmatrix} \mathbf{g}_1^T \\ \mathbf{g}_2^T \\ \vdots \\ \mathbf{g}_{\mathcal{D}}^T \end{bmatrix}, \tag{2.7}$$

where $\mathbf{g}_j$ is the sensitivity kernel corresponding to the j^{th} data station, ω_j, and where there are a total of $\mathcal{D}$ observations.

It should be clear that Equation (2.4) linearly approximates Equation (2.3) in the neighborhood of the discrete model, **m**, and that **G** is thus a linear approximation of the nonlinear forward operator, g. The Jacobian and its generalized inverse, **G** and $\mathbf{G}^*$, respectively, are the algebraic operators upon which optimal experimental design typically operates, either explicitly or implicitly. This is because these two matrices

govern the transformations or mappings back and forth between data and model space. In a sense, $\mathbf{G}$ and $\mathbf{G}^*$ contain *information* on how these transformations are actuated, an idea that is revisited later in this chapter. If the information in $\mathbf{G}$ and $\mathbf{G}^*$ is 'strong', the mapping between M and D is strong. If $\mathbf{G}$ is singular, the mapping between model and data space is not two-way and $\mathbf{G}^*$ is technically undefined. However, this is typically circumvented by imposing additional 'external' information in the form of so-called *model regularization*, which enforces an assumption about the earth model such as smooth spatial variability. In point of fact, as will be seen in the next section, researchers have proposed several different measures of experiment 'quality', and all of these operate in some fashion on $\mathbf{G}$ or $\mathbf{G}^*$.

Nonlinear least squares inversion starts with an initial guess in model space and iteratively describes a trajectory through model space as updates are generated:

$$\begin{aligned} \Delta\mathbf{m}_n &= \mathbf{G}_n^* \Delta\mathbf{d}_n \\ \mathbf{m}_{n+1} &= \mathbf{m}_n + \Delta\mathbf{m}_n \end{aligned}, \tag{2.8}$$

where $\mathbf{m}_0$ is a user-specified initial model guess and where $\mathbf{G}_n^*$ and $\mathbf{d}_n$ are respectively the Jacobian matrix and residual vector at the n^{th} iteration, both of which are implicitly dependent on the n^{th}-iteration model vector, $\mathbf{m}_n$. For nonlinear inversion, an experimental design should ideally be optimal not only with respect to the true solution (which is typically unknown) but also desirably with respect to each of the models through which the inversion passes on its trajectory to the true solution. This objective cannot be practically satisfied because it requires foreknowledge of the inversion's trajectory through model space, but it serves to highlight the necessity for OED to be realistically posed. This point cannot be overstressed; the majority of published studies have simply used true models (Maurer *et al.*, 2000), homogeneous models (e.g. Stummer *et al.*, 2004), or checkerboard models (Curtis, 1999b) for designing geophysical experiments. We do not claim that such approaches are without merit, only that they are unrealistic for actual field exercises; it is no more probable that we know the true earth model *a priori* than it is that the earth is homogeneous or a checkerboard. Such exercises are certainly instructive and guide us to deeper insights into the geophysical OED endeavor, which is, after all, still only in its nas-

cence. Nonetheless, the purpose of optimal experimental design should be to create surveys that are individually tailored to the *specific* target being queried, not to generic or unrealistic models. The models above should only be used as starting points, and subsequent experiments should then be *adapted* to the target as refining information becomes available.

These concepts are exemplified in a hypothetical, 2-parameter problem shown in Figure 2-3. Panel (a) shows an 'inversion trajectory' through model space for a nonlinear inverse problem with noisy data. The green dot indicates the initial model guess for which an optimal experiment has ostensibly been designed. Based on the discussion above, the initial model could be, for example, homogeneous or a checkerboard (e.g., but *not* the true model because this can never realistically be known!). The grey region around the initial model delineates the neighborhood for which the designed experiment is 'optimal', bearing in mind that the experiment has been designed with respect to a *linear approximation* of the forward operator. The final solution (red dot) is not within this 'optimal design' region and therefore the experimental design cannot be said to be optimal with respect to it. As regards Panel (a), it would be ideal if it were possible to create an experiment that was optimal at all points along the inversion trajectory, but as was pointed out above, this presumes foreknowledge of the inversion trajectory, which is impossible. In point of fact, the inversion trajectory is *dependent* on the experiment, and if we alter the experiment we alter the trajectory. So even if we knew the trajectory in Panel (a), it would pertain only to the experiment that was used to perform *that* inversion. Any alteration of this experiment, through design, would create a *different, unknown* trajectory for which there would be no guarantee of optimality along its path. This explains why *nonlinear* experimental design, as it might be called, would be very difficult, if not impossible.

Figure 2-3.b shows a hypothetical example of what the inversion trajectory for this simple 2-parameter might look like if we were to implement some sort of adaptive experimental design. Figure 2-3.a demonstrated that, owing to the limitations imposed by linear approximation of Equation (2.3), an experiment designed for a model significantly different from the true model could not guarantee experiment op-

timality in the neighborhood of the true model. If we designed a new, optimal experiment with respect to the converged solution in Figure 2-3.a (red dot), it is hypothetically possible to improve inversion accuracy (defined as the model rms error between the true solution and the inversion model) and ensure that the designed experiment produces a region of optimality (grey area) more likely to encompass the true solution. Effectively, this new experiment would be deliberately *adapted* to the true model (or at least to one in the neighborhood of the true model), rather than to an arbitrary model such as a homogeneous or checkerboard earth. A well-designed experiment would hypothetically produce an inversion model closer to the true model as well as a region of optimality that encompasses both the inversion model and the true model.

Many researchers have underemphasized that (2.4) is a linear approximation of (2.3) and is thus only valid in a small neighborhood about the current point in model space. This also means that **G** is only valid in this neighborhood (an important distinction for nonlinear least squares inversion). Curtis and Spencer (1999c) were the first to point this out in the literature, and van den Berg et al. (2003) are the only researchers to date who have explicitly addressed this nonlinearity.

So far, we have developed some of the basic foundations for thinking about geophysical OED, but in the preceding developments, definitions of 'survey optimality', 'data quality' and the 'strength' of the mapping between data and model space were not set forth. The following section introduces a short history of contemporary OED for earth science problems. Within this section, the concept of design optimality will come into focus, paving the way for the introduction of a novel OED methodology that is the showpiece of this entire work.

2.2 Historic Objective Functions and OED Algorithms

In this section, a number of historic OED objective functions are reported followed by a mathematical discussion that elucidates their meaning or intention. The section then

closes with a discussion of the characteristics of OED problems that has induced previous reasearchers to employ global search optimization algorithms.

Curtis (1999a) identified five OED objective functions from the literature, all of which operate on the *eigenspectrum* of $\mathbf{G}^T\mathbf{G}$:

$$\lambda\left(\mathbf{G}^T\mathbf{G}\right) \equiv \left\{\lambda_i \mid \mathbf{G}^T\mathbf{G}\mathbf{e}_i = \lambda_i \mathbf{e}_i,\ \forall i = 1,...,N\right\}. \tag{2.9}$$

This list of objective functions includes:

$$\Theta_1 = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_1} = \frac{\mathrm{tr}\left(\mathbf{G}^T\mathbf{G}\right)}{\left\|\mathbf{G}^T\mathbf{G}\right\|_2}, \text{ (Curtis \& Snieder, 1997)} \tag{2.10}$$

$$\Theta_2 = \log \lambda_k \quad \text{for pre-defined fixed } k, \text{ (Barth \& Wunsch, 1990)} \tag{2.11}$$

$$\Theta_3 = k, \text{ such that } \lambda_k > \delta \text{ for some pre-defined tolerance } \delta, \text{ (Curtis, 1999a)} \tag{2.12}$$

$$\Theta_4 = \sum_{i=1}^{N} \frac{-1}{\lambda_i + \delta} = -\mathrm{tr}\left[\left(\mathbf{G}^T\mathbf{G} + \delta\mathbf{I}\right)^{-1}\right], \text{ (Maurer \& Boerner, 1998a)} \tag{2.13}$$

$$\Theta_5 = \log\left\|\mathbf{G}^T\mathbf{G}\right\|_\delta = \sum_{i=1}^{N} \gamma_i, \text{ where } \gamma_i \equiv \begin{cases} \log \lambda_i & \text{if } \lambda_i \geq \delta \\ \text{Penalty} & \text{if } \lambda_i < \delta \end{cases}, \text{ (Rabinowitz \& Steinberg, 1990; Steinberg } \textit{et al.}\text{, 1995)} \tag{2.14}$$

Other objective functions that have recently been proposed include:

$$\Theta_6 = \prod_{i=1}^{N} \lambda_i = \det\left(\mathbf{G}^T\mathbf{G}\right), \text{ (Narayanan } \textit{et al.}\text{, 2004)} \tag{2.15}$$

$$\Theta_7 = \frac{\sum_{i=1}^{\mathcal{D}}\sum_{j=1}^{\mathcal{M}} \alpha_j \left|G_{ij}\right|}{\mathcal{D}\sum_{j=1}^{\mathcal{M}} \alpha_j}, \text{ (Furman } \textit{et al.}\text{, 2004),} \tag{2.16}$$

where $\mathcal{M}$ is the number of model parameters, $\mathcal{D}$ is the number of data observations, and α_j is a weighting term corresponding to the j^{th} model parameter that allows the user to 'focus' an experimental design on particular parameters or regions. Additionally, Stummer *et al.* (2004) and Wilkinson *et al.* (2006a; 2006b) have proposed somewhat complicated objective functions that effectively attempt to maximize the similarity between the *model resolution matrices* of a designed experiment and the (hypo-

thetical) experiment of all possible observations. Briefly, the model resolution matrix is most generally defined as

$$\mathbf{R} = \mathbf{G}^{*}\mathbf{G} \tag{2.17}$$

where $\mathbf{G}^{*}$ is the generalized inverse of **G** (Backus & Gilbert, 1970; Menke, 1989). The resolution matrix is a square symmetric matrix, having the same number of dimensions as the model vector, **m**. Its columns (or rows) specify the degree to which each model parameter in **m** can be resolved. If the i^{th} column of **R**, which corresponds to the i^{th} model parameter in **m**, takes a value of 1 at its i^{th} index and zeros elsewhere, we say the that the i^{th} model parameter is perfectly resolvable, given the experiment. Otherwise, the i^{th} model parameter is obviously *not* perfectly resolvable (Backus & Gilbert, 1970), and this means that the inversion is ill-posed and must rely on some form of model regularization. Hence, the resolution matrix will be the identity matrix whenever the inverse problem is well-posed and something other than the identity matrix when it is ill-posed.

It is instructive to consider the meaning of these various objective functions listed in Equations (2.10) - (2.16), in order to establish a framework for the reader to understand contemporary OED, which will also clarify the motivation for the novel OED method introduced in this research. As most of these objectives depend on the eigenspectrum (Equation (2.9)), a natural starting point is to discuss the Singular Value Decomposition (SVD) and Spectral Decompositions of a Jacobian matrix, **G**, and its auto-inner product, $\mathbf{G}^{T}\mathbf{G}$, respectively.

2.2.1 The Singular Value Decomposition of G

The SVD of a sensitivity matrix, **G**, is

$$\mathbf{G} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{T}, \tag{2.18}$$

where **U** and **V** are orthonormal bases spanning data and model space, respectively, (Lanczos, 1956; Golub & Van Loan, 1996; Strang, 2003) and $\boldsymbol{\Sigma}$ is the diagonal matrix of singular values, σ_i, of **G**. To appreciate that **U** and **V** respectively span data and model space, observe that **U** is a square matrix whose columns and rows, by defini-

tion, are orthonormal. If **G** has $\mathcal{D}$ rows, each of which corresponds to the sensitivity kernel of a single datum, then **U** has dimensions of $\mathcal{D} \times \mathcal{D}$. Hence, **U** contains a set of mutually orthonormal columns that necessarily span $\mathcal{D}$-dimensional data space. By the same argument, if **G** has $\mathcal{M}$ columns, each of which corresponds to the sensitivities of a single model parameter to all data observations, then **V** has dimensions of $\mathcal{M} \times \mathcal{M}$. Ergo, **V** contains a set of mutually orthonormal columns that span $\mathcal{M}$-dimensional model space. We now avail ourselves of two facts: (1) any matrix of orthonormal vectors is a rotation matrix and is therefore length-preserving and (2) $\mathbf{U}^T\mathbf{U} = \mathbf{I}$. Therefore, without loss of generality, an equivalent form of Equation (2.4) is

$$\mathbf{\Sigma}\Delta\mathbf{m}' = \Delta\mathbf{d}', \tag{2.19}$$

where $\Delta\mathbf{m}' \equiv \mathbf{V}^T\Delta\mathbf{m}$ and $\Delta\mathbf{d}' \equiv \mathbf{U}^T\Delta\mathbf{d}$. In effect, Equation (2.19) expresses the same mathematical relationship as Equation (2.4), except that both data and model vectors have been rotated into a different coordinate system. In this transformed system, the sensitivity kernels of each observation (the rows of **G**) have been rotated such that their interactions (or cross-talk, if you will) have been zeroed out. Equation (2.19) expresses a set of *linearly independent* equations, such that the j^{th} component of $\Delta\mathbf{m}'$ interacts solely with to the j^{th} component of $\Delta\mathbf{d}'$. Though $\Delta\mathbf{m}'$ and $\Delta\mathbf{d}'$ do not obviously relate to real quantities in the physical world, this coordinate system provides a mathematical artifice allowing us to perceive the 'strength' of the connection between data and model space. In this coordinate system, the larger is σ_j (the j^{th} element along the diagonal of **Σ**), the stronger is the connection between the j^{th} elements of $\Delta\mathbf{m}'$ and $\Delta\mathbf{d}'$. Hence, the singular values of **G**, contained along the diagonal of **Σ**, *inform* the strength of connection or bridge between data and model space.

2.2.2 The Spectral Decomposition of $\mathrm{G}^{\mathbf{T}}\mathrm{G}$

The leap from the SVD of **G** to the spectral decomposition of $\mathbf{G}^T\mathbf{G}$ is a small one, for it is straightforward to see that if **G** is decomposed as in Equation (2.18) then $\mathbf{G}^T\mathbf{G}$ is given by

$$\mathbf{G}^T\mathbf{G} = \mathbf{V}\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\mathbf{V}^T . \tag{2.20}$$

The same arguments that applied above apply here, except that (again without loss of generality) Equations (2.4) and (2.19) are now equivalently expressed as

$$\boldsymbol{\Sigma}^T\boldsymbol{\Sigma}\Delta\mathbf{m}' = \Delta\mathbf{d}'' , \tag{2.21}$$

where $\Delta\mathbf{d}'' \equiv \boldsymbol{\Sigma}^T\mathbf{U}^T\Delta\mathbf{d}$. Effectively, Equation (2.19) is the same as (2.21), except that (2.19) has been left-multiplied by an additional $\boldsymbol{\Sigma}^T$ term. Customarily, the spectral decomposition denotes the eigenvalue matrix as $\boldsymbol{\Lambda}$, where

$$\boldsymbol{\Lambda} = \boldsymbol{\Sigma}^T\boldsymbol{\Sigma} ; \tag{2.22}$$

hence the i^{th} *eigenvalue* of $\mathbf{G}^T\mathbf{G}$ is seen to be equivalent to the square of the i^{th} *singular value* of $\mathbf{G}$:

$$\lambda_i \equiv \sigma_i^2 . \tag{2.23}$$

Consequently, the notion of the *information* that the singular values of $\mathbf{G}$ provide about the connection between data and model space is essentially identical to the notion of the information that the eigenvalues of $\mathbf{G}^T\mathbf{G}$ provide. The only difference is that, in the latter case, the singular values have been squared.

It is worth mentioning that $\Delta\mathbf{d} = \mathbf{d}_{obs} - \mathbf{d}_{pred}$, where $\mathbf{d}_{obs}$ is the vector of observed data and $\mathbf{d}_{pred}$ is the vector of predicted data for model $\mathbf{m}$. The observed data vector will typically contain data noise so that $\mathbf{d}_{obs}$ can be partitioned into

$$\mathbf{d}_{obs} = \mathbf{d}_{true} + \mathbf{e} , \tag{2.24}$$

where $\mathbf{d}_{true}$ is the 'true' data, without noise, and $\mathbf{e}$ is the vector of data noise. Consequently,

$$\begin{aligned} \Delta\mathbf{d}' &= \mathbf{U}^T\Delta\mathbf{d} = \mathbf{U}^T\left(\mathbf{d}_{true} - \mathbf{d}_{pred}\right) + \mathbf{U}^T\mathbf{e} \\ \Delta\mathbf{d}'' &= \boldsymbol{\Sigma}^T\Delta\mathbf{d}'' = \boldsymbol{\Sigma}^T\mathbf{U}^T\left(\mathbf{d}_{true} - \mathbf{d}_{pred}\right) + \boldsymbol{\Sigma}^T\mathbf{U}^T\mathbf{e} \end{aligned} , \tag{2.25}$$

showing how data noise in the original, untransformed system is projected into the transformed systems. In the SVD case, data errors project onto $\mathbf{U}$, and in the spectral decomposition case, they project onto $\mathbf{U}\boldsymbol{\Sigma}$. In terms of least squares, the more these projections 'reject' the energy in $\mathbf{e}$, the more accurate the solution to the inverse prob-

lem will be, because, in a real sense, the undesirable information in **e** is not permitted to map from data space to model space.

2.2.3 Understanding Historic OED Objective Functions

With the insights gathered by studying the singular- and eigenspectra (the ordered sets of diagonals of $\mathbf{\Sigma}$ and $\mathbf{\Lambda}$, respectively) of $\mathbf{G}$ and $\mathbf{G}^T\mathbf{G}$, respectively, we are now poised to examine the significance of the OED objective functions posed in Equations (2.10) - (2.15).

Before we begin this discussion, it is useful to make a brief detour to learn about so-called *power means* (Bullen, 2003). A p-mean, or power mean of order p, of a set of numbers, $\mathbf{x}$, is of the form

$$m_p(\mathbf{x}) \equiv \left(\frac{1}{n} \sum_{j=1}^{n} x_j^p \right)^{1/p}, \tag{2.26}$$

where x_j is the j^{th} element in a set or vector, $\mathbf{x}$. Equation (2.26) is a generalization of the 'mean' and encompasses such averages as the arithmetic, harmonic and geometric means. To see this, consider the case where $p = 1$, which gives rise to the arithmetic mean, or the case where $p = -1$, which gives rise to the harmonic mean, or the case where $p = 2$, which gives rise to the root-mean-square. The case where $p = 0$ can be approached through the theory of limits to show that it gives rise to the geometric mean (Bullen, 2003). Other 'means' that can be calculated by this formula include m_∞ and $m_{-\infty}$, which can be shown to equal the maximum and minimum of the set $\mathbf{x}$, respectively (Bullen, 2003). One of the intriguing (and useful) properties of p-means is that they are monotonic increasing as a function of p (Bullen, 2003), when the values in $\mathbf{x}$ are nonnegative. That is, $m_p \le m_{p+1}$ for all p, so long as the elements of $\mathbf{x}$ are strictly nonnegative. We can appreciate this by observing that, for a set of nonnegative numbers, the harmonic mean is always less than or equal to the geometric mean, and the geometric mean is always less than or equal to the arithmetic mean. The monotonicity of Equation (2.26) in p is therefore relevant to singular or eigenspectra, which by definition comprise a set of nonnegative scalars.

Based on the preceding facts, the reader may have already intuited that power means differ in their sensitivity to the values in **x**, depending on the power, p, to which its elements have been raised. Indeed, negative power means ($p < 0$) are biased toward the small values in **x**, while positive power means are biased toward large values, until, in the limit as p approaches $\pm\infty$, the minimum or maximum values of **x** respectively attain (Bullen, 2003).

Several of the historic objective functions in Equations (2.10) - (2.15) can be cast in terms of the power means of eigenspectra, particularly functions Θ_1, Θ_4, Θ_5 and Θ_6. Let us examine how this is so and what that means in terms of experimental design. Recall that Equations (2.19) and (2.21) express the same relationship between data and model as (2.4) but in a rotated reference frame that allows us to more easily perceive the 'strength' of the connection between the two spaces from which the data and model vectors derive. We treat the singular or eigenvalues of **G** or $\mathbf{G}^T\mathbf{G}$ expressed in these reference frames, respectively, as correspondent with the *information* that the experiment provides about the connection between data and model space. The greater are the singular values, σ_j, or their squares, λ_j, the greater is the connection *informing* the mapping between data and model space.

From an experimental design standpoint, it is of course impossible to control the individual singular or eigenvalues simultaneously. It should not be assumed that an observation sensitivity kernel (a single row of **G**) communicates its information to only one singular value in **G**, which could only be the case if the sensitivity kernel were exactly parallel to one singular vector in **V**. Rather, a sensitivity kernel typically *distributes* its 'information' over many (possibly all) singular values because multiple singular vectors will generally span its components. This shows that experimental design exercises cannot, in general, exert precise control over individual singular- or eigenvalues. For example, the addition of one observation to an experiment obviously entails the introduction of its sensitivity kernel as a new row in **G**, and this sensitivity kernel will generally not be exactly parallel to only one singular vector in **V**; instead, the information in the sensitivity kernel ends up being distributed between

multiple singular values, in proportion to the degree of parallelism the sensitivity kernel has with each of the singular vectors in **V**.

It is a consequence of the foregoing discussion that the historic objective functions listed above were primarily posed in terms of the complete set of eigenvalues or the eigenspectrum. There are a couple exceptions to this, and they will be discussed in due course. In a sense, these objective functions pose various measures of *global* information or sensitivity, not distinguishing between individual eigenvalues but lumping them together in various ways, which returns us to power means. Objective functions Θ_1, Θ_4 and Θ_6 are all seen to be variations on a theme of power means, for they can be equivalently represented as

$$\Theta_1 = \sum_{i=1}^{N} \frac{\lambda_i}{\lambda_1} = \frac{N}{\lambda_1} m_1(\boldsymbol{\lambda}) \tag{2.27}$$

$$\Theta_4 = -\sum_{i=1}^{N} \frac{1}{\lambda_i + \delta} = -\left(N m_{-1}(\boldsymbol{\lambda} + \delta)\right)^{-1} \tag{2.28}$$

and

$$\Theta_6 = \prod_{i=1}^{N} \lambda_i = \left(m_0(\boldsymbol{\lambda})\right)^N . \tag{2.29}$$

Because these are objective functions, we seek to find their maxima. Hence, without loss of generality these expressions can be simplified to

$$\Theta_1' = \frac{1}{\lambda_1} m_1(\boldsymbol{\lambda}) \tag{2.30}$$

$$\Theta_4' = m_{-1}(\boldsymbol{\lambda} + \delta) \tag{2.31}$$

and

$$\Theta_6' = m_0(\boldsymbol{\lambda}) . \tag{2.32}$$

These simplified objective functions have the same stationary points as those above and therefore could be substituted in the optimization problem if one wished. In this context, we are trying to understand what these objectives mean. Cast in terms of power means, it is evident that these three objectives equate experiment optimality with various averages of the eigenspectrum. As discussed above, if we think of the

eigenvalues of $\mathbf{G}^T\mathbf{G}$ as an information bridge that connects data and model spaces, then *these objectives propose to maximize the average information an experiment provides*. Each objective function does this a bit differently. Based on our knowledge of power means, it should be clear that the (normalized) arithmetic mean in (2.30) is more sensitive to large eigenvalues than small ones, while the geometric and harmonic means in (2.32) and (2.31), respectively, are progressively more sensitive to small eigenvalues (and less so to large ones). From an inversion point of view, the latter two objectives would be better at guarding the inverse problem from the destabilizing influence of small eigenvalues than the first, but the choice of OED objective is up to the user and should depend on the particulars of the inverse problem.

Θ_5 (Equation (2.14)) is similar to the geometric mean of the eigenspectrum because the product of N numbers is easily converted into a sum of their logarithms by taking the natural log of the product. The natural logarithm is monotonic increasing; so stationary points of the logarithm of an objective function are equal to the stationary points of the objective function itself. Hence, up to the special penalty terms that Rabinowitz and Steinberg (1990; 1995), Θ_5 is identical to Θ_6 and Θ_6' (Equations (2.15) and (2.32), respectively), which has been discussed above.

Objective functions Θ_2 and Θ_3 (Equations (2.11) and (2.12)) do not look at averages of the eigenspectrum (information averages) but rather at a single, indexed eigenvalue in the spectrum. Barth and Wunsch's (1990) $\log \lambda_k$-objective seeks an experiment that simply maximizes the k^{th} eigenvalue, where k might correspond with the smallest eigenvalue or, more precisely, the smallest non-zero eigenvalue (which is relevant for rank-limited sensitivity matrices). Curtis's (1999a) k-objective approach is somewhat similar, except that it attempts to maximize the *number* of eigenvalues greater than a prescribed tolerance, δ. Thinking again in terms of eigenvalues as measures of information, both these methods attempt to maximize the total information that an experiment provides, but they do so by disregarding all eigenvalues with indices greater than k. In this manner, designed experiments will be more biased toward the eigenvalues that *can* be made large, rather than those that cannot. These

methods intentionally sacrifice or ignore the information that experiments provide about the most insensitive model parameters, because this information (their sensitivities) will always be irremediably small. Philosophically, it is as if to say 'it is better to expend time designing experiments that maximize *useful,* rather than useless, information'. Effectively, there is no point trying to design an experiment that concerns itself with maximizing the sensitivity of a model parameter whose sensitivity can never exceed 10^{-15}, say; the model parameter will always be numerically irresolvable.

Objective function Θ_7 (Equation (2.16)) expresses a weighted-average cumulative sensitivity, an alternate measure of global sensitivity or information, where the absolute values of each model parameters' sensitivities (columns of **G**) are summed up and then weighted according to focusing criteria (Furman *et al.*, 2004). If no focusing is desired, all weight factors are set to unity. For comparative reasons, we consider here only the case where these weight factors *are* set to unity, so that Θ_7 quantifies truly *global* information like the preceding objective functions. In this case, Θ_7 simplifies to

$$\Theta_7' = \frac{1}{\mathcal{M}\mathcal{D}} \sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{M}} \left| G_{ij} \right| \tag{2.33}$$

As an aside, the Frobenius norm of **G** is denoted as

$$\left\| \mathbf{G} \right\|_F \equiv \sqrt{\sum_{i=1}^{\mathcal{D}} \sum_{j=1}^{\mathcal{M}} \left| G_{ij} \right|^2}\,, \tag{2.34}$$

(Golub & Van Loan, 1996) which bears a close resemblance to (2.33). The square of the Frobenius norm can be expressed equivalently in a number of useful ways, such as

$$\begin{aligned} \left\| \mathbf{G} \right\|_F^2 &= \sum_{i=1}^{\mathcal{M}} \sum_{j=1}^{\mathcal{D}} \left| G_{ij} \right|^2 \\ &= \mathrm{tr}\left(\mathbf{G}^T \mathbf{G} \right) \\ &= \sum_{n=1}^{N} \lambda_n \\ &= N\, m_1\left(\boldsymbol{\lambda} \right) \end{aligned}\,, \tag{2.35}$$

where $\mathrm{tr}(\bullet)$ is the trace operator and λ_n is the n^{th} of N eigenvalues of $\mathbf{G}^T\mathbf{G}$ (Golub & Van Loan, 1996). But for the square term, Equation (2.33) and (2.35) are identical,

showing that the objective function of Furman *et al.* (2004) is strikingly similar to the arithmetic mean of the eigenvalues. Therefore, based on the preceding discussion, we surmise that this objective function must be more sensitive to large eigenvalues than to small ones. In any case, this objective function is another example of measuring experiment optimality according to an average of the global information (sensitivity) an experiment provides.

As mentioned above, resolution-matrix-based objective functions (Stummer *et al.*, 2004; Wilkinson *et al.*, 2006a; Wilkinson *et al.*, 2006b) have also recently been introduced. These objective functions particularly address the situation where the inverse problem *cannot* be well posed, which means that the resolution matrices their objective functions operate upon are not identity matrices.

These are sophisticated, technical objective functions that cannot easily be recapitulated in a single expression, and the interested reader is referred to the listed citations. Nonetheless, it is possible to briefly, verbally describe them. The primary distinction between these and other OED objective functions is that they compare candidate experimental designs with the hypothetical experiment comprising all possible observations, under the assumption that this 'all-encompassing' experiment is the best (though impractical) experiment to query a target site. The purpose of these objective functions boils down to minimizing the disparity between the 'quality' of the designed experiment and the 'all-encompassing' experiment. This is actualized through minimizing a weighted difference between the resolution matrix of a candidate experimental design and the resolution matrix of the 'all-encompassing' experiment.

These resolution-matrix-based approaches are not particularly approachable through eigenanalysis except to observe that the *comparative* approach the authors have developed, which looks at the disparity between the 'quality' of a candidate design versus the 'quality' of an 'all-encompassing' design, does in a sense compare their eigenspectra. The more the eigenspectrum of the candidate design matches that of the 'all-encompassing' design, the higher is its quality metric. In other words, the more the eigenspectrum of a designed experiment matches that of the 'all-

encompassing' design, the closer it approaches to the 'ideal' information afforded by the 'all-encompassing' experiment.

2.2.4 Historic OED Optimization Algorithms

At heart, all geophysical OED exercises come down to some sort of optimization problem. An objective function that quantifies experiment 'quality' must be posed, and some sort of solver algorithm must then be applied to find its stationary points. Complicating is the fact that a solver algorithm necessarily depends on the particulars of the objective function, and not all solver algorithms are equal in the speed at which they converge. Now that we have an intuitive sense for what the several historic objective functions in the previous section were crafted to achieve, we now turn our attention to the solver algorithms that can find their stationary points.

This discussion shall be brief, for there are only a few optimization algorithms that are readily applicable to the objective functions in Equations (2.10) - (2.16). The minima/maxima of nearly all the historic objective functions listed above can only reasonably be found by using so-called *global search strategies*, such as the Genetic and Simulated Annealing algorithms (Smith *et al.*, 1992). For example, Curtis and Snieder (1997), Maurer and Boerner (1998a), Curtis (1999a; 1999b), Furman *et al.* (2004) have all used genetic algorithms, while and Barth and Wunsch (1990) and Hardt and Scherbaum (1994) have used simulated annealing.

The reason for using stochastic optimization techniques to find stationary points of these objective functions, rather than using faster gradient-following approaches (e.g., least squares, steepest descent, etc.), stems from two simple facts.

First, by the chain rule, the partials of an eigenvalue-dependent objective function, with respect to a data observation are of the form

$$\frac{\partial \Theta(\boldsymbol{\lambda})}{\partial \omega_j} = \frac{\partial \Theta}{\partial \boldsymbol{\lambda}} \frac{\partial \boldsymbol{\lambda}}{\partial \omega_j}, \tag{2.36}$$

where $\boldsymbol{\lambda}$ is the vector of eigenvalues of $\mathbf{G}$ (eigenspectrum) and ω_j is the j^{th} observation in an experiment. The partials of Θ with respect to $\boldsymbol{\lambda}$ are easy enough to determine,

but it is not at all obvious how one derives the partials of an *eigenvalue* with respect to an observation, making Equation (2.36) intractable.

Second, there are frequently experimental design situations where the partial derivatives of Θ with respect to ω_j are simply undefined, because an infinitesimal perturbation of ω_j is itself undefined. For example, a geoelectrical exercise where the relative positions of electrodes are fixed precludes the concept of an infinitesimal perturbation of a data station. Examples of this include resistivity borehole logging tools, whose electrodes are in fixed relative positions, or 2D surface resistivity surveys where the electrodes are positioned at equidistant intervals. Clearly, such cases preclude the notion of minute changes in observation position. Instead, the OED optimization problem is *combinatoric*; there are a finite (but usually large) number of *discrete* experiments to choose from. The set of all such experiments defines a *discrete experimental space*, where the notion of gradients is undefined. Currently, such combinatoric optimization problems can only be solved using stochastic search algorithms like those mentioned above. Of course, one can also conceive of experimental design situations where experimental space is continuous; that is, infinitesimal changes in the position of sources and receivers are permitted; but, in the case of geoelectrical problems for example, where two sources and two receivers are needed to make each observation (because of the dipolar nature of the EM domain), the design problem still has a combinatoric overtone because each data observartion entails the *combination* of four electrode positions.

Though stochastic (global) search methods are best for handling local minima in the objective function landscape, they are computationally expensive. The search spaces upon which they work can be amazingly large. For example, if we wish to collect just 20 from a set of 2000 available observations, there are ${}_{2000}C_{20} > 10^{47}$ possible combinations to choose from. Exhaustively evaluating the OED objective function for this many experiment combinations is clearly impossible. While genetic and simulated annealing algorithms make the problem tractable, efficiently reducing the number of objective function evaluations by many orders of magnitude, they still require hundreds of thousands or even millions of such evaluations (Smith *et al.*, 1992).

When we further consider that many of the historic objective functions require either complete or partial eigenanalysis, whose computations go as $O(n^3)$ (Golub & Van Loan, 1996) or $O(n^2)$ (Curtis *et al.*, 2004), it is easy to appreciate that these optimization techniques can take a very long time to converge, for these computations must be executed perhaps millions of times. Surprisingly few authors have actually reported the CPU times it took to run their optimization algorithms (perhaps because these times were so large!). Of those who did, CPU times of between 1 and 24 hours have been reported for OED problems implemented by genetic algorithms (Curtis & Snieder, 1997; Wilkinson *et al.*, 2006b). From a practical standpoint, OED techniques that require hours or days to establish optimal experiments are of limited use. Such lengthy design times prohibit their use in many real-world field exercises. For example, large-scale geoelectrical water/geothermal prospecting exercise that need to image an unknown subsurface over many square kilometers in a short period of time could not reasonably employ most current methods of OED. On the other hand, scenarios where lengthy OED computation times would be acceptable would be in long-term monitoring operations or in planning oilfield operations where rig time is far more expensive than computation time.

2.3 Novel Optimal Experimental Design Objective Functions

Two novel OED objective functions are proposed in this section. These functions attempt to satisfy the dual objectives of maximizing the information magnitude and complementarity provided by designed experiments. It will be seen that, in a sense, information magnitude ensures a strong link between data and model space, while information complementarity ensures that this link is well balanced, preventing unnecessary bias toward the most well-resolved model parameters. Most importantly, a deterministic greedy design algorithm based on these concepts is introduced that sequentially builds experiments, one observation at a time. This approach allows optimal experiments to be designed in a matter of minutes or even seconds (depending on

the size of the inverse problem), thus bringing geophysical OED one step closer to real-world applicability.

2.3.1 Information Magnitude and Complementarity

Recall from previous discussion that we have been associating the term *information* with the sensitivities of the forward operator contained in the Jacobian matrix, **G**. Particularly, we may think of the sensitivity kernel of an observation (Equation (2.6)) as being a multidimensional information vector corresponding to the observation. In semi-qualitative terms, the degree of linear independence of one sensitivity kernel with respect to other sensitivity kernels is a measure of the relative *information complementary* of the corresponding observation. To see this, consider two data stations whose sensitivity kernels are highly correlated; though they are separate observations, the high correlation of their sensitivity kernels means they provide essentially the same information about the target. In contrast, two stations whose sensitivity kernels are minimally correlated provide *complementary information*. Calling this idea 'complementarity' is *a propos* both logically and mathematically. Logically, two things are complementary if each 'supplies' something the other lacks. Mathematically, two angles are complementary if they sum to 90 degrees; and if the angle between two sensitivity kernels is 90 degrees, their correlation is zero and they are linearly independent. Hence, simply put, information complementarity addresses the mutual orthogonality, or linear independence, of the sensitivity kernels in an experiment.

Cast in this light, it is straightforward to appreciate the meaning of *information magnitude*. If complementarity deals with the mutual orthogonality of sensitivity kernels, then magnitude is simply the length or vector-norm of a sensitivity kernel. The greater is the norm of its sensitivity kernel, the greater is the relative information magnitude that the corresponding observation supplies connecting data and model space.

Taken together, the relative magnitude and complementarity of the information vector of a data station, with respect to other data stations, is a useful measure of its relative importance in an experiment. In the end-member case, an ideal experiment

would create a sensitivity matrix that was an orthogonal basis; that is, the inner products of all information vectors (the sensitivity kernels along the rows of the Jacobian) would yield an identity matrix (up to a multiplicative factor). In such an ideal case, each datum would provide a wholly unique piece of information about the target site, in such a way that the magnitude of its information vector would be equal to the magnitudes of all other information vectors.

What is being described above relates in part to the *resolution* and *information density matrices*:

$$\begin{aligned} \mathbf{R} &\equiv \mathbf{G}^{*}\mathbf{G} \\ \mathbf{D} &\equiv \mathbf{G}\mathbf{G}^{*} \end{aligned}, \tag{2.37}$$

(Backus & Gilbert, 1970). An ideal experiment would cause both matrices to be the identity. However, this is not the whole story. It is well known that the SVD of a matrix can be *truncated* so that only the left and right singular vectors corresponding to nonzero singular values are retained. This truncated SVD is identical to the full SVD in the sense that both matrix products equal the original matrix:

$$\mathbf{G} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^{T} = \tilde{\mathbf{U}}\tilde{\boldsymbol{\Sigma}}\tilde{\mathbf{V}}^{T}, \tag{2.38}$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are the columns of **U** and **V**, respectively, corresponding to nonzero singular values and where $\tilde{\boldsymbol{\Sigma}}$ is a square matrix (a submatrix of $\boldsymbol{\Sigma}$) containing the nonzero singular values of **G** (Lanczos, 1956; Golub & Van Loan, 1996). Consequently, both **R** and **D** can be expressed in terms of their truncated singular value decompositions (Equation (2.18)):

$$\begin{aligned} \mathbf{R} &\equiv \mathbf{G}^{*}\mathbf{G} = \tilde{\mathbf{V}}\tilde{\mathbf{V}}^{T} \\ \mathbf{D} &\equiv \mathbf{G}\mathbf{G}^{*} = \tilde{\mathbf{U}}\tilde{\mathbf{U}}^{T} \end{aligned}, \tag{2.39}$$

again where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are respectively the truncated left and right singular vectors of **G** (Lanczos, 1956; Menke, 1989; Tarantola, 2005). Singular vectors are unit vectors, so neither the resolution matrix nor the information density matrix contains information on the singular values of **G**. Hence, while these matrices inform the *linear independence* of data stations, they contain no detail on the relative 'strength' of the information bridging model parameters and data; i.e., they address information com-

plementarity but not magnitude. As alluded to previously, that information is embedded in the singular values of **G**. Ergo, like the historic objective functions discussed above, our objective functions must also somehow incorporate the singular spectrum, either implicitly or explicitly.

2.3.2 Establishing the OED Objective Functions

According to the previous development, our experimental design objective functions must comprise two measures of quality: (1) the linear independence and (2) the magnitude of each data station's information vector or sensitivity kernel. Put in geometric terms, each row of **G** is an information vector for one observation; the objective is to simultaneously maximize both the magnitude and the mutual orthogonality of these vectors.

As will be seen below, the key distinction to this objective is that, unlike nearly all the historic objective functions discussed previously, it can be maximized without calculating the SVD of **G** at all (or, at least without having to initially calculate it), which is an $O\left(n^3\right)$ operation. Rather, we propose an objective function whose main computational expense is in performing vector-space projections. Additionally, it will be shown that this objective function is amenable to *sequential* experimental design via a greedy optimization algorithm. In contrast to the historic design strategies, which require that the entire experiment be designed all at once, our optimization technique executes sequentially, adding observations to a *base experiment* one at a time.

2.3.3 Objective Function #1: Basic Algorithm

The first proposed objective function is most easily expressed by outlining the algorithm into which it is embedded (see Flowchart 2-1); each algorithm step is annotated below.

1-i. The initial observation goes into the *base experiment,* which will be denoted Ω_b, where $\Omega_b \subset \{\Omega \mid \text{the set of all permitted observations}\}$. This first observation is the one whose sensitivity kernel is of maximal length (as will be shown later, the choice for this

initial observation maximizes an alternate but equivalent expression of the experimental design objective function). Construct the sensitivity matrix for Ω_b and denote this $\mathbf{G}_b$. Define Ω_c to be the *complement* of Ω_b; i.e., $\Omega = \Omega_b \cup \Omega_c$. Hence, Ω_c is the set of remaining *candidate observations*. Finally, define $\mathbf{G}_c$ to be the sensitivity matrix for Ω_c. It follows by these definitions that $\mathbf{G}^T = \left[\mathbf{G}_c^T \vdots \mathbf{G}_b^T\right]$, where **G** is the sensitivity matrix of all permitted observations, corresponding to Ω.

1-ii. Increment *n*.

1-iii. Evaluate the candidacy of each observation remaining in Ω_c for addition to the base experiment, Ω_b. For each row in $\mathbf{G}_c$ (the sensitivity kernels of the candidate observations remaining in Ω_c), subtract from it its projection onto the row space spanned by $\mathbf{G}_b$, and then evaluate the norm of the residual. The residual is perpendicular to the row space spanned by $\mathbf{G}_b$, so it represents the *complementary information* the corresponding observation would potentially bring to the base experiment. The magnitude of this residual therefore quantitatively specifies the strength of this complementary information.

Mathematically, this metric is expressed as

$$\Theta(k) = \left\| \left(\mathbf{I} - \mathbf{P}_b\right) \mathbf{g}_c^{(k)} \right\|, \tag{2.40}$$

where $\mathbf{P}_b$ is the *row-space projection matrix* of $\mathbf{G}_b$, defined

$$\Theta(k) = \left\| \left[\mathbf{I} - \mathbf{G}_b^T \left(\mathbf{G}_b \mathbf{G}_b^T\right)^{-1} \mathbf{G}_b\right] \mathbf{g}_c(k) \right\|, \tag{2.41}$$

and $\mathbf{g}_c^{(k)}$ denotes the sensitivity kernel of the k^{th} candidate observation in Ω_c and is just the k^{th} row of $\mathbf{G}_c$:

$$\mathbf{g}_c^{(k)} \equiv \mathbf{G}_c(k,:)^T, \tag{2.42}$$

expressed in colon notation. Note that Equation (2.41) can be executed rapidly by left-multiplying the projection matrix with $\mathbf{G}_c^T$ and taking the square root of the sum of the squares of the columns of the resulting matrix. This avoids using FOR loops.

1-iv. Once all candidate observations have been evaluated according to (2.40), the one that maximizes $\Theta(k)$ is added to the base experiment.

1-v. Choice of stopping criterion is up to the user. Here, we specify a maximum number of observations, $\mathcal{D}$. Other possibilities might include: a threshold term on the smallest eigenvalue (or any eigenvalue of specified index); an information improvement threshold requiring that at least one candidate observation evaluates (according to (2.40)) to above some specified threshold; et cetera.

1-vi. If the stopping criterion is met, the OED algorithm may exit. Otherwise, go to step (*ii*) and repeat. ■

Equation (2.40) can be expressed in a number of equivalent ways. It can be simplified (for heuristic if not for practical purposes) to

$$\Theta(k) = \left\| \mathbf{N}_b^T \mathbf{g}_c^{(k)} \right\|, \tag{2.43}$$

where $\mathbf{N}_b$ denotes the null space of the Jacobian of the base experiment, $\mathbf{G}_b$. Bear in mind that the null space of a matrix is the *orthogonal complement* of its row space (Strang, 2003). That is, the null space of a matrix is perpendicular to its row space. Viewed in this way, the OED objective function is seen to favor candidate observations whose sensitivity kernels are of maximal length after projection onto the null space of $\mathbf{G}_b$; in other words, candidate observations in Ω_c that provide the highest-magnitude complementary information are chosen. The simple cartoon in Figure 2-4 helps to visualize what the basic design algorithm is doing.

Alternately, observe that if $\mathbf{P}_b$ is expanded by substituting the SVD of $\mathbf{G}_b$ for $\mathbf{G}_b$ itself (let $\mathbf{G}_b = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T$) and then simplified, the projection matrix equals the model resolution matrix. That is

$$\mathbf{P}_b = \mathbf{R}_b = \tilde{\mathbf{V}}\tilde{\mathbf{V}}^T \tag{2.44}$$

where $\tilde{\mathbf{V}}$ denotes the truncated right singular vectors of $\mathbf{G}_b$ (those associated with nonzero singular values). Hence, another equivalent form of $\Theta(k)$ is

$$\Theta(k) = \left\| \left(\mathbf{I} - \mathbf{R}_b \right) \mathbf{g}_c^{(k)} \right\|, \tag{2.45}$$

which, by corollary, means that the null space of $\mathbf{G}_b$ equals the identity minus the resolution matrix.

Lastly, and most practically, one can avoid explicitly calculating $\mathbf{P}_b$ or $\mathbf{R}_b$ all together (which involves either the SVD or a matrix inversion) by taking advantage of Gram-Schmidt orthonormalization (Golub & Van Loan, 1996; Strang, 2003). Each time a new data station is added to the base experiment, we can use the Gram-Schmidt method on its sensitivity kernel to create a unit residual vector that is orthogonal to the space spanned by the sensitivity kernels of the data stations already included in the base experiment. Like Eq. (2.40), Gram-Schmidt works by subtracting the projection of a new vector onto the subspace spanned by the extant Gram-Schmidt vectors (making the residual orthogonal) and then normalizing the residual to unit length (making it orthonormal) (Golub & Van Loan, 1996). In this fashion, we can build an orthonormal matrix, $\mathbf{\Gamma}_b$, which spans exactly the same row space as $\mathbf{G}_b$. Here is how this works. Suppose the sensitivity kernel of the first data station in the base experiment is $\mathbf{g}_b^{(1)}$. Its transpose occupies the first row of $\mathbf{G}_b$, but we wish to employ Gram-Schmidt, so we normalize $\mathbf{g}_b^1$ to have unit length and place the transpose of this vector in the first row of the orthonormal matrix, $\mathbf{\Gamma}_b$. Now suppose the sensitivity kernel for the next data station we add is $\mathbf{g}_b^{(2)}$. We execute Gram-Schmidt as follows:

$$\mathbf{r} = \left(\mathbf{I} - \mathbf{\Gamma}_b^T \left(\mathbf{\Gamma}_b \mathbf{\Gamma}_b^T\right)^{-1} \mathbf{\Gamma}_b\right) \mathbf{g}_b^{(2)}, \tag{2.46}$$

and define

$$\boldsymbol{\gamma} = \frac{\mathbf{r}}{\|\mathbf{r}\|}, \tag{2.47}$$

so $\boldsymbol{\gamma}$ is a unit vector orthogonal to the space spanned by the Gram-Schmidt vectors in $\mathbf{\Gamma}_b$. We then augment $\mathbf{\Gamma}_b$ by appending $\boldsymbol{\gamma}$ as a new row vector:

$$\mathbf{\Gamma}_b := \begin{bmatrix} \mathbf{\Gamma}_b \\ \boldsymbol{\gamma}^T \end{bmatrix}, \tag{2.48}$$

and Eqs. (2.46) - (2.48) are repeated each time a new data station is added to the base experiment. There are two reasons for employing the Gram-Schmidt method. First, it has already been pointed out that $\mathbf{\Gamma}_b$ is an orthonormal matrix that spans exactly the same row space as $\mathbf{G}_b$. Thus, $\mathbf{\Gamma}_b$ can be substituted for $\mathbf{G}_b$ in (2.40):

$$\Theta(k) = \left\| \left(\mathbf{I} - \boldsymbol{\Gamma}_b^T \left(\boldsymbol{\Gamma}_b \boldsymbol{\Gamma}_b^T \right)^{-1} \boldsymbol{\Gamma}_b \right) \mathbf{g}_c^{(k)} \right\|. \tag{2.49}$$

Second, $\boldsymbol{\Gamma}_b$ is orthonormal, so

$$\boldsymbol{\Gamma}_b \boldsymbol{\Gamma}_b^T = \mathbf{I}; \tag{2.50}$$

hence, (2.46) can be simplified to

$$\mathbf{r} = \left(\mathbf{I} - \boldsymbol{\Gamma}_b^T \boldsymbol{\Gamma}_b \right) \mathbf{g}_b^{(2)}, \tag{2.51}$$

and (2.49) simplifies to

$$\Theta(k) = \left\| \left(\mathbf{I} - \boldsymbol{\Gamma}_b^T \boldsymbol{\Gamma}_b \right) \mathbf{g}_c^{(k)} \right\|. \tag{2.52}$$

In words, the computationally expensive inverse in (2.40), which is needed to perform row-space projections, can be avoided by using the Gram-Schmidt orthonormalization described above. In practice, this technique is the most efficient way of executing our greedy, sequential OED algorithm.

A graphical example is provided in Figure 2-5 to offer insight into this OED objective function. It is cast in terms of the singular value decomposition of $\mathbf{G}_b$ for experiments of increasing size so the reader can visualize the concepts laid out. The example is for a borehole resistivity problem where 10 electrodes have been deployed to query a heterogeneous earth (a thorough examination of the borehole OED problem is offered in a later chapter).

2.3.4 Mathematical Significance of the Basic Algorithm

The basic design algorithm outlined above addresses the design problem in terms of information complementarity and magnitude, but it is instructive to understand what the algorithm is doing at a more fundamental mathematical level.

First, according to the OED theory laid out above, the sensitivity kernel of the next observation chosen for the experiment, denoted $\hat{\mathbf{g}}$, is of maximal length after projection onto the null space of $\mathbf{G}_b$, which is the Jacobian of the base experiment, and therefore maximizes the objective function

$$\Theta = \hat{\mathbf{g}}^T \left(\mathbf{I} - \mathbf{G}_b^T \left(\mathbf{G}_b \mathbf{G}_b^T \right)^{-1} \mathbf{G}_b \right) \hat{\mathbf{g}}, \tag{2.53}$$

where it should be remarked that Equation (2.53) is just the square of Equation (2.40). The question is: what is the mathematical significance of (2.53)? Below, we will prove that the expression in (2.53) is mathematically equivalent to the ratio of two determinants.

We must first make a short detour through a little linear algebra. Consider a block matrix:

$$\mathbf{A}' = \begin{bmatrix} \mathbf{A} \\ \mathbf{B} \end{bmatrix} \tag{2.54}$$

where $\mathbf{A} \in \mathbb{R}^{m \times n}$ and $\mathbf{B} \in \mathbb{R}^{p \times n}$ such that $m + p \le n$. Its auto-outer product is

$$\mathbf{A}'\mathbf{A}'^{T} = \begin{bmatrix} \mathbf{A}\mathbf{A}^{T} & \mathbf{A}\mathbf{B}^{T} \\ \mathbf{B}\mathbf{A}^{T} & \mathbf{B}\mathbf{B}^{T} \end{bmatrix}. \tag{2.55}$$

We want to find a simple expression for the determinant of $\mathbf{A}'\mathbf{A}'^{T}$ that depends only on **A** and **B**:

$$\det\left(\mathbf{A}'\mathbf{A}'^{T}\right) = \det\left(\begin{bmatrix} \mathbf{A}\mathbf{A}^{T} & \mathbf{A}\mathbf{B}^{T} \\ \mathbf{B}^{T}\mathbf{A}^{T} & \mathbf{A}^{T}\mathbf{A} \end{bmatrix}\right). \tag{2.56}$$

We transform $\mathbf{A}'\mathbf{A}'^{T}$ by the following multiplication:

$$\begin{bmatrix} \mathbf{A}\mathbf{A}^{T} & \mathbf{A}\mathbf{B}^{T} \\ \mathbf{B}\mathbf{A}^{T} & \mathbf{B}\mathbf{B}^{T} \end{bmatrix}\begin{bmatrix} \mathbf{I} & -\left(\mathbf{A}\mathbf{A}^{T}\right)^{-1}\mathbf{A}\mathbf{B}^{T} \\ \mathbf{0} & \mathbf{I} \end{bmatrix} = \begin{bmatrix} \mathbf{A}\mathbf{A}^{T} & \mathbf{0} \\ \mathbf{B}\mathbf{A}^{T} & \mathbf{B}\left(\mathbf{I} - \mathbf{A}^{T}\left(\mathbf{A}\mathbf{A}^{T}\right)^{-1}\mathbf{A}\right)\mathbf{B}^{T} \end{bmatrix}. \tag{2.57}$$

Taking the determinant of (2.57) yields

$$\det\left(\mathbf{A}'\mathbf{A}'^{T}\right)\det\left(\begin{bmatrix} \mathbf{I} & -\left(\mathbf{A}\mathbf{A}^{T}\right)^{-1}\mathbf{A}\mathbf{B}^{T} \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right) = \det\left(\begin{bmatrix} \mathbf{A}\mathbf{A}^{T} & \mathbf{0} \\ \mathbf{B}\mathbf{A}^{T} & \mathbf{B}\left(\mathbf{I} - \mathbf{A}^{T}\left(\mathbf{A}\mathbf{A}^{T}\right)^{-1}\mathbf{A}\right)\mathbf{B}^{T} \end{bmatrix}\right). \tag{2.58}$$

Taking advantage of the fact that the determinant of a block upper (or lower) triangular matrix is the product of the determinants of its diagonal blocks (Golub & Van Loan, 1996), the second and third determinants in Equation (2.57) reduce to

$$\det\left(\begin{bmatrix} \mathbf{I} & -\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\mathbf{B}^T \\ \mathbf{0} & \mathbf{I} \end{bmatrix}\right)=1$$

$$\det\left(\begin{bmatrix} \mathbf{A}\mathbf{A}^T & \mathbf{0} \\ \mathbf{B}\mathbf{A}^T & \mathbf{B}\left(\mathbf{I}-\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\right)\mathbf{B}^T \end{bmatrix}\right)=\det\left(\mathbf{A}\mathbf{A}^T\right)\det\left(\mathbf{B}\left(\mathbf{I}-\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\right)\mathbf{B}^T\right) \quad (2.59)$$

Substituting (2.59) into (2.58), we have the general case that

$$\det\left(\mathbf{A}'\mathbf{A}'^T\right)=\det\left(\mathbf{A}\mathbf{A}^T\right)\det\left(\mathbf{B}\left(\mathbf{I}-\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\right)\mathbf{B}^T\right). \quad (2.60)$$

If we take the special case where $\mathbf{B} = \mathbf{a}^T$, Equation (2.60) reduces to

$$\det\left(\mathbf{A}'\mathbf{A}'^T\right)=\det\left(\mathbf{A}\mathbf{A}^T\right)\left[\mathbf{a}^T\left(\mathbf{I}-\mathbf{A}^T\left(\mathbf{A}\mathbf{A}^T\right)^{-1}\mathbf{A}\right)\mathbf{a}\right]. \quad (2.61)$$

Now, letting

$$\mathbf{G}_b'=\begin{bmatrix}\mathbf{G}_b \\ \hat{\mathbf{g}}^T\end{bmatrix}, \quad (2.62)$$

and subsituting $\mathbf{G}_b$ for **A,** $\mathbf{G}_b'$ for $\mathbf{A}'$, and $\hat{\mathbf{g}}$ for **a**, Equation (2.61) says that

$$\det\left(\mathbf{G}_b'\mathbf{G}_b'^T\right)=\det\left(\mathbf{G}_b\mathbf{G}_b^T\right)\left[\hat{\mathbf{g}}^T\left(\mathbf{I}-\mathbf{G}_b^T\left(\mathbf{G}_b\mathbf{G}_b^T\right)^{-1}\mathbf{G}_b\right)\hat{\mathbf{g}}\right]. \quad (2.63)$$

Rearranging yields

$$\frac{\det\left(\mathbf{G}_b'\mathbf{G}_b'^T\right)}{\det\left(\mathbf{G}_b\mathbf{G}_b^T\right)}=\hat{\mathbf{g}}^T\left(\mathbf{I}-\mathbf{G}_b^T\left(\mathbf{G}_b\mathbf{G}_b^T\right)^{-1}\mathbf{G}_b\right)\hat{\mathbf{g}}, \quad (2.64)$$

from which it is plain that the OED objective function (Equation (2.53)) can alternately be written

$$\Theta(k)=\frac{\det\left(\mathbf{G}_b'\mathbf{G}_b'^T\right)}{\det\left(\mathbf{G}_b\mathbf{G}_b^T\right)}=\mathbf{g}_c^T(k)\left(\mathbf{I}-\mathbf{G}_b^T\left(\mathbf{G}_b\mathbf{G}_b^T\right)^{-1}\mathbf{G}_b\right)\mathbf{g}_c(k), \quad (2.65)$$

where $\mathbf{g}_c(k)$ is the k^{th} candidate from the candidate set, Ω_c. In words, the basic OED objective function expresses the ratio of the determinant of the *augmented* Jacobian over the determinant of the *base* Jacobian. In seeking the best candidate observation

via our method of projections, the OED algorithm is implicitly maximizing ratio of the next determinant to the preceding one.

Incidentally, the nullspace term betwixt the sensitivity kernels in (2.65) is idempotent; therefore, without loss of generality we see that maximizing (2.65) is equivalent to maximizing

$$\Theta(k) = \left\| \left(\mathbf{I} - \mathbf{G}_b^T \left(\mathbf{G}_b \mathbf{G}_b^T \right)^{-1} \mathbf{G}_b \right) \mathbf{g}_c(k) \right\|, \tag{2.66}$$

which is identically the objective function expression first seen in (2.43).

As an aside, a consequence of the fact that the design algorithm attempts to sequentially maximize the determinant of the Jacobian matrix is that the initial observation in the base experiment must be the one whose sensitivity kernel is of maximal length (see Flowchart 2-1, Step *1-i*). After all, the determinant of the auto-outer product of a sensitivity kernel equals its squared length, so the algorithm self-consistently chooses the observation whose sensitivity kernel is of maximum length.

Now that we understand the OED objective function to be maximizing the ratio of the deteriminants of the Jacobians of augmented and base experiments, we turn our attention to its significance from a design standpoint. The determinant of the outer product of a Jacobian matrix with itself is a measure of an experiment's *global sensitivity* to the model parameters. To see this, consider the fact that the model covariance matrix is given by

$$\mathbf{C}_m = \sigma_d^2 \left(\mathbf{G}^T \mathbf{G} \right)^{-1} \tag{2.67}$$

where σ_d^2 is the data variance (*Note: here, we assume uncorrelated data noise that is statistically invariant from data station to data station. A more general form of* $\mathbf{C}_m$ *that does not impose these assumptions exists, but the purposes of exposition, Equation (2.67) will suffice.*) The eigenvalues of $\mathbf{C}_m$ are

$$\lambda_j(\mathbf{C}_m) = \frac{\sigma_d^2}{\lambda_j\left(\mathbf{G}^T \mathbf{G} \right)}. \tag{2.68}$$

Up to zero eigenvalues, the eigenvalues of $\mathbf{G}\mathbf{G}^T$ and $\mathbf{G}^T\mathbf{G}$ are identical. Therefore, the determinant of $\mathbf{G}\mathbf{G}^T$ is the product of the nonzero eigenvalues of both $\mathbf{G}^T\mathbf{G}$ and $\mathbf{G}\mathbf{G}^T$. Thus,

$$\det\left(\mathbf{G}\mathbf{G}^T\right) = \prod_{j=1}^{n} \lambda_j = \frac{\sigma_d^{2n}}{\det\left(\mathbf{C}_m\right)}. \tag{2.69}$$

The determinant of $\mathbf{C}_m$ is a measure of the volume of parameter uncertainty at a particular point in model space (Narayanan *et al.*, 2004). Equation (2.69) thus tells us that the determinant of the outer product of the Jacobian with itself is a measure of the overall data sensitivity to the model parameters. The larger it is, the more sensitive are the experimental data to the model. This measure is also particularly sensitive to small eigenvalues because they cause the determinant to become small, immediately indicating that the experimental data are insentive to one or more parameters. Therefore, an experiment that maximizes $\det\left(\mathbf{G}\mathbf{G}^T\right)$ minimizes the last expression in (2.69) and therefore minimizes the volume of parameter uncertainty. It was demonstrated that the OED methodology maximizes the ratio of the determinants of the next and current experiments; thus, it is now clear that *this method produces a sequence of observations, each of which minimizes the parameter uncertainty volume with respect to the observations preceding it*.

Under particular conditions, the design algorithm, which sequentially maximizes (2.65), can produce a *globally* optimal experiment, where global optimality is defined as maximizing the determinant of the $\mathbf{G}\mathbf{G}^{\mathrm{T}}$ over the space of all possible experiments of the same size. If the sensitivity kernels of the all observations are linearly independent then $\mathbf{G}\mathbf{G}^T$ is a diagonal matrix whose eigenvalues are identically the elements along the diagonal. At any iteration of the sequential design procedure, the design algorithm maximizes (2.65), which can be alternately written as

$$\Theta = \frac{\prod_{j=1}^{n+1} \lambda_j'}{\prod_{j=1}^{n} \lambda_j}, \tag{2.70}$$

where λ'_j and λ_j are the eigenvalues of $\mathbf{G}'_b\mathbf{G}'^T_b$ and $\mathbf{G}_b\mathbf{G}^T_b$, respectively. But if $\mathbf{G}'_b\mathbf{G}'^T_b$ and $\mathbf{G}_b\mathbf{G}^T_b$ are diagonal matrices, taking advantage of (2.55), it follows that

$$\lambda'_j = \lambda_j \quad \text{for } j = 1,\ldots,n\,. \tag{2.71}$$

Hence, Equation (2.70) simplifies to

$$\Theta = \lambda'_{n+1}\,. \tag{2.72}$$

Therefore, at each sequence step, the design algorithm picks the observation that maximizes the next eigenvalue in the eigenspectrum of $\mathbf{G}'_b$, and since this choice does not affect the preceding eigenvalues in $\mathbf{G}_b$, the algorithm maximizes the determinant of $\mathbf{G}'_b\mathbf{G}'^T_b$ over the space of all possible experiments of $n + 1$ observations.

In general, however, the design methodology does not guarantee global optimality as defined here. Consider the following simple example: we have three available observations, of which two will be chosen for experimentation. Suppose the Jacobian matrix for all three observations is

$$\mathbf{A} = \begin{bmatrix} \mathbf{a}_1^T \\ \mathbf{a}_2^T \\ \mathbf{a}_3^T \end{bmatrix} = \begin{bmatrix} -1.45 & 2.21 & 0.01 \\ -2.32 & 0.74 & 0.27 \\ 0.71 & 1.81 & 1.10 \end{bmatrix}, \tag{2.73}$$

which yields

$$\mathbf{A}\mathbf{A}^T = \begin{bmatrix} 7 & 5 & 3 \\ 5 & 6 & 0 \\ 3 & 0 & 5 \end{bmatrix}. \tag{2.74}$$

The design algorithm will first choose the observation corresponding to $\mathbf{a}_1$ because

$$\det\left(\mathbf{a}_1\mathbf{a}_1^T\right) = 7 \geq \det\left(\mathbf{a}_2\mathbf{a}_2^T\right) = 6 \geq \det\left(\mathbf{a}_3\mathbf{a}_3^T\right) = 5\,.$$

Now the design algorithm has two choices for the second observation, and it will choose the one that produces the larger determinant of the following two submatrices:

$$\begin{bmatrix} 7 & 5 \\ 5 & 6 \end{bmatrix} \text{ and } \begin{bmatrix} 7 & 3 \\ 3 & 5 \end{bmatrix}.$$

The design algorithm will pick the observation corresponding to $\mathbf{a}_3$ because

$$\det\left(\begin{bmatrix} 7 & 5 \\ 5 & 6 \end{bmatrix}\right) = 17 \leq \det\left(\begin{bmatrix} 7 & 3 \\ 3 & 5 \end{bmatrix}\right) = 26 .$$

Thus, according to the design methodology, the first and third observations are chosen as forming an optimal experiment. However, notice that the determinate corresponding to the second and third observations is

$$\det\left(\begin{bmatrix} 6 & 0 \\ 0 & 5 \end{bmatrix}\right) = 30 ,$$

which is larger than the determinant for the 'designed' experiment. Therefore, in general, the design procedure does not produce globally optimal experiments. Rather, it offers a compromise between the computation time of OED and global optimality, by finding sub-optimal experiments in a fraction of the time.

2.3.5 Maximum Attainable Rank

The reader may have detected an oversight in Objective Function #1. It is a mathematical fact that the rank of $\mathbf{G}_b$ can never exceed the number of model parameters. To see this, note that the rank of any matrix is always less than or equal to the lesser of it number of rows and columns:

$$\operatorname{rank}(\mathbf{A}) \leq \min(R, C), \tag{2.75}$$

where $\mathbf{A}$ has R rows and C columns (Golub & Van Loan, 1996; Strang, 2003). Logically, if the earth has been discretized into $\mathcal{M}$ model cells, no matter how many observations we choose to populate the base experiment with, the rank of $\mathbf{G}_b$ can never be greater than $\mathcal{M}$.

The preceding fact is significant if $\mathcal{D}$, the maximum number of observations we wish to design our experiment with, exceeds $\mathcal{M}$. For simplicity, let us say that $\dim(\mathbf{G}_b) = \mathcal{D}\text{-}1 \times \mathcal{M}$ with $\mathcal{D}\text{-}1 \geq \mathcal{M}$ and that $\operatorname{rank}(\mathbf{G}_b) = \mathcal{M}$. We wish to add a new observation to the base experiment, but there are exactly $\mathcal{M}$ nonzero singular values that

correspond to the $\mathcal{M}$ right singular vectors in $\mathbf{V}$; hence, $\mathbf{G}_b$ has no nullspace (because the nullspace of any matrix consists of the right singular vectors associated with zero singular values). Therefore, $\Theta(k)$, as defined in Equation (2.43), evaluates to zero for all candidate observations in Ω_c. Consequently, when the sequential design algorithm reaches the *maximum attainable rank* (MAR) of the Jacobian matrix, some alteration of the objective function is needed to evaluate the candidacy of additional observations.

In later chapters, the definition of maximum attainable rank will be expanded to address ill-posed inverse problems, where the rank of the Jacobian matrix cannot even reach $\mathcal{M}$, no matter how many data observations are collected.

2.3.6 Objective Function #2: Addressing Maximum Attainable Rank

It was explained that if the number of desired observations, $\mathcal{D}$, is greater than the maximum attainable rank of the sensitivity matrix, Equations (2.40) - (2.45) will evaluate to zero for all n greater than or equal to the MAR. This follows because once $\mathbf{G}_b$ attains its maximum possible rank, it spans the same row space as $\mathbf{G}$ (the Jacobian of *all* permitted data stations); thus any candidate, $\mathbf{g}_c^{(k)}$, which by definition resides in the row space spanned by $\mathbf{G}$, must reside wholly in the row space spanned by $\mathbf{G}_b$, making it perpendicular to the nullspace of $\mathbf{G}_b$. Therefore, if $\mathcal{D}$ is larger than the MAR of $\mathbf{G}$, the objective function must be modified.

Mindful that the basic design objective function is equivalent to a ratio of determinants (per Equation (2.65)), we would like to find a way to evaluate this ratio efficiently when $\mathbf{G}_b$ reaches and exceeds its MAR. When the design algorithm is in the regime of underdetermined problems (the basic algorithm above), where there are fewer observations than unknowns, the determinant is evaluated in terms of the auto-*outer* product of a Jacobian matrix with itself (again, per Equation (2.65)). This guarantees that the determinant is nonzero because there are always exactly as many ei-

genvalues in the outer product as there are observations, and the eigenvalues are all nonzero. However, when the algorithm reaches the well-determined regime, where there are as many or more observations than unknowns, the outer product of a Jacobian with itself gives rise to zero eigenvalues, causing the determinant to be zero. At this point, instead of working with the auto-outer product, we switch to the auto-*inner* product, bearing in mind that the eigenvalues of the auto-outer and auto-inner products of a matrix are identical up to zeros. Therefore, in the well-determined regime, we must alter Equation (2.65) to

$$\Theta = \frac{\det\left(\mathbf{G}_b'^T \mathbf{G}_b'\right)}{\det\left(\mathbf{G}_b^T \mathbf{G}_b\right)}. \tag{2.76}$$

The trick is to find a way of evaluating this expression efficiently so as to maintain the computational advantage of sequential design, requiring us to make another brief algebraic tangent.

Consider the block matrix $\mathbf{A}'$ in Equation (2.54). In lieu of taking its auto-outer product as previously, we take its auto-inner product, yielding

$$\mathbf{A}'^T \mathbf{A}' = \mathbf{A}^T \mathbf{A} + \mathbf{B}^T \mathbf{B}. \tag{2.77}$$

The ratio of determinants of the augmented and base matrices is given by

$$\frac{\det\left(\mathbf{A}'^T \mathbf{A}'\right)}{\det\left(\mathbf{A}^T \mathbf{A}\right)} = \det\left[\left(\mathbf{A}^T \mathbf{A}\right)^{-1}\left(\mathbf{A}^T \mathbf{A} + \mathbf{B}^T \mathbf{B}\right)\right]. \tag{2.78}$$

Distributing through in the argument of the right hand determinant in (2.78), we have

$$\frac{\det\left(\mathbf{A}'^T \mathbf{A}'\right)}{\det\left(\mathbf{A}^T \mathbf{A}\right)} = \det\left[\mathbf{I} + \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{B}^T \mathbf{B}\right]. \tag{2.79}$$

If we take the case where $\mathbf{B}$ is a row vector, $\mathbf{B} = \mathbf{a}^T$, (2.79) becomes

$$\frac{\det\left(\mathbf{A}'^T \mathbf{A}'\right)}{\det\left(\mathbf{A}^T \mathbf{A}\right)} = \det\left[\mathbf{I} + \left(\mathbf{A}^T \mathbf{A}\right)^{-1} \mathbf{a}\mathbf{a}^T\right]. \tag{2.80}$$

Now consider the following block matrix expression:

$$\begin{bmatrix} \mathbf{a}^T & 1 \\ -\mathbf{I} & \mathbf{0} \end{bmatrix} \begin{bmatrix} \left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a} & -\mathbf{I} \\ 1 & \mathbf{a}^T \end{bmatrix} = \begin{bmatrix} 1+\mathbf{a}^T\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a} & \mathbf{0}^T \\ -\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a} & \mathbf{I} \end{bmatrix}. \tag{2.81}$$

Taking the determinant of Equation (2.81), we have

$$\det\left[\mathbf{I}+\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a}\mathbf{a}^T\right] = 1+\mathbf{a}^T\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a}; \tag{2.82}$$

and substituting back into (2.80), we get

$$\frac{\det\left(\mathbf{A}'^T\mathbf{A}'\right)}{\det\left(\mathbf{A}^T\mathbf{A}\right)} = 1+\mathbf{a}^T\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a}. \tag{2.83}$$

It is straightforward to that show Equation (2.83) can be equivalently expressed as

$$\frac{\det\left(\mathbf{A}'^T\mathbf{A}'\right)}{\det\left(\mathbf{A}^T\mathbf{A}\right)} = 1+\left\|\mathbf{A}\left(\mathbf{A}^T\mathbf{A}\right)^{-1}\mathbf{a}\right\|^2. \tag{2.84}$$

Now, letting

$$\mathbf{G}'_b = \begin{bmatrix} \mathbf{G}_b \\ \mathbf{g}_c^T(k) \end{bmatrix}, \tag{2.85}$$

where $\mathbf{g}_c(k)$ is the k^{th} sensitivity kernel in the candidate set, Ω_c, and substituting $\mathbf{G}_b$ for **A,** $\mathbf{G}'_b$ for $\mathbf{A}'$, and $\hat{\mathbf{g}}$ for **a**, Equation (2.84) says that

$$\Theta(k) = \frac{\det\left(\mathbf{G}_b'^T\mathbf{G}'_b\right)}{\det\left(\mathbf{G}_b^T\mathbf{G}_b\right)} = 1+\left\|\mathbf{G}_b\left(\mathbf{G}_b^T\mathbf{G}_b\right)^{-1}\mathbf{g}_c(k)\right\|^2. \tag{2.86}$$

Because the objective is to maximize Equation (2.86), we can simplify it without loss of generality to

$$\Theta(k) = \left\|\mathbf{G}_b\left(\mathbf{G}_b^T\mathbf{G}_b\right)^{-1}\mathbf{g}_c(k)\right\|. \tag{2.87}$$

Equation (2.87) expresses the same objective (a ratio of determinants) as Equation (2.66) except that it is arranged to handle the well-determined case, whereas the latter was arranged to handle the underdetermined case.

In light of the foregoing, Objective Function #1 can now be modified to accommodate $\mathcal{D}$ greater than the MAR. Flowchart 2-2 elucidates the idea; each step of the algorithm is annotated below.

2-i. Same as Step *1-i*.

2-ii. Same as Step *1-ii*.

2-iii. Same as Step *1-iii*.

2-iv. Same as step *1-iii* and *2-iii* except the objective function is altered per Equation (2.87) to avoid zero eigenvalues.

2-v. Same as Step *1-iv*.

2-vi. Same as Step *1-v*.

■

2.3.7 Objective Function #3: Adaptive Optimal Experimental Design

Either Objective Function #1 or #2 can be employed to perform a sort of *ad hoc* adaptive OED, wherein a preliminary data set is collected and inverted to provide an initial earth model; this working model can then be used to design a tailored experiment whose purpose is to minimize parameter uncertainty by maximizing resolution.

An alternate and possibly more elegant approach would be to design the experiment in real-time, as data are being collected and inverted. Flowchart 2-3 diagrams this idea; each step of the algorithm is annotated below.

3-i. Collect n initial observations. These initial observations can derive from a standard survey or can be designed to optimally query a homogeneous target. Make an initial guess for the earth model, $\mathbf{m}_0$, perhaps starting with a homogenous earth. Set the algorithm counter to $n = 0$.

3-ii. Increment the counter, $n := n + 1$.

3-iii. Perform one or two iterations of a nonlinear least squares inversion, using some type of regularization to ensure stability. Update the model to $\mathbf{m}_{n+1}$.

3-iv. Check *n* to determine which Step to follow next. Define $\mathcal{D}$ as the desired number of observations to populate the experiment with and n_{max} as the maximum number of iterations the coupled design/inversion algorithm is permitted ($n_{max} \geq \mathcal{D}$). If $n < \mathcal{D}$, proceed to Step (*v*), the OED step; if $n \geq \mathcal{D}$, proceed to Step (*vii*), the stopping criterion. Once *n* exceeds$\mathcal{D}$, the experimental design pathway (Steps (*v*) and (*vi*)) is deactivated and the algorithm proceeds thereafter as an ordinary NLLS inversion.

3-v. Use Objective Function #1 or #2 (depending on whether *n* is greater than the MAR) to determine the next observation to be added to the experiment. Bear in mind that $\mathbf{G}_b$, $\mathbf{G}_c$ and $\mathbf{G}$ change at each iteration because the earth model, $\mathbf{m}_n$, changes. This obviously significantly increases computational expense.

3-vi. Collect the new observation prescribed in Step *3-v*.

