MONTE CARLO METHODS IN GEOPHYSICAL INVERSE PROBLEMS

Malcolm Sambridge
Research School of Earth Sciences
Institute of Advanced Studies
Australian National University
Canberra, ACT, Australia

Klaus Mosegaard
Niels Bohr Institute for Astronomy
Physics and Geophysics
University of Copenhagen
Copenhagen, Denmark

Received 27 June 2000; revised 15 December 2001; accepted 9 September 2002; published 5 December 2002.

[1] Monte Carlo inversion techniques were first used by Earth scientists more than 30 years ago. Since that time they have been applied to a wide range of problems, from the inversion of free oscillation data for whole Earth seismic structure to studies at the meter-scale lengths encountered in exploration seismology. This paper traces the development and application of Monte Carlo methods for inverse problems in the Earth sciences and in particular geophysics. The major developments in theory and application are traced from the earliest work of the Russian school and the pioneering studies in the west by *Press* [1968] to modern importance sampling and ensemble inference methods. The paper is divided into two parts. The first is a literature review, and the second is a summary of Monte Carlo techniques that are currently popular in geophysics. These include simulated annealing, genetic algorithms, and other importance sampling approaches. The objective is to act as both an introduction for newcomers to the field and a comprehensive reference source for researchers already familiar with Monte Carlo inversion. It is our hope that the paper will serve as a timely summary of an expanding and versatile methodology and also encourage applications to new areas of the Earth sciences. *Index Terms:* 3260 Mathematical Geophysics: Inverse theory; 1794 History of Geophysics: Instruments and techniques; 0902 Exploration Geophysics: Computational methods, seismic; 7294 Seismology: Instruments and techniques; *Keywords:* Monte Carlo nonlinear inversion numerical techniques
Citation: Sambridge, M., and K. Mosegaard, Monte Carlo Methods in Geophysical Inverse Problems, *Rev. Geophys, 40(3)*, 1009, doi:10.1029/2000RG00089, 2002.

1. INTRODUCTION

[2] *Hammersley and Handscomb* [1964] define *Monte Carlo* methods as "the branch of experimental mathematics that is concerned with experiments on random numbers." (A glossary is included to define some commonly used terms. The first occurrence of each is italicized in text.) Today, perhaps, we would modify this definition slightly to "experiments making use of random numbers to solve problems that are either probabilistic or deterministic in nature." By this we mean either the simulation of actual random processes (a probabilistic problem) or the use of random numbers to solve problems that do not involve any random process (a deterministic problem). The origin of modern Monte Carlo methods stem from work on the atomic bomb during the Second World War, when they were mainly used for numerical simulation of neutron diffusion in fissile material, that is, a probabilistic problem. Later it was realized that Monte Carlo methods could also be used for deterministic problems, for example, evaluating multidimensional integrals. Early successes came in the fields of operations research: *Thomson* [1957] describes a Monte Carlo simulation of the fluctuations of traffic in the British telephone system.

[3] In the 50 years since the modern development of Monte Carlo methods by Ulam, von Neumann, Fermi, and Metropolis, they have been applied to a large array of problems in the physical, mathematical, biological, and chemical sciences (see *Hammersley and Handscomb* [1964] for an early but still very readable account of their origins and uses). Although the phrase "Monte Carlo method" was first used by *Metropolis and Ulam* [1949], there are documented examples of essentially the same principles being applied much earlier. *Kelvin* [1901] had described the use of "astonishingly modern Monte Carlo techniques" (as noted by *Hammersley and Handscomb* [1964]) in a discussion of the Boltzmann equation. Earlier still, *Hall* [1873] recounts numerical experiments to determine the value of π by injured officers during the American Civil War. This procedure consisted of throwing a needle onto a board containing parallel straight lines. The statistics of number of times the needle intersected each line could be used to estimate π. The usefulness of Monte Carlo type of numerical experiments was therefore known well before the beginning of the

Copyright 2002 by the American Geophysical Union.

8755-1209/02/2000RG000089*$15.00*
1009, doi:10.1029/2000RG000089

century; however, their systematic development and widespread use had to wait for the arrival of the electronic computer.

[4] The direct simulation of probability distributions is at the basis of all Monte Carlo methods. The early work of *Metropolis et al*. [1953] was the first to show how to sample a space according to a Gibbs-Boltzmann distribution, using simple probabilistic rules. Today the development of Monte Carlo techniques and the underlying statistical theory is a large and active area of research [*Flournay and Tsutakawa*, 1989]. Earth scientists have embraced the use of Monte Carlo methods for more than 30 years. This paper traces some of those developments and, in particular, the use of Monte Carlo methods in *inverse problems*, where information is to be inferred from indirect data, for example, estimating the variations of seismic wave speed at depth in the Earth from observations at the surface. Real geophysical observations are often noisy and incomplete and always imperfectly constrain the quantities of interest. Monte Carlo techniques are one of a number of approaches that have been applied with success to geophysical inverse problems. Over the past 15 years the range of problems to which they have been applied has grown steadily. The purpose of this review paper is to summarize the role played by Monte Carlo methods in (mainly) geophysical *inversion* and also to provide a starting point for newcomers to the field.

