Recent advances in optimized geophysical survey design

GEOPHYSICS, VOL. 75, NO. 5 (SEPTEMBER-OCTOBER 2010); P. 75A177–75A194, 13 FIGS., 1 TABLE.
10.1190/1.3484194

Recent advances in optimized geophysical survey design

Hansruedi Maurer[1], Andrew Curtis[2], and David E. Boerner[3]

ABSTRACT

Survey design ultimately dictates the quality of subsurface information provided by practical implementations of geophysical methods. It is therefore critical to design experimental procedures that cost effectively produce those data that maximize the desired information. This review cites recent advances in statistical experimental design techniques applied in the earth sciences. Examples from geoelectrical, crosswell and surface seismic, and microseismic monitoring methods are included. Using overdetermined 1D and 2D geoelectrical examples, a minor subset of judiciously chosen measurements provides a large percentage of the information content theoretically offered by the geoelectrical method. In contrast, an underdetermined 2D seismic traveltime tomography design study indicates that the information content increases almost linearly with the amount of traveltime data (source-receiver pairs) considered until the underdeterminancy is reduced substantially. An experimental design study of frequency-domain seismic-waveform inversion experiments reveals that a few optimally chosen frequencies offer as much subsurface information as the full bandwidth. A nonlinear experimental design for a seismic amplitude-versus-angle survey identifies those incidence angles most important for characterizing a reservoir. A nonlinear design example shows that designing microseismic monitoring surveys based on array aperture is a poor strategy that almost certainly leads to suboptimal designs.

INTRODUCTION AND HISTORICAL BACKGROUND

Geophysical surveys estimate specific aspects of the earth's geological structure and composition. Such problems usually reduce to estimating the distribution of petrophysical properties of the subsurface from measured data. To facilitate geological interpretation, numerous standardized data-acquisition procedures have been established, and data processing and inversion algorithms have been developed to extract estimates of the desired subsurface parameters. The importance of new processing, analysis, and inversion tools is widely appreciated and is the subject of considerable research effort. However, it seems that much less attention has been paid to honing the design of data-acquisition procedures (the survey design). This paper focuses on ways to maximize the information content of geophysical-survey data sets.

An acceptable survey design should ensure acquisition of those data that best resolve specific subsurface features or parameters of interest (Maurer and Boerner, 1998b; Curtis and Maurer, 2000). This goal is important because no amount of subsequent data processing or analysis can ever compensate for inadequate or missing data that could have contributed significantly to resolving geological targets. Appropriate survey design is therefore critical to justify the cost of the experiment in terms of the robustness, accuracy, and precision of recovered geological information.

Philosophically, we desire an optimal data set that best resolves specific subsurface petrophysical properties of the subsurface model (e.g., composition, locations of discontinuities, porosity, and fluid saturation over a wide or a more focused subsurface volume) from the acquired data. Yet in practice in most geophysical experiments, model parameters are nonlinearly related to the observable data, making it extremely computationally demanding to calculate resulting constraints on parameter values. Also, the inverse problem of constraining model-parameter estimates from data observations is often ill posed (no solution exactly satisfies the data) and may not have a unique solution. Hence, the process of deciding which data to observe to optimally constrain the interpretation of petrophysical parameters is not straightforward.

One simple, often-used approach to the problem of insufficient or inadequate data is to acquire as much data as possible, including potentially redundant data. However, cost is often a major consideration in geophysical surveying, implying that we need to find means to acquire optimal data in terms of information content while maintaining a favorable cost/benefit ratio. Many geophysical-survey de-

Manuscript received by the Editor 25 January 2010; revised manuscript received 11 May 2010; published online 14 September 2010.
[1]ETH Zürich, Zürich, Switzerland. E-mail: maurer@aug.ig.erdw.ethz.ch.
[2]University of Edinburgh, School of GeoSciences, Grant Institute, Edinburgh, U. K. E-mail: andrew.curtis@ed.ac.uk.
[3]Geological Survey of Canada, Ottawa, Canada. E-mail: dboerner@nrcan.gc.ca.

signs used in the past have been heuristic in the sense that they were generated from a combination of highly simplified theoretical investigations, repeated simulations with numerical or analog models of very simple geological situations (e.g., a perfectly stratified earth), and experience gained from actual field surveys. Often, standard designs have been applied solely because they were required to use existing analysis and inversion tools.

Although these approaches have been extremely successful in many cases, application of heuristic survey designs can be of dubious value in dealing with logistical or instrumentation constraints, or with complex subsurface environments. Moreover, most survey designs were constructed to obtain a complete image of the subsurface, and they may not be appropriate for surveys conducted to resolve specific types or subsets of subsurface petrophysical properties (e.g., those from a specific depth interval of interest). With the current capability for simulating the geophysical response of highly complex geological models and almost arbitrary survey designs, it would seem appropriate to explore whether better subsurface information can be obtained from nonstandard approaches to experimental configurations.

Qualitatively, the benefit of a geophysical survey can be defined as the resulting net increase in resolution of the model parameters of interest. For example, an expected outcome of a geophysical survey might be to determine only the depth to a particular interface. Alternatively, one may wish to resolve all physical-property variations within a restricted 3D volume. The extent to which these expectations are realized will dictate the success or failure of the experiment. Within this seemingly broad spectrum of possibilities lies a common element: the reliability of the parameter information obtained by inverting the geophysical data. We define the quantitative benefit of a geophysical experiment to be directly proportional to the resolution or accuracy of the parameters necessary to answer specific questions of interest. Resolution and accuracy can be estimated formally via application of linear or nonlinear inverse theory, so conceptually it is possible to find desirable survey layouts with optimization algorithms.

Figure 1 illustrates our notion of the relationship between the benefits and costs of performing a geophysical experiment. Although this graph is qualitative, we expect the behavior of the limits to be generally correct. For example, there is no benefit before data are acquired, although fixed costs are certainly incurred in mobilizing equipment and preparing for data acquisition. Additionally, benefits are likely to be subject to the concept of diminishing returns such

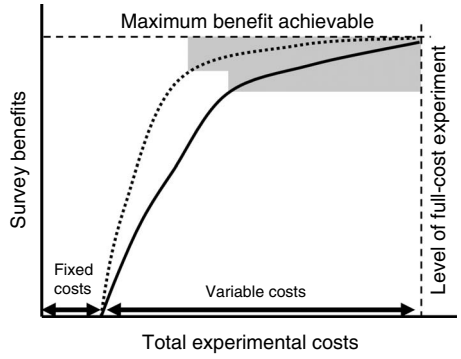

Figure 1. Schematic representation of cost/benefit relationships. Shading is area of diminishing returns. Dotted line is optimized experimental design. Solid line is standard experimental design.

that ever-increasing data acquisition is likely to result in increasingly redundant data rather than continually accruing benefits. Aside from the behavior at the limits of zero and infinite data, the details of cost/benefit curves are difficult to predict, primarily because of the generally nonlinear relationship between data and model parameters. Nonlinearity effectively means that the values of the model parameters influence the information content, and this makes it difficult to characterize, a priori, the information content of specific observed data.

To understand Figure 1 more fully, consider a seismic crosshole survey. In this case, the solid line might represent those designs consisting of increasing numbers of evenly spaced shots and receivers. The variable cost for such designs will be approximately proportional to the amount of data collected and hence inversely proportional to the shot and receiver spacing. On the other hand, one might find survey designs that maximize cost savings for a given level of subsurface feature resolution or that maximize the benefit achievable at each given cost. Each of these objectives would result in a unique cost/benefit curve where the cost/benefit ratio has been optimized such that the curve is shifted to a higher benefit level for any particular survey expenditure (dashed curve in Figure 1).

In reality, there is no single 2D graph equivalent to Figure 1 because multiple dimensions of data-acquisition parameters could be varied for any given geophysical experiment, each defining a different cost/benefit function. Constructing optimized cost/benefit functions for multidimensional formulations is thus founded on the theory of statistical experimental design (SED). (For a geophysics-related tutorial on SED, see Curtis [2004a, 2004b].) SED techniques were originally developed for optimizing industrial manufacturing processes. Cox (1958) formulated some of the first ideas about experimental design, and Fedorov (1972) is credited with the first example of optimal experimental design by Atkinson and Donev (1992), who also document much of the history of important developments in industrial systems.

G. Taguchi developed and implemented design methods (often called Taguchi methods [Taguchi, 1987]) across a broad spectrum of Japanese industrial processes such that final products are as robust as possible to variations in the manufacturing steps. Over only a few years, this led to a revolutionary increase in the quality of Japanese manufacturing. Although Taguchi's methods were often simple, he systematized an approach to quality control that did not exist previously. Their simplicity makes these methods so generally applicable.

The first attempts at optimized experimental design in geophysics were devoted to earthquake-location problems (Kijko, 1977; Rabinowitz and Steinberg, 1990; Hardt and Scherbaum, 1994), ocean tomography (Barth and Wunsch, 1990), geoelectrical soundings (Glenn and Ward, 1976), and magnetotelluric investigations (Jones and Foster, 1986). Building on ideas presented in these early papers, statistical experimental design has been continuously refined.

Originally, statistical experimental design was treated as an optimization problem that could be solved with global optimizers such as genetic algorithms (e.g., Hardt and Scherbaum, 1994), simulated annealing (e.g., Barth and Wunsch, 1990), or multilevel coordinate search (e.g., Ajo-Franklin, 2009). Unfortunately, global optimizers were not yet efficient for larger-scale problems. Therefore, Curtis et al. (2004) introduced a sequential approach that begins with several design parameters (e.g., each describing a datum to be recorded) that are reduced in a stepwise fashion by removing the most redundant datum at each step. This algorithm is effective in designing simple

tomographic surveys and (micro-)seismic monitoring surveys with many design parameters. Curtis and Wood (2004) showed that the same algorithm can be used to design optimal interrogation strategies to obtain geological information from human experts, and Stummer et al. (2004) introduced sequential algorithms that work in the opposite direction, beginning from a minimal design and incrementing it with the most informative datum at each step; they applied this to design geoelectric experiments. Similar approaches were presented by Wilkinson et al. (2006) and Coles and Morgan (2009).

To control computational requirements, many experimental design studies approximate nonlinear relationships between earth-model parameters and the observed data with linear functions. The rationale is that any inherent nonlinearity in the physical system is of second-order importance for small perturbations of model parameters. Although in many cases this approximation may be valid, a more rigorous and robust framework for tackling concurrent or sequential experimental design optimization is offered by nonlinear design methods. Algorithms for nonlinear problems are generally computationally expensive and thus have been restricted to fairly specific problems that can be parameterized with relatively few (~10) parameters. Examples include designing surface seismic receiver density or data-processing strategies for optimal amplitude variation with offset/amplitude variation with angle (AVO/AVA) analysis (van den Berg et al., 2003, 2005; Guest and Curtis 2009, 2010) or designing optimal receiver locations for earthquake or microseismic monitoring surveys (Winterfors and Curtis, 2008). Hyperparameterization methods, recently applied in linearized geophysical problems (Ajo-Franklin, 2009), are being extended successfully to fully nonlinear design methods applicable for full 2D or 3D seismic surveys (Guest and Curtis, 2010).

With the increasing success of statistical experimental design for geophysical applications, numerous methodological studies and application improvements have been published (see Table 1). In this contribution, we first review the theory of statistical experimental design. Then we show a few examples that demonstrate the benefits and limits of applying experimental design concepts to geophysical problems. In the concluding section, we critically review the achievements made over the past decade and outline potentially fruitful avenues of future research.

THEORY

The most important requirement in selecting the experimental parameters for a geophysical survey design is to be clear in specifying the geological and operational objectives. Geological objectives may range from being specific (locate the oil/water contact close to a planned well trajectory) to vague (create an image of some subsurface volume to identify and locate anomalous features that may be of interest). Operational objectives can often be more precisely defined and generally are based on the desire to minimize or control survey costs and the risk of operational failure.

Once fully declared, all survey objectives are encoded into a single mathematical objective function that is intentionally chosen to have a minimum when a survey design best meets the desired objectives. Selecting the optimal survey design can then be achieved with one of several numerical minimization algorithms that vary parameters controlling the design until the minimum is attained.

Of the two components of the objective function, the operational objectives are not easily amenable to generic analysis because they reflect a variety of highly specific influences, including price fluctuations, the specific survey equipment used, contractor experience, overhead costs, geographic location, and even weather conditions and season. Although accommodating operational objectives is a

Table 1. Categories and representative published works of methodological studies and application improvements of statistical experimental design for geophysical applications.

Application	Representative published work
Seismic tomography	Ajo-Franklin, 2009 Brenders and Pratt, 2007 Curtis and Snieder, 1997 Curtis, 1999a, 1999b Maurer et al., 2009 Sirgue and Pratt, 2004
Seismic reflection surveys	Liner et al., 1999
AVO/AVA/AVAz surveys or processing strategies	Coles and Curtis, 2010 Guest and Curtis, 2009, 2010 van den Berg et al., 2003
Earthquake or microseismic location surveys	Curtis et al., 2004 Hardt and Scherbaum, 1994 Lomax et al., 2009 Rabinowitz and Steinberg, 1990 Steinberg et al., 1995 Winterfors and Curtis, 2008, 2010
Electric or electromagnetic surveys	Coles and Morgan, 2009 Coscia et al., 2008 Furman et al., 2007 Hennig et al., 2008 Loke and Wilkinson, 2009 Maurer and Boerner, 1998a Maurer et al., 2000 Oldenborger and Routh, 2009 Stummer et al., 2002, 2004 Wilkinson et al., 2006
Other methodological advances	Curtis and Wood, 2004 Haber et al., 2008 Routh et al., 2005

critically important task in geophysical surveying that typically requires great skill and experience, for this paper we assume that an operational objective exists and that a single (potentially quite complex) cost model can be used within the survey-design context. We therefore focus on methods to achieve the other set of objectives — evaluating and optimizing the information contained in the survey data, subject to constraints from the specified cost model.