3-vii. The stopping criterion is a matter of user choice. The criterion specified here is a data rms error threshold coupled with a maximum permitted number of iterations, n_{max}.

3-viii. If either stopping criterion is met, the algorithm exits.

■

This technique requires model regularization since the inverse problem will be ill posed as long as the number of observations is less than the number of model parameters (Note: it will *always* be ill posed if the number of parameters exceeds the MAR). The main contingency that can destabilize this method is that the experimental design phase is dependent on the current state vector, $\mathbf{m}_n$, which is in turn dependent on the type of inversion being used (the current model vector depends on inversion algorithm details such as regularization scheme, Lagrange multipliers, data and model rescaling, constraints, and so forth). Hypothetically, this method can converge to a local minimum or fail to converge all together if the experimental design subroutine does not prescribe observations that maximize resolution with respect to the true, unknown model, or any model in the neighborhood of the true model.

2.3.8 Applications

In the following chapters, these experimental design techniques are employed in three different geoelectrical scenarios. Not every OED methodology developed here is used in every chapter, because different geophysical scenarios call for different approaches. Furthermore, additional aspects of our OED method are presented in the context of these chapters, instead of being presented here. This was done for the pedagogical reason that these insights would be more intelligible in the context of actual geophysical problems, rather than if they had been divorced from their applications.

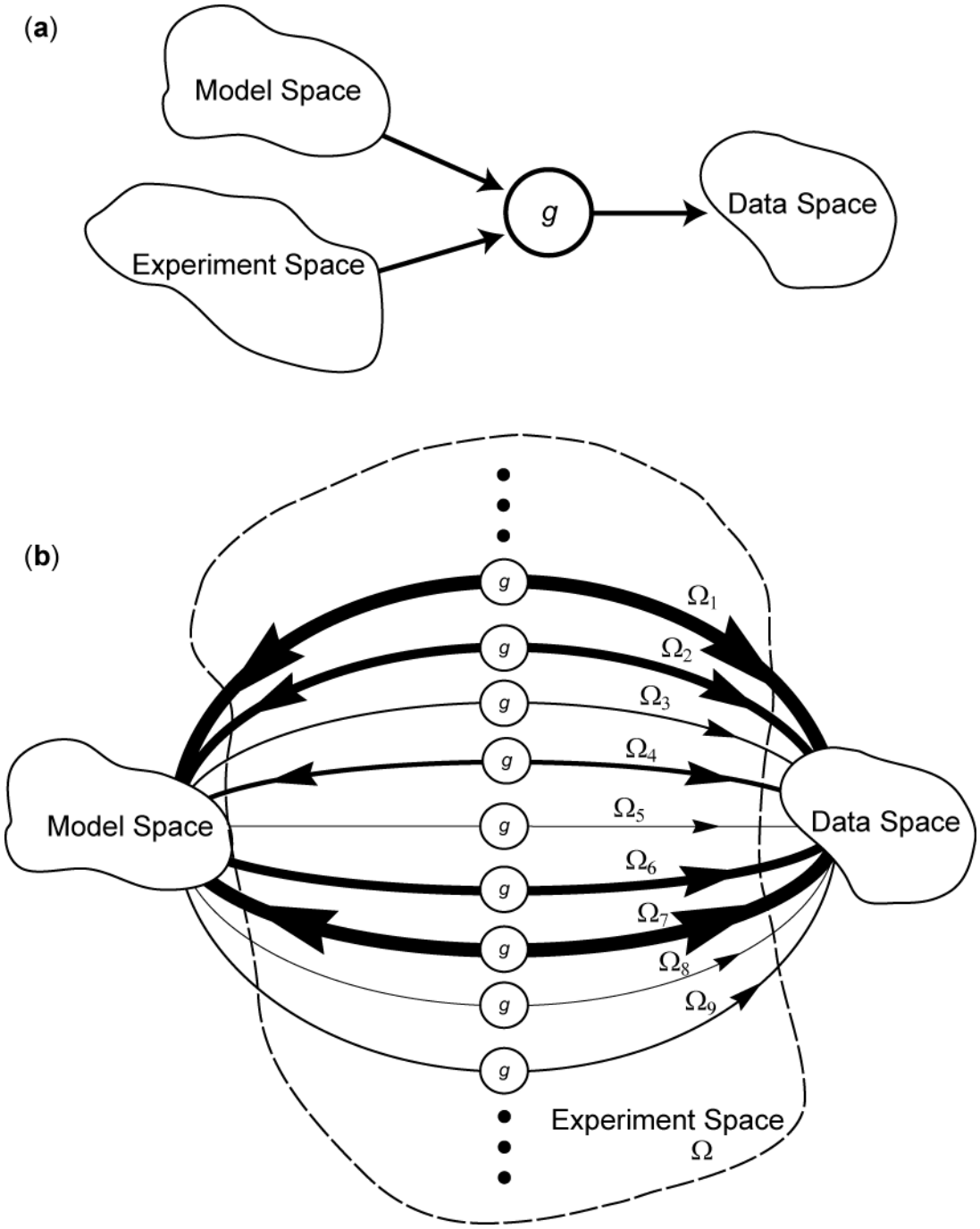


Figure 2-1 Cartoon depictions of the relationships between data-, model-, and experimental-space. (a) Points in model- and experimental-space are mapped by the forward operator, g, to data space. This is an ambiguous way of describing the mapping because it does not distinguish between invertible and noninvertible mappings. (b) An alternate way of representing the mapping from model- to data-space. Each experiment, denoted Ω_n, creates a unique mapping, via g, between model- and data-space. The strength of the mapping is represented by its line weight, and two-way arrows represent invertibility, while one-way arrows represent non-invertibility.

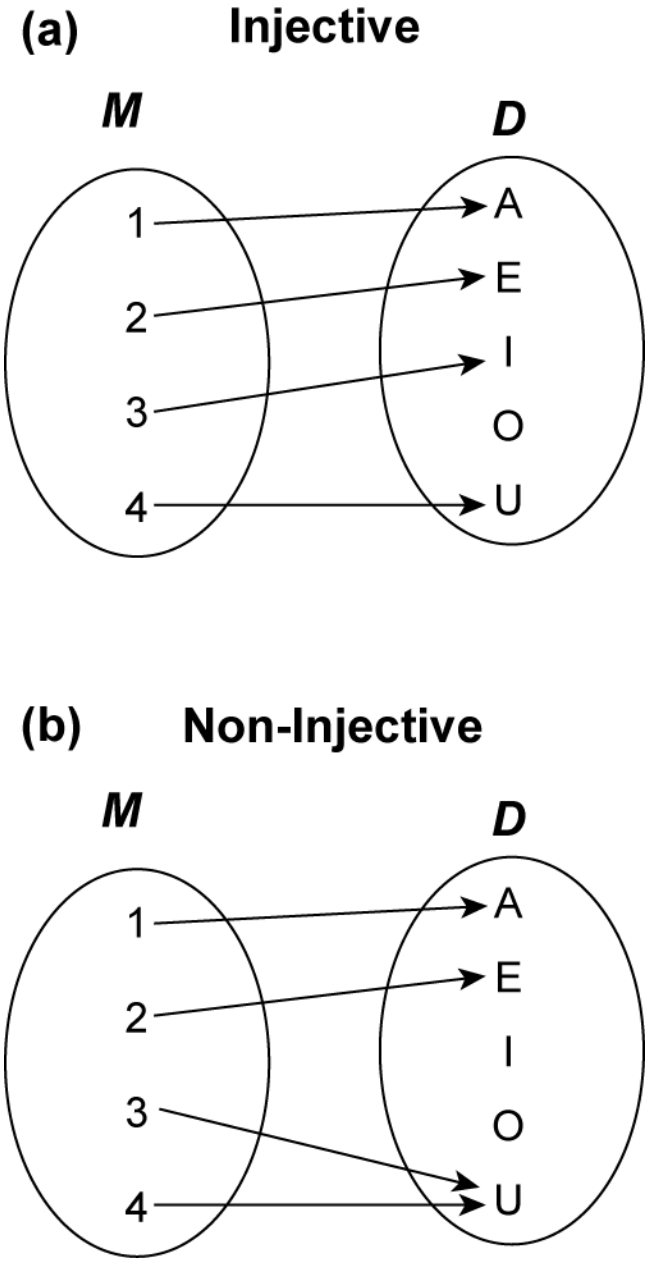


Figure 2-2 Cartoon examples of injective and non-injective mappings. Panel (a) shows an injective mapping between *M* and *D*. Injectivity does not require that every point in *D* has a mate in *M*, but it is a sufficient condition to ensure the invertibility of the mapping. This is reasonable from a geophysical standpoint, as one can easily come up with data sets that cannot possibly be observed in the real world. Panel (b) shows an example of a non-injective (non-invertible) mapping. In this situation, models 3 and 4 both map to data set U, which means that an inverse mapping from U could not distinguish whether the data derived model 3 or 4. Hence, non-injectivity is synonymous with non-invertibility and model nonuniqueness.

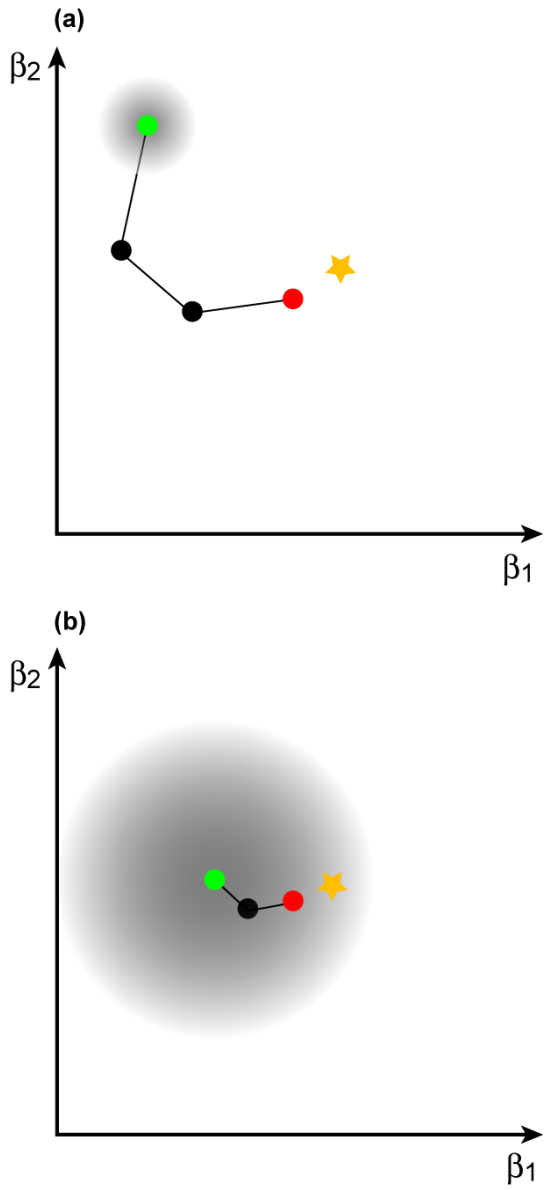
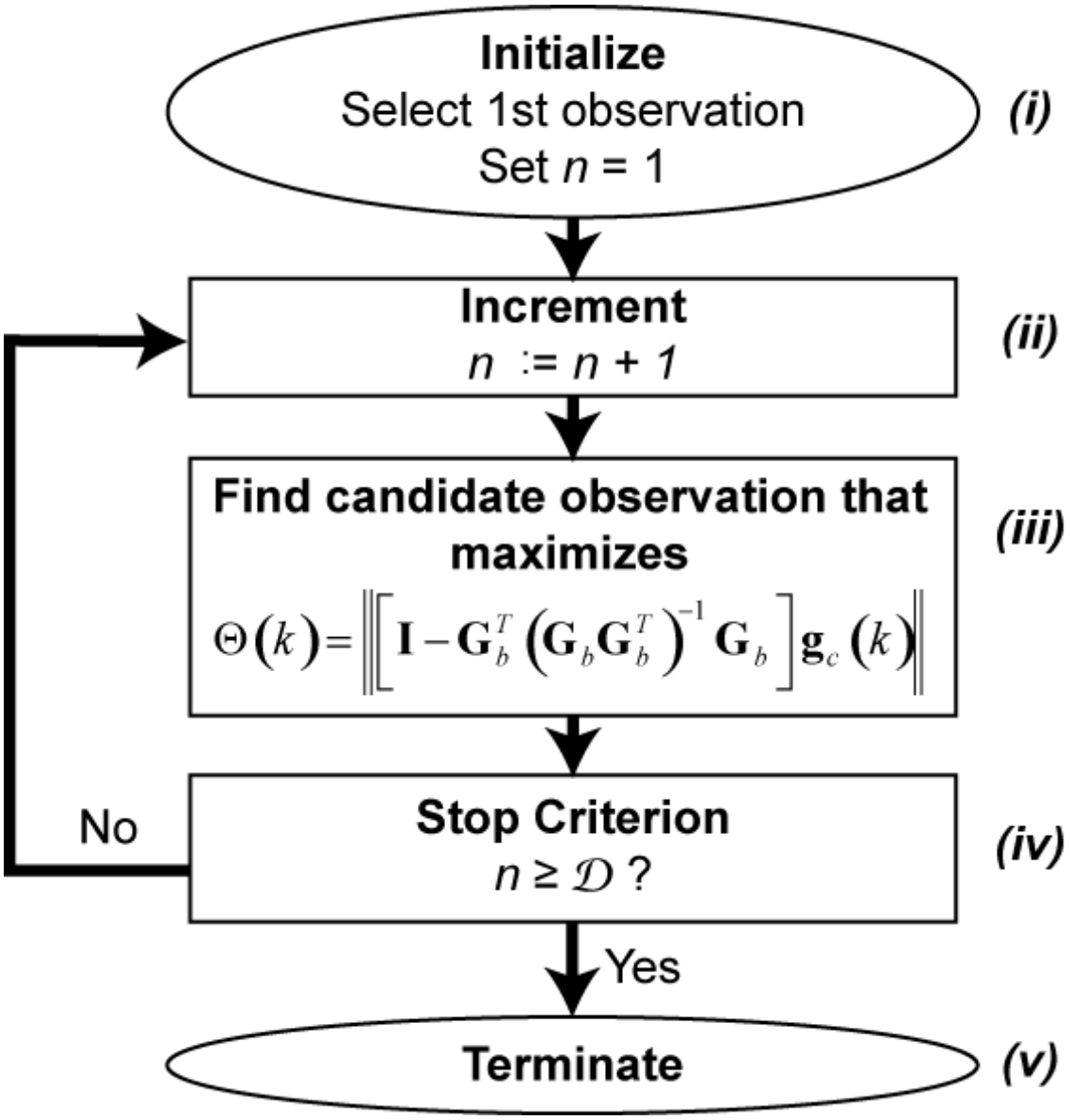


Figure 2-3 Example of 'inversion trajectories', the paths through model space that a nonlinear least squares algorithm follows as it converges on the solution. The green dot is the initial model guess, the red dot is the converged solution and the black dots indicate 'way points' along the trajectory to the converged solution. The yellow star indicates the true model, which typically does not minimize the data rms error when data are noisy. Panel (a) depicts a situation where an experiment has been designed to be optimal with respect to an initial model guess far from the true solution. The grey region depicts a hypothetical neighborhood around the initial model guess for which the designed experiment is 'optimal'. Panel (b) depicts the situation where an experiment has been optimally adapted to an inversion model (green dot) that is assumed to be close to the true solution. In this case, the designed experiment is hypothesized to create a region of experiment optimality with a higher likelihood of containing both the inversion model and the true model.



Flowchart 2-1 Basic algorithm for performing sequential experimental design.

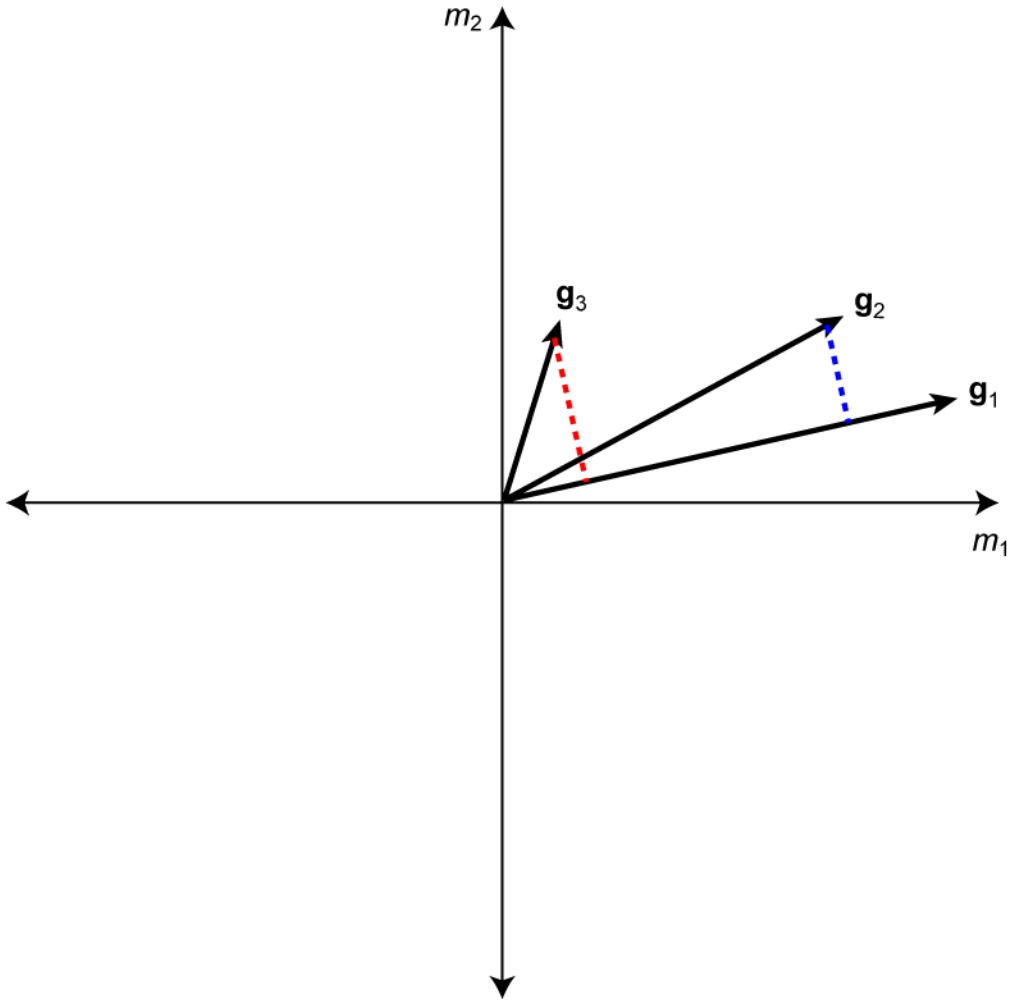


Figure 2-4 Visualization of the basic OED algorithm. If $\mathbf{g}_1$ is the sensitivity kernel of the first observation in the base experiment, the algorithm determines the lengths of the senstivitiy kernels of each candidate observation after projection onto the null space of $\mathbf{g}_1$ (which is orthogonal by definition to the space spanned by $\mathbf{g}_1$). The colored dashed lines represent these projections. The algorithm chooses the candidate whose projection is of maximal length in this orthogonal direction. In this case, the algorithm would choose $\mathbf{g}_3$. Observe that $\mathbf{g}_2$ is longer than $\mathbf{g}_3$, but its projection onto the null space of $\mathbf{g}_1$ is smaller, so the algorithm does not choose it.

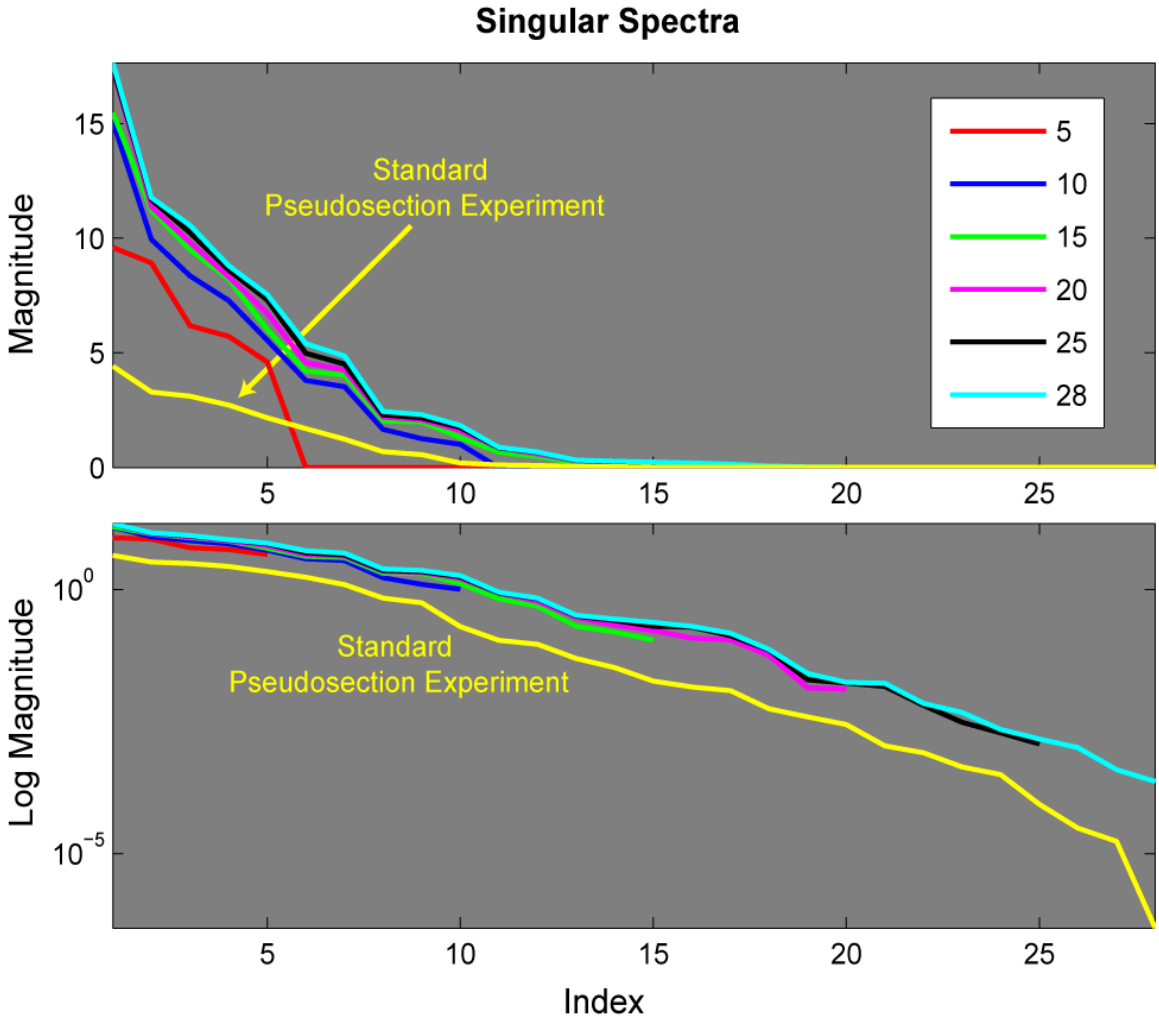
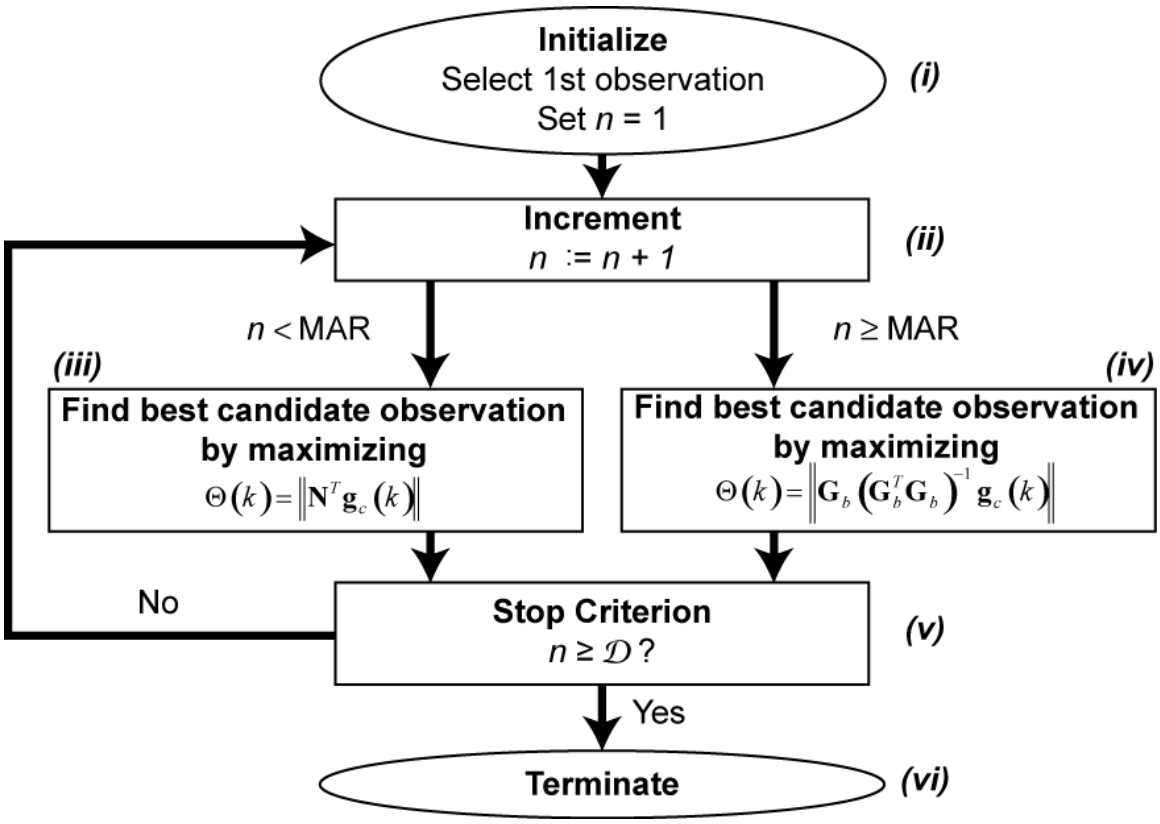
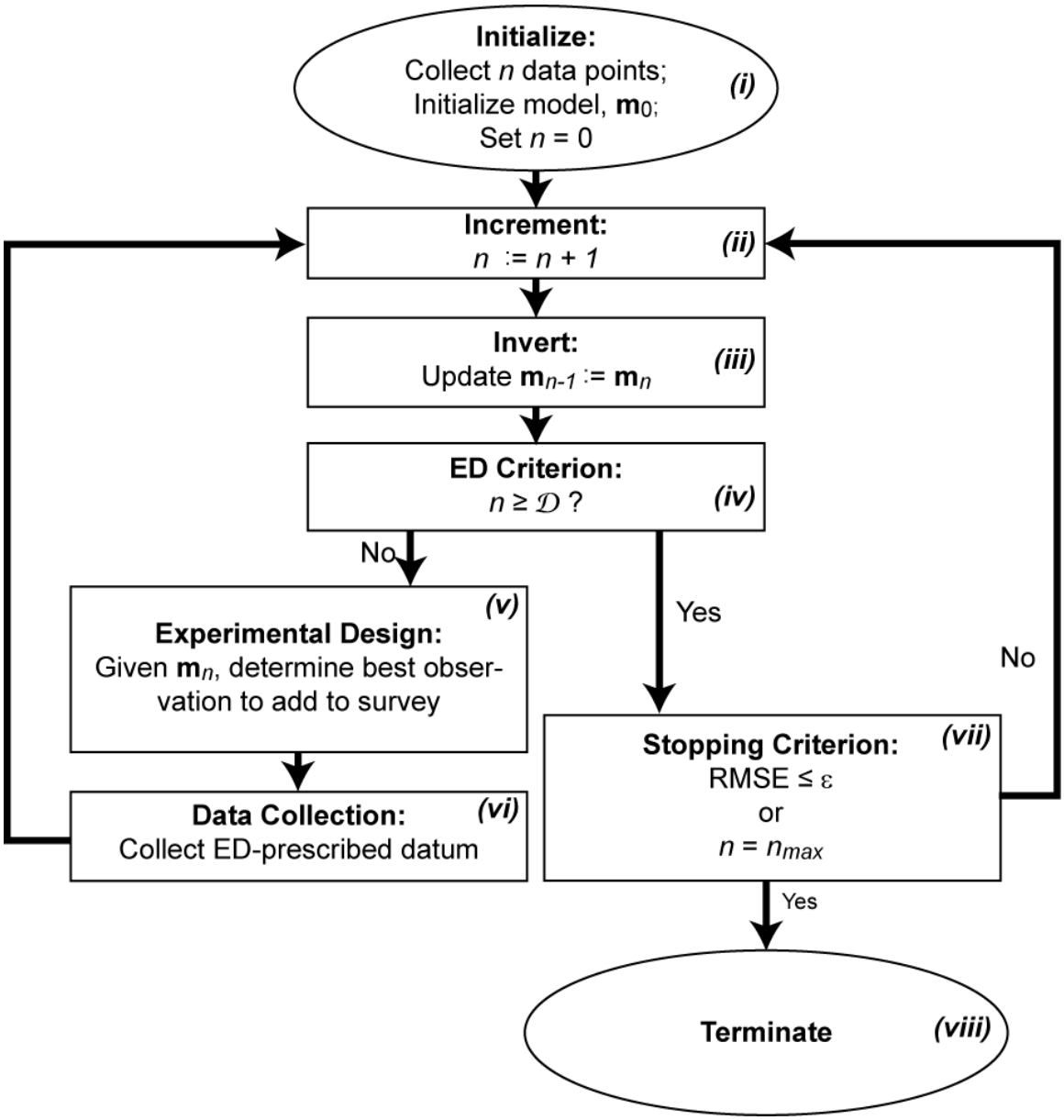


Figure 2-5 Demonstration showing how the experimental design objective function affects the singular spectrum of $\mathbf{G}_b$ as a function of iteration number. Shown are the singular spectra of $\mathbf{G}_b$ at design iterations 5,10, 15, 20, 25, and 28 (the iteration number equals the number of observations currently in the base experiment). This example derives from a borehole resistivity setting (which is fully explored in Chapter 4) where 10 electrodes have been deployed to query a heterogeneous earth model. The details of the borehole problem are irrelevant here; what is important is the behavior of the singular spectra as more observations are added to the base experiment. The top and bottom panels are identical but for the scale along the y-axis. Also shown is the singular spectrum for a standard resistivity survey called the Pseudosection survey (again, the particular details are irrelevant; what is relevant is how the spectra of the designed experiments compare with the spectrum of a widely-used, standard experiment).

Observe that as the number of observations increases all singular values increase. This is because the sensitivities of the added observations are not strictly linearly independent with respect to their predecessors in the base experiment. Also observe that the spectra of the designed experiments are categorically greater than the spectrum for the standard Pseudosection survey.



Flowchart 2-2 Modification of the basic OED algorithm that addresses experiments whose number of observations may exceed the maximum attainable rank of the Jacobian matrix. When the number of observations is less than the maximum attainable rank, this algorithm is identical to the Basic Algorithm in Flowchart 2-1.



Flowchart 2-3 Algorithm describing a possible method for performing adaptive optimal experimental design. This is an integrated form of OED, wherein the operations of data collection, inversion, and experimental design are executed cyclically, until convergence. $\mathcal{D}$ specifies the number of observations the designed experiment will have and n_{max} specifies the maximum number of iterations the super-algorithm is permitted before termination.

Chapter 3

One-Dimensional DC Resistivity

3.1 Introduction

One-dimensional resistivity is among the oldest geoelectrical methods. It assumes that the earth can be approximated as a series of infinite, horizontal layers of differing thicknesses and resistivities. One-dimensional models impose the strongest assumptions on the lithological structure of the earth and therefore are mainly useful in situations where it is reasonable to apply such assumptions, such as might be the case in alluvial settings or when seeking the depth to the water table. Despite the strong restrictions 1D methods impose on the earth model, they are very appealing to the exploration geophysicist because they are easy to deploy and execute in the field and because inversion of their data is usually quite rapid, owing to a relatively small number of parameters as compared with 2D or 3D methods.

This chapter examines optimal experimental design applied to surface one-dimensional resistivity problems. The sequential OED technique developed previously is adapted and applied. We briefly develop the numerics used for 1D resistivity forward modeling and inversion, from which several issues germane to the 1D problem are identified that affect experimental design. Additional aspects of the experimental design problem, which are better addressed without using the OED methodology, are also explored.

A primary objective of this research is to investigate the 1D resistivity problem in the presence of data noise. There are three sources of data error for 1D resistivity: (1) *random errors* due to things like unknown electric fields in the ground and instrument precision; (2) *systematic errors* due to localized heterogeneities in the earth, which cannot be accounted for by using a 1D earth model; and (3) *electrode placement errors*, which impart error by misrepresenting the assumed positions of the electrodes. Random errors in the data are easily addressed, and indeed, several investigations herein are devoted to exploring OED with respect to this kind of noise. Data errors due to local heterogeneities, however, cannot be addressed because 1D modeling simply does not account for any but vertical heterogeneity (Beard & Morgan, 1991). Electrode placement errors are more difficult to appreciate and quantify than random errors, but they can significantly impact data quality. We examine the electrode placement problem in detail and arrive at a surprising and novel discovery.

Several separate research topics are compiled herein, and a slightly unorthodox format has been adopted to present this research. Rather than presenting holistic sections on theory, methodology, results and discussion for the entire chapter, each 'subtopic' is self-contained, with pertinent theory, methodology and so forth being presented in context. These subtopics collectively hang upon the framework of optimal experimental design for 1D resistivity, which is the main topic of this chapter. The subtopics are presented in the following order: (1) Modeling and Inversion; (2) Determining the Number of Layers; (2) Continuous Versus Discrete Experimental Design Search Spaces; (4) Positioning Survey Electrodes; (5) General Surveys for Unknown Earth Models; (6) Adaptive Optimal Experimental Design. Following these sections, a Conclusion is provided that integrates and summarizes the results of the various topic areas.

3.2 Modeling and Inversion

We adopt the exact, analytic solution for multiple horizontal resistivity layers reported by Parasnis (1997). Parasnis' solution is actually a succinct synopsis of the original

solution, which was introduced by Koefoed *et al.* (1979) for modeling apparent resistivity as a function of intrinsic resistivity and layer thicknesses over multiple pancake layers:

$$\rho_a = f\left(\boldsymbol{\rho}, \mathbf{h}, \omega\right), \tag{3.1}$$

where ρ_a is apparent resistivity, expressed here explicitly as a function of layer resistivities and thicknesses in the vectors $\boldsymbol{\rho}$ and $\mathbf{h}$, respectively, and a data station, ω, which is some 4-electrode configuration (a transmitter dipole and a receiver dipole). The details of the method shall not be repeated here (the interested reader is referred to the aforementioned works). Equation (3.1) is used to forward model 1D resistivity and to establish the inverse problem, which employs the Jacobian matrix, populated with the partial derivatives of ρ_a with respect to $\boldsymbol{\rho}$ and $\mathbf{h}$. This approach to 1D resistivity modeling and inversion is called the *variable-thickness method* and has been identified as one of the most robust for accurately resolving model parameters (Simms & Morgan, 1992).

The format for the 1D resistivity problem is shown in Figure 3-1. A sequence of n layers with differing resistivities, ρ_j, and layer thicknesses, h_j, overlies a half-space of resistivity ρ_{n+1}. Four electrodes are positioned collinearly and symmetrically about the center axis of the sounding to collect a datum. It is permissible to employ this symmetry because there is no lateral heterogeneity in the model, which would otherwise require asymmetric electrode configurations. The outer electrodes are positioned a distance L from the center of the sounding and the inner electrodes are each positioned a distance l. Traditionally, current electrodes form the outer pair while potential electrodes form the inner pair, but reciprocity allows for these roles to be interchanged.

The Schlumberger and Wenner electrode configurations (see Figure 3-2) have traditionally been the most widely used for *vertical electrical soundings* (VES), though the Schlumberger spread is noted to be the superior of the two (Zohdy, 1990). The main distinction between the two spreads is that the ratio of the outer to inner electrode spacing for the Schlumberger spread is usually set to be greater than 5

($L/l > 5$), while the Wenner spread fixes this ratio to be 3 ($L/l = 3$). The reason Schlumberger spreads have typically employed *L/l* ratios of 5 or greater is because 1D apparent resistivity modeling has historically employed *approximations* to the full nonlinear forward model function (e.g., Mundry, 1980; Arnason, 1984), and *L/l* > 5 ensured modeling accuracy to within 3%.

However, the nonlinear forward function in this work has not been approximated so we are not restricted to this convention. Denoting a current dipole *AB* and a potential dipole *MN*, a resistivity observation is made by injecting a current, *I*, across *AB* and measuring the resulting potential, *V*, across *MN*. *V* and *I* are converted to apparent resistivity, ρ_a, by

$$\rho_a = k\frac{V}{I}, \tag{3.2}$$

where *k* is the so-called *geometric factor*, which accounts for the relative positions of the four electrodes. The geometric factor is given by

$$k = 2\pi\left(\frac{1}{\overline{AM}} - \frac{1}{\overline{AN}} - \frac{1}{\overline{BM}} + \frac{1}{\overline{BN}}\right)^{-1} \tag{3.3}$$

(Parasnis, 1997) for resistivity measurements made on the surface, where the 'overbars' are length operators. Equation (3.3) can be simplified for Schlumberger and Wenner spreads to

$$k_S = \pi\frac{L^2 - l^2}{2l} \tag{3.4}$$

and

$$k_W = \frac{4\pi}{3}L, \tag{3.5}$$

respectively. The geometric factor should not be too large because it causes the potential measurement to be small, approaching instrument resolution (Stummer *et al.*, 2004). Sharma (1997) recommended a threshold of *k* at around 5500 m, though the choice of geometric factor must ultimately depends on the resolution of the particular field instrument and the magnitude of expected electromagnetic noise at a particular site. A heuristic cartoon is provided in Figure 3-3, plotting the ratio of the predicted

voltage over the injected current as a function of the geometric factor for a homogeneous earth. Geometric factor considerations are at our discretion for synthetic studies. However, to retain a measure of real-world practicality, the geometric factor threshold of 5500 m is adopted herein to comply with assumed limitiations of instrument precision and to minimize EM data noise, which amplified by large k values. However, though it addresses realistic noise problems for resistivity surveys, we acknowledge that this threshold is arbitrary in the context of synthetic trials.

Regardless the particular type of electrode spread used, a series of unique Ll combinations are used to collect the resistivity sounding data set. As a rule of thumb, large L spacings are required to penetrate deeper into the earth. For a homogeneous earth, it has been shown (Telford *et al.*, 1990) that the fraction of the total injected current that flows horizontally below depth z is given by

$$\frac{I_x}{I} = 1 - \frac{2}{\pi}\tan^{-1}\left(\frac{z}{L}\right). \tag{3.6}$$

To ensure that at least half the current flows below a depth, z, the outer electrode spacing, L, must be equal to or greater than z (see Figure 3-4). However, much larger L/z ratios are recommended (as great as $L/z = 5$, which ensures that approximately 85% of the current flows below the depth, z) to maximize sensitivity to the deepest formations. If the maximum depth of investigation is chosen to be 50 m, for example, it is recommended that the largest outer electrode spacing be set 250 m from the center of the sounding, if possible. Owing to the nonlinearity of (3.6), 1D surveys typically distribute the outer electrodes exponentially. This is exemplified by the Schlumberger and Wenner surveys seen in Figure 3-2, which are characteristic of VES surveys.

Data inversion is carried out using nonlinear least squares formalism. The Jacobian matrix of the forward model is calculated with respect to layer resistivities and thicknesses at each iteration of the inversion. Additionally, the problem is log-rescaled, to ensure the positivity of resistivities and layer thicknesses and to improve convergence times. The inversion is stabilized by parameter damping, which is con-

trolled by a Lagrange multiplier whose value is governed by the Levenberg-Marquardt algorithm.

One distinction of 1D resistivity as it is modeled here is that the discretization of the earth is not predetermined, in contrast to the pre-specified meshes that are used in 2D and 3D problems. Because the 1D problem treats layer thicknesses as unknowns, earth discretization is a free variable. The point is that the design procedure must accommodate not only the resistivities of each of the layers, but their thicknesses as well.

3.3 Determining the Number of Layers

In order to execute our sequential experimental design procedure, it is important to determine the correct number of layers the target site should be modeled with. *A priori*, we cannot know how many layers to fit the data with, and there is no guarantee that the number can be deduced by inspecting the raw field data. There is a large literature on methods of model determination, but the purpose here is to demonstrate that we generally *can* determine the number of layers, thereby facilitating the OED exercise, which depends on a working earth model. The particular method by which we determine the number of layers is not so relevant as just showing that we *can* perform the exercise.

Simms and Morgan (1992) proposed that a series of inversions with successively more layers be executed, keeping track of the model and data errors as a function of the number of layers. The following development paraphrases that paper. Both model and data errors should decrease as the number of layers increase, which is predicted along the diagonal line in Figure 3-5 (reproduced from (Simms & Morgan, 1992)). The point at which the model and data rms errors are closest to the origin is taken as the best model. Afterwards, the addition of more layers cannot significantly reduce the data rmse and these models represent the set of nonunique solutions. However, these nonunique solutions typically induce larger model rms errors, as predicted along the horizontal line Figure 3-5. Eventually, the number of layers becomes

so large that the inverse problem cannot resolve all parameters and is ill posed, as predicted along the vertical line in Figure 3-5.

Simms and Morgan's method was examined by repeating one of their examples. A 2-layer model over a halfspace was created. The top layer was 1 m thick with a resistivity of 1 Ωm; the second layer was 5 m thick with a resistivity of 0.2 Ωm; and the substratum had a resistivity of 1 Ωm. 20 data points were synthesized at logarithmic outer electrode spacings of approximately 6 positions per decade, from 1 to 1000 m. Inner electrode spacings were forced to have a geometric factor no greater than 5500 m. The data were contaminated with 5% Gaussian noise. The model and synthetic field data are shown in Figure 3-6. The data were inverted for increasing numbers of layers, and percent rms errors were calculated for both model and data. The results are shown in Figure 3-7. The numbers next to each model/data pair indicate the number of layers for which the data were inverted. This result nicely validates the heuristic that Simms and Morgan put forth. The correct number of layers (over the half-space) is two, which is evident in the cross plot. As predicted, the data error is essentially invariant for models using more than two layers, while the model error increases.

In practice, real-world problems preclude calculation of the *model* error, because the true model is unknown. In many cases, inspection of the data rmse curve is sufficient to determine the correct number of layers. The curve will typically decrease rapidly, asymptoting at the point where the correct number of layers is tried; afterwards, the curve flattens, indicating that additional layers do not significantly improve the solution. Figure 3-8 shows a real field sounding that was collected by the ERL in Curaçao in 2001, during a water prospecting campaign. The data were inverted for one to four layers. The data misfit curve indicates that the appropriate model has two layers over an infinite half-space.

Simms and Morgan (1992) also developed an F-test that discriminates between competing layer models. The F-test is a statistical method for determining whether two sample variances originate from the same distribution. It is useful for model discrimination because it provides an easy way to determine whether the data rms errors

for two models are statistically significantly different. If an F-test between two models indicates that their errors *are* statistically different (at a confidence level, α), the model with the larger misfit is rejected; otherwise, the model with larger number of parameters is rejected. The F metric is defined as

$$F \equiv \frac{s_1^2 / m}{s_2^2 / n}, \tag{3.7}$$

where s_1^2 and s_2^2 are sample variances, and m and n are the number of degrees of freedom of the two samples (typically equal to the number of elements in each sample). It is conventional, though not required, to place the larger of the two sample variances in the numerator, making F > 1. Equation (3.7) is modified for our purposes by substituting the data ms errors of two competing models for s_1^2 and s_2^2, and by observing that $m = n$ because the same number of observations are used for both models:

$$F = \frac{\left\| \mathbf{d}_{obs} - \mathbf{d}_{pred}^1 \right\|^2}{\left\| \mathbf{d}_{obs} - \mathbf{d}_{pred}^2 \right\|^2}. \tag{3.8}$$

This ratio is compared with an F-table at a specified confidence level, α, to ascertain whether the difference between rms errors is statistically significant.

F-tests were conducted on the various layer models in the preceding example. 20 observations were made, so the degrees of freedom were $m = n = 20$. Confidence levels for the F value for samples with 20 degrees of freedom are 2.1242 with 95% confidence, 1.6023 with 85% confidence, and 1.3580 with 75% confidence. Table 3-2 reports the percent rms errors for data and model as well as the F values between successive layer models. The F value between the one- and two-layer models greatly exceeded the 95% confidence level, so the 2-layer model is chosen over the 1-layer model. The F value between the 2- and 3-layer models, however, fails to be significant even at the 75% confidence level, and none of the remaining F values are significant either. Therefore, the F-test indicates that the best model is the 2-layer model, which we know to be correct. It is especially noteworthy that the F-test picks the 2-layer model because the data errors for models with 3 or more layers are actually smaller. For real-world applications, we would not know the model error reported in

the first row of Table 3-2, and while the combination of the two error terms allows us to easily discriminate between models, the F-test has clearly distinguished itself as a robust discriminator when the data are the only information available.

In light of the preceding discussion, it is hereafter assumed that the correct layer model can be identified, freeing us to focus on other details of optimal experimental design in the coming sections.

One closing remark is needed regarding nonuniqueness. There are several causes of nonuniqueness. Underdetermined problems are nonunique because there are fewer observations than parameters. The principle of equivalence, which is dramatically exemplified in gravimetry for example, gives rise to nonuniqueness because different earth models can beget precisely the same data, no matter how many observations are made. Poor model resolution, attendant to an inadequately designed experiment or to the physical nature of a problem, also causes nonuniqueness. 1D resistivity is subject to all these kinds of nonuniqueness, though to a far smaller degree than 2D and 3D methods. In fact, most 1D resistivity problems can usually be well posed. That is to say, there is a unique solution to the inverse problem (provided not too many layers are used) as indicated in Figure 3-5. The nonuniqueness shown in that figure differs from the types listed above; it comes from treating the parameterization of the earth as a free variable.

Ultimately, most 1D resistivity problems can be well posed as long as one uses some sort of model discrimination, like the F-test, and so long as the experiment is adequate for the target site.

A corollary of well-posed inverse problems is this: noiseless data can always be inverted to yield the true solution (provided the correct model functional has been chosen). Whereas optimal experimental design for ill-posed problems must address modeling errors caused both by noisy data *and* resolution limitations, *OED for well-posed problems can focus strictly on minimizing the effects of data noise*. Thus, a primary objective of later investigations is to demonstrate that our sequential OED technique can produce data sets for 1D resistivity that have superior noise rejection capabilities.

3.4 Continuous Versus Discrete Experimental Search Spaces

The formulation of the 1D resistivity forward and inverse modeling, as implemented here, permits electrodes to be placed anywhere, continuously along the surface. This is in contrast to the restrictions imposed by using 2D and 3D transmission line network analogs (e.g., Madden, 1972; Zhang *et al.*, 1995; Shi, 1998), which permit electrodes to be placed only at discrete input/output nodes of an equivalent resistor-network. Consequently, 1D resistivity OED has an infinite number of 4-electrode combinations to choose from, whereas 2D/3D OED has a finite number. Consequently, the update subroutine of our sequential OED algorithm (which picks the next best observation to add to the base experiment) no longer searches a finite set of permitted observations but an infinite one. The update subroutine must not only pick the *combination* of four electrodes to be used but also the *positions* of those electrodes.

The update subroutine of our greedy, sequential OED algorithm chooses the observation whose sensitivity kernel, after projection onto the null space of the base Jacobian matrix, is of maximal length, ensuring that it contributes maximal complementary information. If there are a finite number of permitted observations, it is permissible to exhaustively evaluate each according to this objective. If there are infinitely many, we are forced to use a different optimization approach, since an exhaustive search is impossible. In this research, we utilize an evolutionary algorithm called differential evolution (see Appendix A for details) for determining the optimal observation to be added to the base experiment.

3.5 Survey Electrode Placement

The *xy*-symmetry of one-dimensional earth modeling permits a VES observation to be made using electrodes symmetrically positioned about the center of the survey, typically with the current dipole on the outside and the potential dipole within (as in Figure 3-1). This allows us to cast the experimental design problem in terms of *L-l*

pairs, rather than x_1-x_2-x_3-x_4 4-tuples, where x_i denotes the position of the i^{th} electrode, independent of the others. The main thrust of this section is to examine the effects of electrode placement on data noise and thereby on errors in inverse modeling.

Field technicians cannot place survey electrodes with infinite accuracy, and some error will be introduced into the data by placement errors. These errors are most pronounced when any two electrodes are close together. The percent error in the geometric factor, as a function of these *placement* or *sampling errors*, is examined in detail to ascertain the degree to which this kind of noise contaminates the data.

As described in Equation (3.2) and shown in Figure 3-1, an apparent resistivity field datum derives from measured currents and voltages as well as the geometric factor of 4-electrode data stations. Of these three quantities, the geometric factor is not actually *measured*. It is rather *calculated* using Equation (3.3), under the assumption that the spread electrodes have been placed at their appropriate positions. In other words, calculation of the geometric factor assumes there has been no error in placing the electrodes. In practice, infinitely precise electrode placement is impossible for a variety of reasons. There may be an obstacle where the electrode needs to be placed, the field tape measure may get jostled as the survey is being laid out, an electrode might be driven into the ground at an angle, or the field technician might accidentally set an electrode slightly off its desired position. The purpose of this investigation is to quantify the effects of electrode misplacement to better understand how much or how little this kind of error ultimately affects data inversion.

Multiple Monte Carlo investigations were performed to determine the statistics of the error introduced by electrode misplacement in calculating the apparent resistivity. Various combinations of outer and inner electrode spacing were tried; for each *Ll* combination, 10^5 realizations of electrode misplacement were generated and the apparent resistivities were calculated, creating large samples from which to quantify the expected error. Error was quantified by determining the *coefficient of variation* for each *Ll* combination. The coefficient of variation is a statistic that measures the *relative variability* of a random variable by the ratio of its standard deviation to its mean,

$$c = \frac{\sigma}{k}. \tag{3.9}$$

The coefficient of variation provides a normalized measure of the expected scatter about the mean of a random variable. It is particularly useful in this Monte Carlo exercise because it is directly proportional to the 95% confidence limits for the apparent resistivity when these limits are expressed in terms of percent error. To see this, note that the 95% confidence interval, in terms of percent error, is given by

$$\text{C.I.}(95\%) \equiv \frac{\pm \Delta k(95\%)}{k} = \frac{\pm 1.96\sigma}{k} = \pm 1.96c, \tag{3.10}$$

for an underlying random variable that is normally distributed (which is the case for these Monte Carlo exercises). Equation (3.10) shows that the percent-error confidence limits are directly proportional to the coefficient of variation. Two separate cases of misplacement error were considered: one assuming that 95% of all placement errors fall between ±1 cm of the desired position; and the other assuming that 95% of errors fall between ±2 cm. All placement errors were normally distributed with zero mean.

The results of all Monte Carlo simulations are synopsized in the two panels shown in Figure 3-9. The simulation statistics have been organized into families of curves. Each curve shows the coefficient of variation for fixed L and is plotted against the ratio of the inner to outer electrode spacing, l/L, which is bounded between 0 and 1. Observe that the expected error decreases as the outer electrode spacing, L, becomes large. This is explained by noting that the geometric factor should be less sensitive to placement errors when the relative distances between electrodes are large. The same explanation applies when we trace the curves from left to right, for l/L increasing from 0 to 1. Again, the smallest errors in the geometric factor occur when the electrodes are relatively far from one another, whereas the largest errors occur when any pair of electrodes is close together. Also shown in these panels are the ±10%, ±5%, and ±2.5% error confidence levels calculated according to Equation (3.10). These give an idea the values of L and l one should choose if one wishes to avoid percent data errors (due to electrode misplacement) in excess of some threshold.

One of the most surprising aspects in Figure 3-9 is the position where each of the error curves goes through its minimum. The dots in both panels indicate these positions. Though there is some scatter due to finite sampling, the mean value of these minima is $l/L = 0.43 \pm 0.01$, where 0.01 is the standard deviation. This suggests that there is a unique value for the inner electrode spacing that, as a function of the outer electrode spacing, minimizes data error due to electrode misplacement. This unusual discovery precipitated an investigation to determine whether this ratio could be derived analytically. The following derivation indicates that a value very close to this ratio does indeed minimize the error due to electrode misplacement.

An expression for the geometric factor, equivalent to that given in Equation (3.3), is

$$k = \frac{2\pi}{\frac{1}{x_2 - x_1} - \frac{1}{x_3 - x_1} - \frac{1}{x_4 - x_2} + \frac{1}{x_4 - x_3}}, \tag{3.11}$$

where

$$\begin{aligned} \overline{AM} &= L - l = x_2 - x_1 \\ \overline{AN} &= L + l = x_3 - x_1 \\ \overline{BM} &= L + l = x_4 - x_2 \\ \overline{BN} &= L - l = x_4 - x_3 \end{aligned} \tag{3.12}$$

and where $x_1 < x_2 < x_3 < x_4$, which ensures that the potential dipole is *internal* to the current dipole. This electrode configuration (the potential dipole internal to the current dipole) forces k to be strictly positive. Now consider the gradient of the natural logarithm of k with respect to the electrode positions, x_1, x_2, x_3 and x_4

$$\mathbf{g} \equiv \nabla_{\mathbf{x}} \ln k = \begin{bmatrix} \frac{1}{k}\frac{\partial k}{\partial x_1} \\ \frac{1}{k}\frac{\partial k}{\partial x_2} \\ \frac{1}{k}\frac{\partial k}{\partial x_3} \\ \frac{1}{k}\frac{\partial k}{\partial x_4} \end{bmatrix}, \tag{3.13}$$

where $\mathbf{x} = \left[x_1, x_2, x_3, x_4\right]^T$. The elements of **g** specify the instantaneous percent change in *k* with respect to the four electrode positions, because $\left(\partial_{x_i} k\right)/k \approx \Delta k / \left(k \Delta x_i\right)$. A useful perspective on **g** is to note that its elements are the normalized sensitivities of the geometric factor to perturbations in electrode positioning. The 'perturbations', in this case, are random fluctuations in each electrode's position, owing to placement error. Therefore, to minimize the expected percent error in *k*, the elements of **g** must be minimized. However, the signs of the partials in (3.13) are ambiguous, but this is overcome by squaring the elements of **g**. Thus, the general minimization problem is to find $\hat{\mathbf{x}}$ satisfying

$$\hat{\mathbf{x}} = \arg\min_{\mathbf{x}} \left\| \mathbf{g}\left(\mathbf{x}\right) \right\|^2 . \tag{3.14}$$

To execute the above derivation, we must first determine **g** with respect to **x**, but it is convenient to skip a step here: rather than express **g** with respect to **x**, we evaluate **g** and then make the following substitutions for **x**: $x_1 = -L, x_2 = -\alpha L, x_3 = \alpha L, x_4 = L$. These substitutions force the electrodes to be symmetrically situated about the center of the survey and give rise to the following expression for **g**:

$$\mathbf{g} = \begin{bmatrix} \dfrac{1}{L\left(\alpha^2 - 1\right)} \\ \dfrac{\alpha^2 + 1}{2L\alpha\left(1 - \alpha^2\right)} \\ \dfrac{\alpha^2 + 1}{2L\alpha\left(\alpha^2 - 1\right)} \\ \dfrac{1}{L\left(1 - \alpha^2\right)} \end{bmatrix}. \tag{3.15}$$

The preceding substitutions also effect a modification of Equation (3.14). Rather than finding $\hat{\mathbf{x}}$ that satisfies Equation (3.14), we now seek $\hat{\alpha}$ that does the same, using Equation (3.15). Equation (3.14) now takes the form

$$\hat{\alpha} = \arg\min_{\alpha \in [0,1]} \left(\frac{1 + 6\alpha^2 + \alpha^4}{2L^2\alpha^2\left(\alpha^2 - 1\right)^2} \right), \tag{3.16}$$

which is satisfied over the interval $0 \le \alpha \le 1$ by

$$\alpha \cong 0.429195 . \tag{3.17}$$

This value of α was determined using a symbolic equation solver in Mathematica on Equation (3.16) (*Note: For the interested reader, a thorough derivation of these mathematics is provided in Appendix B*). This validates the observation from Figure 3-9 that the *l/L* ratio is indeed constant (regardless of the outer electrode spacing) with a value of ~0.43, which is very close to the estimated minima seen in Figure 3-9.

The preceding concept was tested by Monte Carlo simulation. A Schlumberger survey, customarily used by the ERL, was compared with a survey designed using Equation (3.17). The two surveys are tabulated in Table 3-1, columns 1 and 3. The outer electrode spacing of the designed survey was set to be identical to that for the Schlumberger survey, with the inner electrode spacings set to be 0.43*L*. Although not reported in Table 3-1, the geometric factors of these surveys are both below the required threshold of 5500 m mentioned earlier. Five hundred realizations of electrode misplacement were generated for each survey, from which noisy field data were synthesized and inverted. The placement errors were normally distributed with zero mean and a standard deviation of ~0.005 m, causing 95% of the errors to be within ±1 cm. The underlying earth model, shown in

Figure 3-10, was a 2-layer model with a 10 m thick surface layer of 200 Ωm, a 10 m thick intermediate layer of 50 Ωm, and a 400-Ωm basement.

The relative performance of the two surveys was compared by numerically approximating the mean absolute percent error (MAPE) between the true and predicted sounding curves for each realization. The MAPE metric is given by

$$\text{MAPE} = \frac{1}{z_{\max}} \int_0^{z_{\max}} \left| \frac{\rho(z) - \tilde{\rho}(z)}{\rho(z)} \right| dz , \tag{3.18}$$

where $\rho(z)$ is the true model profile and $\tilde{\rho}(z)$ is the predicted model profile, and where $z_{\max}$ is set to the maximum desired depth of investigation. Note that the outputs for our 1D inversion are resistivities, ρ_j, and layer thicknesses, h_j. These are converted into a discrete approximation of the continuous profile, $\tilde{\rho}(z)$, in order to nu-

merically approximate Equation (3.18). Cumulative distributions of the MAPE are shown for both surveys in Figure 3-11. Also shown are the 90% confidence levels for both surveys. The Designed and Schlumberger surveys produce MAPEs less than ~20% and ~34%, respectively, with 90% confidence. More important than this is the fact that the cumulative frequency distribution of modeling errors for the Designed survey is almost entirely to the left of the distribution for the Schlumberger survey. This indicates that the frequency distribution of model errors for the Designed survey is shifted to the left (smaller model errors) relative to the frequency distribution of model errors for the Schlumberger survey. Hence, the Designed survey produces a statistically significant reduction in model error.

The preceding demonstration validates the hypothesis that, to minimize the effects of electrode misplacement on inversion, one can place the inner electrodes at 43% of the outer electrode spacing. It also confirms the analytic derivation in Equations (3.11) - (3.17).

Notably, the Wenner spread (Figure 3-2) employs a constant ratio of $l/L = 1/3$, which is closer to the observed (and analytically derived) $l/L = 0.43$ shown here than the ratio for Schlumberger spreads, which is typically $l/L \leq 1/5$. So, while the Wenner spread has been in disfavor historically (Zohdy, 1990), given modern inversion practices, it may be that the Wenner spread is in some ways superior to the Schlumberger spread.

In closing this section, it is important to recapitulate. The demonstrations and analyses here have shown that the inner electrode spacing should simply be set to 43% of whatever is the outer electrode spacing. *This is true if the only source of data noise is expected to derive from electrode misplacement*. As will be seen in subsequent sections, if the primary source of data noise is random fluctuations due to unknown EM sources, the '43% Rule' does not apply. In point of fact, if we wished to perform 1D resistivity OED that simultaneously addressed *both* noise sources – electrode misplacement and random EM fields – the design algorithm would have to be altered such that both noise types were accounted for. This scenario is not addressed in this research, but it is conjectured that this would be a significantly harder design

problem. This derives from the fact that electrode misplacement errors, which ultimately affect calculation of the geometric factor, k, necessarily *multiply* random errors in a potential observation. This is evident from inspection of Equation (3.2), for if we replace k with $\tilde{k} = k + k_e$, where k_e is the error in calculating the geometric factor due to electrode misplacement, and if we replace V with $\tilde{V} = V + V_e$, where V_e is error in the potential measurement due to random fluctuations of an EM field, the apparent resistivity calculation clearly depends on the *product* of k_e and V_e. This ultimately means that the diagonal elements of the data-error covariance matrix differ and are observation-dependent. It is probable that our sequential OED technique could be adapted to scenarios where the magnitude of data noise varies from observation to observation, but this prospect remains to be addressed.

3.6 General Surveys for Unknown Earth Models

In Chapter 1, we briefly described two general types of optimal experimental design, *unspecific* and *specific* OED. Unspecific OED was the term used to describe experimental design problems when no model information is available on a target site. In such cases, it is reasonable to start from the assumption that the earth is homogeneous. The purpose of unspecific OED therefore is to design an experiment optimized for a homogeneous earth, on the assumption that such an experiment can serve as a general survey in unknown areas. We are of course free to perform additional optimized surveys afterward, to improve imaging accuracy, and it should be clear that these surveys derive from *specific* OED, because a working model is now *specified*.

We would like to design a pseudo-optimal 1D resistivity experiment for situations where no information is available on the true earth model, and we begin by assuming a homogeneous earth. Recall that our sequential OED technique essentially operates on the sensitivity kernels of permitted observations. The following discussion demonstrates that a number of practical problems arise pertaining to these sensitivity kernels when a homogeneous earth model is assumed.

First, there is only one model parameter for a 1D homogeneous earth; there is only one resistivity – the resistivity of the halfspace – and there are no layer thicknesses. This means that *the observation sensitivity kernels are scalars, not vectors*. The key feature of our sequential OED method is that it attempts to simultaneously maximize information complementarity and magnitude, but information complementarity relates to the orthogonality of sensitivity kernel, and orthogonality is undefined for one-dimensional vectors. Consequently, the sequential OED method is inapplicable, because information complementarity is undefined. In effect, as there is only one parameter, only one observation would be needed to determine it. But the objective is to create an experiment that is robust not only for homogeneity but for heterogeneity as well. As stated at the beginning of this section, we would like to design an experiment that can be used as a primary survey when no knowledge of the earth structure is available, and a survey with only one observation would be absurd.

A potential workaround would be to introduce the artifice of a multilayer earth with identical layer resistivities. Still, this poses practical problems. The partial derivatives of apparent resistivity with respect to layer resistivities are dependent on the layer thicknesses, even though the resistivities are identical. In other words, two different layer models will give rise to different sensitivity kernels, even though the earth is actually identical for both models. Additionally, the partial derivatives of apparent resistivity with respect to layer thicknesses will all be zero because perturbing the thickness of a layer whose resistivity is identical to adjacent layers will not affect the predicted data. Consequently, the artificial partitioning of a homogeneous earth results in arbitrary sensitivity kernels. Because different partitionings create different sensitivity kernels, different experimental designs will arise for arbitrary partitionings. Clearly, it is impractical to set up the OED problem by this artifice.

The foregoing problems with observation sensitivity kernels imply that our OED technique must be abandoned for 1D homogeneity. Instead, an analytic approach is put forth. We propose that a pseudo-optimal experiment for a homogeneous earth will be one whose current fractions, as described in Equation (3.6) (see Figure 3-4), are linearly distributed over some user-defined range, as shown in Figure 3-12. Intui-

tively, this ensures that each depth interval from the surface to the maximum desired investigation depth, z_{max}, is apportioned an equal current fraction. Mathematically, we are specifying that the ratio, I_x/I, should be a linear function of j, where j is the j^{th} observation from a set of n observations:

$$\frac{I_x}{I}(j) = a\,j + b \quad \text{for} \quad j = 1, 2, ..., n\,, \tag{3.19}$$

where a and b are linear coefficients that are yet to be determined. Now, if we define L_{min} as the minimum permitted outer electrode spacing (this will be user specified), L_{max} as the maximum permitted electrode spacing (user specified), and z_{max} as the maximum desired depth of investigation (user specified), a system of two equations arise,

$$\begin{aligned} \frac{I_x}{I}(j=1) &= a + b = 1 - \frac{2}{\pi}\tan^{-1}\left(\frac{z_{\max}}{L_{\min}}\right) \\ \frac{I_x}{I}(j=n) &= a\,n + b = 1 - \frac{2}{\pi}\tan^{-1}\left(\frac{z_{\max}}{L_{\max}}\right) \end{aligned}, \tag{3.20}$$

that are solved for a and b, producing

$$\begin{aligned} a &= \frac{1}{n-1}\left[\tan^{-1}\frac{z_{\max}}{L_{\max}} - \tan^{-1}\frac{z_{\max}}{L_{\min}}\right] \\ b &= \frac{1}{n-1}\left[\tan^{-1}\frac{z_{\max}}{L_{\min}} - \tan^{-1}\frac{z_{\max}}{L_{\max}}\right] \end{aligned}. \tag{3.21}$$

Combining (3.21), (3.19) and (3.6) and solving for L as a function of the j^{th} observation, we end up with the following prescription for outer electrode spacings:

$$L(j) = \frac{z_{\max}}{\tan\nu(j)}, \quad \text{for } j = 1, 2, ..., n\,, \tag{3.22}$$

where

$$\begin{aligned} \nu(j) = &\frac{j}{n-1}\left[\tan^{-1}\frac{z_{\max}}{L_{\max}} - \tan^{-1}\frac{z_{\max}}{L_{\min}}\right] \\ &+ \frac{1}{n-1}\left[n\,\tan^{-1}\frac{z_{\max}}{L_{\min}} - \tan^{-1}\frac{z_{\max}}{L_{\max}}\right] \end{aligned}. \tag{3.23}$$

Below, we examine the foregoing prescription for outer electrode spacing to determine whether it can indeed produce superior data. Three surveys were compared in a Monte Carlo simulation over the synthetic target in
Figure 3-10. The three surveys were the Schlumberger, Wenner and 'Designed' surveys shown in Table 3-1, columns 1, 2 and 4. The inner electrode spacings for the Designed survey were permitted to be greater than or equal to 0.15 m or no less than would cause the geometric factor to exceed 5500 m. The former restriction was employed in deference to Figure 3-9 where it was shown that extremely small l/L ratios give rise to large errors in calculating apparent resistivity field data. The latter was employed in deference to the limits of instrument resolution discussed previously. Five hundred realizations of noisy data were synthesized for each survey and inverted. The noise for each realization was 5% Gaussian.

Figure 3-13 synopsizes the results of this Monte Carlo investigation. Panels *a*-*e* show the cumulative frequencies of percent error between the true and predicted model parameters, ρ_1, ρ_2, ρ_3, h_1 and h_2. These panels also show the 95% confidence intervals for each parameter with respect to the three experiments. Differences between the confidence intervals for the five parameters are small enough between the three surveys to argue that their performances are essentially identical, though the 'Designed' survey distinguishes itself with respect to the ρ_2 and h_2 parameters where it produces expected modeling errors 2 to 3% smaller than the Schlumberger and Wenner surveys. The three surveys also appear to perform similarly when we examine the cumulative frequency of total modeling error shown in Panel *f* (this panel shows the cumulative distribution of MAPEs, as defined in (3.18)). All three surveys produce basically the same 95% confidence level for mean absolute percent error. From these results, it is concluded that any of the three surveys would be adequate as a primary survey over an unknown target. Surprisingly, the Wenner survey performs as well as the Schlumberger survey, which apparently contradicts the rule of thumb that the Schlumberger is the better of the two (Zohdy, 1990). This can possibly be explained by noting that the above rule of thumb originated prior to the widespread use of 1D resistivity inversion (note, the original edition of Zohdy's manuscript was writ-

ten in 1974). The primary argument favoring Schlumberger surveys was that they display greater peak-to-peak variation in apparent resistivity *data* than does Wenner over the same interval. However, from an inversion standpoint, a data set with greater maximum data variability than another does not necessarily imply that it will produce a superior inversion model. A second explanation could simply be that the Wenner and Schlumberger surveys happen to perform equally well on this particular earth model. A third explanation could be that the gradient of the error landscape (the data rms error as a function of the model parameters) is relatively small in the neighborhood of the true solution, no matter which of the three surveys is used. Because the inversion algorithm implements a version of nonlinear least squares, ubiquitously shallow gradients in proximity to the true solution would cause all three surveys to have similar modeling errors, forcing the conclusion that the three work equally well.

Whatever the case, it is apparent that the 'Designed' survey is generally no more robust than the Schlumberger or Wenner surveys when no model information is available. Therefore, it is admissible to use any of the three as a primary survey over an unknown target. As mentioned at the beginning of this section, we are free to optimize additional experiments once the primary survey has produced a working earth image. This would be an example of what we term Two-Stage (or Multi-Stage) Adaptive OED, which is examined below.

3.7 Sequential Versus Global Design Algorithms

In Chapter 2 it was explained that the novel design methodology can be implemented by a greedy algorithm, which sequentially adds observations one at a time that are (locally) optimal with respect to the observations currently in the base experiment. This algorithm differs from traditional OED optimization algorithms that operate on entire experiments at once, such as the genetic and simulated annealing algorithms, which were previously identified as being widely used in OED exercises. This difference in optimization strategy has two significant aspects: (1) the sequential technique executes in a fraction of the time needed for global search strategies and (2) because the

sequential method chooses updates that are *locally* optimal (that is, they are optimal with respect the observations preceding them in an experiment but not necessarily with antecedent ones), it does not guarantee that the resulting experiments are *globally* optimal. Hence, the novel sequential design method sacrifices global experiment optimality for increased computation speed.

The question is: how suboptimal are sequentially designed experiments as compared with globally designed ones, and how much timesavings can they be expected to produce? In this section we examine a simple 1D resistivity OED problem and compare the results that arise from sequential and global optimization strategies. Because we are comparing two optimization algorithms, this section addresses their abilities to maximize the objective function. In other words, the comparison is based on the efficiency and robustness of each algorithm strictly with respect to the OED objective function, not with respect to the subsequent designed experiments' abilities to produce high quality data. Whether the OED objective function is itself appropriate for experimental design exercise is addressed elsewhere.

At each iteration of the sequential design method, Equation (2.65) shows that design objective is equivalent to maximizing the ratio of the determinants of the Jacobians of the augmented and base experiments. Because the base experiment is fixed at any given iteration, maximizing this ratio is equivalent to just maximizing the determinant of the augmented Jacobian. Therefore, the appropriate objective function to consider for the global search strategy is just the determinant of the Jacobian of the experiment.

The two-layer earth model in Figure 3-10 was used as the basis for desiging experiments of 5, 10, 15 and 20 observations via the novel sequential method and via a global search strategy executed by a genetic algorithm, operating on the determinant of the Jacobian matrix. The results are tabulated in Table 3-3. Observe that the sequential design method used a decreasing fraction of the global CPU time (the time it took for the global search algorithm to converge) as the number of observations increases. Furthermore, observe that the relative opimality of sequentiallly designed

experiments (measured as a fraction of the global optimum produced by the genetic algorithm) increased as the number of observations increased.

A crossplot of the fractional optimality of sequential designs versus the fractional CPU time is shown in Figure 3-14 to help visualize the results in Table 3-3. This figure plainly shows that the sequential design method closes the gap between itself and global design methods with regard to the optimality of designed experiments for increasing numbers of observations, reaching nearly 92% of global optimality at 20 observations. Coupled with this is the fact that the sequential method significantly reduced relative CPU time, requiring only ~1% as much CPU time as the global search strategy for 20 observations. Though this comparison looks at only one design scenario, it establishes the tantalizing hypothesis that the sequential design strategy not only asymptotes to the global optimum for large numbers of observations but that it also does so at an increasingly efficient rate as compared with global search strategies.

Figure 3-15 provides a sense for the relative optimality of sequentially and globally designed surveys as compared with randomly generated surveys. Ten thousand surveys of 20 observations were randomly generated and the determinants of their resulting Jacobians were tallied. The histogram shows the probability distribution of the optimality ($\log_{10} \det \mathbf{G}^T\mathbf{G}$) of randomly generated surveys as well as the optimality of sequentially and globally designed ones of 20 observations. Even with log scaling along the *x*-axis, it is quite clear that both designed surveys are beyond the right tail of the distribution, demonstrating that they do indeed produce statistically significant solutions to the design objective function.

3.8 Adaptive Optimal Experimental Design

The objective of the following research is to determine whether our sequential OED algorithm could significantly improve 1D resistivity data quality. Because the number of inversion parameters for 1D resistivity is small (as compared with 2D and 3D methods, for example), the inverse problem can normally be well posed, as long as

there are no resolution issues brought about by extremely thin layers or by trying to invert for layers deeper than the practical resolution limits of the experiment (governed by the maximum outer electrode spacing). Of course, parameter resolution is always an issue for surface geoelectrical methods because it drops off rapidly as a function of depth. Thus, even though a 1D inverse problem may technically be well posed, it may still suffer from the effects of poor conditioning of the Jacobian matrix.

One of the most pronounced effects of a poorly conditioned inverse problem is large modeling error due to noise in the data. This is evident by considering the *model covariance matrix*:

$$\mathbf{C}_m = \sigma_d^2 \left(\mathbf{G}^T \mathbf{G} \right)^{-1}, \tag{3.24}$$

where $\mathbf{G}$ is the Jacobian and σ_d^2 is the expected variance of the data noise (This formulation of the model covariance matrix assumes that data errors are uncorrelated and have the same variance. More general expressions for $\mathbf{C}_m$ exist). When $\mathbf{G}$ is poorly conditioned, one or more of its singular values are near zero. Upon inverting, these small singular values blow up, causing the diagonals of $\left(\mathbf{G}^T \mathbf{G} \right)^{-1}$ and $\mathbf{C}_m$ to be large. The diagonal of $\mathbf{C}_m$ contains the expected variance of each model parameter (as a function of noise in the data of variance σ_d^2) about its true value; if these variances are large, there is a large uncertainty associated with parameter estimates, and large uncertainties mean that significant biases can arise in the predicted model. Consequently, the prime objective of OED for 1D problems is to minimize model errors induced by data error as it propagates through the inverse.

In the following investigation, we examine an adaptively optimized 1D experiment. We approach the imaging problem in two stages. In the first stage, a Schlumberger survey was used to synthesize a noisy data set over the target site shown in Figure 3-10. These data were inverted to produce the working model shown in Figure 3-16. Using this model as input to the sequential OED algorithm developed in Chapter 2, we then designed an experiment optimized for the working model (see Table 3-1, column 5). In the second stage, a new noisy data set was synthesized based on

the optimum experiment and a second inversion was performed. This is an example of Two-Stage Adaptive OED. In all cases, data noise was 5% Gaussian.