[5] This paper consists of two parts. The first is a literature review, which describes the origins and major developments in the use of Monte Carlo methods for geophysical inverse problems. It is hoped that this will give an overview of the field to the newcomer and act as a source of references for further study. The second part of the paper is intended as more of a tutorial. Here we describe some of the details of how to use modern Monte Carlo methods for inversion, parameter estimation, optimization, uncertainty analysis, and *ensemble inference*. We have tried to emphasize the limitations as well as the usefulness of Monte Carlo–based methods and also to highlight some of the trends in current research. In addition to an extensive bibliography and glossary of common terms we have also included a list of world wide web addresses where (at the time of writing) further material, computer code, and other information can be found. It is hoped that this will serve as a starting point for the interested reader to explore this active interdisciplinary research field for themselves.

2. A BRIEF HISTORY OF MONTE CARLO INVERSION IN GEOPHYSICS

2.1. Beginnings of Monte Carlo Inversion

[6] In the summer of 1966 the third international symposium on Geophysical Theory and Computers was held at Cambridge, United Kingdom. The subsequent proceedings were published a year later as a special issue of the *Geophysical Journal of the Royal Astronomical Society* and contain some classic papers. One of these is the now famous article by *Backus and Gilbert* [1967], which, along with several others by the same authors [*Backus and Gilbert*, 1968, 1970], established the foundations of geophysical inverse theory. In this paper it was shown that nonuniqueness was a fundamental property of geophysical inverse problems; that is, if any *Earth model* could be found to satisfy "gross Earth data," then an infinite number of them would exist. In the same paper it was shown how this nonuniqueness could be exploited to generate unique models with special properties as an aid to interpretation. In the same volume is a paper by *Keilis-Borok and Yanovskaya* [1967] (describing earlier work in the USSR), which was the first to introduce Monte Carlo inversion methods into geophysics. From that date the use of Monte Carlo inversion techniques has become widespread in geophysics, but interestingly enough their initial appeal was that they offered a way of dealing with the nonuniqueness problem.

[7] At this time *Monte Carlo inversion (MCI)* meant generating discrete Earth models in a uniform random fashion between pairs of upper and lower bounds, which were chosen a priori. Each generated Earth model was tested for its fit to the available data and then accepted or rejected. The final set of accepted Earth models were used for interpretation [*Press*, 1970b]. As the computational power became available in the latter part of the 1960s, Monte Carlo inversion became feasible for some important problems in seismology. The first applications were to the inversion of seismic body-wave travel times (compressional and shear) and 97 eigenperiods of the Earth's free oscillations for variations in the Earth's compressional (α), shear (β) wave velocities, and density (ρ) as a function of depth [*Press*, 1968; *Wiggins*, 1969; *Press*, 1970a, 1970b].

[8] The main appeal of MCI was that it avoided all assumptions of linearity between the observables and the unknowns representing the Earth model upon which most previous techniques relied. In addition, it was thought that a measure of uniqueness of the solutions would be obtained by examining the degree to which the successful models agreed or disagreed [*Press*, 1968]. The original Monte Carlo paper by *Keilis-Borok and Yanovskaya* [1967] introduced the "hedgehog" inversion (attributed to V. Valius and later published by *Valius* [1968]), which sought to map out a region of *acceptable models* in *parameter space*. This was done by deterministically sampling all models in the vicinity of an acceptable model, which had previously been determined by MCI. The whole process could then be repeated many times over. This approach was later used in the estimation of upper mantle Q structure from Rayleigh wave attenuation [*Burton and Kennett*, 1972; *Burton*, 1977] and in other surface-wave dispersion studies [*Biswas and Knopoff*, 1974].

[9] Shortly after its introduction, criticisms of Monte

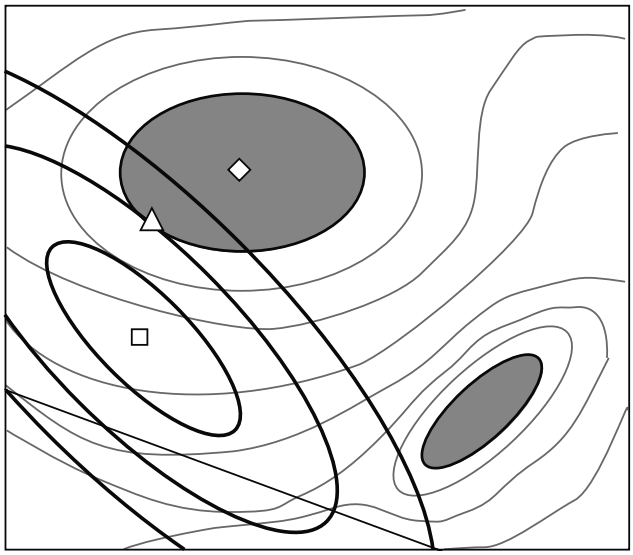

Figure 1. Contours of a data misfit function, $\phi(\mathbf{m})$, (shaded) in the parameter space of a nonlinear problem. The two shaded areas represent the regions of acceptable data fit, while the darker elliptical lines are contours of some regularization function, $\psi(\mathbf{m})$. The diamond represents the model with the best data fit and is distinct from the triangle, which is the data-acceptable model with least ψ. The square is the model with minimum ψ, but it does not satisfy the data.