In practice, extracting quantitative subsurface information from geophysical data requires inversion or inference methods. These methods effectively transform observed geophysical data into constraints on a specific set of parameters that describe a model of subsurface petrophysical properties (e.g., reservoir porosity, fluid properties). Invariably, some subsurface parameters are well constrained by data (e.g., often those in the near surface), but others may be completely unconstrained. Designing geophysical surveys that produce data which best achieve the survey objectives requires a comprehensive understanding of (1) the inversion or inference procedure and (2) how we encode the specific goals of the experiment into the design objective function. We therefore now review aspects of inverse theory that are indispensable in order to follow the development of design theory.

Inverse theory

Once any geophysical survey has been completed, a data set $\mathbf{d}$ will be available for geological interpretation. Let $\mathbf{m}$ denote the subsurface model consisting of a vector of spatially defined petrophysical parameters in the parameter space $\mathbf{M}$. Inversion is the process used to translate information in $\mathbf{d}$ to constrain estimates of $\mathbf{m}$.

An essential component of inversion is the ability to calculate the geophysical data response of any particular subsurface model. This is the so-called "forward problem," which we express in terms of the function $\mathbf{f}$ that relates $\mathbf{m}$ to a vector of predicted data: $\mathbf{d} = \mathbf{f}(\mathbf{m})$.

It is useful to represent uncertainty in $\mathbf{m}$ and $\mathbf{d}$ with probability density functions (PDFs; Tarantola, 2005). Let $p(\mathbf{d})$ be the PDF describing the uncertainty in $\mathbf{d}$ from all uncorrected measurement errors (from instrumental bias, noise, and any other artifacts). Similarly, $p(\mathbf{m})$ is the so-called prior PDF, representing all information about $\mathbf{m}$ existing before (prior to) inversion. The mathematical formulation of the forward problem may also contain known limitations in expressing the physical relationships between $\mathbf{d}$ and $\mathbf{m}$, for example, using linearized approximations to nonlinear relations; so we use $\Theta(\mathbf{d},\mathbf{m})$ to denote the PDF describing this potentially uncertain relationship.

Let's say that if $\mathbf{f}(\mathbf{m})$ is a complete and accurate representation of the forward problem physics, the conditional PDF $\Theta(\mathbf{d}|\mathbf{m})$ (the probability distribution of $\mathbf{d}$ when $\mathbf{m}$ is fixed at a particular value) is described by $\Theta(\mathbf{d}|\mathbf{m}) \propto \delta[\mathbf{d} - \mathbf{f}(\mathbf{m})]$, where δ is the Dirac delta function. Also, if the forward problem is assumed to place only the minimum constraints possible on $\mathbf{m}$, the PDF describing this state of information about $\mathbf{m}$ is called the homogeneous (or null) distribution, represented by $\mu(\mathbf{m})$. No PDF exists that describes zero information, but some information about $\mathbf{m}$ always exists in practice (the finiteness, positivity, or possible range of parameter values, for example).

With these assumptions, the joint PDF describing the forward physics is given by

$$\Theta(\mathbf{d},\mathbf{m}) = k'\,\delta[\mathbf{d} - \mathbf{f}(\mathbf{m})]\mu(\mathbf{m}) \tag{1}$$

for normalizing constant k' (Tarantola and Valette, 1982). A solution to the inverse problem is found by combining the information in the data $p(\mathbf{d})$, the PDF expressing existing or prior knowledge about the model $p(\mathbf{m})$, and the accuracy of the forward model $\Theta(\mathbf{d},\mathbf{m})$ relating $\mathbf{m}$ and $\mathbf{d}$ (Tarantola and Valette, 1982; Tarantola, 2005). This is achieved in a probabilistic framework by constructing a PDF Q describing the total resultant state of posteriori (postsurvey) information:

$$Q(\mathbf{d},\mathbf{m}) = k\frac{p(\mathbf{d})\Theta(\mathbf{d},\mathbf{m})p(\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})}, \tag{2}$$

where k is a normalizing constant and $\mu(\mathbf{d},\mathbf{m})$ is the homogeneous or null distribution over data $\mathbf{d}$ and parameters $\mathbf{m}$.

The joint PDF in equation 2 contains all information pertinent to constraining estimates of $\mathbf{m}$ (from existing knowledge of the earth, the observed data, and the physics underlying the geophysical experiment). The final, posteriori state of information about the parameters of interest $\mathbf{m}$ is given by integrating over $\mathbf{d}$ (essentially summing the PDFs of $\mathbf{m}$ for every possible data set $\mathbf{d}$ consistent with measurements) to obtain the marginal PDF:

$$Q(\mathbf{m}) = kp(\mathbf{m})\int_{\mathbf{D}}\frac{p(\mathbf{d})\Theta(\mathbf{d},\mathbf{m})}{\mu(\mathbf{d},\mathbf{m})}d\mathbf{d}. \tag{3}$$

Equation 3 is the general, probabilistic solution to the inverse problem from the available data because it describes the uncertainty (PDF) characterizing the parameters $\mathbf{m}$, given all available information about the physics and recorded data. The integral in equation 3 is called the likelihood function $L(\mathbf{m})$, which measures how well any model $\mathbf{m}$ explains data $\mathbf{d}$. Hence, dropping constant k, we obtain

$$Q(\mathbf{m}) \propto p(\mathbf{m})L(\mathbf{m}). \tag{4}$$

Uncertainty in the data $p(\mathbf{d})$ is often assumed to follow a Gaussian distribution, described by mean $\mathbf{d}_0$ and covariance matrix $\mathbf{C}_\mathrm{d}$. Further, if uncertainties in the forward model $\Theta(\mathbf{d},\mathbf{m})$ are negligible, then Θ takes the form in equation 1. It is also usually assumed that $\mu(\mathbf{d},\mathbf{m})$ is a uniform (constant) distribution within some reasonable bounds. In a system of discrete data and model parameters, the likelihood function then becomes approximately

$$L(\mathbf{m}) = \exp\left\{-\frac{1}{2}[\mathbf{d}_0 - \mathbf{f}(\mathbf{m})]^\mathrm{T}\mathbf{C}_\mathbf{d}^{-1}[\mathbf{d}_0 - \mathbf{f}(\mathbf{m})]\right\}, \tag{5}$$

where $L(\mathbf{x})$ provides the relative probability of any set of parameter values $\mathbf{m}$ being the true values, given the current data measurements. Integrating over all of $\mathbf{M}\times\mathbf{D}$ to find k in equation 3 is often computationally intractable, so we usually must use the product of the prior $p(\mathbf{m})$ and the nonnormalized likelihood $L(\mathbf{m})$ given in equation 4 as sufficient information about the solution to the inverse problem.

Ultimately, the solution to the probabilistic inverse problem is a posteriori PDF, which allows one to extract information about $\mathbf{m}$ and uncertainty in the estimation. If $\mathbf{f}$ is nonlinear or $p(\mathbf{d})$ is non-Gaussian, PDF $Q(\mathbf{m})$ may be highly irregular and include multiple local maxima. Algorithms used to estimate Q must therefore then be made robust to such occurrences, typically by pseudorandom (Monte Carlo) sampling of $\mathbf{M}$. For each sample $\mathbf{m}_i$, the value proportional to $Q(\mathbf{m}_i)$ is calculated using equation 4. A good overview of useful Monte Carlo algorithms is included in Press et al. (2007).

Sampling all pertinent parts of the model space is often computationally unfeasible when there are too many parameters, prior infor-

mation with which to constrain the range of model-parameter estimates is too limited, or the forward relation in Θ is highly nonlinear. In such cases, it is usual to approximate the posteriori distribution $Q(\mathbf{m})$ to create a tractable inverse problem solution, usually by approximating $\mathbf{f}(\mathbf{m})$ with a linear function. However, when the forward relationship $\mathbf{f}$ is truly nonlinear, iterative application of linear approximations complicates the interpretation of inversion results. Inaccuracies in the linear approximations cannot be observed in the computational results from the linearized model alone and will only be apparent from comparisons with nonlinear models. Nevertheless, many geophysical problems are currently intractable without linear approximations and hence are often used.

Linearized approximations

Linear approximations are covered in many excellent texts and summary papers (e.g., Matsu'ura and Hirata, 1982; Menke, 1984), so we only describe key concepts that are relevant for experimental design. A linear approximation is derived for model-parameter values in the region around some reference model $\mathbf{m}_0$. This reference model is usually the mean or the maximum-likelihood model from the prior model PDF. If the derivatives of $\mathbf{f}$ evaluated at any model $\hat{\mathbf{m}}$ are denoted by the matrix $\mathbf{F}$, where the ijth element is

$$F_{ij}\hat{\mathbf{m}} = \left[\frac{\partial f_i}{\partial m_j}(\hat{\mathbf{m}})\right], \tag{6}$$

initially where $\hat{\mathbf{m}} = \mathbf{m}_0$, then the linearized approximation to $\mathbf{f}$ is

$$\mathbf{d} - \mathbf{d}_0 = \mathbf{F}(\mathbf{m} - \mathbf{m}_0) + O(\mathbf{m} - \mathbf{m}_0)^2, \tag{7}$$

where $\mathbf{d}_0 = \mathbf{f}(\mathbf{m}_0)$. The higher-order terms $O(\mathbf{m} - \mathbf{m}_0)^2$ are ignored in the linear approximation and thus represent the error associated with the linear assumption. Once ignored in the mathematical formulation, any errors associated with these terms will not be apparent without additional calculations using nonlinear models. Nevertheless, this approximation is usually valid in some vicinity of $\mathbf{m}_0$; from this point on in this subsection, we ignore these higher-order terms.

Because the reference model $\mathbf{m}_0$ is known, the linearized inverse problem reduces to estimating the vector of model parameter perturbations $\mathbf{m}^{\text{est}} = \mathbf{m} - \mathbf{m}_0$. A corresponding inverse operator $\mathbf{F}^{-1}$ that estimates the model parameters by minimizing the discrepancy between $\mathbf{d}$ and $\mathbf{d}_0$ can be written as

$$\mathbf{m}^{\text{est}} = \mathbf{F}^{-1}(\mathbf{d} - \mathbf{d}_0). \tag{8}$$

In geophysical problems, $\mathbf{F}$ is generally singular or nearly singular. This is the result of inherent ambiguities and/or insufficient and imprecise data. Hence, $\mathbf{F}^{-1}$ must be approximated.

There are several options for approximating $\mathbf{F}^{-1}$, the most popular being to include regularization constraints that reduce the effective number of degrees of freedom by enforcing desired behavior on the model parameters (e.g., see Matsu'ura and Hirata, 1982). In this paper, we consider the commonly used iterative Gauss-Newton scheme:

$$\mathbf{m}^{\text{est}}_{i+1} = (\mathbf{F}^T\mathbf{C}_\mathbf{D}^{-1}\mathbf{F} + \mathbf{C}_\mathbf{M}^{-1})^{-1}\mathbf{F}^T[(\mathbf{d} - \mathbf{d}_0) + \mathbf{F}\mathbf{m}^{\text{est}}_i], \tag{9}$$

where i is the iteration number ($\mathbf{m}^{\text{est}}_0$ is set to the initial reference model $\mathbf{m}_0$); $\mathbf{C}_\mathbf{D}$ is the data covariance matrix that describes all measurement uncertainties of $\mathbf{d}$; $\mathbf{C}_\mathbf{M}$ is the a priori model covariance matrix, which allows regularization constraints, such as requiring specific spatial variations (e.g., smoothness, roughness, minimum variation from the reference model) in model parameter values to be enforced. Importantly, regularization should be selected such that it represents our prior expectations of the posteriori model-parameter variations. A least-squares (l_2-norm) misfit objective function is implicit in equation 9. To maintain assumptions behind the linearized approximation, the optimization procedure to estimate $\mathbf{m}$ is applied iteratively so that the sensitivities contained in $\mathbf{F}$ in equation 6 are re-evaluated at the best estimate of the model-parameter vector $\hat{\mathbf{m}} = \mathbf{m}^{\text{est}}_i + \mathbf{m}_{i-1}$ at iteration $i + 1$.

The quality of the result of an inversion of a truly linear forward model can be appraised by examining the model resolution matrix $\mathbf{R}$ and the posteriori model parameter covariance matrix $\mathbf{C}$, defined as

$$\mathbf{R} = (\mathbf{F}^T\mathbf{C}_\mathbf{D}^{-1}\mathbf{F} + \mathbf{C}_\mathbf{M}^{-1})^{-1}\mathbf{F}^T\mathbf{C}_\mathbf{D}^{-1}\mathbf{F}, \tag{10}$$

$$\mathbf{C} = (\mathbf{F}^T\mathbf{C}_\mathbf{D}^{-1}\mathbf{F} + \mathbf{C}_\mathbf{M}^{-1})^{-1} \tag{11}$$

(e.g., Menke, 1984), where $\mathbf{F}$ in equations 10 and 11 is evaluated using equation 6 with $\hat{\mathbf{m}}$ set to the final estimate of parameters $\mathbf{m}$ from the iterative optimization procedure. The resolution matrix $\mathbf{R}$ relates the final estimated model parameters $\mathbf{m}_{\text{est}}$ to the true model-parameter perturbations $\mathbf{m}_t$ because, by substituting $\mathbf{d} - \mathbf{d}_0 \approx \mathbf{F}(\mathbf{m}_t - \mathbf{m}^{\text{est}}_i)$ in equation 9,

$$\mathbf{m}_{\text{est}} \approx \mathbf{R}\mathbf{m}_t. \tag{12}$$

Of particular interest are the diagonal elements of $\mathbf{R}$: values close to zero indicate poorly resolved model parameters, and values close to one indicate well-resolved model parameters.

The posteriori covariance matrix in equation 11 translates data uncertainties $\mathbf{C}_\mathbf{D}$ into the space of model parameters and combines them with the prior parameter uncertainties to estimate posteriori covariances. Off-diagonal elements indicate the degree to which parameter estimates remain correlated postinversion (i.e., large values indicate pairs of parameters that are not independently resolvable). Diagonal elements are the variances of individual parameter estimates, and small values indicate well-resolved parameters.