The performance of the designed experiment was evaluated by comparing it with Schlumberger and Wenner soundings used as second-stage surveys. A Monte Carlo simulation was executed with 500 realizations of noisy data that were inverted by the three second-stage surveys. The results are shown in Figure 3-17.

Panels *a*-*e* show the 95% confidence limits for the five model parameters (ρ_1, ρ_2, ρ_3, h_1 and h_2). Except for ρ_1, the designed experiment produces smaller ranges of expected error for the parameters than the Schlumberger and Wenner surveys. However, ρ_1 has the smallest range of expected error of the five parameters, never exceeding ± 1%. Despite the fact the designed experiment does not constrain this parameter as well as the other two surveys, the magnitude of the error is of negligible importance. Significantly, the optimized experiment reduces the range of error for ρ_2 and h_2 (Panels *b* and *e*) by approximately 15%, compared with the Schlumberger and Wenner surveys. These two parameters are clearly the least well resolved of the five, so a reduction in their uncertainties will have a large effect on the total modeling error. It is notable that the error distributions for ρ_2 and h_2 are nearly identical. This indicates that these two parameters are highly correlated, which will generally destabilize the inversion and slow convergence. Despite the evident improvement of the optimized survey over the other two surveys, even the designed survey apparently fails to break the correlation between ρ_2 and h_2. This illustrates a fundamental fact about surface resistivity inversion: resolution limitation is an incontrovertible physical reality that cannot be completely eradicated, even by careful survey design (Madden & Mackie, 1989; Madden, 1990). The most OED can do is minimize resolution limitation by reducing parameter correlation.

Panel *f* in Figure 3-17 compiles the modeling errors into a single metric. Shown are the cumulative distributions of mean absolute percent error (Equation (3.18)) for the three surveys. Also shown are two 'benchmark' lines: the 90% confidence limit (horizontal, black, dashed line) and the MAPE from stage-one model (vertical, black, dashed line). The other dashed lines project the intersections of the error curves with

the benchmarks onto the appropriate axes. The 90% confidence limit for the designed survey is ~49% modeling error, while it is ~83% for the Schlumberger and Wenner surveys. This is a significant reduction in the expected total model error and is taken as evidence that the designed survey produces superior data quality. Additionally, ~86% of modeling errors for the designed survey are less than the error for the stage-one model, indicating that the optimized experiment produces a statistically significant reduction in modeling errors arising from data noise (this is seen by tracing the horizontal, red, dashed line to the frequency axis). In contrast, only ~62% and ~58% of model errors produced by the Schlumberger and Wenner surveys, respectively, were less than the one-stage model error.

Withal, Figure 3-17 offers compelling evidence that our sequential OED method does indeed produce statistically superior data quality in the presence of data noise, as compared with standard 1D resistivity surveys. An alternate way of presenting these Monte Carlo exercises is to show the average model profiles with error bars for each survey. Figure 3-18 shows the average model profiles and error bars (two standard deviations), as well as the true model and Stage-One model for comparison. All three experiments produce average model profiles that visually match better with the true model than does the Stage-One model. Additionally, all three experiments produce roughly equal (ands small) model errors for the resistivities in the two overburden layers as well as in the halfspace, though careful scrutiny shows that the adaptively Designed survey does indeed produce smaller modeling errors with respect to the thickness of the surface layer and the resistivity of the intermediate layer. The adaptively Designed survey, however, very evidently reduced the model uncertainty associated with the thickness of the intermediate layer, as compared with the Schlumberger and Wenner soundings. Lastly, the Designed survey also produced an average model that more closely matches the true earth model.

It is instructive to examine the reasons why this OED technique can produce data sets with superior noise rejection characteristics. In Chapter 1, we described a greedy OED algorithm that designs experiments by sequentially adding observations to a base experiment. Additional observations are chosen according to their relative ‘fit-

ness' with respect to the base experiment. Fitness is qualitatively defined with respect to two objectives: maximum information content and maximum information complementarity. The observation from the set of permitted observations that maximizes these two objectives is chosen as the next to be added to the base experiment. The fitness of a candidate observation is quantized by taking the vector norm of the projection of its sensitivity kernel onto the null space of the Jacobian of the base experiment. By projecting a sensitivity kernel onto the null space, only those components of the kernel that are orthogonal to the Jacobian of the base experiment are preserved (it is a residual of sorts). The orthogonality condition thus addresses information complementarity, while the vector norm of the projection handles information magnitude.

In short, we augment the Jacobian of the base experiment by appending a row vector (sensitivity kernel) whose length is maximal along a direction orthogonal to the space spanned by the Jacobian. From this point of view, it should be clear that the added row is treated like an eigenvector and its magnitude is treated like an eigenvalue. An arbitrary eigenvector is orthogonal to the space spanned by all other eigenvectors; and the projection of a sensitivity kernel onto the null space of the Jacobian is orthogonal to the Jacobian, so it is orthogonal to the *eigenvectors* of the Jacobian. This shows that the projection is like an eigenvector; and if the projection is like an eigenvector, it follows immediately that its magnitude is like an eigenvalue. Based on this discussion, *maximizing the fitness metric is an attempt to maximize each of the singular values of the Jacobian matrix of the final experiment*.

An alternate expression of the model covariance matrix in (3.24) is

$$\mathbf{C}_m = \sigma_d^2 \, \mathbf{V}\boldsymbol{\Sigma}^{-2}\mathbf{V}^T , \tag{3.25}$$

where $\mathbf{V}$ contains the right singular vectors of $\mathbf{G}$ (which incidentally span model space) and where $\boldsymbol{\Sigma}^{-2}$ is a diagonal matrix containing the inverse squared singular values of $\mathbf{G}$ (e.g., Backus & Gilbert, 1970; Menke, 1989; Tarantola, 2005). To minimize modeling errors induced by data noise, we should minimize the diagonal elements of $\mathbf{C}_m$ as much as possible. In light of the foregoing discussion, it is now clear that the sequential OED method does this: it attempts to maximize the singular values

of **G**, which in turn minimizes the elements of $\boldsymbol{\Sigma}^{-2}$, causing the diagonal of $\mathbf{C}_m$ to be as small as possible. Note that **V** is an orthonormal basis, and all orthonormal bases are rotation matrices, so even though $\mathbf{C}_m$ is not directly proportional to $\boldsymbol{\Sigma}^{-2}$ it is true that

$$\text{trace}\left(\mathbf{C}_m\right) = \sigma_d^2 \,\text{trace}(\boldsymbol{\Sigma}^{-2}), \tag{3.26}$$

because rotation matrices are length preserving. This proves that the sequential OED method does in fact seek to maximize data quality by maximizing a data set's noise rejection characteristics.

To illustrate the preceding development, Table 3-4 lists the diagonal elements of the model covariance matrices for the Schlumberger, Wenner and Designed experiments investigated in this study. As the diagonal of $\mathbf{C}_m$ contains expected parameter variances, the square root has been taken to express parameter uncertainties in terms of standard deviations. Additionally, each standard deviation has been normalized by the parameter value, so the tabulated entries express the expected degree of relative parameter uncertainty (this is called the coefficient of deviation, which is the ratio of the standard deviation to the mean). Parameter uncertainties were all evaluated with respect to the stage-one model (Figure 3-16), which was the initial guess for the stage-two inversion, and with respect to the true model. Looking at the normalized parameter uncertainties for the stage-one model, the designed survey clearly has the lowest uncertainties, except for ρ_1 and h_1. However, the uncertainties for these two parameters are so low in general that it is inconsequential that the designed survey produces larger uncertainties. The same pattern is seen for the true model. Importantly, the designed survey shows the greatest improvement for parameters with the most uncertainty. This provides another piece of evidence supporting the conclusion that the sequential OED method produces data sets with superior noise rejection capability. In addition, it is apparent that the design method redistributes parameter uncertainties, sacrificing some of the certainty of the well-resolved parameters to improve the certainty of less well-resolved ones.

The last issue to consider with regard to our greedy OED method is computation time. For the technique to be useful, it must be reasonably applicable in the field.

Figure 3-19 shows the CPU time for experimental designs from 1 to 36 observations. All CPU times were clocked on a dual 2GHz laptop with 2GB RAM. Two points are notable: first, the CPU time is basically $O(n)$, essentially linear with respect to the number of observations; second, the total CPU time is on the order of minutes or even seconds, depending how large a survey is desired. These two points signify that our greedy OED algorithm imposes minimal computational demands, and there is no reason the method cannot be practically implemented in real-world settings. However, one cannot escape the fact that the survey electrodes must be repositioned for an adaptive OED exercise and this will slow down field operations significantly.

3.9 Integration and Summary

This chapter has explored various aspects of optimal experimental design applied to 1D surface resistivity. A large portion of the document has been dedicated to compiling and expanding our knowledge of the 1D resistivity problem in order to lay the groundwork for the sequential OED method introduced in Chapter 1. The following remarks summarize and integrate our research to provide a macroscopic view of 1D experimental design as it has been investigated and developed herein.

3.9.1 Modeling and Inversion

The one-dimensional resistivity forward and inverse problems have been approached using a variable-thickness modeling scheme first introduced by Simms and Morgan (1992). This method is particularly robust because it uses a small number of parameters (as compared with fixed layer-thickness approaches that large numbers of layers and resistivities). Small numbers of parameters are desirable for OED because the optimization algorithm executes quickly. Moreover, the inverse problem can usually be well posed when there are only a few model parameters, and this equates with improved model resolution as well as improved data noise rejection characteristics.

3.9.2 Determining the Number of Layers

A critical issue for 1D resistivity is to determine the correct number of layers a target site should be modeled with. One-dimensional resistivity is unusual in this sense, because not only do we need to identify earth parameters that minimize the data rmse, we must also identify the correct modeling function. A two-layer model has a different modeling function than a three-layer model, for example. Consequently, we must discriminate between competing layer models. If the layer-model is incorrect, experimental design runs the risk of optimizing for a scenario significantly different than the true model. The ramification of which would be that the designed experiment might bias against the true model, precluding correct identification.

We addressed this problem by employing Simms and Morgan's F-test methodology (Simms & Morgan, 1992), and we conclusively demonstrated that the correct layer-model could be identified. This is of practical significance for adaptive OED, where an experiment is tailored to a working model that has been created by inverting an initial data set. Subsequent investigations assumed that the correct layer model could always be found, allowing these investigations to avoid the distraction of model discrimination.

3.9.3 Continuous Versus Discrete Experimental Search Spaces

Because of the way the 1D resistivity problem has been formulated, electrodes are permitted to be placed anywhere on the continuum of the survey line. Thus, there are infinite combinations of 4-electrode configurations available for experimentation. This means that the experimental design search space is infinite, not discrete. OED search spaces for 2D and 3D resistivity, in contrast, are discrete (for the transmission network analogue) because the electrodes can only be placed at pre-determined positions. To search an infinite space of permitted observations, a modification was made to the sequential OED algorithm. The algorithm was initially designed to exhaustively evaluate each candidate observation in a discrete set. Instead, a differential evolution algorithm (see Appendix A) was substituted for the exhaustive search. The

DE algorithm quickly searches the experiment space for the best candidate observation to be added to the base experiment.

The main issue with reverting to a search algorithm on an infinite search space, as opposed to performing an exhaustive perusal of a discrete search space, is additional computation expense. However, because the number of observations for 1D resistivity is normally small, the computation time for experimental design was shown to be linear with respect to the number of observations desired. It was shown that a reasonably sized experiment could be designed in a matter of minutes or seconds, which is certainly practical for real-world implementation.

3.9.4 Survey Electrode Placement

This investigation was slightly off topic with regards to our sequential OED method. However, it was included because its results were both novel and unexpected. We examined the data errors created by accidental electrode misplacement. It was shown that electrode misplacement creates a nontrivial error in calculating the apparent resistivity of an observation. Because the inversion operates on the apparent resistivities created by each observation in an experiment, electrode misplacement errors ultimately compromise the quality of the inversion model. A careful analysis of these errors showed that the optimal placement of survey electrodes is such that the inner electrode spacing should be approximately 43% of the outer electrode spacing, regardless how large the expected placement errors are.

It was also seen that data errors decrease as a function of increasing outer electrode spacing. Additionally, data errors are largest whenever any two electrodes in a 4-electrode observation are close together.

It is worth pointing out that our sequential OED technique would never predict this outcome. The reason is most easily appreciated by considering the model covariance matrix expressed in Equation (3.24) or (3.25). An analysis of the noise rejection characteristics of our sequential design algorithm showed that an optimally designed experiment minimizes the diagonal of the model covariance matrix by minimizing the

diagonal of $\left(\mathbf{G}^T\mathbf{G}\right)^{-1}$. That is, the design algorithm operates exclusively on the Jacobian matrix, **G**. This approach treats the noise in the data (which is represented by σ_d^2 in Equations (3.24) and (3.25)) as *independent* of the observation. By contrast, the electrode misplacement approach treats the noise in the data as *dependent* on the observation. Hence, whereas the objective of our sequential design algorithm is to minimize the diagonal of $\left(\mathbf{G}^T\mathbf{G}\right)^{-1}$, the objective for electrode placement error is to minimize σ_d^2. An interesting follow-on study would be to attempt experimental design with respect to both considerations: electrode misplacement (observation-dependent) and random data noise (observation-independent).

3.9.5 General Surveys for Unknown Earth Models

A general survey, pseudo-optimal for any reasonable earth model, would be of significant practical use, either as a standalone survey or as the primary survey in a two-stage adaptive OED framework. A general survey would be used in cases where no useful knowledge of the earth structure is available. We argue, that absent any knowledge of the earth structure, it is reasonable to start from a homogeneous assumption. However, it was shown that our OED algorithm could not usefully be applied to a homogeneous earth because of difficulties relating to the partials of the forward operator with respect to the model parameters.

In lieu of our sequential design approach, it was hypothesized that a survey that evenly distributed the horizontal fraction of the total current over a specified depth interval would produce superior data quality. A comparative examination of the Schlumberger, Wenner and Designed surveys over a heterogeneous target however indicated that the three surveys performed nearly identically in the presence of data noise. It was therefore concluded that any of the three surveys would be adequate as a general survey.

A surprising result was the fact that the Wenner survey performed as well as the Schlumberger. Common wisdom maintains that the Schlumberger survey produces

superior data. However, this wisdom was never based on data *inversion*, but rather on the inspection *field data*. It could be that this rule of thumb needs revision in light of modern inversion practices. Inspection of parameter uncertainties shown in Table 3-4 suggests that the Schlumberger survey does indeed perform better than the Wenner survey; but the percent difference between the two never exceeds ~20%, which is not insignificant but neither is it particularly *significant*.

3.9.6 Sequential Versus Global Design Algorithms

It was shown that the sequential design methodology does indeed produce suboptimal experiments compared to global search methods. However, the sequential method approaches global optimality as the number of observations increases, reaching more than 90% optimality in the study conducted here. Additionally, the fractional CPU time required for the sequential method decreases significantly as the number of observations increase, reaching about only 1% in this study.

The preceding two facts indicate that the novel sequential design method produces extremely high-quality, albeit suboptimal, data quality in a fraction of the time needed to design experiments by stochastic search methods. As always, there is a trade-off that end users must decide for themselves. Is it worth the large additional computational expense needed to execute a stochastic search to eke out the last 10 to 20% of optimality that an experiment potentially could have or would it be acceptable to live with 80 to 90% optimality achieved in 1 to 5% of the computation time? Clearly, this an issue of diminishing returns and one in which Pareto's Principle might be invoked. A large additional input would be required to realize only a small/modest increase in output. Ultimately, the fact that the sequential method closes the gap in terms of optimality between itself and global search methods as experiment sizes increase argues strongly for its application in real-world problems. It is hypothesized that the sequential method may indeed produce global optima as the number of observations continues to increase, but this must be tempered by the fact that OED is subject to a general law of diminishing returns (as shown elsewhere in this thesis), which asserts that arbitrarily large random surveys can perform as well as deliberately designed ones.

An interesting possibility would be to create a hybrid optimization technique that couples the two strategies. In the first stage, a sequential design could be performed to produce an experiment close to, but not quite, optimal. This experiment would then be used to seed a global search algorithm, allowing the algorithm to start in the neighborhood of the optimal experiment and possibly greatly reducing convergence times.

3.9.7 Adaptive Optimal Experimental Design

Having previously examined various relevant issues for 1D resistivity OED, the last investigation in this chapter finally turned its attention to the application of our sequential OED method on 1D problems. A two-stage approach was adopted, where an initial noisy data set was inverted and then an optimal survey was designed based on the working model from the stage-one model. Noisy data were generated for the optimal survey and a second inversion was conducted. An examination of modeling errors showed that the designed survey significantly outperformed both the Schlumberger and Wenner surveys, which were also used as second stage surveys for comparison. This exemplified the fact that our design methodology does indeed produce data sets with superior noise rejection characteristics.

An analysis looking at how our design technique endows experiments with superior noise cancellation properties showed that it produces model covariance matrices with smaller elements along the diagonal. As these elements correspond to the expected parameter uncertainties in the presence of data noise, smaller values imply that the inversion models produced by designed experiments are more accurate or, equivalently, less uncertain.

It was also observed that the designed experiment forfeited some of the resolution for highly resolved parameters in favor of accentuating the resolution of less well resolved parameters. Philosophically, this is good situation. It would of course be ideal if we could simply decrease the uncertainty (which equates with resolution) associated with poorly resolved parameters, but the nature of the 1D resistivity forward problem apparently precludes this. The next best thing would be to trade some of the

resolution of the well-resolved parameters for improved resolution of the least well-resolved parameters, which is the behavior that has been observed.

Overall, the two-stage AOED technique has shown itself to be a useful means of reducing model uncertainties in the presence of data noise.

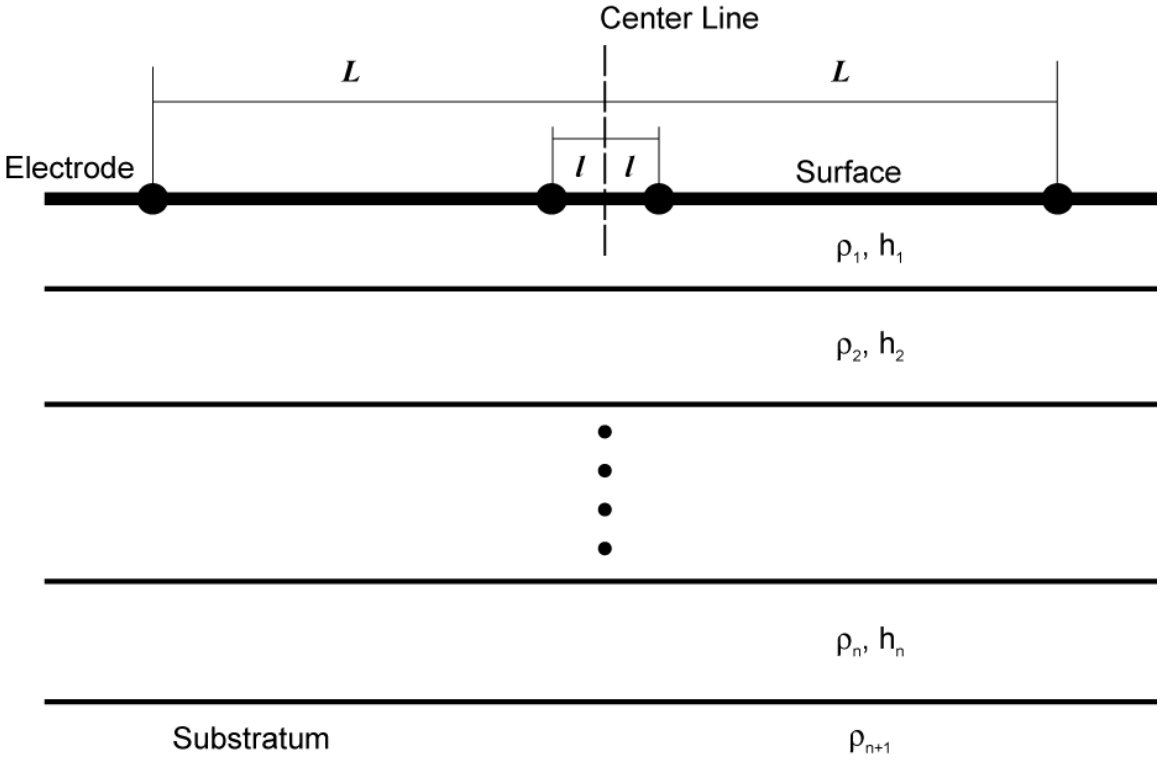


Figure 3-1 General 1D resistivity model. The earth is modeled as a series of infinite parallel layers, each having a unique resistivity, ρ_j, and layer thickness h_j. The bottom 'layer', labeled 'Substratum', is an infinite half-space. Also shown is a typical quadrupole configuration for querying the site. It is convenient to situate the sounding electrodes symmetrically about the centerline of the sounding, as shown. This can be done because there is no lateral heterogeneity in the model, which would otherwise necessitate asymmetric electrode configurations. L specifies the outer electrode spacing and l specifies the inner spacing. Typically, the outer electrodes are current-bearing and the inner ones are potential-measuring, but reciprocity permits these roles to be reversed. Several different Ll combinations are used to collect a data set.

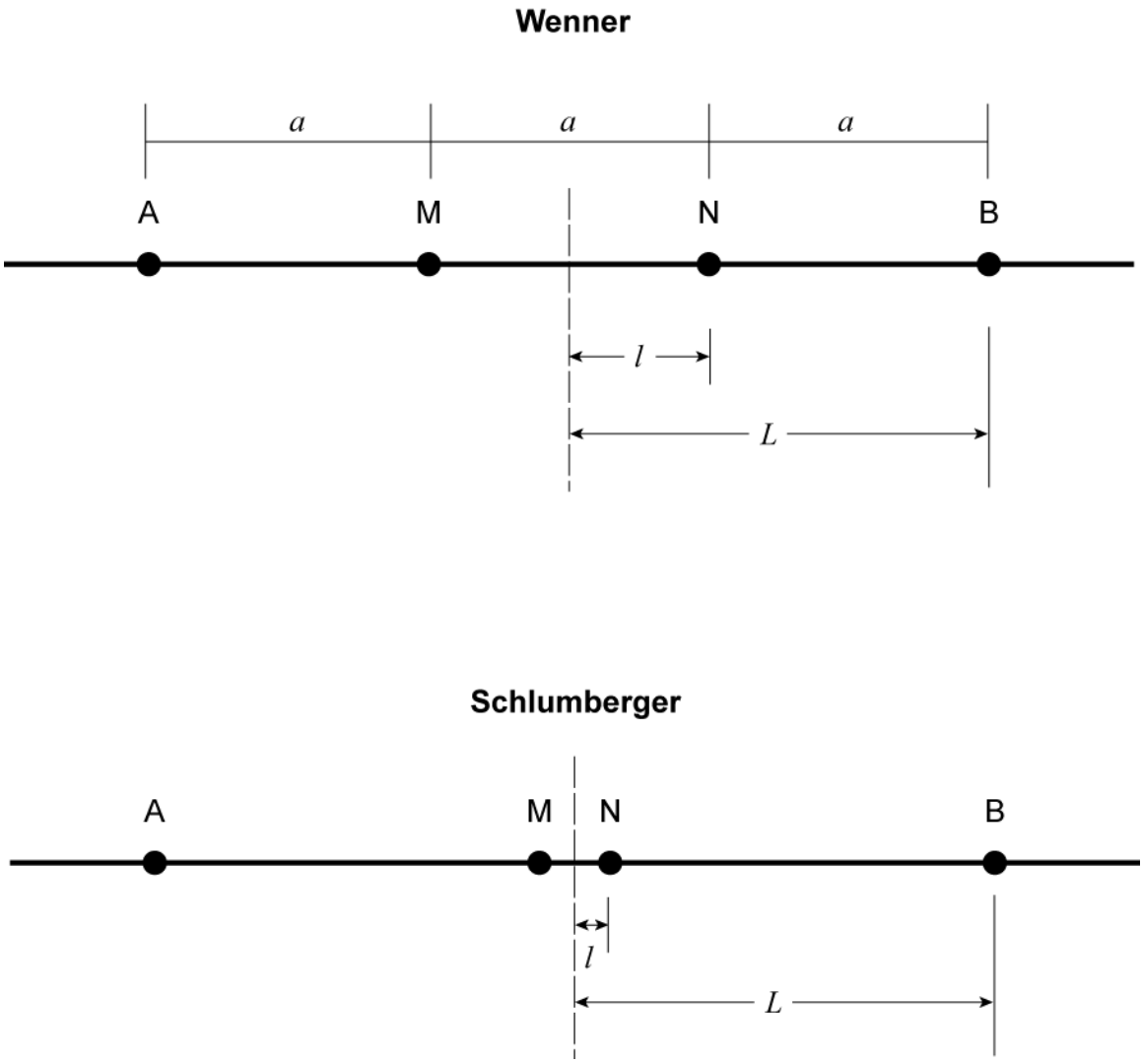


Figure 3-2 Schlumberger and Wenner electrode arrangements. The Wenner arrangement forces adjacent electrode spacings to be equal, causing the outer electrode spacing, L, to be 3 times the inner spacing, l. The Schlumberger arrangement customarily forces the outer electrode spacing, L, to be five times greater than the inner spacing, l.

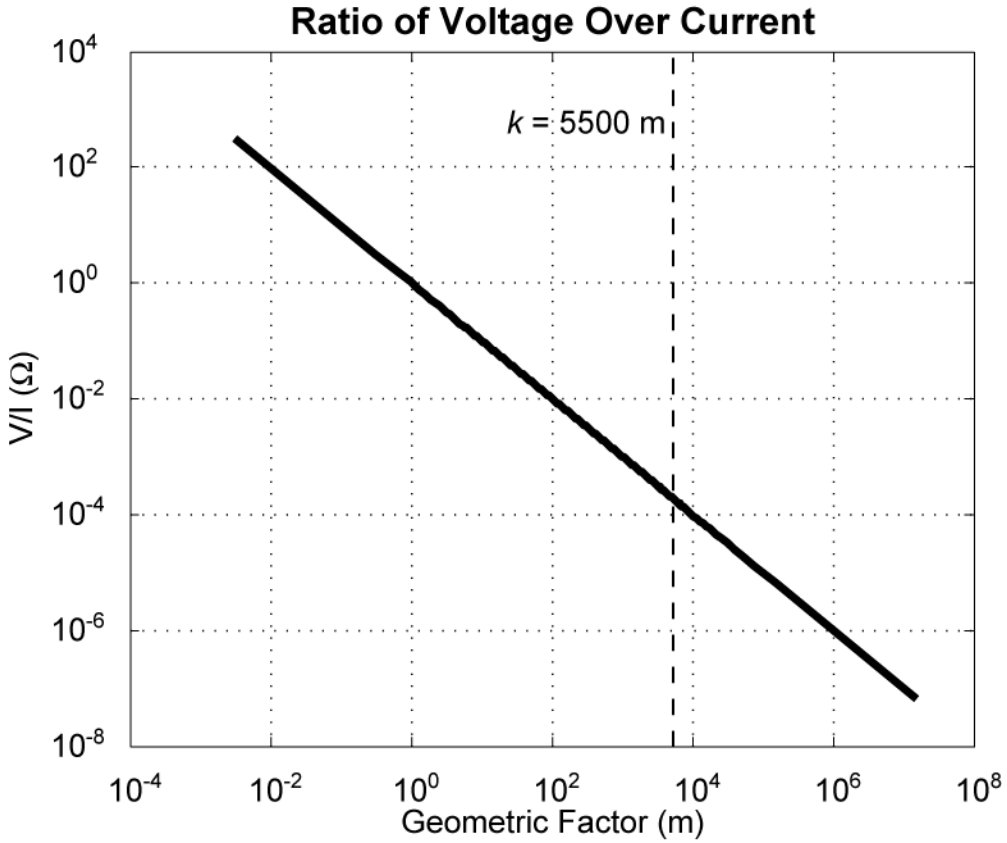


Figure 3-3 An example of the ratio of predicted voltage over injected current (which is the *apparent resistance* of an observation) as a function of the geometric factor, k, for a homogeneous earth. Also shown is a recommended geometric factor threshold of 5500 m (Sharma, 1997). Note that, at k = 5500 m, the ratio of V over I is nearly 10^{-4}. This means the magnitude of the voltage observation would be approximately only one ten thousandth of the magnitude of the injected current. A voltage only one ten thousandth the magnitude of the injected current approaches the instrument precision of most modern resistivity meters. Moreover, any random EM noise will also be added to the potential observation and then be amplified by the geometric factor, according to Equation (3.2).

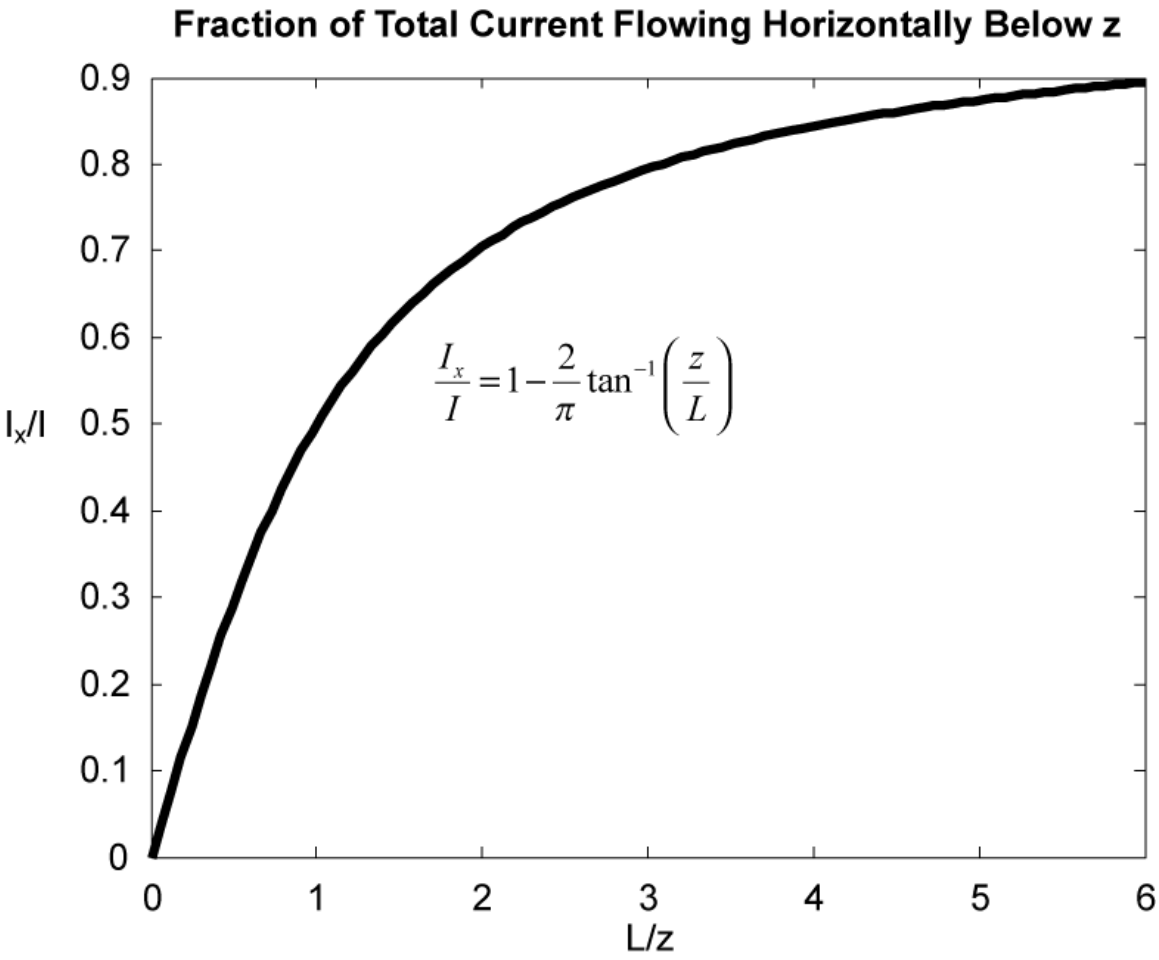


Figure 3-4 This graphic shows the fraction of the total current that flows horizontally below depth *z* for a homogeneous earth (after (Telford *et al.*, 1990)). The outer electrode spacing has been normalized by the depth scale.

Schlumberger		Wenner		Design 1		Design 2		Design 3	
L	I	L	I	L	I	L	I	L	I
0.90	0.15	0.90	0.30	0.90	0.39	1.00	0.15	1.00	0.15
1.25	0.15	1.25	0.40	1.25	0.54	5.00	0.15	1.00	0.15
1.70	0.15	1.70	0.55	1.70	0.73	9.10	0.15	1.00	0.15
2.30	0.15	2.30	0.75	2.30	0.99	13.20	0.15	13.30	0.15
3.15	0.15	3.15	1.05	3.15	1.35	17.60	0.15	13.40	0.15
4.00	0.15	4.00	1.35	4.00	1.72	22.20	0.15	13.80	0.15
6.00	0.15	6.00	2.00	6.00	2.58	27.20	0.20	13.90	0.15
8.00	1.25	8.00	2.65	8.00	3.43	32.60	0.30	20.30	0.15
11.00	1.25	11.00	3.65	11.00	4.72	38.60	0.45	31.30	0.30
15.00	1.25	15.00	5.00	15.00	6.44	45.30	0.60	31.40	0.30
20.00	1.25	20.00	6.65	20.00	8.58	53.20	0.80	31.90	0.30
26.00	4.00	26.00	8.65	26.00	11.16	62.50	1.10	32.10	0.30
37.00	4.00	37.00	12.35	37.00	15.88	73.80	1.55	34.70	0.35
50.00	4.00	50.00	16.65	50.00	21.46	88.20	2.20	89.80	2.30
75.00	4.00	75.00	25.00	75.00	32.19	107.30	3.30	92.10	2.40
100.00	4.00	100.00	33.35	100.00	42.92	134.20	5.15	94.90	2.55
175.00	9.00	175.00	58.35	175.00	75.11	175.90	8.80	107.90	3.30
250.00	18.00	250.00	83.35	250.00	107.30	250.00	17.75	250.00	17.75

Table 3-1 List of all 1D surveys examined in this chapter. Outer electrode spacing is given by L and inner electrode spacing by l. 'Design 1' fixed $l = 0.43L$ and was designed for superior resistance to electrode misplacement errors. 'Design 2' was created as a general survey for unknown earth models. 'Design 3' was an adaptively optimized experiment based on an initial earth model created by inverting Schlumberger data over a heterogeneous target.

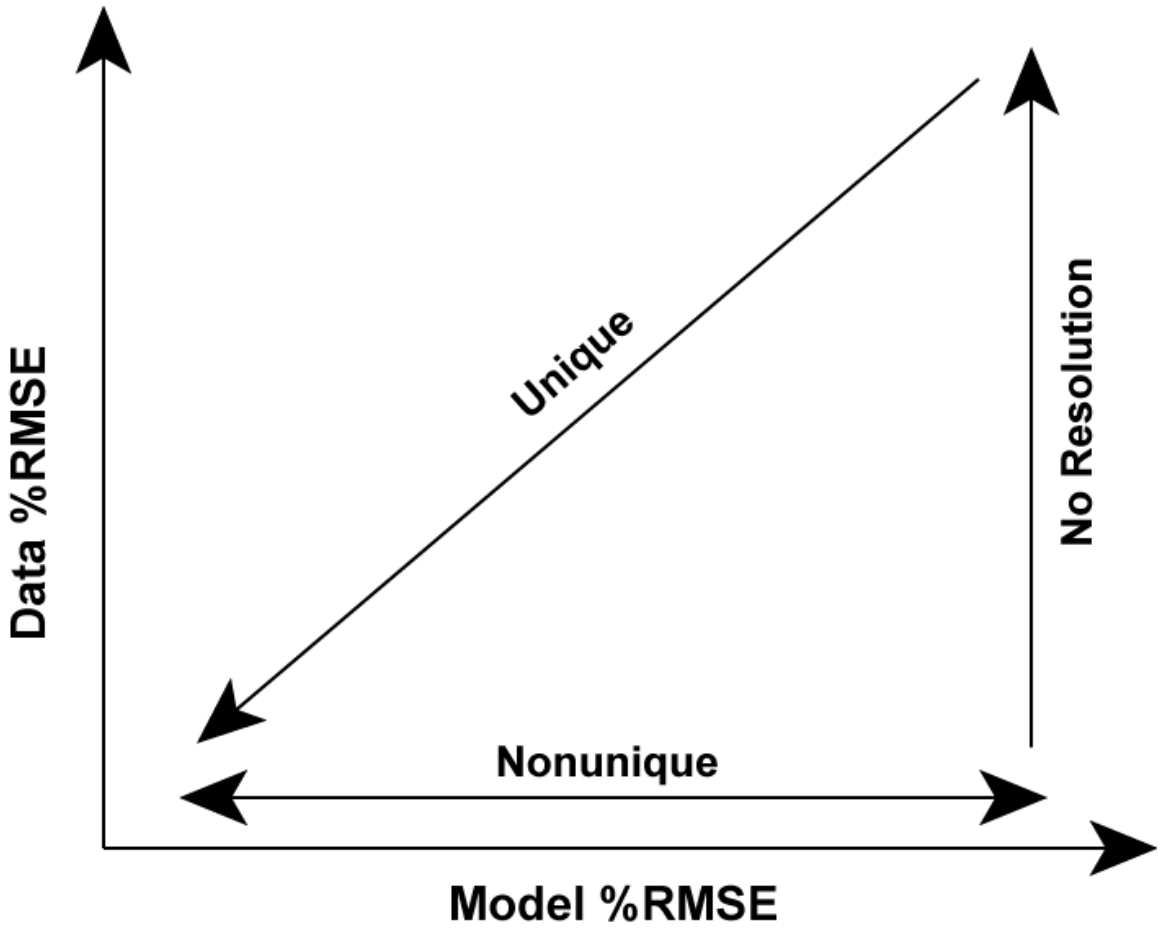


Figure 3-5 The predicted path of data and model errors as a function of the number of layers (after (Simms & Morgan, 1992)). As the number of layers increases, both data and model errors decrease along the line labeled 'Unique'. This line is so labeled because there is hypothetically a single, unique model that minimizes the data error (for a specified number of layers). The point where the path inverts is the point where data and model errors are both minimized, and this point indicates the correct layer model. As the number of layers increase beyond this point, the data error will not significantly reduce, but model errors will increase. This is a region of nonuniqueness, because there are multiple models that equally minimize the data error, which is the definition of nonuniqueness. Eventually, the number of layers becomes so large that the inverse problem is ill-posed, which is denoted along the line labeled 'No Resolution'.

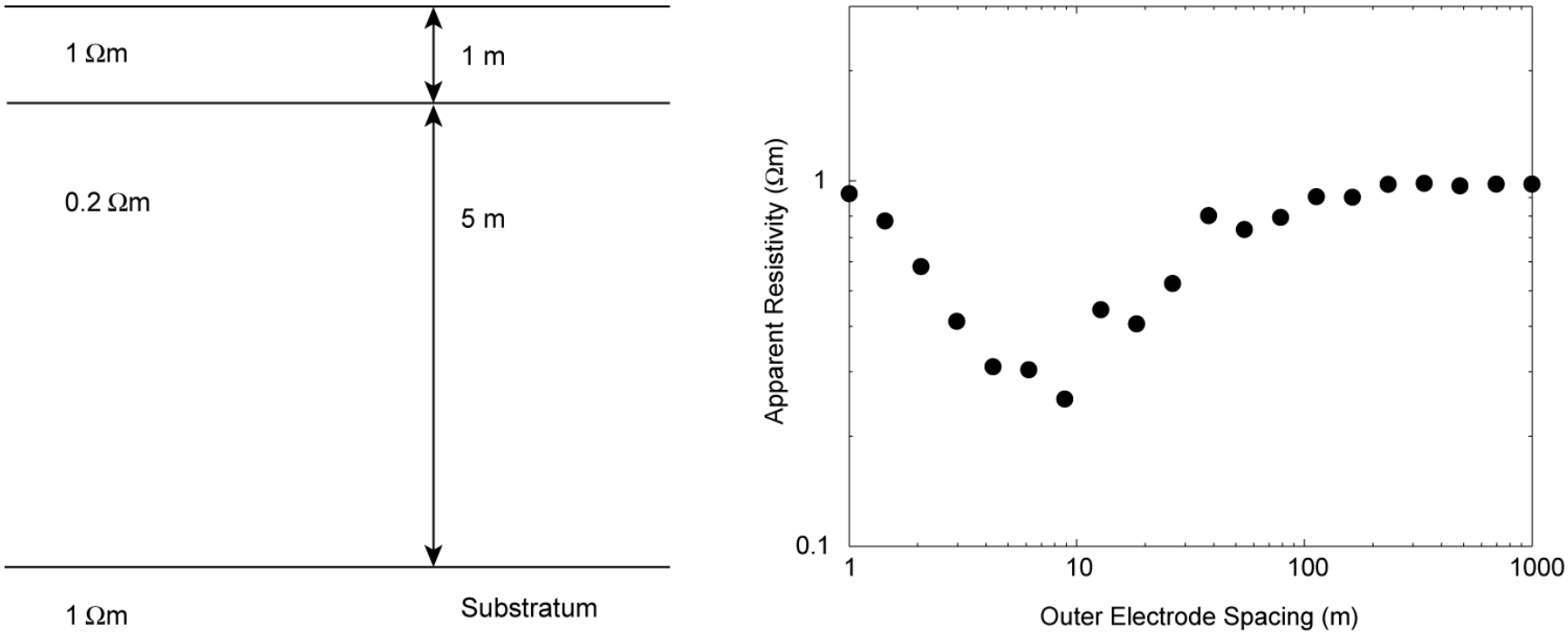


Figure 3-6 1D earth model (left) and 20 synthetic apparent resistivity data points (right) querying the model. Outer electrodes were logarithmically distributed over three decades. 5% Gaussian noise was added to the data.

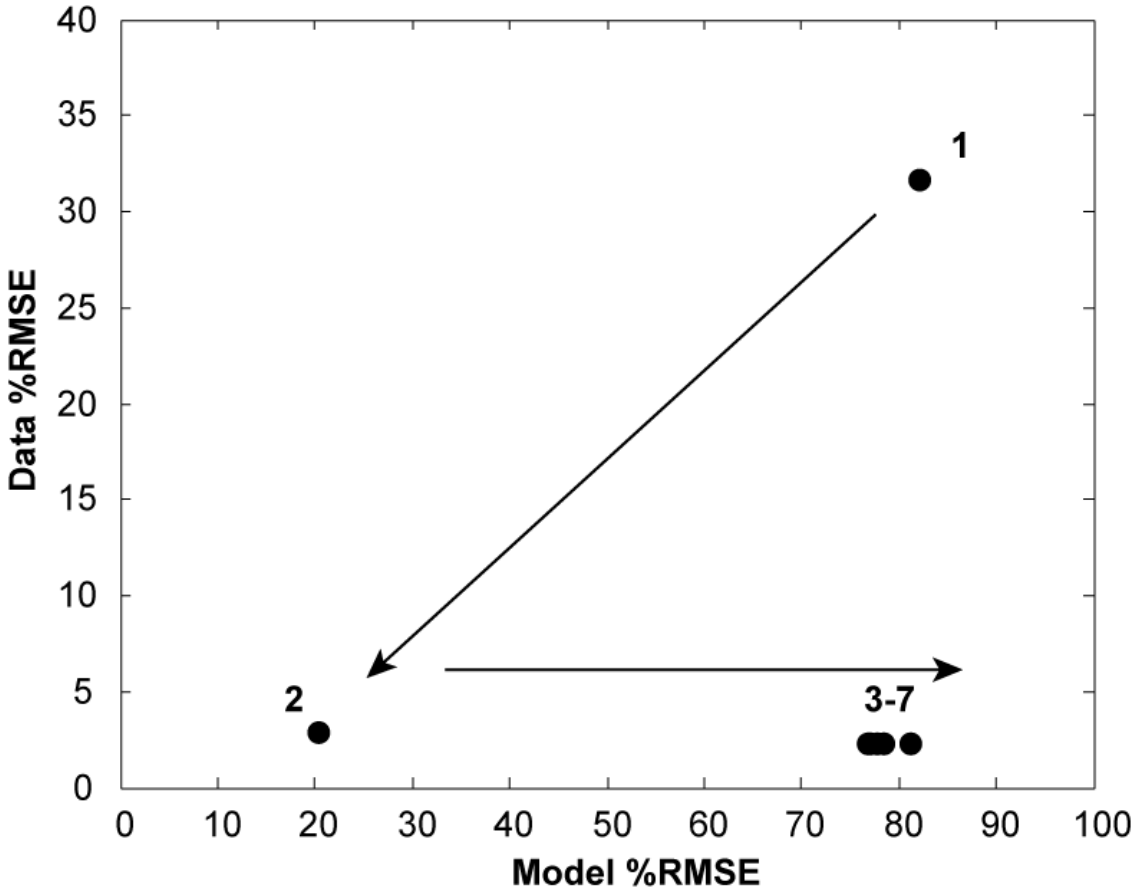


Figure 3-7 Cross plot of data percent rms error versus model percent rms error. As predicted, the phase plot goes through a minimum for the two-layer model, which is the correct layer model. Thereafter, model errors increase while data error remains fairly constant. This is the region of nonuniqueness, where multiple models equally minimize the data error.

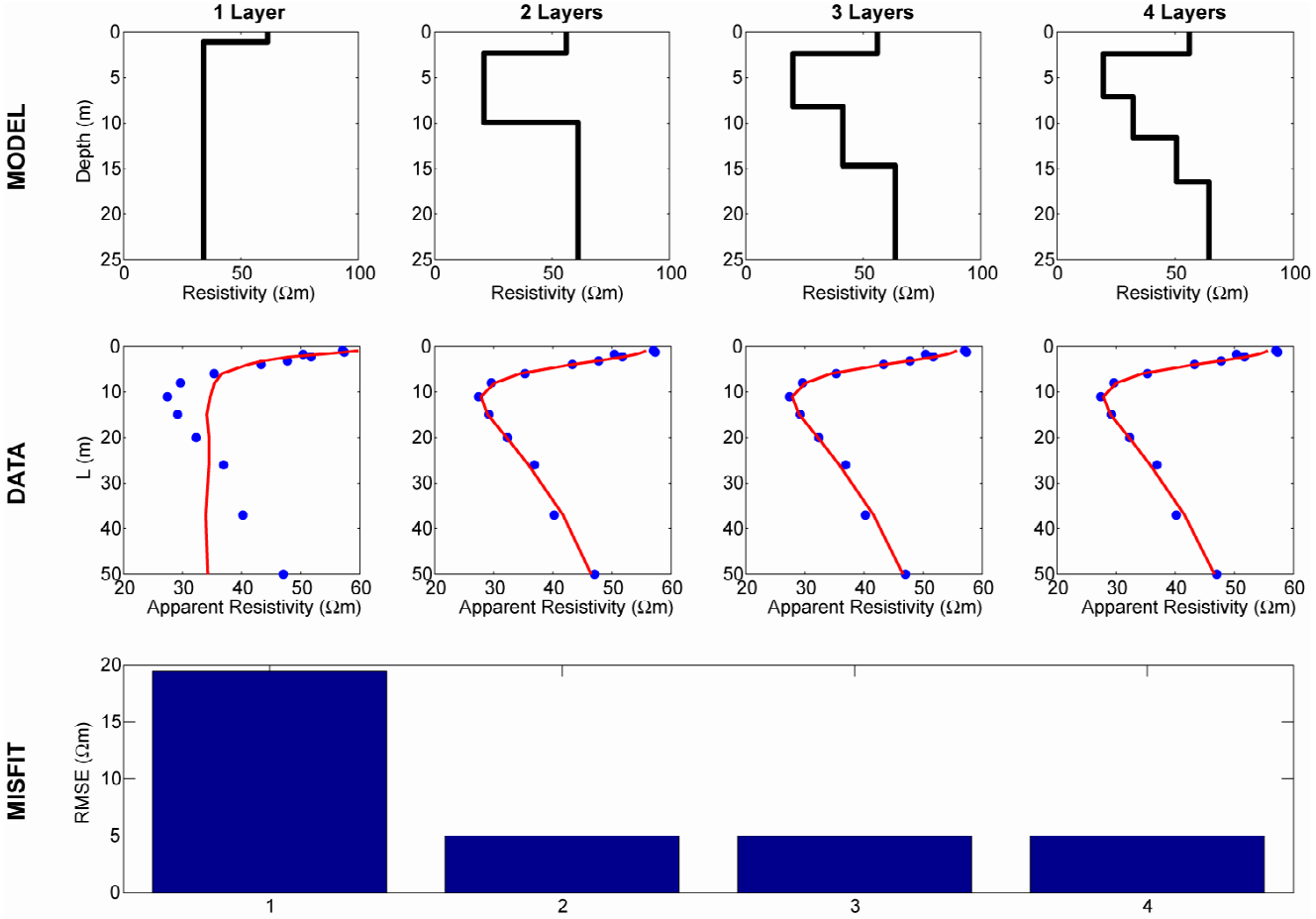


Figure 3-8 Inversion models (top), data and fit (middle), and data rms errors (bottom) for models from 1 to 4 layers. Apparent resistivity data are from a Schlumberger sounding conducted in Curaçao in 2001. The data rms error drops rapidly from the one-layer to the two-layer model and then levels off thereafter. Absent any other information or analysis, the two-layer model should be chosen.

Number of Layers	1	2	3	4	5	6	7
Model %RMSE	82.19%	20.44%	81.33%	77.68%	76.85%	77.15%	78.39%
Data %RMSE	31.59%	2.80%	2.27%	2.27%	2.27%	2.27%	2.26%
F (w.r.t. previous model)		93.83	1.33	1	1	1	1.02

Table 3-2 Percent data and model rms errors and F-values reporting the ratios of the mean squared errors for models of successive numbers of layers. To illustrate, the first F-value is the ratio of the mean squared error of the one-layer model to the mean squared error of the two-layer model; the next F-value is the ratio of the data mse of the two-layer to three-layer model, and so on. Because twenty data were used, F-test confidence levels were calculated for samples both with 20 degrees of freedom. The F-test is significant at 95% confidence if the F-value exceeds 2.1242; for 85% confidence, the F-value must exceed 1.6023; for 75% confidence, the F-value must exceed 1.3580. From F-values listed above, it is clear that the two-layer model is the correct model. Interestingly, note that the percent data rms error is monotonic non-increasing from 1 to 7 layers. The F-test has allowed us to statistically identify the proper layer-model despite non-increasing data errors.

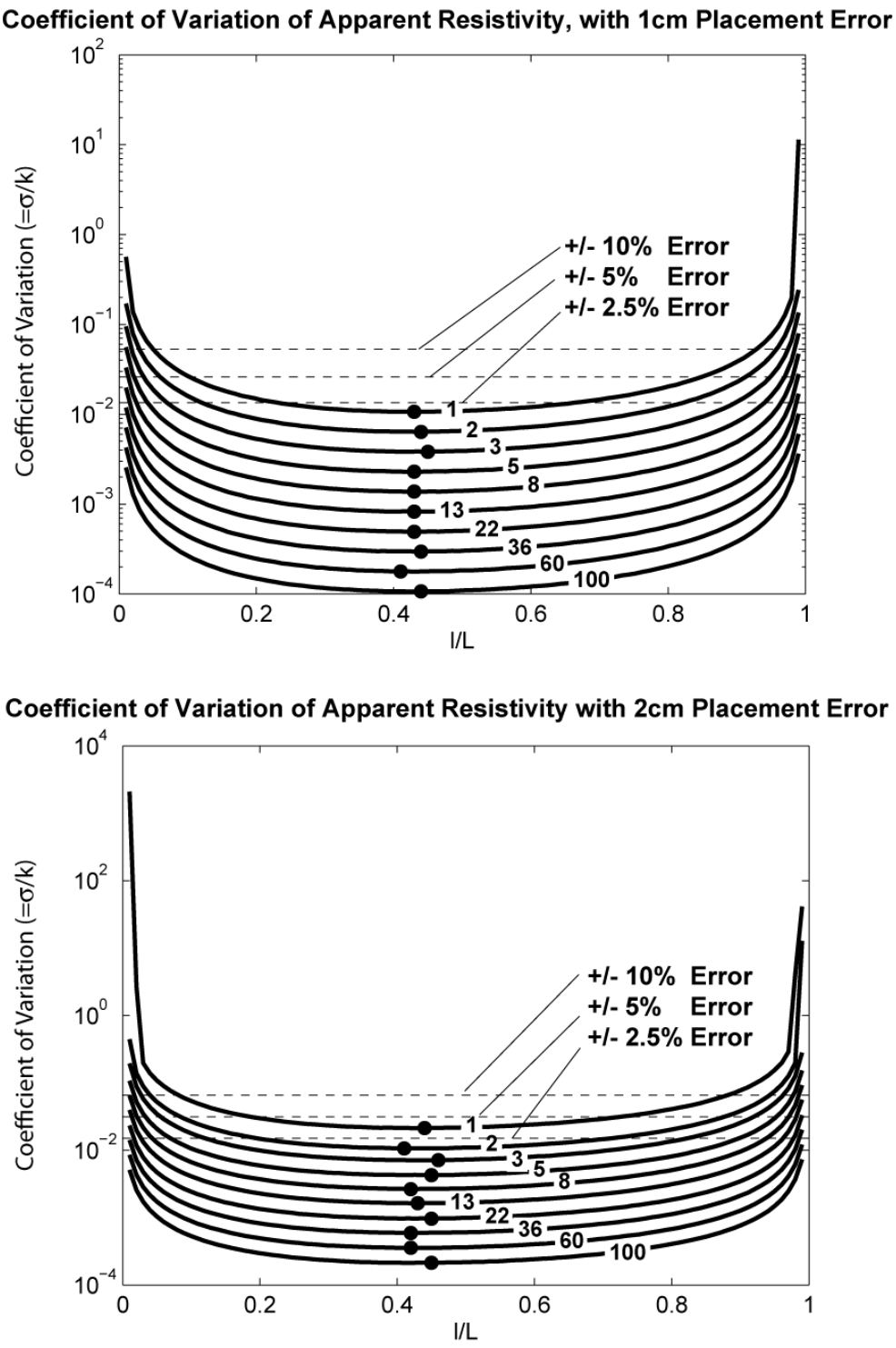


Figure 3-9 Families of curves showing expected errors in apparent resistivity estimates due to errors in electrode placement. The curves plot the coefficients of variation (standard deviation over the mean) of apparent resistivity versus the ratio of the inner electrode to outer electrode spacing. The curves are organized according to fixed *L*; for example, the curve labeled '2' illustrates coefficients of variation for *Ll*-

pairs where the outer electrode spacing is fixed at 2 m. The top panel shows results for electrode placement errors that are ± 1 cm, with 95% confidence; the lower panel is for placement errors of ± 2 cm with 95% confidence. Dashed lines show thresholds of ± 10%, ± 5% and ± 2.5% data error, for reference. Two notable points are: (1) expected data errors decrease as the outer electrode spacing increases and (2) the curves all appear to go through minima (black dots) at the same point, where $l/L \cong 0.43 \pm 0.01$.

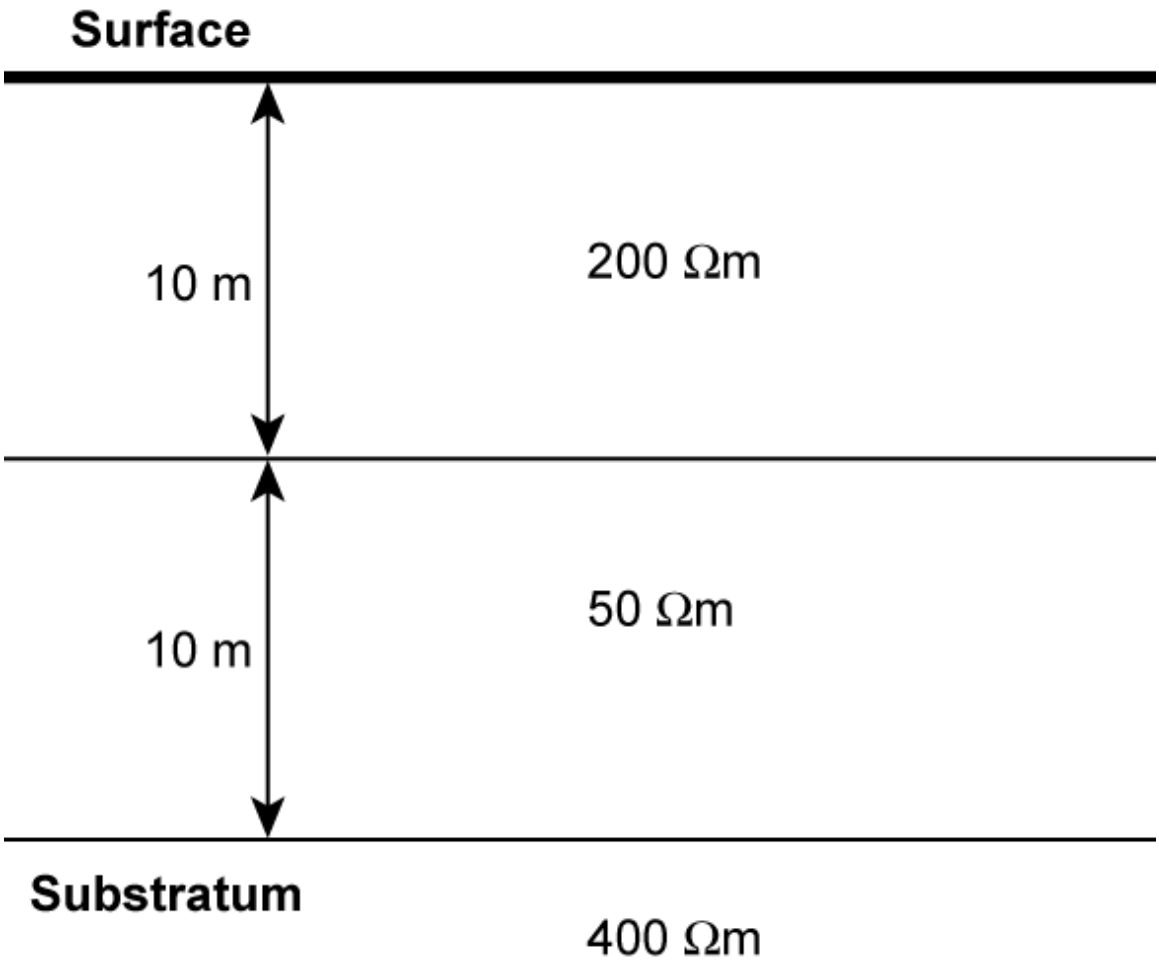


Figure 3-10 Two-layer model with a conductive intermediate layer sandwiched between two more resistive layers.

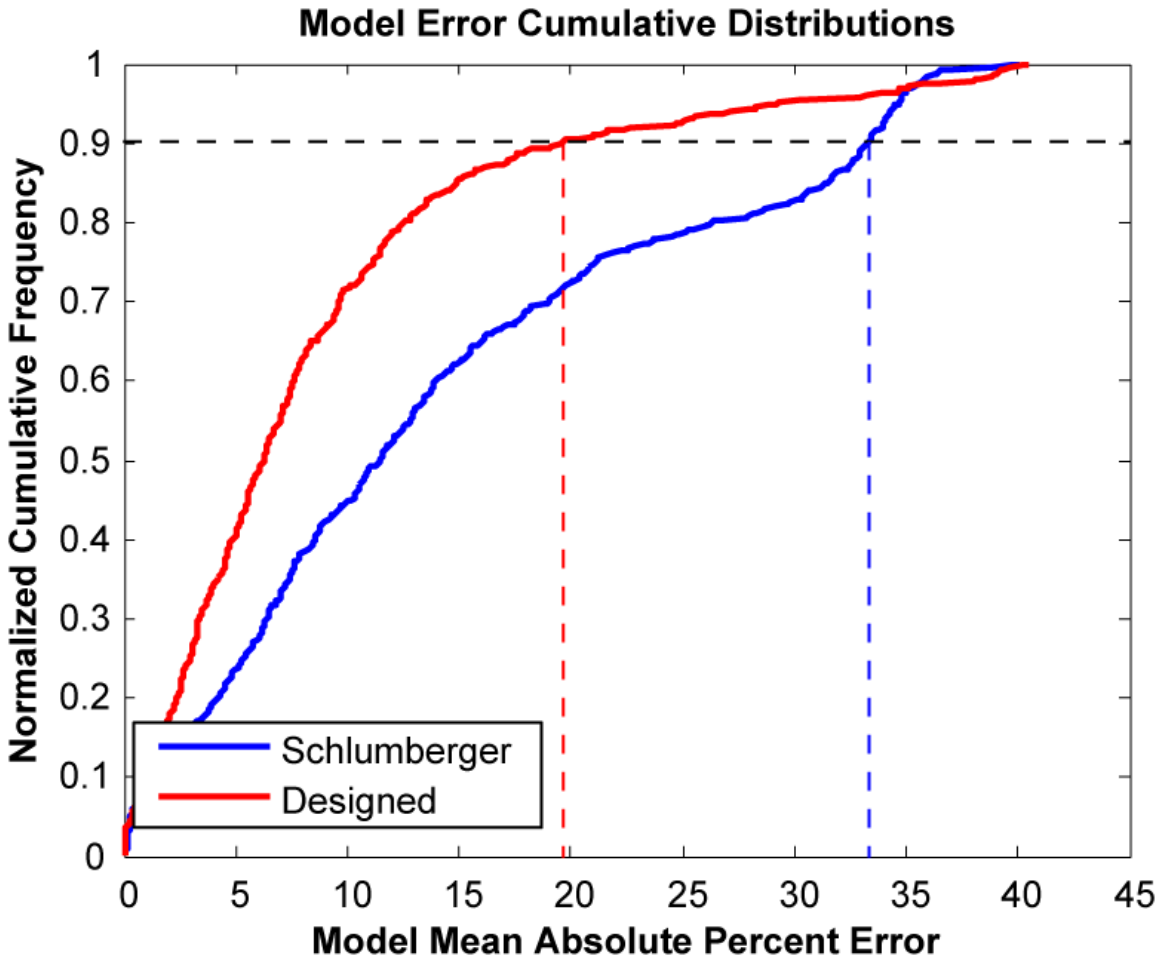


Figure 3-11 Model error cumulative distributions for a Monte Carlo comparison of the Schlumberger survey and a survey designed to minimize data errors created by electrode misplacement. Errors are reported as mean absolute percent error between the true model and the inversion model. At 90% confidence (horizontal dashed line) the designed survey produces modeling errors no greater than ~20% (vertical, red dashed line), whereas the Schlumberger survey produces modeling errors no less than ~34% (vertical, blue dashed line).

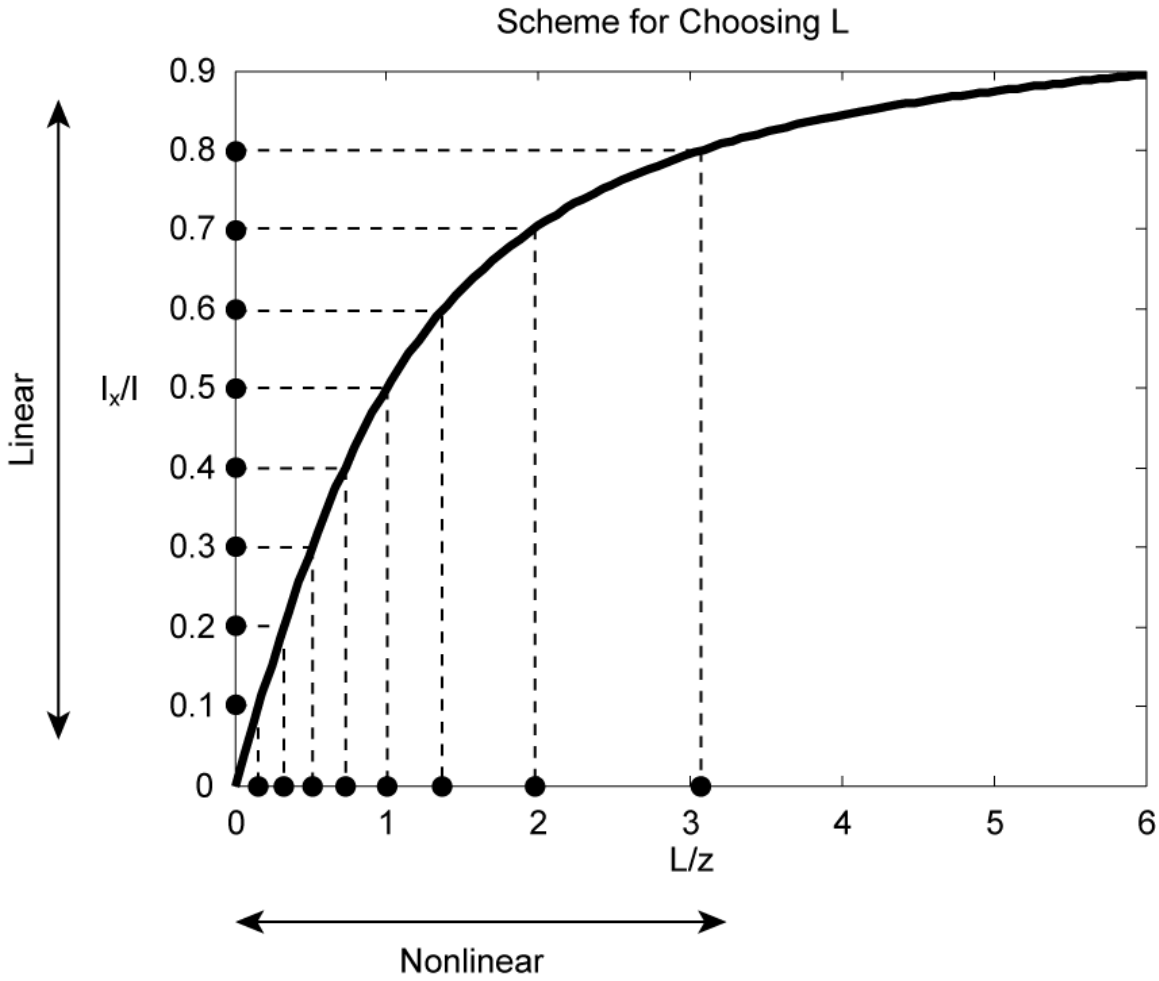


Figure 3-12 Cartoon depicting a pseudo-analytical scheme for choosing electrode positions for a general survey, to be used over unknown targets. The solid curve (after (Telford *et al.*, 1990)) expresses the fraction of the horizontal current, I_x/I, flowing below depth z, as a function of the outer electrode spacing (which has been normalized by z). It was hypothesized, that if one picks a set of I_x/I terms that increase linearly, the derived experiment would provide equal coverage over the depth column of interest. The horizontal dashed lines indicate an example of a set of linear I_x/I picks, and the vertical dashed lines show the prescribed L/z positions. If a maximum investigation depth of z_{max} is desired, one can easily convert the prescribed L/z positions into L positions.

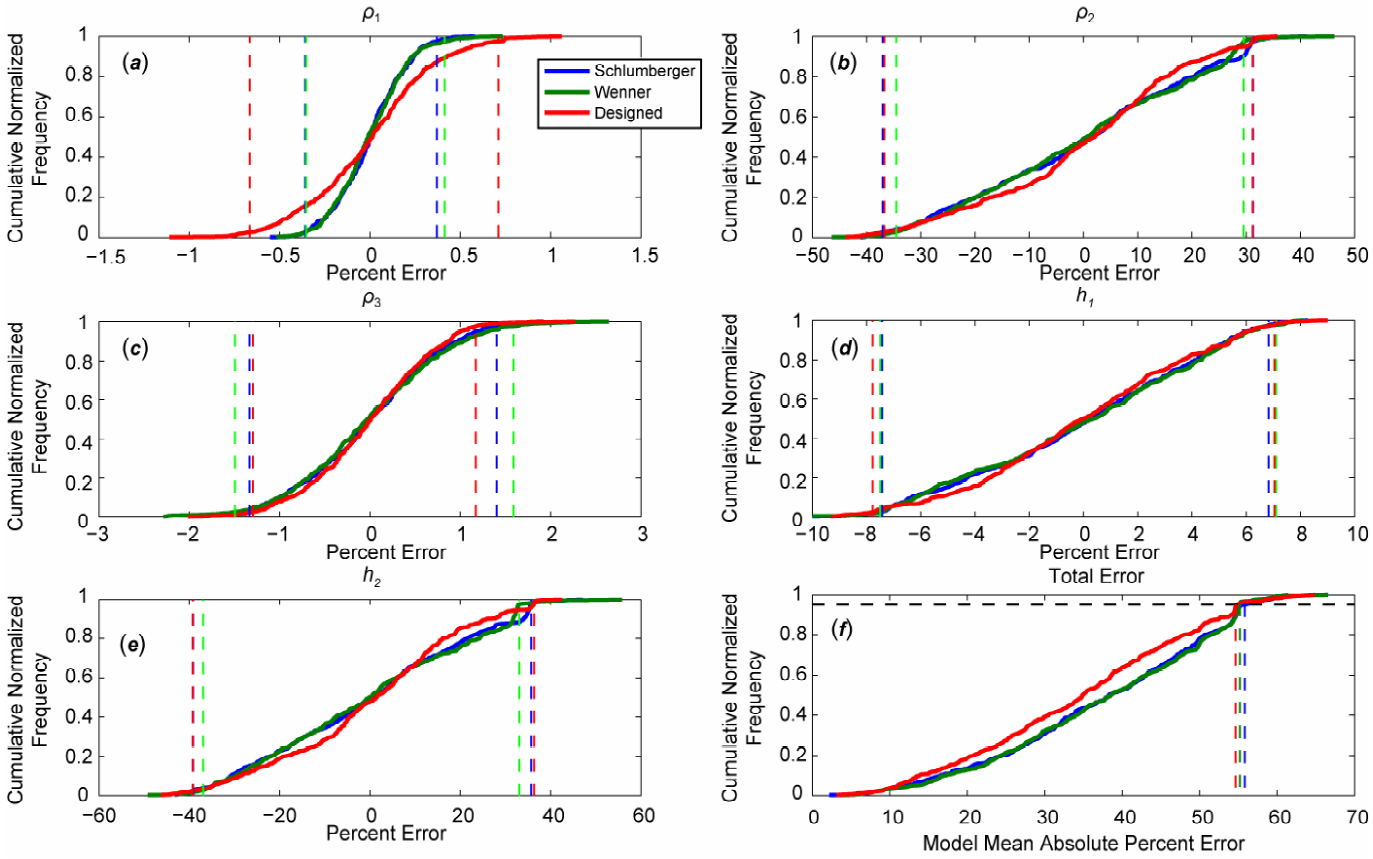


Figure 3-13 Panels *a-e* show the cumulative distributions of percent error between the true and predicted model parameters for a Monte Carlo comparison of Schlumberger, Wenner and Designed surveys. The designed survey was hypothesized to be a superior as a general survey. Also shown in these panels are the 95% confidence intervals (vertical dashed lines). Panel *f* shows the cumulative distributions and 95% confidence levels (dashed lines) of mean absolute percent errors between the true and predicted models for the three surveys. From these results, it is concluded that performances of the three surveys are essentially identical.

Sequential Versus Global Optimization Algorithms

Number of Observations	Algorithm	CPU Time (s)	Fraction of Global CPU Time	det(G^TG)	Fraction of Global Optimumn
5	Global	82	0.33	5.36	0.58
	Sequential	27		3.09	
10	Global	398	0.14	164.9	0.85
	Sequential	57		140.7	
15	Global	1850	0.05	1298	0.89
	Sequential	92		1154	
20	Global	10881	0.01	5481	0.92
	Sequential	129		5031	

Table 3-3 Comparison of sequential and global optimization algorithms for a 1D resistivity OED problem. The model for which designs were executed is shown in Figure 3-10. The underlying objective was to maximize the determinant of the (auto-inner product) of the Jacobian matrix, $\det \mathbf{G}^T\mathbf{G}$. Fractional CPU Times and Fractional Optimality are also shown, reporting the ratios of sequential-over-global CPU time and sequential-over-global optimality (as measured by $\det \mathbf{G}^T\mathbf{G}$), respectively.

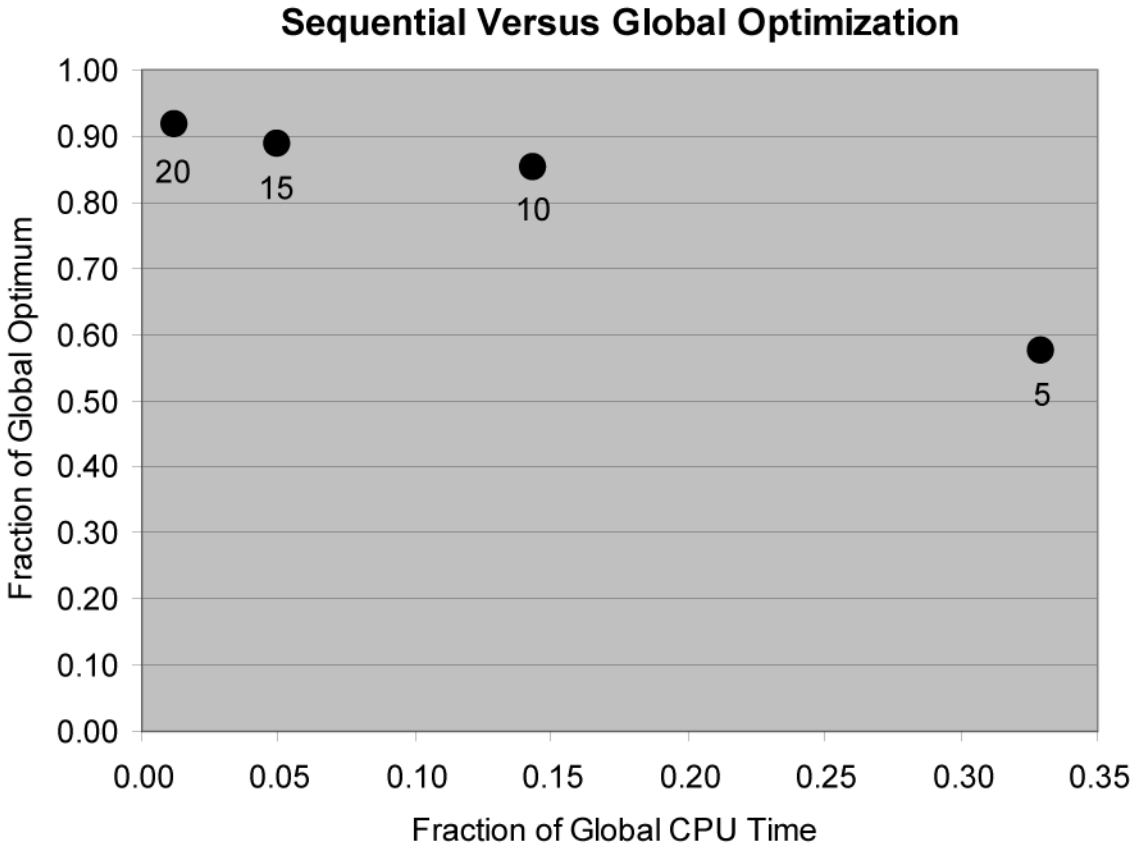


Figure 3-14 Graphical synopsis of results tabulated in Table 3-3. Shown is a crossplot of the sequential-over-global optimality ratio versus the sequential-over-global CPU time ratio. Numbers under each point indicate the number of observations used in the designs.