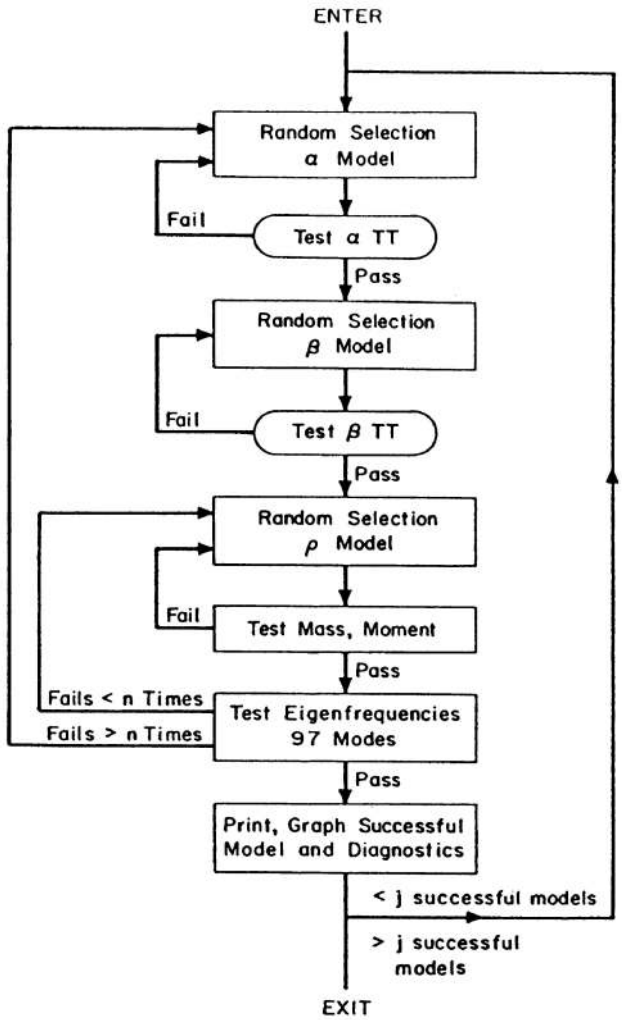

Figure 2. The flow chart of the early Monte Carlo algorithm used by *Press* [1968]. Note that the 1-D Earth model had to satisfy data constraints on travel times, eigenfrequencies, and mass and moment of inertia of the Earth before passing into the output population (see Figure 3). (From *Press* [1968].)

Carlo inversion followed. One problem was that it is never known whether sufficient number of models had been tested. It was always possible that acceptable models may exist that bear no resemblance to the satisfactory models that had been obtained; hence the real Earth may lay outside of the estimated "nonuniqueness bounds." An uncomfortable possibility was that the acceptable models might form multiple unconnected "islands" in parameter space (see Figure 1). An MCI approach might miss some of these islands altogether. (In the work of *Press* [1968], 5 million Earth models were tested, and just 6 were found that passed all data tests. See Figures 2 and 3). In practice, this meant that sets of upper and lower bounds estimated by MCI could not be literally interpreted as "hard" bounds on, say, velocity or density as a function of depth. For this reason *Press* [1970b] refers to his estimated envelope of acceptable Earth models as "a guide to hypotheses rather than firm conclusions."

[10] An approach developed by *Anderssen and Senata* [1971, 1972] went some way to answering these criticisms. They developed a statistical procedure for estimating the reliability of a given set of nonuniqueness bounds. Their method was subsequently applied to the inversion of seismic and density profiles by a number of authors [*Worthington et al.*, 1972, 1974; *Goncz and Cleary*, 1976].

[11] Another criticism of MCI, argued by *Haddon and Bullen* [1969], was that the successful models generated were likely to contain unnecessary complexity (e.g., the typical small-scale oscillations that had been obtained in velocity or density depth profiles). This was because the likelihood of generating a parametrically simple model was very small, and hence MCI results were biased toward physically unrealistic Earth models. One way this difficulty was addressed was by seeking families of "uncomplicated" Earth models, with acceptable fit to data. *Wiggins* [1969] devised a parameterization for 1-D velocity profiles that allowed one to impose velocity, velocity gradient with depth, and velocity curvature bounds simultaneously. This technique has been used in a number of areas since [e.g., *Cary and Chapman*, 1988; *Kennett*, 1998]. *Anderssen et al.* [1972] extended the earlier work of *Anderssen and Senata* [1972] to include constraints on the form of the Earth models generated by MCI. They noted that the resulting set of parameter bounds obtained by MCI would be affected by the constraints imposed on the Earth model. For example, if the gradient of a density profile were constrained tightly over a