Finally, we again note that equations 10 and 11 are only valid for truly linear forward models. When the models contain nonlinearities, interpretation of the resolution offered by a data set requires a more sophisticated approach (e.g., Stark, 2008), and experience shows that estimates of linearized posteriori uncertainty in $\mathbf{C}_\mathbf{D}$ are usually severely underestimated. In such cases, the quality of posteriori parameter estimates should be quantified by calculating or sampling the posteriori parameter PDF using equations 3–5 if computationally possible.

Design theory

Geophysical survey design theory consists of methods to select the data-acquisition parameters such that model parameters of interest are resolved optimally. The same theory has been used to optimize the model parameterization to best represent information contained in the data (e.g., Curtis and Snieder, 1997). However, here we focus on the former mode of application.

The inverse problem solutions in equations 3 and 9 are constrained by $\mathbf{d}$, by the PDF of prior information on $\mathbf{m}$, and by forward-problem physics Θ relating $\mathbf{d}$ and $\mathbf{m}$. Survey-design methods influence the form of this inverse problem and hence its solution by changing which data should be recorded. The goal of the design procedure is to collect those data such that the pertinent information de-

scribed by solution $Q(\mathbf{m})$ is maximized. The design problem is therefore a macro-optimization problem where, prior to the survey taking place, we optimize (design) the inverse problem that we expect to solve after the survey has occurred.

The selected design is the one that maximizes some objective function. Ignoring cost and logistics for the moment, this objective function is usually taken to be some measure of expected information:

$$J(\xi) = E_{\mathbf{m}_t}\{I[Q(\mathbf{m});\xi,\mathbf{m}_t]\}. \quad (13)$$

Here, ξ is a vector describing the design (e.g., source and receiver locations, shot fold, particular equipment to be used), $I[Q(\mathbf{m});\xi,\mathbf{m}_t]$ is a measure of the information contained in the resulting posteriori PDF $Q(\mathbf{m})$ for ξ when the true model parameters are given by $\mathbf{m}_t$, and the statistical expectation operator $E_{\mathbf{m}_t}$ averages I over the distribution of all possible values for the true model $\mathbf{m}_t$, which (by our prior knowledge) is expected to be distributed according to the prior distribution $p(\mathbf{m})$. The value $J(\xi)$ should be maximized. If, instead, a minimization problem is desired (e.g., to combine with cost, which also should be minimized), then the negative of the measure in equation 13 can be used.

Within the expectation in equation 13, the design criterion takes account of all possible potential values $\mathbf{m}_t$, their prior PDF $p(\mathbf{m})$, and the corresponding data (including their uncertainties) expected to be recorded for each model [uncertainties are accounted for within $Q(\mathbf{m})$]. To calculate the expectation usually requires integration over a far greater proportion of the model and data spaces $\mathbf{M}$ and $\mathbf{D}$ compared to the solution of the inverse problem after a particular data set has been recorded [where $p(\mathbf{d})$ is fixed and limited by actual measurements and hence $Q(\mathbf{m})$ is generally more tightly constrained]. Consequently, experimental design is usually far more computationally costly than solving any particular inverse problem postexperiment.

For this reason, design methods that capitalize on linearized approximations to the model-data relationship $\Theta(\mathbf{m},\mathbf{d})$ similar to those described above that are used for inversion have by necessity often been used for designing surveys (e.g., Rabinowitz and Steinberg, 1990; Steinberg et al., 1995; see Curtis, 2004a). Nonprobabilistic methods (which do not explicitly consider the prior probability distribution) have also been used (e.g., Maurer and Boerner 1998a; Curtis, 1999a, 1999b; Stummer et al., 2004; Coles and Morgan, 2009). Truly nonlinearized design methods that optimize the objective function in equation 13 have been developed in geophysical problems only relatively recently (van den Berg et al., 2003, 2005; Curtis, 2004b; Winterfors and Curtis, 2008; Guest and Curtis, 2009, 2010).

The key to understanding any particular design method is to understand (1) whether the physics describing the forward model are approximated (e.g., linearized) or whether the full physics are considered, (2) which information measure I is used in equation 13, and (3) how the macro-optimization is achieved. In recent years, developments have occurred in all three areas within geophysical survey-design applications.

Information measures and (non)linear physics

Shannon information

A critical concept in information theory is that the information content of an uncertain (noisy) process is determined by the process entropy (Shannon, 1948). If $\mathbf{X}$ is a random variable that takes a value $\mathbf{x}$ which varies probabilistically with PDF $p(\mathbf{x})$, then the (Shannon) information I_{Shan} is defined by

$$I_{\text{Shan}}[p(\mathbf{x})] + c = -\text{ent}(\mathbf{X}) = \int_{\mathbf{x}} p(\mathbf{x})\log[p(\mathbf{x})]d\mathbf{x}, \quad (14)$$

where c is a constant. The entropy $\text{ent}(\mathbf{X})$ is a measure (in the sense of an expected value) of the information in the PDF $p(\mathbf{x})$. In his well-known paper of 1948, Shannon shows that entropy is the only measure with a certain set of desirable properties (e.g., linear additivity of the information associated with independent pieces of information).

By setting $I = I_{\text{Shan}}$ in equation 13, the design process will maximize an objective function that measures the expected amount of Shannon information in the posteriori probability distribution $Q(\mathbf{m})$, i.e., in the PDF of $\mathbf{m}$ after the survey has taken place. The constant c in equation 14 is irrelevant and can be set to zero for design purposes because maximizing I_{Shan} for $c = 0$ will also maximize it for any other value of c.

The main limitation with this design approach is computational. Unless the PDF is nonzero over a finite range, numerically evaluating the entropy to approximate the integral in equation 14 is computationally intensive. In addition, for values of $\mathbf{x}$ with small probabilities $p(\mathbf{x})$, $\log[p(\mathbf{x})]$ is very large and negative, and it often varies rapidly with $p(\mathbf{x})$, meaning many samples are needed to obtain a robust approximation of the integral to estimate the entropy or Shannon information.

Although setting $I = I_{\text{Shan}}$ in equation 13 requires substantial effort to sample each $Q(\mathbf{m})$ adequately, in nonlinear problems the situation is even worse. By examining equation 3, it is apparent that sampling $Q(\mathbf{m})$ is equivalent to solving an inverse problem. When the forward physical relationship deviates significantly from being linear, inversion is performed using full Monte Carlo inverse methods or it proceeds iteratively by solving a sequence of linearized models using equations such as equation 9; either process can be extremely demanding computationally. Furthermore, the expectation operator in equation 13 requires us to evaluate the Shannon information for the full range of posteriori PDF distributions of $Q(\mathbf{m})$ that we are likely to encounter postsurvey (i.e., we would need to invert every possible data set that might be recorded in the experiment, or at least a representative selection of them). Without additional insight, this approach could only be implemented for simple experimental design problems (e.g., one or two model parameters and data).

The computational limitations inherent in this formulation can be reduced significantly by Shewry and Wynn's (1987) breakthrough: under certain conditions, the Shannon information can be obtained by maximizing the entropy in the data space $\text{ent}[p(\mathbf{d})]$ rather than in the model space $\text{ent}[Q(\mathbf{m})]$. Shewry and Wynn's conditions require that $p(\mathbf{m})$ and the data uncertainties, on each individual datum in $\mathbf{d}$ be independent of the survey design (see Curtis [2004b] for a tutorial illustrating this conceptual step). Computing $\text{ent}[p(\mathbf{d})]$ requires an estimate of the PDF $p(\mathbf{d})$, which can be calculated by projecting $p(\mathbf{m})$ into data space through the forward model by taking samples $\mathbf{m}_i$ of $\mathbf{m}$ according to $p(\mathbf{m})$ and calculating $\mathbf{d}_i = \mathbf{f}(\mathbf{m}_i)$. The resulting set $\{\mathbf{d}_i\}$ will be distributed according to $p(\mathbf{d})$. Evaluating $\text{ent}[p(\mathbf{d})]$ only requires evaluating $\mathbf{f}$ rather than solving the inverse problem.

Even with the simplifications afforded by the assumptions of Shewry and Wynn (1987), the computational challenge of experimental design that maximizes Shannon information remains significant for many geophysical problems. However, if implemented with appropriate Monte Carlo and optimization algorithms, these methods can be computationally tractable for several important geophysical methods, including designing surface seismic surveys and optimizing data-processing strategies (e.g., Guest and Curtis, 2009, 2010).

For linear problems, maximizing Shannon information is not dependent on the true model (see Curtis, 2004a); hence, it is unnecessary to calculate the expectation in equation 13. Maximizing Shannon information is then also equivalent to maximizing the so-called *D*-criterion (different design criteria are assigned different alphabetical names; see Atkinson and Donev [1992] for an overview). In statistical literature the *D*-criterion is usually defined to be the determinant $\det[\mathbf{F}^T\mathbf{F}]$; but as in equation 11, it can be extended to take the form $\det[\mathbf{F}^T\mathbf{C}_\mathbf{D}^{-1}\mathbf{F}]$ to incorporate data uncertainties, or to $\det[\mathbf{F}^T\mathbf{C}_\mathbf{D}^{-1}\mathbf{F} + \mathbf{C}_\mathbf{M}^{-1}]$ to include the effects of model regularization. Alternatively, using a nonprobabilistic approach, Matsu'ura and Hirata (1982) provide a formalism that allows $\mathbf{F}$ to be consistently weighted such that $\det[\mathbf{F}^T\mathbf{F}]$ accounts for variations in confidence in different data or for focusing the information measure on different parameters of particular interest.

Measures of variance

Variance is defined as the expected squared deviation of a random variable, say, X, about its mean:

$$V(X) = \int_X E[(X - E(X))^2]dX, \tag{15}$$

which measures the expected degree of variation of X. We often desire solutions to geophysical problems that have minimum possible variance on the model parameters because such solutions will be expected to be constrained around some limited range of the model parameter space M.

A principal limitation in using Shannon information to design nonlinear problems is that it does not necessarily discriminate between designs that result in high or low variance in the expected inversion solution. Optimal designs might therefore generate maximum Shannon information in the inverse problem solution yet retain high variance in that solution.

To understand this counterintuitive statement, consider a PDF $p_1(x)$ that allows x to have only one of two distinct values, one or three, each with equal probability, and a second PDF $p_2(x)$ that allows x to have only the values zero and four with equal probability. Both PDFs contain identical Shannon information because they both allow only two possible values for x; hence, in this sense, x is equally tightly constrained by either PDF: $I_{\text{Shan}}[p_1(x)] = I_{\text{Shan}}[p_2(x)]$. However, solution $p_2(x)$ has a higher variance than $p_1(x)$ around the common mean value of two. We would usually prefer a design that would give postsurvey solution $p_1(x)$ over one that would give $p_2(x)$ because the latter more tightly constrains the range of values of x, even though the number of possible values that x can assume is the same in each case.

For this reason, in nonlinear problems, it is often desirable to create designs that will minimize some measure of spread, such as expected variance ($I = V$ in equation 13), rather than designs that only maximize Shannon information ($I = I_{\text{Shan}}$). There is also a distinct computational advantage in avoiding the calculation of an integrand containing a logarithm as in equation 14, as explained above. Hence, the design problem is generally numerically more tractable for variance than for Shannon information. Nevertheless, we confront the usual numerical challenge known as the *curse of dimensionality*, which states that computational effort required for integration increases geometrically as the dimensionality of the integrand increases (Curtis and Lomax, 2001). Hence, it is necessary to look for "smart" methods to approximate the variance or to define less-burdensome measures of spread that are efficient for nonlinear problems. In the following, we explain one useful alternative derived by Winterfors and Curtis (personal communication, 2010) that is analytically related to variance and is so sufficiently computationally efficient that it has been used to design multisensor microseismic monitoring surveys.

Variance or spread is fundamentally related to the distance between different parameter values (points in model space $\mathbf{M}$) that are ambiguous with respect to (i.e., are not discriminated by) recorded data. Given two points $\dot{\mathbf{m}} \neq \ddot{\mathbf{m}}$ in $\mathbf{M}$ (the dots are indices), it is possible to create various measures of how likely these are to give rise to the same observation — and hence be indistinguishable given such an observation. This is determined by the extent to which their respective data-space probability densities $p(\mathbf{d}|\dot{\mathbf{m}},\xi)$ and $p(\mathbf{d}|\ddot{\mathbf{m}},\xi)$ overlap, where ξ is a vector that defines the survey design. These distributions each describe the probability of recording any data vector $\mathbf{d}$ if the true parameter values were represented by model $\dot{\mathbf{m}}$ or $\ddot{\mathbf{m}}$, respectively. The most straightforward option for such a measure is

$$S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi) = \int_{\mathbf{d}\in\Omega} p(\mathbf{d}|\dot{\mathbf{m}},\xi)p(\mathbf{d}|\ddot{\mathbf{m}},\xi)dD. \tag{16}$$

This defines a so-called bifocal measure (Winterfors and Curtis, 2008), simultaneously focusing on two points $(\dot{\mathbf{m}},\ddot{\mathbf{m}})$ in parameter space instead of only one, which is the most common approach (e.g., measure I_{Shan} in equation 14).

The measure in equation 16 does, however, have two disadvantages: (1) $S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi)$ is always high for $\dot{\mathbf{m}} = \ddot{\mathbf{m}}$, even though this case does not contribute to uncertainty in estimates of the model parameters $\mathbf{m}$, and (2) the unit of $S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi)$ is the same as of a probability density $p(\mathbf{d})$ over data space, implying that $S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi)$ will increase with decreased observational uncertainty (the opposite ideally should be the case for a useful measure of parameter uncertainty).