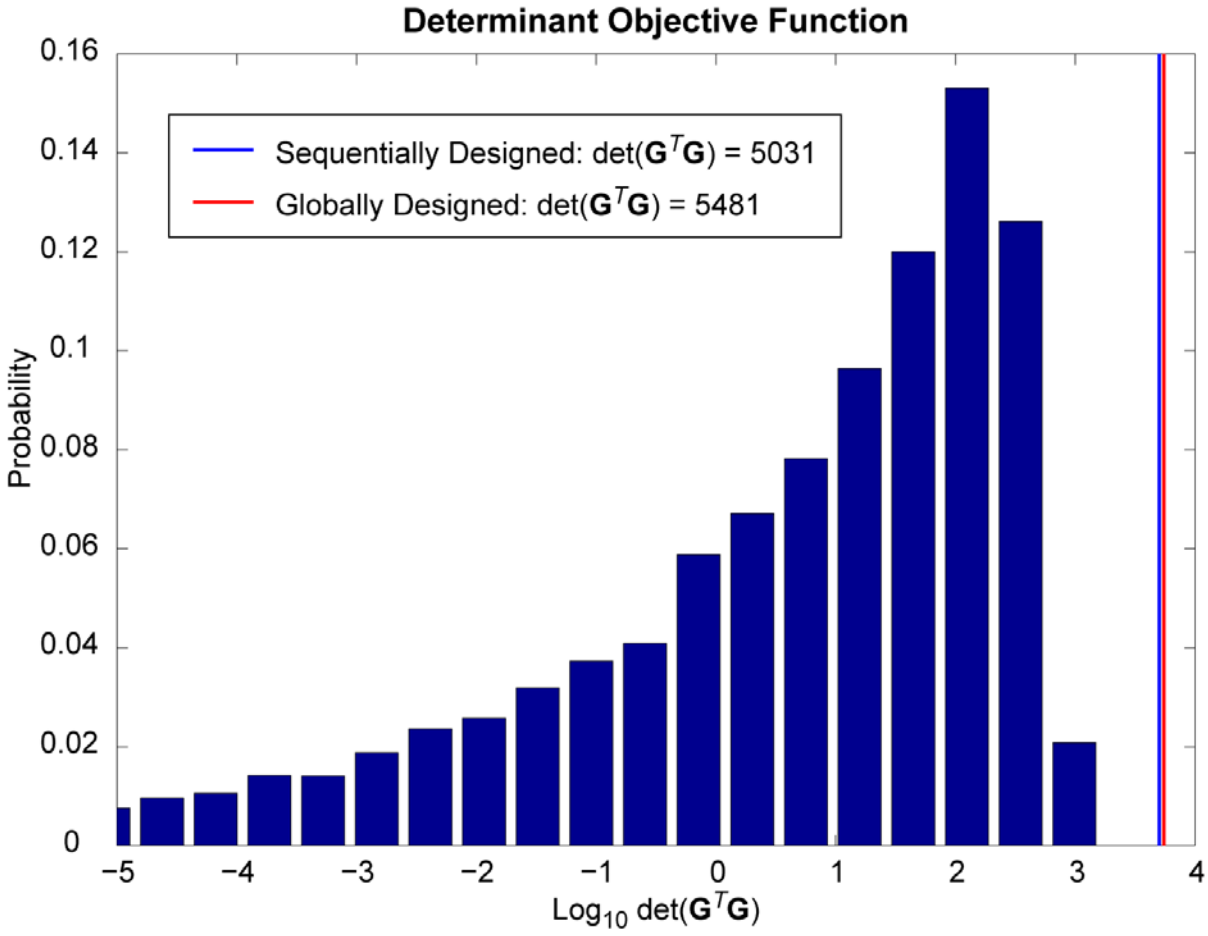


Figure 3-15 Comparison of 20-observation surveys. The bar plot shows the probability distribution of $\det \mathbf{G}^T\mathbf{G}$ for 10,000 randomly generated surveys. Also shown is $\det \mathbf{G}^T\mathbf{G}$ of the sequentially (blue line) and globally (red line) designed surveys. Note the *x*-axis is log scaled to accommodate the many orders of magnitude in the determinants of randomly generated surveys.

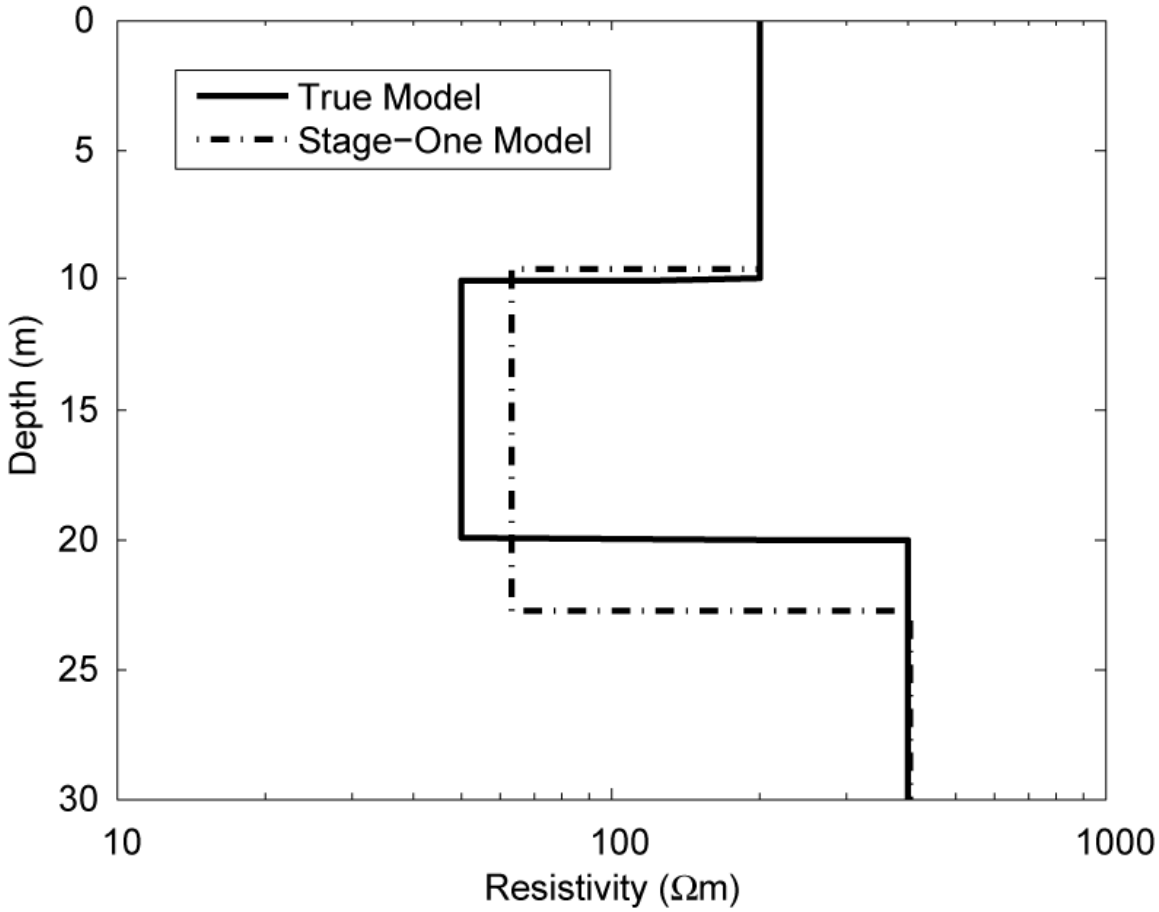


Figure 3-16 True earth model (solid line) and stage-one inversion model (broken line) from a noisy Schlumberger data set. Five percent Gaussian noise had been added to the Schlumberger data.

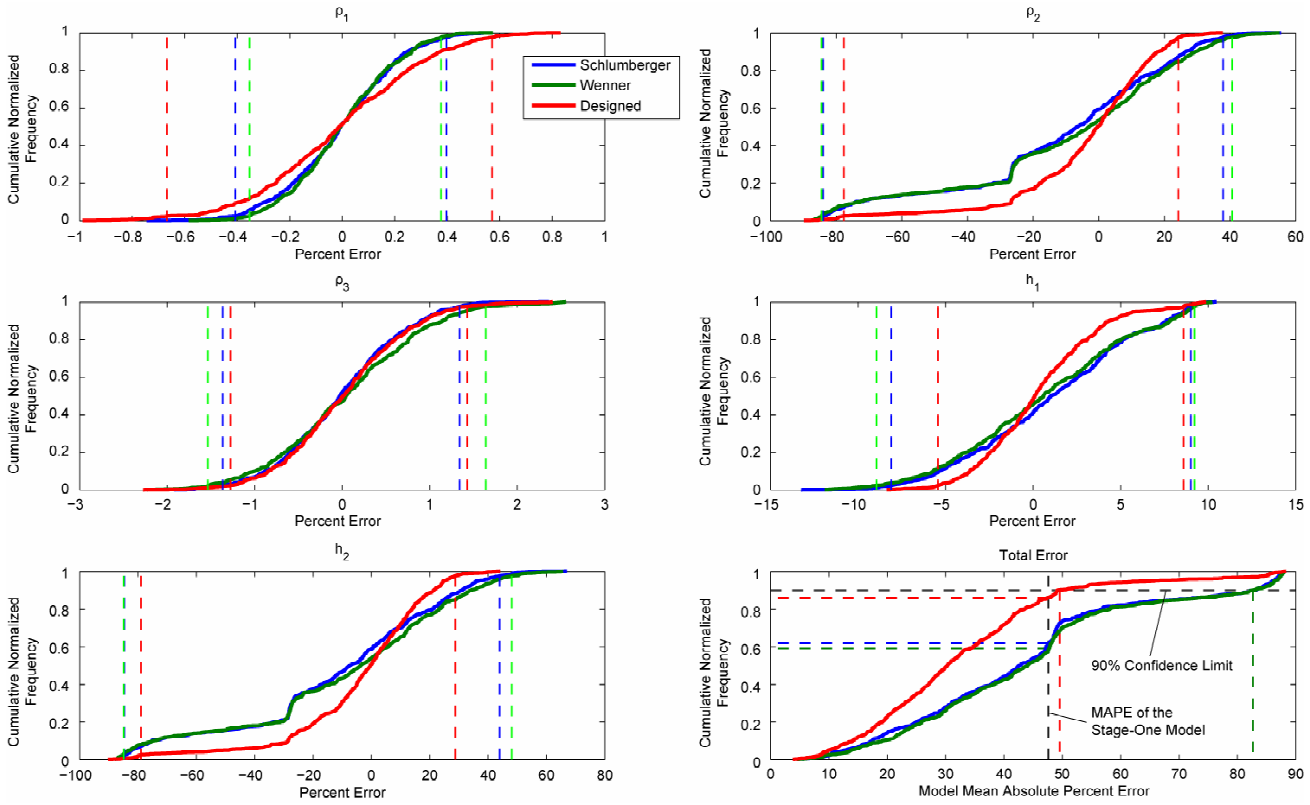


Figure 3-17 Cumulative error distributions for a Monte Carlo comparison of three surveys (Schlumberger, Wenner and Designed) used as second-stage inversions. Panels *a*-*e* show percent errors for the five model parameters and 95% confidence intervals (vertical, colored dashed lines). Panel *f* shows the cumulative distributions of mean absolute percent errors for the three surveys. Also shown in Panel *f* are the 90% confidence levels (horizontal, black dashed line and colored vertical lines projected from the intersections of the 90% confidence level and the error curves to the *x*-axis) and the MAPE of the stage-one inversion model (vertical, black dashed line and colored horizontal lines projected from the intersections of the stage-one MAPE line and the error curves to the *y*-axis). From these it is seen that the designed survey produces expected % modeling rms errors of less than

~49% with 90% confidence, while the Schlumberger and Wenner surveys produce significantly poorer modeling errors, being only less than ~83% with 90% confidence. It is also seen that ~86% of modeling errors for the designed survey are less than the error for the stage-one inversion, while only ~60% of the Schlumberger and Wenner errors are less than the stage-one error.

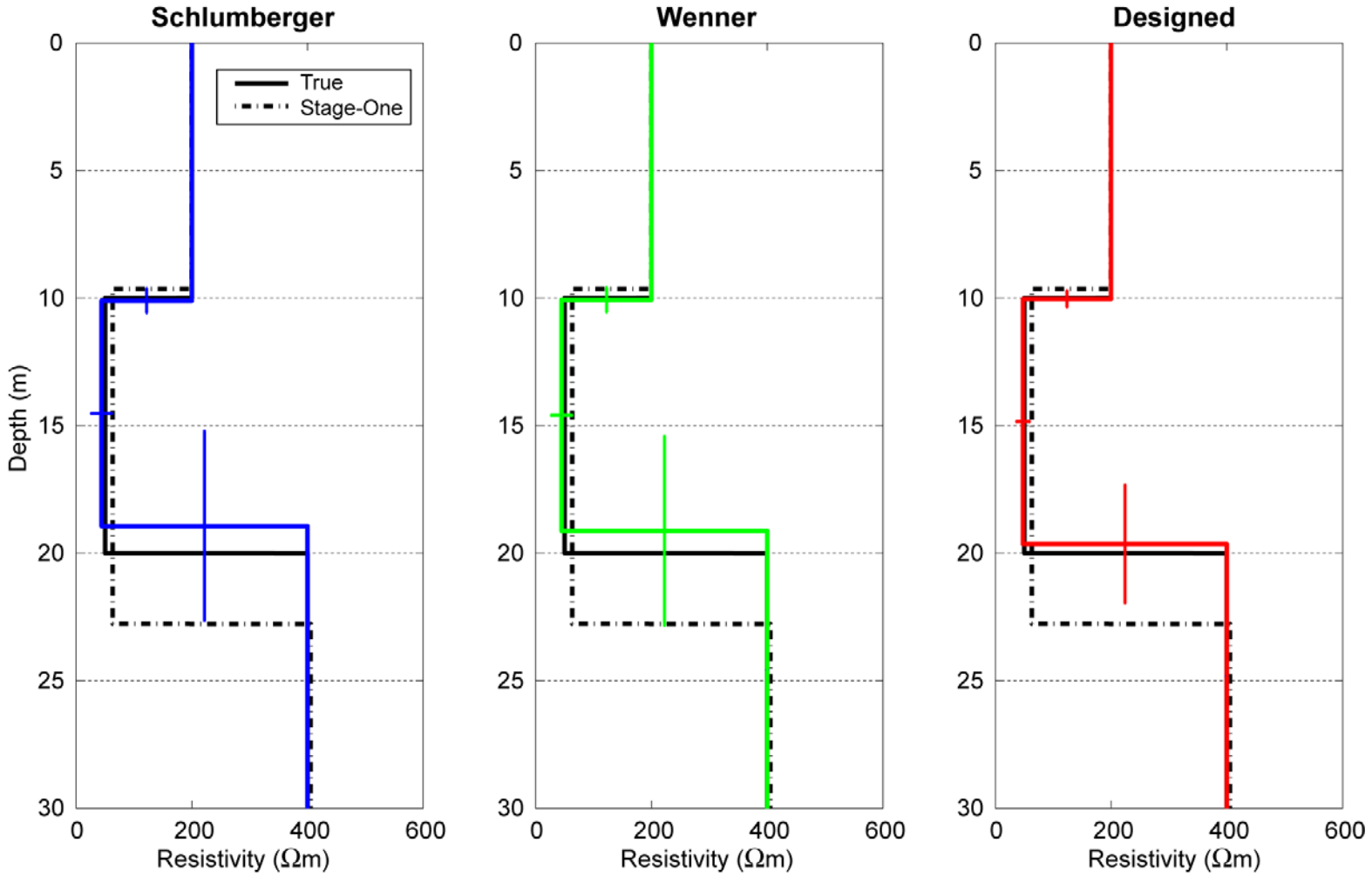


Figure 3-18 Monte Carlo results for three experiments used for a second-stage inversion, all starting with an initial model guess from the 'Stage-One' inversion model. In all cases, data were contaminated with 5% Gaussian noise and inversion models were

generated for 500 such realizations. Shown are the true model profile (solid black line), the 'Stage-One' inversion model (broken black line), and the Monte Carlo average models for the Schlumberger (blue), Wenner (green) and adaptively Designed (red) surveys. Also shown are Monte Carlo error bars at two standard deviations, to give an idea of the degree of parameter uncertainty each experiment produced.

	Stage-One Model			True Model		
	Schlumberger	Wenner	Designed	Schlumberger	Wenner	Designed
ρ_1	0.0019	0.0019	0.0029	0.0019	0.0019	0.0029
ρ_2	0.1260	0.1447	0.0796	0.2767	0.3175	0.1767
ρ_3	0.0073	0.0088	0.0075	0.0072	0.0086	0.0074
h_1	0.0379	0.0430	0.0257	0.0491	0.0557	0.0326
h_2	0.1476	0.1702	0.0946	0.3038	0.3492	0.1956

Table 3-4 Normalized parameter uncertainties predicted by taking the square root of the diagonal of the model covariance matrix and dividing by the model parameter values (Note: this is the coefficient of variation). The 'stage-one model' column compares parameter uncertainties for the Schlumberger, Wenner and adaptively optimized surveys evaluated at the stage-one model. The 'true model' column shows the same, except evaluated at the true model. The designed survey generally produces smaller parameter uncertainties, particularly with respect to the least well-resolved parameters, ρ_2 and h_2. However, it trades resolution of the well-resolved parameters, ρ_1 and h_1, to do this. Nonetheless, uncertainties for the well-resolved parameters are so small that this trade is reasonable.

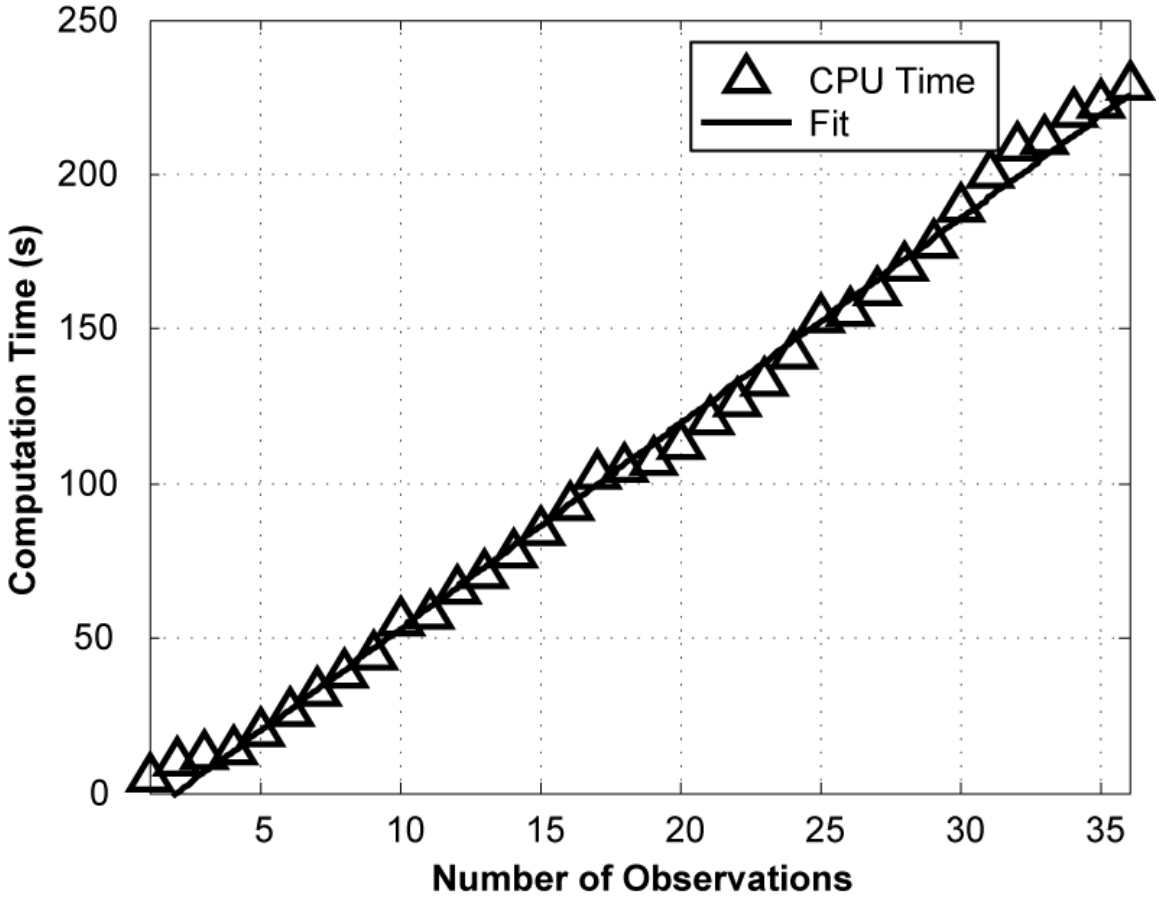


Figure 3-19 CPU times for experiments from 1 to 36 observations designed using our sequential OED algorithm. All CPU times were clocked with a dual 2GHz laptop with 2GB RAM.

Chapter 4

Single Borehole DC Resistivity

4.1 Introduction

Researchers have examined various aspects of axially symmetric borehole DC resistivity modeling and inversion for more than two decades (e.g., Yang & Ward, 1984; Zemanian & Anderson, 1987; Liu & Shen, 1991; Zhang & Xiao, 2001; Spitzer & Chouteau, 2003; Wang, 2003). These methods model the earth with a set of rectangular prisms that are treated as being axially symmetric about the borehole. Axial symmetry imposes a strong assumption on the lithological structure of the earth, for there are only a limited number of scenarios where this symmetry attains, particularly horizontal layering. Despite the strong restriction, axially symmetric geoelectrical imaging is useful in settings where data can only be collected from a single borehole, despite the limitation that the azimuthal position of a discrete 3D anomaly cannot be pinpointed because of the non-directionality of source and receiver electrodes.

This chapter examines optimal experimental design applied to single-borehole DC resistivity. The borehole problem is cast in cylindrical coordinates, and the forward and inverse problems are discretized using a transmission network analogy (Swift, 1971; Madden, 1972; Zhang *et al.*, 1995; Shi, 1998) adapted to cylindrical coordinates. We briefly develop the numerics used for the borehole resistivity forward and inverse problems, from which several issues pertinent to the design of experiments in

the borehole are identified. Additionally, some effort is spent circumscribing the OED problem with respect to electrical tomography problems, which can serve to expedite design times not only for borehole problems but any geoelectrical scenario where electrode positions are fixed on a pre-determined grid. Multiple experimental design objective functions proposed in Chapter 2 are examined, all of which were solved using our greedy, sequential design algorithm. Inversion results have been produced for all investigations in order to compare optimally designed experiments with randomly generated and standardized ones. Finally, several adaptive OED techniques are explored.

This chapter comprises several research areas, and a slightly unorthodox format has been adopted to present this research. Rather than presenting holistic sections on theory, methodology, results and discussion for the entire chapter, each 'subtopic' is self-contained, with pertinent theory, methodology and so forth being presented in context. These subtopics comprise the framework of optimal experimental design for single-borehole DC resistivity. The subtopics are presented in the following order: (1) Modeling and Inversion; (2) Electrical Resistivity Tomography and Experimental Design; (3) Comparing the Three Quadrupole Configurations; (4) Random Experiments Versus Designed Experiments; (5) Two-Stage, Adaptive Experimental Design; (6) In-Line, Adaptive Optimal Experimental Design; (7) Small Batch, Adaptive Experimental Design; (8) Noisy Data and Experimental Design. Following these sections, a Summary is provided that integrates and summarizes the results of the various topic areas.

4.2 Modeling and Inversion

The single-borehole resistivity problem places both transmitting and receiving electrodes along the borehole axis. As explained above, the azimuth of a resistivity feature cannot be resolved in this formulation, so the forward and inverse models are simplified by forcing resistivity to be azimuthally independent. Consequently, cylindrical coordinates were adopted, with resistivity being radially symmetric about the

borehole axis (Figure 4-3). In this manner, the borehole resistivity problem can be treated as a –pseudo-2D tomography problem as an effect of this simplification. This model may be oversimplified for many field situations, but the purpose of this paper is to investigate *optimal experimental design* so it is reasonable to consider problems that do not unnecessarily complicate or mask the effects specifically attributable to designed experiments.

The pseudo-2D borehole resistivity forward and inverse problems were treated using an adaptation of the transmission line network analogy introduced by Swift (1971) and Madden (1972) and employed by Zhang *et al.* (1995) and Shi (1998). This method approximates the earth as a discrete mesh of impedances connected together in a network and subject to Kirchoff's First and Second Laws. Mixed boundary conditions, introduced by Zhang et al. (1995), were used at the boundaries of the model grid. Solving the transmission network equations, one can model the potentials observed at electrodes along the borehole string owing to current-sources applied elsewhere along the string. Moreover, the sensitivities of these data, with respect to the formation resistivities in the model cells, can also be determined from the transmission network equations, yielding the Jacobian matrix, **G**, which is used for experimental design and for data inversion. The interested reader is referred to the preceding citations for exact details

The inversion program used in this chapter employed nonlinear least squares formalism with an L^1-norm constraint on the gradient of the model parameters. The linearized forward problem is denoted

$$\mathbf{G}\Delta\mathbf{m} = \Delta\mathbf{d}\,, \tag{4.1}$$

where **G** is the Jacobian matrix, $\Delta\mathbf{m}$ is the model update vector, and $\Delta\mathbf{d}$ is the difference between observed and predicted data. Formally, the inversion objective was given by

$$\Theta = \left\|\Delta\mathbf{d} - \mathbf{G}\Delta\mathbf{m}\right\|_2^2 + \lambda\left\|\nabla\mathbf{m}\right\|_1, \tag{4.2}$$

where λ is a Lagrange multiplier governing the trade-off between the two competing objectives. The Levenberg-Marquardt algorithm (Levenberg, 1944; Marquardt, 1963)

was used to govern the tradeoff factor, λ, between data rms error and the L^1 model-smoothness constraint.

4.3 Electrical Resistivity Tomography and Experimental Design

To optimize the use of computational resources for ERT-based experimental design, it is worthwhile to first examine the character of the experimental space, with a particular interest in finding shortcuts that can expedite the design algorithm.

Daily *et al.* (2004) and Stummer *et al.* (2004) have shown that for a set of N electrodes there exist 3_NC_4 (thrice 'N choose 4') unique combinations of quadrupole transmitter/receiver combinations. To understand this, observe that there are $_NC_4$ unique combinations of four electrodes from a set of N. There are 24 ways of permuting four electrodes, which suggests that there should be a total of 24_NC_4 different transmitter/receiver combinations. However, mindful that both transmitter and receiver are dipoles, there are only $6 = {}_4C_2$ unique combinations from any set of 4 electrodes (i.e., once the transmitter dipole is specified, the receiver dipole is also automatically specified, and vice versa) so the total number of transmitter/receiver pairs appears to be 6_NC_4. Finally, reciprocity stipulates that the roles of transmitter and receiver are interchangeable, so the true number of transmitter/receiver pairs comes out to be 3_NC_4 as has been indicated in the literature.

The nature of ERT experimental space, which comprises these 3_NC_4 transmitter/receiver combinations, becomes a bit clearer. If 20 electrodes have been deployed, then there are 14,535 unique transmitter/receiver combinations or data stations. If the geophysicist chooses to design an experiment with S data stations, there are $_{14,535}C_S$ for the designer to choose from; this is the size of the space over which the designer's algorithm must search. To get some sense of the size of this search space, if S = 200 then there are approximately 10^{457} surveys in the search space! It should

also be clear that this is a discrete search space. Infinitesimal changes in a survey are undefined, as are continuous transitions from one survey to the next. This is why researchers have been forced to use so-called global search algorithms that are suited for discrete search spaces (unlike faster algorithms that rely on the gradient information).

The discussion above outlines the extent to which current research has elucidated ERT experimental space. It is nevertheless possible to delve deeper. To do so it is necessary to draw attention to two facts. The first concerns the factor of three in $3_N C_4$. It specifies the three unique ways in which any four electrodes can be arranged into transmitter and receiver dipoles (accounting for reciprocity and polarity switching). These three transmitter/receiver combinations are named *internal*, *external*, and *straddling* and are shown in Figure 4-1. The second fact is this: any quadrupole transmitter/receiver data station can be expressed as a unique superposition of four pole-pole data stations. To demonstrate, let A and B be transmitter electrodes and let M and N be receiver electrodes. The potential difference measured across M and N due to the current transmitted across A and B is expressed as

$$\phi_{\mathrm{ABMN}} \equiv \phi_{\mathrm{AM}} - \phi_{\mathrm{AN}} - \phi_{\mathrm{BM}} + \phi_{\mathrm{BN}}, \tag{4.3}$$

where ϕ_{AM} is the theoretical potential at receiver pole M due to a current injected at transmitter pole A.

It is evident therefore that the $3_N C_4$ data stations that comprise experimental space are all linear combinations of the ${}_N C_2$ possible pole-pole stations. Experimental space can therefore be algebraized by indexing the set of all pole-pole data stations for N electrodes and creating an *incidence* or *picker matrix*, **P**, which linearly combines the indexed pole-pole stations to create quadrupole stations. In obedience to (4.3), each row of **P** consists of zeros except for four entries, which are set to '1s' and '-1s' and which are indexed to pick the appropriate pole-pole data stations necessary to create particular quadrupole stations. **P** has dimensions $3_N C_4 \times {}_N C_2$ ($3_N C_4$ for the total number of transmitter/receiver stations and ${}_N C_2$ for the total number of pole-

pole data stations). Viewed in this way, **P** is seen to be a map between pole-pole and quadrupole experimental spaces.

With experimental space algebraized, additional insight arises. Because all quadrupole data stations are linear combinations of pole-pole stations, it follows that the vector space spanned by quadrupole data stations can have no more dimensions than the space spanned by pole-pole stations. The proof is simple: **P** is a matrix that maps pole-pole stations to quadrupole stations and the rank of a rectangular matrix cannot exceed the lesser of its dimensions. **P** has $3\,{}_N C_4$ rows, one for each possible quadrupole station, and ${}_N C_2$ columns, one for each possible pole-pole station, ergo

$$\operatorname{rank}(\mathbf{P}) \le \min\left(3\,{}_N C_4\,,\,{}_N C_2\right) = {}_N C_2 = \frac{1}{2} N\left(N-1\right). \tag{4.4}$$

This an important, fundamental result. The general linearized forward problem expressed in Eq. (4.1) can be modified to explicitly incorporate **P** as follows:
Let

$$\mathbf{G}^{(2)}\Delta\mathbf{m} = \Delta\mathbf{d}^{(2)} \tag{4.5}$$

define the forward linear problem for pole-pole data stations, where $\mathbf{G}^{(2)}$ and $\Delta\mathbf{d}^{(2)}$ are respectively the pole-pole sensitivity matrix and pole-pole data vector. The quadrupole forward problem is then defined as

$$\mathbf{G}^{(4)}\Delta\mathbf{m} = \Delta\mathbf{d}^{(4)}, \tag{4.6}$$

where

$$\mathbf{G}^{(4)} \equiv \mathbf{P}\mathbf{G}^{(2)} \text{ and } \Delta\mathbf{d}^{(4)} \equiv \mathbf{P}\Delta\mathbf{d}^{(2)}. \tag{4.7}$$

A straightforward result from linear algebra specifies that the rank of the product of two matrices cannot exceed the lesser of their ranks; hence,

$$\begin{aligned} &\operatorname{rank}\left(\mathbf{G}^{(4)}\right) = \operatorname{rank}\left(\mathbf{P}\mathbf{G}^{(2)}\right) \le \min\left[\operatorname{rank}(\mathbf{P}), \operatorname{rank}\left(\mathbf{G}^{(2)}\right)\right] \le \operatorname{rank}(\mathbf{P}) = {}_N C_2 \\ &\therefore \qquad \operatorname{rank}\left(\mathbf{G}^{(4)}\right) \le \frac{1}{2} N\left(N-1\right) \end{aligned}. \tag{4.8}$$

In words, it is impossible for the quadrupole sensitivity matrix to span a space larger than ${}_N C_2$ dimensions. For example, if 20 electrodes are deployed, the rank of $\mathbf{G}^{(4)}$ cannot exceed 190, *no matter the number of model parameters or the number or data*

points collected. This is one of the reasons why electrical resistivity tomography is so difficult: the inverse problem is almost always ill posed, even when more data points have been collected than there are parameters. More to the point, if there are more than ${}_N C_2$ model parameters, $\mathbf{G}^{(4)}$ will be singular, and this is why resistivity inversion traditionally relies on regularization and constraint methods.

The ill posedness of resistivity inversion is even worse than is suggested above. We have empirically observed (though not yet proven) that

$$\operatorname{rank}\left(\mathbf{P}\right) \leq {}_N C_2 - N = \frac{1}{2} N\left(N-3\right); \tag{4.9}$$

hence

$$\operatorname{rank}\left(\mathbf{G}^{(4)}\right) \leq \frac{1}{2} N\left(N-3\right) \tag{4.10}$$

For a 20-electrode array, this means that the *maximum attainable rank* of $\mathbf{G}^{(4)}$ is 170. The idea of a maximum attainable rank becomes important in the following sections, where we pose several new experimental design objective functions.

The algebraic perspective outlined above suggests that a more thorough consideration of the three quadrupole configurations (internal, external and straddling) be made. For example, can experiments be designed with only one configuration type, usefully reducing experimental space and expediting design? Two pieces of evidence are considered: the cumulative sensitivity of all electrode configurations of similar type (using a homogeneous earth) and the rank of the associated picker matrix. Cumulative sensitivities were calculated by

$$\sum_{i \in \Omega_{type}} \sum_{k=1}^{K} \left| P_{ik} G_{kj}^{(2)} \right|, \tag{4.11}$$

where $K = {}_N C_2$ and Ω_{type} is the set of all row-indices of $\mathbf{P}$ whose associated quadrupole configurations are of one type (internal, external or straddling). The associated picker matrix is the submatrix P_{ik} for all $i \in \Omega_{type}$. For simplicity, these submatrices are denoted $P^{(I)}$, $P^{(E)}$ and $P^{(S)}$. The rank of a picker submatrix is important because, as explained above, the rank of $\mathbf{G}^{(4)}$ cannot exceed the rank of the picker submatrix,

so picker submatrices should ideally have the maximum attainable rank, $N(N-3)/2$. In other words, the rank of a picker submatrix governs the rank of the Jacobian matrix and thereby the attainable resolution of the inverse problem. Figure 4-2 shows the cumulative sensitivities and ranks of $\mathbf{P}^{(I)}$, $\mathbf{P}^{(E)}$ and $\mathbf{P}^{(S)}$. Based on this figure, it appears that the internal- and straddling-type configurations are considerably superior to the traditional external-type, which is the most widely used quadrupole configuration in resistivity tomography. This would suggest that significant improvements in data quality can be realized simply by using configuration types other than the external type. This hypothesis is explored more fully in following sections.

4.4 Comparing the Three Quadrupole Configurations

Three quadrupole data station types were identified above, internal, external and straddling (see Figure 4-1). This investigation explored the hypothesis that the external-type configuration would give rise to experiments having the poorest data quality and that the internal- and straddling-type configurations would provide relatively superior data quality. A Monte Carlo approach was adopted wherein *random* experiments consisting of configurations of only one type were generated and their data inverted. Fifty realizations were created for each configuration type to provide a statistical measure of the expected data quality. The model rms error between the true and inverted models was used as a quality metric. Experiments from 28 to140 data stations were investigated to derive *performance curves* as a function of observation number. All synthetic data were noiseless for these exercises.

It was speculated that the performance of the three configuration types would behave asymptotically as the number of observations increased, indicating a law of diminishing returns. Additionally, it was also speculated that the performances for the internal- and straddling-type configurations would be similar to one another and superior to the external-type performance curve. All inversions were run for exactly 20 iterations to expedite the investigation and to ensure the consistency of comparison.

The discretized model in Figure 4-4 was used for these comparisons. A background of 100 Ωm was used with two embedded anomalies a conductive anomaly (A) of 20 Ωm and a resistive anomaly (B) of 500 Ωm. Ten borehole electrodes were deployed from the surface to a depth of 9 meters at equispaced 1m intervals to query the target. A 26×16 irregular mesh was used (including boundary condition cells), with cell sizes increasing proportionally as their distance from the borehole array.

The results are synopsized in Figure 4-5. As predicted, random experiments comprising either solely internal or straddling quadrupoles outperform random ones using the external configuration, particularly when more than 28 observations are used. For fewer than 28 observations, external-configuration experiments tend to outperform internal-configuration ones but cannot be said to outperform straddling-configuration ones, which have a nontrivial probability of producing very small relative model rms values. Also as expected, the performance of experiments comprising internal and straddling configurations are comparable, as predicted.

Figure 4-2 helps to explain these performance curves. As shown in that figure, the maximum attainable rank of a Jacobian matrix comprising only external-type quadrupoles was 28 (for 10 electrodes). The MAR indicates the maximum number of nonzero singular values the Jacobian can have and thus relates to the resolution capabilities of the experiment because the trace of the resolution matrix equals the rank of the Jacobian f(Lanczos, 1956). Therefore, no more than 28 model parameters could be uniquely resoved by external-type experiments, no matter how large they were permitted to be. This explains why the performance curve for the external configuration is essentially independent of the number of observations; no matter how many observations there are (from 28 to 140), no more than 28 model parameters will ever be uniquely resolvable, and the flatness the performance curve illustrates this fact. In contrast, internal-type and straddling-type experiments could resolve as many as 34 and 35 model parameters, respectively. Therefore, the resolving power of external and straddling quadrupole surveys is greater. This also explains why their performance curves decrease and then asymptote (once their resolution limits are met) over the interval from 28 to 140 observations.

The performance curves for internal and straddling configurations ramp down quickly, asymptoting around 50 to 60 observations and remaining fairly constant thereafter, as additional observations are added. This is taken as evidence that experimentation does obey a law of diminishing returns; that is, as an experiment increases in size, ever-larger numbers of observations must be added to usefully improve data quality (as measured by model rms error). This is a point of common sense, but it establishes the compelling prospect that *compact experiments can be deliberately designed, whose performance is comparable with much larger random experiments*.

The possibility of collecting small, information-dense data sets, rather than large, information-sparse ones is in itself sufficient motivation for the serious geophysical inversion practitioner to systemically design experiments whenever possible. It is noteworthy that the Pseudosection survey (one of the most common quadrupole geoelectrical surveys in use today) consists solely of external-configuration data stations. In light of the preceding results, it should be clear that the Pseudosection survey is inadequate to produce robust data and its use for data collection and inversion should be avoided henceforth.

4.5 Random Experiments Versus Designed Experiments

One way of assessing whether our experimental design methodology produces superior data quality is to compare designed experiments with randomly generated ones. In this section, optimal experiments were designed based on a homogeneous earth model, using the design strategies described in Chapter 2. A series of trials were executed for random and designed experiments to quantify their relative performances as a function of the number of observations used. As a result of preceding investigations, designed experiments used only straddling-type electrode configuration. This reduced the search space and thereby expedited design times.

Experiments using from 28 to 140 observations were designed or randomly generated, and noiseless synthetic data were simulated for each survey, querying the heterogeneous model in Figure 4-4. Each synthetic data set was then inverted to produce an inversion model. To gather statistics, twenty-five random experiments were realized for each number of observations. 'Standard' experiments of 28 and 140 observations were also considered, as these were expected to outperform random experiments. The 'standard' experiment of 28 observations was the *Pseudosection survey* and the one of 140 was the *ERL survey* (see the Appendices C and D for details on these two surveys).

Once again, a law of diminishing returns was hypothesized. That is, as the number of observations increased, the difference between the qualities of designed and random experiments was expected to decrease. The quality metric adopted for comparing the relative performances of all experiments was the model rms error.

Figure 4-6 graphs the experiment performances (model rms errors versus number of observations). A scatter plot of the final model rms errors are shown for all random experiments, as well as the designed experiment performance curve and the performance for two standardized experiments. The prime feature of Figure 4-6 is that designed experiments produced categorically lower model rms errors (higher data quality) than the expected performance of comparably sized random or standard experiments.

Notice in Figure 4-6 that, occasionally, a random experiment produced a smaller model rmse than the designed experiment of the same size. One possible explanation has to do with our greedy sequential OED algorithms. These algorithms are intentionally crafted to be *expeditious*, and they consequently produce experiments that cannot be guaranteed to be *globally* optimal. By choosing to minimize our design objective function through a sequence of *locally optimal*, rather than globally optimal, observations, design times can be substantially reduced. However, the trade-off is that a set of locally optimal observations does not necessarily 'add up' to a globally optimal experiment. Globally optimal experiments may exist, but they can only be found by using the much slower global search algorithms discussed previously. A truly *globally*

optimal experiment would hypothetically give rise to data whose inversion would produce a vanishingly small model rms error (for noiseless data). The probability that a random experiment could produce a lower model rms error than this hypothetical experiment would literally be zero. The issue is the amount of time needed to discover this globally optimal experiment. Rather than expending hours or days seeking global optimality, why not design experiments that produce excellent, though suboptimal, results for a fraction of the computational expense? This is the so-called *Pareto Principle*, which asserts that '80% of the consequences stem from 20% of the causes'. In this context, we assume we can achieve something like 80% experiment optimality for 20% of the computational effort (though, in fact, it is more like 1% or 0.1% of the computational effort).

An alternate explanation for the occasional random experiment outperforming a designed one in Figure 4-6 is that the designed experiments were crafted based on a *homogeneous* earth model. As was discussed in Chapter 2, the optimality of a designed experiment, with respect to one model, does not guarantee its optimality with respect to another, distant model.

It is also evident from Figure 4-6 that this experimental design methodology obeys a law of diminishing returns. The designed experiment performance curve asymptotes within the first 40 observations or so, and the expected performance curve for random experiments continues to slowly decrease, reducing the difference between the performances of designed and random experiments. Therefore, the clear benefit of OED is that it produces high-quality, compact, smart data sets with dense complementary information. It is also inferred from Figure 4-6 that there is a critical number of observations beyond which experimental design is no longer worthwhile, because it becomes increasingly likely that a random experiment could perform as well as a designed one. Consequently, if time and money are of material importance, it behooves the exploration geophysicist to design their experiments rather than relying on standardized survey geometries or random surveys. The small computational expense our greedy design algorithms require only strengthens that argument.

A ‘piano role’ cartoon depicts the experiment designed with 28 observations (Figure 4-7). There is no discernible pattern in this set of 28 data stations, which is unfortunate because if there were it would have been possible to write down a simple recursive formula for creating ‘optimal’ designs using any number of electrodes. As it is, it appears that we must resort to computational design methods.

Figure 4-8 shows the CPU time for experimental designs of increasing numbers of observations. All computations were carried out on an HP laptop with dual 2GHz processors and 2GB RAM. The key point is that our novel sequential OED methodology executes in a matter of seconds. This is in contrast to traditional global search methods that must evaluate their objective functions literally hundreds or thousands more times, and which thus require significantly more CPU time. Thus, our sequential methodology provides a significant contribution to optimal experimental design theory and significantly closes the gap between the theory and practical application of OED.

As pointed out above, the designed experiments in this study were created assuming a homogeneous half-space. In this respect, we should refrain from calling them *optimal* because they have not been tailored to the particular target they query. However, the fact that experiments designed for a homogeneous half-space and employed on a heterogeneous target produce such superior results is testimony to how much room for improvement there is in executing geoelectrical surveys.

4.6 Two-Stage, Adaptive Optimized Experimental Design

This study investigated the utility of performing an initial survey/inversion followed by an experimental design phase and a second inversion. In this respect, the procedure can be thought of as two-stage: Stage One collects and inverts a standardized data set, producing a working model or image of the earth; Stage Two designs an experiment tailored for the working model and then collects and inverts a second data set, using the preliminary model as an initial guess for the inversion algorithm. A

Pseudosection data set was initially generated and inverted to produce the working model. Experiments were then tailored for that model. Data for the new experiments were then generated and inverted, and the resulting image misfit was measured to evaluate performance. Experiments of 28 and 140 observations were designed and compared with the Pseudosection and ERL surveys. To determine whether the performance of designed experiments showed *bona fide* improvement a Monte Carlo study was conducted using 100 randomly generated 28-observation experiments for Stage Two. A performance histogram was then created, offering a probabilistic perspective on the relative performance of designed experiments.

The results are shown in Figure 4-9 (the inversion result for the 140-observation ERL survey is also shown for comparison). This simple form of adaptive OED performs very well. Panel (**a**) shows that the 28-observation designed experiment reduced model rms error by over an order of magnitude compared with the initial survey. Importantly, the designed survey captured the correct shape of the 500 Ωm resistive anomaly, which was misrepresented by the Pseudosection image. Compared with the 28-observation examples in Figure 4-6 (indeed, compared with any size experiment in Figure 4-6) this small, adapted experiment performed exceptionally. Panel (**b**) shows the result for a 140-observation designed experiment. As expected by the law of diminishing returns, the model rms error for this designed experiment is not significantly better than its smaller cousin in Panel (**a**), suggesting that the smaller experiment would suffice in view of financial considerations. Both designed experiments outperform the ERL survey shown at right in Panel (**b**).

The results of this investigation are very exciting: small, adapted experiments can significantly improve data quality. In Panel (**a**), only 56 observations were made (28 each for the Pseudosection and designed surveys). The ERL survey, by comparison, used 140 observations without producing a better image. The total computation time for the two inversions and one design phase was 125 seconds for the 28/28 trial (60 seconds for each inversion and 5 seconds for the design phase) and 160 seconds for the 28/140 trial (60 seconds for each inversion and 40 seconds for the design phase). The conclusion is obvious: rather than collect a large data set, which poses a financial

burden, this simple, two-stage adaptive technique can be employed to produce superior imaging at a fraction of the cost.

To determine whether the 28-observation design was statistically superior or just a product of chance, it was compared with 100 random experiments of 28 observations. As with the designed experiment, the inversions of all random survey data were initialized with the same Pseudosection image. The histogram in Figure 4-10 shows the outcome; the adaptively designed experiment outperformed 95% of all random surveys, demonstrating that the AOED procedure produced experiments whose superiority is statistically significant.

One reason that the designed experiments performed so well is that the initial model plugged into the 2nd stage inversion is the Pseudosection image. In previous inversions, where no prior information was available, the starting model was a homogeneous half-space. In this case, the inversion algorithm did not have far to travel through model space, nor was it in jeopardy of falling into local minima or diverging. This statement is true of *designed experiments* but not the random ones. The Monte Carlo performance histogram (Figure 4-10) was bimodal, with the higher mode peaking around a model rms error of 30 Ωm. This indicates that some fraction of the random experiments actually caused divergence. It is concluded that optimized experiments perform so well because they were deliberately, systematically *designed* to perform well. Even though the Pseudosection earth model was incorrect, it was close enough to the true model that the tailored surveys performed outstandingly.

4.7 In-Line, Adaptive Optimized Experimental Design

Of the trials performed in this chapter, this was the most ambitious. The prospect of performing *real-time* or *in-line experimental* design was examined. Complete details of the methodology are outlined in Chapter 2, but a brief recap is offered here. This technique operates by sequentially executing experimental design *in line* with data inversion; as the earth model comes into focus, so to speak, additional observations

are added one at a time and inverted under the supposition that they have been optimized to 'steer' the inversion toward the true model. The method is thus seen to be an integrated super-algorithm that cyclically alternates between a model-update phase (through partial inversion) and an experimental-design phase. Each time the model vector is updated (via one or two iterations of a NLLS inversion), an additional observation is prescribed for the base experiment, using our design methodology; the observation is then collected and a new model update is made, completing one cycle. The super-algorithm continues this process until one or more stopping criteria are met.

Using this integrated design strategy, experiments of 28 and 140 data stations were designed. Standard Pseudosection and ERL surveys were also considered as a means of comparison. As previously, the experiment quality metric used to assess relative survey performance was the model rms error. During the inversion step, the nonlinear inversion algorithm was allowed two iterations, increasing the probability that the working model had incorporated the information provided by each new observation. Ten initial data stations were used to seed the algorithm, which were specified by designing a 10-observation experiment optimized for a homogenous earth.
It was initially thought that this technique would outperform all techniques investigated in this chapter, both in terms of model rms error and possibly in overall computation time. But, as will be seen, this supposition proved to be incorrect.

Figure 4-11 shows the inversion images for Trials 1 and 2, where 28- and 140-observation designs were attempted, respectively. Panel (**a**) shows the 28-observation results and Panel (**b**) shows the 140-observation results. The standard surveys (Pseudosection and ERL surveys) were executed as though they had been sequentially designed (like the designed surveys), to account for the effects of the serialized inversion. This ensured that the comparison between designed and standard experiments would be unbiased by the inversion method.

Neither designed experiment performed particularly well with regard to comparably sized standard surveys. The in-line AOED experiments shown here were not only unremarkable compared with standardized surveys, but they performed more poorly

than – or, at best, comparably to – the experimental designs in Figure 4-6, where a homogeneous earth model was used as the design target. Their performance was also significantly inferior to the two-stage designed experiments above (see Figure 4-9). Moreover, the need to recalculate the Jacobian in the design step, after each observation is added, significantly increased the computational burden of this method. The 28-observation trial executed in a little more than 600 seconds, and the 140-observation trial took more than 3000 seconds, compared with 125 and 160 seconds, respectively, for the Two-Stage approach above.

Figure 4-12 offers a different perspective. It shows the performance curves for standard and designed experiments as a function of observation number, as the experiments were being built. While both designed and standard surveys fared poorly at 28 observations, the designed experiment improved rapidly thereafter, asymptoting by ~50 observations. The standard survey improved a bit but leveled off until ~75 observations where it finally improved to a level comparable with the designed survey. An explanation for the rapid and drastic improvement of the designed experiment around 30 observations is that the number of experimental observations has approached the maximum attainable rank ($_{10}C_2 - 10 = 35$) for this problem. This is a matter of course, because the objective function deliberately favors additional observations whose information is maximally complementary with respect to the base experiment. Thus, it is conjectured that in-line adaptive OED might be expedited by simply specifying a number of observations equal to or slightly greater than the maximum attainable rank of the problem.

There are at least three significant drawbacks to in-line adaptive OED as it has been demonstrated here. First, it is computationally demanding. Each time the design step is executed, it requires recalculation of the sensitivity matrix of all permitted data stations, **G**, because the earth model has been updated. As the size of the inverse problem increases, significantly more computational expense will be required to aggregate the designed experiment. Second, the design step depends on the current working model, which in turn depends on the inversion algorithm. If the inversion begins to diverge, the design step will add observations optimized for the diverging

model, possibly further destabilizing the inversion and setting up a negative feedback. Third, it is confounding to determine an appropriate regularization schedule. The Levenberg-Marquardt algorithm is quite effective for traditional inverse problems where the data set remains unchanged from start to finish, but there is an added degree of complexity in trying to tune λ as the data set itself changes. Though it is not discussed in this document, a great deal of time was spent tuning the whole in-line procedure to produce any kind of useful result.

4.8 Small Batch, Adaptive Optimal Experimental Design

Given the results of In-Line AOED section above, a slight modification to the in-line algorithm was implemented to try to improve performance. Steps ***3-iii*** and ***3-v*** in Flowchart 2-3 were altered as follows:

3-iii'. Perform *between three and seven iterations* of a nonlinear least squares inversion, using some type of regularization to ensure stability. If at any time between the third and seventh iteration the data rms error increases, terminate the inversion, update the model to $\mathbf{m}_{n+1}$ and proceed to Step ***3-iv***.

3-v'. Use Objective Function #1 or #2 (depending on whether n is greater than the maximum attainable rank) to determine the next η observation to be added to the experiment.

■

The difference between Step *3-iii'* and ***3-iii*** is that more iterations were allowed before the updated model was passed on, hopefully allowing the model to approach closer to the true model before the design step was executed. The difference between *3-v'* and ***3-v*** was that η observations, rather than only one, were added to the experiment before the inversion step was revisited. The purpose for *3-v'* is similar to that for *3-iii'*, namely to increase improvements to the model before it is passed to the next step of the super-algorithm. This modified algorithm might be called *small batch adaptive AOED* because it builds the experiment in groups or batches of observations, rather

than by single additions. The ultimate purpose of these modifications was twofold: 1) reduce the computational expense associated with the design step (which requires re-calculation of the sensitivity matrix) and 2) improve the final model rms error.

An example trial was executed that built an optimal experiment in groups of seven observation ($\eta = 7$), starting with 7 initial or seed observations, to a total of 56 observations. As specified in Step *3-iii'*, the nonlinear inversion algorithm was permitted between 3 and 7 iterations before passing control to the design step. For comparison, the ERL and Pseudosection surveys were also treated as being designed in a batch-built manner. As before, the reason for this was to remove the effects of the serial inversion in the comparison between standard and designed experiments.

Figure 4-13 displays the result for the 56-observation AOED trial. By the terminus of this method, it had achieved a model rms error comparable to the asymptote seen in Figure 4-12. The total computational time for this was 261 seconds, less than half the time required to execute the In-Line AOED approach for 28 observation above. Thus, one of the two stated purposes of this exercise has been met: the computational burden has been reduced without sacrificing image quality (as measured by the model rms error). A different viewpoint on the comparison between Batch and In-Line AOED is provided by Figure 4-14, which shows the performance curves of both methods as a function of the number of observations. The shapes of the two curves are similar, but the batch method reduced model rms error earlier, suggesting that this technique may be more robust, especially when small numbers of observations are used.

The standard Pseudosection and ERL surveys were also executed in this batch approach to ascertain how they behaved in comparison with the designed experiment. For the sake of brevity, the inversion images are not shown, but the final model rms errors were 68.09 Ωm and 4.55 Ωm, respectively, for the Pseudosection and ERL surveys. Clearly, the batch-designed experiment of 56 observations produced a final model rms error comparable to that for the ERL survey, but the ERL survey used 140 observations, so we conclude that this batch method does indeed produce smart, condensed data sets.

It is concluded from this exercise that small batch AOED is an improvement over the In-Line technique discussed previously. While the final rms errors were comparable, the computational burden was significantly reduced by more than a factor of two in the batch method. Additionally, it is surmised that the batch method is less prone to divergence than the In-Line method, which is of considerable concern for these types of techniques.

At the end of the day, though the batch method has performed well, we would not necessarily recommend it for real-world application. The procedure incorporates data collection, experimental design, and inversion into one super-algorithm that can ostensibly be deployed to great advantage in the field. After all, if the geophysical explorer could collect and invert an optimal data set in real-time, creating an optimal image right in the field, it would represent a significant advance in exploration methods. However, there are a number of algorithm hyperparameters (the regularization schedule, the number of inversion iterations allowed per batch, and the batch size itself) that must be tuned in order to get good results. If these hyperparameters are not properly tuned, the entire integrated procedure can diverge or converge to local minima, resulting in wasted time and money. In short, there is no fail-safe way to prevent these last two AOED procedures from going awry. Nonetheless, additional research in this area is needed.

4.9 Noisy Data and Experimental Design

Having now examined the effects of optimal experimental design with respect to noiseless data scenarios for borehole DC resistivity, it is natural to turn our attention to its application in realistic, noisy-data situations.

In this section, the Two-Stage Adaptive OED method discussed above was implemented in the presence of noisy data. A stage-one model, shown in Figure 4-15, which came from a Pseudosection survey of 28 noisy observations over the heterogeneous target in Figure 4-4, was used as the experimental design earth model. Stage-two experiments of 140 observations were designed using ‘All’ and ‘Straddling’ quad-

rupoles, and these were compared with the ERL survey, which was also used as a second stage survey. A series of Monte Carlo simulations was performed where 50 realizations of 3% Gaussian noise were added to the data sets for each survey and the contaminated data were then inverted.

The results from these Monte Carlo simulations were a bit lackluster and afforded an instructive viewpoint on the limitations of our design methodology. In light of the results reported above on the Two-Stage AOED method, it was expected that these noisy-data studies would further solidify evidence that the 2-stage adaptive design strategy generally produces smarter data sets, with superior inversion modeling characteristics. Figure 4-16 shows the average second stage models for the 'ERL', 'All' and 'Straddling' experiments of 140 observations (Panels a - c). Visually, the three results are nearly indistinguishable, though it appears that the average 'Straddling' image has better identified the magnitude of the resistive anomaly. The right hand panels in this figure (Panels d – f) show 'Monte Carlo Parameter Uncertainties', which are evaluated by the expression

$$E_{ij} = 100\sqrt{\frac{1}{N}\sum_{k=1}^{50}\left(\frac{\rho_{ij} - \hat{\rho}_{ij}^{(k)}}{\rho_{ij}}\right)^2} \tag{4.12}$$

where $\hat{\rho}_{ij}^{(k)}$ is the inversion-model resistivity of the ij^{th} cell for the k^{th} noise realization and ρ_{ij} is the true resistivity of the ij^{th} cell. The parameter uncertainties are therefore like standard deviations about the true model. Curiously, though the average 'ERL' model looks to be the worst of the three images, its largest parameter uncertainties (Panel d) are concentrated directly around the resistive anomaly. Whereas the 'All' and 'Straddling' designs both create larger uncertainties (Panels e and f) elsewhere in the target window but smaller uncertainties around the resistive anomaly. This derives from the fact that the design methodology qualitatively seeks to equalize the resolving power of an experiment throughout the target window. As was seen in the previous chapter on 1D resistivity, depending on the earth model, the design method may sacrifice the sensitivity of some well-resolved parameters to increase the sensitivity of other less well-resolved ones. Clearly, the most poorly resolved parameters

for this particular earth model are the ones with large resistivity values, and it is plain from Figure 4-16.d-f that the designed experiments have 'deliberately' and somewhat successfully reduced their uncertainty.

It is difficult to determine from Figure 4-16 whether the 2nd stage designed experiments have truly produced smarter data compared with the 2nd stage ERL survey. An alternate representation of these Monte Carlo simulations is shown in Figure 4-17. These 'performance curves' are just the normalized cumulative frequency of modeling errors produced by each experiment in the Monte Carlo exercises. There are two obvious features in this figure. First, the distributions of modeling errors for designed surveys are *generally* to the left of those for the ERL survey, meaning that the designed surveys generally produce smaller errors than the ERL survey. But – and this is the second feature – a portion of the 'performance curves' for the designed surveys also fall significantly to the *right* of the ERL curve, corresponding to large model % rms errors. These two facts signify that the probability distributions of modeling errors for the designed surveys are left-skewed and peak at smaller values than the probability distribution for the ERL survey; but the probability distributions for the designed experiments also have long right-sided tails that extend significantly beyond the right-hand tail for the ERL survey. This means the designed experiments *generally* produce smarter data than the ERL survey but at the apparent risk of occasionally actually producing 'dumber' data.

An alternate analysis looks at the comparative probabilities that one experiment will outperform another. To do this, we evaluated all permutations of the differences in Monte Carlo modeling errors between two experiments and determined the average number of times the difference was positive or negative. Mathematically, this is expressed as

$$p = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_{ij} \,, \tag{4.13}$$

where

$$\delta_{ij} = \begin{cases} 1 \text{ if } E_i^{(A)} - E_j^{(B)} < 0 \\ 0 \text{ otherwise} \end{cases} \tag{4.14}$$

and $E_i^{(A)}$ and $E_i^{(B)}$ are the i^{th} and j^{th} Monte Carlo model % rms errors for Experiments A and B, respectively. p approximates the probability that Experiment A will produce a smaller model % rms error than Experiment B; and $1 - p$ approximates the complementary probability that Experiment B will outperform Experiment A. These probabilities do not indicate *by how much* one experiment is expected to outperform another, but this is simply determined by taking the difference between their average model % rms errors:

$$\Delta E^{(AB)} = \frac{1}{m}\sum_{i=1}^{m} E_i^{(A)} - \frac{1}{m}\sum_{j=1}^{m} E_j^{(B)} \tag{4.15}$$

Table 4-1 compiles the comparative probabilities of producing smaller modeling error and average differences in modeling error between the three experiments. Each column in Panel (a) indicates the probability that the survey in the column header will produce smaller modeling errors than the competing surveys in the row headers. So, for example, the 2nd stage 'All' design has a 42% chance of producing a smaller model % rms error than the 2nd stage 'Straddling' design. Inspection of the comparative probabilities in Panel (a) indicates that both designed experiments are more than 50% likely to produce smaller model errors than the ERL survey. Had these probabilities been 50% or less, it would have signified that the designed surveys were statistically no better (or even worse) for producing high quality data than the ERL survey. However, neither the 'All' nor 'Straddling' designs produced comparative probabilities with the ERL survey that were vastly greater than 50%, being only 61% and 72%, respectively. Thus, while the designed surveys clearly have a statistically proven advantage over the ERL survey, the advantage is arguably small, begging the question whether it is worthwhile to undertake design at all in this case. When one further considers that the average differences in model % rms error between the designed surveys and the ERL survey was at most only 2.2%, it becomes harder to claim that the statistical advantage that the designed surveys apparently have is anything but

a paper tiger. Still, one wonders whether, through some modification of the design objective function, the long right-sided tails seen in Figure 4-17 could be eradicated, allowing us to unequivocally proclaim that designed experiments always produce smarter data. It is because of these tails that the average model % rms errors for all surveys are nearly equal and that the probabilities of producing smaller modeling errors are not larger for the designed experiments.

It is worth briefly noting that Figure 4-16 and Figure 4-17 and Table 4-1 all indicate that the 'Straddling' design outperformed the 'All' design, as was noted above. This fact is carefully scrutinized in the next chapter.

In closing this section, an outstanding question remains: why did the noiseless 2-stage AOED example above (Figure 4-9) produce over an order of magnitude improvement in overall modeling accuracy (measured by model % rms error), as compared with the ERL survey, while the noisy Monte Carlo cases examined here showed almost negligible comparative improvements? There are two possible explanations, which are not necessarily mutually exclusive.

The first explanation concerns the model regularization scheme employed by the inversion. All inversions carried out in this chapter used an L^1-smoothness constraint on the spatial gradient of the model, and the strength of this constraint has been controlled by a Lagrange trade-off parameter (per Equation (4.2)). These borehole resistivity inversions are inherently ill posed because there are more model parameters than the maximum attainable rank of the Jacobian matrix; hence, some form of model regularization is a necessity. When the data are noiseless, the Lagrange multiplier must be large enough to promote inversion stability but not so large as to overly bias the inversion model. In a manner of speaking, we can allow the Lagrange multiplier to become small (in the Levenberg-Marquardt sense) as the inversion converges to let the data 'speak for themselves', because there is no error in the data to derail the inversion process.

However, when the data are noisy, the smoothness constraint performs not only as a stabilizing functional but also as a sort of noise filter. Ill-posedness, by definition, means that one or more model parameters are practically irresolvable and that the in-

verse problem gives rise to nonunique solutions. A poorly resolved parameter is very insensitive to the data, which means, from a forward modeling point of view, that it can be perturbed by a large amount without significantly altering the predicted data. Turning this idea around, it follows that if there is noise in the observed data, poorly resolved parameters could take on extravagantly large or small values in an attempt to minimize the data rmse. It is for this reason that model regularization is especially critical to ill posed inverse problems faced with noisy data. The only way to prevent poorly resolved parameters from 'blowing up' or vanishing in the presence of noisy data is to rely more heavily on the regularization constraint, which in this case enforces an assumption of local smoothness. It is in this sense that we argue that the smoothness constraint must also act as a data noise filter.

In short, out of necessity the Lagrange multiplier in Equation (4.2) was kept significantly larger (two to three orders of magnitude, in fact) for these noisy data trials than for the preceding noiseless trials. However, regularization methods bias the final inversion result, so if the Lagrange trade-off parameter between data rmse and model smoothness is kept large, the final inversion image is biased toward smoothness more than toward a small data rmse.

The reason designed surveys did not perform more spectacularly in the presence of data noise now becomes clear. The design methodology is concerned with the information provided by the *experiment*, which can be approached through the singular values of the Jacobian matrix, **G**. But if the inversion has been heavily biased (out of necessity) toward smooth models, to protect against the ill effects of data noise, this information is in the *constraint matrix* (which is represented by the second term in Equation (4.2)), not in **G**. The design methodology does not address the model regularization term and therefore does not produce surveys with particularly outstanding data quality as compared with standard surveys.

In light of the foregoing discussion, a second possible explanation for the lackluster performance of designed surveys in the presence of data noise may have to do with the design objectives. Recall from Chapter 2 that the design technique can be broken into two stages, depending on whether the base experiment contains more or less ob-

servations than the maximum attainable rank of the Jacobian matrix (see Flowchart 2-1 and Flowchart 2-2 for details). The second stage deals with the addition of observations after the number of observations in the base experiment has exceeded the maximum attainable rank (Flowchart 2-2). The objective in this second stage is to increase the magnitude of the smallest singular values in the Jacobian of the base experiment. Let us take a brief detour to consider the actual inversion step that minimizes Θ in Equation (4.2) with respect to $\Delta\mathbf{m}$:

$$\Delta\mathbf{m} = \left(\mathbf{G}^T\mathbf{G} + \lambda\mathbf{L}^T\mathbf{L}\right)^{-1}\left(\mathbf{G}^T\Delta\mathbf{d} - \mathbf{L}^T\mathbf{L}\mathbf{m}\right) \tag{4.16}$$

where $\mathbf{G}$ is the Jacobian matrix, $\mathbf{L}$ is the linearization of the L1-smoothness operator, $\Delta\mathbf{d}$ is the difference between observed and predicted data, $\Delta\mathbf{m}$ is the model update vector, $\mathbf{m}$ is the current model vector and λ is the Lagrange trade-off parameter between data rmse and model smoothness. From an inversion standpoint, we are concerned with the eigenvalues inside the matrix inverse in (4.16), and it should be evident that these eigenvalues depend on the magnitude of λ. If λ is very large, the eigenvalues of the inverse matrix are dominated by the eigenvalues of $\mathbf{L}^T\mathbf{L}$; and conversely, if λ is very small, the eigenvalues of $\mathbf{G}^T\mathbf{G}$ dominate. This means, in particular, that the small nonzero eigenvalues of $\mathbf{G}^T\mathbf{G}$ will be 'swamped out' by the eigenvalues of $\mathbf{L}^T\mathbf{L}$ whenever λ is large. And because the 2nd stage of the design algorithm attempts to maximize these small eigenvalues in $\mathbf{G}^T\mathbf{G}$, the effort is essentially wasted, for the eigenvalues of $\lambda\mathbf{L}^T\mathbf{L}$ overpower the small eigenvalues in $\mathbf{G}^T\mathbf{G}$ (Figure 4-18). Hypothetically, the observations in a designed experiment that attempt to maximize the smallest eigenvalues in $\mathbf{G}^T\mathbf{G}$ are rendered useless (uninformative) when the model regularization term is dominant. It might in fact be better to choose additional observations that maximize the *large* eigenvalues in $\mathbf{G}^T\mathbf{G}$. Such observations would at least ostensibly inform the inversion because they would be expected to affect the eigenvalues of $\mathbf{G}^T\mathbf{G}$ above a 'water level' imposed by $\lambda\mathbf{L}^T\mathbf{L}$.

4.10 Integration and Summary

This chapter has examined multiple methods of optimal experimental design applied to a heterogeneous earth model, cast in a pseudo-2D borehole DC resistivity framework. As well, key issues pertaining to fixed-electrode electrical resistivity tomography were explored, from which several useful insights were elucidated that not only inform geophysical experimental design but ostensibly the ERT method itself. Several subtopics were considered throughout this chapter; below is a brief synopsis highlighting the main features and/or results from each of subtopic.

4.10.1 Modeling and Inversion

The single-borehole DC resistivity forward and inverse problems have been addressed using the transmission line network analogy (e.g., Swift, 1971; Madden, 1972; Zhang *et al.*, 1995; Shi, 1998). With only one borehole in which to place survey electrodes, the azimuthal position of resistivity features was irresolvable, allowing the earth model to be simplified to axially symmetric cylindrical coordinates. All inversions implemented model regularization via the L1-norm of the spatial gradient of the model. This permitted inversion models to have sharp contrasts because contrasts are not as heavily penalized by an L1-norm as they would be by an L2-norm smoothness constraint.

4.10.2 Electrical Resistivity Tomography and Experimental Design

Careful examination of the electrical resistivity tomography problem has produced several significant and novel results that are important not only for design but for the ERT problem itself.

- *The maximum attainable rank of the Jacobian matrix for an arbitrary experiment comprising quadrupole data stations is* $_{N}C_{2} - N$*, where N is the total number of survey electrodes.*

It was shown that the maximum attainable rank of the Jacobian matrix for an arbitrary quadrupole experiment is ${}_N C_2 - N$, where N is the total number of survey electrodes. The concept of maximum attainable rank clarifies why geoelectrical inverse problems are often ill posed – even if there are more data than model parameters, if the maximum attainable rank is less than the number of model parameters, the problem is still ill posed. Maximum attainable rank is also pivotal to the experimental design objective functions introduced in this work.

- *The straddling-type electrode configuration is superior for geoelectrical surveys.*

Three quadrupole configuration types were identified: external, internal, and straddling. It was empirically shown that, for 10 survey electrodes, the maximum attainable ranks (MARs) of Jacobians comprising quadrupole observations of only one type were 28, 34 and 35 for the external, internal and straddling configurations, respectively. The MAR for the Jacobian of all possible quadrupole observations is also 35. The MAR is diagnostic of the resolution limitations of an experiment, for it quantifies the maximum number of nonzero eigenvalues the linearized inverse problem can have, and this number is almost always smaller than the number of model cells for which a 2D resistivity inversion attempts to solve. As a rule of thumb, therefore, one should seek to design experiments that maximize the MAR, indicating that straddling-type quadrupole configurations are the best of the three configuration type. The relative magnitude and distribution of cumulative sensitivities of the three configuration types (based on a homogeneous model) correlated with their MARs; to wit, external-type configurations produced the poorest cumulative sensitivity, while the internal and straddling types produced comparable and larger cumulative sensitivities. It was subsequently demonstrated by Monte Carlo exercises that experiments consisting of either straddling or internal configurations do indeed outperform experiments comprising

only internal configurations, especially as the number of observations in the experiment becomes large.

- *The Pseudosection survey produces poor data quality and should perhaps be discontinued.*

  The Pseudosection survey is made only of external-type electrode configurations, and this configuration type was identified as producing the poorest data quality of the three configuration types – it lacks both sensitivity and resolution. This was plainly demonstrated by Monte Carlo exercises that consistently showed the Pseudosection survey performing worse than nearly any other experiment of the same size.

- *The straddling-type configuration can be used exclusively for experimental design, expediting the design algorithm.*

  The MAR of the Jacobian for experiments consisting only of straddling-type electrode configurations was shown to equal the MAR of the Jacobian of all data stations, and the cumulative sensitivity of all straddling-types was the best of the three configuration types. Consequently, it was hypothesized that this configuration type could be used exclusively in the design enterprise, expediting the design algorithm by reducing the number of permitted observations by two thirds. All experimental designs in this paper comprise only this configuration type, and the results summarized below support its exclusive use.

4.10.3 Random Experiments Versus Designed Experiments

A series of Monte Carlo exercise were executed to compare randomly generated experiments of different numbers of observations with designed and standard experiments. The designed experiments were created assuming a homogeneous earth model. All trials used noiseless data.

- *Designed experiments produce superior data sets.*

  Based on model rms errors, it was shown that designed experiments produced statistically significant improvements in data quality as compared with random and standardized surveys, even though the designed experiments were created assuming a homogeneous earth. It is therefore concluded that the methods of OED developed in this work represent a *bona fide* advance in the design of experiments for borehole resistivity.

- *Designed experiments are suboptimal but statistically more likely to produce high-quality data sets than random or standard surveys.*

  Because the methods of experimental design laid out in this research are sequential (greedy), they do not necessarily produce truly optimal experiments. It was seen that there is a small (but nonzero) probability that an arbitrary random experiment could produce a better inversion result than a designed one. The only way to produce truly optimal experiments is to use global search strategies, which are prohibitively expensive for real-world applications. Instead, the *Pareto Principle* is invoked, where it is qualitatively argued that our design methods can achieve ~80% optimality for ~20% of the work.

- *Experimental design is subject to a law of diminishing returns*

  It was shown that geophysical OED obeys a law of diminishing returns. That is, as the number of observations increases, relative improvements in the performance of designed experiments diminish. Consequently, as the number of observations becomes large, the utility of designing an experiment (rather than randomly generating one) diminishes. Therefore, the primary benefit of experimental design methods lies in the fact that they produce compact, smart data sets with high in-

formation density, which is of critical importance from a financial and temporal standpoint. The maximum attainable rank should impinge on where this performance transition occurs. Based on the Monte Carlo exercises, this threshold might be approximately one and a half to two times the MAR. Beyond that point, improvements in designed experiment performance were fairly small.

4.10.4 Two-Stage AOED

This investigation studied the possibility of performing experimental design on an earth model that was created by an initial Pseudosection inversion. It was a first attempt to deliberately *adapt* an experiment to a heterogeneous model, which would then be used as an initial guess in a second inversion. To confirm that this form of adaptive OED actually produced superior results, a comparative Monte Carlo simulation was conducted wherein 100 randomly generated experiments were used as the second stage experiment.

- *Two-stage adaptive OED produced exemplary results, surpassing experimental designs based on homogeneous earth models, and surpassing standard surveys.*

  Two-Stage AOED produced the best results of any investigation in this chapter. The model rms error for the 28/28 trial (28 observations in the Pseudosection survey and 28 in the designed experiment) was well over an order of magnitude smaller than that for the Pseudosection survey itself; and the result for the 28/140 trial (28 observations for the Pseudosection survey and 140 for the designed experiment) was twice as good as that for the 28/28 trial. That the 28/140 result was only twice as good as the 28/28 trial further demonstrates the law of diminishing returns. Based on these results, it was speculated that a reasonably small second-stage experiment (no bigger than the MAR) could be designed and executed that would produce superior data quality and inversion results. In both cases, total computation did not exceed 3 minutes (for two inversions and one design). The

ERL survey was also used as a second-stage experiment but did not produce results significantly better than if it had started with a homogeneous initial guess.

