Optimizing the design of geophysical experiments: Is it worthwhile?

ANDREW CURTIS, Schlumberger Cambridge Research, Cambridge, U.K.
HANSRUEDI MAURER, ETH, Zurich, Switzerland

Coordinated by Guillaume Cambois

Determining the structure, composition, and state of the earth's subsurface from measured data is the principal task of many geophysical experiments or surveys. Standard procedures involve the recording of appropriate data sets followed by the application of data analysis techniques to extract the desired information. Whereas the importance of new tools for the analysis stage of an experiment is well recognized, much less attention seems to be paid to improving the data acquisition. A measure of the effort allocated to data analysis research relative to that devoted to data acquisition research is presented in Figure 1. Since 1955 there have been more than 10 000 publications on inversion methods alone, but in the same period only 100 papers on experimental design have appeared in journals. Considering that the acquisition component of an experiment defines what information will be contained in the data, and that no amount of data analysis can compensate for the lack of such information, we suggest that greater effort be made to improve survey planning techniques. Furthermore, given that logistical and financial constraints are often stringent and that relationships between geophysical data and model parameters describing the earth's subsurface are generally complicated, optimizing the design of an experiment may be quite challenging. Here, we present a short review of experimental design procedures that optimize the benefit of a field survey, such that maximum information about the target structures is obtained at minimum cost.

Experimental design techniques. Statistical experimental design techniques were popularized by G. Taguchi who, as a government advisor, developed and implemented design methods—often called Taguchi methods (Taguchi, 1987)—across a broad spectrum of Japanese industrial processes such that final products were as robust as possible to fluctuations in the manufacturing process. Over only a few years this led to a revolution 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. It is their simplicity that makes these methods so generally applicable.

Editor's note: This article was published in EOS Forum*, 16 May 2000, Vol. 81, No. 20, copyright by the American Geophysical Union. It has been slightly modified to conform to* TLE *style.*

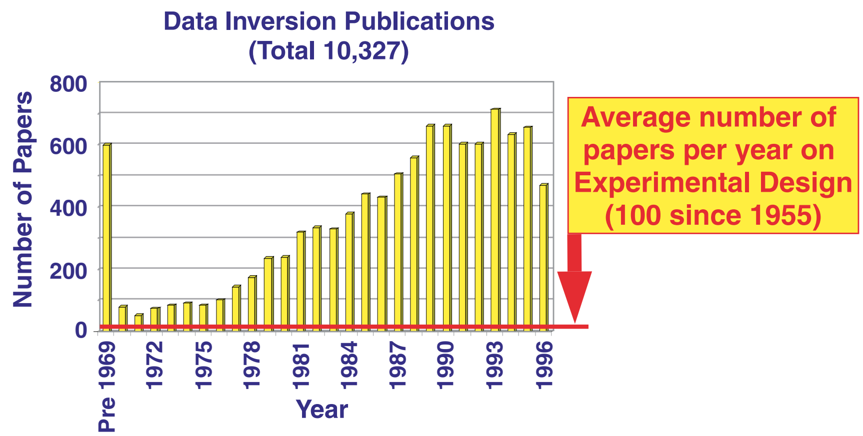

Figure 1. Number of papers published on inversion methods and experimental design as a function of year.

The theoretical framework of statistical experimental design techniques is based on clear definitions of the:

- experimental goals,
- experimental constraints, and
- well-defined physical relationships between the data and the model parameters.

Experimental goals may be extremely diverse. Situations where statistical design techniques have been used in geophysics include (1) determining optimal locations of seismometers with the goal of locating earthquakes with minimum uncertainty (e.g., Rabinowitcz and Steinberg, 1990); (2) designing source/receiver geometries for ship-based tomography that optimally detects subsurface velocity anomalies (Barth and Wunsch, 1990); (3) designing surveys for electromagnetic tomography from the ground surface with the goal of constraining optimally the subsurface conductivity structure (Maurer and Boerner, 1998); and (4) planning crosswell seismic tomographic surveys that illuminate the interwell structure optimally (Curtis, 1999).