Winterfors and Curtis (personal communication, 2010) show that one way to overcome the first problem is to multiply $S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi)$ by the squared distance between $\dot{\mathbf{m}}$ and $\ddot{\mathbf{m}}$ (assuming that $\mathbf{D}$ is equipped with a distance metric d). The second problem can be addressed by dividing $S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi)$ by a measure T_D of average observational probability density:

$$T_D = \int_{\mathbf{d}\in D} p(\mathbf{d}|\xi)p(\mathbf{d}|\xi)dD. \tag{17}$$

This results in the ambiguity measure:

$$R(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi) = \frac{d^2(\dot{\mathbf{m}},\ddot{\mathbf{m}})}{T_D} S(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi). \tag{18}$$

To create a global measure of the ambiguity of an investigation technique design, it is necessary to take the expectation of $R(\dot{\mathbf{m}},\ddot{\mathbf{m}},\xi)$

over all possible point pairs in model parameter space, with respect to the prior parameter PDF $p(\mathbf{m})$:

$$W(\xi) = \int_{\dot{\mathbf{m}} \in M} \int_{\ddot{\mathbf{m}} \in M} p(\dot{\mathbf{m}}) p(\ddot{\mathbf{m}}) R(\dot{\mathbf{m}}, \ddot{\mathbf{m}}, \xi) dM dM. \quad (19)$$

The expected observational ambiguity $W(\xi)$ is thus a measure of the average ambiguity of all possible observations, given a survey design and a prior distribution over parameter values. Therefore, $W(\xi)$ can be used to evaluate or design surveys described by ξ prior to the acquisition of any observations.

Furthermore, $W(\xi)$ relates to the expected posteriori variance in a simple manner: inserting equation 17 into 18, applying Bayes' rule, and changing the order of integration gives

$$W(\xi) = \frac{2}{T_D} \int_{\mathbf{d} \in D} p^2(\mathbf{d}|\xi) V[\mathbf{m}|\mathbf{d}, \xi] dD, \quad (20)$$

where, if the distance metric d is the standard Euclidian distance, then $V[\mathbf{m}|\mathbf{d}, \xi]$ is equal to the variance of the posteriori model space PDF, given the data $\mathbf{d}$ from a survey with design ξ (Winterfors and Curtis, personal communication, 2010). Thus, $W(\xi)$, as defined in equation 19, is the expected variance of the posteriori PDF, weighted by $p(\mathbf{d}|\xi)$.

The advantage offered by optimizing using this measure of ambiguity rather than the variance itself is that to calculate $W(\xi)$ using equations 16–19 requires only forward-model calculations (calculations of the PDF of $\mathbf{d}$ given $\mathbf{m}$ on the right-hand side of equations 16 and 17). Calculation of the variance V in equation 20 requires calculating the expected variance of the posteriori PDF of $\mathbf{m}$ given $\mathbf{d}$, which requires the inverse problem solution. Similar to Shewry and Wynn (1987), this approximation obviates the need to solve the inverse problem, helping to reduce the effect of the curse of dimensionality. Winterfors and Curtis (personal communication, 2010) show that using $W(\xi)$ is particularly computationally advantageous when data uncertainties can be represented by closed-form PDFs such as Gaussian, Poisson, or Laplacian distribution functions.

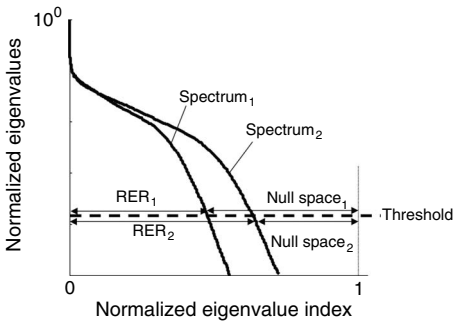

Figure 2. Representation of two normalized eigenvalue spectra of the Hessian matrix $\mathbf{F}^T\mathbf{C}_D^{-1}\mathbf{F}$ (equation 9). The horizontal axis is normalized with respect to the total number of model parameters, and the vertical axis is normalized with respect to the largest eigenvalues. Dashed line indicates the threshold level that defines the relative eigenvalue range (RER). $RER_{1,2}$ indicate measures of quality of the two surveys.

Minimizing $W(\xi)$ does not seem to fit into the general framework of equation 13 because the integration is over the data space D instead of the model space M. To optimize with respect to the ambiguity measure, we have to replace $J(\xi)$ by $W(\xi)$ in equation 13. However, as shown, $W(\xi)$ can be thought of as a reweighted alternative to V over the model space, and V does fit into the framework of equation 13 by setting $I = V$.

In linear models with normally distributed uncertainties, the posteriori variance $V[\mathbf{m}|\mathbf{d}, \xi]$ will be constant with respect to the observation $\mathbf{d}$. As a consequence, optimizing $W(\xi)$ will then be exactly equivalent to optimizing the expected posteriori variance. In linear statistical experimental design literature, optimizing designs such that their expected posteriori variance is minimized is known as the A-optimality criterion. This corresponds to minimizing the sum of the variances (diagonal elements of C in equation 11).

Other information measures

In subsequent examples, we use two other measures related to the D-criterion. Besides the determinant of the matrix $[\mathbf{F}^T\mathbf{C}_D^{-1}\mathbf{F}]$, one may also consider its eigenvalue spectrum. For the perspective of optimizing the resolution of the complete suite of model parameters, an optimized design should result in a small condition number (ratio of the largest to the smallest eigenvalue) of $[\mathbf{F}^T\mathbf{C}_D^{-1}\mathbf{F}]$. This can be achieved by minimizing

$$\sum_i \frac{1}{\lambda_i + \delta}, \quad (21)$$

where λ_i are the eigenvalues and δ is a small positive constant (see Curtis [1999b] for an extensive discussion of this and related measures).

Another possible choice is to define a measure that is related to the unresolved part of the model space (null space). The choice of a survey layout governs the structure of the matrix $\mathbf{F}$, and close inspection of equation 9 indicates that the reliability of the parameter estimates $\mathbf{m}^{est}$ depends primarily on our ability to invert the matrix $(\mathbf{F}^T\mathbf{C}_D^{-1}\mathbf{F} + \mathbf{C}_M^{-1})$. Without the regularization constraints in $\mathbf{C}_M^{-1}$, this matrix would likely be singular, such that its determinant would be zero and its condition number would be infinite. The sensitivities in $\mathbf{F}$ represent the information content offered by a particular survey design and $\mathbf{C}_M^{-1}$ indicates our preconceived ideas on the subsurface structure (e.g., closeness to a prior model estimate or that it should have smooth spatial variations), so it is certainly advisable to maximize the contribution of $\mathbf{F}^T\mathbf{C}_D^{-1}\mathbf{F}$ and to minimize the influence of $\mathbf{C}_M^{-1}$.

Figure 2 shows two typical eigenvalue spectra of $\mathbf{F}^T\mathbf{C}_D^{-1}\mathbf{F}$ as they may arise from a geophysical inversion experiment. The vertical axis is logarithmically scaled and normalized with respect to the largest eigenvalue of the corresponding spectra, and the horizontal axis is normalized by the total number of eigenvalues (number of model parameters). Because of the finite precision of numerical computations, eigenvalues are rarely identical to zero, even when the matrix is singular. Therefore, a threshold must be introduced below wherein the eigenvalues are considered to be insignificant (dashed line in Figure 2). The intersections of the eigenvalue spectra with the threshold line indicate the portions of the resolved model space and the unresolved null space. We define the range of significant eigenvalues, the relative eigenvalue range (RER), to be the horizontal coordinate of the intersection point. This provides a simple

and intuitive means for quantifying the quality of a particular survey design. For example, the hypothetical survey design 2 in Figure 2 is superior to survey design 1.

Macro-optimization methods

Ideally, when designing an experiment, one searches exhaustively for the design that optimizes the design objective function in equation 13. A variety of minimization algorithms are available for this task. The most popular method historically has been the Detmax algorithm (Mitchell, 1974). However, other authors have used genetic algorithms (Hardt and Scherbaum, 1994; Curtis and Snieder, 1997; Maurer and Boerner, 1998a; Curtis, 1999a, 1999b), simplex algorithms (Winterfors and Curtis, 2008), simulated annealing (Barth and Wunsch, 1990), or multilevel coordinate search (Ajo-Franklin, 2009). Conceptually, any global optimizer can be used. The most suitable choice is problem dependent.

When there are multiple design parameters, searching the entire design space (the space of all permissible combinations of design parameters) may be computationally unfeasible. Sequential design strategies have been developed in many linear design studies (e.g., Curtis et al., 2004; Stummer et al., 2004; Coles and Morgan, 2009), and Guest and Curtis (2009, 2010) introduce a sequential design strategy applicable to nonlinear problems. Such algorithms generally increment the number of elements within the design vector ξ for each iteration:

$$\xi_j = \arg\max[J(\xi_j)], \text{ such that } \xi_{j-1} \text{ is fixed}, \quad (22)$$

where $\xi_j = \{\xi^1, \ldots, \xi^j\}$, with ξ^i an element of the design vector ξ, and where J is the objective function in equation 13. The new optimal design ξ_j combines the design defined at the previous iteration ξ_{j-1}, augmented by the single datum defined by ξ^j that maximizes (minimizes) the objective function given that ξ_{j-1} remains fixed. In this way, the work required to design an experiment with n design parameters is reduced from searching an n-dimensional design space to n separate searches of 1D design spaces. Although at each iteration the design is only locally optimal, Guest and Curtis (2009) show that for nonlinear AVO/AVA design problems, the locally optimal design is almost identical to the globally optimal design (for problems in which the globally optimal design could be calculated).

The choice of which specific algorithm to use should in principle be affected by the so-called no-free-lunch (NFL) theorems (Wolpert and Macready, 1997). These state that no single optimization algorithm is ideally suited for all objective functions (and conversely, no single objective function is ideally suited to be minimized by all optimization algorithms). D. Coles and A. Curtis (personal communication, 2010) examine the influence of NFL theorems on linearized statistical experimental design by comparing several sequential optimization algorithms on three quite different design objective functions. They show that within this limited context, a clear ranking is made between the optimization algorithms: almost regardless of the objective function, the best algorithm allows incremental augmentation and reduction of the design at each iteration. Hence, somewhat surprisingly, Coles and Curtis do not observe the effect of NFL theorems within this limited context.

EXAMPLES

Example 1: Designing geoelectrical experiments

Geoelectrical methods have been applied with great success for many years (e.g., Butler [2005] and references therein). Until the early 1990s, interpretation of geoelectrical data was performed predominantly in terms of layered-earth models, and the measurements were conducted almost exclusively with standardized electrode configurations such as Schlumberger, Wenner, or dipole-dipole (e.g., Telford et al., 1990).

Figure 3a shows a Schlumberger electrode configuration. The measuring dipole is centered in the configuration, and the two electrodes, where current is injected, are each deployed at a distance r from the center. Schlumberger experimental design thus consists of

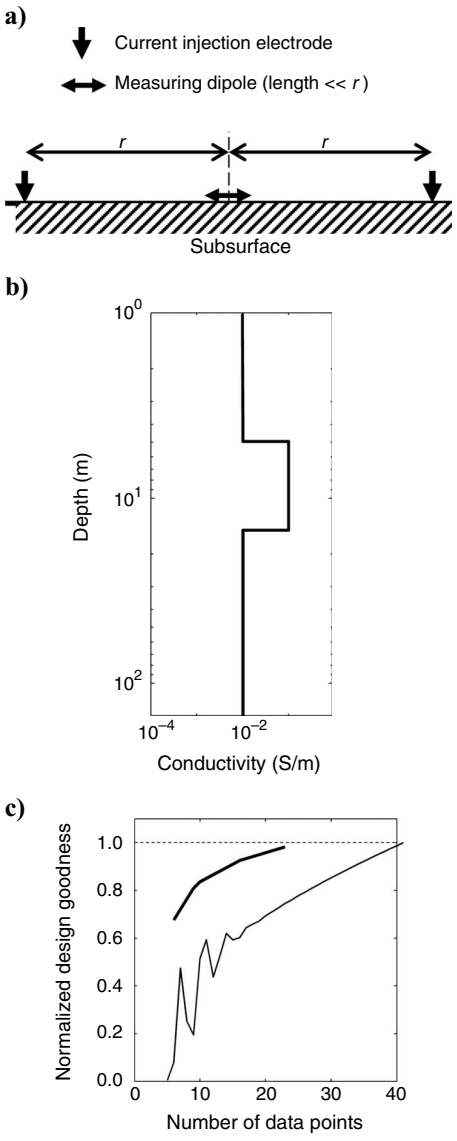

Figure 3. A 1D geoelectrical survey design. (a) Schematic representation of a Schlumberger sounding. (b) Subsurface model used for the experimental design. (c) Cost/benefit curves for a design where the current electrode separation is increased geometrically (thin solid curve) and optimized design (thick solid curve). The dotted line indicates the normalized goodness of 1.0.

varying a single parameter, the distance r between the current source and sink. Traditionally, r is varied over a suite of regularly log-spaced distances. The resulting data are then inverted for layer conductivities and layer thicknesses.

A typical layered-earth model with a conductive middle layer is shown in Figure 3b. Our goal is to constrain all five model parameters (three layer conductivities and two layer thicknesses) in an optimal fashion. The data space consists of 40 distances r, logarithmically spaced over a range of 1 – 10,000 m.

Intuitively, one might expect optimal layouts to include more-or-less equally log-spaced recording distances, as suggested by the sensitivity design studies of Oldenburg (1978). Figure 3c shows cost/benefit curves for equally log-spaced experiments (sequentially adding larger r values to the design) and optimally designed experiments using a genetic algorithm (see Maurer et al. [2000] for details). The optimized experiment is found by maximizing the design-goodness measure in equation 21. Benefit is expressed as a normalized goodness (goodness of a particular experiment divided by goodness of the complete data set of 40 distances), and costs are assumed to be proportional to the number of measurements.

For the equally log-spaced design, the curve in Figure 3c exhibits some erratic behavior. This indicates that rigidly incrementing r logarithmically can provide data that contain only limited information for resolving the five model parameters. The goodness can even decrease when more data points are added by following a rigid log-space sampling strategy. This is caused when particularly critical distances might be missed using a rigid sampling scheme. In the optimized design, the goodness increases monotonically and reaches the level of the full-scale experiment at about 20 data points. The result suggests that conventional Schlumberger DC resistivity soundings might be made substantially more cost effective by a judicious choice of r values needed to resolve the earth model. Maurer et al. (2000) also discuss strategies for optimized layouts suitable for a range of different subsurface models.