Additionally, the Monte Carlo exercises showed that the 28/28 designed experiment produced a result better than 95% of 28/28 trials conducted with a randomly generated second stage experiment. This establishes beyond question that the 2-stage AOED methodology is creating significantly superior data *by design* and is not a statistical anomaly. The primary benefit of this technique is that, for reasonably small additional cost, model accuracy can be considerably improved.

4.10.5 In-Line and Small Batch AOED

These investigations explored the possibility of performing experimental design in real-time or *in line* with data collection and inversion, creating a geophysical technique that would allow the user to collect and invert a high-quality data sets in the field. Two approaches were considered. Both started with a small initial experiment and one adaptively built the experiment one observation at a time while the other adaptively built the experiment in small batches.

- *In-line experimental design does not outperform 2-stage design and requires more CPU time.*

  Despite the potential benefits that an integrated experimental design approach seemingly offers geophysical exploration, our investigations showed that it produces a final earth image inferior to the 2-stage method. Moreover, the in-line technique required more than twice as much CPU time. This method does demonstrate the law of diminishing returns, so some decrease in computational expense can be made by prudently choosing the number of observations. Even so, the method still requires more CPU time than the 2-stage method, and without producing a superior image.

- *Small Batch OED performed comparably with in-line design, using a smaller number of observations but still did not outperform 2-stage design. While small batch AOED required less CPU time in total, it still required more than the 2-stage AOED method.*

  The small batch approach produced comparable model rms error to the in-line AOED method. To its credit, this technique did so at less computational expense – especially with regard to the number of times the Jacobian matrix must be recalculated. Additionally, it was argued that a batch approach would be less prone to inversion divergence or to local minima. At the end of the day however, this method still has higher computational overhead than the 2-stage method, and like the in-line approach, it did not produce a superior image.

- *Both in-line and small batch AOED require considerable hyperparameter tuning, making them risky to deploy in the field.*

  With some effort, it may be possible to tune both the in-line and small batch AOED algorithms to perform better than has been shown in this research. But the need for such careful tuning counts against these methods. Fully integrated and automated geophysical techniques like in-line and small batch AOED should be usable 'straight out of the box' so to speak.

  Because the experimental design phase of these techniques is dependent on the current model vector, which is in turn dependent on the inversion algorithm, there is an unaddressed possibility that the entire procedure can catastrophically diverge, with no safety mechanism to counteract divergence. This translates into the potential for considerable wasted time and effort, which runs counter to the spirit of optimal experimental design.

4.10.6 Noisy Data and Experimental Design

A series of Monte Carlo exercises was undertaken to evaluate the utility of a Two-Stage AOED approach in the presence of 3% Gaussian noise added to the data. Two second-stage experiments were designed using our sequential design algorithm on a Stage-One model produced by the Pseudosection survey. The performances of the designed surveys were compared with the performance of the ERL survey, which was also employed as a second stage survey.

While the designed 2nd stage experiments were statistically shown to outperform the ERL survey, the margin of improvement they offered was considerably smaller than was seen in the noiseless case discussed above. In point of fact, the designed surveys were generally more likely to produce smaller modeling errors in the presence of noisy data, but the distribution of their modeling errors also had a significant right-sided tail, such that they occasionally produced significantly larger modeling errors. Thus, on average, the designed and standard surveys all appeared to have nearly identical performances.

Two reasons were posited for the significantly poorer performance of 2nd stage designed experiments in the presence of data noise (as opposed to their performances in the noiseless data case). First, it was necessary to rely more heavily on the smoothness constraint in the presence of data noise to prevent poorly resolved parameters from taking on physically unreasonable values. This undermined the utility of designed experiments, which were tailored to maximize the information that the *experiments* provided on the earth model, not the information *imposed* by a model smoothness constraint. Second, as a consequence of the extra reliance placed on the model regularization scheme, the experimental design objectives were countermanded, as they were geared toward maximizing the small eigenvalues of the Jacobian matrix. All but the largest eigenvalues were 'swamped out' by the eigenvalues of the constraint matrix, rendering ineffective the observations in the experimental design that were specifically chosen to maximize the small eigenvalues of the Jacobian. It was hypothesized that, in such ill posed cases as those examined here, it might be better to design experiments whose observations preferentially increase the *largest*

eigenvalues of the Jacobian, above the water level of the eigenspectrum of the constraint matrix. This hypothesis remains to be investigated.

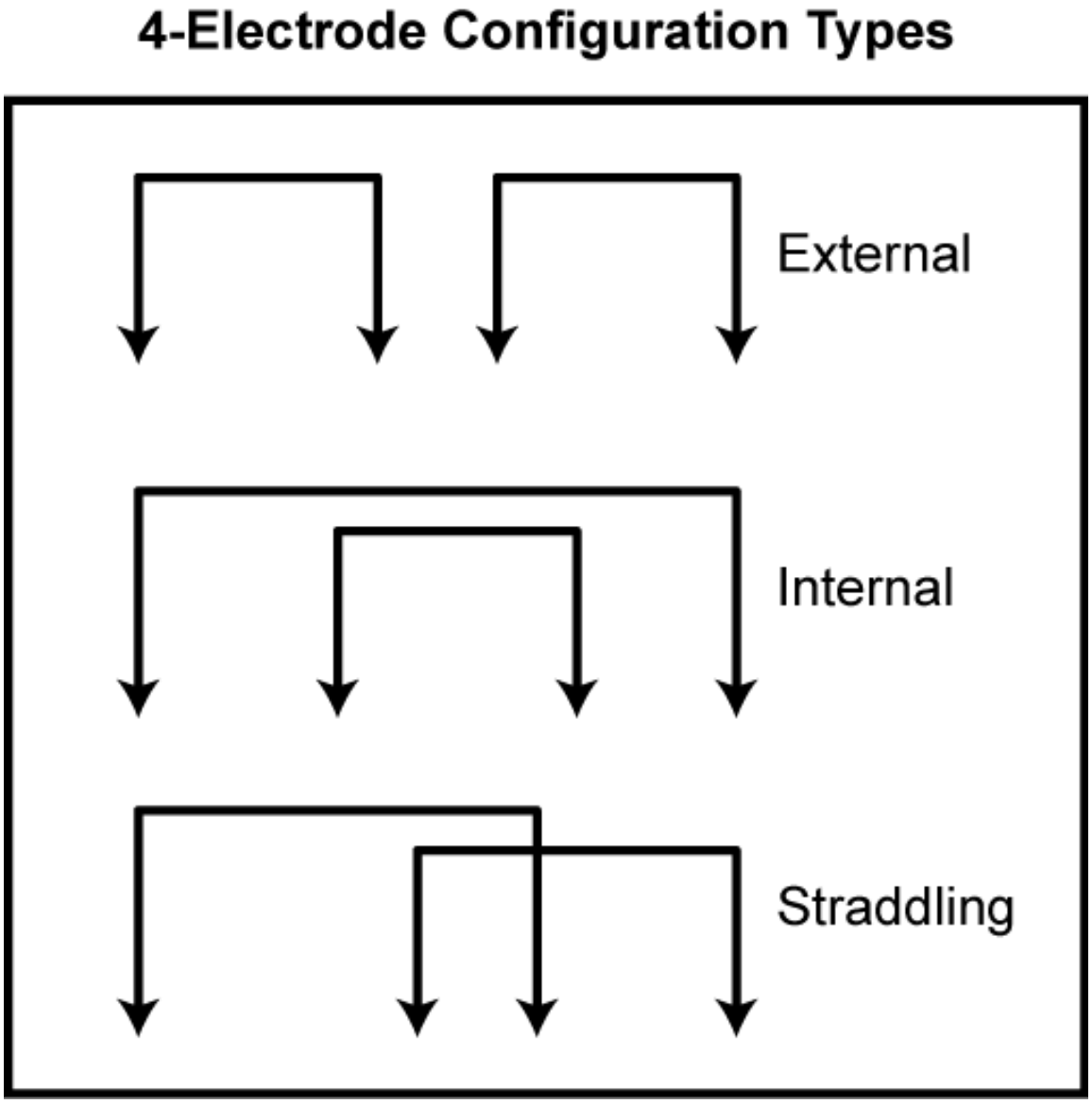


Figure 4-1 The three unique quadrupole configuration types: external, internal, and straddling.

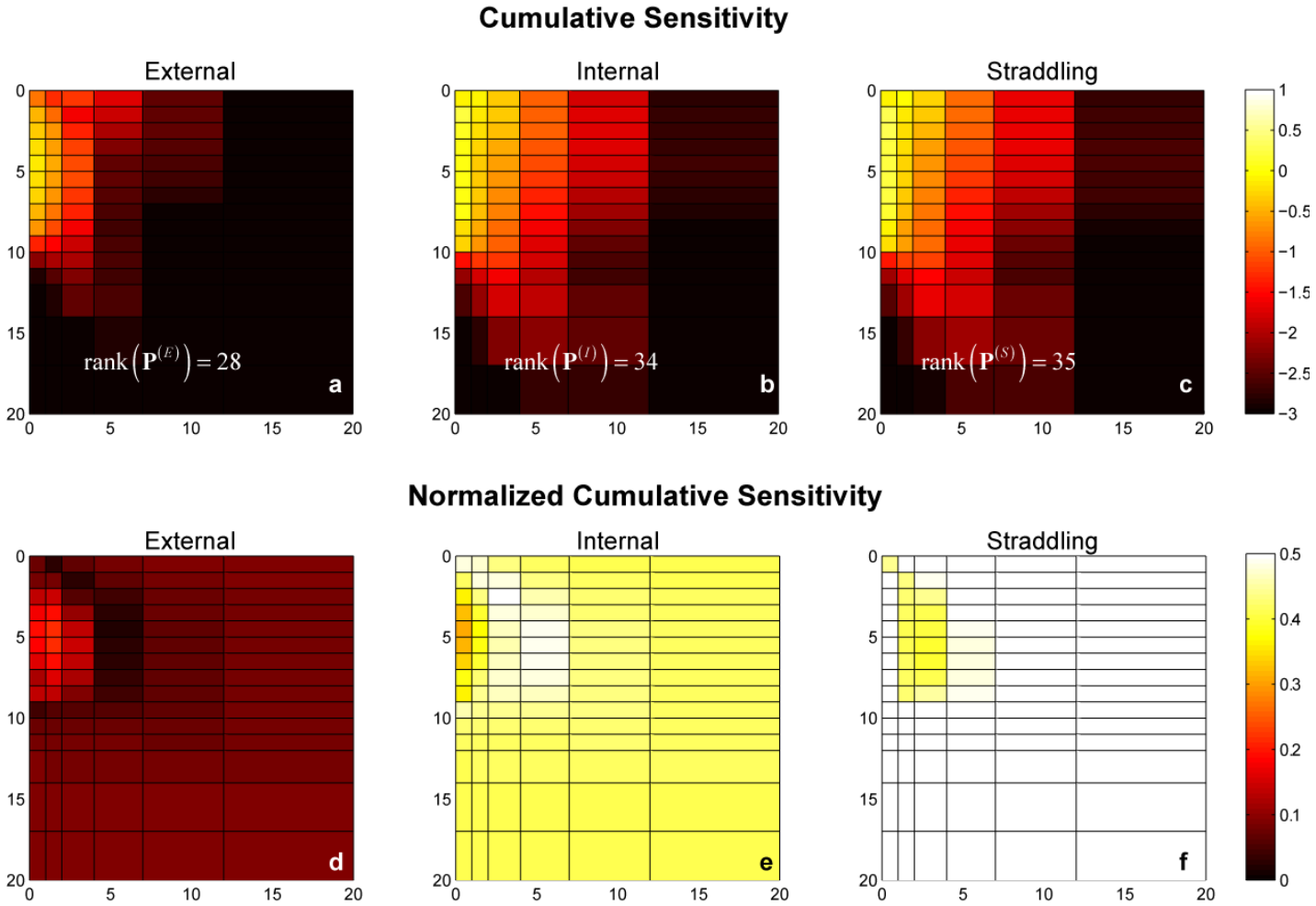


Figure 4-2 Cumulative absolute sensitivities of all electrode configurations of the same type. Panels a, b, and c show $\log_{10}$ cumulative sensitivities; panels d, e, and f show cumulative sensitivities normalized by the cumulative sensitivity of all permitted observations. Also shown are the ranks of the associated picker matrices that select all configurations of each type.

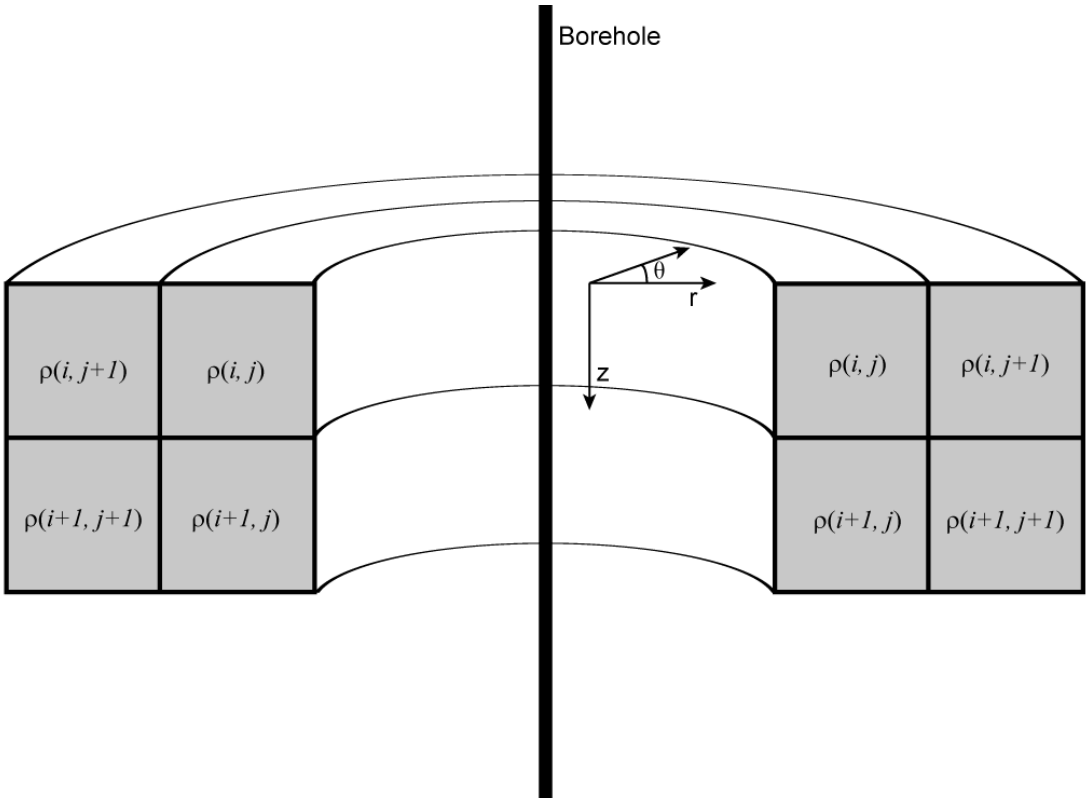


Figure 4-3 Cartoon describing the borehole model. Because one borehole is used, resistivity is treated as being azimuthally invariant as shown above. Consequently, the problem is reduced from 3D to 2D.

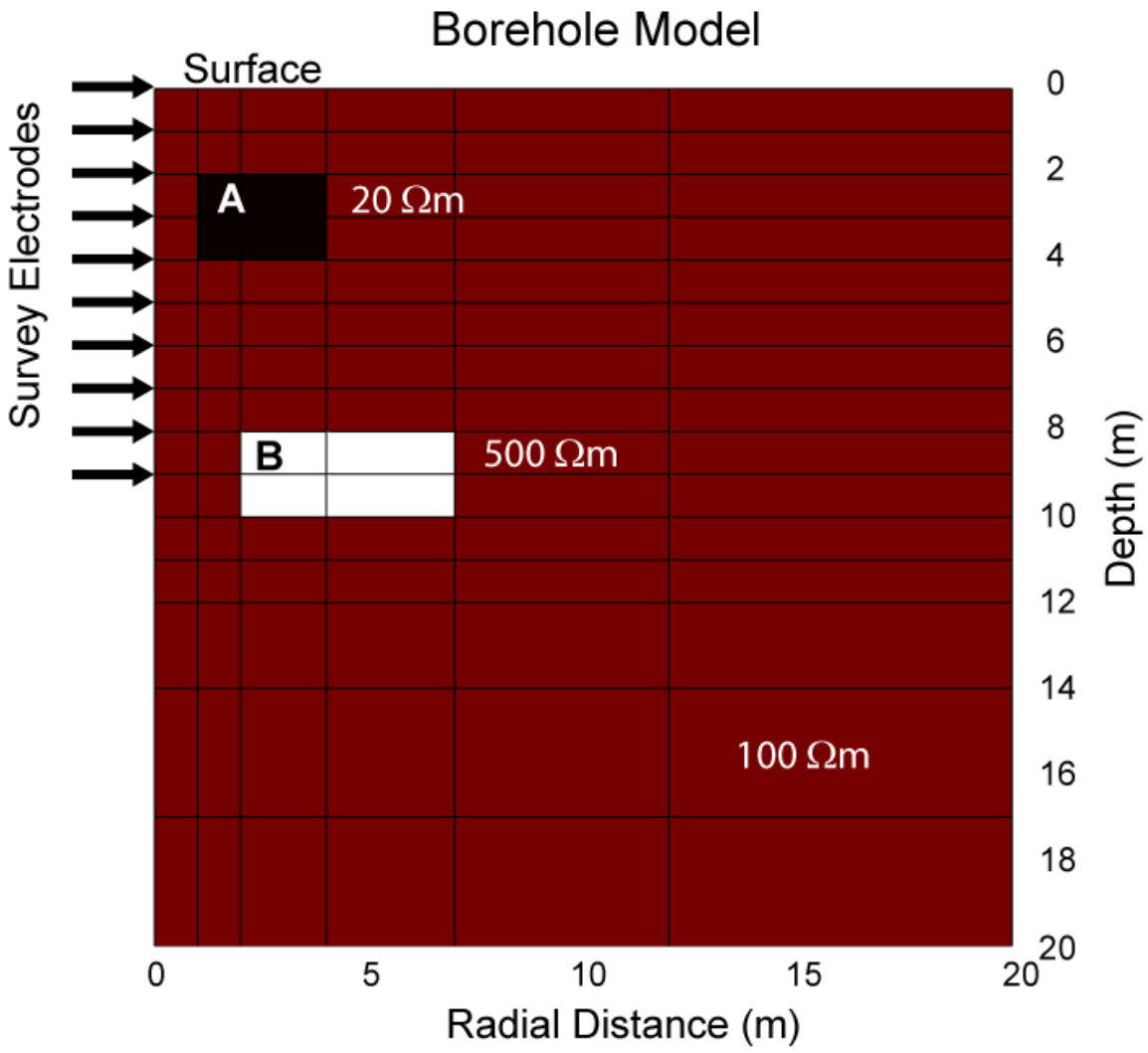


Figure 4-4 Resistivity model used in this chapter. Ten electrodes (arrows at left) were placed at equispaced intervals of 1m along the borehole, from the surface to a depth of 9 meters. The background resistivity was set to 100 Ωm and anomalies A and B were set to 20 Ωm and 500 Ωm, respectively. The discretized model extends beyond what is shown here because boundary blocks were needed to ensure modeling accuracy. The model used a total of 416 parameters.

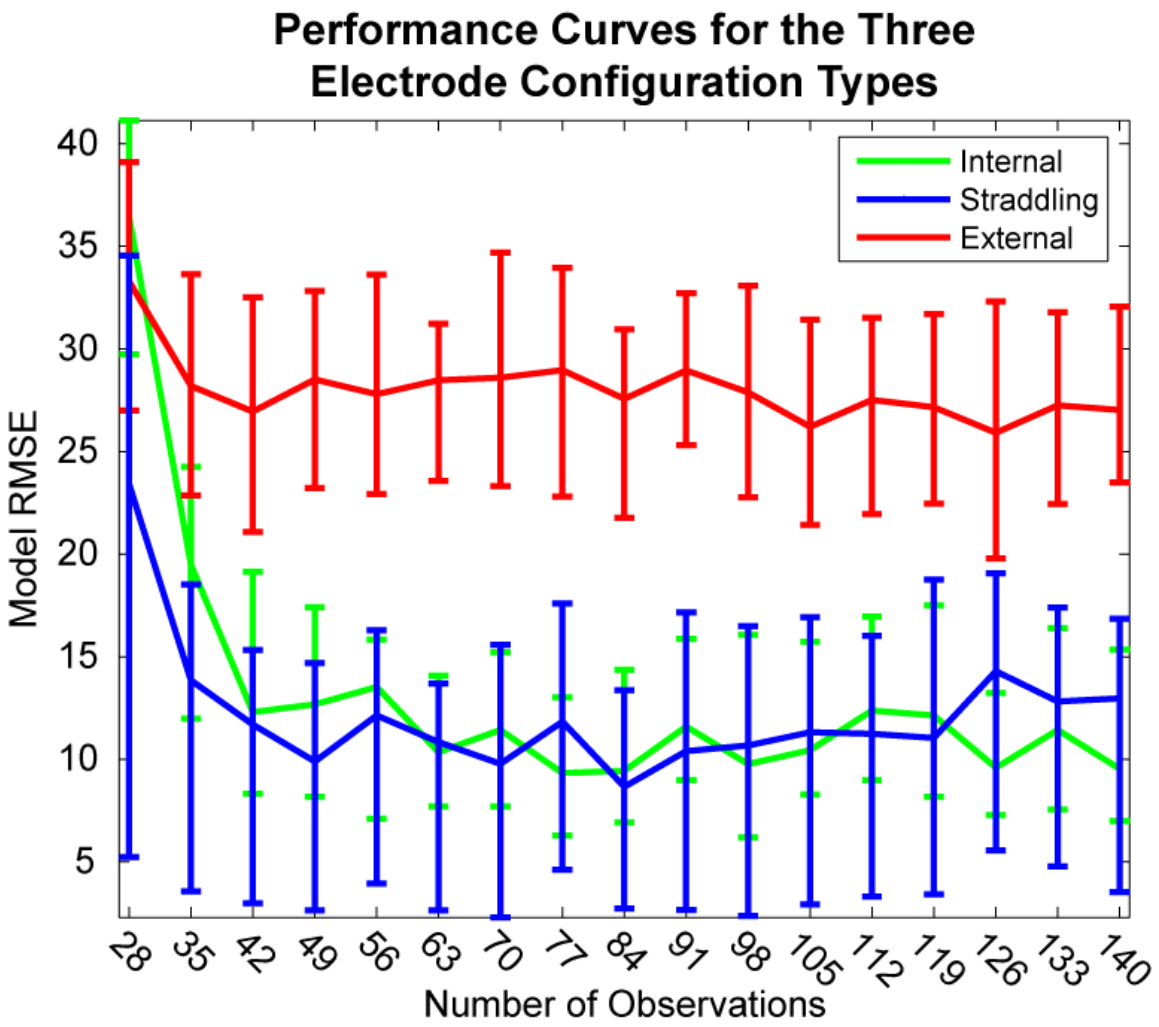


Figure 4-5 Performance of the three electrode configuration types based on a Monte Carlo investigation. Curves show the median model rms error for 50 randomly realized experiments using one of the three configuration types. Also shown are 'error bars' illustrating the 25th and 75th percentiles.

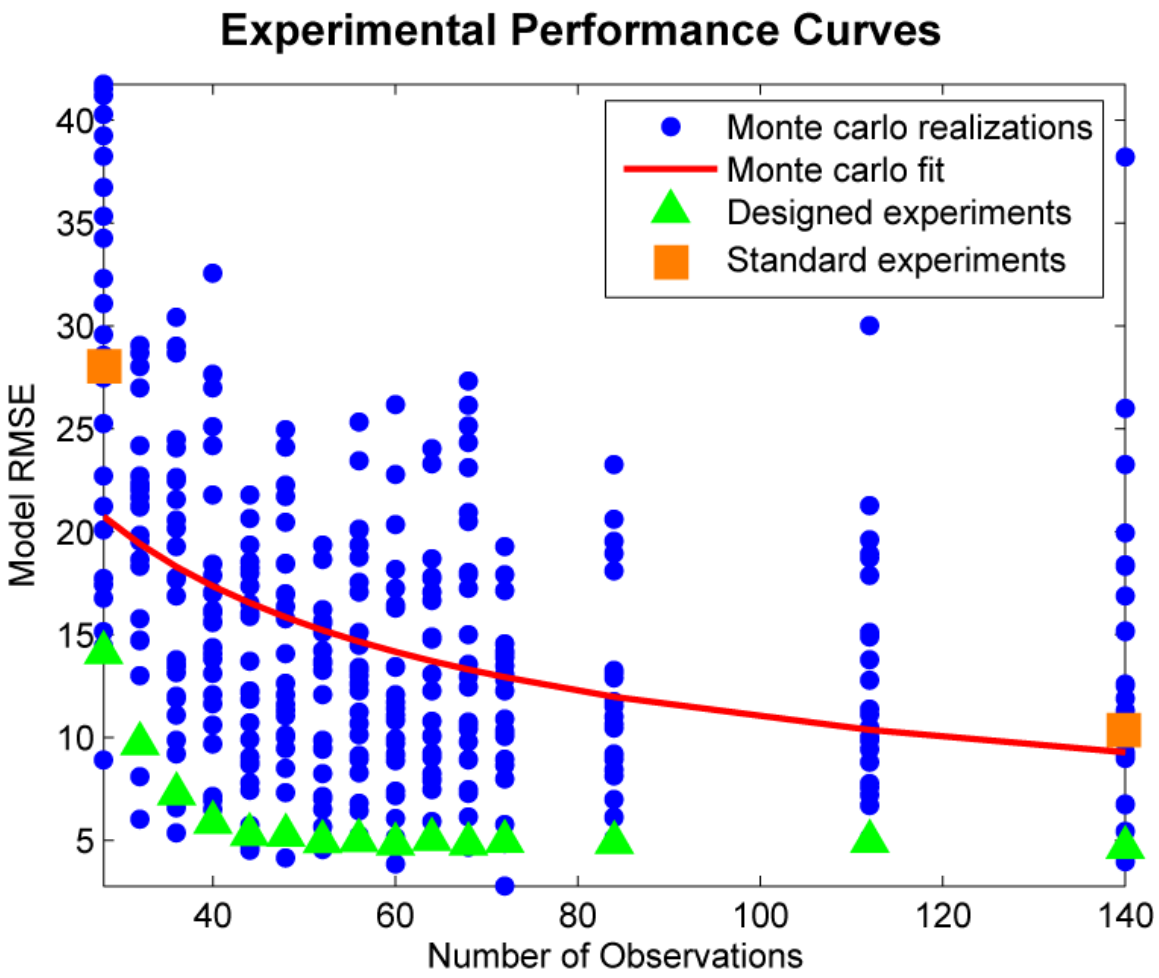


Figure 4-6 Performances for random, designed, and standardized surveys. Shown are model rms errors as a function of number of observations. The performance of randomly generated experiments (Monte Carlo) is shown in blue scatter plot, with an expected performance curve shown in solid red. The performance of designed experiments is shown by the green triangles, and the performance of two standardized surveys (Pseudosection and ERL) are shown by the orange squares.

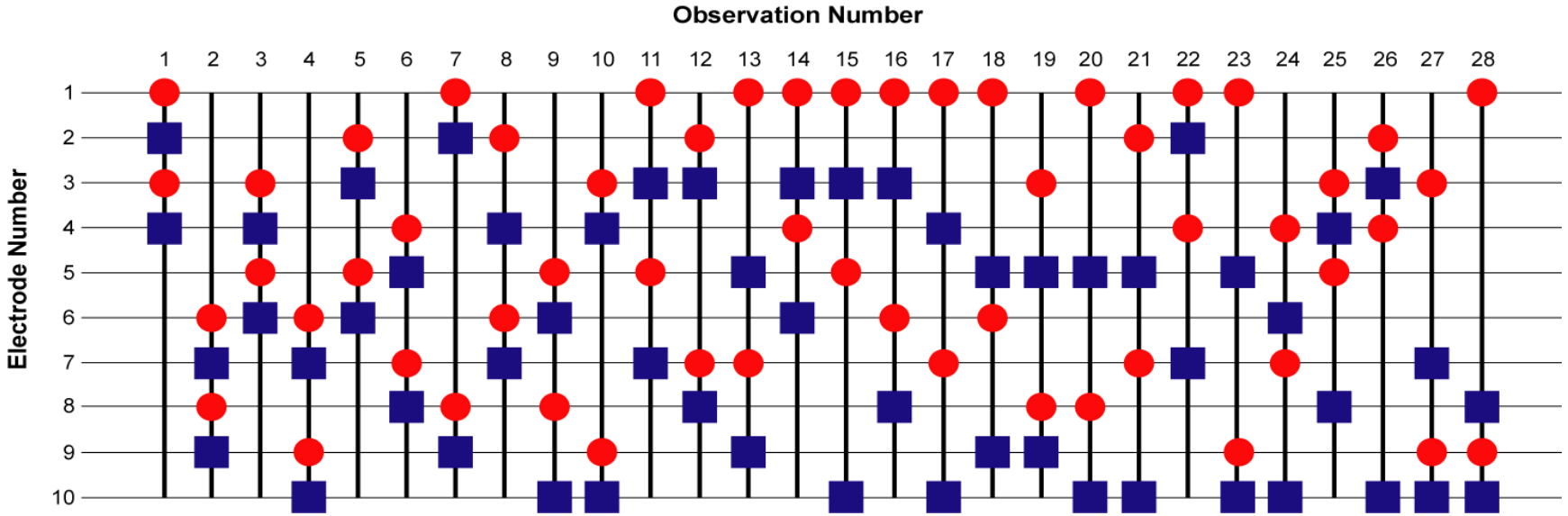


Figure 4-7 'Piano role' cartoon showing a designed experiment of 28 observations based on a homogeneous earth model. Red circles specify the positions of the transmitter dipoles and blue squares indicate the positions of the receiver dipoles. The order of the observations, from left to right, is the order in which the sequential design algorithm included them in the experiment. There is no readily apparent pattern in this set of data stations. This is unfortunate because a discernible pattern would be amenable to upscaling or downscaling with the number of borehole electrodes, allowing us to create a simple recursive formula for surveys without needing to resort to computational design.

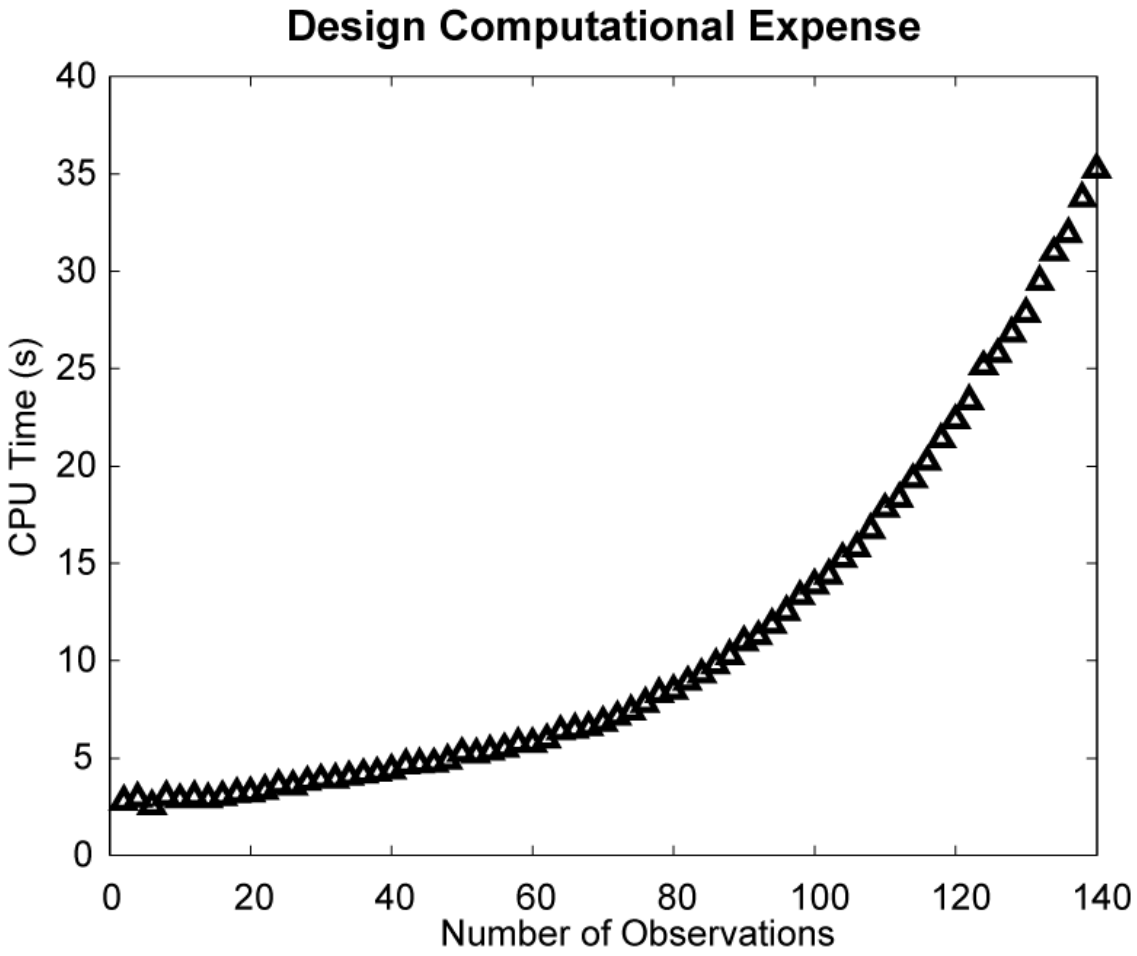
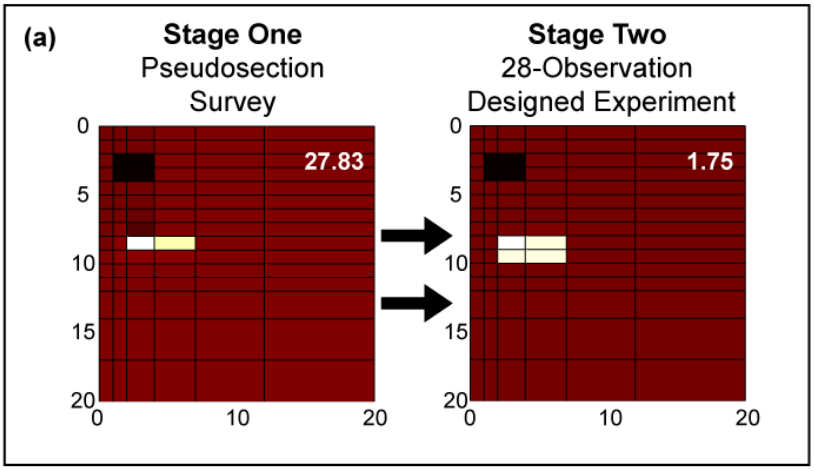


Figure 4-8 Computational expense for the sequential OED algorithm.



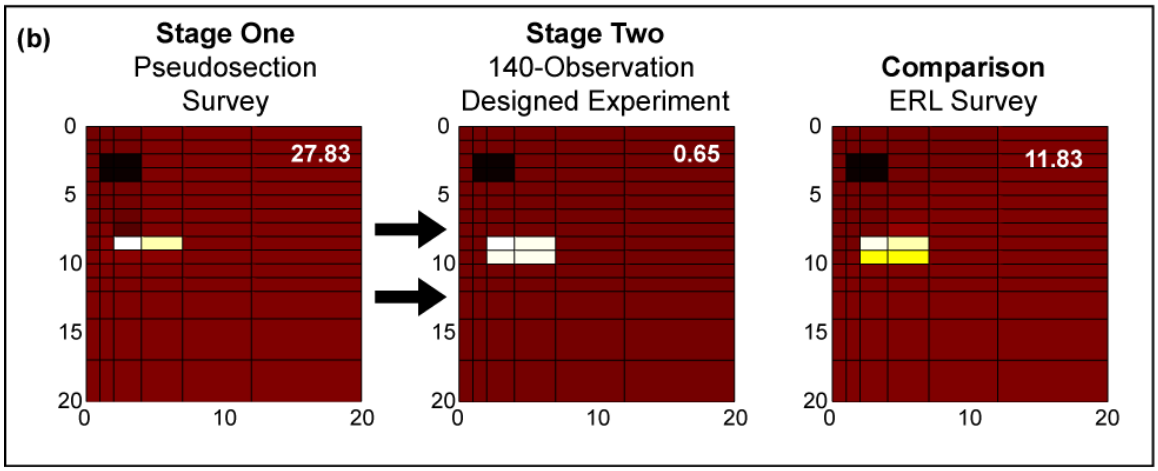


Figure 4-9 Results of two Two-Stage Adaptive Optimal Experimental Design trials. The first stage of this procedure inverts an initial dataset (Pseudosection survey) and the second stage designs an optimal experiment based on the inversion image. The second data set is then inverted, completing the procedure. Final model rms errors are shown in the top right corner of all panels. (a) The 28/28 trial (28 observations for the Pseudosection survey, 28 for the designed experiment). (b) The 28/140 trial, as well as the inversion result for the standard ERL survey for comparison.

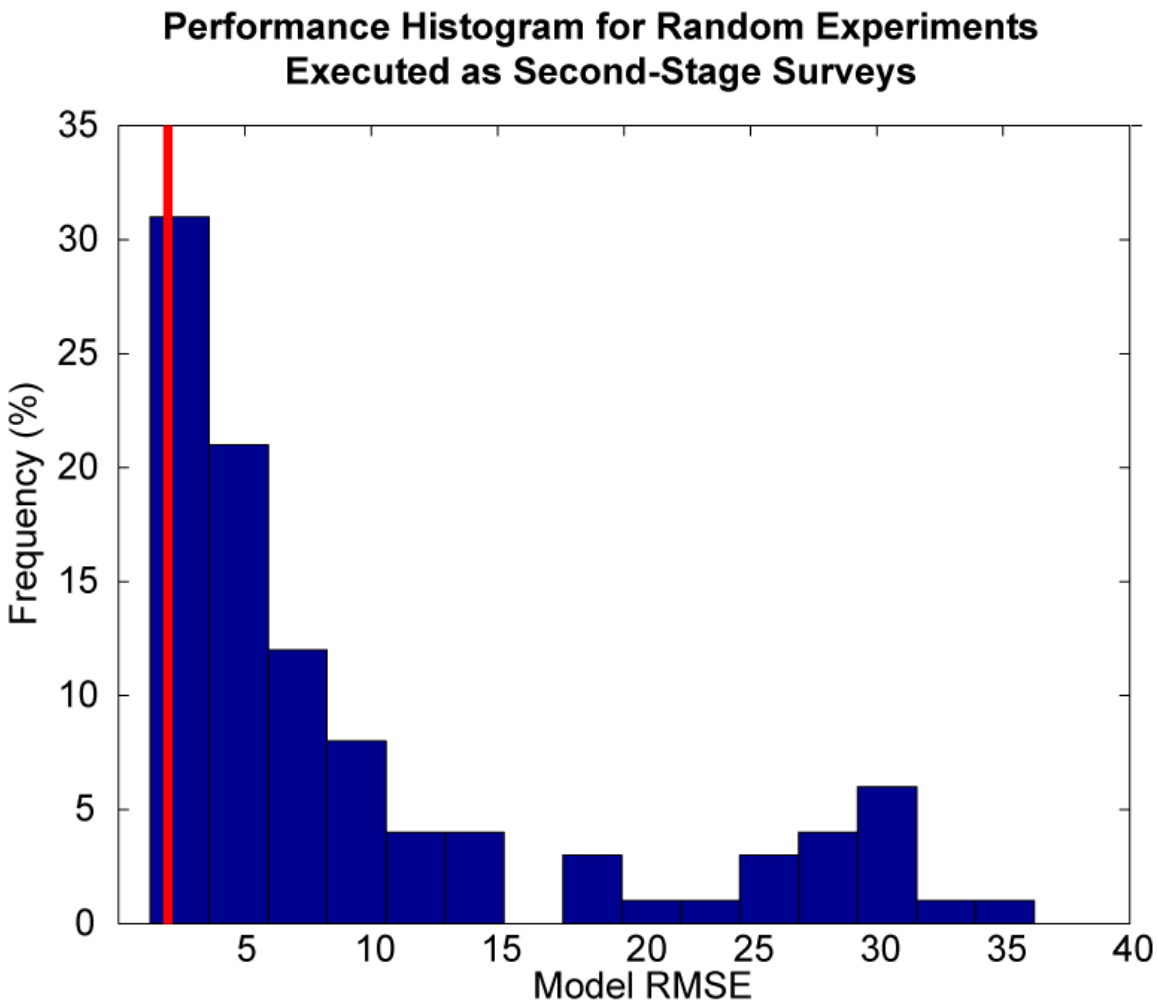


Figure 4-10 Histogram showing the frequency distribution of model rms errors for random experiments of 28 observations executed as second-stage surveys, compared with an adapted experiment also executed as a second-stage survey. The input model for all inversions was the heterogeneous model created in the first stage inversion of Pseudosection data (shown in Figure 4-9). The red line shows the model rms error attained for the adaptively designed experiment of 28 observations. The model rms error of the designed experiment is in the lowest 5th percentile, meaning it outperformed 95% of the random experiments.

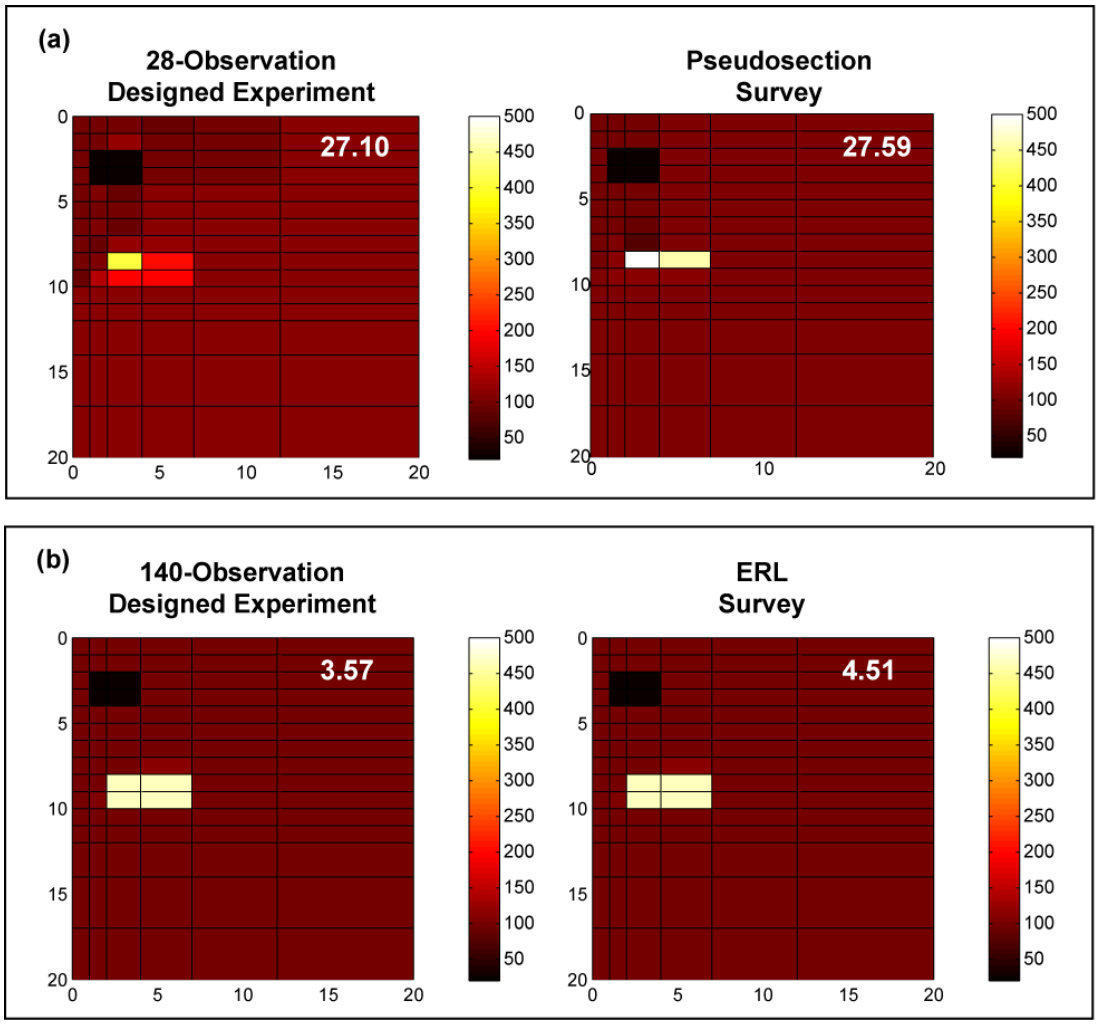


Figure 4-11 Results of two In-Line Adaptive Optimal ED trials. This AOED procedure starts with a small, initial dataset, which is partially inverted to produce an updated model; the model is then used in an experimental design phase to determine a single observation to be added to the experiment; the augmented data set is then partially 're-inverted', and the procedure continues cyclically, until convergence. Final model rms errors are shown in the top right corner of each panel. Inversion results for standardized surveys of equal size are shown for comparison. Note: the Pseudosection and ERL comparisons offered in this figure were executed as though the design algorithm had designed them. In this fashion, the effects of the 'serial' inversion implicit in this technique were accounted for. (a) 28-observation design and the Pseudosection survey comparison. (b) 140-observation design and the ERL survey comparison.

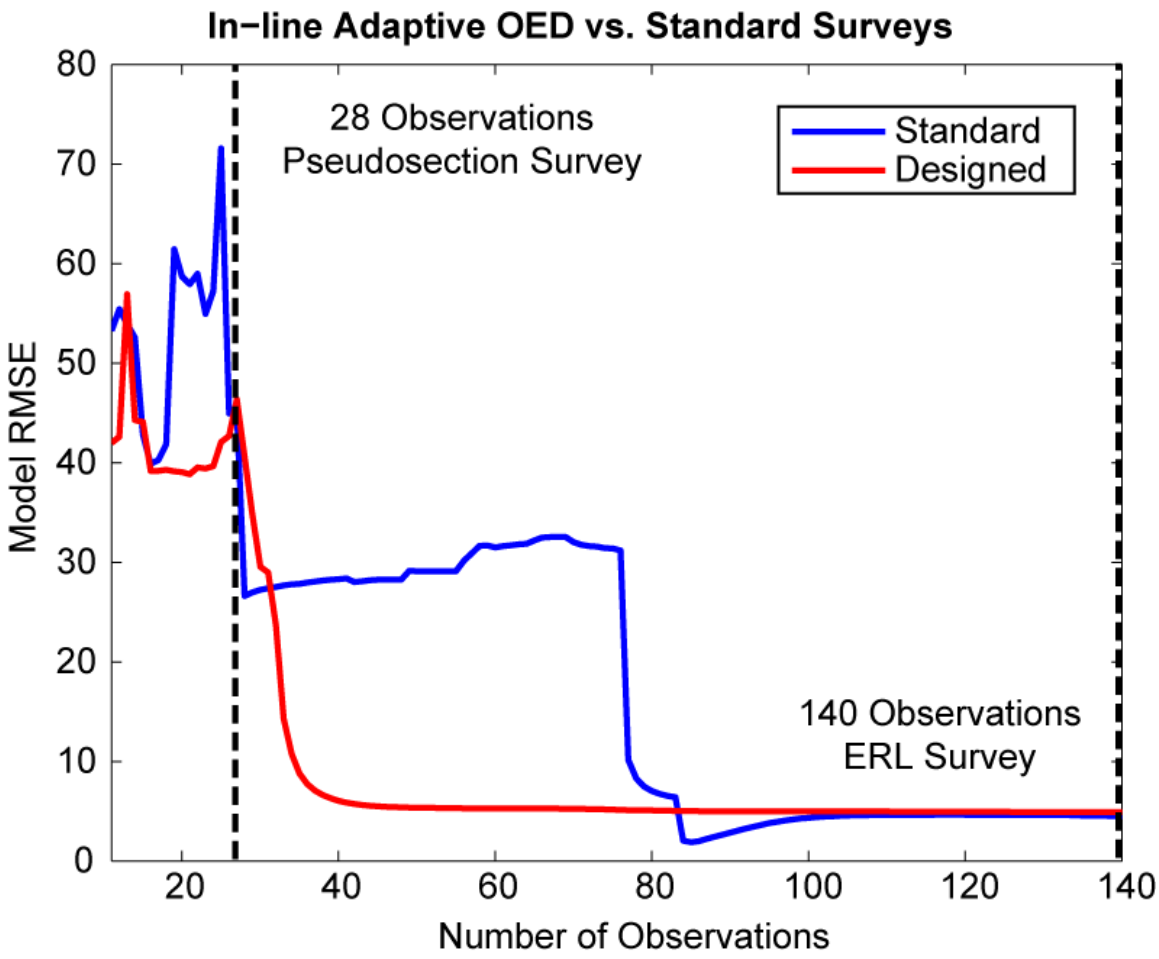


Figure 4-12 Performance curves for inline, adaptively designed surveys and standardized surveys, as a function of the number of observations. The dashed lines indicate the 28- and 140-observation positions, which correspond respectively to the Pseudosection and ERL surveys.

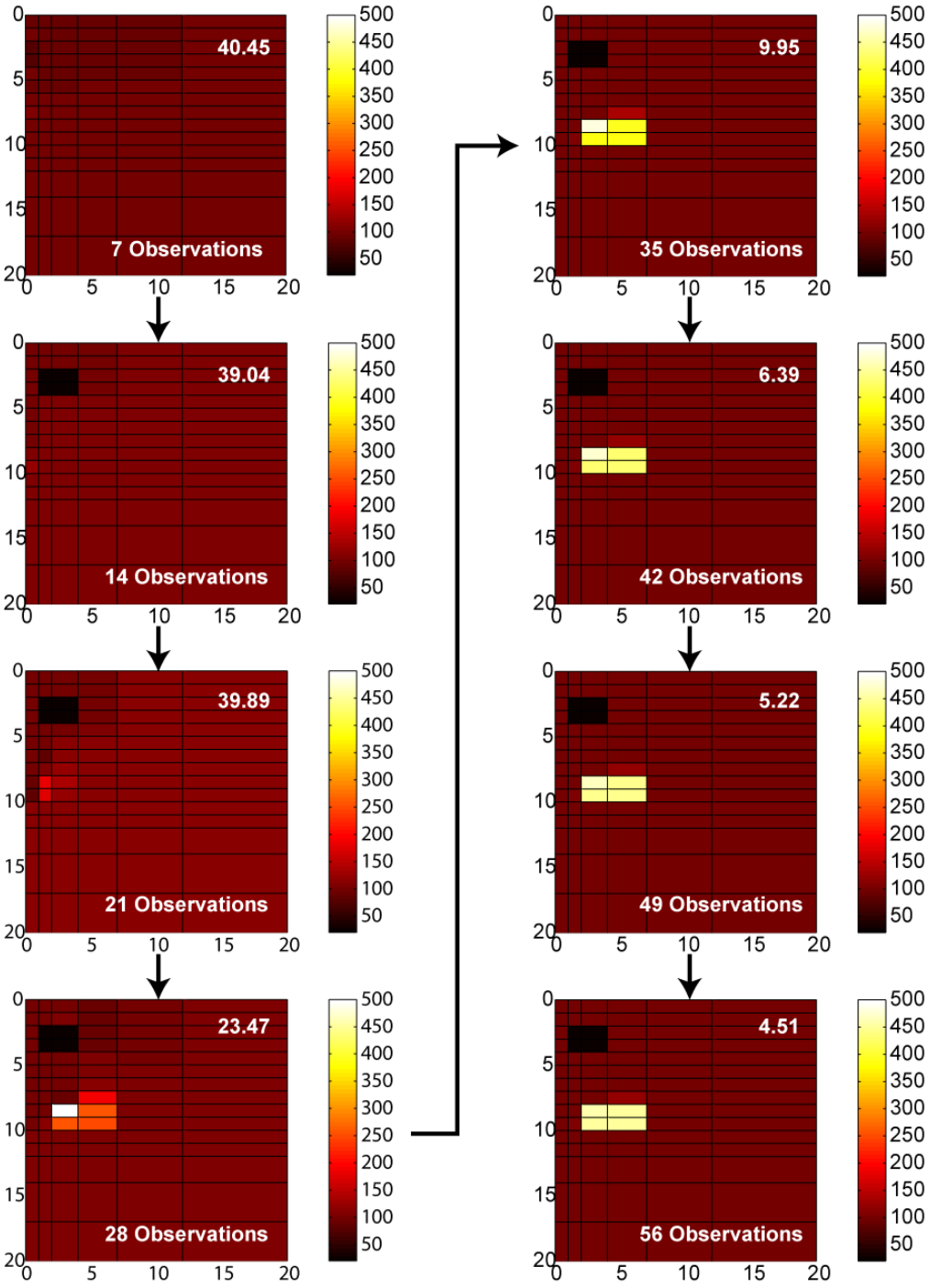


Figure 4-13 Small Batch AOED trial. A 56-observation experiment was designed in batches, following the same guidelines as the Inline AOED procedure exemplified in Figure 4-11, except observations were added in groups of seven, rather than singularly.

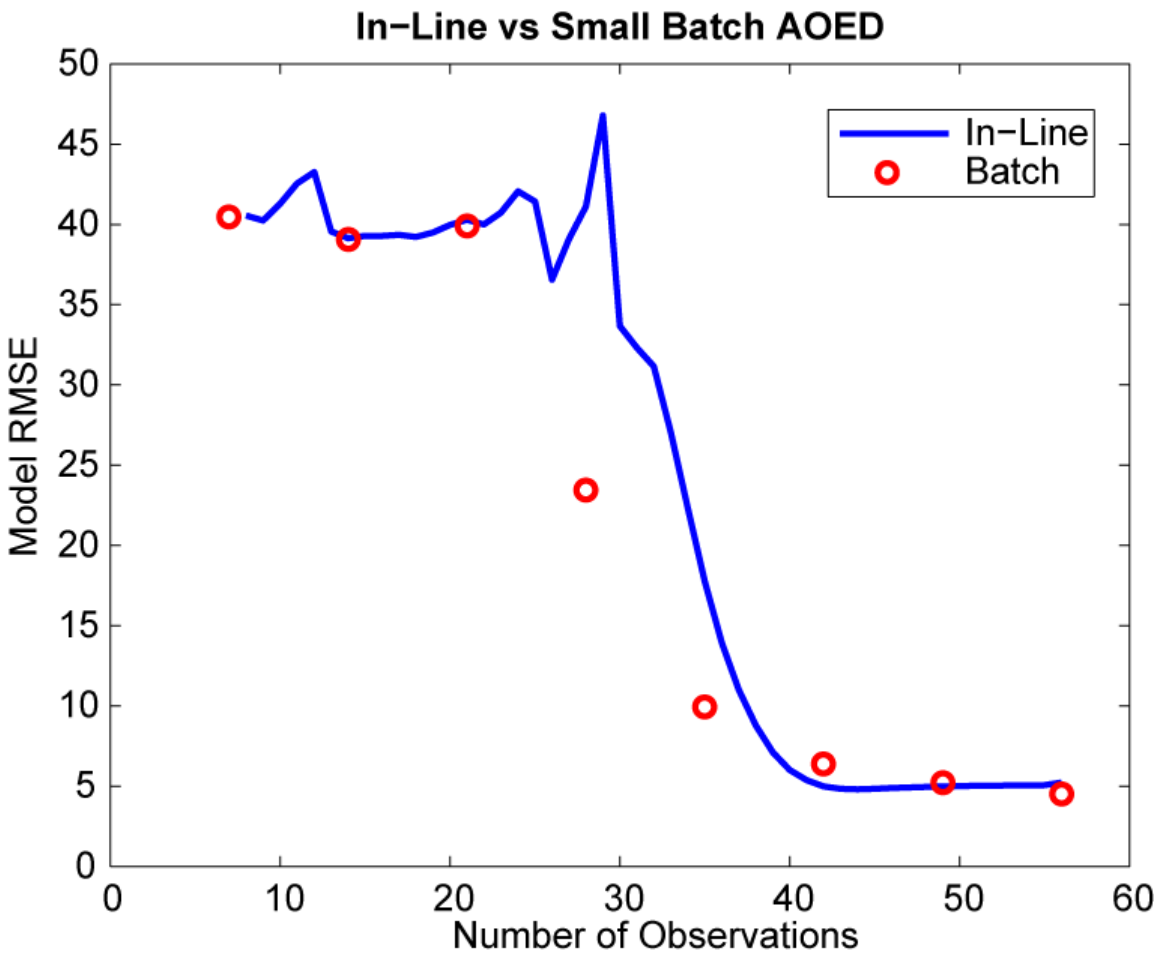


Figure 4-14 Performance curves for Inline and Small Batch AED, as a function of number of observations.

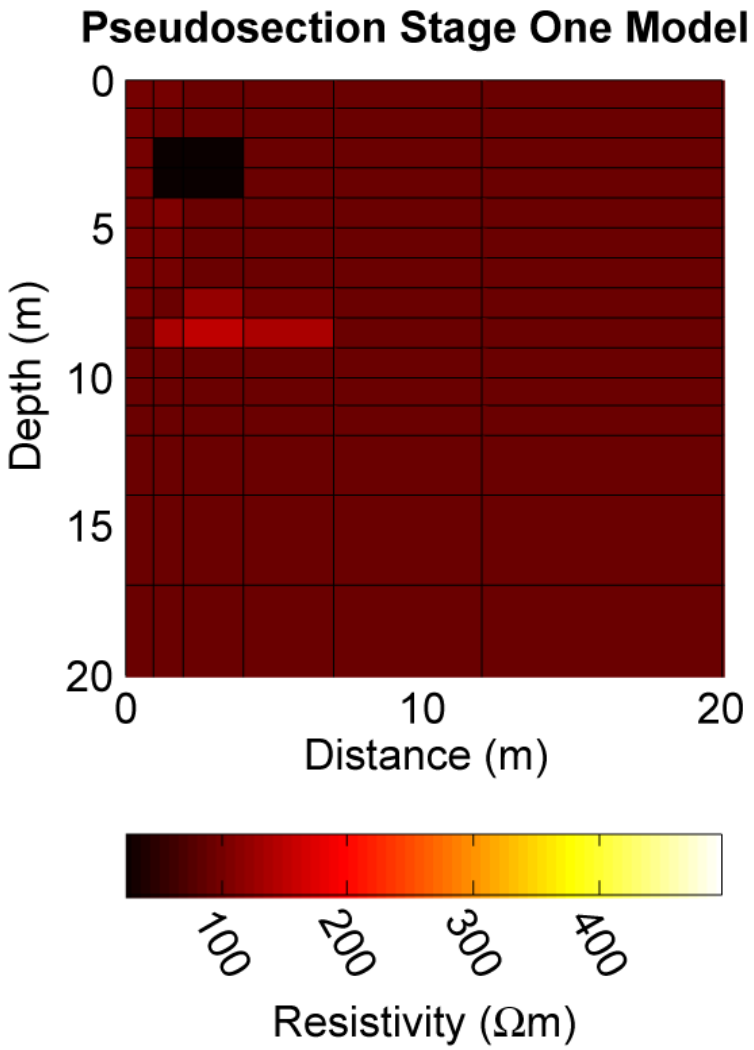


Figure 4-15 Stage one model generated by inverting Pseudosection data (28 observations) contaminated with 3% Gaussian noise. This was the working model used to design experiments for a Stage-Two inversion.

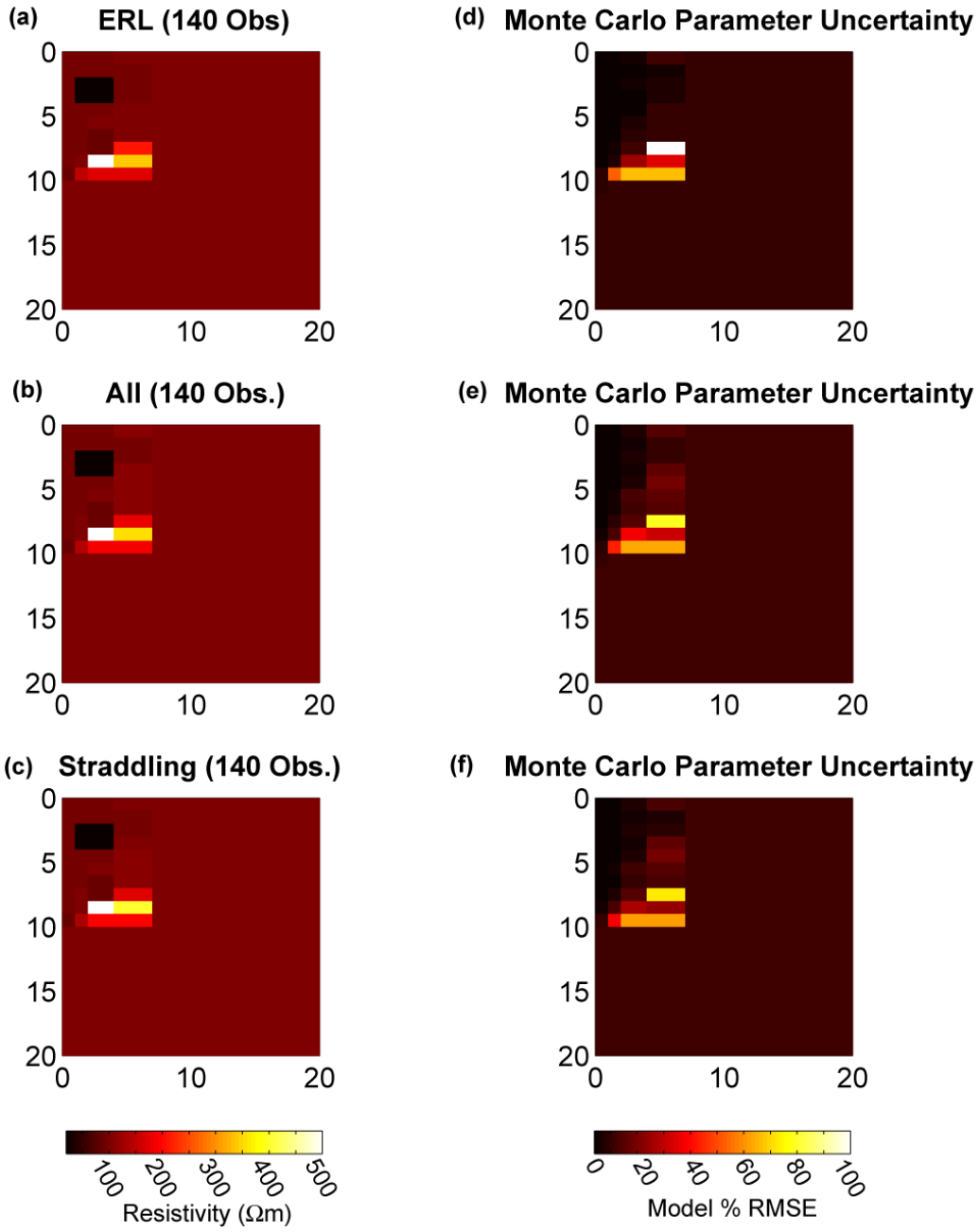


Figure 4-16 Average models (left hand panels) and model uncertainties (right hand panels) for Monte Carlo simulations of 3% Gaussian noise contaminating 'ERL', 'All' and 'Straddling' data sets of 140 observations in a Two-Stage AOED exercise. 50 realizations of noise were generated for each case. Parameter uncertainties were estimated by calculating the model % rms error between the true model (Figure 4-4) and the inversion models for all 50 realizations for each experiment.

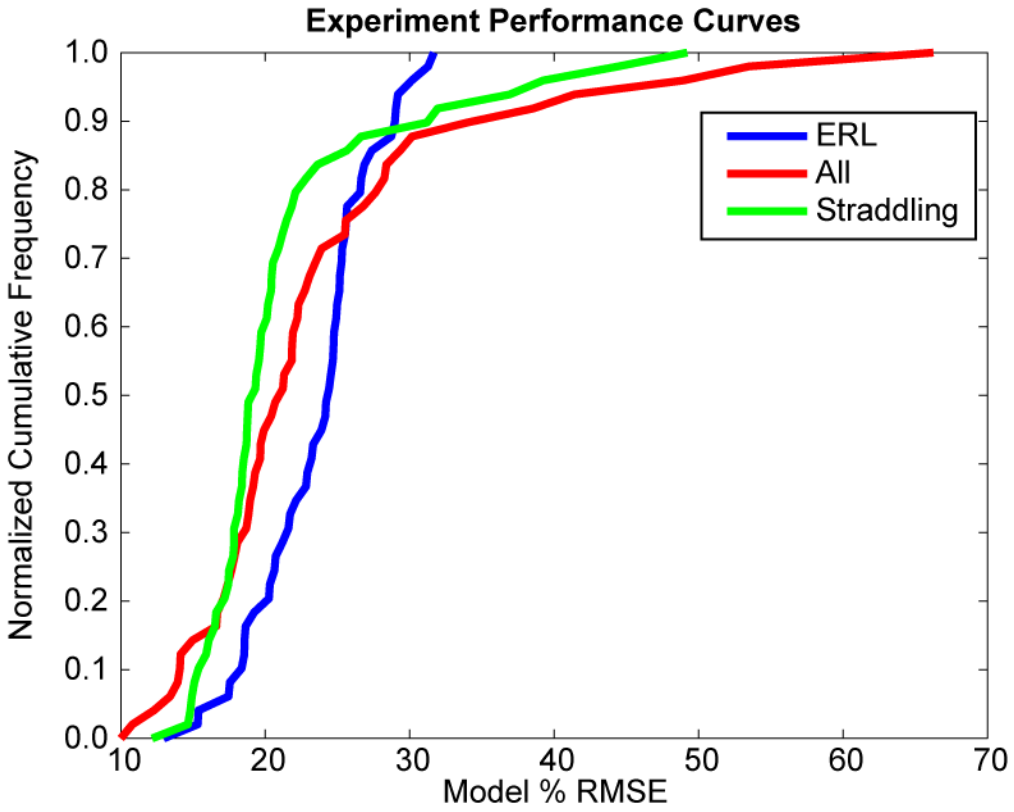


Figure 4-17 Normalized cumulative frequency of modeling errors for Monte Carlo simulations of noisy data for 'ERL', 'Straddling' and 'All' experiments of 140 observations executed as stage-two inversions.

(a) **Comparative Probabilities of Producing Smaller Modeling Error**

	ERL	All	Straddling
ERL		0.61	0.72
All	0.39		0.58
Straddling	0.28	0.42	

(b) **Average Difference in Modeling Error**

	ERL	All	Straddling
ERL		0.0%	-2.2%
All	0.0%		-2.2%
Straddling	2.2%	2.2%	

Table 4-1 Monte Carlo comparison of the 'ERL', 'Straddling' and 'All' surveys using 140 observations as 2nd stage inversions. (a) Comparative probabilities of producing smaller model % rms error between the three designed experiments according to Equation (5.20). Each column records the probabilities that the experiment in that column will produce a smaller model % rms error than the experiments in the competing rows. (b) Difference between average model % rms errors for the three experiments.

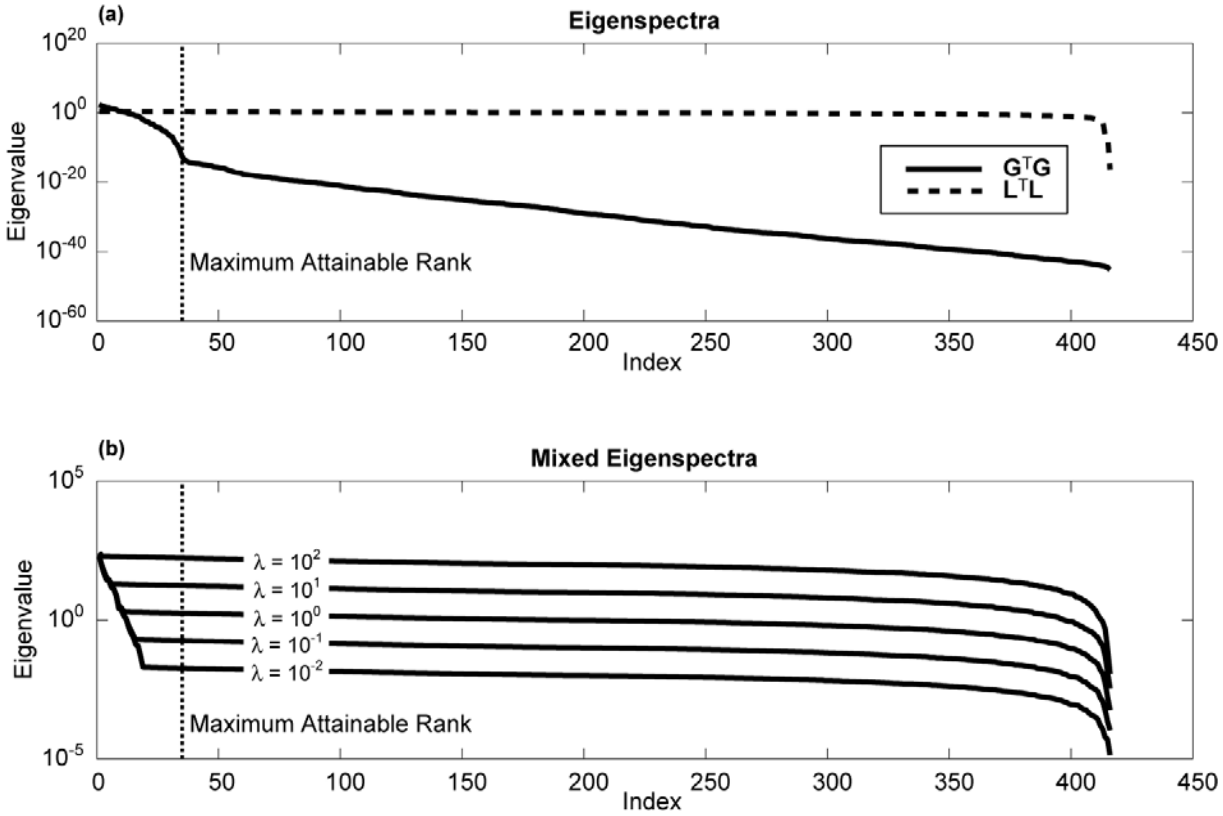


Figure 4-18 (a) Typical eigenspectra for $\mathbf{G}^T\mathbf{G}$ and $\mathbf{L}^T\mathbf{L}$ for a 10-electrode borehole inverse problem. Also shown is the maximum attainable rank for Jacobian matrices for 10-electrode surveys (vertical dashed lines). (b) Eigenspectra for the sum $\mathbf{G}^T\mathbf{G} + \lambda\mathbf{L}^T\mathbf{L}$ evaluated for different values of Lagrange multiplier, λ. Notice that $\mathbf{L}^T\mathbf{L}$ dominates the eigenspectra, even at indices less than the maximum attainable rank. This means the smoothness constraint effectively renders irrelevant the eigenvalues in $\mathbf{G}^T\mathbf{G}$ that are less than the corresponding eigenvalues in the eigenspectrum of $\lambda\mathbf{L}^T\mathbf{L}$. From an OED perspective then, any effort to maximize the eigenvalues of $\mathbf{G}^T\mathbf{G}$ falling below the 'water level' of the eigenspectrum of $\lambda\mathbf{L}^T\mathbf{L}$ is essentially wasted.

Chapter 5

Two-Dimensional Surface DC Resistivity

5.1 Introduction

Two-dimensional surface resistivity is a well-established geoelectrical method for producing cross-sectional images of earth resistivity. Forward and inverse modeling typically assumes that the earth can either be approximated by a series of infinite, rectangular bars perpendicular to the strike of the survey (e.g., Tripp *et al.*, 1984; Mcgillivray & Oldenburg, 1990; Loke & Barker, 1995), or as a finite 3D grid where semi-infinite off-axis model cells are used as boundary blocks to satisfy far-field boundary conditions (e.g., Mackie *et al.*, 1988; Shi, 1998; Rodi & Mackie, 2001). Two-dimensional models impose less stringent restrictions on the distribution of earth media than 1D models, and they can therefore model a wider variety of lithologic scenarios, such as subsurface cavities, subterranean rivers, mineral deposits and engineering structures.

This chapter explores optimal experimental design applied to surface two-dimensional resistivity problems. The sequential OED technique developed previously is adapted and applied to a number of simple case studies. Below, we briefly develop the numerics used for 2D resistivity forward/inverse modeling before moving

on to design applications. As will be seen, 2D modeling requires a large number of model cells, increasing the computational burden of optimized experimental design. Consequently, substantial effort is spent attempting to find ways to expedite the design algorithm. Of all the results compiled herein, the discussion and development of novel design expedients may well be the most important contribution of this research.

This chapter catalogs several semi-independent research topics; so, rather than presenting holistic sections on theory and methodology for the entire document, each topic is self-contained, with pertinent theoretical and methodological details being presented contextually. These topics are offered in the following order: (1) Modeling and Inversion; (2) Expediting Design; (3) Homogeneous Designs; and (4) Adaptive, Optimal Experimental Design. Following these sections, concluding statements are provided that integrate and summarize the results from these topical areas.

5.2 Modeling and Inversion

Two-dimensional surface resistivity modeling and inversion was executed using a transmission line network analogy (Swift, 1971; Madden, 1972; Zhang *et al.*, 1995; Shi, 1998) in which the earth is discretized and modeled by a network of resistors whose resistances depend on intrinsic earth resistivities. The interested reader is referred to the above citations for further details. Mixed boundary conditions, introduced by Zhang et al. (1995), were used at the boundaries of the model grid. Despite the efficiency of this boundary condition method, additional *boundary blocks* were still needed to ensure the accuracy of modeled potentials within the target area. As these additional boundary blocks slow forward and inverse computations, their size was increased logarithmically as the distance from the electrical survey to minimize their numbers. Figure 5-1 shows an example of the model grid and target window for a survey of 20 electrodes.