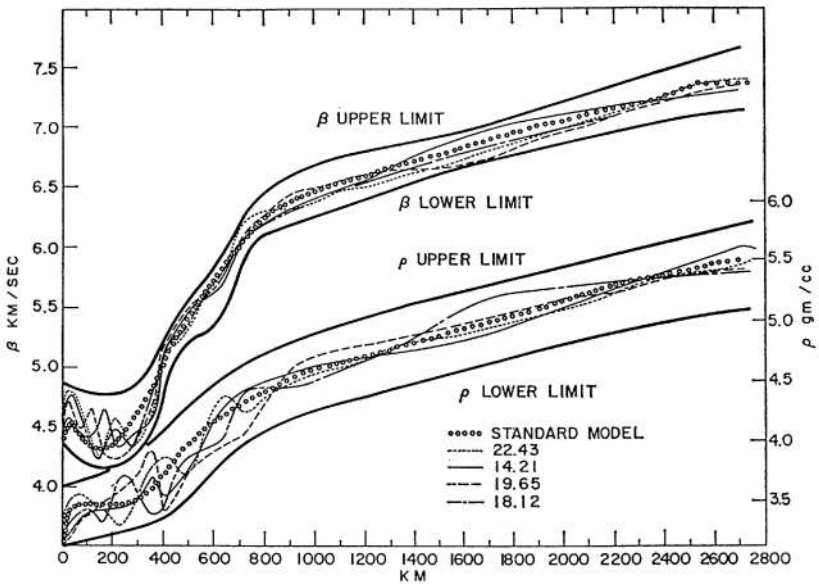

Figure 3. The six seismic and density Earth models that passed all tests shown in Figure 2 from the 5 million generated (from *Press* [1968]).

particular depth range, then this would result in relatively narrow bounds on the density, giving the false impression that the average density over the depth range was well determined. Clearly, care had to be used when interpreting MCI results obtained under smoothness constraints.

2.2. Monte Carlo Techniques Fall Out of Favor

[12] During the 1970s, attention moved away from Monte Carlo inversion and toward linear inverse problems and the use of prior information to resolve nonuniqueness (often referred to as ill posedness in linear problems) [*Wiggins*, 1972; *Jackson*, 1972; 1979]. Linear inversion techniques became popular and were applied widely (for a recent summary see *Snieder and Trampert* [1999]). Uniform Monte Carlo searching of parameter spaces was thought to be too inefficient and too inaccurate for problems involving large numbers of unknowns, for example, 50–100. (Note that in the earlier work of *Press* [1968] and *Wiggins* [1969] it was possible to increase efficiency, by testing "partial models" against subsets of the data, and thereby reject many unacceptable models early on. Figure 2 shows an outline of Press's original MCI algorithm where this is employed.) Nevertheless, uniform random search methods still found applications. In addition to the regional and global travel time studies, other applications of MCI have included electromagnetic induction [*Anderssen*, 1970], Rayleigh wave attenuation [*Mills and Fitch*, 1977; *Mills and Hales*, 1978], regional magnetotelluric studies [*Hermance and Grillot*, 1974; *Jones and Hutton*, 1979], estimation of mantle viscosities [*Ricard et al.*, 1989], and estimation of plate rotation vectors [*Jestin et al.*, 1994].

[13] An attractive feature of discrete linear inversion schemes was that estimates of resolution and model covariance could be obtained [*Franklin*, 1970; *Jordan and Franklin*, 1971; *Wiggins*, 1972]. In this case, resolution measures the degree by which model parameters can be independently determined (from each other), while model covariance measures the degree by which errors in the data propagate into uncertainty in the model parameters. Together they allow assessment of confidence bounds and trade-offs between parameters, which can be very useful in analyzing inversion results.

[14] A difficulty with linearized estimates of resolution and covariance is that they are based on local derivative approximations, evaluated about the best data fitting model, and as such can become less accurate, as the data-model relationship becomes more nonlinear. This can often lead to an underestimate of uncertainty and hence overconfidence in results. Around the same time as applications of discrete inverse theory were becoming widespread, it was shown how Monte Carlo sampling techniques could also be used to determine resolution estimates but without the need for invoking derivative approximations [*Wiggins*, 1972; *Kennett and Nolet*, 1978].

[15] The influence of nonlinearity can vary considerably between problems. For example, earthquake hypocenter location using travel times of seismic phases is often described as weakly nonlinear [see *Buland*, 1976] (although examples exist of the failure of linearization even in this case [e.g., *Billings et al.*, 1994; *Lomax et al.*, 2000]). In contrast, the estimation of seismic velocity structure from high-frequency seismic (body) wave forms can be highly nonlinear [see *Mellman*, 1980; *Cary and Chapman*, 1988]. In the latter case, subtle changes in velocity structure can significantly influence the details of the observed seismograms.