Experimental constraints are usually unique for each experiment. They define the range of possible experimental designs. A key limitation is the accessibility of the investigation area, because it is usually not possible to place sources and receivers everywhere. Other constraints are imposed by instrumental specifications, such as frequency and dynamic range. Finally, each survey is ultimately limited by financial constraints.

Physical relationships between the observed data and the properties of the earth impose theoretical limitations with respect to uniqueness and resolution power. Potential field methods, for example, suffer from inherent ambiguities, and some diffusive EM techniques are known to have limited spatial resolution.

Having specified the experimental goals, constraints, and physical model-data relationships, the general recipe of statistical experimental design can be described as follows:

1) Define the data space that includes all realizable survey configurations.
2) Define the model space that includes the complete range of plausible earth models.
3) Find the survey configuration that constrains the subsurface model(s)

optimally with respect to the experimental goals, given the constraints and the physical relationships.

Finding an optimal configuration requires a clear definition of the term "optimal." For example, a good survey configuration should lead to high accuracy and reliability of the model parameter estimates. Moreover, the configuration should be realizable easily and require minimal financial effort. In some cases, for example, expected postsurvey model parameter accuracy and reliability might be specified by means of linear inverse theory (model covariance and resolution), whereas the financial effort might be considered to be proportional to the number of data receivers deployed. Other factors may also contribute to the quality of a particular survey configuration. The only requirement is that these factors can be expressed by a mathematical expression.

Generally a huge number of possible survey configurations needs to be considered. Global optimization schemes, such as genetic algorithms or simulated annealing have proven to be efficient means to search for the best survey configuration.

A few examples. A straightforward application of statistical experimental design techniques is in designing an optimal seismic crosswell tomography experiment. Traveltimes of seismic energy between sources in one well and receivers in another (with either sources or receivers on the ground surface) are used to constrain the interwell