All inversions in this chapter were carried out by nonlinear least squares formalism. In particular, Jacobian matrices were explicitly tabulated and inverted, though faster techniques of inversion have been introduced, such as the Conjugate Gradients

method (e.g., Mackie & Madden, 1993b; Zhang *et al.*, 1995; Shi, 1998; Rodi & Mackie, 2001), that do not require the explicit computation of Jacobians. However, our design methodology *requires* Jacobian matrices, so it was determined that all data inversions should employ them as well. Inversions were carried out using a dual objective function requiring the simultaneous minimization of both data rms error and model roughness. The model regularization was implemented by means of the L1-norm of the spatial gradient of the model, in contrast to the more common L2-norm, which is also known as basic Tikhonov Regularization (Tikhonov & Arsenin, 1977). Mathematically, the objective function for the inversions is

$$\Theta(\mathbf{m}) = \|\mathbf{d} - \mathbf{G}\mathbf{m}\|_2^2 + \lambda \|\nabla \mathbf{m}\|_1, \tag{5.1}$$

where **d** is the data vector, **G** is the Jacobian matrix, **m** is the model vector and λ is a Lagrange multiplier governing the trade-off between the two competing objectives. By using the L1-norm instead of the L2-norm, sharper inversion images could be created because the L1-norm is less sensitive to large resistivity contrasts between adjacent model cells. For these iterative nonlinear least squares inversions, the Levenberg-Marquardt algorithm was used to govern the Lagrange multiplier (Levenberg, 1944; Marquardt, 1963).

5.3 Expediting Design: Reducing the Size of the Jacobian Matrix

Poisson's equation governs the electrostatic domain for resistivity problems. It is a continuous differential equation of state, and to practically model and invert 2D resistivity data, the earth is discretized into cells of constant resistivity. For discretized modeling to accurately reflect solutions to Poisson's equation, boundary conditions must be satisfied, which usually entails either or both the Neumann and Dirichlet BCs (e.g., Pelton *et al.*, 1978; Tripp *et al.*, 1984) or some type of mixed boundary condition (Dey & Morrison, 1979; Zhang *et al.*, 1995), depending on the problem set up. To actualize these boundary conditions, additional *boundary blocks* are added around

the edges of the discretized target area. These boundary blocks are necessary to achieve accurate modeling results, but they are nuisance parameters from an inversion standpoint because we must solve for their resistivities despite the fact that these resistivities are of no interest. The efficient, mixed boundary conditions introduced by Zhang *et al.* (1995) were used for the forward and inverse modeling in his work, but even so, with additional left, right and bottom boundary blocks, the total number of resistivity cells needed for an array of 20 electrodes was nearly 1100 (21 row cells x 51 column cells).

From the perspective of our OED method, an inverse problem in 1100 parameters is large. Calculating the nullspace of the Jacobian of the base experiment (which must be performed each time a new observations is to be added) is still within our present-day computational means, but the nullspace matrix will have approximate dimensions of $1100 \times (1100 - n)$, where n is the number of observations currently in the base experiment. Not only will calculating this matrix take time, but also its storage in memory approaches present-day limitations (for workstations, personal computers, etc.). Naturally, one wonders whether there is a way to expedite OED computations by reducing the size of the Jacobian matrix and thereby the size of the nullspace matrix.

Below, two different methods are introduced for usefully reducing the size of the Jacobian matrix for the purposes of experimental design. The first proceeds by deliberately *truncating* the Jacobian matrix according to a user-specified threshold term. The second proceeds by *compressing* the information in the Jacobian matrix. Both techniques are fully developed initially, and the section then concludes with a comparative examination to assess their relative utility.

5.3.1 Truncating the Jacobian Matrix, Method 1

The first and most straightforward method for reducing the size of the OED Jacobian matrices introduces the idea of truncation. Because the boundary blocks in Figure 5-1 (e.g.) are outside the target area where the survey electrodes are placed, their sensi-

tivities are relatively small. This, of course, means that some form of model regularization is needed to prevent the resistivities in these model cells from diverging to arbitrarily large or small values, and indeed the inversion method described above uses a smoothness constraint to impose regularization. It also means that little can be done to improve the sensitivity of these cells to the data. Consequently, the simplest way of reducing the size of the Jacobian matrix would be to remove from it all columns that correspond to boundary blocks, leaving only those columns that correspond to model cells in the target window.

5.3.2 Truncating the Jacobian Matrix, Method 2

A second method of reducing the size of Jacobian matrices is an alternate method of truncation. To motivate this method, consider the following hypothetical Jacobian matrix:

$$\mathbf{G}_{hyp} = \begin{bmatrix} 0 & 1 & 0 & 4 & 1 \\ 0 & 2 & 0 & 3 & 3 \\ 0 & 3 & 0 & 2 & 2 \\ 0 & 4 & 0 & 1 & 4 \end{bmatrix}. \tag{5.2}$$

The singular value decomposition of $\mathbf{G}_{hyp}$ is

$$\mathbf{G}_{hyp} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T, \tag{5.3}$$

where

$$\mathbf{U} \cong \begin{bmatrix} 0.38 & 0.76 & 0.16 & -0.50 \\ 0.52 & 0.21 & -0.66 & 0.50 \\ 0.43 & -0.06 & 0.73 & 0.50 \\ 0.61 & -0.61 & -0.09 & 0.50 \end{bmatrix}$$

$$\mathbf{\Sigma} \cong \begin{bmatrix} 8.77 & 0 & 0 & 0 \\ 0 & 3.48 & 0 & 0 \\ 0 & 0 & 0.98 & 0 \\ 0 & 0 & 0 & 0 \end{bmatrix} \quad . \tag{5.4}$$

$$\mathbf{V} \cong \begin{bmatrix} 0 & 0 & 0 & -1.00 & 0 \\ 0.60 & -0.41 & 0.69 & 0 & 0 \\ 0 & 0 & 0 & 1.00 & 0 \\ 0.53 & 0.85 & 0.05 & 0 & 0 \\ 0.61 & -0.33 & -0.72 & 0 & 0 \end{bmatrix}$$

The rank of $\mathbf{G}_{hyp}$ equals the number of its nonzero singular values, 3. Because two columns of $\mathbf{G}_{hyp}$ are all zeros, they only make the matrix larger, without adding any information. This is clear not only by noting that $\text{rank}\left(\mathbf{G}_{hyp}\right) = 3$ but also by observing that last column of $\mathbf{V}$ is all zeros. In fact, we can truncate $\mathbf{G}_{hyp}$ by removing these zero columns. This does not change the singular values, which means that the same information is still present but in a usefully reduced form. Consider the truncated submatrix of $\mathbf{G}_{hyp}$,

$$\tilde{\mathbf{G}}_{hyp} = \begin{bmatrix} 1 & 4 & 1 \\ 2 & 3 & 3 \\ 3 & 2 & 2 \\ 4 & 1 & 4 \end{bmatrix}. \tag{5.5}$$

Its SVD is

$$\tilde{\mathbf{G}}_{hyp} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T, \tag{5.6}$$

where

$$\tilde{\mathbf{U}} \cong \begin{bmatrix} 0.38 & 0.76 & 0.16 & -0.50 \\ 0.52 & 0.21 & -0.66 & 0.50 \\ 0.43 & -0.06 & 0.73 & 0.50 \\ 0.61 & -0.61 & -0.09 & 0.50 \end{bmatrix}$$

$$\tilde{\mathbf{\Sigma}} \cong \begin{bmatrix} 8.77 & 0 & 0 \\ 0 & 3.48 & 0 \\ 0 & 0 & 0.98 \\ 0 & 0 & 0 \end{bmatrix} \quad . \tag{5.7}$$

$$\tilde{\mathbf{V}} \cong \begin{bmatrix} 0.60 & -0.41 & 0.69 \\ 0.53 & 0.85 & 0.05 \\ 0.61 & -0.33 & -0.72 \end{bmatrix}$$

Observe that: $\tilde{\mathbf{U}} = \mathbf{U}$; $\tilde{\mathbf{\Sigma}} = \mathbf{\Sigma}$, except that the zeroes column has been removed; and $\tilde{\mathbf{V}}$ is a submatrix of **V**, with the first and third rows and fourth and fifth columns removed. In words, the singular values of $\tilde{\mathbf{G}}_{hyp}$ remain unchanged, which means the information contents of the two matrices are identical. However, the right singular vectors in **V** (which comprise an orthonormal basis that spans model space) have been reduced from 5 dimensions to 3 dimensions. This is a method for reducing a high dimensional space to a more manageable one, and is known in the literature by multiple name, including as principal component analysis (e.g., Jolliffe, 2002). In fact, the singular vectors in $\tilde{\mathbf{V}}$ are *parallel* to the associated singular vectors in **V** in the 5-dimensional space spanned by **V** (i.e., we augment each vector in $\tilde{\mathbf{V}}$ by placing zeroes before the first element and between the first and third elements to make them 5-dimensional). So not only are the singular values the same, but the relevant singular vectors in **V** and $\tilde{\mathbf{V}}$ point in exactly the same directions in the original model space.

From an OED standpoint, the first and third parameters (corresponding to the first and third columns of $\mathbf{G}_{hyp}$) are essentially irresolvable no matter what observations are made. Therefore, it makes little sense to expend effort trying to design an experiment to resolve these parameters. Our OED technique operates on the nullspace of the Jacobian matrix, and this null space contains one or more columns of the **V** matrix. Hence, if we reduce the size of **V** by truncating $\mathbf{G}_{hyp}$ to remove poorly resolved

parameters, the experimental design algorithm can execute more quickly, without biasing the design toward these irremediably poorly resolved parameters. The important point is that the dimension of right singular vectors in **V** can be usefully reduced without significantly altering the information content of the Jacobian matrix.

Because many boundary blocks are modeled at great distances from the survey electrodes, their sensitivities are small. Of course, this is one of the reasons why model regularization is critical for 2D resistivity problems; the sensitivities of far-away cells are so small that nearly any resistivity value can be placed in these cells without seriously affecting the predicted data. In effect, the columns of the Jacobian matrix corresponding to boundary blocks will have values near zero, while the columns corresponding to cells directly beneath the survey will not. This is a situation very similar to the simple exposition given in Equations (5.2) - (5.7) except the 'zero' columns will not be identically zero but nearly so. If we deliberately disregard the boundary blocks in the design procedure, we can reduce the size of the Jacobian matrix and expedite design. In a sense, one can think of this as *focusing* the design on parameters we actually wish to resolve, those in the target window. A similar idea was introduced by Curtis (1999b), though it did not address focusing in terms of boundary blocks but in terms of a *subset* of the cells *within* the target window.

No column of the Jacobian matrix for a real 2D resistivity problem is ever precisely a zero vector, so the remaining issue is to establish a criterion for retaining or discarding columns. In essence, we must determine which columns are 'close enough' to zero vectors that they can be discarded for the purposes of experimental design. This problem is approachable by employing a thresholding criterion.

Consider a homogeneous earth discretized into 21 × 51 model cells (Figure 5-1.a) and queried by 20 electrodes (Figure 5-1.b). We first calculate the Jacobian matrix of all possible 4-electrode observations and then the *cumulative sensitivity* of each column, expressed by its L2-norm, which gives rise to a spatial distribution of cumulative sensitivities shown in Figure 5-2.a. The L2-norm is a useful way to quickly determine how close to zero the elements in any column of the Jacobian are (*Note: The sensitivities in Figure 5-2.a have been log rescaled (see the colorbar at right) to high-*

light the many orders of magnitude over which they vary). Because the cumulative sensitivities in Figure 5-2.a are the square root of the sums of squared sensitivities, no corresponding column element in the Jacobian matrix can exceed the values shown. In other words, a cell whose cumulative sensitivity is 10^{-4} implies that the magnitude of no element in the corresponding column of the Jacobian exceeds 10^{-4}. Hence, Figure 5-2.a presents an upper bound on the magnitude of the elements in the corresponding columns of the Jacobian. We can use these cumulative sensitivities to threshold the columns of the Jacobian. A threshold, α, can be specified so that any column whose cumulative sensitivity is less than this will be discarded for the OED exercise. For example, if we choose $\alpha = 10^{-4}$, only columns of the Jacobian matrix whose cumulative sensitivities exceed 10^{-4} will be retained for the design exercise. This kind of threshold is arbitrary, however, because it does not account for the *relative* magnitudes of the cumulative sensitivities; if all sensitivities in the Jacobian were on the order of 10^{-4}, nearly all columns would be discarded for the design exercise.

Instead, we have adopted a percentile-based threshold approach. If we divide each cumulative sensitivity by the sum of *all* cumulative sensitivities, we produce a *relative* measure of each model cell's cumulative sensitivity with respect to the experiment using all possible 4-electrode observations. The relative cumulative sensitivities are expressed as

$$S_j = \frac{\sqrt{\sum_i G_{ij}^2}}{\sum_j \sqrt{\sum_i G_{ij}^2}}, \tag{5.8}$$

where **G** is here treated as the Jacobian of all possible 4-electrode observations. S_j conveys the relative cumulative sensitivity of the j^{th} model cell with respect to the overall cumulative sensitivity of *all* model cells queried by all possible 4-electrode observations. For the design exercise, we retain only those model cells whose S_j values at or above the α^{th} percentile. For example, Figure 5-2.b shows the spatial distribution of cumulative sensitivities that are at or above the 99.9^{th} percentile (red re-

gion). In short, 99.9% of all the sensitivity available for experimentation is relegated to a small (red) region directly beneath the survey.

5.3.3 Compressing the Jacobian Matrix

Rather than *truncating* Jacobian matrices, an alternate idea is to *compress* them. Below, we develop a mathematical trick that allows the size of the Jacobian of the base experiment to be compressed for the purpose of experimental design.

To begin, consider the Jacobian matrix of all possible pole-pole arrays over a homogeneous medium, denoted $\mathbf{G}^{(2)}$, where the '(2)' superscript signifies that two electrodes are used to make each observation – one source and one receiver pole (*Note: homogeneity is not required for the following mathematics, but a demonstration of the upcoming ideas is made using a homogeneous earth*). For a survey of N electrodes, there are ${}_N C_2 = N(N-1)/2$ unique pole-pole array combinations, so $\mathbf{G}^{(2)}$ must have $N(N-1)/2$ rows. Let there be M model cells, and assume that there are at least as many of them as pole-pole arrays (which will almost always be the case once boundary blocks are added). Then the rank of $\mathbf{G}^{(2)}$ must be equal to or less than $N(N-1)/2$:

$$\text{rank}\left(\mathbf{G}^{(2)}\right) \le N(N-1)/2, \tag{5.9}$$

because the rank of any matrix cannot exceed the lesser of its number of columns and rows. That is, $\mathbf{G}^{(2)}$ has $N(N-1)/2$ rows and M columns, where $M > N(N-1)/2$, so its rank does not exceed $N(N-1)/2$. Now consider the singular value decomposition of,

$$\mathbf{G}^{(2)} = \mathbf{U}\boldsymbol{\Sigma}\mathbf{V}^T. \tag{5.10}$$

It follows that there are only $N(N-1)/2$ nonzero singular values in $\boldsymbol{\Sigma}$ because this is the formal definition of rank. Without loss of generality, $\mathbf{G}^{(2)}$ can equally be expressed by its truncated SVD:

$$\mathbf{G}^{(2)} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T, \tag{5.11}$$

where $\tilde{\mathbf{U}}$ and $\tilde{\mathbf{V}}$ are truncations of **U** and **V**, retaining only the first $N(N-1)/2$ columns, and where $\tilde{\mathbf{\Sigma}}$ is the truncation of **Σ**, retaining the first $N(N-1)/2$ rows and columns.

Here is where the mathematical trick comes in. We transform $\mathbf{G}^{(2)}$ by right multiplying it with $\tilde{\mathbf{V}}$, creating a transform matrix

$$\tilde{\mathbf{G}}^{(2)} = \mathbf{G}^{(2)}\tilde{\mathbf{V}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}\tilde{\mathbf{V}}^T\tilde{\mathbf{V}} = \tilde{\mathbf{U}}\tilde{\mathbf{\Sigma}}. \tag{5.12}$$

Whereas the dimensions of $\mathbf{G}^{(2)}$ are

$$\dim\left(\mathbf{G}^{(2)}\right) = {}_N C_2 \times M, \tag{5.13}$$

the dimensions of $\tilde{\mathbf{G}}^{(2)}$ are reduced to

$$\dim\left(\tilde{\mathbf{G}}^{(2)}\right) = {}_N C_2 \times {}_N C_2, \tag{5.14}$$

recalling that $M > {}_N C_2 = N(N-1)/2$.

$\tilde{\mathbf{V}}$ is an incomplete orthonormal basis. It does not span all of model space, but it does span the same model subspace spanned by $\mathbf{G}^{(2)}$. Moreover, and this is the critical point, any set of orthonormal vectors (no matter whether they form a complete basis or not) form a rotation matrix, and rotation matrices have the nice property that they preserve the lengths and angles between vectors (e.g., Golub & Van Loan, 1996; Jolliffe, 2002; Strang, 2003). In short, in Equation (5.12), $\tilde{\mathbf{V}}$ rotates the row-vectors in $\mathbf{G}^{(2)}$ but it preserves their lengths and the relative angles between them. In other words, the L2-norm of any row-vector in $\tilde{\mathbf{G}}^{(2)}$ equals the L2-norm of the corresponding row in $\mathbf{G}^{(2)}$, and the angle between any two row-vectors in $\tilde{\mathbf{G}}^{(2)}$ equals the angle between the corresponding rows in $\mathbf{G}^{(2)}$. But Equation (5.14) indicates that there are fewer columns in $\tilde{\mathbf{G}}^{(2)}$ than in $\mathbf{G}^{(2)}$, so the act of right multiplying $\tilde{\mathbf{V}}$ with $\mathbf{G}^{(2)}$ *compresses* the information in the row-vectors of $\mathbf{G}^{(2)}$ to the more compact row-vectors in $\tilde{\mathbf{G}}^{(2)}$ *without any loss or distortion of information*.

The Jacobian of all quadrupole combinations is given by

$$\mathbf{G}^{(4)} = \mathbf{P}\mathbf{G}^{(2)}, \tag{5.15}$$

where **P** is a sparse *picker matrix* of ±1s that linearly combines pole-pole arrays to create quadrupole arrays. The transform in Equation (5.12) commutes without any loss of generality to produce the transform matrix for $\mathbf{G}^{(4)}$:

$$\tilde{\mathbf{G}}^{(4)} = \mathbf{P}\tilde{\mathbf{G}}^{(2)} = \mathbf{P}\mathbf{G}^{(2)}\tilde{\mathbf{V}} = \mathbf{G}^{(4)}\tilde{\mathbf{V}} \tag{5.16}$$

As before, whereas the dimensions of $\mathbf{G}^{(4)}$ are

$$\dim\left(\mathbf{G}^{(4)}\right) = 3_N C_4 \times M, \tag{5.17}$$

the dimensions of $\tilde{\mathbf{G}}^{(4)}$ are reduced to

$$\dim\left(\tilde{\mathbf{G}}^{(4)}\right) = 3_N C_4 \times {}_N C_2, \tag{5.18}$$

recalling that $3_N C_4$ is the total number of unique 4-electrode combinations available for a survey using *N* electrodes (Daily *et al.*, 2004; Stummer *et al.*, 2004). Because the information in the row-vectors of $\tilde{\mathbf{G}}^{(2)}$ is losslessly compressed, the row-vectors in $\tilde{\mathbf{G}}^{(4)}$ are also losslessly compressed, as they are just linear combinations of the rows in $\tilde{\mathbf{G}}^{(2)}$. To appreciate this, consider the last term in Equation (5.16); $\tilde{\mathbf{V}}$ rotates the row-vectors of $\mathbf{G}^{(4)}$, but it preserves lengths and angles just as it did for $\tilde{\mathbf{G}}^{(2)}$.

The two key 'metrics' in our sequential OED technique are information magnitudes and complementarity, both of which have been preserved under the simple compressive, linear transform described above. Consequently, the lossless compression of Jacobian matrices can greatly expedite the design algorithm, because the nullspace matrices are reduced from dimensions of $M \times M - n$ to only ${}_N C_2 \times {}_N C_2 - n$, where *n* is the number of observation in the base experiment. For example, for a 20-electrode survey with 21 × 51 model cells, the nullspace matrix of an uncompressed *n*-observation Jacobian has dimensions of 1071 × (1071 – *n*). The nullspace matrix of the corresponding *compressed* Jacobian has dimension of 190 × (190 – *n*). The design algorithm must calculate the nullspace and multiply it by all remaining observation

sensitivity kernels each time a new observation is to be added to the base experiment. Clearly, the significant reduction in size of the nullspace matrix will profoundly expedite the design process.

5.4 To Truncate, To Compress or Neither

To assess the competing Jacobian reduction schemes above, four experiments were designed for a 20-electrode survey querying a homogeneous earth discretized into 21 × 51 model cells (see Figure 5-1). One survey was designed using the 'Unaltered Jacobian'; the second was designed using the 'Truncated Jacobian Method 1', truncating the Jacobian to retain only the sensitivities of 19 × 9 model cells in the target window; the third was designed using the 'Truncated Jacobian Method 2', truncating at the 99^{th} percentile, similar to Figure 5-2.b; the third was designed using the 'Compressed Jacobian Method'. The 'Unaltered Jacobian' had dimensions of 14,535×1071, the 'Truncated Jacobian Method 1' dimensions of 14,535×171, the 'Truncated Jacobian Method 2' dimensions of 14,535×339, and the 'Compressed Jacobian' dimensions of 14,535×190.

CPU times were recorded for optimal experimental designs from 1 to more than 700 observations for each of the design methods and are reported in Figure 5-3. Relative to the 'Compressed Jacobian' and 'Truncated, Method 2' methods, which are equally the fastest, the 'Truncated, Method 1' and 'Unaltered' methods take an average of 2.93 and 24.14 times longer to execute for designs of the same size. Clearly, all Jacobian reduction schemes significantly improve design times relative to the 'Unaltered Jacobian' method, making them desirable alternatives for the OED exercise. It is also worth noting that the design times for all reduction schemes execute on the order of seconds to minutes, whereas the 'Unaltered' approach executes on the order of tens of minutes to hours. This suggests that any of the three reduction techniques is truly executable in the field, with only a small amount of time required to design optimal experiments.

Expedited design times mean little if they do not produce experiments that rival or exceed the performance of experiments designed with the 'Unaltered Jacobian' method, however. A Monte Carlo examination was therefore undertaken to compare the quality of the expedited designs compared with the 'Unaltered' design. Experiments of 700 observations were designed by each technique for the homogeneous earth in Figure 5-1 and were used to synthesize field data over the heterogeneous target shown in Figure 5-4.a. For each of the three designs, their synthetic field data were contaminated with 50 realizations of 3% Gaussian noise and inverted. The average models for each design are shown in Figure 5-4, Panels **b** – **e**. Also shown in these panels are the model % rms errors at the 10th, 50th and 90th percentiles (in brackets) for the Monte Carlo exercises, to provide an idea of the distribution of modeling errors for each design. These percentiles were evaluated based on the 50 model % rms errors for each Monte Carlo simulation, where the model % rmse is formulated in the usual way:

$$E^{(k)} = 100\sqrt{\frac{1}{M}\sum_{i,j}\left(\frac{\rho_{ij} - \hat{\rho}_{ij}^{(k)}}{\rho_{ij}}\right)^2}\ , \tag{5.19}$$

where $\hat{\rho}_{ij}^{(k)}$ is the inversion-model resistivity of the ij^{th} cell for the k^{th} noise realization, ρ_{ij} is the true resistivity of the ij^{th} cell and M is the total number of model cells. Comparison of the three percentile levels shows that the 'Compressed Jacobian' method outperforms both the 'Unaltered, Method 1' and 'Truncated Jacobian' Methods for this earth model. In addition, the 'Compressed Jacobian' outperforms the 'Truncated Jacobian Method 2' except for producing a larger median model % rmse.

The performances of the 'Truncated, Method 1' and 'Unaltered Jacobian' designs are more similar, though it is inferred that the distribution of modeling errors for the 'Truncated, Method 1' design has a more pronounced right-sided tail because its 90th percentile is larger than that for the 'Unaltered' design. Additionally, it is also inferred that the *mode* of modeling errors for the 'Truncated' design is smaller than that for the 'Unaltered' design because the median value (50th percentile) of the 'Truncated, Method 1' design is smaller. Therefore, it appears that the 'Truncated Method

1' method produces modeling errors whose distribution is more left-skewed than the distribution for the 'Unaltered' method, but its distribution also has a longer right-sided tail. This means that the 'Truncated, Method 1' design has a higher probability of producing a smaller model % rmse than the 'Unaltered' design, but that occasionally it will produce an outlier with a larger modeling error than the 'Unaltered' design. Interestingly, the performances of the 'Compressed' and 'Truncated, Method 2' techniques are similar to one another. The 'Compressed' method produces only slightly smaller 10^{th} and 90^{th} percentiles, while also producing a modestly larger median model % rms error. It is possible that, with samples of only 50 realizations, the margin of error in determining these percentiles is large enough that the two methods perform nearly identically. However, visual inspection of the average models in Panels **d** and **e** seems to indicate that the 'Compressed' method does a better job delineating the boundaries and estimating the magnitude of the conductive anomaly, and this is taken as evidence that this method is in fact slightly better than the 'Truncated, Method 2' technique.

From these, it is concluded that the 'Compressed Jacobian' technique is the best of the four design methods. Not only does it have the smallest design times, it also produces the smallest modeling errors. In light of the preceding discussion, it is further concluded that both 'Truncated Jacobian' techniques are superior to the 'Unaltered Jacobian' method. Although the 'Truncated, Method 1' technique does have a higher probability of producing larger modeling errors, its smaller design time arguably outweighs the added risk of poorer performance. In short, however, the 'Compressed Jacobian' technique appears to be the one to use for 2D surface resistivity OED.

An alternate way of analyzing these Monte Carlo simulations is to establish comparative probabilities that one experiment will outperform another. To do this, we subtract the Monte Carlo modeling errors of Experiment B from Experiment A in every possible combination and determine the average number of times the difference is positive or negative. Mathematically, this is expressed as

$$p = \frac{1}{m^2} \sum_{i=1}^{m} \sum_{j=1}^{m} \delta_{ij} \,, \tag{5.20}$$

where

$$\delta_{ij} = \begin{cases} 1 \text{ if } E_i^{(A)} - E_j^{(B)} < 0 \\ 0 \text{ otherwise} \end{cases} \tag{5.21}$$

and $E_i^{(A)}$ and $E_i^{(B)}$ are the i^{th} and j^{th} Monte Carlo model % rms errors for Experiments A and B, respectively. p approximates the probability that Experiment A will produce a smaller model % rms error than Experiment B; and $1 - p$ approximates the complementary probability that Experiment B will outperform Experiment A. These probabilities do not indicate *by how much* one experiment is expected to outperform another, but this is easily determined by taking the difference between their average model % rms errors:

$$\Delta E^{(AB)} = \frac{1}{m}\sum_{i=1}^{m} E_i^{(A)} - \frac{1}{m}\sum_{j=1}^{m} E_j^{(B)}. \tag{5.22}$$

Table 5-1.a reports the comparative probabilities that each survey in this Monte Carlo exercise will produce smaller model % rms errors relative to the others, according to Equation (5.20). The table is organized column-wise, meaning each column contains probabilities that the experiment in the column header will produce smaller model errors than the experiments in the row headers. For example, there is a 36% chance that the 'Unaltered Jacobian' design will produce a smaller model error than the 'Compressed Jacobian' design. Table 5-1.b reports the difference in average model % rms error between experiments, according to Equation (5.22). Again, the table is organized column-wise; so on average the 'Unaltered Jacobian' design produces model % rms errors 10.7% larger than the 'Compressed Jacobian' design (*Note: this not 10.7%* relative *to the 'Compressed' design, but in absolute terms; that is, if the 'Compressed Jacobian' method produces a model % rms error of 20% then the 'Unaltered Jacobian' method is expected to produce an error of 30.7%*). These two tables clearly show that the 'Unaltered Jacobian' method produces a design that is *least probable* to produce smaller model errors than the other methods and that it in fact produces model errors between 5 and 10% *larger* than the other methods. Furthermore, the 'Truncated, Method 1' approach produces a design that is more than

50% probable to produce smaller model errors than any of the other techniques. However, though the 'Truncated, Method 1' approach produces a design that is 59% likely to induce smaller model errors than the 'Compressed' method, the average difference between these two techniques indicates that the 'Compressed Jacobian' technique produces model % rms errors that are 2.7% smaller. From this fact, it is more apparent that the 'Truncated, Method 1' technique produces designs that are prone to large outliers in model error. Indeed, the 'Compressed Jacobian' technique produces a design whose average model % rms error is smaller than any of the other techniques. This corroborates the previous conclusion that the 'Compressed Jacobian' technique is the best of the four design expedients.

A final remark is required before concluding this section. The reader may wonder whether the experiments designed by these competing techniques produce the same experiment, for the Monte Carlo results suggest they do not. Indeed, the three experimental designs were all different, though not entirely disjoint. It is impractical to tabulate or to sensibly plot the three experiments because they each contain 700 observations. However, it was observed that, of their 700 observations, the 'Unaltered' and 'Compressed' designs shared 213 observations in common, while the 'Unaltered' and 'Truncated, Method 1' designs shared 232 common observations and the 'Truncated, Method 1' and 'Compressed' designs shared 212 common observations. Of these three, they shared 167 common observations. It is speculated that these reduction schemes produce different experiments (especially the 'Compressed' method, which was shown to contain the same information, but compressed, as the unaltered Jacobian) partially because of numerical errors arising from the calculation of large nullspace matrices. Small errors in the elements of nullspace matrices could be amplified when these matrices multiply the sensitivity kernels of candidate observations during the design update step. This would occasion errors in the relative fitness of candidate observations, with the possible consequence that the 'wrong' observation is sometimes added to the base experiment.

5.5 Homogeneous Designs

The design of a scientific experiment typically requires some foreknowledge of its outcome insomuch as one can make such a prediction. Without this foreknowledge, experimentation must proceed by trial and error, with the experimenter optimizing the experiment over successive trials until useful results are obtained. This is no less true for geophysical exploration. Without prior knowledge of the earth's structure, it is impossible to design an optimal experiment for a *particular* target site because the experiment cannot be adapted to the site. Absent such prior information, the best we can do is to start from a homogeneous earth assumption. Afterward, if necessary, a second (third, etc.) experiment can be designed that is optimally sensitive to the particular distribution of earth media at the site.

To begin, we consider the design of 2D surface resistivity experiments for a homogeneous earth. The sequential OED method described in Chapter 2 was used to design a set of experiments optimized for a homogeneous earth model. In light of the developments in the preceding section, the 'Compressed Jacobian' method was used for all designs, as this was demonstrated to be a superior design technique. A 20-electrode spread was simulated over a homogeneous half space with electrodes at 1 m equispaced intervals. Daily *et al.* (2004) and Stummer *et al.* (2004) have shown that for a set of N electrodes there exist $3_N C_4$ (thrice 'N choose 4') unique combinations of 4-electrode transmitter/receiver combinations. Thus a total of 14,535 (three time 20 choose 4) 4-electrode data stations are available with twenty survey electrodes.

5.5.1 Expediting Design: Designing with Quadrupoles of One Type

14,535 is a significant number of candidate observations, and the design algorithm, which evaluates the fitness of *all* candidates *every* time an observation is added to the base experiment, can take some time to execute. However, as was observed in Chapter 4, these 14,535 candidate observations can be divided into three groups, according to the positioning of their electrodes. These three groups were identified as the *inter-*

nal, *external* and *straddling* quadrupole configurations. It was illustrated in Chapter 4 that the cumulative sensitivities of the three configuration types differ, with the internal and straddling types providing the greatest overall sensitivity. Moreover, the ranks of the Jacobian matrices of all observations of each type have been observed to differ. Based on these empirical results, the *maximum attainable ranks* of the Jacobian matrices are predicted as follows:

$$\begin{aligned} \text{rank}\left(\mathbf{G}_{ext}\right) &\le {}_{N-2}C_2 = \frac{1}{2}\left(N-2\right)\left(N-3\right) \\ \text{rank}\left(\mathbf{G}_{int}\right) &\le {}_{N-1}C_2 - 2 = \frac{1}{2}\left(N-1\right)\left(N-2\right) - 2, \\ \text{rank}\left(\mathbf{G}_{str}\right) &\le {}_{N-1}C_2 - 1 = \frac{1}{2}\left(N-1\right)\left(N-2\right) - 1 \end{aligned} \tag{5.23}$$

where $\mathbf{G}_{ext}$ is the Jacobian of all internal quadrupole observations, $\mathbf{G}_{int}$ is the Jacobian of all internal quadrupoles, $\mathbf{G}_{str}$ is the Jacobian of all straddling quadrupoles, and N is the number of electrodes used in the survey. To test the hypothetical ranks in Equation (5.23), the three Jacobians were generated for a homogeneous earth queried with 20 electrodes. For a 20-electrode survey, the ranks of $\mathbf{G}_{ext}$, $\mathbf{G}_{int}$ and $\mathbf{G}_{str}$ should be less than or equal to 153, 169 and 170, respectively (*Note: the ranks are predicted to be* less than or equal to *these values because, in real-world applications, the spatial resistivity distribution may preclude the Jacobian from reaching its hypothetical maximum attainable rank.*) Figure 5-5 shows the cumulative sensitivities for three experiments, each consisting of all quadrupoles of a single type. Also shown in the figure are the ranks of the Jacobian matrices, which indeed correspond with their predicted values. Notice that the straddling quadrupoles produce the greatest overall sensitivity throughout the tomogram window, followed closely by the internal quadrupoles, and trailed distantly by the external quadrupoles. This is the same result as is seen in the Chapter 4 and is taken as additional proof that 2D resistivity surveys should employ many, if not all, straddling quadrupoles. Additionally, it is also observed in the Chapter 4 that the rank of the Jacobian of all observations is predicted to be ${}_{N}C_2 - N$, which equals ${}_{N-1}C_2 - 1$, the rank of $\mathbf{G}_{str}$. Therefore, the set of straddling quadrupoles alone achieves the maximum attainable rank of the Jacobian matrix.

That is, the model space spanned by the Jacobian of all *straddling* observations is identical to the model space spanned by the Jacobian of *all* observations. Put yet another way, the sensitivity kernel of any non-straddling quadrupole is a linear combination of the sensitivity kernels of straddling quadrupoles, and therefore provides no information that the set of straddling quadrupoles itself does not provide.

An alternate perspective on internal, external and straddling quadrupoles is offered in Figure 5-7. Shown are the continuous sensitivity distributions for the three quadrupole configurations of the same four electrodes. Clearly, the external configuration produces the weakest sensitivity, while the internal and straddling configurations penetrate much more deeply and equally into the formation. In point of fact, the straddling configuration produces slightly better sensitivity at depth, while the internal configuration produces slightly better sensitivity at intermediate distances from the array. This helps explain why the external configuration is so poor for resistivity inversions and corroborates the hypothesis that the straddling configuration should produce the least rank-limited Jacobians, for it penetrates furthest into the formation, maximizing the resolving power of experiments using this configuration.

Based on the foregoing demonstration, it is concluded that experimental design can be simplified by opting to design surveys using only the straddling configuration, rather than using all types. This reduces the number of candidate observations by two thirds (for the 20-electrode case, there are 14,535 candidates in all, but only 4845 straddling quadrupoles). Hence, a design algorithm that optimizes surveys using straddling quadrupoles should be approximately 67% faster than one using all types. This hypothesis is validated in Figure 5-6, which shows the CPU times for experiments designed using all candidate observations (14,535 candidates) and using only straddling type observations (4845 candidates). Computation times are shown for experiments from 1 to over 700 observations. All CPU times were clocked using a dual 2 GHZ laptop with 2GB RAM. The mean percent difference between the two CPU curves is ~56%, which is less than the predicted 67%, but nonetheless a significant timesavings. Clearly, our sequential design technique can be greatly expedited by taking advantage of the superior quality of straddling type quadrupoles.

5.5.2 Demonstration with Noiseless Data

At this point, there is compelling evidence that 2D resistivity OED can be further expedited by truncating the set of candidate observations to include only straddling quadrupoles. To be thorough, we examined the performance of designed experiments using *all* observations versus ones designed using external, internal or straddling quadrupoles. Hereafter, we prefix the designed experiments with the labels 'All', 'Internal', 'External' and 'Straddling' to indicate the set of quadrupoles from which they were designed. In this investigation, the 'Standard' Pseudosection and ERL surveys were also considered, to provide benchmark performance levels, and five random experiments were also tried. All designed experiments were based on a homogeneous earth assumption and were given varying numbers of observations. However, the objective is to determine whether experiments designed for a homogeneous earth are satisfactory to query unknown heterogeneous targets, so the earth model in this investigation was heterogeneous. Figure 5-8 shows the earth model upon which all experiments were tried; it simulates a large cavity (5000 Ωm) embedded in a 500 Ωm medium above the water table (200 Ωm).

Figure 5-9 shows the performance curves (modeling error) for various experiments used to synthesize and invert noiseless synthetic data deriving from the heterogeneous model in Figure 5-8. If we first consider the relative performances of the 'External' (blue), 'Internal' (magenta) or 'Straddling' (green) designs, it is immediately evident that the external quadrupole configuration produces the largest modeling errors and the straddling and internal configurations produce comparable errors (though internal appears to outperform straddling when fewer observations are used and vice versa when more observations are used). This reaffirms the analysis in this and Chapter 4, which concluded that the straddling and internal quadrupoles are superior for experimental design than are the external quadrupoles. This also reaffirms *the reason that Pseudosection data produce such poor image quality is because they are created only using external-type quadrupoles*.

Compared with 'Internal', 'External' and 'Straddling' designs, the 'All' experiments (orange) perform better, no matter the number of observations used. Therefore, while the set of all straddling quadrupoles have been shown to span the same model space as the set of all possible observations, it initially appears that the flexibility to design experiments from the set of all possible observations still produces the best overall results. In such a case, the primary trade-off comes from the additional computation time it takes execute the design algorithm as demonstrated in Figure 5-6. If additional CPU time were an issue, it would be better to trade a little design quality by using only the straddling quadrupoles in exchange for faster design times. However, subsequent trials will show that experiments designed with straddling quadrupoles, *in the presence of data noise*, actually produce data sets with superior noise rejection characteristics.

Also shown in Figure 5-9 is the performance curve for randomly generated experiments (purple) of varying numbers of observations. Observations were randomly selected from the set of all possible observations. First, notice that the random experiments perform better than the External *designed* experiments. This provides yet another piece of evidence against the external quadrupole configuration. Second, notice that the difference between the performances of 'Internal', 'Straddling' and 'Random' experiments reduces as the number of observations becomes large. This demonstrates the hypothesized law of diminishing returns: as the number of observations becomes large, the relative benefit of *designing* experiments diminishes. Indeed, Figure 5-9 shows that designed experiments produce the greatest relative improvement in modeling errors (especially designs from the set all observations as compared with random experiments) when small numbers of observations are used. As the number of observations becomes large, it is inferred that nearly any experiment (except 'External') will do as well as any other; there is no need to expend additional time designing an optimal experiment if it is only marginally better than a random one. When one considers that larger experiments take longer to design (as per Figure 5-6), this law of diminishing returns is even more accentuated. Small designed experiments produce the largest relative improvements in modeling error and take the

least time to design, whereas large experiments produce the smallest relative improvements but take the most time to design. Thus, the primary benefit of our OED technique is for designing compact, information-dense 2D resistivity surveys.

Figure 5-9 also shows the performance of the 'Standard' Pseudosection, ERL and Decimated ERL surveys (red dots), which for a 20-electrode layout use 153, 383 and 765 observations, respectively. The Pseudosection survey, consisting of only external quadrupoles, performs more poorly than even a Random experiment of the same size and performs comparably with the 'External' designs. Accordingly, exploration geophysics should foreswear further use of the Pseudosection survey, except perhaps as a preliminary survey to be followed by a designed survey adapted to the target site. The ERL survey proves to be an exemplary experiment, producing a model error comparable to the 'All' design. If an adapted design will not be used secondarily, the ERL survey is commended as an excellent general survey.

5.5.3 Demonstration with Noisy Data

The examples in this section have so far only considered experimental performances in the presence of noiseless data. To be thorough, we also examine the real-world situation of noisy data. As the 'Straddling' and 'All' designs have performed best in the preceding example, they are compared with the Pseudosection and ERL surveys of 153 and 765 observations, respectively. A Monte Carlo exercise was undertaken wherein synthetic data for each survey were contaminated with 50 realizations of 3% Gaussian noise and inverted. Figure 5-10 and Figure 5-11 show the average inversion models (over 50 noise realizations) for experiments of 153 and 765 observations, respectively. Also shown in these figures are the 'Monte Carlo Parameter Uncertainty' (right-hand panels) and modeling errors at the 10^{th}, 50^{th} and 90^{th} percentiles (bracketed numbers beneath the 'M.C. Model % RMSE' header in the left-hand panels, see Equation (5.19)). The 'Monte Carlo Parameter Uncertainty' is defined as the % rms error of each model cell over all 50 noise realizations:

$$E_{ij} = 100\sqrt{\frac{1}{N}\sum_{k=1}^{50}\left(\frac{\rho_{ij} - \hat{\rho}_{ij}^{(k)}}{\rho_{ij}}\right)^2}\ , \tag{5.24}$$

where $\hat{\rho}_{ij}^{(k)}$ is the inversion-model resistivity of the ij^{th} cell for the k^{th} noise realization and ρ_{ij} is the true resistivity of the ij^{th} cell.

As expected, both the 'Straddling' and 'All' designs of 153 observations (Figure 5-10) produce significantly smaller model errors than the Pseudosection data. In point of fact, only 10% of Pseudosection inversions produced model % rms errors smaller than 37.30%, which is larger than the *median* % model rms errors for the 'Straddling' and 'All' designs. Moreover, at least 90% of the modeling errors for both designed surveys were smaller than the *median* error for the Pseudosection. These observations are further exemplified in the first columns of Table 5-2.a-b, which indicate that there is no more than an 11% probability that the Pseudosection survey will produce a smaller model error than either the 'All' or 'Straddling' surveys and will, on average, produce model % rms errors 31% larger. By now, it has been well established that the Pseudosection survey, which uses only external type quadrupoles, is a very poor resistivity survey, so it comes as no surprise that the designed surveys outperform it. Indeed, the 'Monte Carlo Parameter Uncertainties' displayed in the right-hand panels in Figure 5-10 show that the Pseudosection survey produces considerably larger modeling errors throughout the target window than either designed survey.

Having again confirmed that the Pseudosection survey is inadequate to ensure data quality, let us now compare the performances of the 'Straddling' and 'All' designs in Figure 5-10. The 10^{th} percentile of the modeling error for the 'Straddling' design is greater than that for the 'All' design, and the 90^{th} percentile is less. This indicates that the 'Straddling' design produces less variability in modeling errors than the 'All' design. Looking at the 'Straddling' and 'All' columns in Table 5-2, the 'All' design is 61% likely to produce a smaller model % rmse than the 'Straddling' design (Panel (a)), but on average the expected model errors produced by the two are identical (0.0% difference in Panel (b)). From these two pieces of evidence, it appears that the 'All' design simply produces more variability in model % rmse than the 'Straddling'

design, so while there is higher probability that the 'All' design will produce a smaller model error, this is counteracted by the fact it will also occasionally produce much larger model errors, such that the average modeling errors between the two are basically equal. This is corroborated to some degree by considering the spatial distribution of expected modeling errors in the 'Monte Carlo Parameter Uncertainty' panels (right-hand panels in Figure 5-10). The 'All' design produces larger model errors in the vicinity of the hypothetical 'cave' and the 'Straddling' design produces larger errors in the hypothetical 'water table'. These two error regions apparently counterbalance one another so that the final expected model % rms errors for the two designs are equal. Nonetheless, the 'All' design does have a higher probability of producing the smaller model error; thus, all else literally being equal, the 'All' design is the best survey of the three designs. Hence, when a relatively small number of observations are used, it may be worthwhile to design experiments using all possible quadrupole types rather than limiting the design to only straddling quadrupoles. For designs of 153 observations, Figure 5-6 indicates that it takes only about a minute to design an experiment from the set of all quadrupoles and about half a minute for experiments designed from the restricted set of straddling quadrupoles. For such small experiments, there is debatably little difference between expending half or a whole minute optimizing an experiment so one might as well design from the set of all quadrupoles. Of course, for larger numbers of parameters, the factor-of-two difference in design time between 'Straddling' and 'All' designs becomes more important because design time naturally scales with the number of parameters. Ultimately, the user must determine whether it is worthwhile to sacrifice data quality in exchange for shorter computation time.

Turning now to designs of 765 observations (Figure 5-11), the situation has evened out somewhat. Both designed experiments still produce smaller model errors at the 10^{th}, 50^{th} and 90^{th} percentiles than the ERL survey, but the differences are far less pronounced than they were for the designs of 153 observations in Figure 5-10. Visual inspection of the 'Parameter Uncertainties' in Figure 5-11 also clearly indicates that the expected errors in the target window are considerably smaller for designed

experiments than for the ERL survey. Table 5-3 synopsizes the Monte Carlo exercise. While the average differences in modeling error are much smaller than they were for the preceding example of 153 observations (particularly with respect to the 'standard' ERL survey), these two tables clearly show that the ERL survey does not perform as well as the two designed surveys. First, the ERL survey only has 19% and 36% probabilities of producing smaller model errors than the 'Straddling' and 'All' designs, respectively. Second, the ERL survey produces average model errors at least 7.8% larger than the two designed experiments. From these facts, it is evident that the ERL survey produces greater variability in modeling error than either designed experiment, indicating that the experimental designs produce data with better noise rejection characteristics than the ERL survey. As regards the relative performances of the 'Straddling' and 'All' designs of 765 observations, the 'Straddling' design is 71% likely to produce a smaller model error than the 'All' design, according to Table 5-3.a, and model % rms errors 3.5% smaller, according to Table 5-3.b. The superiority of the 'Straddling' design is further borne out by the percentiles in Figure 5-11, which indicate that the 'Straddling' design produces smaller model errors at all three percentile levels. Visual inspection of the average 'Straddling' and 'All' models, as well as the 'Parameter Uncertainties' also strongly support a conclusion that the 'Straddling' survey produces data with superior noise rejection characteristics. Hence, it is concluded that the 'Straddling' design is the best of the three experiments of 765 observations in this Monte Carlo exercise.

5.5.4 A Curious Reversal

An interesting reversal has occurred in Figure 5-11 and Table 5-3, with respect to the performances of the 'Straddling' and 'All' designs. In Figure 5-10 and Table 5-2, the 'All' design of 153 observations was observed to outperform the competing 'Straddling' design of the same size, but the opposite is observed for experiments of 765 observations. In the 765-observation case (Figure 5-11), model error percentiles for the 'Straddling' design were smaller than those for the 'All' design, and the corresponding 'Parameter Uncertainties' errors in the right-hand panels show that the

'Straddling' design produced smaller expected errors in the target area. Moreover, the Monte Carlo synopses in Table 5-2 and Table 5-3 show the same pattern, with the 'All' survey appearing to be superior for small numbers of observations and the 'Straddling' design being superior for large numbers. This result seems incongruous, for while the set of straddling quadrupoles has been shown to span the same model space as that spanned by the set of all quadrupoles, one would expect that the greater variety of observation combinations afforded by using all possible quadrupoles would ensure a superior experimental design.

A possible explanation for this incongruity is the nature of our greedy design algorithm. A greedy algorithm, by definition, sequentially finds *locally* optimal updates to a multivariable optimization problem under the assumption that the final solution will be (nearly) globally optimal. Contextually, 'locally optimal' means each observation in a design is 'optimal' with respect to the sequence of observations preceding it in the base experiment but not necessarily with any antecedent observations. In this sense our design methodology is greedy and deterministic. This determinism means that a design of 153 observations is automatically a subset of a design of 765; in the latter, the leading 153 observations in the sequence of 765 are identically the observations in the former. This makes the situation all the more perplexing because Figure 5-10 indicates that the 'All' design of 153 observations outperformed the 'Straddling' design, but as additional observations were added to these two experiments their relative performances apparently reversed.

Remember that the design algorithm chooses the candidate whose sensitivity kernel is of maximal length after projection onto the nullspace of the Jacobian of the base experiment. Additional observations added to the 'All' design of 153 observations have satisfied this objective, but perhaps, in a manner of speaking, they have satisfied the objective too well. Here is the idea. The projection of a sensitivity kernel onto the *nullspace* of the Jacobian matrix is equivalent to subtracting from it its projection onto the *Jacobian matrix itself*. That is,

$$\mathbf{N}^T\mathbf{g} = \mathbf{g} - \mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{g}, \tag{5.25}$$

where **G** is the Jacobian matrix, **N** is the nullspace matrix of **G** and **g** is the sensitivity kernel of a candidate observation. The term, $\mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{g}$, in Equation (5.25) contains the components in **g** that project onto the space spanned by the Jacobian matrix. The information in $\mathbf{G}\left(\mathbf{G}^T\mathbf{G}\right)^{-1}\mathbf{G}^T\mathbf{g}$ is not in the nullspace of **G** but in the space spanned by **G**. Therefore, it is not *new* information but a repetition of information already provided by the base experiment. The design algorithm ignores this term because it only concerns itself with the residual, after this term has been subtracted from the sensitivity kernel. Effectively, the algorithm decomposes a sensitivity kernel into vectors of 'old' and 'new' information, but it evaluates candidacy based solely on the 'new' information vector. However, each time an observation is added to the base experiment, the so-called 'old' information in the sensitivity kernel still exists and accentuates the information already provided by the base experiment. These ideas are demonstrated in the diagram in Figure 5-12. The vector, $\mathbf{g}_{old}$, contains those components of **g** that reside in the space spanned by **G** and $\mathbf{g}_{new}$ contains the components of **g** that reside in the nullspace of **G** (which, by definition, is orthogonal or *complementary* to the space spanned by **G**). Therefore, the sum of the auto-inner products of $\mathbf{g}_{old}$ and $\mathbf{g}_{new}$ equals the inner product of **g** with itself, showing that the energy in **g** is *partitioned* between 'old' and 'new' information.

The preceding development provides a framework to explain the apparent superiority of the 'Straddling' design of 765 observations. As the 'All' and 'Straddling' designs sequentially amass observations, it could be that the observations added to the 'All' design provide less 'old' information than their counterparts for the 'Straddling' design. So, while additional observations in the 'All' design may add more 'new' information than their counterparts in the 'Straddling' design, they may also be adding *less* 'old' information. In other words, there may be some benefit imparted by the straddling quadrupole configuration in that each observation added to the 'Straddling' design better accentuates 'old' information. Because this is not explicitly part of the experimental design objective, it is an ancillary benefit created by designing with straddling quadrupoles.

The preceding concepts are examined in Figure 5-13. Panel (a) shows the partitioning of energy between 'Old' and 'New' information for consecutive observations in both the 'All' and 'Straddling' designs. Though the two sets of curves become jagged as the number of observations increase, the 'Straddling' design appears to match the 'All' design with respect to the amount of 'new' information each observation brings to the experiment. However, the observations in the 'Straddling' design clearly contribute more 'old' information for large numbers of observations. This assessment is corroborated in Panel (b), where the two sets of curves have been integrated. In this panel, it is apparent that the 'All' design produces slightly more 'new' information than the 'Straddling' design, which explains why the 'All' design of 153 observations performs better. But the 'Straddling' design quite evidently produces more 'old' information as the number of observations becomes large. This means that the 'Straddling' design will overtake the 'All' design in terms of data quality because, ultimately, the former produces greater total information.

5.5.5 Finding a Pattern in Designed Experiments

Overall, it has been demonstrated that experiments designed for a *homogeneous* earth and employed on a heterogeneous site outperform comparably sized Random experiments and perform as well or better than Standard experiments like the Pseudosection and ERL surveys, regardless whether the data are noisy or clean. The next question of interest is whether the observations in these designs can be organized in a manner to reveal a pattern that can be easily communicated. If such a pattern could be discerned, it would be of significant utility because it would allow us to easily scale experiments up or down according to how many electrodes are used. In other words, we would have a general survey whose data quality could be assured on mathematical grounds.

Unfortunately, this pattern recognition exercise is harder than it might seem. Consider a simple experiment of only three observations as shown in Figure 5-14 (red dipole injects the source current, blue dipole measures the potential difference). From a purely physical point of view, the rules of electrostatics permit any pair of observa-

tions to be summed or differenced to produce the third ($\phi_3 = \phi_1 + \phi_2$, $\phi_2 = \phi_3 - \phi_1$, and $\phi_1 = \phi_3 - \phi_2$). This simple experiment technically produces only two bits of unique information since knowledge of any two observations implies knowledge of the third. The natural question is: if only two observations are needed to produce these two bits of information, does it matter which two we choose? For the sake of exposition, let us assume that the three observations query a homogeneous earth. Now let us consider the sensitivity kernels of the three observations. From the point of view of the inverse problem, two pairs of observations are equivalent if the angles between their sensitivity kernels and the sums of their magnitudes are equal. This stems from the fact that the angle between sensitivity kernels measures the complementarity of the information they provide and the sum of their magnitudes measures their 'total strength'. If two pairs of observations are equivalent, they should produce the same information complementarity and magnitude. Table 5-4 shows the magnitudes of, and angles between, the sensitivity kernels for the three observations. Clearly, neither the sums of magnitudes nor the angles between any two sensitivity kernels are equal. Therefore, while the *physics* allows that any pair of observations provides the same information (insomuch as they can be combined to produce the third observation), the *inversion* does not make this allowance. From an inversion standpoint, the first and third observations combine to produce the '*strongest*' information (sum of magnitudes), while the first and second combine to produce the most *complementary* information.

Returning to the pattern recognition problem, we would like to be able to inspect a designed experiment to discern a simple pattern in the quadrupoles it chooses. For example, for each current dipole in the experiment, we might hope to see a set of receiver dipoles that can be described by a simple translation of adjacent electrodes from left to right. However, in light of the preceding discussion, there is no reason to assume that the design algorithm will choose a set of observations that lends itself to such a simple description. Indeed, because data inversion submits to a stricter definition of information equality than the physics does, it would be extraordinary if designed experiments were so easily described. To prove the point, Figure 5-15 shows

the first 30 observations for the 'All' designed experiment. A discernible pattern arguably exists for the first 17 observations, but no pattern can easily be seen thereafter. Thus, while it would be useful if designed experiments (for homogeneous media) were amenable to simple description, it appears that if any pattern exists, it would take considerable effort to find it. Exacerbating this situation is the fact that we do not know *a priori* what pattern we should be looking for. Perhaps a pattern recognition algorithm or a clustering algorithm could aid this situation.

5.6 Adaptive Optimal Experimental Design

The preceding section explored the optimization of experiments designed to query a homogeneous earth but deployed on a heterogeneous target. Thus far, it has been statistically demonstrated that such 'general' experiments reliably produce smaller modeling errors than random experiments or 'standard' experiments such as the Pseudosection and ERL surveys. Naturally, the next topic of interest is the design of optimal experiments for heterogeneous targets. In particular, it would be instructive to ascertain whether optimal designs for site-specific heterogeneity can produce statistically smaller modeling errors than experiments that have simply been designed to query a homogeneous earth.

The exploration geophysicist often has no prior knowledge of the spatial distribution of the material property they attempt to image. This was the motivation for designing optimal experiments for a homogeneous earth; absent any knowledge of the true earth structure, the most neutral starting assumption is that there is no structure. Once an initial image of the target has been generated by inversion, it may be worthwhile to design a *second*, site-specific experiment that is *adapted* to the local heterogeneity. This 'second stage' survey could hypothetically reduce parameter uncertainties by generating a smarter data set with superior noise rejection characteristics, as compared with the initial experiment used to produce a 'first stage' image. This would allow us to produce a more accurate image of the target, with less uncertainty.

What is being described here is an example of Two-Stage, Adaptive OED introduced in Chapter 2 and first examined in Chapter 4.

Alternately, one can attempt real-time adaptive experimental design by toggling between design and inversion stages, as described in Chapter 2 and first investigated in the 'Small Batch' example in Chapter 4. In this section, both Two-Stage and Small-Batch AOED are implemented.

5.6.1 150-Observation Adaptive OED

A Small-Batch approach was executed as follows: 1) an initial survey of 50 observations was designed based on a homogeneous earth assumption; 2) data were synthesized (and were either subsequently left clean or contaminated with random noise) and partially inverted, with the inversion algorithm being allowed greater than four but no more than ten iterations; 3) using the updated earth model, an additional 10 observations were prescribed to the experiment, using the design algorithm; 4) the inversion step (Step 2) was revisited; 5) the algorithm then toggled between the inversion and design steps (Steps 2 through 4) until a total of 150 observations were generated. A noiseless and a noisy data example are considered in this section.

A note on implementing inversion constraints for Small Batch AOED: Equation (5.1) indicates that the inversion objective function comprises a data rms error objective and an L_1 model smoothness objective; the Lagrange multiplier, λ, governs the tradeoff between the two; for the noiseless data case, the tradeoff parameter was controlled by the Levenberg-Marquardt algorithm and, each time an inversion stage was initiated, the final value of λ from the previous inversion stage was used as the initial value in the current stage; for the noisy data case, λ was also governed by the LM algorithm but it was always reset to its maximum value at the beginning of each inversion stage, to aid inversion stability.

Figure 5-16 shows the results for a 150-observation, real-time adaptively designed survey with clean data. The true model is show in Figure 5-8. The technique does astonishingly well, converging to an excellent earth image in approximately 120 ob-

servations, and taking approximately 10 minutes to execute the coupled design/inversion procedure. For comparison, consider the performance curves in Figure 5-17. Observe that the adaptive technique produces a final model % rms error over an order of magnitude smaller than the largest homogeneously designed survey and the ERL survey.

This simple example provides compelling evidence that adaptive design techniques can be used in the field to good advantage. However, there is a very important caveat. The inversion step of this integrated approach is dependent on the model regularization scheme, which we have previously explained is an L_1-smoothness constraint. Customarily, one employs a Lagrange multiplier to govern the tradeoff between the data rmse objective and the smoothness objective and this tradeoff variable is typically dynamic, changing from iteration to iteration in a nonlinear least squares inversion (here we use Levenberg-Marquardt to control the dynamics). However, tradeoff parameter dynamism was not intended for situations where the number of data points changes from iteration to iteration. As additional data observations are added to a least squares problem, the relative weight of data error (which is formally expressed as the sum of squared data esiduals) in the inverse problem increases with respect to the smoothness constraint. Therefore, it becomes difficult to dynamically vary the tradeoff parameter in an integrated design/inversion algorithm because one cannot know *a priori* at what level to set the Lagrange multiplier at the beginning of each inversion stage. In effect, if the integrated design/inversion procedure starts to diverge, additional observations will be prescribed for a model that is moving *away* from the true earth model, which in turn may further destabilize the inversion, creating a catastrophic negative feedback.

The preceding caveat is particularly exemplified in the presence of noisy data. Noisy data can heavily bias the inversion model, especially in the initial stages of a real-time adaptive procedure when only small numbers of data have been collected. If the algorithm begins to diverge in these early stages, there may be little to prevent runaway divergence. Figure 5-18.a shows a final inversion image for the adaptive procedure in the presence of noisy data and Figure 5-18.b shows the model % rms er-

ror of the image; recall that the same adaptive algorithm was used as that for the results in the noiseless data scenario in Figure 5-16 and Figure 5-17 except the Lagrange multiplier was reset to its maximum value at the beginning of each inversion step. Though this represents a single realization of data noise, the model % rmse is significantly larger than that for comparably sized surveys shown in Figure 5-10 (see right hand panels), which were produced by inverting noisy data using homogeneously designed surveys.

For real-time adaptive procedures to be practical, some safeguard must be created to prevent divergence. Otherwise, the field operator risks wasting a good deal of time and money running a real-time adaptive scheme whose end product is worthless, possibly forcing the operator to start all over again. Perhaps a form of automatic ridge regression could be used to control the Lagrange multiplier. This method automatically picks an ideal trade-off parameter at each iteration of a nonlinear LS inversion (though it is computationally more expensive to control the tradeoff parameter with this technique) and should therefore be capable of dealing with changing numbers of observations.

5.6.2 153-Observation Two-Stage AOED

The Two-Stage AOED approach was implemented on the heterogeneous model in Figure 5-8. A first stage inversion image (Figure 5-19) was produced by deploying the 'All' design of 153 observations optimized for a homogenous earth. This image derived from the inversion of synthetic field data contaminated with 3% Gaussian noise, yielding a reported model % rmse of ~31%. This model was used to design a series of adapted, optimal experiments to be used in a second stage inversion. All second stage inversions started with the earth model in Figure 5-19 as an initial guess. Designed experiments of 153 and 765 observations were created, to facilitate comparison with the Pseudosection and ERL surveys of those respective sizes. Furthermore, designed experiments were created using both the 'Straddling' and 'All' methods described previously in this work.

In deference to the reality of data noise, all second stage data sets were contaminated with noise. This permitted us to directly compare the performances of homogeneous designs in the previous section with the heterogeneous designs in this section, in the presence of the data noise. Accordingly, a series Monte of Carlo simulations like those performed above were executed here. 50 realizations of 3% Gaussian noise were added to the data for each experimental design, and the contaminated data were then inverted and their images and modeling errors stored.

Figure 5-20 graphically shows the results of Monte Carlo simulations of second stage inversions carried out using 153 observations. As has been amply shown throughout this document the performance of the Pseudosection survey is inferior, and that fact is no less evident in this figure. Both adapted designs produce smaller modeling errors at the 90^{th} percentile than the Pseudosection survey produces at the 10^{th} percentile! A glance at the first columns in Table 5-5.a-b further confirms this; the Pseudosection survey has no more than a 3% probability of producing smaller model errors than either the adapted 'Straddling' or 'All' designs (Panel (a), labeled 'Straddling (Het.)' and 'All (Het.)'), and the average difference in model % rms error is no less than 36%, which is even worse than the expected difference in model % rmse between the Pseudosection and homogeneous 'Straddling' and 'All' designs (Panel (b), labeled 'Straddling (Hom.)' and 'All (Hom.)').

Of much more interest is a comparison of the model error percentiles between Figure 5-10 and Figure 5-20 for the 'All' and 'Straddling' designs. The 10^{th}, 50^{th} and 90^{th} percentiles for the model % rms error of the adapted designs (Figure 5-20) were between ~3% and ~6% less than those for the homogeneous ones (Figure 5-10). Indeed, columns 2 - 3, rows 4 - 5 in Table 5-5.b show that the average differences in model % rms error between homogenous and heterogeneous designs were approximately between 5% and 8%. In addition, the same columns and rows in Table 5-5.a show that the adapted experiments were no less than 60% likely to produce smaller model % rms errors than the homogeneous designs, and particularly, the adapted 'All' design ('All (Het.)') was nearly 90% likely to produce a smaller modeling error than the homogeneous 'Straddling' design ('Straddling (Hom.)'). Lastly, a visual compari-

son of the 'Monte Carlo Parameter Uncertainties' for the 'Straddling' and 'All' design in Figure 5-10 and Figure 5-20 clearly show that parameter uncertainties within the target window were smaller for the both adapted designs. This body of evidence strongly indicates that experiments that have been deliberately adapted to site-specific structure produce smaller modeling errors and less parameter uncertainty than 'standard' experiments or even experiments that have been optimized for a homogeneous earth.

Before moving on to examine the Monte Carlo simulations of 765 observations, we compare the performances of the adapted 'Straddling' and 'All' designs of 153 observations. Recall from the discussion in the previous section that the homogeneous 'All' design outperformed the homogeneous 'Straddling' design. Figure 5-20 and Table 5-5 show the same pattern for adapted experiments. Table 5-5 indicates that the adapted 'All' design is 68% likely to produce a smaller model % rmse than the adapted 'Straddling' design, and the average difference in model % rmse is 2.6%, in favor of the 'All' design. Additionally, the model error percentiles for the 'All' design in Figure 5-20 are less than the corresponding percentiles for the 'Straddling' design, and the Monte Carlo parameter uncertainties for the 'All' design are evidently smaller as well, particularly in the vicinity of the hypothetical cave. This further substantiates the conclusion that geoelectrical experiments should be designed from the set of all possible quadrupoles when a small number of observations are to be made.

5.6.3 765-Observation Two-Stage AOED

As was the case in the Homogeneous Designs section above, the disparities in the performances of standard and designed experiments became smaller for larger numbers of observations. Figure 5-21 graphically shows the results of the Monte Carlo simulations of second stage inversions executed using 765 observations. Again, as was seen in the homogeneous case previously, both adaptively designed experiments produced smaller model error percentiles at all levels compared with the ERL survey. The Monte Carlo parameter uncertainties in the right-hand panels also clearly show that the designed surveys were much better adapted to data noise rejection than was the

ERL survey. The first columns in Table 5-6 further substantiate these findings. Compared with the adapted surveys ('Straddling (Het.)' and 'All (Het.)'), the ERL survey was no more than 22% likely to produce a smaller model % rmse, and in fact produced an average model % rmse at least 11% greater than either designed survey. Hence, it has been shown that our adaptive OED method produces smart data sets with superior noise cancellation properties compared with the standard Pseudosection and ERL surveys, no matter whether the designs use 153 observations or 765.

Next, we examine the relative performances of experiments designed using a homogeneous earth model versus those adapted to the target site. Comparing the model % rms error percentiles for designed experiments in Figure 5-11 and Figure 5-21, it is evident that the adapted designs generally produced smaller error percentiles than did the homogeneous designs (except for the 90th percentile for the 'Straddling' designs). Significantly, the Monte Carlo parameter uncertainties in the right-hand panels in these two figures show that the adapted designs substantially reduced uncertainty in the target window compared with the homogeneous designs. This suggests that the adapted designs produced less variability in modeling error than did the homogeneous designs; hence, the adapted designs were less prone to data noise. This assertion is partly confirmed by considering columns 2 –3, rows 4 – 5 in Table 5-6.a-b. With the exception of the 'All (Het.)'/'Straddling (Hom.)' comparison, the adapted designs were no less than 61% likely to produce smaller model % rms errors than the homogeneous designs, and they produced average model % rms errors between ~2% and ~5% smaller than the homogeneous designs. Compared with the relative differences in performance seen for the 153-observation case just above, the relative improvements in performance for these 765-observation experiments is smaller. Once again, we see the effects of diminishing returns: the larger experiments are allowed to become, the less benefits there are in deliberately *designing* them.

Interestingly, the *homogeneous* 'Straddling' design ('Straddling (Hom.)') and the *adapted* 'All' design ('All (Het.)') perform almost identically, according to Table 5-6. Based on the 'Curious Reversal' discussion in the preceding section, this is not entirely surprising. It was shown in Figure 5-13 that, as the number of observations in-

creases, designs restricted to only straddling quadrupoles produce more 'old information' than do designs that pick from all quadrupoles. Hence, while 'All' designs initially produce more information than 'Straddling' designs, 'Straddling' designs eventually overtake the 'All' designs. This ultimately translates into smaller parameter uncertainties, as has been shown in Figure 5-11 and Figure 5-21. Despite the fact that the homogeneous 'Straddling' design of 765 observations was designed for the wrong earth model, it still apparently produced more information about the true, *heterogeneous* target than did the adapted 'All' design.

This leads naturally to the last necessary comparison before ending the section: the relative performances of adapted 'Straddling' and 'All' designs of 765 observations. Figure 5-21 shows that the model error percentiles for the adapted 'Straddling' design are less than those for the adapted 'All' design, except at the 90^{th} percentile. More importantly, visual inspection of the Monte Carlo parameter uncertainties in the right-hand panels plainly shows that uncertainties are generally smaller throughout the target window for the 'Straddling' design. Table 5-6 confirms this observation, indicating that there is a 60% likelihood that the adapted 'Straddling' design will produce a smaller model error and that its expected model % rmse is 1.8% smaller than that for the adapted 'All' design. This shows that, like the foregoing homogeneous examples, 'Straddling' designs of large numbers of observations outperform their competing 'All' counterparts. Therefore, the hypothesis still stands: for the sequential design method developed in this work, 'All' designs should be used for small experiments and 'Straddling' designs should be used for large ones.

5.7 Integration and Summary

This chapter has presented several case studies in 2D surface resistivity OED, focusing on the sequential optimal experimental design methodology developed in Chapter 2. Many of the issues germane to 4-electrode or quadripolar geoelectrical experiments have already been introduced and discussed in the Borehole OED Chapter, which we have taken advantage of here, freeing us to examine other aspects of the 2D

surface resistivity OED problem. Particularly, the 2D resistivity examples in this chapter used a large number of modeling cells and a large number of fixed survey electrodes, making the optimal ED problem especially challenging because computational expense scales with the number of model parameters and with the number of permitted observations. Below is a summary of the exercises carried out.

5.7.1 Modeling and Inversion

The two-dimensional resistivity forward and inverse problems have been addressed using the transmission line network analogy (e.g., Swift, 1971; Madden, 1972; Zhang *et al.*, 1995; Shi, 1998). Of particular interest in this chapter was the necessity to use a considerable number of boundary blocks to implement far-field boundary conditions and to ensure accurate modeling of electrical potentials in the vicinity of the survey. Though they were essentially nuisance parameters, it was necessary to invert for the resistivities in the boundary blocks, causing the inverse problem to be ill posed. In effect, the inversion had to solve for 1100 parameters while having a maximum attainable rank of only 170 (for 20-electrode surveys). More is said about this ill-posedness below. All inversions herein implemented model regularization via the L1-norm of the spatial gradient of the model. This permitted inversion models to have sharp contrasts because contrasts are not as heavily penalized by an L1-norm as they would be by an L2-norm smoothness constraint.

5.7.2 Expediting Design: Reducing the Size of the Jacobian Matrix

Aside from the design methodology itself, one of the most important aspects of this research may turn out to be the discovery of useful ways to shorten OED computation times. As mentioned above, all OED examples in this chapter had to contend with designs for a large number of model parameters (nearly 1100) as well as a large number of permissible quadripolar observations (more than 14,000). As our sequential OED method is a Jacobian-based technique, and a Jacobian must respectively have as

many rows and columns as observations and parameters, it follows that design times increase in proportion to these two. While our sequential, greedy design algorithm has shown itself to be quite fast in general, even this algorithm is impeded by the increased scope of a large design problem.

We examined three alternative methods of reducing the size of the Jacobian matrix in an attempt to expedite design times. Two truncation methods were introduced. The simplest just eliminated all columns of the Jacobian matrix corresponding to boundary blocks, arguing that their sensitivities were irremediably small and could not be redeemed no matter how many observations were made. The second method truncated the Jacobian matrix according to a percentile-based approach. Only those model cells whose cumulative sensitivities were above a user-specified percentile, α, would be retained. The underlying argument behind this approach was that we wished to retain as much overall sensitivity as possible in the design exercise. By choosing a large percentile threshold, say 99.9%, we would ensure that the set of retained model parameters would be responsible for 99.9% of all sensitivity that a data set could ever have to the model parameters. The third reduction scheme cannot be called a truncation method. Rather than simply eliminating certain parameters (columns) from the Jacobian matrix, it is possible to *compress* the number of columns in the Jacobian matrix by taking advantage of its singular value decomposition and the fact that the inverse problem is rank limited. It was shown that this act *losslessly* compresses the information that each observation provides. In brief, the sensitivity kernel of each observation occupies one row of the Jacobian matrix, and these sensitivity kernels can be losslessly compressed from vectors in 1071 dimensions to vectors in 170 dimensions. Moreover, because this compression uses an orthonormal basis (the truncated right singular vectors of the Jacobian), the magnitudes and angles between sensitivity kernels are preserved, which is critical for our design algorithm.

These three Jacobian-reduction methods were implemented in the design algorithm and compared with an experimental design that did not avail itself of any expediting strategy. It was shown that the 'Compressed Jacobian' and the percentile-based 'Truncated Jacobian' methods expedited design by a factor of more than 20, and the

simple 'Truncated Jacobian' approach expedited design times by a factor of nearly 3. At the end of the day, it is important to point out that the three reduced-Jacobian design methods execute in a few minutes, which makes them all realistically accessible in real-world field situations. Importantly, it was shown that the three reduction schemes were all more likely to produce a smaller model % rmse than the 'Unaltered Jacobian' method, and they also all produced smaller average model % rms errors. Ultimately, the 'Compressed Jacobian' method was shown to produce not only the shortest design times but also the smallest average model % rms errors, making it the most useful of the three proposed design expedients.

5.7.3 Homogeneous Designs

In this section, an examination was undertaken to determine whether experimental designs optimized for a *homogeneous* earth could be usefully applied to a heterogeneous target. The primary benefit of such an investigation would be to demonstrate that surveys designed for a homogeneous half-space could perform as well or better than either random or standard surveys. If such were the case then it would be better to use these designed experiments than the standard surveys in widespread use today.
Four topical areas were considered in this section: (1) expediting design times by designing with quadrupoles of only one type; (2) noiseless data trials; (3) noisy data trials; and (4) finding patterns in designed experiments.

Expediting design times by designing with only a single quadrupole type has already been examined in the Borehole OED Chapter, and Figure 5-5 reaffirms the assertion from that chapter that the best single quadripolar type is the straddling arrangement, followed closely by the internal arrangement and distantly by the external arrangement. Because the straddling-type quadrupole is observed to be the best of the quadrupole types, this chapter examined surveys designed using only this quadripolar type as compared with designs picking from all quadrupole types. Using only straddling quadrupoles (rather than all possible quadrupole types for OED), design times could be reduced by more than a factor of 2. But the important issue was whether faster design times also corresponded with equal or better inversion performances.

Interestingly, it was shown that when small numbers of observation were used, the 'All' design performed better than the 'Straddling' design, but when large numbers of observations were used, the 'Straddling' design did better. This was explained by noting that 'All' designs initially produced more information about the target area when small numbers of observations were used, but owing to the nature of our greedy design algorithm, 'Straddling' designs eventually produced more information about the target area as the number of observations became large. Therefore, as it imposes little additional computational expense, it is recommended that small designs should be created using all possible quadrupoles, and large designs should use only straddling quadrupoles, as these will not only ensure shorter design times but smarter data.

While the noiseless data examples in the Homogeneous Design section (Figure 5-9) indicated that designs using all quadripolar types produced smaller model errors than any other type of experiment, it was important to consider the performances of designed experiments in the presence of realistic data noise. The preceding discussion relates to this concern and has demonstrated that the actual situation, in the presence of data noise, suggests that 'Straddling' designs outperform 'All' designs when large numbers of observations are collected.

Though it would be useful if homogeneous designed experiments could be communicated in a simple manner, preferably using some sort of basic recursion, it was explained that this is an as yet unattainable goal. The physical law of superposition for electrostatics allows that the information of two observations can be linearly combined to produce a third observation or piece of information, but it was shown *from an inversion standpoint* that the information that this third observation produces is not necessarily equivalent to the information produced by the two observations from which it derived. Therefore, designed experiments can apparently give rise to unrecognizable observation patterns that contravene our wish to categorize them into simple, communicable packets.