The simple layered-earth experiment shown in Figure 3 is interesting conceptually but of limited practical relevance. Today, it is more common that geoelectrical data are acquired using multi-electrode systems (e.g., Griffiths and Turnbull, 1985; Stummer et al., 2002). Such data sets are inverted tomographically (e.g., Loke and Barker, 1996) to create 2D and 3D subsurface images, which provide substantially more realistic information about the earth compared with layered-earth models. The data space of such experiments can be very large. For an n-electrode array, there exist $n(n-1)(n-2)(n-3)/8$ nonreciprocal configurations. For experiments using 30, 50, and 100 electrodes, one conceptually could have 82,215; 690,900; and 11,763,675 data measurements. These types of experiments are extremely difficult to design because there are no heuristic means by which to determine the type and number of configurations that would provide favorable cost/benefit ratios.

Statistical experimental design, as applied to the simple Schlumberger sounding example, is computationally too expensive for 2D or 3D earth models because the goodness function would have to be evaluated many times during global optimization. Therefore, Stummer et al. (2004) propose a sequential design strategy, whereby an initial measurement configuration is chosen and then sequentially augmented until the desired benefit level is reached. The initial data set can be one of the standard electrode configurations or even a single set of injecting and measuring bipoles. The choice of the next electrode configuration in the sequence is governed by examining which has the largest potential of adding information content. Stummer et al. (2004) define an incremental information measure based on the linear independence of a candidate configuration with respect to the configurations already selected. Furthermore, they consider the relative in-

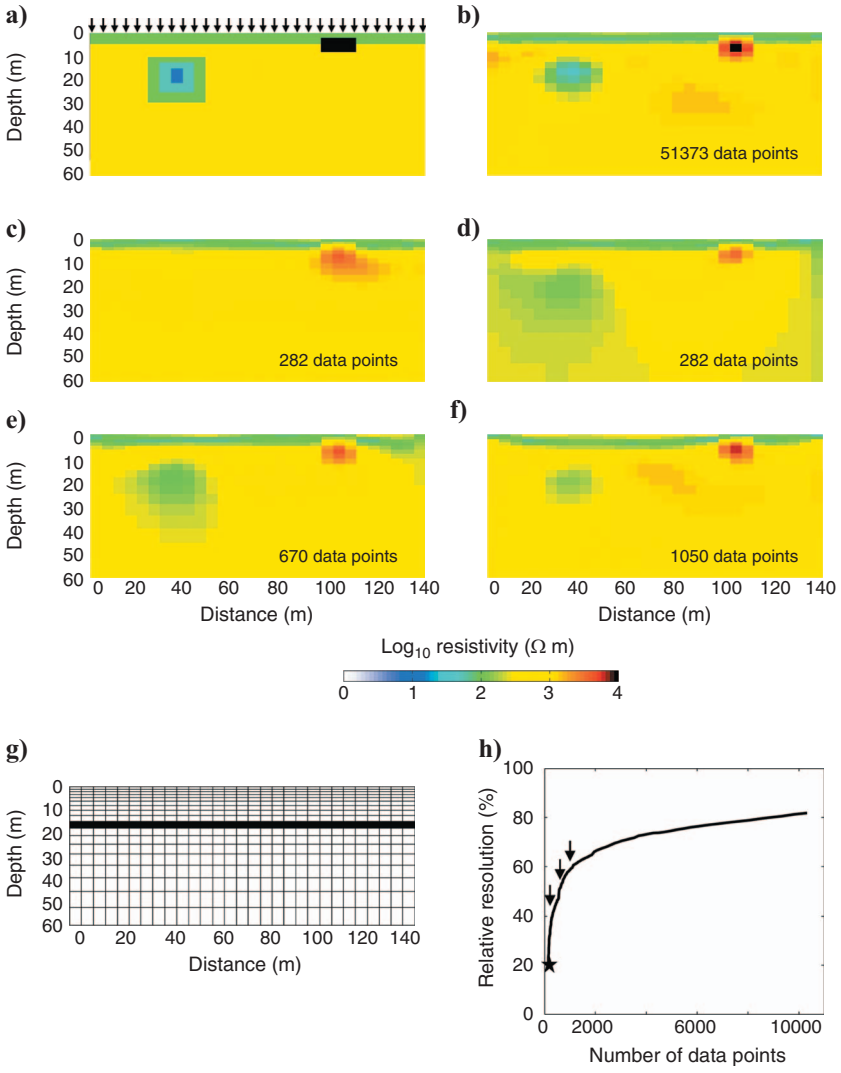

Figure 4. A 2D geoelectrical survey design. (a) True subsurface model and position of electrodes (vertical arrows). (b) Inversion result using the comprehensive data set. (c) Inversion results obtained with a combination of Wenner and dipole-dipole data. (d–f) Inversion results with the optimized data sets. (g) Depth range (filled black area) for which the cost/benefit curve in (h) is constructed. The star in (h) denotes the initial dipole-dipole data set, and the vertical arrows indicate the results for the optimized data sets shown in (d–f).

crease of the formal resolution (defined via the model resolution matrix; see equation 10) within the individual model cells. There is some latitude on how to define the value of incremental information. For example, Wilkinson et al. (2006), Coscia et al. (2008), and Loke and Wilkinson (2009) use measures based entirely on the increase of the formal resolution.

The 2D example shown in Figure 4 is adapted from Stummer et al. (2004). Figure 4a shows the true subsurface model and a deployment of a 30-electrode array. From all possible configurations of injection and measuring dipoles, we exclude those with crossed injecting and current dipoles and configurations with unfavorable geometric factors, leaving 51,337 out of a possible 82,215 measurement configurations. Stummer et al. (2004) denote this reduced suite of configurations as a comprehensive data set. Inversion of a comprehensive data set leads to the image shown in Figure 4b. It reflects the total information content offered by the complete data space. For comparison, Figure 4c shows the result using a combination of all possible Wenner and dipole-dipole configurations (subject to the same geometric factor restrictions as for the comprehensive data set: 282 configurations). At shallow depths, the tomograms in Figure 4b and c are comparable, but other electrode configurations are apparently needed to resolve the deep conductive feature between $x = 20$ and 40 m and $d = 10$ and 30 m that is recovered in the inversion of the complete data set.

Although Figure 4b and c demonstrates that traditional configurations do not recover the full information content offered by the geoelectrical method, collecting a comprehensive data set would not be cost effective. Determining the appropriate measurement configurations requires optimized experimental design. We follow the sequential design strategy of Stummer et al. (2004) and start with a dipole-dipole data set (147 data points); then we add successively more configurations. At 282 data points (the same number as the combined Wenner/dipole-dipole data set), the information content has already improved (compare the inversion results shown in Figure 4c and d). Successively adding further data leads to dramatic improvements of the recovered images (Figure 4e and f) using only a minor fraction (2%) of the complete data set.

The image-quality increase can be quantified by analyzing the increase of the formal resolution. For example, imagine we are particularly interested in resolving the depth range of the conductive anomaly. To quantify this, we sum the normalized resolution (corresponding diagonal elements of the model resolution matrix expressed as a percentage of the comprehensive data-set experiment) of the cells highlighted in Figure 4g and plot the average relative resolution as a function of the number of data points. The initial dipole-dipole data set provides only approximately 20% of the total information content in this depth range. When further configurations are added, the relative resolution increases quickly. At about 1000 data points, the cost/benefit curve enters into the realm of diminishing returns. In fact, the image in Figure 4f is already quite comparable to those of the comprehensive data set, and it is very similar to images constructed with 2000, 5000, and 10,000 data points (not shown).

The sequential experimental design procedure (Figure 4) is calculated using sensitivities for a homogeneous half-space to represent the state of knowledge prior to an experiment. One might expect that using this approximation when the current-flow patterns (and thus also the sensitivities) are disturbed by the presence of the conductivity anomalies of the true model (Figure 4a) would lead to suboptimal design results. Surprisingly, this is not the case, as demonstrated in Stummer et al. (2004). This also indicates that linear experimental design is adequate for such types of problems; the much more expensive nonlinear methods are not expected to provide superior results.

Example 2: Designing seismic crosshole experiments

Traveltime tomography

Figure 5a illustrates a simple seismic experiment to resolve the geologic structure between two parallel boreholes. Shots (red dots) can be placed in the borehole on the left or at the surface, whereas receivers (blue circles) are installed in the borehole to the right or at the surface. All useful shot-receiver combinations can be subdivided into three different categories: (i) shots in the left borehole and receivers in the right borehole, (ii) shots in the left borehole and receivers at the surface, or (iii) shots at the surface and receivers in the right borehole. With these subsets, we form three data spaces: $\mathbf{D}_1 = (\mathrm{i})$, $\mathbf{D}_2 = (\mathrm{i}) + (\mathrm{ii})$, and $\mathbf{D}_3 = (\mathrm{i}) + (\mathrm{ii}) + (\mathrm{iii})$. The geologic properties of the medium between the boreholes are assumed to be homogeneous, meaning that the rays between shots and receivers can be modeled with straight lines for computing the traveltime sensitivities. Each of the three data sets includes 20 shot and receiver positions, which results in 400 traveltimes (1200 traveltimes for all three data sets). The model cells have a width of 2.5 m, and the total number of cells is thus $40 \times 40 = 1600$.

Figure 5b shows the eigenvalue spectra of $\mathbf{F}^T\mathbf{F}$ (the data covariance matrix $\mathbf{C}_D^{-1}$ is assumed to be a unity matrix in this example) for

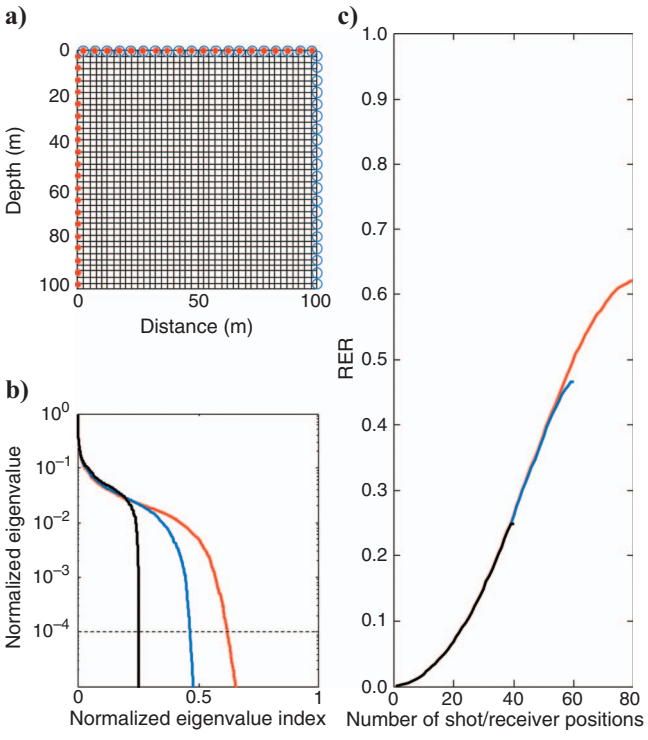

Figure 5. Survey design for seismic traveltime crosshole tomography. (a) Experimental layout. Red dots indicate shot, and blue circles indicate receiver positions. The model grid is depicted with black lines. (b) Normalized eigenvalue spectra for **A** sources in left borehole and receivers in right borehole (black), **B** sources in left borehole and receivers in right borehole and at the surface (blue), and **C** sources in left borehole and at the surface and receivers in right borehole and at the surface (red). (c) Cost/benefit curves for optimized experiments using data space **A** (black), **B** (blue), and **C** (red).

$\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ when all shot and receiver positions are considered. The tomographic imaging plane is subdivided in a relatively fine mesh of model cells (Figure 5a), so the problem is strongly underdetermined (more model parameters than data) and exhibits a significant null space. Using a threshold value of 10^{-4} for the normalized eigenvalues (dashed line, Figure 5b) leads to RER values (see also Figure 2) of about 0.25, 0.47, and 0.62 for $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$, respectively. This indicates that placing shots and/or receivers at the surface increases the information content of the data significantly (see also Curtis, 1999a, 1999b).

Our goal is to use statistical experimental design to determine if it is possible to obtain an RER value for a subset of shot and receiver positions that is similar to that for the complete data set $\mathbf{C}$. We initially select a single shot and receiver position and then add shot or receiver positions sequentially. All possible source-receiver pairs for the current selection of sources and receivers are considered. The choice of each subsequent shot or receiver position is made such that the resulting RER value is increased maximally by testing every possible source and receiver as a possible candidate. Although this approach requires many eigenvalue decompositions to be performed, the computations do not impose a major problem ($\sim$2 hours are required to run the design study in Figure 5 on a standard workstation). For constructing cost/benefit curves, we make the simplifying assumption that adding a shot or receiver position is equally expensive. Costs are thus proportional to the total number of shot and receiver positions, and benefit is represented by the RER value.

Figure 5c shows the experimental design results. The black curve indicates the RER, when shot and receiver positions contained in data space $\mathbf{D}_1$ are considered. Initially, the RER increases only slowly and then exhibits a more-or-less linear increase until it reaches its maximum level of about 0.25. The results for $\mathbf{D}_2$ (blue curve) and $\mathbf{D}_3$ (red curve) are similar. Unlike the slightly overdetermined geoelectrical examples (Figure 4c and h), the area of diminishing returns in this underdetermined crosshole imaging experiment is quite small. Thus, depending on the survey's objectives, it may be worthwhile to include as many shot and receiver positions as possible in the experimental design. The sequence of shot and receiver position choice depends on the selection of the initial shot and receiver positions, but the development of the RER values (Figure 5c) is virtually identical for any choice of the initial configuration.

For strongly underdetermined problems, the cost/benefit relationship can be expected to be fairly linear, with little evidence of diminishing returns. However, adding more shot and receiver positions to the data space would leave only a small underdetermined component, and the curves in Figure 5c would flatten out at some point. There is a hint of this effect for all three curves in Figure 5c. The law of diminishing returns therefore also applies to this type of experimental design, but the base data sets $\mathbf{D}_1$, $\mathbf{D}_2$, and $\mathbf{D}_3$ have been chosen such the cost/benefit curves cover only the linearly increasing part.