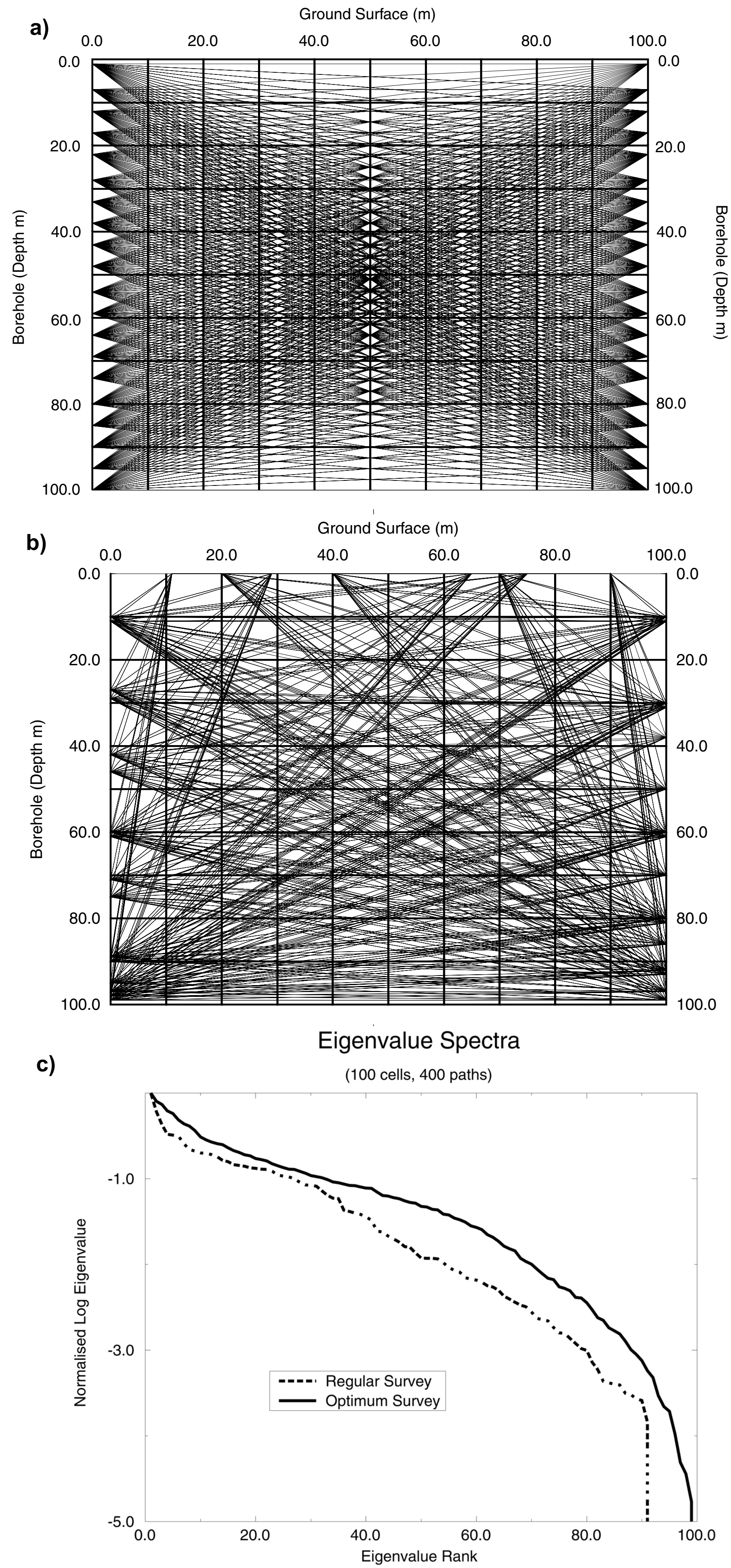

Figure 2. A crosswell tomographic experiment. (a) Left and right hand sides represent wells, top the ground surface. Velocity model to be constrained consists of regularly spaced square cells outlined by bold lines. Thin lines show predicted raypaths between 20 seismic sources and 20 receivers (hence 400 raypaths), each distributed evenly down the length of each well. The velocity model is assumed to be approximately constant over this 100 m interval. (b) Similar to (a), but this time using an optimized geometry of seismic sources and receivers. (c) Logarithm of eigenvalue spectra of the tomographic inverse problems corresponding to the experimental designs depicted in (a)-dashed line and (b)-solid line, normalized by the largest eigenvalue in each case.

seismic velocity structure. In a first step, a large number of source and receiver positions are specified within the wells and at the surface. This defines the data space that comprises all possible source-receiver combinations and hence all possible seismic raypaths. The optimization algorithm then has to identify an optimal subset of these rays.

Assume that the experimental goals are to obtain maximum information about the interwell velocity structure, which in turn is represented by a parameterized model. Such information will be obtained, postexperiment, by linearly inverting the measured crosswell seismic traveltimes to constrain the model. A measure of the amount of model information available from any particular survey design can be constructed from the eigenvalues of the inverse problem: The eigenvalues are mathematical quantities that describe how well any particular linear combination of the model parameters can be constrained by the data. Eigenvalues are directly related to model resolution and uncertainty. In this experiment we would like to measure information about the entire model space, so we are equally interested in maximizing all of the eigenvalues (e.g., Maurer and Boerner 1998; Curtis, 1999). If we had been interested in maximizing information about a particular part of the model then we would try to maximize eigenvalues that were related to those particular model parameters (Curtis 1999).

An example is depicted in Figure 2. Sources can be positioned either at the ground surface or in the left hole, whereas receivers can be placed at the surface or in the right hole. Assuming an approximately homogeneous velocity model (hence straight raypaths) and that the true velocity model can be represented approximately using square model cells with side length 10 m, an even distribution of 20 source and 20 receiver positions (Figure 2a) results in an eigenvalue spectrum of the corresponding tomographic inverse problem as shown by the dashed line in Figure 2c. An optimized configuration is depicted in Figure 2b. Its superiority is demonstrated with the corresponding eigenvalue spectrum in Figure 2c (solid line) since all eigenvalues have been increased. Also note that only a few receivers are required at the surface, whereas both the source and receiver density increase with depth down the length of each well. Here we have considered a fixed number of source and receiver positions. Therefore, the quality of a design is defined only by the corresponding eigenvalue spectrum. Alternatively, one could simultaneously maximize the model information and minimize the number of rays, which leads to cost-optimized survey layouts.