5.7.4 Adaptive, Optimal Experimental Design

The last investigation executed in this chapter examined the adaptive design of optimal experiments tailored to a specific site. The main objective in this investigation was to determine whether there was any quantifiable benefit to adapting 2D surface resistivity experiments to a specific heterogeneous site, using our sequential design methodology, rather than just using standard surveys or experiments optimized for a homogeneous earth.

A 'Two-Stage' AOED method was adopted for these exercises, as this method had previously been shown in the Chapter 4 to produce the fastest and most reliable results. It was definitively shown in this chapter that adapted 2D surface resistivity 'Straddling' and 'All' designs outperformed homogeneously designed surveys as well as standard Pseudosection and ERL surveys.

These AOED studies also reaffirmed the observation that optimal ED is subject to a law of diminishing returns. The relative improvement in data quality was more pronounced when small numbers of observations were used, and as the number of observations increased, these relative improvements decreased.

Additionally, as was observed in the concluding remarks on Homogeneous Designs, 'All' designs outperformed 'Straddling' designs when small numbers of observations were used, but the situation reversed for large number of observations. The same result was observed in the adapted design studies. Therefore, it is concluded that this is, at least, an invariant feature of our sequential design method, and perhaps a general feature of experimental 2D resistivity design problems.

5.7.5 Conclusion

Withal, it has been shown that our sequential optimal experimental design methodology produces 2D surface resistivity experiments whose data are *smart* in the sense that they produce smaller average model errors and smaller parameter uncertainties than competing random or standard surveys. Moreover, our greedy design algorithm achieves these results at a computational expense on the order of only minutes. At

most, our technique requires a second data collection and inversion, and at least, it only requires a single data collection and inversion, depending on how important the practitioner perceives accurate and certain imaging to be.

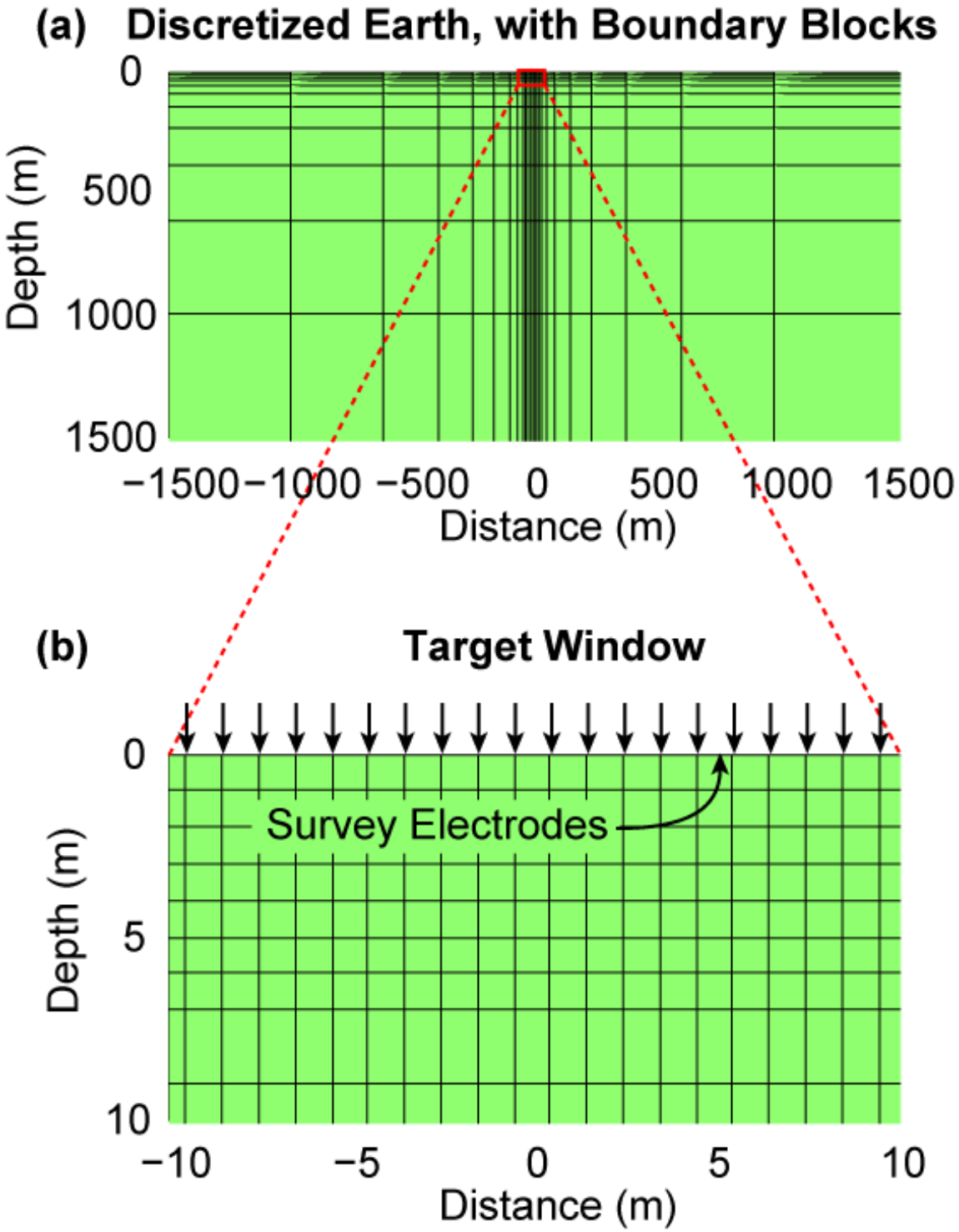


Figure 5-1 (a) Example of the discretization of a homogeneous earth and the target window (red box) within which we are concerned with properly modeling resistivity. (b) Magnification of the target window.

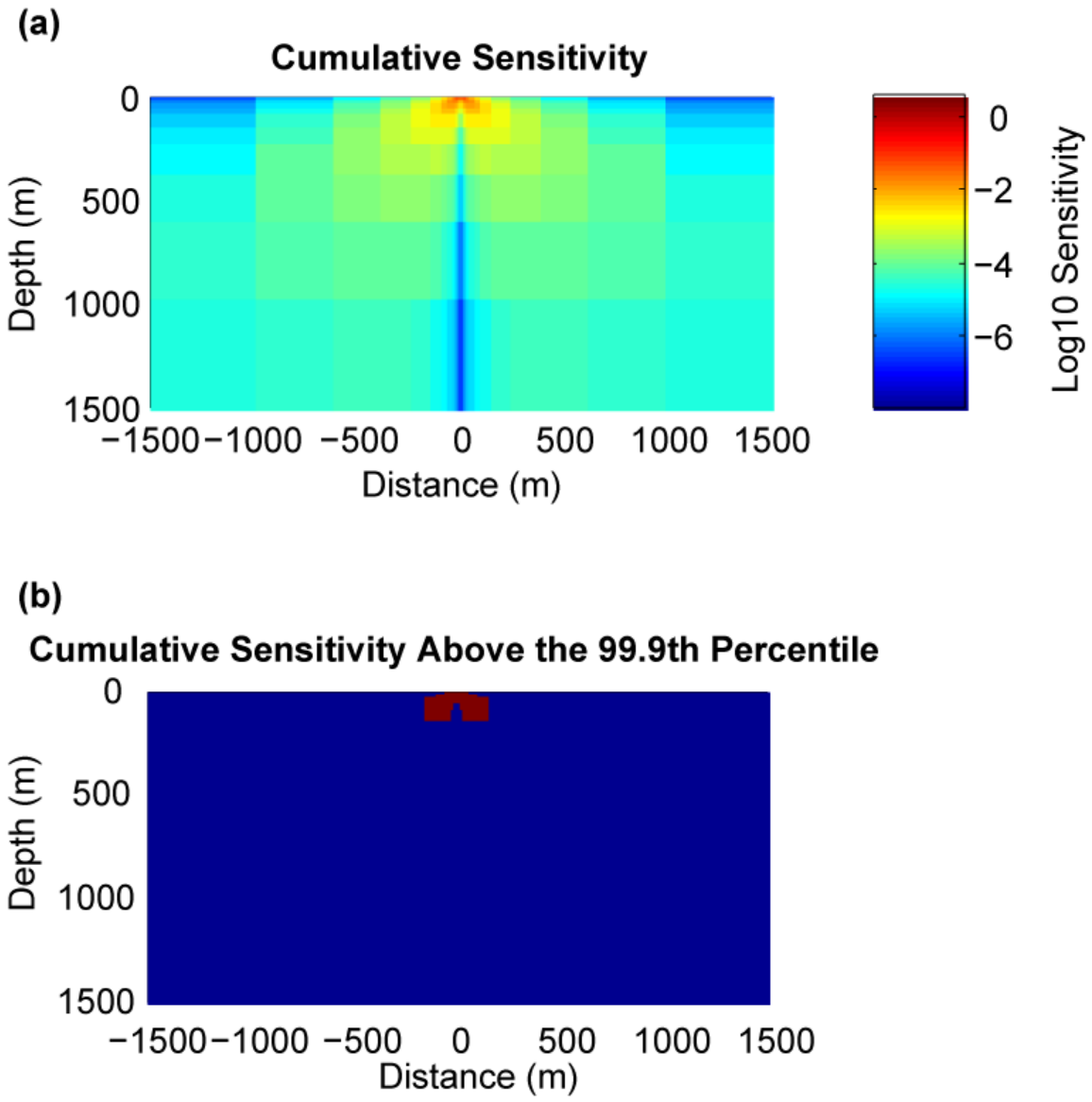


Figure 5-2 (a) Cumulative absolute sensitivities ($\log_{10}$) and (b) a binary image showing the spatial distribution of cumulative sensitivities at or above the 99.9th percentile (red region). That is, 99.9% of all sensitivity that the complete set of all observations affords is relegated to the small red region.

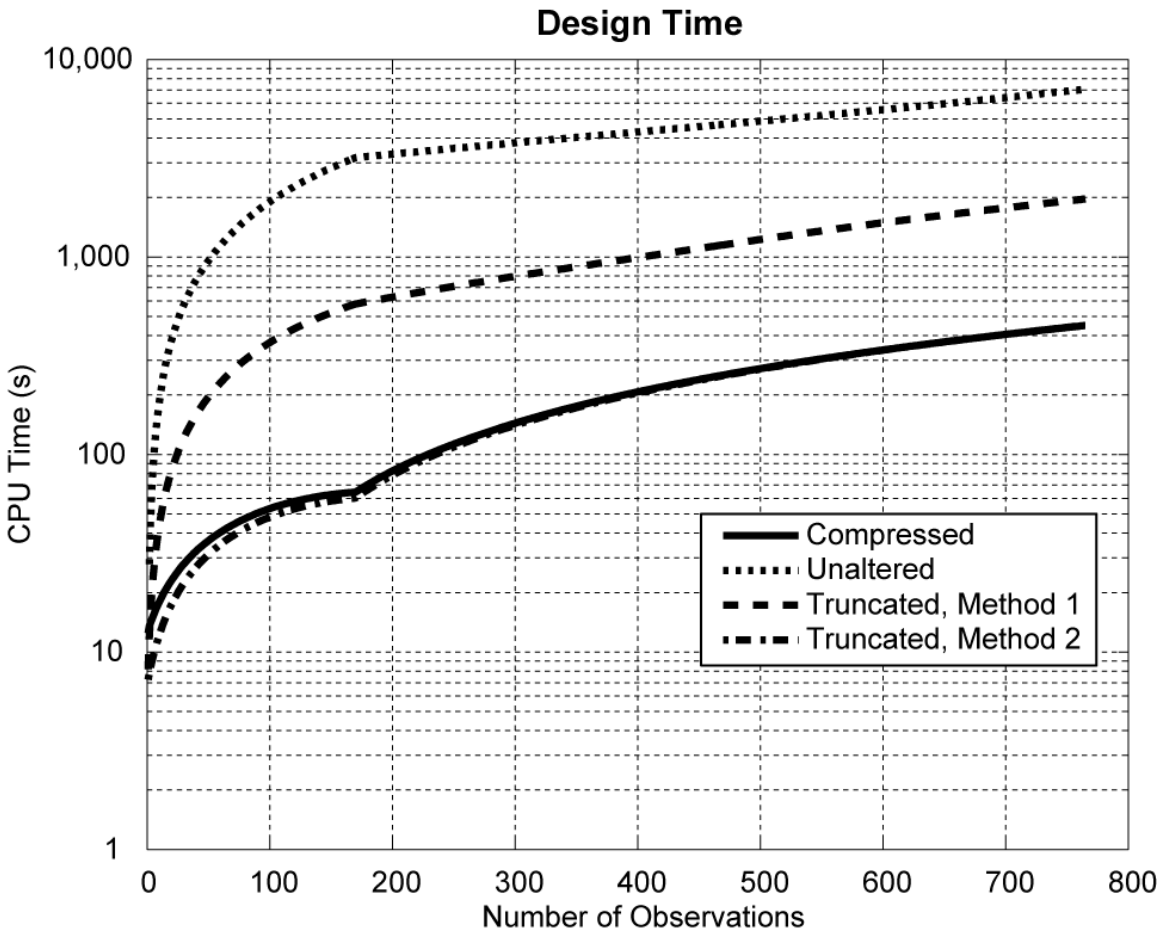


Figure 5-3 CPU times for experiments designed using the 'Unaltered Jacobian', 'Truncated Jacobian, Method 1', 'Truncated Jacobian, Method 2' (at the 99th percentile), and the 'Compressed Jacobian' methods. Note that the 'Truncated Jacobian, Method 2' and 'Compressed Jacobian' methods produced nearly identical CPU times. Also, note that the 'Unaltered' technique took nearly 2 hours (~7,000 seconds) to produce the largest experiment, while comparably sized 'reduced Jacobian' methods took ~0.5 hours and ~7.5 minutes. All CPU times were clocked on a dual 2GHz laptop with 2GB RAM.

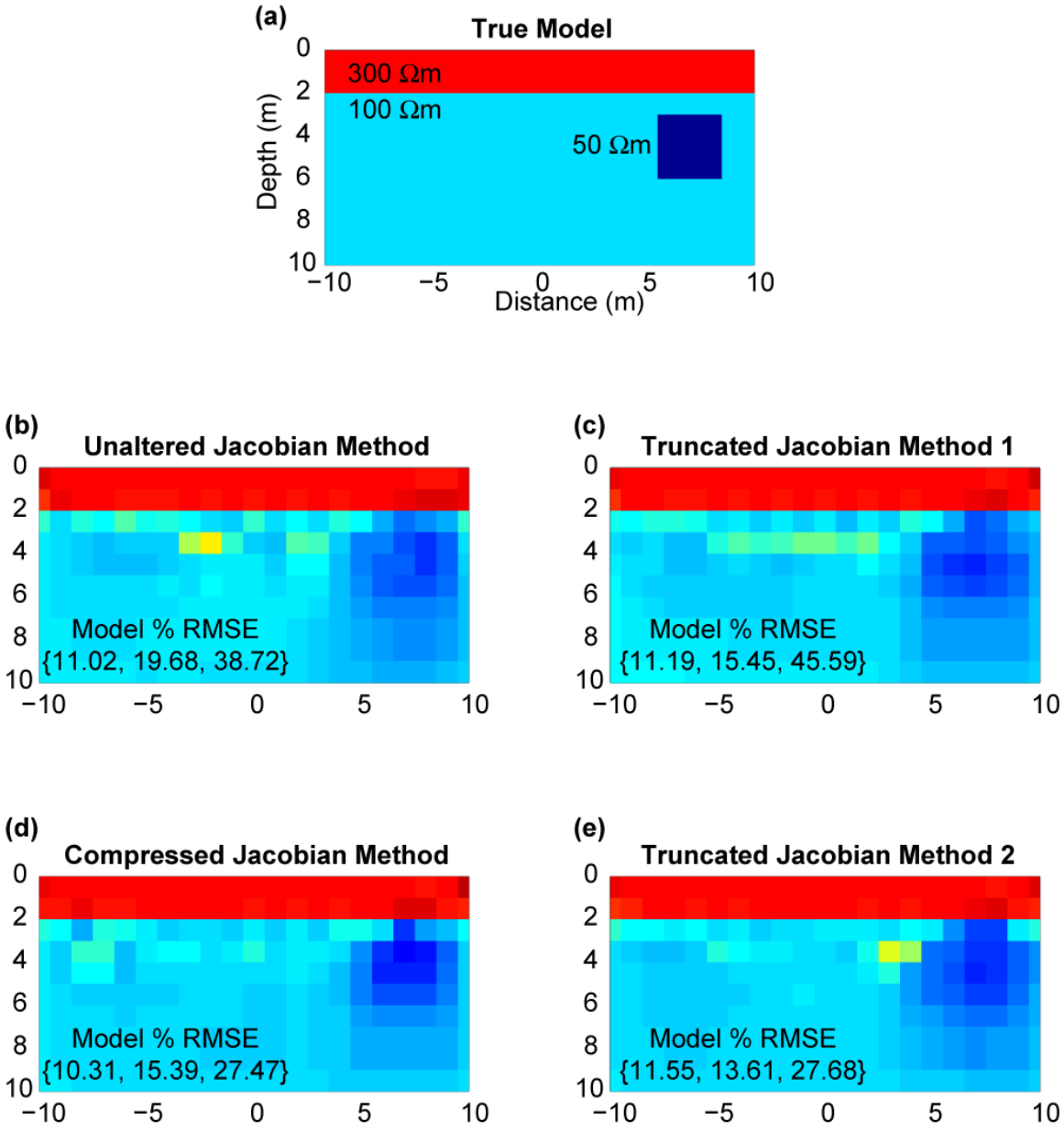


Figure 5-4 Heterogeneous model (a) and the average Monte Carlo models for the 'Unaltered Jacobian' method (b), the 'Truncated Jacobian, Method 1' (c), the 'Compressed Jacobian' (d), and the 'Truncated Jacobian, Method 2'. The bracketed numbers beneath the 'Model % RMSE' header in each panel specify model % rms errors at the 10th, 50th and 90th percentiles (generated by the Monte Carlo simulations). For example, the median model % rmse of the 'Compressed Jacobian' technique is 15.39% while it is 19.68% for the 'Unaltered Jacobian' technique.

Probability of Producing a Smaller Model % RMS Error

	Unaltered	Truncated, Method 1	Truncated, Method 2	Compressed
Unaltered		0.68	0.60	0.64
Truncated, Method 1	0.32		0.40	0.41
Truncated, Method 2	0.40	0.60		0.53
Compressed	0.36	0.59	0.47	

Difference in Average Model % RMS Error

	Unaltered	Truncated, Method 1	Truncated, Method 2	Compressed
Unaltered		-8.0%	-5.1%	-10.7%
Truncated, Method 1	8.0%		2.9%	-2.7%
Truncated, Method 2	5.1%	-2.9%		-5.5%
Compressed	10.7%	2.7%	5.5%	

Table 5-1 (a) Comparative probabilities of producing smaller model % rms error between the four design methodologies according to Equation (5.20). Each column records the probabilities that the design technique of that column will produce an experiment with smaller model % rms errors than the experiments produced by the design techniques in the competing rows. (b) Difference between average model % rms errors for Monte Carlo simulations of noisy field data for experiments designed by the four competing techniques.

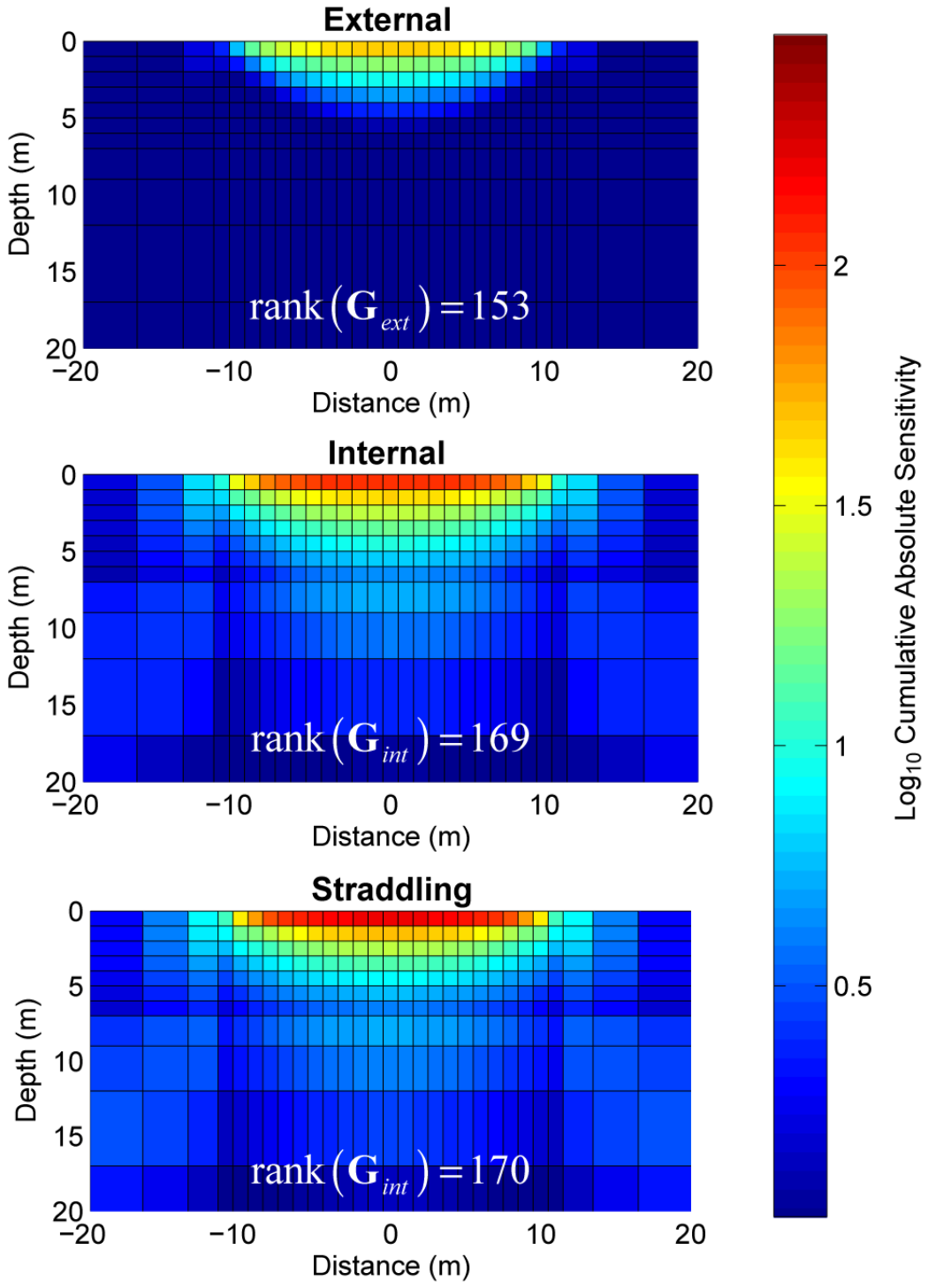


Figure 5-5 Cumulative absolute sensitivities ($\log_{10}$) of the three 4-electrode configurations – external, internal and straddling – querying a homogeneous earth. 20 survey electrodes were simulated, and as predicted in Equation (5.23), the ranks of the Jacobian matrices comprising all observations of one quadrupole type are 153, 169 and 170 for the external, internal and straddling configurations, respectively.

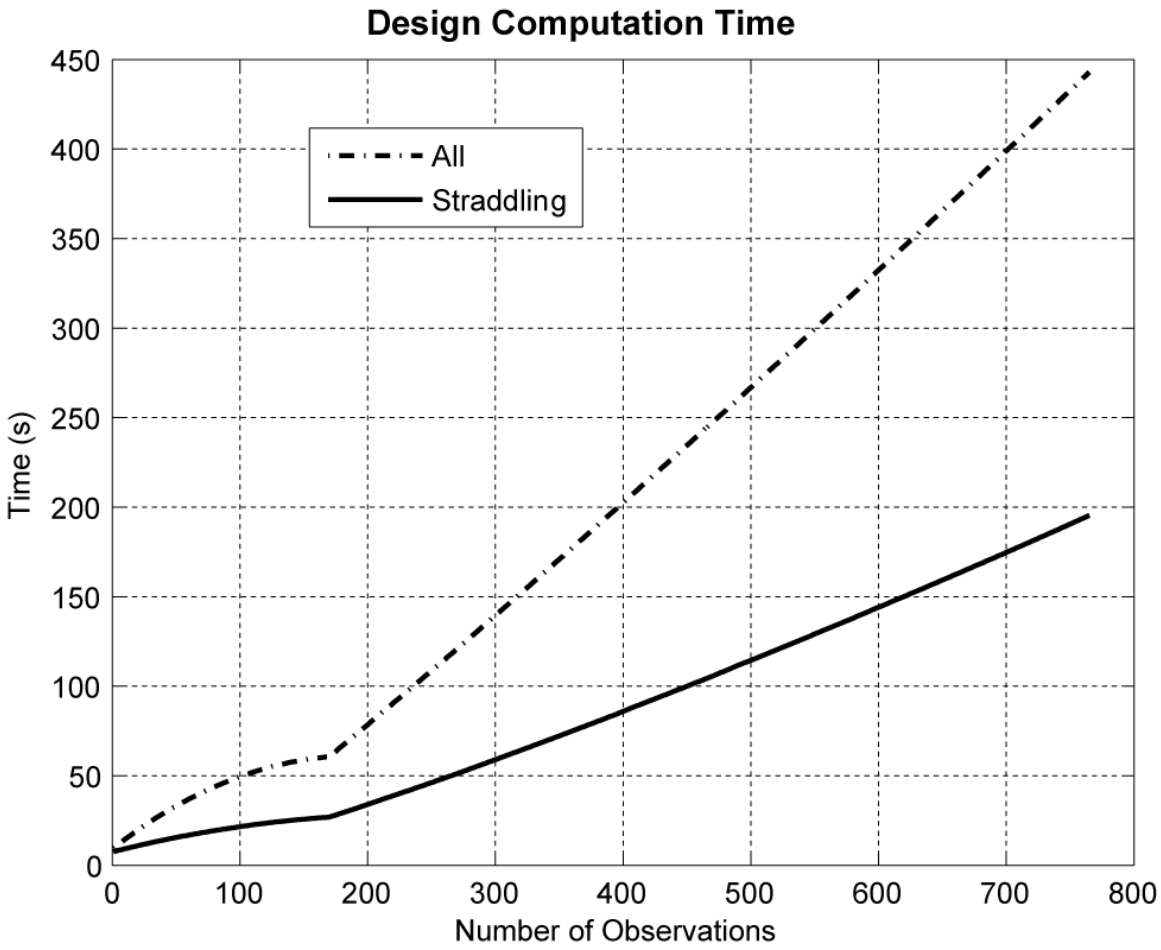


Figure 5-6 CPU times for experiments that picked from the set of 'All' quadrupoles (broken line) and from the smaller set of 'Straddling' quadrupoles (solid line). The 'Straddling' designs executed more than twice as fast on average. All CPU times were clocked on a dual 2GHz laptop with 2GB RAM.

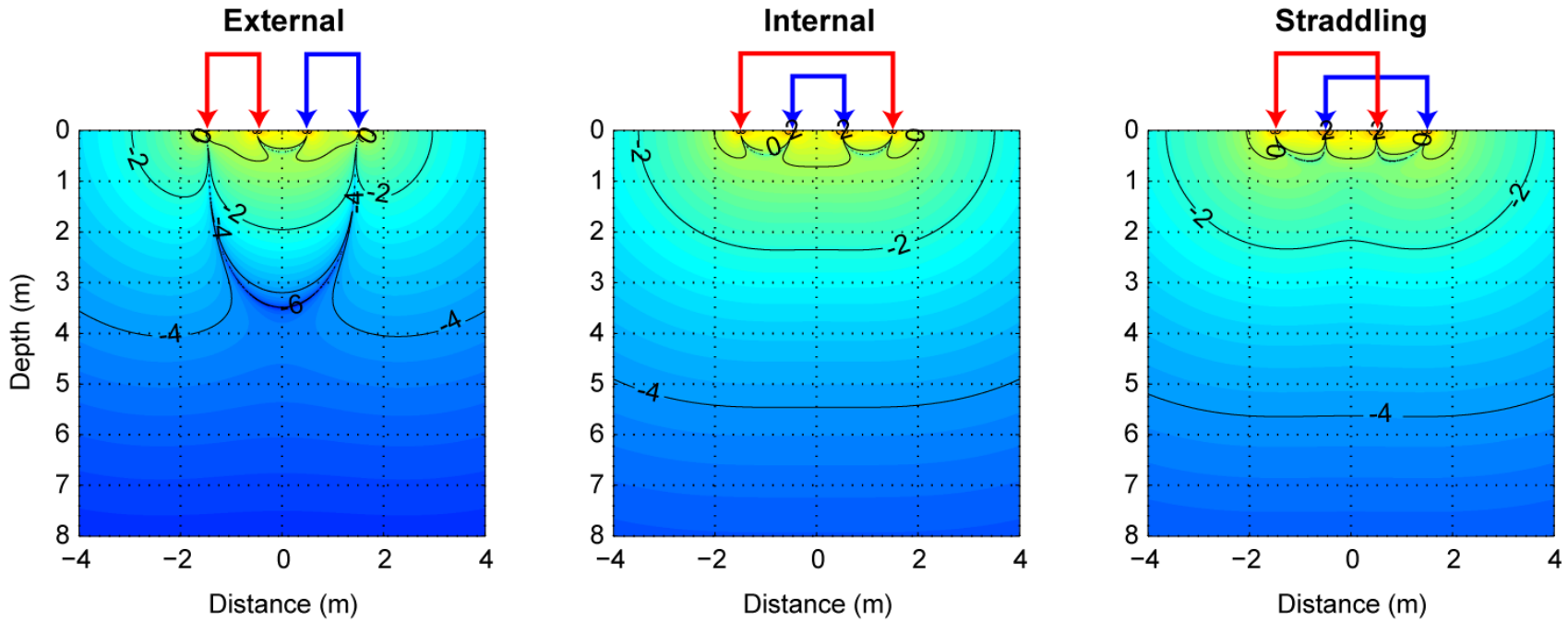


Figure 5-7 Continuous $\log_{10}$ absolute sensitivities for external, internal and straddling quadrupole configurations of the same four electrodes, querying a homogeneous earth (after (Loke & Barker, 1995)). Colorscale is the same for all three panels and contours have been added for reference. Note that the sensitivity of the external configuration drops off most rapidly as a function of distance from the array. Internal and straddling configurations distribute their sensitivities more similarly, but the straddling configuration penetrates a bit deeper while the internal configuration produces slightly better sensitivity nearer to the array (note the difference in the '-2' contour).

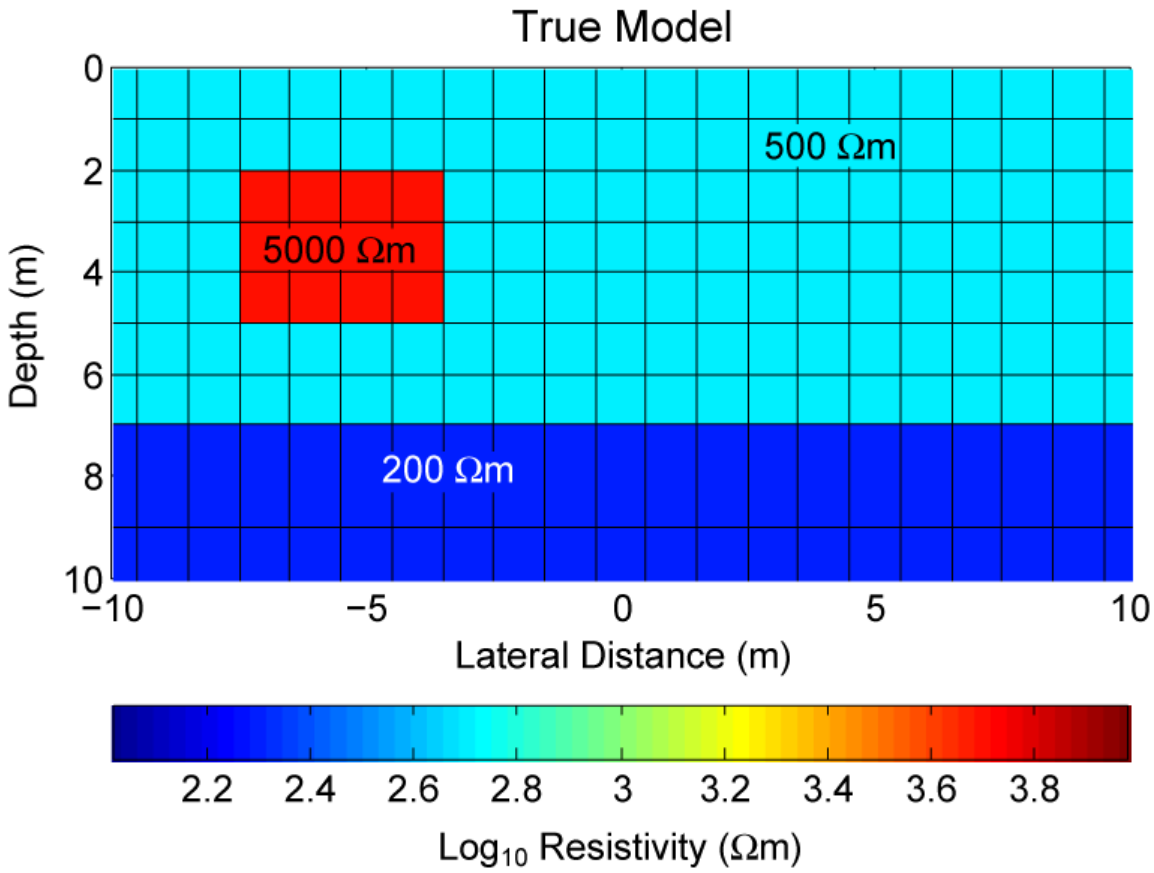


Figure 5-8 A hypothetical cave model (5000 Ωm) situated above a water table (200 Ωm).

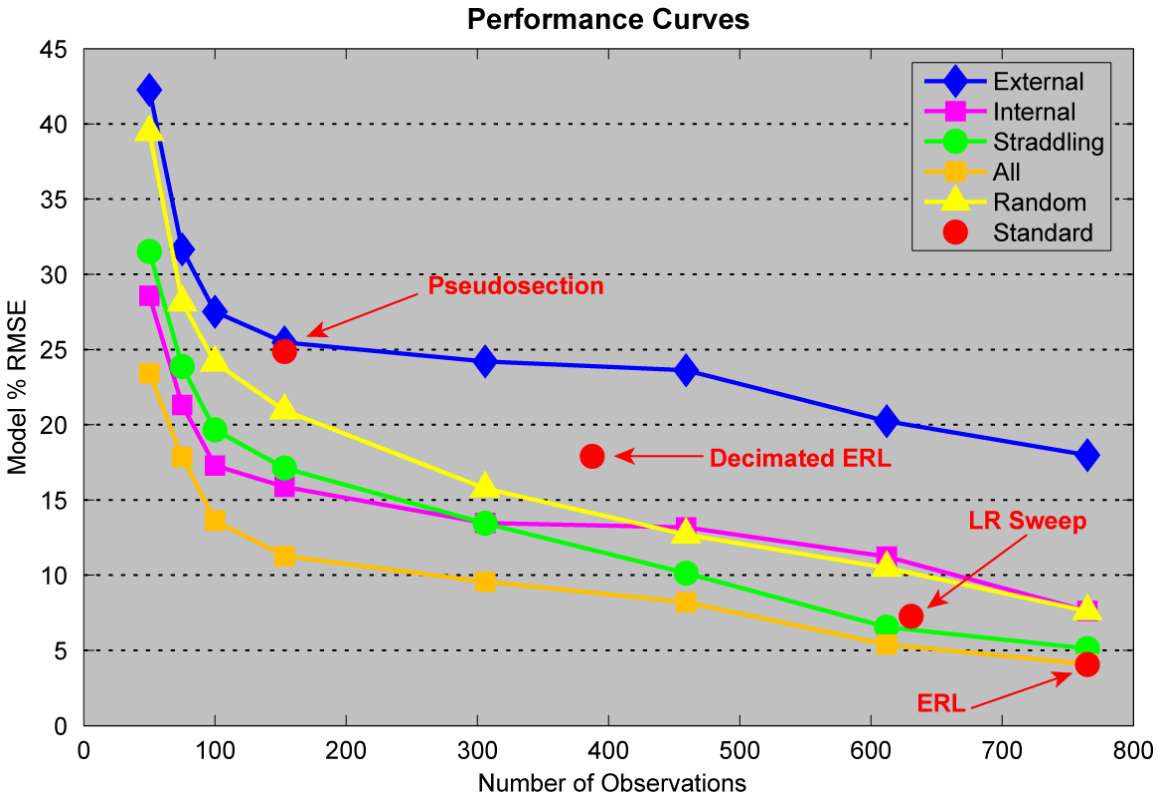


Figure 5-9 Performance curves for standard, random and 'homogeneously' designed experiments of varying sizes. Model % rms errors for the various experiments are plotted against the number of observations used. All experiments queried the hypothetical cave model in Figure 5-8 and the data were noiseless in all cases. The Pseudosection and ERL surveys are identified in the figure as well as a 'Decimated ERL' survey (every other observation removed from the ERL survey) and a 'Left-Right' Sweep' (the ERL Survey minus the Pseudosection survey) are shown betwixt.

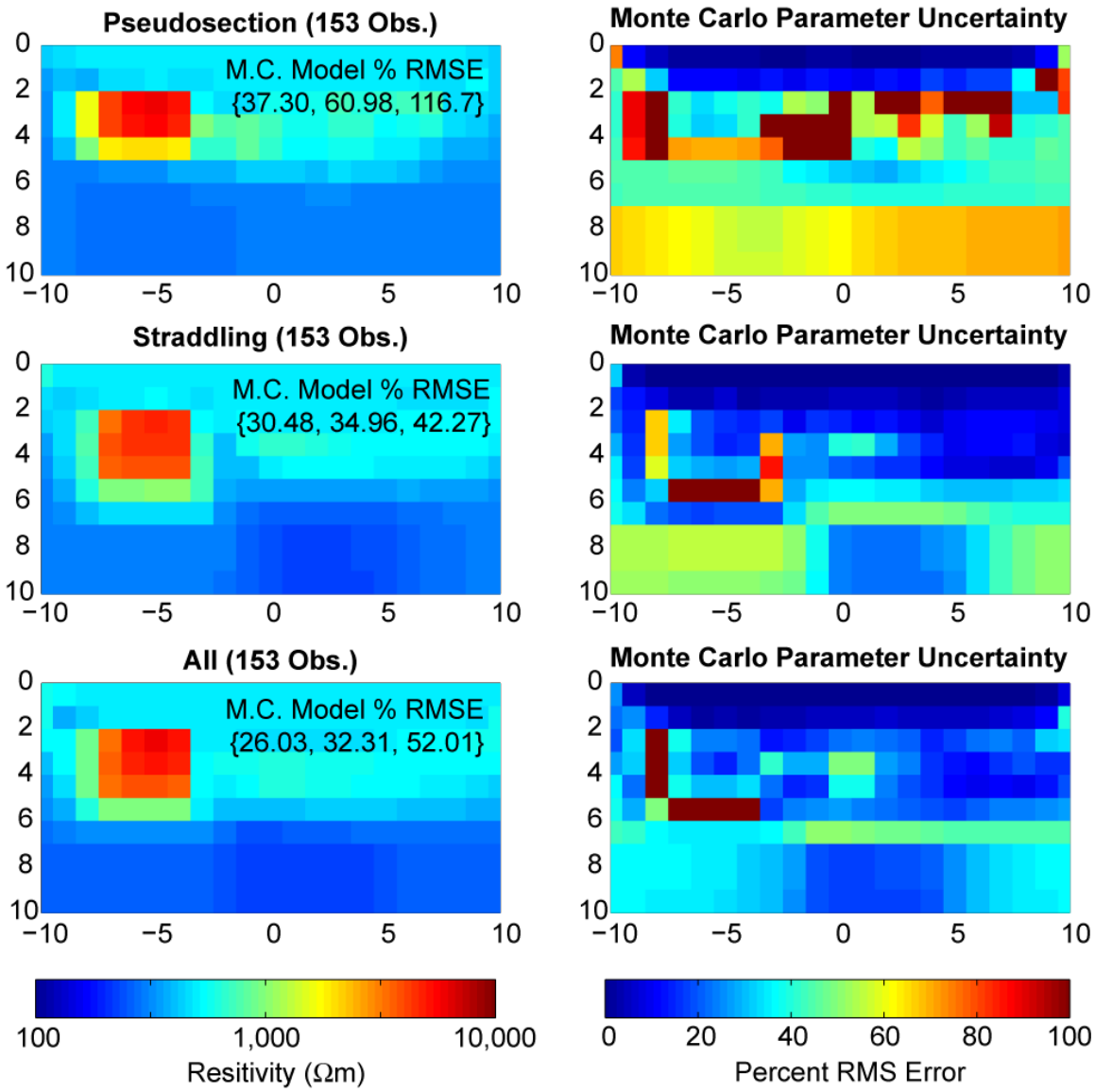


Figure 5-10 Graphical synopses of Monte Carlo simulations for data inversions with 50 realizations of 3% Gaussian noise for the three experiments, 'Pseudosection', 'Straddling' and 'All'. All experiments queried the cave model in Figure 5-8 with 153 observations. The left-hand panels show the average Monte Carlo models for each experiment over all 50 noise realizations. The bracketed numbers beneath the 'M.C. Model % RMSE' headers in each of the left-hand panels specify model % rms errors at the 10^{th}, 50^{th} and 90^{th} percentiles (generated by the Monte Carlo simulations). The right-hand panels show the 'Monte Carlo Parameter Uncertainties', measured as % rms error according to Equation (5.24), for the three experiments.

(a) **Probability of Producing Smaller Model % RMS Error**

	Pseudosection	Straddling	All
Pseudosection		0.90	0.89
Straddling	0.10		0.61
All	0.11	0.39	

(b) **Difference in Average Model % RMS Error**

	Pseudosection	Straddling	All
Pseudosection		-31.0%	-31.0%
Straddling	31.0%		0.0%
All	31.0%	0.0%	

Table 5-2 Monte Carlo comparison of two optimally designed experiments using 153 observations. The Pseudosection survey was also considered for completeness. All designs were based on a homogeneous earth assumption. (a) Comparative probabilities of producing smaller model % rms error between the three designed experiments according to Equation (5.20). Each column records the probabilities that the experiment in that column will produce a smaller model % rms error than the experiments in the competing rows. (b) Difference between average model % rms errors for the three experiments.

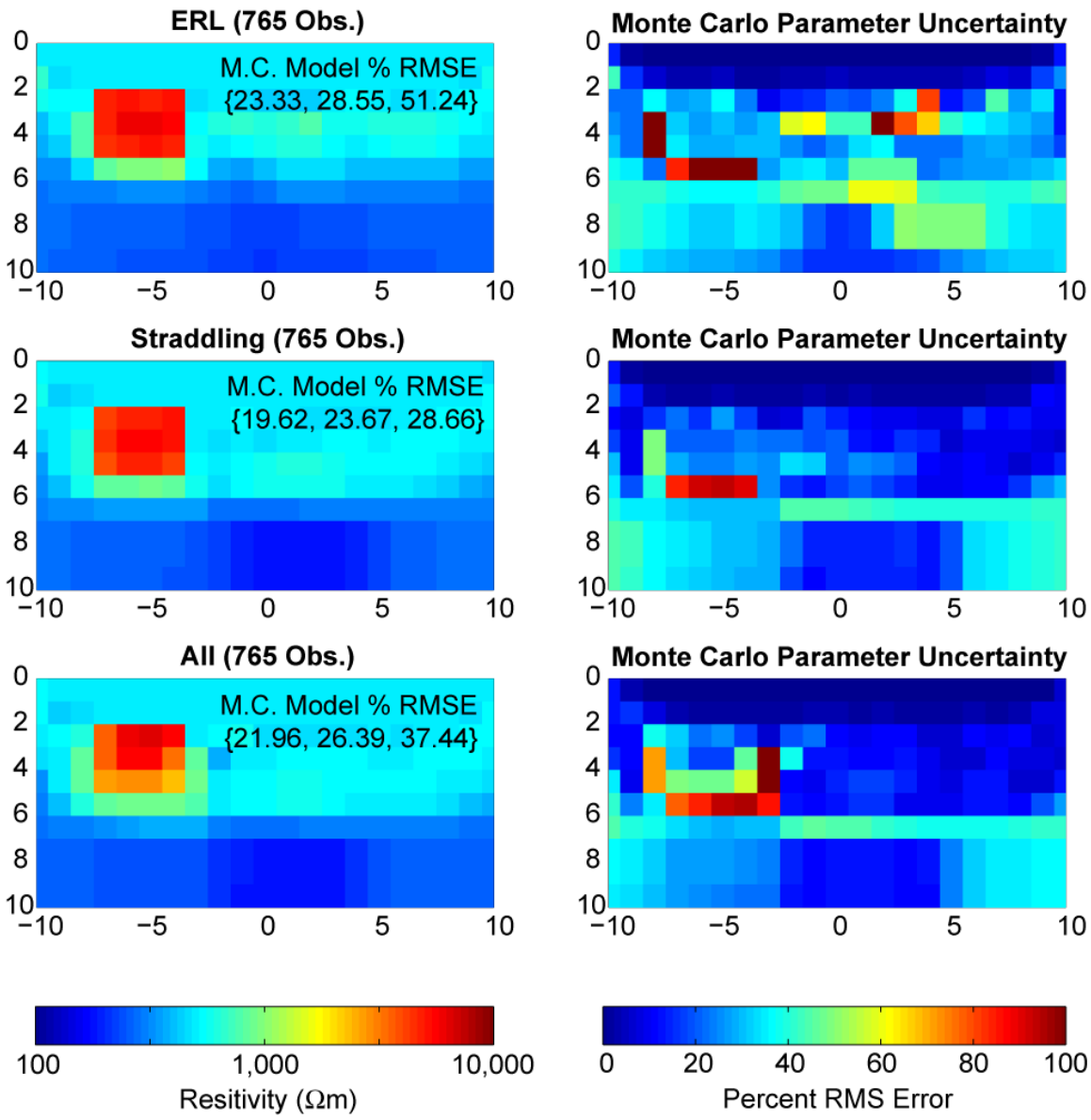


Figure 5-11 Graphical synopses of Monte Carlo simulations for data inversions with 50 realizations of 3% Gaussian noise for the three experiments, 'ERL', 'Straddling' and 'All'. All experiments queried the cave model in Figure 5-8 with 765 observations. The left-hand panels show the average Monte Carlo models for each experiment over all 50 noise realizations. The bracketed numbers beneath the 'M.C. Model % RMSE' headers in each of the left-hand panels specify model % rms errors at the 10^{th}, 50^{th} and 90^{th} percentiles (generated by the Monte Carlo simulations). The right-hand panels show the 'Monte Carlo Parameter Uncertainties', measured as % rms error according to Equation (5.24), for the three experiments.

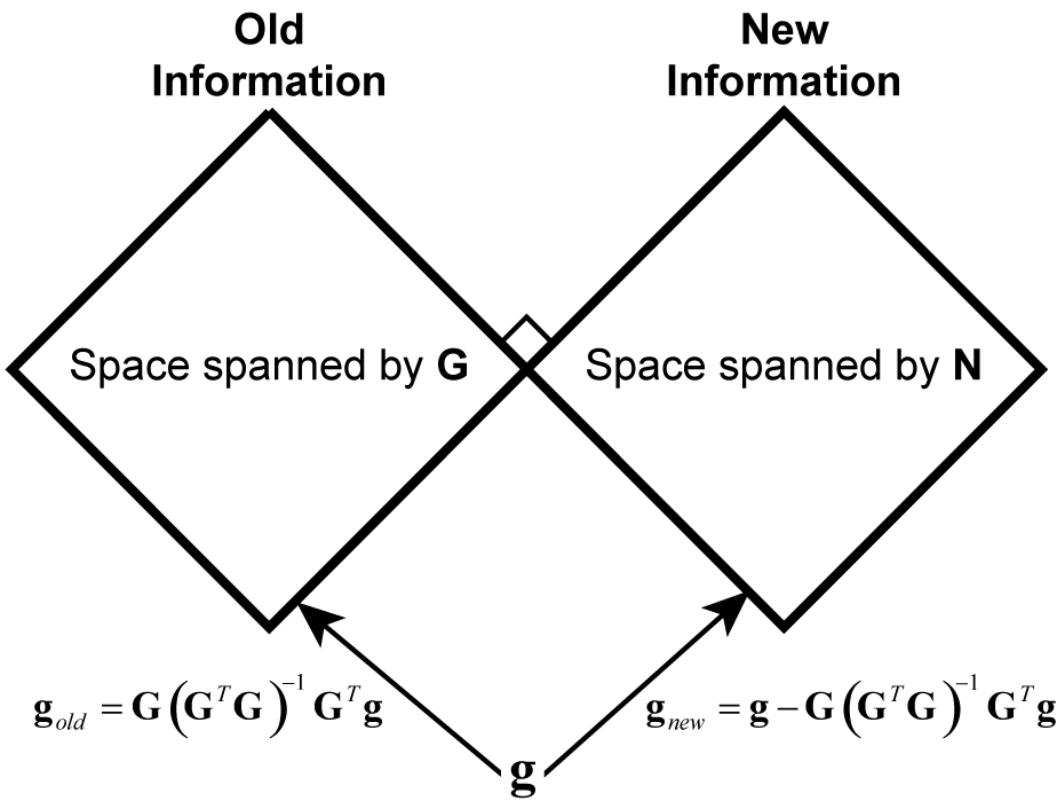
(a) **Probability of Producing Smaller Model % RMS Error**

	ERL	Straddling	All
ERL		0.81	0.64
Straddling	0.19		0.29
All	0.36	0.71	

(b) **Difference in Average Model % RMS Error**

	ERL	Straddling	All
ERL		-11.3%	-7.8%
Straddling	11.3%		3.5%
All	7.8%	-3.5%	

Table 5-3 Monte Carlo comparisons of two optimally designed experiments using 765 observations. The ERL survey was also considered for completeness. All designs were based on a homogeneous earth assumption. (a) Comparative probabilities of producing smaller model error between three designed experiments according to Equation (5.20). Each column records the probabilities that the experiment in the column header will produce smaller model errors than the experiment in the row headers. (b) Difference between average model % rms errors for the three experiments.



$$\mathbf{g}^T\mathbf{g} = \mathbf{g}_{new}^T\mathbf{g}_{new} + \mathbf{g}_{old}^T\mathbf{g}_{old}$$

$\mathbf{g}_{old}^T\mathbf{g}_{old}$: Energy in **g** mapping to the space spanned by **G**

$\mathbf{g}_{new}^T\mathbf{g}_{new}$: Energy in **g** mapping to the space spanned by **N**

Figure 5-12 Cartoon depicting the partitioning of 'Old' and 'New Information' into orthogonal subspaces. As our sequential design algorithm builds the experiment, 'old information' is defined as that information which is already available to the base experiment, and this is expressed by the model space spanned by **G**, the Jacobian matrix of the base experiment. 'New information' must be complementary to 'old information' and is therefore defined as that information which exists in a model subspace orthogonal to the 'old information'; this is expressed by the nullspace of **G**, denoted **N**. The sensitivity kernel of a candidate observation, **g**, can therefore be completely decomposed into components which reside in the 'old information' space, $\mathbf{g}_{old}$, and which reside in the 'new information' space, $\mathbf{g}_{new}$. Because the union of the spaces spanned by **G** and **N** spans all of model space, it follows that $\mathbf{g}^T\mathbf{g} = \mathbf{g}_{new}^T\mathbf{g}_{new} + \mathbf{g}_{old}^T\mathbf{g}_{old}$ and that $\mathbf{g}_{new}^T\mathbf{g}_{new}$ is the energy in **g** that maps to 'new information' space and $\mathbf{g}_{old}^T\mathbf{g}_{old}$ is the energy that maps to 'old information' space.

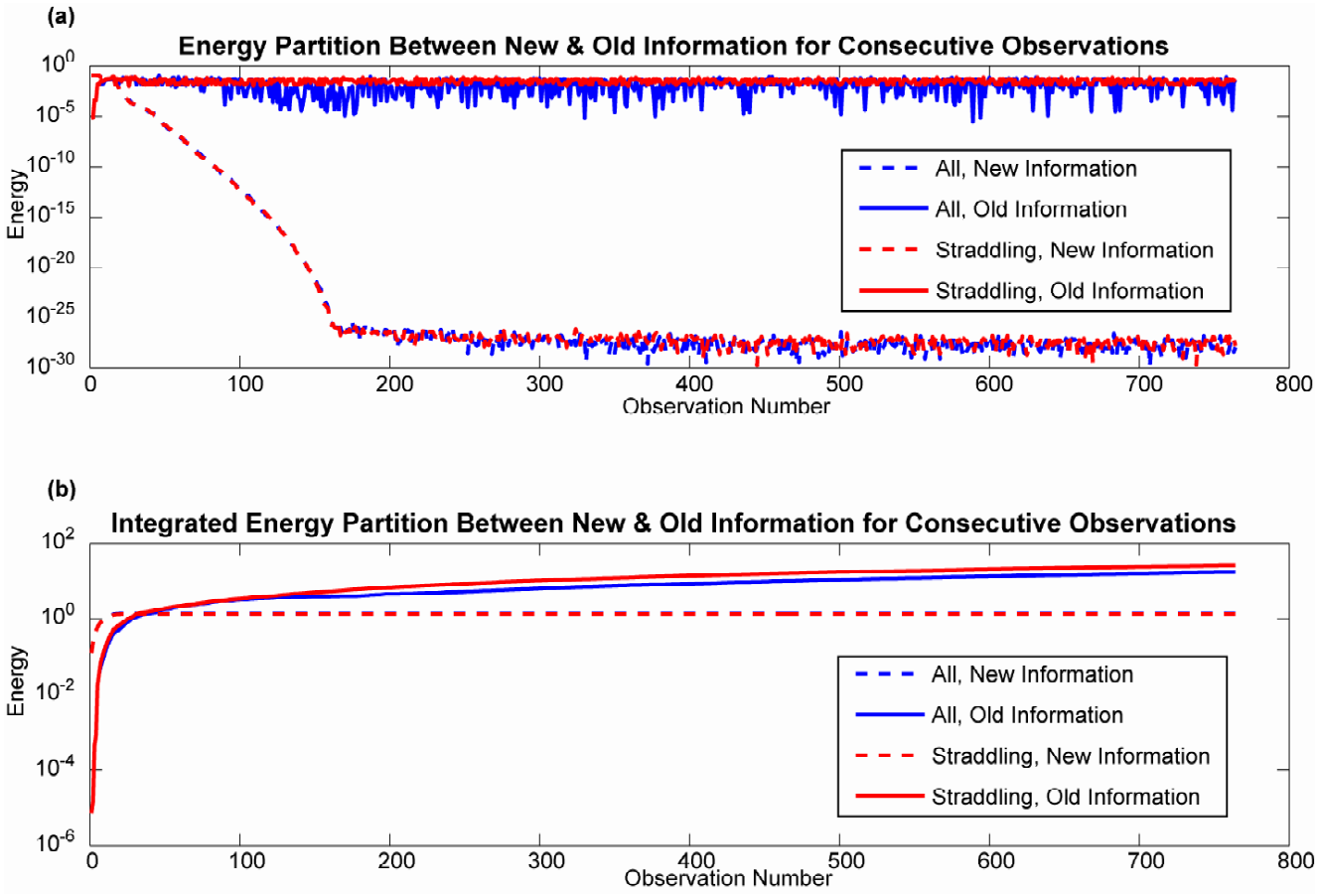


Figure 5-13 (a) Partitioning of the energy in the sensitivity kernels of consecutive observations in 'Straddling' and 'All' designs (red and blue lines, respectively) between 'New' information (solid lines) and 'Old' information (dashed lines). Observe that no new information is available after the 170th observation, which happens to be the maximum attainable rank of any Jacobian for this inverse problem. (b) Integrals of the 'Old' and 'New' information curves in Panel (a). Notice that both 'Straddling' and 'All' designs contribute equal amounts of 'new' information, but the 'Straddling' design contributes more 'old' information as experiments become larger.

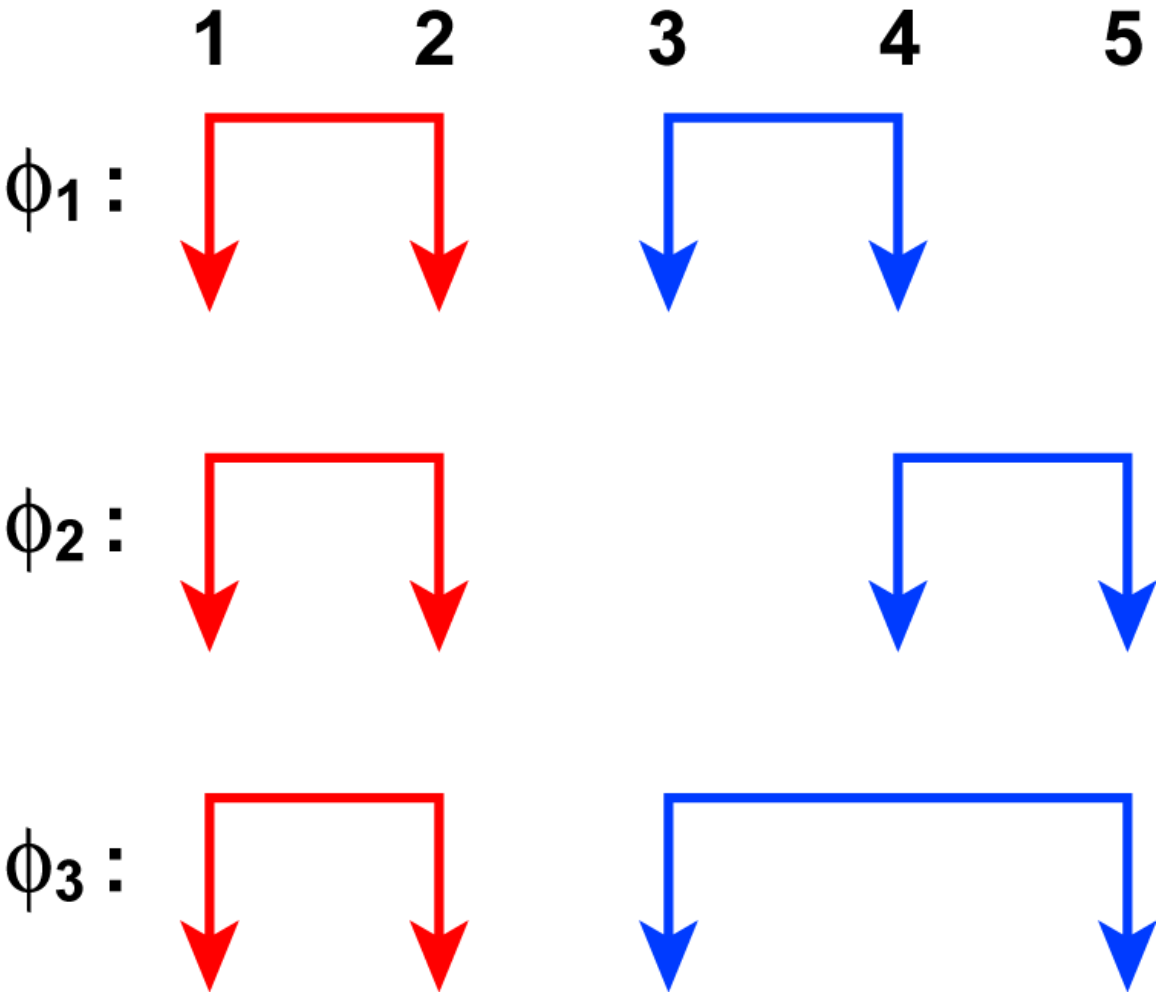


Figure 5-14 A Simple hypothetical experiment of three observations. Red indicates the current-source dipoles and blue indicates the potential-receiver dipoles (though reciprocity permits these roles to be reversed). Because electrostatic problems obey the law of superposition, the potentials of any two observations can be combined to produce the third; that is, $\phi_3 = \phi_1 + \phi_2$, $\phi_2 = \phi_3 - \phi_1$, and $\phi_1 = \phi_3 - \phi_2$. Hence, these three observations can only produce *two* bits of unique information about any target, because the information of any one of these observations is dependent on the information provided by the other two.

Magnitude & Angle Between Sensitivity Kernels

	Magnitude	Angle (°)		
		$\partial\phi_1/\partial\boldsymbol{\rho}$	$\partial\phi_2/\partial\boldsymbol{\rho}$	$\partial\phi_3/\partial\boldsymbol{\rho}$
$\partial\phi_1/\partial\boldsymbol{\rho}$	0.0899	0		
$\partial\phi_2/\partial\boldsymbol{\rho}$	0.0394	69.21	0	
$\partial\phi_3/\partial\boldsymbol{\rho}$	0.1103	19.53	49.68	0

Table 5-4 The magnitudes of, and angles between, the sensitivity kernels for the three hypothetical observations in Figure 5-14. All sensitivity kernels were based on querying a homogeneous earth.

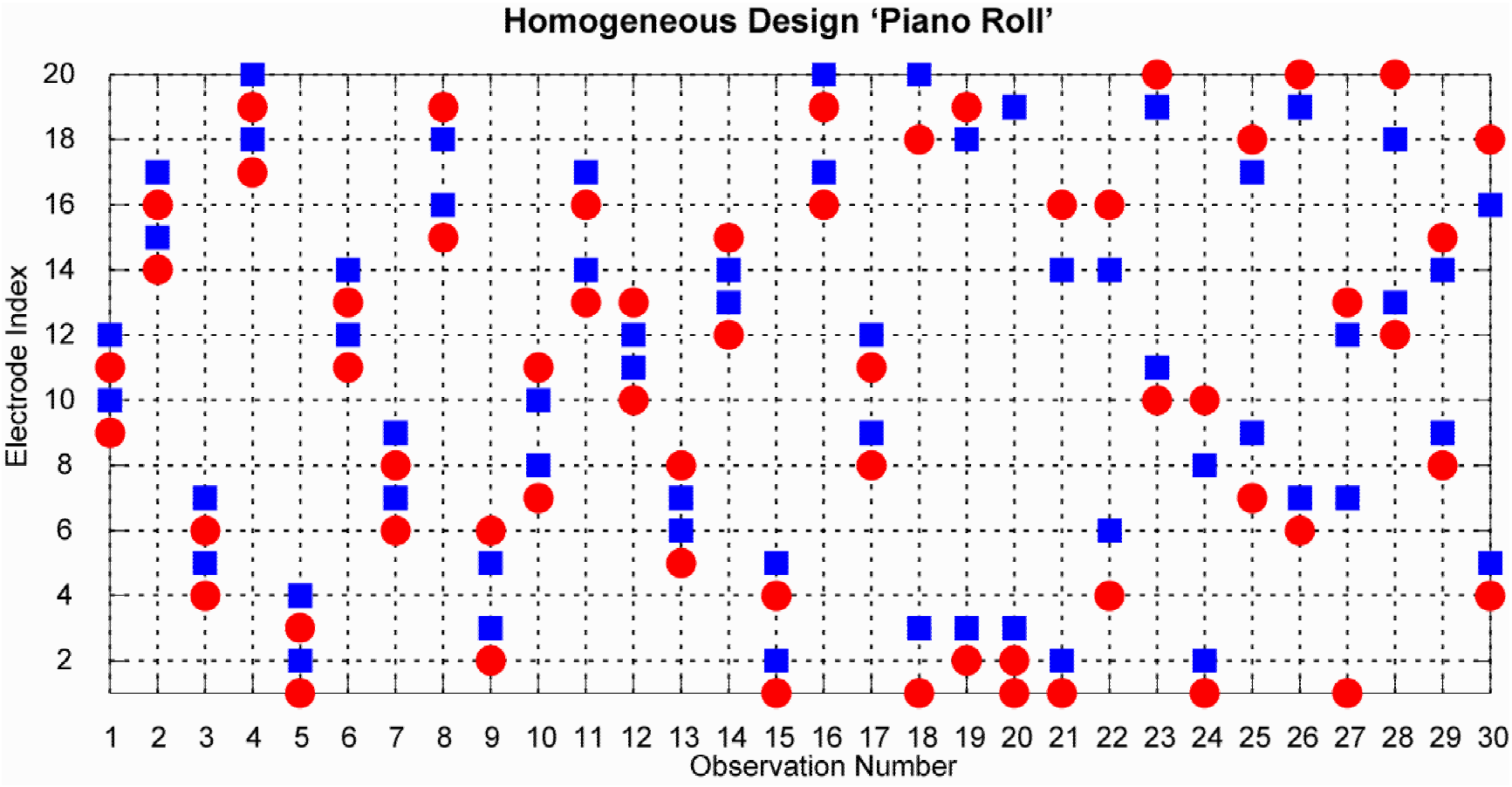


Figure 5-15 'Piano role' cartoon showing the first 30 observations in the 'All' designed experiment based on a homogeneous earth model. Red circles specify the positions of the transmitter dipoles and blue squares indicate the positions of the receiver dipoles. The order of the observations, from left to right, is the order in which the sequential design algorithm included them in the experiment. There is no readily apparent pattern in this set of data stations. This is unfortunate because a discernible pattern would be tractable to upscaling or downscaling with the number of electrodes.

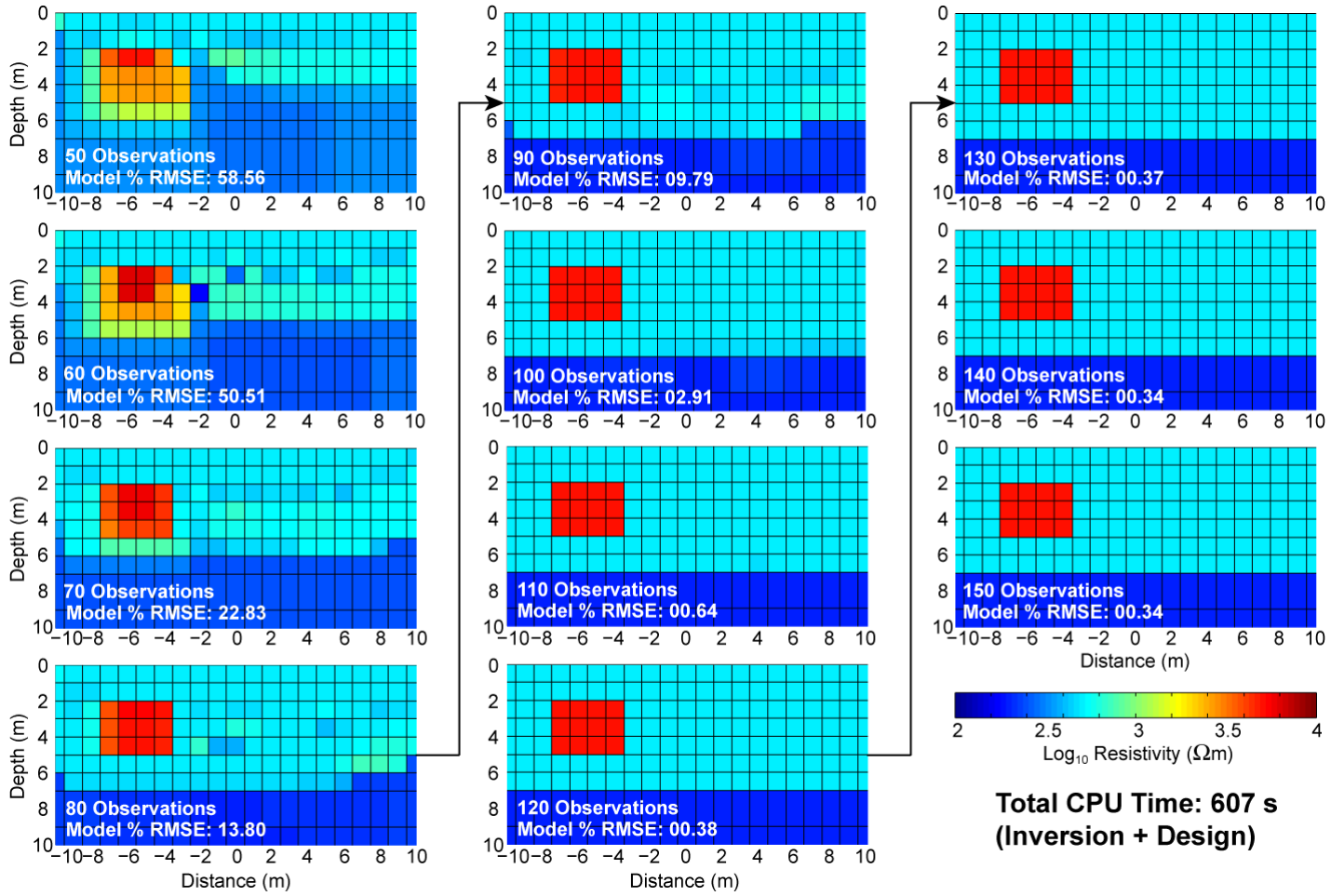


Figure 5-16 Small-Batch Adaptive OED demonstration with noiseless data. 50 initial observations were designed based on a homogeneous assumption and partially inverted (between 4 and 10 inversion iterations). Additional 'small batches' of 10 observations were prescribed by designing with the current heterogeneous earth model, until 150 observations were accumulated.

Each design stage was followed by a partial inversion stage. Total CPU time for designs and inversions together was approximately 10 minutes.

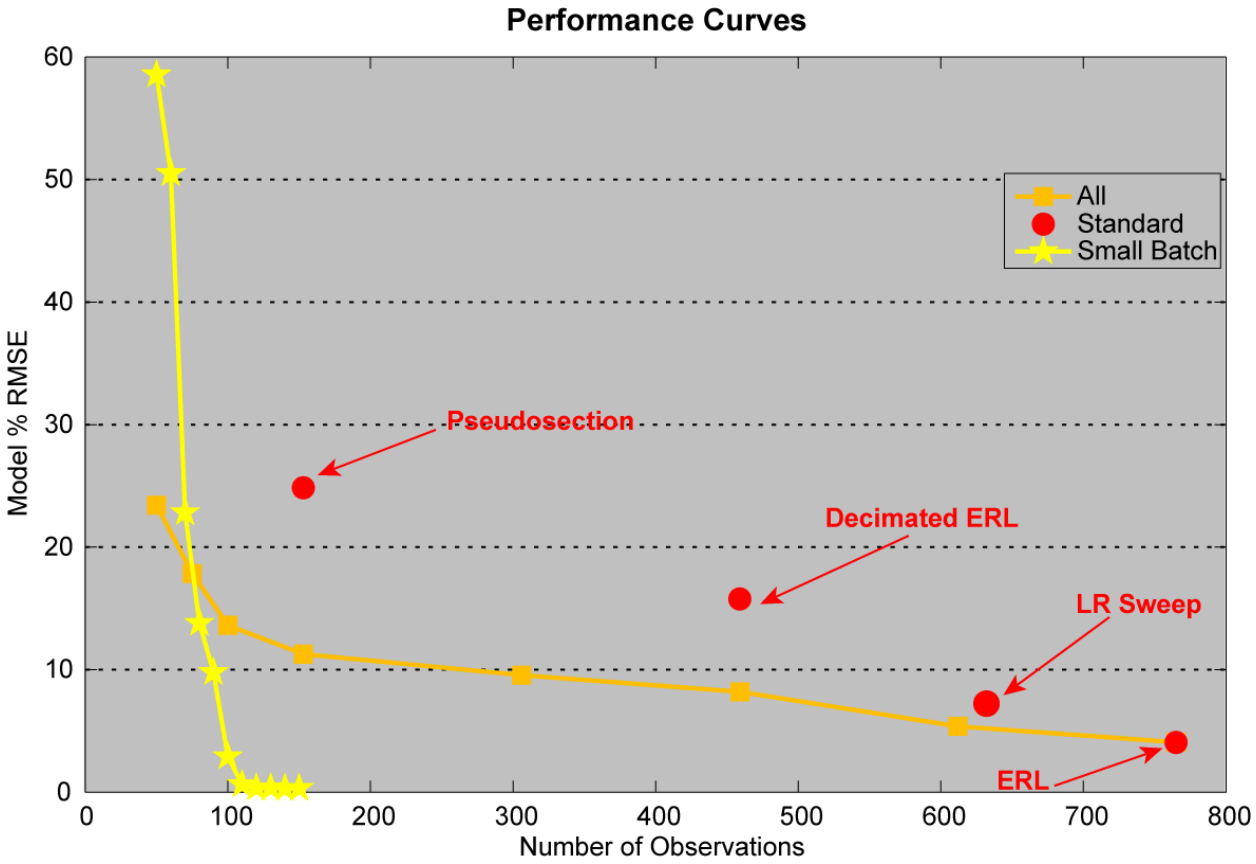


Figure 5-17 Performance curve for the Small Batch AOED method (model % rms errors in Figure 5-16) as compared with the performance curve of homogeneously designed surveys of varying numbers of observations (picking from 'All' available quadrupoles). Also shown are the performances for various 'Standard' surveys.