[16] Once nonlinearity is taken into account, it can be

useful to view the process of inversion in terms of optimization in a high-dimensional parameter space. Usually, some *objective function* is devised that measures the discrepancy between observables and theoretical predictions from a model. The precise nature of the optimization problem to be solved can vary considerably. For example, one might seek to minimize an objective function based solely on a measure of fit to data [e.g., *Cary and Chapman*, 1988] or one based on a linear combination of data fit and model regularization (for a discussion see *Menke* [1989]). A constrained optimization problem can be produced with the addition of (explicit) constraints on the unknowns [e.g., *Sabatier*, 1977a, 1977b, 1977c; *Parker*, 1994], or the data itself might enter only in these constraints, while the objective function represents regularization on the model. (This is often called extremal inversion; for examples, see *Jackson* [1976], *Parker* [1977], *Constable et al*. [1987] and *Parker* [1994]). In some cases these formulations are equivalent, and, in general, the most appropriate one will depend on the particular problem in hand and the questions being asked of the data.

[17] In linear problems it is common to use a quadratic criterion to measure data misfit. This leads to least squares techniques and an ellipsoidal objective function in parameter space (see Figure 4a). Most (discrete) linearized inversion techniques correspond to a gradient-based optimization algorithm, for example, steepest descent, conjugate gradients, Newton-Raphson (see *Gill et al*. [1981] and *Press et al*. [1992] for summaries]). Linearized inversion techniques can be applied to weakly nonlinear problems and in some cases highly nonlinear ones, when a good enough guess at the solution is available in advance. As the nonlinearity of the data/model relationship increases, a data misfit objective function can become more complex, and hence optimization can become more difficult. Common descriptions involve terms like "narrow valleys," "abrupt (nonsmooth) gradient changes," and multiple minima, although in truth not too many of these structures have been looked at in any detail. (See *Scales et al*. [1992] and *Billings et al*. [1994] for some rather nasty examples.)

[18] An example of a multimodal objective function arising from the residual statics problem is shown in Figure 4b [see *Deng and Scales*, 1999]. Here the data fit is measured by a *probability density function* (*PDF*) which reflects the alignment of phases. The global maximum of this function corresponds to the best data fit model, and locating it would be difficult with techniques based on gradients or matrix inversion, unless one started near the solution (in the "basin" of the global solution). In any case, for this type of objective function it might well be argued that the model corresponding to the globally maximum was of little use by itself, and inference should be based on the class of all acceptable models, if they could be found.

[19] From here on we will use the generic terms "misfit" and objective function to describe a combination of data misfit and perhaps a smoothing or regularization term, indeed any multidimensional function that might be optimized in an inversion procedure. Usually, one would minimize a *misfit function* (possibly under constraints), but as shown by Figure 4b we would seek to maximize a probability density function. (The latter arise in a probabilistic treatment of inversion problems discussed in section 3.3.)

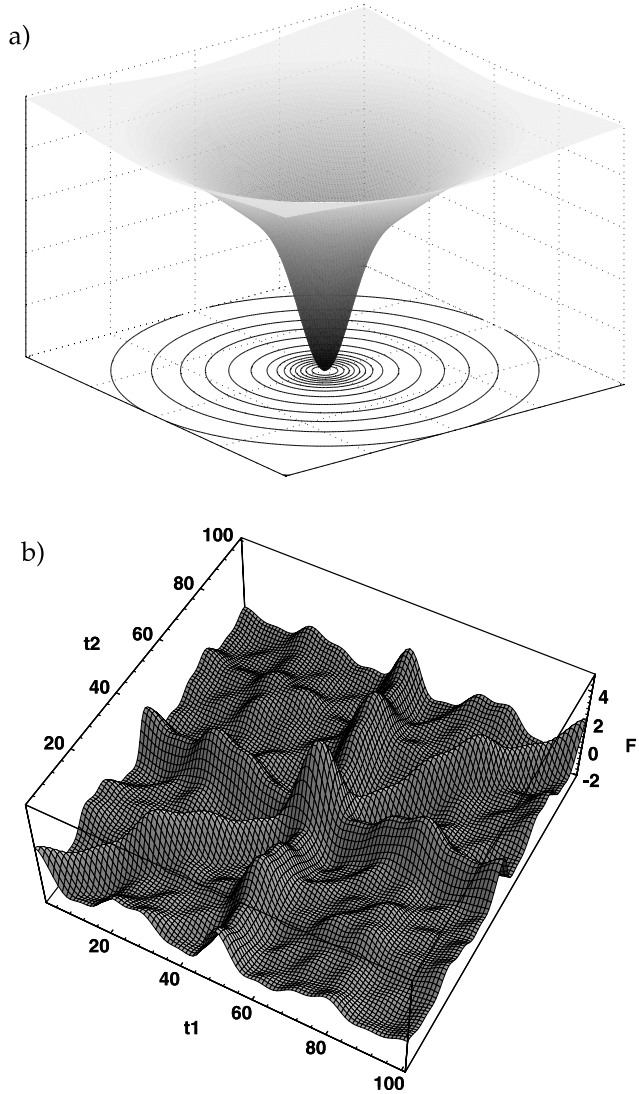

Figure 4. (a) Simple smoothly varying data misfit function with a single global minimum representing a best fit data model. Contours are projected onto the lower plane and are elliptical. (b) Probability density function for a two-dimensional residual statics problem after *Deng and Scales*, 1999. Here the global maximum represents the most likely model. Note that even in this "simple" problem, multiple maxima are observed with valleys and ridges inclined to the parameter axes.