The underdetermined nature of the problem is also the reason for the apparent contrast of our results with those of Curtis (1999b). The ratio of source-receiver spacing to model cell width in Figure 5 is 2.0 (5-m source-receiver spacing, 2.5-m cell width; see Figure 5a), whereas the comprehensive data set in Curtis (1999b) has a corresponding ratio of 0.5. Hence, adding data (i.e., raypaths) in the Curtis (1999b) example rapidly increases the information content by virtue of obtaining data (sampling) from more model cells.

Finally, Figure 5c demonstrates that improving the cost/benefit ratio does not substantially depend on having shots or receivers at the surface. If the surface locations would be truly significant, the blue and red curves (shots and receivers at the surface and in boreholes) in Figure 5c should produce RER values that are significantly above the value of the black curve (shots and receivers only in boreholes) at 40 shot-receiver positions (all shots and receivers in the boreholes), which is not the case. Clearly, this conclusion is only valid for the overall conditioning of the inversion problem. If the near-surface structures would be of primary interest, the contribution of the surface receivers would be significant.

Frequency-domain waveform inversion

In addition to traveltime information, the frequency-dependent behavior of seismic wave propagation can be used to infer the elastic properties of the subsurface area of interest (e.g., Pratt, 1999). However, a critical aspect of designing frequency-domain waveform experiments is the choice of frequencies. The importance of selecting the frequencies for measurement to improve inversion results is discussed, for example, by Sirgue and Pratt (2004). Here, we present an alternative approach that is further described in Maurer et al. (2009).

The experimental design, shown in Figure 6a, includes 31 shots positions in the left hole and 31 receiver positions in the right hole. The model area is 30×30 m and consists of a medium exhibiting a stochastic velocity distribution with a standard deviation of 100 m/s, a correlation length of 8 m, and a fractal dimension of 0.5. Additionally, positive and negative square-shaped velocity anomalies with dimensions 2×2 m^2 were added to the stochastic background model. Here, we make an acoustic approximation so that only P-wave velocities V_P are considered in the model space. As a source function, we consider a Ricker wavelet with a center frequency of 700 Hz. The data space includes all source and receiver positions that are fixed, leaving us to focus on the choice of frequencies. We seek to select an optimized subset from 30 equally spaced frequencies between 100 and 1500 Hz.

The inversion methodology is similar to that described for the traveltime tomography, except that we start with a single frequency and add other frequencies sequentially. Because seismic waveform inversions are known to be prone to nonlinear effects that can trap the iterative inversion in local minima, it is common practice to establish an initial model using only low frequencies (e.g., Brenders and Pratt, 2007). Therefore, we choose the lowest frequency (100 Hz) for initializing the statistical experimental design algorithm.

A simple (nonoptimized) experimental design strategy would be to always add the next-higher frequency until the entire data space is selected. The corresponding eigenvalue spectra of such an approach are shown in Figure 6b, where each line indicates the eigenvalue spectrum resulting from adding each sequential frequency. The development of the RER values (using a threshold value of 10^{-10}) is displayed in Figure 6d as a dashed line. The information content increases steadily; after about 18 frequencies, the RER curve reaches the area of diminishing returns and flattens out.

Selecting the frequencies on the basis of the RER criterion (choosing the next frequency such that the RER increases maximally) results in eigenvalue spectra as shown in Figure 6c and in RER development as shown with the solid line in Figure 6d. After only three frequencies, we almost reach the RER value for the full-scale experiment using all frequencies. After the initial frequency of 100 Hz, the algorithm selects the highest frequency (1500 Hz). The third fre-

quency is also chosen from the higher end of the spectrum, but it is not the second-highest frequency (see Maurer et al. [2009] for more details).

Results of the tomographic inversions using all 30 frequencies or only the three optimal frequencies selected by the experimental design are shown in Figure 7. On the basis of the results in Figure 6, one would expect the results to be similar, which is indeed the case. In the central part and in the low-velocity regions where the formal resolution is quite good, the individual features are well resolved. Only in the high-velocity region near the upper edge of the model do both tomograms suffer from resolution problems.

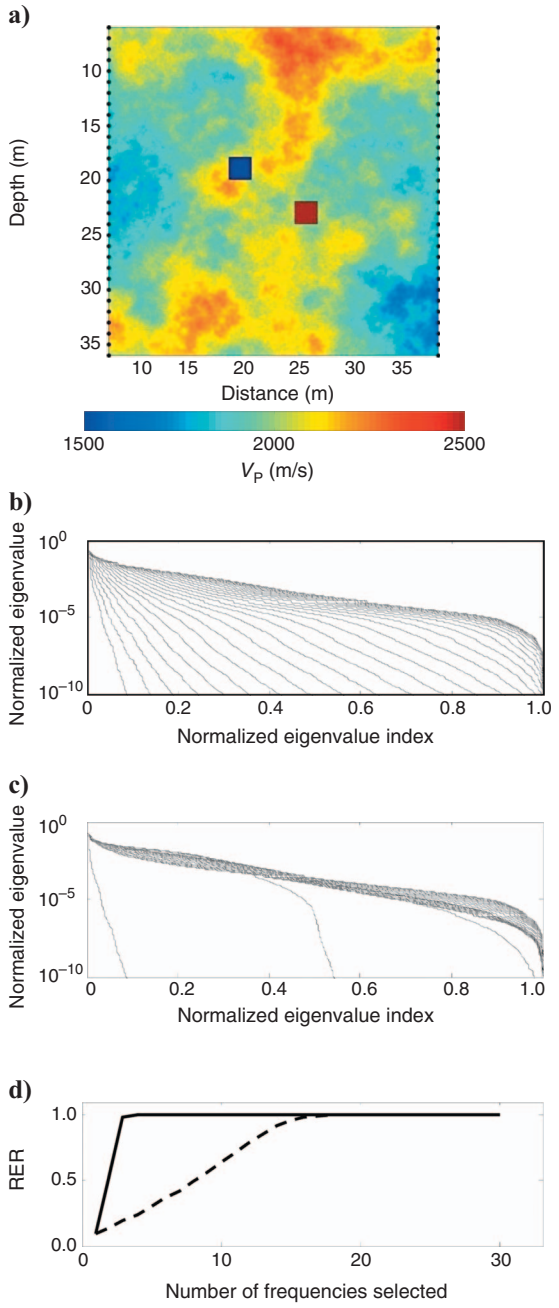

Figure 6. Frequency-domain waveform inversion design. (a) True velocity model and shot and receiver positions. (b) Normalized eigenvalue spectra for progressively adding higher frequencies. (c) Normalized eigenvalue spectra for adding optimized frequencies. (d) Cost/benefit curves for adding progressively higher frequencies (dashed line) and optimized frequencies (solid line).

Example 3: Nonlinear experimental design — Designing AVA surveys

The amplitude of a seismic wave reflected from a subsurface boundary between two geological layers at depth (Figure 8) is a function of the wave's incident angle at the boundary, the density ρ_i, and the intrinsic properties of the elastic media (i.e., for isotropic media, the P-wave velocity α_i, and S-wave velocity β_i) of both layers $i = 1,2$. The recorded amplitudes of the reflected waves (after geometric spreading effects have been accounted for) are given by the solution to the Zoeppritz equations (e.g., Yilmaz, 2000).

If the upper-layer parameters are known, it is possible to obtain information about the elastic media properties α_2 and β_2 and the density ρ_2 of the lower layer (e.g., a reservoir) from measurements of the reflected P-wave amplitude at a range of incidence angles i_1. The forward model in this case takes the form

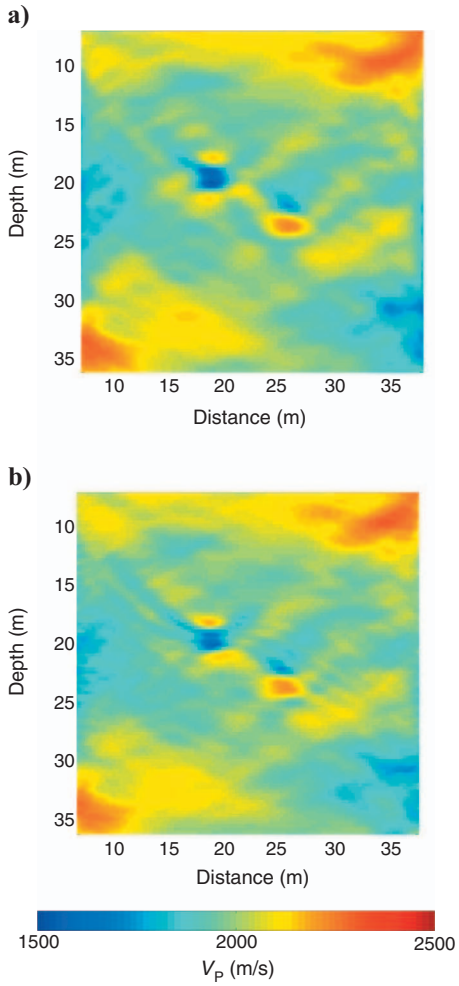

Figure 7. Frequency-domain waveform inversion results using (a) all 30 frequencies and (b) three optimized frequencies.

$$A_1(i_1) = f_x(\alpha_2,\beta_2,\rho_2;i_1),$$

where A_1 is the amplitude of the reflected P-wave and f_x is a function that represents a solution of the highly nonlinear Zoeppritz equations.

The experimental design problem addressed by van den Berg et al. (2003, 2005) attempts to find the offset between the source and a single receiver that maximizes Shannon information about the lower layer, taking account of the full nonlinearity of f_x. They direct Monte Carlo sampling of the prior model PDF, evaluating and maximizing the data-space entropy using the method of Shewry and Wynn (1987). Although it is true that maximizing Shannon information can lead to large variances, this does not occur in the van den Berg et al. (2003, 2005) AVO example.

Guest and Curtis (2009) introduce sequential algorithms that add one receiver at a time to the one-receiver design, using a schema similar to that in equation 22. They develop a novel adaptive Monte Carlo method that, for each design parameter added, successively excludes parts of design space over which the entropy was estimated with low uncertainty to be small. Large computational efficiency gains result, as compared to van den Berg et al.'s methods. This allows optimal survey designs to be found for a survey with multiple ($\sim$10) offsets, which is enough to identify simple patterns of receiver spacing in optimal designs that could be used as rules of thumb (heuristics) for more complex designs.

Guest and Curtis (2010) apply this method to find the optimal set of incident angles (i_1) at the interlayer interface that best constrains the elastic properties of a simulated reservoir from recorded P-wave reflection coefficients; source-receiver offsets at the earth's surface corresponding to these optimal angles at the reflector can then be found with ray-tracing algorithms. Processing algorithms can be optimized to focus attention (e.g., noise-reduction methods and time spent by data processors) on data recorded around those offsets.

By using the shaly-sand model of Goldberg and Gurevich (1998) that relates porosity, clay content, fluid saturation, and a range of other petrophysical properties to V_P and V_S and to density, Guest and Curtis (2009, 2010) find that the optimal design depends significantly on the reservoir's porosity. This is easily understood by examining the critical angles from a forward problem for oil reservoirs of low porosity (uniform prior PDF is 10%–20%), medium porosity (uniform prior PDF is 20%–30%), and high porosity (uniform prior PDF is 30%–40%). Figure 9 shows the corresponding P-wave reflection coefficient ranges and demonstrates that the critical angle changes significantly with the range of possible porosities. In particular, the critical angle is never reached by any survey design for a high-porosity substrate.

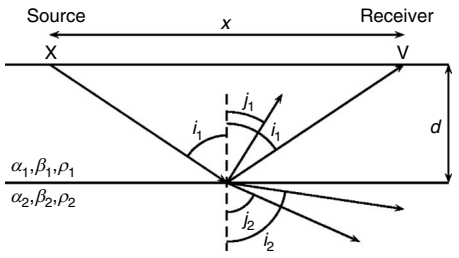

Figure 8. Geometry of an AVA experiment with a single interface at depth d. The distance between the source X and the receiver V is called the offset x. At the interface, P-wave energy incident at angle i_1 is split into a reflected P-wave (angle i_1) and a P-S wave conversion (angle j_1) and is also transmitted into the second layer as a P-wave (angle i_2) and P-S-wave conversion (angle j_2). Amplitudes of each wave are given by the Zoeppritz equations. Properties of the subsurface are density ρ, P-wave velocity α, and S-wave velocity β in layers 1 and 2.

Using the sequential nonlinear design methods with the novel Monte Carlo method, an optimal set of angles is derived to maximize the Shannon information about the parameters of the lower layer. Figure 10 shows the optimal designs for 10 incident angles selected from the range 0°–90° at 0.5° intervals. Three design optimizations are performed, one for each of the three different porosity ranges in

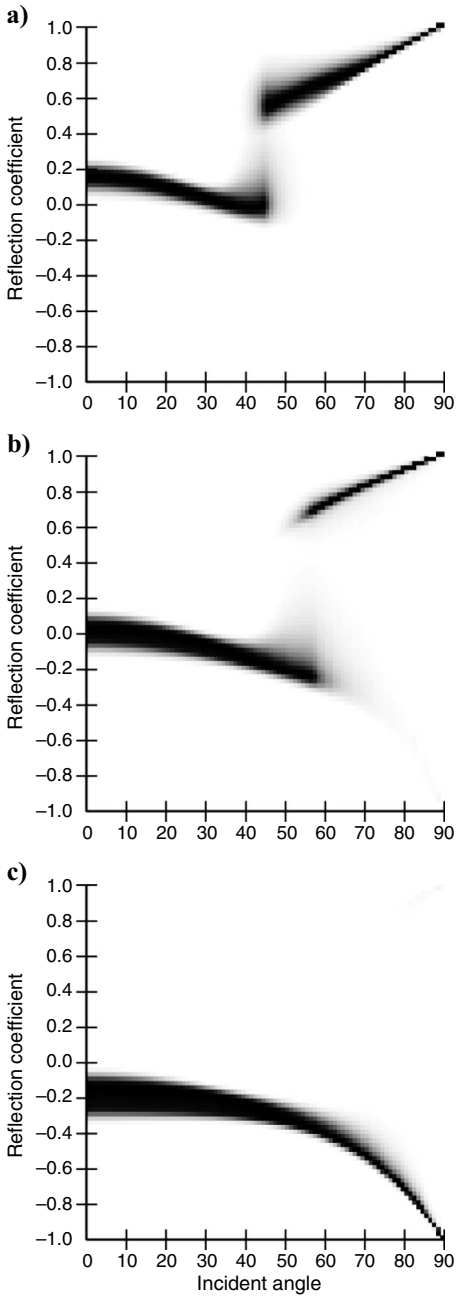

Figure 9. Reflection-coefficient histograms for oil-filled reservoir velocity-density models with varying uniform prior porosity ranges. (a) The low-porosity reservoir (10–20%), (b) the medium-porosity reservoir (20–30%), and (c) the high-porosity reservoir (30%–40%).