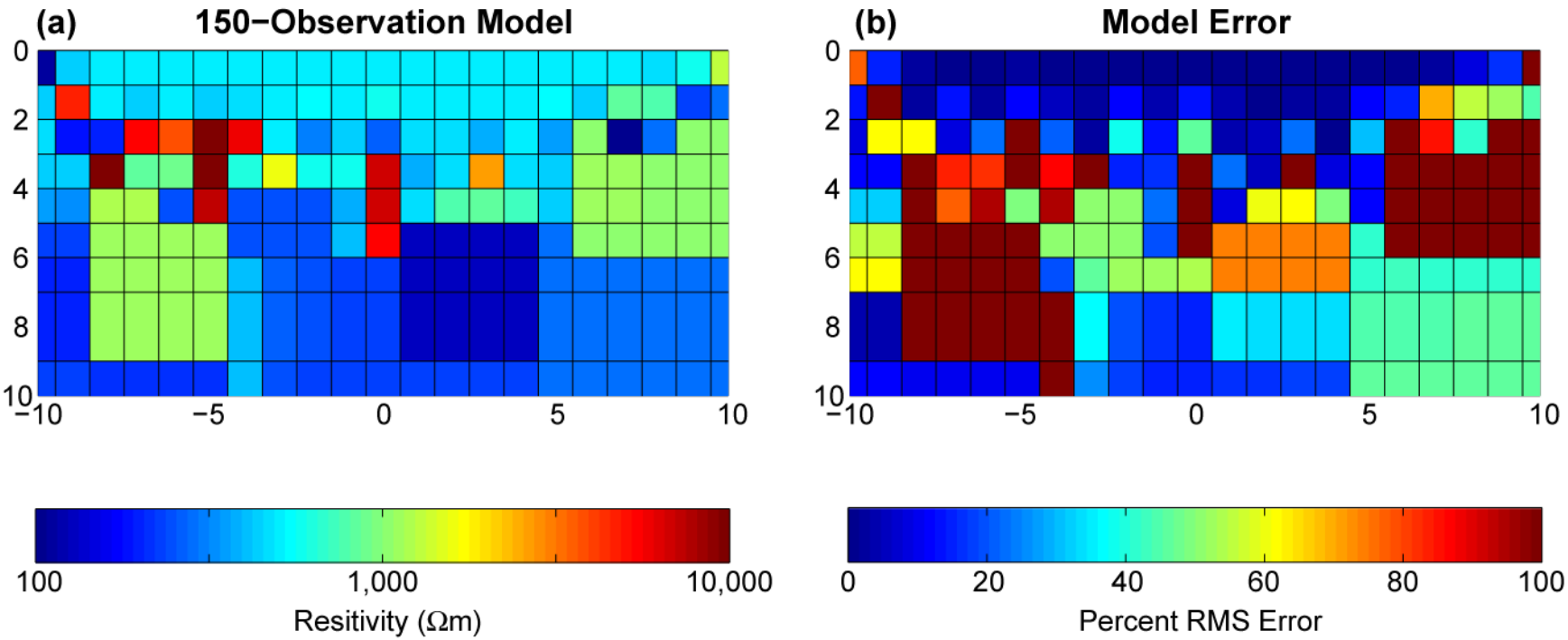


Figure 5-18 Example of Small Batch AOED in the presence of 3% Gaussian noise in the data. Panel (a) shows the final inversion model for one realization of data noise and Panel (b) shows the model % rms error in each model cell. Despite restting the Lagrange tradeoff parameter (per Equation (5.1)) to its maximum value at the beginning of each inversion step, the final inversion model is arguably divergent (see Figure 5-10 for comparison).

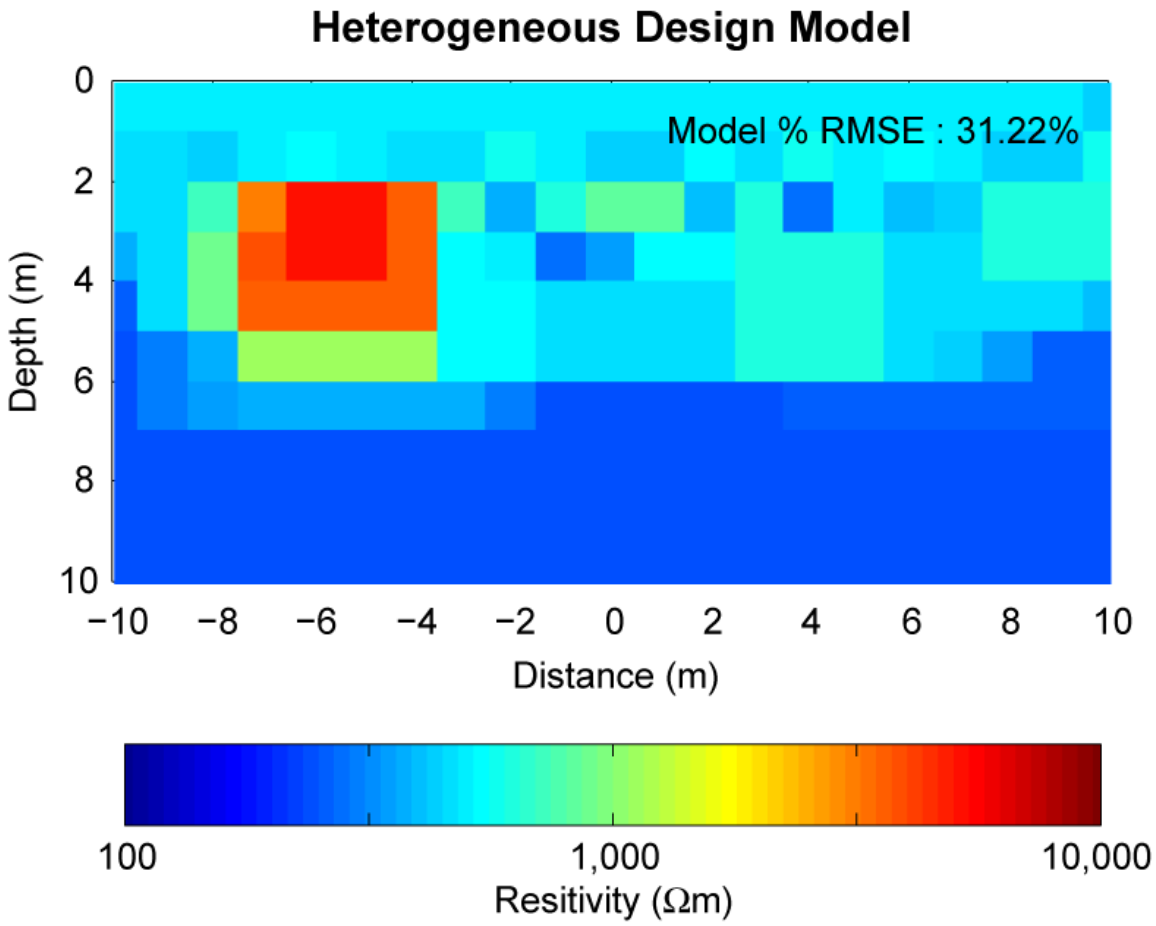


Figure 5-19 Heterogeneous earth model used as the design model for the adaptive, optimal experimental design exercise. The model was generated by inverting noisy synthetic data that had been generated for a designed experiment of 153 observations (using a design that picked from 'All' available quadrupoles) optimized for a homogeneous earth. The model % rms error for this image was 31.22%.

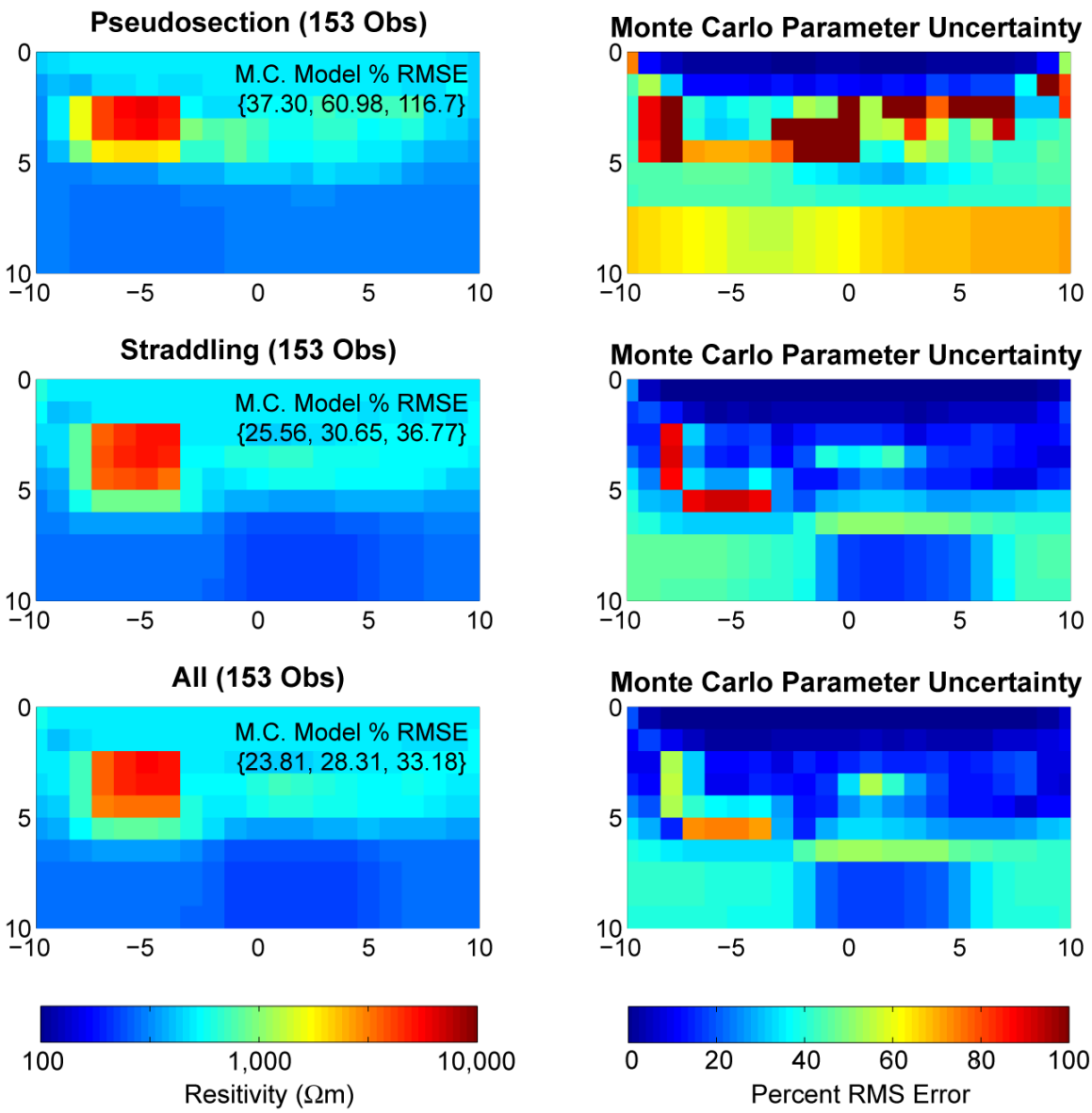


Figure 5-20 Graphical synopses of Monte Carlo simulations for data inversions with 50 realizations of 3% Gaussian noise for the three experiments, 'Pseudosection', 'Straddling' and 'All'. The designed experiments were adapted to the model in Figure 5-19, rather than using a homogeneous one. All experiments queried the cave model in Figure 5-8 with 153 observations. The left-hand panels show the average Monte Carlo models for each experiment over all 50 noise realizations. The bracketed numbers beneath the 'M.C. Model % RMSE' headers in each of the left-hand panels specify model % rms errors at the 10^{th}, 50^{th} and 90^{th} percentiles (generated by the Monte Carlo simulations). The right-hand panels show the 'Monte Carlo Parameter Uncertainties', measured as % rms error according to Equation (5.24), for the three experiments.

(a) **Probability of Producing Smaller Model % RMS Error**

	Pseudosection	Straddling (Het.)	All (Het.)	Straddling (Hom.)	All (Hom.)
Pseudosection		0.97	0.99	0.90	0.89
Straddling (Het.)	0.03		0.68	0.22	0.40
All (Het.)	0.01	0.32		0.12	0.27
Straddling (Hom.)	0.10	0.78	0.88		0.61
All (Hom.)	0.11	0.60	0.73	0.39	

(b) **Difference in Average Model % RMS Error**

	Pseudosection	Straddling (Het.)	All (Het.)	Straddling (Hom.)	All (Hom.)
Pseudosection		-36.1%	-38.7%	-31.0%	-31.0%
Straddling (Het.)	36.1%		-2.6%	5.1%	5.1%
All (Het.)	38.7%	2.6%		7.8%	7.7%
Straddling (Hom.)	31.0%	-5.1%	-7.8%		0.0%
All (Hom.)	31.0%	-5.1%	-7.7%	0.0%	

Table 5-5 Monte Carlo comparisons of five experiments using 153 observations. The experiments labeled '(Hom.)' were designed based on a homogeneous earth, and the ones labeled '(Het.)' were adaptively designed based on the heterogeneous model in Figure 5-19. The Pseudosection survey was also considered for completeness. (a) Comparative probabilities of producing smaller model error between three designed experiments according to Equation (5.20). Each column records the probabilities that the experiment in the column header will produce smaller model errors than the experiment in the row headers. (b) Difference between average model % rms errors for the three experiments.

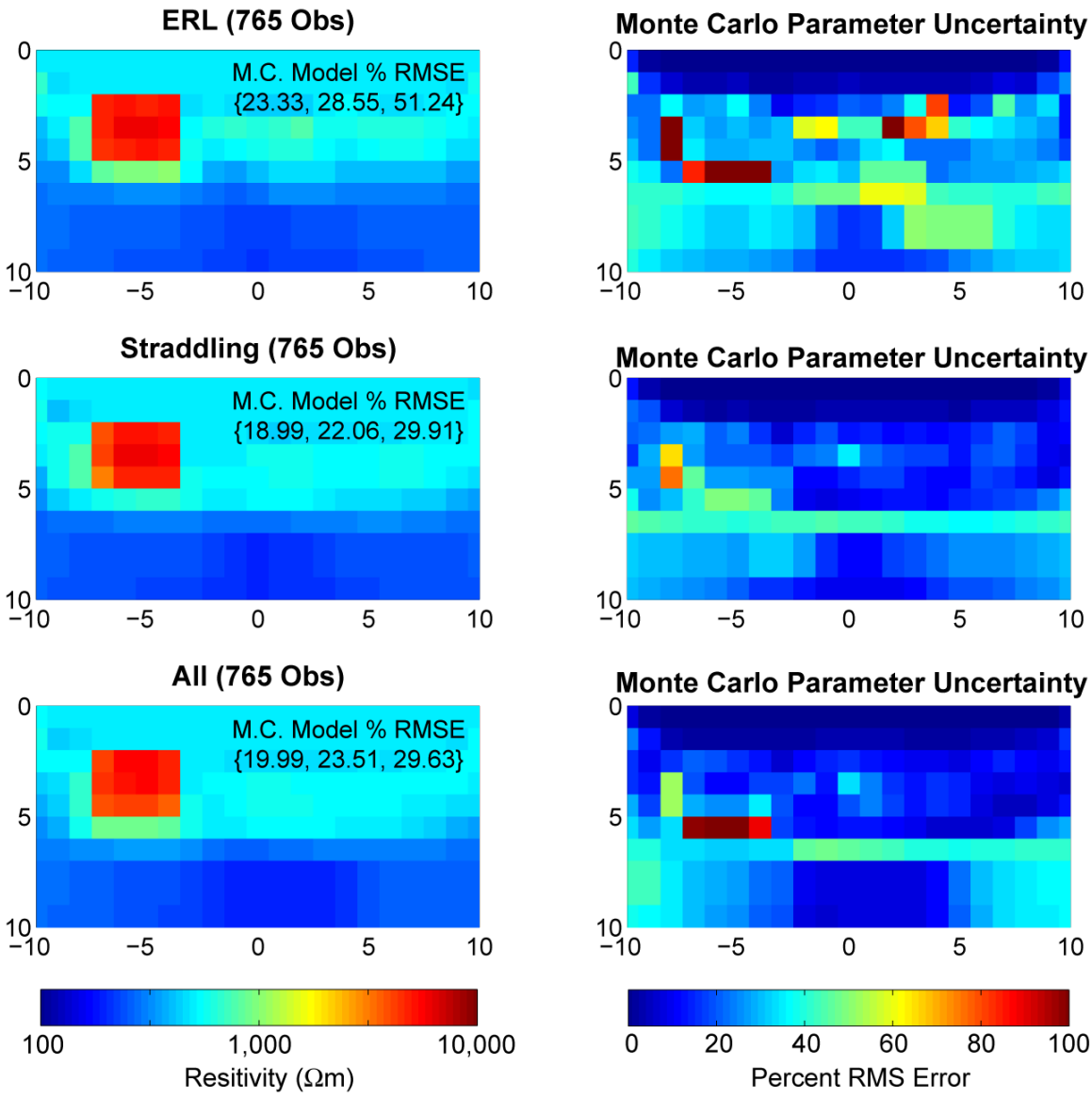


Figure 5-21 Graphical synopses of Monte Carlo simulations for data inversions with 50 realizations of 3% Gaussian noise for the three experiments, 'ERL', 'Straddling' and 'All'. The designed experiments were adapted to the model in Figure 5-19, rather than using a homogeneous one. All experiments queried the cave model in Figure 5-8 with 765 observations. The left-hand panels show the average Monte Carlo models for each experiment over all 50 noise realizations. The bracketed numbers beneath the 'M.C. Model % RMSE' headers in each of the left-hand panels specify model % rms errors at the 10th, 50th and 90th percentiles (generated by the Monte Carlo simulations). The right-hand panels show the 'Monte Carlo Parameter Uncertainties', measured as % rms error according to Equation (5.24), for the three experiments.

(a) **Probability of Producing Smaller Model % RMS Error**

	ERL	Straddling (Het.)	All (Het.)	Straddling (Hom.)	All (Hom.)
ERL		0.84	0.78	0.81	0.64
Straddling (Het.)	0.16		0.40	0.39	0.23
All (Het.)	0.22	0.60		0.51	0.31
Straddling (Hom.)	0.19	0.61	0.49		0.29
All (Hom.)	0.36	0.77	0.69	0.71	

(b) **Difference in Average Model % RMS Error**

	ERL	Straddling (Het.)	All (Het.)	Straddling (Hom.)	All (Hom.)
ERL		-13.1%	-11.3%	-11.3%	-7.8%
Straddling (Het.)	13.1%		1.8%	1.8%	5.3%
All (Het.)	11.3%	-1.8%		0.0%	3.6%
Straddling (Hom.)	11.3%	-1.8%	0.0%		3.5%
All (Hom.)	7.8%	-5.3%	-3.6%	-3.5%	

Table 5-6 Monte Carlo comparisons of five experiments using 765 observations. The experiments labeled '(Hom.)' were designed based on a homogeneous earth, and the ones labeled '(Het.)' were adaptively designed based on the heterogeneous model in Figure 5-19. The ERL survey was also considered for completeness. (a) Comparative probabilities of producing smaller model error between three designed experiments according to Equation (5.20). Each column records the probabilities that the experiment in the column header will produce smaller model errors than the experiment in the row headers. (b) Difference between average model % rms errors for the three experiments.

Chapter 6

Conclusion

This thesis has examined various aspects of geophysical optimal experimental design (OED) applied to DC resistivity problems. A novel method of experimental design was developed and attendant algorithms were created to practically execute survey optimization. Three variations on the design method were introduced: (1) a basic method that allows geophysical surveys to be designed without prior knowledge of earth structure; (2) an adaptive method that designs experiments once some preliminary knowledge of earth structure is known; and (3) a real-time or in-line adaptive method that attempts to design experiments as information on earth structure is being gathered and the earth image is being refined. These methods were variously applied to three unique geoelectrical settings to assess their advantages and drawbacks as compared with standard geoelectrical surveys and random surveys. The three applications settings included 1D surface resistivity, single-borehole (pseudo-2D) resistivity, and 2D surface resistivity.

In the course of investigation, many additional practical insights into the nature of geoelectrical data acquisition and inversion were brought to light that either directly or indirectly affect optimal experimental design. Additionally, both noiseless and noisy data scenarios were examined to demonstrate the strengths and weaknesses of these design techniques in theoretical versus real-world situations.

Several subsections are provided below that synthesize the theory and/or results from the various thesis chapters into comprehensive units that will hopefully provide

a macroscopic picture of the geophysical OED enterprise as it has been posed and investigated in these pages. Following these is a brief discussion of potential future research avenues.

6.1 Modeling and Inversion

Transmission line network analogs were applied to the forward models for the 2D and Borehole DC resistivity problems. These methods required additional boundary blocks to ensure modeling accuracy near the survey electrodes and to impose boundary conditions, which increased the number of model parameters, exacerbating the ill posedness of these inverse problems.

One-dimensional resistivity forward modeling was treated through a general analytic solution for horizontally layered media, permitting the earth to be modeled in terms of layer thicknesses and associated resistivities.

All data inversions were executed using nonlinear least squares formalism. Model regularization was managed through an L^1-norm smoothness constraint applied to the spatial gradient of the model, in conjunction with the Levenberg-Marquardt algorithm, which handled the trade-off between data rmse and model smoothness.

6.2 Theory

A novel optimal experimental design strategy was introduced. The prime characteristic of the technique lies in the treatment of observation sensitivity kernels as information vectors that bridge data and model spaces. In the OED framework, the fitness of an observation that is a candidate for addition to a base experiment is evaluated in terms of the magnitude and complementarity of the information it can provide relative to the base experiment. These two concepts are loosely quantified in terms of the magnitude (vector norm) of an observation's sensitivity kernel after it has been projected onto the null space of the base experiment and the angle its sensitivity kernel makes with respect to the space spanned by the base experiment. It was shown that

this formulation gives rise to an OED objective function equivalent to the ratio of the determinants of the augmented and base Jacobian matrices. The parameter uncertainty volume is a function of the determinant of a Jacobian matrix, so the basic design method attempts to minimize this measure of parameter uncertainty in a sequential fashion.

A second, and equally important, characteristic of the design methodology is that it is implemented using greedy optimization. That is, it approaches the survey optimization problem by sequentially finding observations to be added to a base experiment that are *locally optimal* with respect to the base experiment. This is in direct contrast to most historic OED techniques, which have approached the optimization problem in a global sense, trying to find whole experiments at once that are *globally optimal* with respect to some design objective function. Choosing to optimize experiments through a series of locally optimal updates generally cannot ensure that the final survey is globally optimal for a particular problem, but it does allow the design algorithms to execute in a fraction of the time that would be needed to find global solutions. This means that computationally expensive search algorithms such as the genetic and simulated annealing algorithms can be avoided.

Below are brief descriptions outlining the machinery of three algorithms that implement the preceding ideas.

6.2.1 Basic Design Algorithm

The basic algorithm initializes an experiment with one or more observations – typically the seed is a 'lead' observation whose sensitivity kernel has the largest vector norm – and sequentially adds additional observations. Candidacy for addition to the base experiment is evaluated by determining the L^2-norm of an observations sensitivity kernel onto the null space of the Jacobian of the base experiment. In this manner, the dual objective of maximizing information magnitude and complementarity is achieved.

In all design scenarios, a working model of the earth resistivity distribution is required. If no knowledge of this distribution is available, we start from a homogeneous assumption.

6.2.2 Modified Algorithm for Maximum Attainable Rank

This algorithm builds on the basic design algorithm. When the number of observations in the base experiment reaches the maximum attainable rank of the sensitivity matrix, candidacy for addition to the experiment can no longer be evaluated based on the projection of sensitivity kernels onto the null space of the sensitivity matrix. This is because, at this point, the sensitivity kernels of all candidate observations reside wholly in the space spanned by the Jacobian of the base experiment; i.e., no component of any sensitivity kernels resides in the null space of the Jacobian.

To overcome this problem, the design objective function is altered once the maximum attainable rank has been encountered. Instead of projections onto the null space of the Jacobian, we instead perform projections onto the vector space spanned by the singular vectors of the Jacobian corresponding to the smallest singular values. Candidacy is based on those observation sensitivity kernels that are of maximal length when projected onto this subspace and which therefore contribute the most information to the small singular values, reducing model uncertainty and the inversion instability associated with these small values.

Bringing singular vectors into the evaluation of observation candidacy does increase computational expense, but because the modified algorithm still approaches the problem through a greedy optimization method, it still executes relatively quickly.

6.2.3 Adaptive Design Methods

The adaptive design methods introduced in this work have taken one of two approaches to the design problem.

The simpler of the two performs an initial survey and uses that inversion image as the model submitted to the design algorithm to generate a tailored survey for a second

inversion. This was labeled Two-stage Adaptive OED. Either of the algorithms discussed above are suitable, without modification, for use in this Two-Stage technique.

The more elaborate methods approached the design problem in real-time. Both techniques integrate inversion and design into a super-algorithm that cycles between a design stage and an inversion stage, as new observations are optimally added to the experiment.

One suggested methodology, termed In-Line AOED, approached the real-time design problem by first collecting, and partially inverting, a seed data set. The updated earth model was then submitted to the design algorithm, which found a single optimal observation to add to the base experiment. The new observation would then be collected; a new updated model would be created, again by partial inversion; and the process would then cycle until convergence.

An alternate methodology, termed Small-Batch AOED, worked in a manner similar to the In-Line technique, except instead of adding one observation at a time, small batches were added. This was hypothesized to speed up the overall run-time of the integrated super-algorithm and to ensure more reliable modeling results at the end of the day.

6.3 Continuous Versus Discrete Experimental Spaces

Depending on how the geoelectrical forward and inverse problems were posed, survey electrodes could either be placed anywhere on a continuum or would be restricted to discrete positions. This affected the type of optimization algorithms that could be brought to bear on the survey optimization problem.

One-dimensional resistivity problems were posed in such a way that the survey electrodes could be placed anywhere along a continuous line. This meant that there were literally an infinite number of observations available to the survey designer, or that the experimental search space was infinite. While a continuous experimental search space would suggest that partial derivatives of experiment fitness (evaluated through the design objective function) could be evaluated with respect to electrode

positions, it was shown that the complexity of the design objective function makes their evaluation nearly impossible. Consequently, fast gradient-following optimization algorithms were not available to solve the greedy optimization problem for each additional observation. Instead, an evolutionary optimization method, called Differential evolution, was employed.

The 2D surface and borehole resistivity problems were cast in terms of resistor networks (transmission line network analogy), forcing survey electrode placements to be at discrete pre-defined positions, at nodes of the network. As a consequence, there were finite sets of 4-electrode or quadrupole observations available for experimentation. In this case, partial derivatives of the design objective function with respect to electrode position were undefined, precluding the possibility of gradient-following greedy optimization algorithms, which, in light of previous discussion, would have nonetheless been impossible to find. However, unlike the 1D surface resistivity problem, a finite set of permitted observations allowed us to exhaustively evaluate the fitness of candidate observations.

6.4 Sequential Versus Global Design Algorithms

Sequential design does produce suboptimal experiments compared to global search methods but approaches global optimality as the number of observations increases. In an exercise conducted in this research, the sequential method reached more than 90% of global optimality, which is excellent considering it took ~1/100th the CPU time of a global search algorithm to do so. The sequential method imposes significantly less computational expense, requiring as little as (or less than) 1% of the compuataion time of global search techniques. All of which is to say that sequential design produces extremely high-quality, albeit suboptimal, data quality in a fraction of the time needed to design experiments by stochastic search methods. The trade-off comes in deciding whether an additional 10 or 20% increase in optimality is worth increasing CPU time by more than 2 orders of magnitude. Ultimately, sequential design closes the ‘optimality gap’ between itself and global search methods as experiment sizes in-

crease, arguing strongly for its application in real-world problems. It is hypothesized that the sequential method may indeed produce global optima as the number of observations continues to increase, but this must be tempered by the fact that OED is subject to a general law of diminishing returns, which asserts that arbitrarily large random surveys can perform as well as deliberately designed ones.

6.5 Homogeneous-Earth Experimental Designs

The first avenue of investigation for all optimal experimental design exercises was to consider the case where no *a priori* information was available on the distribution of earth media. Absent this knowledge, a homogeneous earth was assumed for survey optimization exercises. These 'homogeneously' designed experiments were then deployed on heterogeneous targets and compared with random and standard surveys to ascertain whether any benefit could be garnered in survey optimization for the wrong (but 'neutral') earth model.

6.5.1 Random, Standard and Designed Experiments

It was shown in the 2D and borehole resistivity chapters that experiments optimized for a homogeneous earth but deployed on heterogeneous targets were statistically more likely to produce high quality data sets than either random or standard surveys (Pseudosection and ERL surveys) of equal size, where 'quality' connotes small model % rms error. Because the optimization method uses a greedy optimization approach, it cannot guarantee globally optimal surveys, meaning that occasionally a random survey could produce a smaller modeling error than a designed survey, but the probability of generating such a survey at random was generally so low as to preclude adopting a wholly random experiment strategy. Moreover, homogeneously designed surveys outperformed both the standard Pseudosection and ERL surveys. Hence, it is concluded that the sequential survey optimization procedures developed herein do indeed produce smarter data sets in the case where no prior information on the target site is available.

6.5.2 Diminishing Returns

There is one significant caveat to the preceding remarks: geophysical optimal experimental design is subject to a law of diminishing returns. The 2D and borehole examples both showed that as the number of observations in an experiment becomes large, the difference in performance between designed, random and standard surveys diminishes (as measured by the model % rmse). In particular, as designed experiments approached the size of ERL surveys, the relative difference in their performances, while still favoring designed surveys, was only marginally. From a practical point of view, it becomes relevant to ask whether the additional time and computational effort needed to design 'large' surveys is worth the marginal benefit they can give, relative to a ready-made survey like the ERL survey.

Notably, the largest relative improvements in data quality were realized when small designed experiments were compared with equally sized random and standard ones. This is advantageous on two related scores: first, the smaller the experiment is chosen to be, the faster the design algorithm executes; second, by choosing to use small designed surveys that provide compact information-dense data, the entire field operation can be executed more rapidly than by heavy handedly collected large quantities of data pell mell.

6.5.3 Noiseless Versus Noisy Data

The 2D surface resistivity chapter formally examined the use of homogeneously designed experiments deployed on heterogeneous targets in the presence of data noise. To some extent, so did the 1D resistivity chapter, but the design algorithm was not employed to design a 1D homogeneous survey for practical reasons. It was shown that designed surveys were statistically more likely to produce higher quality data than competing standard surveys but that, compared with the differences in performance in the noiseless data cases, the designed experiments were not nearly as superior. This fact is more fully developed below, after concluding remarks have been offered regarding adaptively designed experiments in the presence of noisy data.

6.5.4 Experimental Design Pattern Recognition

The 2D and borehole resistivity chapters showed 'piano roll' depictions of the quadrupoles, in the order they were added to the base experiment, for 'homogeneously' designed experiments. It was hoped that a simple pattern could be deduced governing the selection of optimal quadrupoles, which could be easily communicated. Were this the case, a new general survey would be established that could easily be scaled up or down depending on the number of electrodes used in a survey. Based on demonstrations that homogeneously designed experiments produced superior data quality relative to standard or random experiments, such a general survey would supercede, to some degree, computational optimal experimental design all together.

Unfortunately, no such pattern was readily apparent. Three explanations were offered why this might be so. First, it was shown that observations could be linearly combined to produce new observations. This means that two or more observations in an experiment could be combined to produce a third observation that might be more amenable to the pattern recognition exercise. But, from a mathematical point of view, this third observation does not generally provide the same information between data and model space as was provided by the two observations from which it derived. Consequently, an effort to substitute observations that were not placed in the experiment by design is either difficult or impossible. Such substitutions will likely diminish the quality of the information provided by the designed experiment.

Second, and perhaps most importantly, the pattern recognition exercise is made more difficult by the fact that we do not know *a priori* what pattern we are looking for. The first few observations in designed experiments have a readily identifiable pattern, but this identifiability breaks down quickly thereafter. It would indeed be a prodigious and frustrating exercise to ascertain a pattern in designed surveys when one has no foreknowledge what one is looking for.

Finally, the greedy sequential optimization algorithms employed in this research sacrifice global optimality in exchange for faster computation times. While there may be a readily apparent pattern in a *globally* optimal homogeneous survey, it is likely that the 'locally optimal' surveys produced by our design algorithms do not give rise

to identifiable patterns. It might therefore be worthwhile to expend the effort to design a truly globally optimal experiment for a homogeneous earth, on the hypothesis that a discernible pattern will arise in the resulting experiment.

6.6 Adaptive Optimal Experimental Design

The next level of sophistication in our investigations of optimal experimental design was to adapt experiments to specific targets. Of the three adaptive procedures introduced, the Two-Stage methodology was most promising. It was employed successfully in all three chapters, 1D, 2D and borehole resistivity, in both noiseless and noisy data scenarios. In addition to the fact that the 2-stage method performed as well or better than both the In-Line and Small-Batch AOED methods, it also had the fastest execution times and required less hyperparameter tuning, making it more practical in field applications.

Furthermore, it was clear from the 2D and borehole investigations that adapted surveys produced data quality superior to that produced by the homogeneously designed ones. However, the law of diminishing returns was still evident in the adaptive cases. There comes a point where the user must decide whether it is worthwhile to submit to the additional expense of designing large adapted surveys when large random or standard surveys would suffice to produce comparable data quality.

6.6.1 Noiseless Data Versus Noisy Data

As was seen in the 2D and borehole resistivity studies, homogeneous designs did not show a comparable degree of superiority over random and standard surveys once noise was introduced into the data. The same situation was observed for adapted experimental designs. Though the adapted designs in all three resistivity chapters were shown to produce statistically significant improvements in data quality in the presence of data noise compared with standard surveys, they did not show such a marked degree of improvement as was seen in noiseless data cases. In particular, the borehole AOED example showed that there was a distinct possibility that the designed experi-

ment would actually produce a significantly larger modeling error than the ERL survey, though the probability was not exceedingly high. The 2D resistivity AOED examples did not show this behavior, but it was nonetheless clear that the quality of the data for designed surveys was compromised by noise. The 1D AOED examples also showed that the designed experiments reduced parameter uncertainty relative to Schlumberger and Wenner surveys, but the improvements were not profound. This leads to the possible hypothesis that the quality of data for designed experiments (for noiseless or noisy data) is model dependent. Sometimes an experimental design can greatly improve data quality relative to random or standard surveys and sometimes the improvements will not be so spectacular.

Moreover, all AOED results in the presence of noisy data highlight the fact that geoelectrical problems are ill posed, or at least poorly conditioned, inverse problems. Such problems are always very susceptible to data noise, and this is apparently an irremediable characteristic of geoelectrical imaging methods, no matter whether surveys are optimally designed or not. Therefore, though it is disappointing, it is nonetheless unsurprising that the adaptively optimized geoelectrical surveys, in the presence of data noise, did not perform comparably with their counterparts in the noiseless cases.

Further concluding remarks on optimal experimental design in the presence of noisy data are offered separately below.

6.7 CPU Time

By dint of the fact that the design methods in this research were executed via greedy optimization methods, their CPU times were extraordinarily small. In all three resistivity scenarios, CPU times were typically on the order of seconds to a few minutes. This is in contrast to the computational burden imposed by global search methods, which have been only sporadically reported in the literature and which usually ranged from hours to days. Practically speaking, design algorithms that take hours or days to execute are of only limited use in geophysics. Primarily, they might be used for

monitoring problems, but they certainly would be of little use in near real-time field exercises.

A greedy approach to the design exercise sacrifices global optimality in exchange for fast computation time. And though some of the results reported herein are perhaps not as profound as we would like, the fact is that our design methodology *does* produce superior data sets in nearly all instances. That it can do so for such small additional computational burden makes greedy optimization an important inclusion in the emerging field of optimal experimental design.

6.8 Expediting Design

As the number of model parameters and the size of experiments increase (typically in lockstep), even greedy optimization algorithms run less quickly than might be practical for real-world application. Two ideas were examined that could usefully reduce computation times. The first was relevant particularly to electrical tomography problems and the second related to reduction in the size of the Jacobian matrix, which is ultimately the mathematical construct upon which our experimental design algorithms operate.

6.8.1 External, Internal and Straddling Quadrupoles

In electrical resistivity tomography, the set of all possible 4-electrode or quadrupole observations can be partitioned into three distinct subsets, called 'Internal', 'External' and 'Straddling'. These names derive from the type of quadrupoles the subsets comprise. The 'Internal' configuration signifies that the potential dipole is inside the current dipole; the 'External' configuration signifies that the potential dipole is outside the current dipole; and the 'Straddling' configuration signifies that the potential dipole is half inside half outside the current dipole.

It was shown that the set of all straddling quadrupoles produced the highest cumulative sensitivity within the inversion target window, followed closely by the set of all internal configurations and distantly by the set of all external configurations. Addi-

tionally, it was shown that the rank of the Jacobian of all straddling configurations was equal to the maximum attainable rank (MAR) of the Jacobian of all quadrupoles, that the rank of the Jacobian of all internal configurations was one less than that, and that the Jacobian of all external configurations was N less than the MAR, where N is the number of survey electrodes.

Trials in the 2D and borehole resistivity chapters confirmed that experiments consisting only of straddling quadrupoles do indeed produce data quality comparable to, and sometimes in excess of, experiments designed from all possible quadrupole configurations.

Thus, it was illustrated that the straddling quadrupole configuration was the most superior of the three configuration types and could be exclusively used in design exercises, reducing the set of permitted candidate observations by two-thirds and commensurately increasing design times.

In addition, it was shown repeatedly in both the 2D and borehole chapters that the performance of the Pseudosection survey, which consists exclusively of external quadrupoles (by far the poorest of the three configurations), was exceptionally poor. In point of fact, the Pseudosection survey performed worse than even random experiments of the same size. It is therefore recommended that the Pseudosection survey should be avoided in geophysical exploration.

6.8.2 Truncating and Compressing the Jacobian

As the number of model parameters becomes large, the number of columns in the Jacobian matrix of an experiment increases. The sequential experimental design objectives operate on the sensitivity kernels of observations, and these kernels increase in length as the number of columns in the Jacobian increase. All of which is to say that OED computation time is proportional to the number of model parameters. Three novel approaches for decreasing the number of columns in the Jacobian matrix (strictly for use in the OED exercise) were introduced, which would be useful in reducing design computation time for large imaging problems.

The first method was a simple Jacobian truncation method that expunged all columns of the Jacobian corresponding to boundary blocks outside the target window. The second method was also a truncation approach that expunged columns of the Jacobian based on whether the magnitude of their 'global' sensitivity was above a user-defined threshold. This method was similar to the first, but allowed some boundary cells to remain in the truncated Jacobian so long as their 'global' sensitivity was above the threshold. The third method compressed the Jacobian matrix from $\mathcal{M}$ columns (corresponding to a total of $\mathcal{M}$ model cells) to a number of columns equal to the maximum attainable rank. This last method worked on the principle that all information in the sensitivity kernels resides in the vector space spanned by the right singular vectors of the Jacobian of all observations. Because this Jacobian is necessarily rank limited, the number of right singular vectors corresponding to nonzero singular values must equal the MAR. Therefore, left multiplying with the truncated set of its right singular vectors losslessly compresses the Jacobian matrix. The advantage of this last approach is that all information in the sensitivity kernels is preserved, but in a compressed format where lengths and angles have been preserved.

Comparison of these three Jacobian reduction schemes showed that experimental design executed with the compressed Jacobian method produced the best data quality, followed closely by the simple method of truncation (expunging all boundary blocks from the Jacobian). Moreover, the compressed and simple-truncation methods produced data quality comparable with experiments designed without reducing the Jacobian matrix. The main difference between these two successful reduction schemes, compared with the unaltered Jacobian approach, was that design times were reduced by a factor of more than 20. This is a very significant improvement in computation time and constitutes an important contribution to the theory of experimental design.

6.9 Noisy Data

Though data noise has already been discussed in the context of homogeneous and adapted designs, a separate concluding section is provided here to add a few pertinent final remarks.

6.9.1 Data Noise Due to Electrode Misplacement

In the 1D resistivity chapter, the issue of data noise created by electrode misplacement errors was taken up. Misplacement errors can derive from the finite precision with which electrodes can be placed, simple carelessness, or the need to place the electrode elsewhere because of obstacles.

A Monte Carlo exercise and subsequent statistical analysis showed that for 1D resistivity soundings, the ideal spacing of the inner electrodes is ~43% of the spacing of the outer electrodes. This spacing minimizes the expected error in an apparent resistivity datum owing to electrode misplacement, regardless of the outer electrode spacing. Additionally, expected data errors decrease as the outer electrode spacing increases. This was a very unusual discovery, but the analysis was double checked two different ways and was found to be correct. Among other things, it implies that the Wenner array (inner electrode spacing is 33% of the outer electrode spacing), which has for a long time been disregarded for being inferior to the Schlumberger array, might in fact be superior to the Schlumberger array in cases where electrode misplacement is a concern.

A Monte Carlo trial simulating random electrode misplacements for a 1D resistivity scenario strongly confirmed the unusual '43% Rule'. It was statistically shown that modeling errors for a '43%-survey' were considerably smaller than modeling errors for a competing Schlumberger survey of comparable size.

6.9.2 Random Data Noise

The homogeneous and adapted experimental designs in this research showed a great deal of improvement in data quality over random and standard surveys when data

were noiseless, but these improvements partially evaporated when the data were contaminated with noise. Two related hypotheses were posited to explain this apparent decline in survey performance, and both pertain to the model regularization scheme needed to stabilize the ill-posed/ill-conditioned inverse problems in this work.

First, in the presence of data noise, it was explained that the Lagrange multiplier governing the trade-off between the data rmse and model smoothness had to be kept larger than was the case for noiseless data. The ultimate effect of which was that the inversion result was biased toward an external source of information – the smoothness constraint – for which designed experiments had not been optimized. As a result, designed experiments (which, it should be pointed out, still outperformed their standard and random competitors) did not perform as well as might have been expected and in fact performed much more similarly to random and standard surveys, which were also subject to increased reliance on the smoothness constraint.

Second – and this is really just an extension of the first – the design objective would need to be modified to incorporate the constraint matrix. Because of the heavy reliance on constraints for ill-posed/ill-conditioned inverse problems in the presence of data noise, the eigenspectrum of the constraint matrix actually dominates the eigenspectrum of the Jacobian matrix for all but a handful of the lead eigenvalues. It was hypothesized that the experimental design objectives might need to modified in such a way that they favor the *large* eigenvalues in the Jacobian, rather than the *small*, as these are the only eigenvalues that in effect convey information on the bridge between data and model space.

6.10 Future Work

Below is a non-exhaustive list of future avenues of research in geophysical OED.

6.10.1 ERT Optimal ED Using Current Dipoles

Two-dimensional and borehole resistivity experiments were designed one observation at a time, choosing from a large set of permitted quadrupoles. Rather than designing

ERT experiments in this fashion, it might be possible to simplify the design method by designing using only current dipoles. Each current dipole would automatically come with a suite of electrical potential observations (perhaps via a set of simple translating potential dipoles) that measure the potential field at all remaining electrode positions. This would reduce the 'permitted number of observations' very significantly. In fact, there would just be ${}_N C_2$ current dipoles, which is considerably smaller than the $3\,{}_N C_4$ permitted quadrupole observations. Hence, design time could be extraordinarily efficient, as the number of possible 'observation sets' would be small. Moreover, it might also be much easier to identify a design pattern for a homogeneous earth when the design process is executed in terms of these observation sets.

6.10.2 Nonlinear Optimal Experimental Design

It was explained in the Theory chapter that most contemporary methods of experimental design operate on the Jacobian matrix, which is a linear approximation of the nonlinear forward operator. In this sense, this work, and most historic work, in optimal experimental design are examples of what might generally be termed *approximately linear* experimental design. Contextually, experimental designs are only practically optimal in a neighborhood around the point in model space for which they have been designed. If the true model is outside this conceptual 'neighborhood of optimality', the final inversion result for a designed survey can in no way be claimed to be optimal. This naturally begs the question whether some form of *nonlinear experimental design* might be better suited for design problems.

Instead of working with linear approximations of the forward modeling equation(s), perhaps there is some way of setting up design problems with respect to the nonlinear equation(s). In fact, this author is aware of at least one group who has begun to tackle the fully nonlinear problem (Andrew Curtis, personal communication). Hypothetically, working with the full nonlinear constitutive equations might be useful in designing more robust experiments whose optimality can be expected to encompass

a larger neighborhood in model space than can currently be defined with respect to linear approximations.

6.10.3 An Exhaustive Comparison of Current Design Strategies

As yet, an exhaustive comparison of current design methodologies has not yet been undertaken. This would be a formidable but important next step in the development of geophysical OED theory. It is envisioned that this study would entail a large number of comparative Monte Carlo exercises to compare the various design methods and would also account for computation time and feasibility in field settings. This would require a very significant amount of coding and error checking, for each of the design methodologies has different objectives and may be better suited to one of any number of optimization algorithms.

6.10.4 Incorporating Model Regularization

As was discussed in the concluding remarks on data noise and experimental design, many of the results reported in this research indicated that the model smoothness constraints had an undue (though necessary) influence on inversion results. The design objectives set forth in the Theory chapter were geared toward maximizing the information in a survey that bridges model and data space; they *did not* address the information *imposed* by model regularization, however. In light of the heavy reliance on model regularization for noisy data inversions, the design objective might need to be modified to maximize the *available* information a survey provides. That is, the information in the data that cannot be 'swamped out' or overwhelmed by the information imposed by model constraints. In short, future efforts at experimental design in the presence of noisy data should somehow incorporate the model regularization method into their optimization scheme.

6.10.5 Field Trials

To this author's knowledge, only a small number of actual field trials have been carried out to assess OED methods in the real world. This is forgivable because it is important to make sure the theory is right before going out to collect real data. It is in our opinion time to bring the theory of geophysical OED into the world with real field examples.

6.10.6 Combining Electrode Misplacement Errors and Random Errors in the OED Exercise

An interesting prospect for one-dimensional resistivity OED would be to incorporate the two distinct error sources, electrode misplacement and random potential field fluctuations, into the same OED exercise. This research has looked at optimal survey design with respect to these noise sources independently of one another, but it might be possible to bring them together under the same optimization algorithm. Appendix B concludes with an expression for the expected data noise due to electrode misplacement. If this expression were incorporated into a data-error covariance matrix (which depends on the Jacobian) that also accounted for random field fluctuations, one might have the beginnings of a fine OED technique that addressed both error sources simultaneously.

6.10.7 Continued Effort to Find Design Patterns

The author concedes that efforts to find patterns in designed surveys over homogeneous media were fairly limited. While the author does not relent on any of the arguments concerning the difficulty of finding such patterns, it is certainly possible that the clever application of clustering algorithms or other artificial intelligence methods might go a long way toward identifying any patterns if they exist.

6.10.8 Integrate Sequential and Global Design Strategies

An interesting possibility would be to create a hybrid optimization technique that couples the two strategies. In the first stage, a sequential design could be performed

to produce an experiment close to, but not quite, optimal. This experiment would then be used to seed a global search algorithm, allowing the algorithm to start in the neighborhood of the optimal experiment and possibly greatly reducing convergence times.

Appendix A

Differential Evolution

Differential Evolution (DE) is an optimization technique for finding the critical point of a multivariable function, developed by Price, Storn and Lampinen (2005). As its name implies, DE is an evolutionary algorithm that, similar to the genetic algorithm, evolves a solution that (with high probability) globally minimizes a multivariable function, even for functions with multiple critical points. One of the distinguishing features of DE is that it operates on model vectors directly, rather than on their binary encodings. This is advantageous because it removes resolution limitations imposed by binary representation.

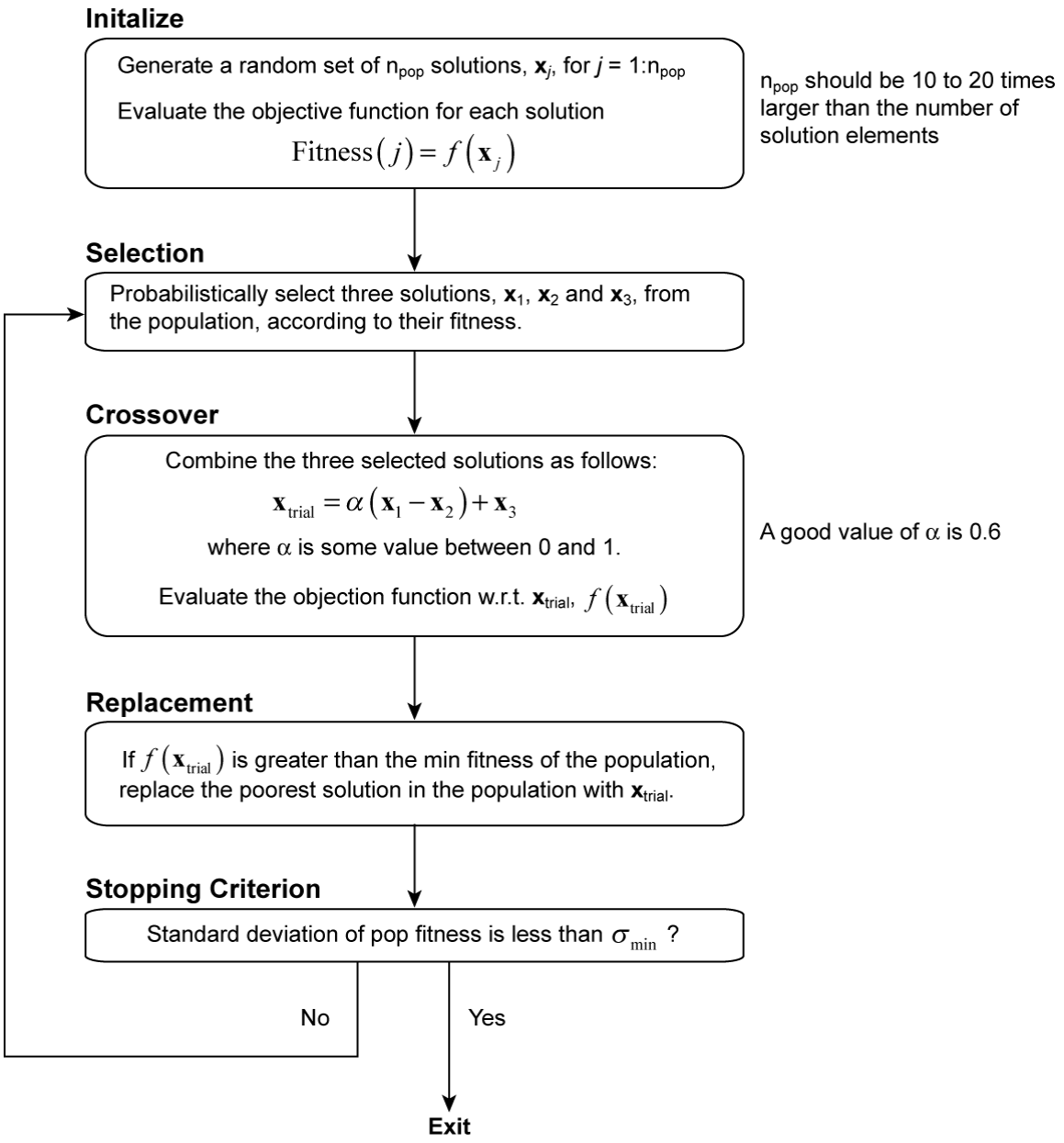
Aside from working with real solutions rather than encoded representations, the primary difference between DE and the basic genetic algorithm is its *crossover* operation. Crossover is effected by picking *three* solutions from the population (using a selection operator that favors the fittest solutions) and combining them to create a new trial solution. The trial solution is created as follows:

$$\mathbf{x}_{trial} = \alpha\left(\mathbf{x}_1 - \mathbf{x}_2\right) + \mathbf{x}_3\,, \tag{A.1}$$

where $\mathbf{x}_1$, $\mathbf{x}_2$ and $\mathbf{x}_3$ are three solutions selected from the population, and α is scalar between 0 and 1. In effect, $\mathbf{x}_1 - \mathbf{x}_2$ is a perturbation that is added to $\mathbf{x}_3$ to produce a trial solution. This perturbation term is customarily scaled by α to expedite convergence.

A nice property of (A.1) is that the magnitude of $\mathbf{x}_1 - \mathbf{x}_2$ becomes small as the population converges on a solution (convergence coincides with population homogenization), so the magnitude of perturbations scales down with convergence. This is similar to simulated annealing, where the magnitude of perturbations are controlled by a reduction schedule.

Flowchart A-1 describes the mechanics of a basic differential evolution algorithm.



Flowchart A-1 Differential Evolution Algorithm (after (Price *et al.*, 2005)).

Appendix B

Minimizing the Effects of Electrode Misplacement

The geometric factor is given by

$$k = 2\pi\left(\frac{1}{AM} - \frac{1}{BM} - \frac{1}{AN} + \frac{1}{BN}\right)^{-1}, \tag{B.1}$$

which is alternately expressed as

$$k = \frac{2\pi}{\frac{1}{x_2 - x_1} - \frac{1}{x_3 - x_1} - \frac{1}{x_4 - x_2} + \frac{1}{x_4 - x_3}}, \tag{B.2}$$

where $x_1 < x_2 < x_3 < x_4$ and

$$\begin{aligned} AM &= L - l = x_2 - x_1 \\ AN &= L + l = x_3 - x_1 \\ BM &= L + l = x_4 - x_2 \\ BN &= L - l = x_4 - x_3 \end{aligned}. \tag{B.3}$$

If the misplacement error for electrode x_j is δx_j, it follows by the Chain Rule that the error in k is approximated by

$$\delta k \cong \frac{\partial k}{\partial x_1}\delta x_1 + \frac{\partial k}{\partial x_2}\delta x_2 + \frac{\partial k}{\partial x_3}\delta x_3 + \frac{\partial k}{\partial x_4}\delta x_4 . \tag{B.4}$$

If we are interested in the percent error, we divide (B.4) by k, producing

$$\frac{\delta k}{k} \cong \frac{1}{k}\frac{\partial k}{\partial x_1}\delta x_1 + \frac{1}{k}\frac{\partial k}{\partial x_2}\delta x_2 + \frac{1}{k}\frac{\partial k}{\partial x_3}\delta x_3 + \frac{1}{k}\frac{\partial k}{\partial x_4}\delta x_4 . \tag{B.5}$$

If we assume that the misplacement errors are random and uncorrelated, it follows from (B.5) that

$$\left(\frac{\delta k}{k}\right)^2 \cong \left(\frac{1}{k}\frac{\partial k}{\partial x_1}\right)^2 (\delta x_1)^2 + \left(\frac{1}{k}\frac{\partial k}{\partial x_2}\right)^2 (\delta x_2)^2 + \left(\frac{1}{k}\frac{\partial k}{\partial x_3}\right)^2 (\delta x_3)^2 + \left(\frac{1}{k}\frac{\partial k}{\partial x_4}\right)^2 (\delta x_4)^2, \quad \text{(B.6)}$$

from which it is seen that $\delta k/k$ is the square root of the sum of the squared percent errors.

If the misplacement errors are random variables that come from the same zero-mean distribution with variance, σ^2, we can determine the expected value of Equation (B.6),

$$\left\langle \left(\frac{\delta k}{k}\right)^2 \right\rangle, \quad \text{(B.7)}$$

by taking advantage of the fact that

$$\left\langle (\delta x_1)^2 \right\rangle = \left\langle (\delta x_2)^2 \right\rangle = \left\langle (\delta x_3)^2 \right\rangle = \left\langle (\delta x_4)^2 \right\rangle = \sigma^2; \quad \text{(B.8)}$$

hence,

$$\left\langle \left(\frac{\delta k}{k}\right)^2 \right\rangle \cong \sigma^2 \left[\left(\frac{1}{k}\frac{\partial k}{\partial x_1}\right)^2 + \left(\frac{1}{k}\frac{\partial k}{\partial x_2}\right)^2 + \left(\frac{1}{k}\frac{\partial k}{\partial x_3}\right)^2 + \left(\frac{1}{k}\frac{\partial k}{\partial x_4}\right)^2 \right]. \quad \text{(B.9)}$$

Note that Equation (B.9) can be simplified to

$$\left\langle \left(\frac{\delta k}{k}\right)^2 \right\rangle = \sigma^2 \left(\nabla_{\mathbf{x}} \ln k \cdot \nabla_{\mathbf{x}} \ln k\right), \quad \text{(B.10)}$$

where $\nabla_{\mathbf{x}}$ is the gradient operator with respect to the electrode positions.

Equation (B.10) expresses the expected squared percent error in the geometric factor due to electrode misplacement (where the misplacements have variance σ^2). If we wish to minimize this expected error, we must choose electrode positions, x_j, that minimize (B.10). The gradient of ln k is

$$\nabla_{\mathbf{x}} \ln k = \left(\frac{1}{x_2 - x_1} - \frac{1}{x_3 - x_1} - \frac{1}{x_4 - x_2} + \frac{1}{x_4 - x_3} \right)^{-1} \begin{bmatrix} \frac{1}{(x_3 - x_1)^2} - \frac{1}{(x_2 - x_1)^2} \\ \frac{1}{(x_4 - x_2)^2} + \frac{1}{(x_2 - x_1)^2} \\ -\frac{1}{(x_3 - x_1)^2} - \frac{1}{(x_4 - x_3)^2} \\ \frac{1}{(x_4 - x_3)^2} - \frac{1}{(x_4 - x_2)^2} \end{bmatrix}. \quad \text{(B.11)}$$

Because the electrodes are placed concentrically about the center of the survey, it is convenient to introduce the following substitutions: $x_1 = -L, x_2 = -\alpha L, x_3 = \alpha L, x_4 = L$. These substitutions allow us to express the entire survey in terms of the outer electrode spacing, L, and a fractional factor, α, which multiplies L to produce the inner electrode spacing. Using these substitutions, Equation (B.11) becomes

$$\nabla_{\mathbf{x}} \ln k = \begin{bmatrix} \frac{1}{L(\alpha^2 - 1)} \\ \frac{1 + \alpha^2}{2\alpha L(1 - \alpha^2)} \\ \frac{1 + \alpha^2}{2\alpha L(\alpha^2 - 1)} \\ \frac{1}{L(1 - \alpha^2)} \end{bmatrix}. \quad \text{(B.12)}$$

Plugging (B.12) into (B.10), we arrive at the following expression for the expected squared percent error of the geometric factor:

$$\left\langle \left(\frac{\delta k}{k} \right)^2 \right\rangle = \sigma^2 \left(\frac{1 + 6\alpha^2 + \alpha^4}{2L^2 \alpha^2 (\alpha^2 - 1)^2} \right), \quad \text{(B.13)}$$

defined over the interval $0 < \alpha < 1$. We need to minimize this expression with respect to α to determine the inner electrode spacing (for fixed L) that produces the smallest error in the geometric factor (recall the inner electrode spacing is given by αL). This is done by setting the derivative of (B.13) with respect to α to zero and solving for α:

$$\frac{\partial}{\partial\alpha}\left\langle\left(\frac{\delta k}{k}\right)^2\right\rangle=\sigma^2\,\frac{\alpha^6+13\alpha^4+3\alpha^2-1}{L^2\alpha^3\left(1-\alpha^2\right)^3}=0\,. \tag{B.14}$$

Equation (B.14) can be simplified to

$$\alpha^6+13\alpha^4+3\alpha^2-1=0 \tag{B.15}$$

without loss of generality, and a root finding algorithm can be employed to find solutions for α (the Mathematica root finder was used here). The roots are

$$\alpha=\{\pm 0.429195,\ \pm 0.652292\,i,\ \pm 3.57194\,i\}, \tag{B.16}$$

and only one root satisfies the condition $0<\alpha<1$,

$$\alpha=0.429195\,. \tag{B.17}$$

Careful algebraic manipulation, using the cubic formula, reveals that the single positive root of (B.15) is given by the expression

$$\alpha=\left(-\frac{13}{3}+\frac{8\sqrt{10}}{3}\cos\left[\frac{1}{3}\left(\pi-\tan^{-1}\frac{3\sqrt{111}}{251}\right)\right]\right)^{1/2}=0.429195\,. \tag{B.18}$$

As an aside, if we take the square root of (B.13) we have the coefficient of variation of the geometric factor. This follows because $\left(\delta k/k\right)^2=\sigma_k^2/\mu_k^2$, where $\delta k^2=\sigma_k^2$ is the expected variance and $k^2=\mu_k^2$ is the expected squared mean of the geometric factor. Taking the square root therefore yields the coefficient of variation, σ_k/μ_k

Appendix C

Pseudosection Survey

The Pseudosection survey for an array of 10 electrodes is tabulated at right. **A** and **B** designate the positive and negative transmitter electrodes, respectively; **M** and **N** designate the positive and negative receiver electrodes, respectively.

	A	B	M	N
1	1	2	3	4
2	1	2	4	5
3	1	2	5	6
4	1	2	6	7
5	1	2	7	8
6	1	2	8	9
7	1	2	9	10
8	2	3	4	5
9	2	3	5	6
10	2	3	6	7
11	2	3	7	8
12	2	3	8	9
13	2	3	9	10
14	3	4	5	6
15	3	4	6	7
16	3	4	7	8
17	3	4	8	9
18	3	4	9	10
19	4	5	6	7
20	4	5	7	8
21	4	5	8	9
22	4	5	9	10
23	5	6	7	8
24	5	6	8	9
25	5	6	9	10
26	6	7	8	9
27	6	7	9	10
28	7	8	9	10

Appendix D

ERL Survey

The ERL Survey for an array of 10 electrodes is tabulated on the next page. The same descriptions for **A**, **B**, **M** and **N** apply as in Appendix C.

	A	B	M	N
1	1	2	3	4
2	1	2	4	5
3	1	2	5	6
4	1	2	6	7
5	1	2	7	8
6	1	2	8	9
7	1	2	9	10
8	2	3	4	5
9	2	3	5	6
10	2	3	6	7
11	2	3	7	8
12	2	3	8	9
13	2	3	9	10
14	3	4	5	6
15	3	4	6	7
16	3	4	7	8
17	3	4	8	9
18	3	4	9	10
19	4	5	6	7
20	4	5	7	8
21	4	5	8	9
22	4	5	9	10
23	5	6	7	8
24	5	6	8	9
25	5	6	9	10
26	6	7	8	9
27	6	7	9	10
28	7	8	9	10
29	1	2	3	4
30	1	2	4	5
31	1	2	5	6
32	1	2	6	7
33	1	2	7	8
34	1	2	8	9
35	1	2	9	10
36	1	3	4	5
37	1	3	5	6
38	1	3	6	7
39	1	3	7	8
40	1	3	8	9
41	1	3	9	10
42	1	4	2	3
43	1	4	5	6
44	1	4	6	7
45	1	4	7	8
46	1	4	8	9
47	1	4	9	10
48	1	5	2	3
49	1	5	3	4
50	1	5	6	7
51	1	5	7	8
52	1	5	8	9
53	1	5	9	10
54	1	6	2	3
55	1	6	3	4
56	1	6	4	5
57	1	6	7	8
58	1	6	8	9
59	1	6	9	10
60	1	7	2	3
61	1	7	3	4
62	1	7	4	5
63	1	7	5	6
64	1	7	8	9
65	1	7	9	10
66	1	8	2	3
67	1	8	3	4
68	1	8	4	5
69	1	8	5	6
70	1	8	6	7
71	1	8	9	10
72	1	9	2	3
73	1	9	3	4
74	1	9	4	5
75	1	9	5	6
76	1	9	6	7
77	1	9	7	8
78	1	10	2	3
79	1	10	3	4
80	1	10	4	5
81	1	10	5	6
82	1	10	6	7
83	1	10	7	8
84	1	10	8	9
85	9	10	1	2
86	9	10	2	3
87	9	10	3	4
88	9	10	4	5
89	9	10	5	6
90	9	10	6	7
91	9	10	7	8
92	8	10	1	2
93	8	10	2	3
94	8	10	3	4
95	8	10	4	5
96	8	10	5	6
97	8	10	6	7
98	7	10	1	2
99	7	10	2	3
100	7	10	3	4
101	7	10	4	5
102	7	10	5	6
103	7	10	8	9
104	6	10	1	2
105	6	10	2	3
106	6	10	3	4
107	6	10	4	5
108	6	10	7	8
109	6	10	8	9
110	5	10	1	2
111	5	10	2	3
112	5	10	3	4
113	5	10	6	7
114	5	10	7	8
115	5	10	8	9
116	4	10	1	2
117	4	10	2	3
118	4	10	5	6
119	4	10	6	7
120	4	10	7	8
121	4	10	8	9
122	3	10	1	2
123	3	10	4	5
124	3	10	5	6
125	3	10	6	7
126	3	10	7	8
127	3	10	8	9
128	2	10	3	4
129	2	10	4	5
130	2	10	5	6
131	2	10	6	7
132	2	10	7	8
133	2	10	8	9
134	1	10	2	3
135	1	10	3	4
136	1	10	4	5
137	1	10	5	6
138	1	10	6	7
139	1	10	7	8
140	1	10	8	9

References

Arnason, K., 1984. The effect of finite potential electrode separation on Schlumberger soundings, *SEG Technical Program Expanded Abstracts*, **3**(1), 129-132

Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross earth data, *Philosophical Transactions - Royal Society of London, Physical Sciences and Engineering*, **226**(1173), 123-192

Barth, N. & Wunsch, C., 1990. Oceanographic Experiment Design by Simulated Annealing, *Journal of Physical Oceanography*, 1249-1263

Bartle, R.G. & Sherbert, D.R., 2000. *Introduction to real analysis*, Wiley, New York

Beard, L.P. & Morgan, F.D., 1991. Assessment of 2-D resistivity structures using 1-D inversion, *Geophysics*, **56**(6), 874-883

Bullen, P.S., 2003. *Handbook of means and their inequalities*, Kluwer Academic Publishers, Dordrecht ; Boston

Curtis, A., 1999a. Optimal experiment design: cross-borehole tomographic examples, *Geophys J Int*, **136**(3), 637-650

Curtis, A., 1999b. Optimal design of focused experiments and surveys, *Geophys J Int*, **139**(1), 205-215

Curtis, A. & Maurer, H., 2000. Optimizing the design of geophysical experiments: Is it worthwhile?, *The Leading Edge*, **19**(10), 1058-1062

Curtis, A., Michelini, A., Leslie, D. & Lomax, A., 2004. A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys, *Geophys J Int*, **157**(2), 595-606

Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics*, **62**(5), 1524-1532

Curtis, A. & Spencer, C., 1999c. Survey design strategies for linearized nonlinear inversion, *SEG Technical Program Expanded Abstracts*, **18**(1), 1775-1778

Daily, W., Ramirez, A., Binley, A. & LeBrecque, D., 2004. The Meter Reader---Electrical resistance tomography, *The Leading Edge*, **23**(5), 438-442

Dey, A. & Morrison, H.F., 1979. Resistivity Modeling for Arbitrarily Shaped 2-Dimensional Structures, *Geophysical Prospecting*, **27**(1), 106-136

Du, D. & Pardalos, P.M., 1998. *Handbook of combinatorial optimization*, Kluwer Academic Publishers, Boston

Forsell, U. & Ljung, L., 2000. Some results on optimal experimental design, *Automatica*, **36**, 749-759

Furman, A., Ferre, T.P.A. & Heath, G.L., 2007. Spatial focusing of electrical resistivity surveys considering geologic and hydrologic layering, *Geophysics*, **72**(2), F65-F73

Furman, A., Ferre, T.P.A. & Warrick, A.W., 2004. Optimization of ERT surveys for monitoring transient hydrological events using perturbation sensitivity and genetic algorithms, *Vadose Zone J*, **3**(4), 1230-1239

Golub, G.H. & Van Loan, C.F., 1996. *Matrix computations*, Johns Hopkins University Press, Baltimore

Hardt, M. & Scherbaum, F., 1994. The Design of Optimum Networks for Aftershock Recordings, *Geophys J Int*, **117**(3), 716-726

Jolliffe, I.T., 2002. *Principal component analysis*, Springer, New York

Koefoed, O., Patra, H.P. & Mallick, K., 1979. *Geosounding principles*, Elsevier Scientific Pub. Co.; distributors for the U.S., Elsevier/North-Holland, Amsterdam ; New York, New York

Lanczos, C., 1956. *Applied analysis*, Prentice-Hall, Englewood Cliffs, N.J.,

Lanczos, C., 1961. *Linear differential operators*, Van Nostrand, London, New York,

Levenberg, K., 1944. A Method for the Solution of Certain Non-Linear Problems in Least Squares, *Quarterly of Applied Mathematics*, **2**, 164-168

Liu, C. & Shen, L.C., 1991. Response of Electromagnetic-Pulse Logging Sonde in Axially Symmetrical Formation, *Ieee T Geosci Remote*, **29**(2), 214-221

Loke, M.H. & Barker, R.D., 1995. Least-Squares Deconvolution of Apparent Resistivity Pseudosections, *Geophysics*, **60**(6), 1682-1690

Mackie, R.L., Bennett, B.R. & Madden, T.R., 1988. Long-Period Magnetotelluric Measurements near the Central California Coast - a Land-Locked View of the Conductivity Structure under the Pacific-Ocean, *Geophysical Journal-Oxford*, **95**(1), 181-194

Mackie, R.L. & Madden, T.R., 1993b. Conjugate Direction Relaxation Solutions for 3-D Magnetotelluric Modeling, *Geophysics*, **58**(7), 1052-1057

Madden, T.M. & Mackie, R.L., 1989. Three-dimensional magnetotelluric modelling and inversion, *P Ieee*, **77**(2), 318-333

Madden, T.R., 1972. Transmission systems and network analogies to geophysical forward and inverse problems, *ONR Technical Report*, **72-73**

Marquardt, D.W., 1963. An Algorithm for Least-Squares Estimation of Nonlinear Parameters, *Journal of the Society for Industrial and Applied Mathematics*, **11**(2), 431-441

Maurer, H. & Boerner, D.E., 1998a. Optimized and robust experimental design: a non-linear application to EM sounding, *Geophys J Int*, **132**(2), 458-468

Maurer, H. & Boerner, D.E., 1998b. Optimized design of geophysical experiments, *The Leading Edge*, **17**(8), 1119

Maurer, H., Boerner, D.E. & Curtis, A., 2000. Design strategies for electromagnetic geophysical surveys, *Inverse Probl*, **16**(5), 1097-1117

Mcgillivray, P.R. & Oldenburg, D.W., 1990. Methods for Calculating Frechet Derivatives and Sensitivities for the Nonlinear Inverse Problem - a Comparative-Study, *Geophysical Prospecting*, **38**(5), 499-524

Menke, W., 1989. *Geophysical data analysis : discrete inverse theory*, Academic Press, San Diego

Mundry, E., 1980. The effect of a finite distance between potential electrodes on Schlumberger resistivity measurements---A simple correction graph, *Geophysics*, **45**(12), 1872-1875

Narayanan, C., Rao, V.N.R. & Kaihatu, J.M., 2004. Model parameterization and experimental design issues in nearshore bathymetry inversion, *J Geophys Res-Oceans*, **109**(C8), -

Parasnis, D.S., 1997. *Principles of applied geophysics*, Chapman and Hall, London ; New York

Pelton, W.H., Rijo, L. & Swift, C.M., 1978. Inversion of 2-Dimensional Resistivity and Induced-Polarization Data, *Geophysics*, **43**(4), 788-803

Price, K.V., Storn, R.M. & Lampinen, J.A., 2005. *Differential evolution : a practical approach to global optimization*, Springer, Berlin

Rabinowitz, N. & Steinberg, D.M., 1990. Optimal configuration of a seismographic network: A statistical approach, *Bulletin of the Seismological Society of America*, **80**(1), 187-196

Rodi, W. & Mackie, R.L., 2001. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, **66**(1), 174-187

Sharma, P.V., 1997. *Environmental and engineering geophysics*, Cambridge University Press, Cambridge, U.K. ; New York, NY, USA

Shi, W., 1998. *Advanced modeling and inversion techniques for three-dimensional geoelectrical surveys*, Department of Earth, Atmospheric, and Planetary Sciences, Massachusetts Institute of Technology, Cambridge, MA

Simms, J.E. & Morgan, F.D., 1992. Comparison of four least-squares inversion schemes for studying equivalence in one-dimensional resistivity interpretation, *Geophysics*, **57**(10), 1282-1293

Smith, M.L., Scales, J.A. & Fischer, T.L., 1992. Global search and genetic algorithms, *The Leading Edge*, **11**(1), 22-26

Spitzer, K. & Chouteau, M., 2003. A dc resistivity and IP borehole survey at the Casa Berardi gold mine in northwestern Quebec, *Geophysics*, **68**(2), 453-463

Steinberg, D.M., Rabinowitz, N., Shimshoni, Y. & Mizrachi, D., 1995. Configuring a seismographic network for optimal monitoring of fault lines and multiple sources, *Bulletin of the Seismological Society of America*, **85**(6), 1847-1857

Strang, G., 2003. *Introduction to linear algebra*, Wellesley-Cambridge, Wellesly, MA

Stummer, P., Maurer, H. & Green, A.G., 2004. Experimental design: Electrical resistivity data sets that provide optimum subsurface information, *Geophysics*, **69**(1), 120-139

Stummer, P., Maurer, H., Horstmeyer, H. & Green, A.G., 2002. Optimization of DC resistivity data acquisition: Real-time experimental design and a new multielectrode system, *Ieee T Geosci Remote*, **40**(12), 2727-2735

Swift, C.M., 1971. Theoretical Magnetotelluric and Turam Response from Two-Dimensional Inhomogeneities, *Geophysics*, **36**(1), 38-52

Tarantola, A., 2005. *Inverse problem theory and methods for model parameter estimation*, Society for Industrial and Applied Mathematics, Philadelphia, PA

Telford, W.M., Geldart, L.P. & Sheriff, R.E., 1990. *Applied geophysics*, Cambridge University Press, Cambridge [England] ; New York

Tikhonov, A.N. & Arsenin, V.I., 1977. *Solutions of ill-posed problems*, Winston; distributed solely by Halsted Press, Washington; New York

Tripp, A.C., Hohmann, G.W. & Swift, C.M., 1984. Two-Dimensional Resistivity Inversion, *Geophysics*, **49**(10), 1708-1717

van den Berg, J., Curtis, A. & Trampert, J., 2003. Optimal nonlinear Bayesian experimental design: an application to amplitude versus offset experiments, *Geophys J Int*, **155**(2), 411-421

Wang, H.N., 2003. Simultaneous reconstruction of geometric parameter and resistivity around borehole in horizontally stratified formation from multiarray induction logging data, *Ieee T Geosci Remote*, **41**(1), 81-89

Wilkinson, P.B., Chambers, J.E., Meldrum, P.I., Ogilvy, R.D. & Caunt, S., 2006a. Optimization of array configurations and panel combinations for the detection and imaging of abandoned mineshafts using 3D cross-hole electrical resistivity tomography, *J Environ Eng Geoph*, **11**(3), 213-221

Wilkinson, P.B., Meldrum, P.I., Chambers, J.E., Kuras, O. & Ogilvy, R.D., 2006b. Improved strategies for the automatic selection of optimized sets of electrical resistivity tomography measurement configurations

doi:10.1111/j.1365-246X.2006.03196.x, *Geophys J Int*, **167**(3), 1119-1126

Yang, F.W. & Ward, S.H., 1984. Inversion of Borehole Normal Resistivity Logs, *Geophysics*, **49**(9), 1541-1548

Zemanian, A.H. & Anderson, B., 1987. Modeling of Borehole Resistivity Measurements Using Infinite Electrical Grids, *Geophysics*, **52**(11), 1525-1534

Zhang, J., Mackie, R.L. & Madden, T.R., 1995. 3-D Resistivity Forward Modeling and Inversion Using Conjugate Gradients, *Geophysics*, **60**(5), 1313-1325

Zhang, Z. & Xiao, J., 2001. Inversions of surface and borehole data from large-loop transient electromagnetic system over a 1-D earth, *Geophysics*, **66**(4), 1090-1096