In the crosswell example, the velocity fluctuations were assumed not too severe. This makes the source-receiver geometry virtually independent of the velocity model, i.e. the problem is approximately linear. In other applications, the optimal experimental configuration may depend on the "true" model parameter values themselves. This is the case for any nonlinear problems, such as the earthquake location problem. Let us assume that we want to design a local seismic network devoted to monitoring and locating earthquakes along a fault zone. An optimal network to locate earthquakes occurring toward one end of the fault may be very different from that required for earthquakes toward the other end. In such situations one might opt to use Bayesian experimental design techniques (Chaloner and Verdinelli, 1995). These techniques account for all prior expectations of where the earthquake is likely to occur, based, for instance, on the locations and magnitudes of previous earthquakes. Assume that we can describe our prior expectations by a probability distribution. This encapsulates the relative likelihood of occurrence at all locations along or around known faults. Using Bayesian design techniques optimizes the expected accuracy of the earthquake location procedure, weighted by the relative likelihood of occurrence at each location. In this way a survey can be designed that is most likely to give well constrained source locations, given all knowledge prior to the earthquake. Likewise, seismic gaps could also be considered in the design process.

Finally, we consider a simple set of experiments to determine a polynomial approximation to the P-T melting threshold of a particular crystalline material. Given a set of melting point determinations at atmospheric pressure, 5 GPa and 10 Gpa, the polynomial coefficients can be estimated. The pressures and melting temperatures recorded may be uncertain due to inaccuracy of equipment used. Hence, the estimates of the polynomial coefficients are also uncertain. Given these results, and the fact that each experiment takes a significant amount of effort so that we wish to minimize the number of trials, we must determine at which pressures the next set of experiments should be performed in order to maximize information about the P-T curve. In fact, such experimental situations can be designed optimally using almost identical statistical experimental design techniques to those described above; one simply reformulates the method of constraining the polynomial coefficients from the data as an inverse problem and follows the procedures already described.

Conclusion—the idea of GEDNET. Because we believe that statistical experimental design techniques may have the potential to improve geophysical measuring practice, we propose to establish a Geophysical Experimental Design NETwork (GEDNET). This should involve an interdisciplinary working group devoted to collecting and distributing information on experimental design techniques and to coordinating future research. A striking feature of experimental design techniques is that the same basic principles are common to a wide range of geophysical applications. With GEDNET, we hope to stimulate exchange of ideas and transfer of information between individual disciplines. Those who are interested should check the Web page *www.gednet.geophys.ethz.ch* and/or subscribe to the email list *gednet@ethz.ch*.

Suggestions for further reading. "Oceanographic experiment design by simulated annealing" by Barth and Wunsch (*Journal of Physical Oceanography*, 1990). "Bayesian experimental design: A review" by Chaloner and Verdinelli (*Statistical Science*, 1995). "Optimal experiment design: Cross-borehole tomographic examples" by Curtis (*Geophysical Journal International*, 1999). "Optimal design of focussed experiments and surveys" by Curtis (*Geophysical Journal International*, 1999). "Optimized and robust experimental design: A nonlinear application to EM sounding" by Maurer and Boerner (*Geophysical Journal International*, 1998). "Optimal configuration of a seismographic network: A statistical approach" by Rabinowitcz and Steinberg (*Bulletin of the Seismological Society of America*, 1990). *Systems of Experimental Design* by Taguchi (Volume 1 1976, Volume 2 1977; translation 1987, UNIPUB, Langham, Maryland).

Corresponding author: A. Curtis, curtis@cambridge.scr.slb.com