2.3. Monte Carlo and Optimization

[20] Optimization procedures that depend on gradient approximations or matrix inversion can suffer from numerical instabilities caused by ill-conditioned matrices

or failure of convergence, etc. These situations usually arise when the objective function is highly irregular, multimodal, and nonsmooth or has discontinuities. Since Monte Carlo methods work by directly sampling a parameter space, they do not rely on the objective function being smooth in any sense, nor do they involve any potentially numerical unstable process, like matrix inversion. In this sense they are inherently stable. If the random search process is inefficient, then convergence to an optimum (or even local) solution may be slow, but it will, nevertheless, continue. With gradient based methods, there is the possibility that no solution will be found because of complete failure of the iterative procedure.

[21] As the available computational power grew in the 1980s and more sophisticated *direct search* methods became available, Monte Carlo techniques came back into vogue for geophysical inversion. At this stage our definition of a Monte Carlo technique needs an update. A more general definition (and the one we shall adopt) is a method that uses *pseudorandom* sampling to search a parameter space to retrieve Earth models or other information about the unknowns of interest. The important change is that the search no longer needs to be uniform. Random sampling from highly nonuniform multidimensional distributions is now included in our definition of a Monte Carlo method. This issue has led to some confusion in the literature. The point to remember is that Monte Carlo does not imply uniform sampling, just as random samples do not imply uniform random samples.

[22] The work of *Rothman* [1985, 1986] introduced into geophysics the technique of simulated annealing, which is a nonuniform Monte Carlo method for *global optimization*. This approach was originally developed to simulate the statistical mechanics of systems in equilibrium but quickly gained attention as a general purpose global optimization method. *Simulated annealing* was originally devised by *Kirkpatrick et al*. [1983] [see also *Geman and Geman*, 1984] and is based on the *Metropolis sampling algorithm* of *Metropolis et al*. [1953]. A reference work is that of *Aarts and Korst* [1989]. The tutorial section of this review paper deals with this topic in more detail.

[23] In the work of *Rothman* [1985, 1986], simulated annealing was applied to the large residual statics problem of exploration seismics. The impressive results stimulated many applications to other geophysical optimization problems. These include the estimation of Earth models from fitting of seismic body waveforms in exploration seismics reflection seismics and migration [*Jakobsen et al*., 1988; *Jervis et al*., 1993; *Varela et al*., 1998; *Landa et al*., 1989; *Mosegaard and Vestergaard*, 1991; *Vestergaard and Mosegaard*, 1991; *Sen and Stoffa*, 1991]; inversion of travel/arrival time data [*Pullammanappallil and Louie*, 1993, 1994]; earthquake location and rupture histories [*Billings*, 1994; *Hartzell and Liu*, 1996]; seismic source and ground motion studies [*Scherbaum et al*., 1994; *Courboulex et al*., 1996; *Ihmle and Ruegg*, 1997]; estimation of crustal structure [*Steck*, 1995; *Zhao et al*., 1996]; deconvolution of seismograms [*Courboulex et al*., 1996; *Calderon et al*., 1997]; thermodynamic modeling [*Bina*, 1998]; groundwater management and remediation [*Dougherty and Marryott*, 1991; *Kou et al*., 1992; *Rizzo and Dougherty*, 1996; *Rogers et al*., 1998]; more residual statics estimation [*Vasudevan and Wilson*, 1991; *Nørmark and Mosegaard*, 1993]; waste disposal site selection [*Muttiah et al*., 1996]; geostatistics [*Datta et al*., 1995; *Goovaerts*, 1996; *Pardo*, 1998]; seismic ray tracing [*Velis and Ulrych*, 1996]; and electromagnetic, resistivity, and magnetotelluric imaging [*Gilbert and Virieux*, 1991; *Dosso and Oldenburg*, 1991; *Dittmer and Szymanski*, 1995]. This is only a subset of the many papers making use of this technique over the past 10 years.

[24] The power of simulated annealing (SA) was that it could be used in cases where the model-data relationship was highly nonlinear and produced multimodal data misfit functions (as in Figure 4b). Simulated annealing remains one of the few approaches specifically designed for global optimization problems that has been successfully applied across several disciplines [*Aarts and Korst*, 1989]. The price of this sophistication is that control parameters (an "annealing schedule") are introduced that govern the characteristics of the sampling, and these had to be determined externally. Subsequent work by *Nulton and Salamon* [1988], *Andresen et al*. [1988], and *Hajek* [1988] showed that a theoretically optimal annealing schedule could be achieved by monitoring the progress of the algorithm and adjusting as necessary [see also *Mosegaard and Vestergaard*, 1991]. Other authors designed more simplistic but quite effective alternatives [*Szu and Hartley*, 1987; *Ingber*, 1989; *Basu and Frazer*, 1990]. Global optimization techniques were here to stay and over the following years became a popular addition to the geophysicist's tool bag.