Figure 9 (all other petrophysical parameters were allowed to vary within broad ranges defined by their respective a priori probability distributions). As the porosity increases, so does the range of angles spanned by the selected angles. For the low-porosity model (Figure 10a), all angles are located within 0°–51°, with each of the two angle clusters located within 11° envelopes. As the porosity range increases, so does the maximum incident angle and so does the angular range of each of the two envelopes within which angles are selected. Note that the survey design for the midporosity reservoir almost precludes the measurements from the low-porosity design for the far-offset envelopes, illustrating the importance of prior porosity information in AVA survey design.

Comparing Figures 9 and 10, we see that the optimal designs use angles to obtain information from (1) near-zero incident angles and (2) the critical angle or from (3) the highest-curvature region in the absence of a critical angle. The algorithm selects incident angles that display the largest sensitivity to perturbations in the elastic parameters of the second layer. These regions display the maximum uncertainty or entropy in the data PDF $p(\mathbf{d})$. Shewry and Wynn (1987) prove that this implies maximum expected Shannon information about the model parameters.

The results from the 10-angle optimization can be generalized. Figure 11 shows the spatially averaged density of receiver positions calculated from the angle positions in Figure 10 for each of the different-porosity oil-saturated reservoirs (Guest and Curtis, 2010). For high-porosity reservoirs (dotted line), a larger proportion of the selected angles to be processed should be placed at high incident angles (around 55°). As the porosity decreases, the highest density range also shifts to lower angles. For all cases, there is a density minimum near 30°, identifying an information minimum around that angle.

Guest and Curtis (personal communication, 2010) are developing the approach further by directly designing the optimal receiver density curves (similar to those in Figure 11) for a surface seismic AVA survey. The receiver density curve can be encapsulated by a very few design parameters (in this case, receiver density at the beginning, middle, and end of the receiver line with linear interpolation between these points), leading to a fairly low (<10) dimensional design problem. Despite this simplicity, if an appropriate density curve is designed, any number of receivers can be placed according to that design. Thus, for the first time, they have designed full 2D seismic surveys using SED methods — in this case, to find designs that maximize AVA reservoir information.

Interestingly the possible angle distribution is limited to 0°–30° (current practice because commonly used approximations to the Zoeppritz equations break down at larger angles), and Guest and Curtis (personal communication, 2010) find that the dependence of the design on porosity is greatly reduced. The strong dependence is derived from large changes in the critical angle with any change in expected porosity; the lack of dependence when the reduced angular range of 0°–30° occurs is because the critical angle is almost always excluded from this angular range in realistic scenarios (see Figure 10), so the strongest dependence of the AVA relations on porosity is removed from consideration. Hence, Guest and Curtis could create a general, approximately optimal, small-angle design, valid for all reasonable porosities and a range of other reservoir properties. A key result is that this design depends strongly on the number of receivers placed, up to about 100 receivers; thereafter, the design is fairly stable with respect to the number of receivers.

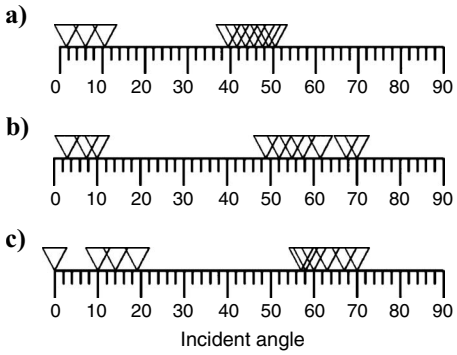

Figure 10. Experimental design for the reservoir model with varying prior porosity ranges using 10 selected traces implementing the Guest and Curtis (2009) method. (a) The final design for the 10%–20% porosity reservoir, (b) the 20%–30% porosity reservoir, and (c) the 30%–40% porosity reservoir. Receivers are allowed to be placed between 0° and 70° incident angle.

Nonlinear experimental design — Microseismic event location surveys

Rather than merely maximizing Shannon information, experimental design theory also can be used to evaluate the differences between designs by understanding the model-resolution capability of geophysical data. Here, we evaluate one simple survey design for locating microseismicity using the bifocal ambiguity measure of equation 19 (Winterfors and Curtis, personal communication, 2010) given in equation 9, and then use the same measure to find an optimal design.

Seismometer arrays are often used to locate and characterize seismicity from fault movement or microseismicity from fracturing that occurs in the earth's subsurface, resulting from naturally occurring stresses caused, for example, by tectonic motion and glacial rebound or from stress changes caused by fluid injection or extraction in subsurface reservoirs. Location is usually achieved by detecting arrival times of seismic waves emanating from each seismic event and by back-projecting the wave energy and hence arrival times through an approximate model representing the earth's subsurface seismic velocity or slowness (the reciprocal of velocity) structure. Methods such as migrating event energy act in a similar manner (back projec-

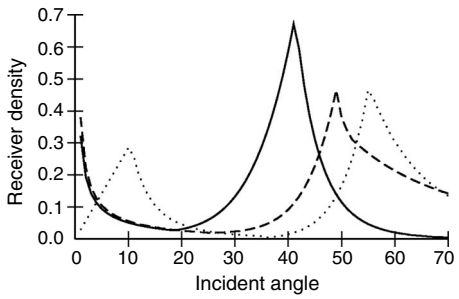

Figure 11. Receiver density plots calculated from the oil-reservoir scenarios considered in Figures 9 and 10. The regression lines have been fitted to 10° moving-average results. The low-porosity (10%–20%) oil-reservoir density is shown by the solid line, the medium-porosity (20%–30%) oil-reservoir density by the dashed line, and the high-porosity (30%–40%) oil-reservoir density by the dotted line.

tion) but use waveform information (rather than only traveltimes) to locate events. This nonlinear inverse problem can result in multimodal posteriori PDFs $Q(\mathbf{m})$ (see Lomax et al. [2009] for a review).

Intuitively, survey designs for locating and monitoring seismicity should maximize geometric aperture of the survey to improve triangulation accuracy; but because of attenuation of the energy in propagating seismic waves, increasing aperture reduces data quality for increased event-to-seismometer distances. Seismometer arrays are often deployed only on the surface (e.g., Rabinowitz and Steinberg, 1990; Steinberg et al., 1995; Winterfors and Curtis, 2008) or only in a borehole (e.g., Curtis et al., 2004). When available, boreholes are sometimes preferred because they offer advantages in terms of signal-to-noise ratio by placing sensors below the weathered near-surface layers of the earth (often highly attenuating to seismic waves), which are assumed to outweigh geometric advantages that might be offered by placing sensors on the surface. However, the question remains as to the quantitative interpretational advantages or disadvantages of monitoring from the surface or from a borehole in a horizontally stratified earth, even ignoring near-surface attenuation effects. Here, we examine this question using a simple example from Winterfors (personal communication, 2010) and based on evaluating the ambiguity measure of Winterfors and Curtis (personal communication, 2010) in equation 19.

Figure 12 is a graphical representation of the ambiguity $W(\xi)$ in a (micro)seismic event location problem given a two-receiver survey design using sensors on the surface of a 2D, horizontally stratified, seismic velocity model. The slownesses shown are the differences between P- and S-wave slownesses in each layer because these can be used to locate events without determining the event-source timing (Tarantola and Valette, 1982). The cross section measures 1000 $\times$ 1000 m, and it is assumed that seismicity could occur anywhere within this cross section with a uniform prior probability. The standard deviation of a priori uncertainty in layer velocities is assumed to be 100 m/s, and the traveltime data uncertainties have standard deviations of 5 ms.

Evaluating the ambiguity measure $W(\xi)$ in equation 19 requires

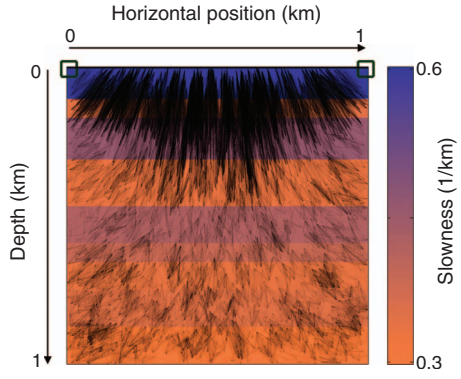

Figure 12. Bifocal ambiguity calculated for the seismic-monitoring design shown by the two receivers (squares) on the surface. Colors show the difference between S- and P-wave seismic slownesses in the subsurface. Black lines connect pairs of locations $\dot{\mathbf{m}}$ and $\ddot{\mathbf{m}}$ in the subsurface that would be indistinguishable given P- and S-wave arrival-time data recorded at the two seismometers. The width (hence darkness) of each line is proportional to the magnitude of the integrand in equation 19, i.e., to the relative contribution of these points to the integral, and hence to precisely how ambiguous the two locations are expected to be postsurvey given recorded data. The value of the integral in this case is $W(\xi) = 2500$ m^2.

that a double integral over model-parameter space be estimated numerically — in this case, using a Monte Carlo method. Winterfors and Curtis (personal communication, 2010) provide several techniques that improve the efficiency of the Monte Carlo method [e.g., recycling forward function evaluations $\mathbf{f}(\dot{\mathbf{m}})$ by also using them for $\mathbf{f}(\ddot{\mathbf{m}})$]. In this case, the forward function $\mathbf{d} = \mathbf{f}(\dot{\mathbf{m}})$ involves solving the eikonal equation to obtain arrival-time data $\mathbf{d}$ at all receiver locations for any event location $\dot{\mathbf{m}}$.

The Monte Carlo scheme considers many pairs of points in parameter space; in Figure 12, each black line connects a pair of locations $\dot{\mathbf{m}}$ and $\ddot{\mathbf{m}}$ in the subsurface encountered in the Monte Carlo integration that would be almost indistinguishable from P- and S-wave arrival-time data recorded at the two surface seismometers, i.e., $\mathbf{f}(\dot{\mathbf{m}}) \approx \mathbf{f}(\ddot{\mathbf{m}})$. The width of each line is proportional to the magnitude of the integrand in $W(\xi)$ in equation 19; that is, width indicates the ambiguity of the two locations in the postsurvey interpretation. Some of the line widths are too small to distinguish clearly, but this is deliberate; only locations that are highly ambiguous have bold lines to indicate that they make the largest contributions to the ambiguity measure $W(\xi)$. Thinner lines connect locations that are still ambiguous but which are expected to be better discriminated by the data and hence provide a lower contribution to $W(\xi)$. In this plot, the density of the black lines therefore indicates the extent and location of model-space ambiguity for a particular survey geometry, and the general orientation of the lines indicates the direction of this ambiguity. Thus, interpreting Figure 12, we see that indistinguishable seismic locations are confined to the top two surface layers for this particular receiver geometry and that the overall direction in which locations cannot be distinguished is approximately vertical. The overall ambiguity calculated for this design is $W(\xi) = 2500$ m^2.

The receiver locations are then optimized such that $W(\xi)$ is minimized using a gradient-descent method. Receivers are allowed to be placed on the surface or in a borehole on one edge of the cross section. Figure 13 shows a similar representation of the ambiguity in the same velocity structure as in Figure 12 but for two optimally located receivers, both of which turn out to be in the borehole. Compared to Figure 12, the ambiguity of event locations (the density of black lines) is far more evenly distributed throughout the cross section. Also apparent is the reduced expected ambiguity for this design, as shown by the reduced line widths; relatively few locations are strongly ambiguous. However, we also see that the distances between ambiguous event locations (line lengths) are generally shorter; hence, we expect that the variance in event locations postsurvey

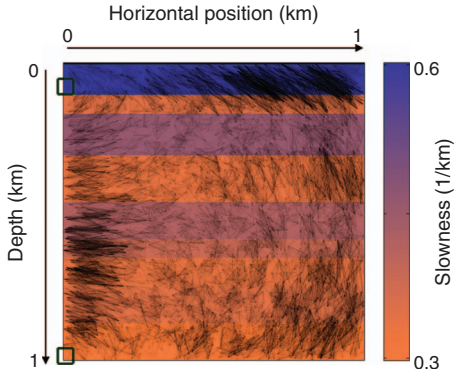

Figure 13. Similar to Figure 12 but for the seismic-monitoring design shown by the two receivers (squares) in the borehole. The value of the integral in this case is $W(\xi) = 1000$ m^2.

should be smaller, as suggested by equation 20. Correspondingly, the overall ambiguity calculated for this design is $W(\xi) = 1000$ m^2.

Nominally, the two-receiver arrays in both surveys have almost the same geometric aperture for locating the events, yet they provide quite different degrees of ambiguity in resulting event locations. Indeed, the better of the two designs actually has lower spatial aperture. The dominant uncertainties in event location are vertical for an array of surface sensors in a horizontally stratified medium. This is because first-arriving waves from the ambiguous event locations take approximately equal time to travel directly to an adjacent high-velocity layer and then follow similar paths horizontally along and within that layer and then up to each seismometer. Hence, data recorded at such seismometers cannot distinguish between the two locations. Using vertically offset borehole receivers above and below such high-velocity layers reduces much of this ambiguity without introducing a similar ambiguity in the horizontal direction because the layers are geometrically orthogonal to the observation well. Thus, first-arriving energy generally cannot travel within any single high-velocity layer to both of the receivers.

Equation 19 provides a useful and intuitive assessment of expected ambiguity in seismic event location and leads to designs that are intuitively reasonable. It is clear from these examples that it would be misleading to design experiments solely on the basis of maximizing apparent aperture without accounting for the effect of wave propagation in horizontally stratified subsurface velocity structures; such a strategy would almost definitely lead to suboptimal designs.