[25] Some time after the appearance of simulated annealing, another direct search method was introduced into geophysics and applied widely. *Genetic algorithms* (GA) were originally devised as a model of adaptation in an artificial system by *Holland* [1975]. An early reference work is by *Goldberg* [1989], and a more recent tutorial is given by *Whitley* [1994]. As was the case with simulated annealing, geophysicists borrowed this technique from another discipline and used it for global optimization. Genetic algorithms fall into the class of Monte Carlo techniques because they also use random numbers to control components of the search.

[26] Genetic algorithms were first used by geophysicists in the early 1990s. A number of papers appeared in quick succession [*Stoffa and Sen*, 1991; *Gallagher et al*., 1991; *Wilson and Vasudevan*, 1991; *Smith et al*., 1992; *Sen and Stoffa*, 1992; *Sambridge and Drijkoningen*, 1992; *Scales et al*., 1992], largely in the area of seismic waveform fitting. Interestingly enough, genetic algorithms were not originally designed as function optimizers, and the range of problems to which they have been applied

is quite broad. (For reviews, see *Davis* [1987], *Goldberg* [1989], and *Gallagher and Sambridge* [1994].) Nevertheless, their main role in geophysics (as in many other disciplines) has been as a global optimization tool. Like simulated annealing, the metaphor underlying genetic algorithms is a natural optimization process, in this case biological evolution. Many variants of genetic algorithms exist (even when applied to optimization). Indeed, they are probably best viewed as a class of methods rather than as a well-defined algorithm. As with simulated annealing, some asymptotic convergence results are known for particular versions [*Davis and Principe*, 1991]. However, all versions involve control parameters, which determine the characteristics of the direct search process, and tuning them for each problem can be nontrivial.

[27] Within a few years of their introduction, genetic algorithms became quite popular within the Earth sciences and were applied in a wide range of areas. Some examples include earthquake hypocenter location [*Kennett and Sambridge*, 1992; *Sambridge and Gallagher*, 1993; *Billings et al.*, 1994; *Wan et al.*, 1997; *Muramatsu and Nakanishi*, 1997]; estimation of focal mechanisms and seismic source characteristics [*Kobayashi and Nakanishi*, 1994; *Zhou et al.*, 1995a; *Sileny*, 1998; *Yu et al.*, 1998]; mantle viscosity estimation [*King*, 1995; *Cadek et al.*, 1998; *Kido et al.*, 1998]; groundwater monitoring and management problems [*McKinney and Lin*, 1994; *Ritzel et al.*, 1994; *Cieniawski et al.*, 1995; *Rogers et al.*, 1995; *Tang and Mays*, 1998]; meteorite classification [*Conway and Bland*, 1998]; seismic anisotropy estimation [*Horne and Macbeth*, 1994; *Levin and Park*, 1997]; near-source seismic structure [*Zhou et al.*, 1995b]; regional, crustal seismic structure and surface wave studies [*Lomax and Snieder*, 1994, 1995a, 1995b; *Drijkoningen and White*, 1995; *Yamanaka and Ishida*, 1996; *Neves et al.*, 1996]; design of microseismic networks [*Jones et al.*, 1994]; fission track dating [*Gallagher*, 1995]; seismic profiling and migration [*Jervis et al.*, 1993; *Jin and Madariaga*, 1993; *Nolte and Frazer*, 1994; *Horne and Macbeth*, 1994; *Boschetti et al.*, 1996; *Docherty et al.*, 1997]; seismic receiver functions studies [*Shibutani et al.*, 1996]; problems in geotectonics [*Simpson and Priest*, 1993]; magnetotelluric inversion [*Everett and Schultz*, 1993]; inversion of potential fields [*Boschetti et al.*, 1997]; conditioning of linear systems of equations [*Curtis and Snieder*, 1997]; seismic ray tracing [*Sadeghi et al.*, 1999]; there are many others. Some studies have involved devising variants of the basic approach and adapting them to the characteristics of individual problems [e.g., *Sambridge and Gallagher*, 1993; *Koper et al.*, 1999].

[28] The question as to whether simulated annealing or genetic algorithms perform better for a particular problem (i.e., more efficiently, more likely to find acceptable or even optimal models, etc.) has been addressed by a number of authors, both within the Earth sciences and elsewhere [see *Scales et al.*, 1992; *Ingber and Rosen*, 1992; *Sen and Stoffa*, 1995; *Horne and Macbeth*, 1998]. Most commonly, these studies compare performance on particular optimization problems, and from these it is difficult to draw general conclusions. Quite clearly, their relative performance varies between applications and also with the particular versions of each method that are being compared. For a recent, very readable, discussion of the types of optimization problem for which they are each suited, see *Gershenfeld* [1999].

[29] A few other global optimization techniques (again originating in other fields) have made fleeting appearances in the geophysical literature. Two notable examples are evolutionary programming [*Minster et al.*, 1995] and Tabu (or Taboo) search [*Cvijović and Klinowski*, 1995; *Vinther and Mosegaard*, 1996; *Zheng and Wang*, 1996]. The former is related to genetic algorithms but was developed quite independently [*Fogel*, 1962; *Fogel et al.*, 1966]. Again, the primary motivation was not optimization but the simulation of complex adaptive systems (see *Gell-Mann* [1994] for a popular discussion). Tabu search is not strictly speaking a Monte Carlo method since it does not make use of random numbers, but it is able to climb out of local minima in misfit functions [*Cvijović and Klinowski*, 1995; *Vinther and Mosegaard*, 1996]. Very recently, a new Monte Carlo direct search technique known as a neighbourhood algorithm (NA) has been proposed, this time developed specifically for sampling in geophysical inverse problems [*Sambridge*, 1999a]. The approach makes use of concepts from the growing field of computational geometry and bares little resemblance to either genetic algorithms or simulated annealing. It is difficult to say much about these approaches with any confidence as experience with geophysical problems is still rather limited.