CONCLUSIONS

Major improvements have been achieved in the field of experimental design during the past 15 years, both conceptually and in practice. For instance, the superiority of nontraditional, optimized electrode configurations for geoelectrical surveys is now well recognized. Modern data-acquisition systems allow arbitrary electrode configurations to be implemented quite easily, so it is expected that collecting geoelectrical data with optimized electrode configurations will soon become common practice.

Despite these successes, several concerns remain to be addressed.

First, most geophysical-survey design studies have focused on optimizing data information content. This is certainly an important first step; but to truly optimize a geophysical experiment, other cost/benefit constraints must be considered. For example, for conducting seismic crosshole measurements, one may use a single borehole geophone or a string with several geophones. A single geophone would be relatively cheap to buy or lease, and it would allow the measurement of quite irregular source-receiver patterns. With a geophone string, many measurements could be performed much more efficiently, but the purchase or rental costs would be higher and there would be less flexibility for choosing source-receiver patterns. Future experimental design algorithms need to consider available instrumentation choices as constraints and factor the associated costs in the optimization process.

Second, nonlinear experimental design is critical for a wide range of geophysical problems because few can truly be considered linear over the range of models considered possible according to prior information. The examples presented in this paper demonstrate the general feasibility of nonlinear experimental design to relatively simple problems, but more research is required before such techniques can be applied routinely to most practical survey-design problems. Nevertheless, for the first time, optimal receiver density for AVA/AVO studies has been calculated for complete 2D seismic surveys. Hence, the emergence and proof of concept of practical, industrial, nonlinear design algorithms may be approaching rapidly.

Third, most survey-design algorithms, particularly nonlinear variants, are computationally expensive. This imposes severe constraints on applications to realistic design problems. Improving the computational efficiency of key modules that are called many times while executing a design algorithm is therefore critical. In linear problems, the computation of the model resolution matrix can be improved dramatically using update formulas and parallel computing. Similar efficiencies will be even more important for nonlinear design algorithms.

Finally, easy-to-use computer interfaces must be designed that allow nonspecialists to develop experimental designs, based on different scenarios involving choices in survey costing and possible variations in petrophysical parameterization of earth models.

ACKNOWLEDGMENTS

This paper includes research results to which many of our colleagues have contributed. The authors especially thank Thomas Guest, Emanuel Winterfors, and Peter Stummer for providing figures and information describing their design methods. We also thank two anonymous reviewers for helpful suggestions that improved the clarity of the manuscript.

REFERENCES

Ajo-Franklin, J., 2009, Optimal experiment design for time-lapse traveltime tomography: Geophysics, **74**, no. 4, Q27–Q40, doi: 10.1190/1.3141738

Atkinson, A. C., and A. N. Donev, 1992, Optimum experimental designs: Clarendon Press.

Barth, N., and C. Wunsch, 1990, Oceanographic experiment design by simulated annealing: Journal of Physical Oceanography, **20**, 1249–1263, doi: 10.1175/1520-0485(1990)020<1249:OEDBSA>2.0.CO;2

Brenders, A. J., and R. G. Pratt, 2007, Efficient waveform tomography for lithospheric imaging: Implications for realistic, two-dimensional acquisition geometries and low-frequency data: Geophysical Journal International, **168**, no. 1, 152–170, doi: 10.1111/j.1365-246X.2006.03096.x

Butler, D. K., ed., 2005, Near-surface geophysics: SEG.

Coles, D., and F. Morgan, 2009, A method of fast, sequential experimental design for linearized geophysical inverse problems: Geophysical Journal International, **178**, 145–158, doi: 10.1111/j.1365-246X.2009.04156.x

Coscia, I., L. Marescot, H. R. Maurer, S. A. Greenhalgh, and A. G. Green, 2008, Experimental design for cross hole electrical resistivity tomography data sets: EAGE Near-Surface Meeting.

Cox, D. R., 1958, Planning of experiments: John Wiley & Sons, Inc.

Curtis, A., 1999a, Optimal design of focussed experiments and surveys: Geophysical Journal International, **139**, no. 1, 205–215, doi: 10.1046/j.1365-246X.1999.00947.x

———, 1999b, Optimal experiment design: Cross-borehole tomographic examples: Geophysical Journal International, **136**, no. 3, 637–650, doi: 10.1046/j.1365-246x.1999.00749.x

———, 2004a, Theory of model-based geophysical survey and experimental design, Part A — Linear problems: The Leading Edge, **23**, 997–1004, doi: 10.1190/1.1813346

———, 2004b, Theory of model-based geophysical survey and experimental design, Part B — Nonlinear problems: The Leading Edge, **23**, 1112–1117, doi: 10.1190/1.1825931

Curtis, A., and A. Lomax, 2001, Prior information, sampling distributions and the curse of dimensionality: Geophysics, **66**, 372–378, doi: 10.1190/1.1444928

Curtis, A., and H. R. Maurer, 2000, Optimizing the design of geophysical experiments — Is it worthwhile? Eos, Transactions, American Geophysical Union, **81**, 224–225.

Curtis, A., A. Michelini, D. Leslie, and A. Lomax, 2004, A deterministic algorithm for experimental design applied to tomographic and microseismic monitoring surveys: Geophysical Journal International, **157**, 595–606, doi: 10.1111/j.1365-246X.2004.02114.x

Curtis, A., and R. Snieder, 1997, Reconditioning inverse problems using the genetic algorithm and revised parameterization: Geophysics, **62**,

1524–1532, doi: 10.1190/1.1444255
Curtis, A., and R. Wood, 2004, Optimal elicitation of probabilistic information from experts, *in* A. Curtis and R. Wood, eds., Geological prior information: Informing science and engineering: Geological Society of London Special Publication 239, 127–145.
Fedorov, V. V., 1972, Theory of optimal experiments: Academic Press Inc.
Furman, A., T. P. A. Ferre, and G. L. Heath, 2007, Spatial focusing of electrical resistivity surveys considering geologic and hydrologic layering: Geophysics, **72**, no. 2, F65–F73, doi: 10.1190/1.2433737
Glenn, W. E., and S. H. Ward, 1976, Statistical evaluation of electrical sounding methods, Part 1: Experimental design: Geophysics, **41**, 1207–1221.
Goldberg, I., and B. Gurevich, 1998, A semi-empirical velocity-porosity-clay model for petrophysical interpretation of P- and S-velocities: Geophysical Prospecting, **46**, 271–285.
Griffiths, D. H., and J. Turnbull, 1985, A multi-electrode array for resistivity surveying: First Break, **3**, no. 7, 16–20.
Guest, T., and A. Curtis, 2009, Iteratively constructive sequential design of experiments and surveys with nonlinear parameter-data relationships: Journal of Geophysical Research, **114**, B04307, doi: 10.1029/2008JB005948
———, 2010, Optimal trace selection for amplitude-variation-with-angle (AVA) processing of shale-sand reservoirs: Geophysics, **75**, no. 4, doi: 10.1190/1.3462291.
Haber, E., L. Horesh, and L. Tenorio, 2008, Numerical methods for experimental design of large-scale linear ill-posed inverse problems: Inverse Problems, **24**, 055012.
Hardt, M., and F. Scherbaum, 1994, The design of optimum networks for aftershock recordings: Geophysical Journal International, **117**, 716–726, doi: 10.1111/j.1365-246X.1994.tb02464.x
Hennig, T., A. Weller, and M. Moller, 2008, Object orientated focussing of geoelectric multielectrode measurements: Journal of Applied Geophysics, **65**, no. 2, 57–64, doi: 10.1016/j.jappgeo.2008.04.007
Jones, A. G., and J. H. Foster, 1986, An objective real-time data-adaptive technique for efficient model resolution improvement in magnetotelluric studies: Geophysics, **51**, 90–97, doi: 10.1190/1.1442043
Kijko, A., 1977, An algorithm for the optimal distribution of a regional seismic network — I: Pure and Applied Geophysics, **115**, 999–1009, doi: 10.1007/BF00881222
Liner, C. L., W. D. Underwood, and R. Gobeli, 1999, 3-D seismic survey design as an optimization problem: The Leading Edge, **18**, 1054–1060.
Loke, M. H., and R. D. Barker, 1996, Practical techniques for 3D resistivity surveys and data inversion: Geophysical Prospecting, **44**, no. 3, 499–523, doi: 10.1111/j.1365-2478.1996.tb00162.x
Loke, M. H., and P. Wilkinson, 2009, Rapid parallel computation of optimised arrays for electrical imaging surveys: EAGE Near-Surface Conference, Proceedings, B11.
Lomax, A., A. Michelini, and A. Curtis, 2009, Earthquake location, direct, global-search methods: *in* R. A. Meyers, ed., Encyclopedia of complexity and system science: Springer.
Matsu'ura, M., and N. Hirata, 1982, General least-squares solutions to quasi-linear inverse problems with a priori information: Journal of Physics of the Earth, **30**, 451–468.
Maurer, H. R., and D. E. Boerner, 1998a, Optimized and robust experimental design: A non-linear application to EM sounding: Geophysical Journal International, **132**, no. 2, 458–468, doi: 10.1046/j.1365-246x.1998.00459.x
———, 1998b, Optimized design of geophysical experiments: The Leading Edge, **17**, 1119.
Maurer, H. R., D. E. Boerner, and A. Curtis, 2000, Design strategies for electromagnetic geophysical surveys: Inverse Problems, **16**, no. 5, 1097–1118, doi: 10.1088/0266-5611/16/5/302
Maurer, H. R., S. A. Greenhalgh, and S. Latzel, 2009, Frequency and spatial sampling strategies for crosshole seismic waveform spectral inversion experiments: Geophysics, **74**, no. 6, WCC11–WCC21.
Menke, W., 1984, Geophysical data analysis: Discrete inverse theory: Academic Press Inc.
Mitchell, T. J., 1974, An algorithm for the construction of 'D-optimal' experimental designs: Technometrics, **16**, no. 2, 203–210, doi: 10.2307/1267940
Oldenborger, G. A., and P. S. Routh, 2009, The point-spread function measure of resolution for the 3-D electrical resistivity experiment: Geophysical Journal International, **176**, 405–414, doi: 10.1111/j.1365-246X.2008.04003.x
Oldenburg, D. W., 1978, The interpretation of direct current resistivity measurements: Geophysics, **43**, 610–625, doi: 10.1190/1.1440840
Pratt, R. G., 1999, Seismic waveform inversion in the frequency domain, Part 1: Theory and verification in a physical scale model: Geophysics, **64**, 888–901, doi: 10.1190/1.1444597
Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery, 2007, Numerical recipes: The art of scientific computing, 3rd ed.: Cambridge University Press.
Rabinowitz, N., and D. M. Steinberg, 1990, Optimal configuration of a seismographic network: A statistical approach: Bulletin of the Seismological Society of America, **80**, no. 1, 187–196.
Routh, P. S., G. A. Oldenborger, and D. W. Oldenburg, 2005, Optimal survey design using the point spread function measure of resolution: 75th Annual International Meeting, SEG, Expanded Abstracts, 1033–1036.
Shannon, C. E., 1948, A mathematical theory of communication: The Bell System Technical Journal, **27**, 379–423, 623–656.
Shewry, M. C., and H. P. Wynn, 1987, Maximum entropy sampling: Journal of Applied Statistics, **14**, no. 2, 165–170, doi: 10.1080/02664768700000020
Sirgue, L., and R. G. Pratt, 2004, Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies: Geophysics, **69**, 231–248, doi: 10.1190/1.1649391
Stark, P. B., 2008, Generalizing resolution: Inverse Problems, **24**, no. 3, 034014, doi: 10.1088/0266-5611/24/3/034014
Steinberg, D. M., N. Rabinowitz, Y. Shimshoni, and D. Mizrachi, 1995, Configuring a seismographic network for optimal monitoring of fault lines and multiple sources: Bulletin of the Seismological Society of America, **85**, 1847–1857.
Stummer, P., H. R. Maurer, and A. G. Green, 2004, Experimental design: Electrical resistivity data sets that provide optimum subsurface information: Geophysics, **69**, 120–139, doi: 10.1190/1.1649381
Stummer, P., H. R. Maurer, H. Horstmeyer, and A. G. Green, 2002, Optimization of DC resistivity data acquisition: Real-time experimental design and a new multielectrode system: IEEE Transactions on Geoscience and Remote Sensing, **40**, 2727–2735, doi: 10.1109/TGRS.2002.807015
Taguchi, G., 1987, Systems of experimental design (volumes 1 and 2 [1976 and 1977] with 1987 translation): UNIPUB.
Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: Society for Industrial and Applied Mathematics.
Tarantola, A., and B. Valette, 1982, Generalized nonlinear problems solved using the least squares criterion: Reviews of Geophysics, **20**, no. 2, 219–232, doi: 10.1029/RG020i002p00219
Telford, W. M., L. P. Geldart, and R. E. Sheriff, 1990, Applied geophysics: Cambridge University Press.
van den Berg, J., A. Curtis, and J. Trampert, 2003, Bayesian, nonlinear experimental design applied to simple, geophysical examples: Geophysical Journal International, **55**, 411–421.
———, 2005, Correction to Bayesian, nonlinear experimental design applied to simple, geophysical examples: Geophysical Journal International, **161**, no. 2, 265–265.
Wilkinson, P., P. Meldrum, J. Chambers, O. Kuras, and R. Ogilvy, 2006, Improved strategies for the automatic selection of optimized sets of electrical resistivity tomography measurement configurations: Geophysical Journal International, **167**, 1119–1126, doi: 10.1111/j.1365-246X.2006.03196.x
Winterfors, E., and A. Curtis, 2008, Numerical detection and reduction of non-uniqueness in nonlinear inverse problems: Inverse Problems, 24, no. 025016.
Wolpert, D. H., and W. G. Macready, 1997, No free lunch theorems for optimization: IEEE Transactions on Evolutionary Computation, **1**, no. 1, 67–82.
Yilmaz, O., 2000, Seismic data processing: SEG.