2.4. Ensemble Inference Rather Than Optimization

[30] The renewed interest in Monte Carlo techniques for global optimization and exploration raised a familiar question, that is, how to make use of the sampling they produced to assess trade-offs, constraints and resolution, in multimodal nonlinear problems. Put another way, how can one use the collection of Earth models generated by a Monte Carlo procedure to do more than estimate a set of "best fitting" parameters. This was, in effect, a return to the questions posed by the first users of Monte Carlo; however, the response adopted by the second generation of practitioners was to take a Bayesian approach. This statistical treatment of inverse problems became well known to geophysicists through the work of *Tarantola and Valette* [1982a, 1982b; see also *Tarantola*, 1987] and had been applied extensively to linearized problems. Monte Carlo techniques allowed an extension of the Bayesian philosophy to nonlinear problems.

[31] *Bayesian inference* is named after *Bayes* [1763], who presented a method for combining prior information on a model with the information from new data. In this formulation of an inverse problem all information is

represented in probabilistic terms (i.e., degrees of belief). Bayesian inference is reasonably general in that it can be applied to linear or nonlinear problems. (It is dealt with in more detail in the tutorial section of this paper.) In short, it combines the prior information known on the model, with the observed data, and produces the *posterior probability density function* (PDF) on the model parameters, which is taken as the "complete solution" to the inverse problem. (Standard references are by *Box and Tiao* [1973], and useful recent books are by *Smith* [1991] and *Gelman et al*. [1995]. Summaries within a geophysical context are given by *Duijndam* [1988a, 1988b] and *Mosegaard and Tarantola* [1995].) The Bayesian approach is not without its criticisms. For example, implicit in its formulation is that one must know the statistical character of all error, or noise, processes in the problem, which can be very difficult in some cases, especially when the theoretical predictions from a model involve approximations. In addition, it is a controversial issue as to whether prior information can be adequately represented probabilistically (see *Scales and Snieder* [1997] and *Gouveia and Scales* [1998] for a discussion). (Note that probabilistic prior information is often called "soft" and differs from strict inequalities on the model parameters, which are referred to as "hard" prior information [see *Backus*, 1988; *Stark*, 1992]).

[32] In a Bayesian approach the *posterior PDF* spans the entire model space. The case where it is a Gaussian can be dealt with effectively using linearized inversion techniques [see *Tarantola*, 1987; *Menke*, 1989]. Linearized techniques use local curvature information on the PDF about its maximum to estimate resolution and trade-offs. If the PDF is a Gaussian, then the local curvature defines the complete structure of the PDF in parameter space. For highly nonlinear problems the posterior PDF can have a complex multimodal shape, arising from the nature of the data fit (*likelihood function*) or perhaps from the inclusion of complex prior information. In this case, global optimization techniques are needed to identify the maximum of the posterior probability density; however, as the complexity of the PDF increases, a single "most probable" model (if one exists) has little meaning (see Figure 4b). Even if one could be found, a linearized treatment of resolution problem would be of little value (essentially because the local information on the PDF's curvature is not representative of the PDF as a whole). In these cases, information on the complete shape of the posterior is needed to produce Bayesian measures of uncertainty and resolution. It is here that Monte Carlo methods have major advantages over linearized (local) methods, since the sampling they produce can be used to calculate Bayesian integrals. Within a Bayesian context then, the emphasis is less on optimization and more on sampling the most probable regions of parameter space as determined by the posterior PDF, a process known as *importance sampling*. (Compare this to the early MCI work where the emphasis was on exploring the acceptable regions of parameter space, as defined by data and prior constraints.)

[33] *Monte Carlo integration* of multidimensional probability distributions is an active area of research in computational statistics (for summaries see *Flournay and Tsutakawa* [1989], *Smith* [1991], *Smith and Roberts* [1993], and *Gelman et al*. [1995]). Over the past 10 years, geophysicists have begun to use *Markov Chain Monte Carlo* (*MCMC*) methods, which directly simulate the posterior PDF, that is, draw random samples distributed according to the posterior PDF, and from these calculate Bayesian estimates of constraint and resolution [see *Koren et al*., 1991; *Mosegaard and Tarantola*, 1995; *Gallagher et al*., 1997; *Gouveia and Scales*, 1998]. It is not surprising that many of these studies arise in seismology and in particular the estimation of Earth structure from high-frequency body waveforms, especially in exploration studies. This is an area where complex multimodal and multidimensional PDFs can result from the discrepancies between observed and predicted seismograms. An example is shown in Figure 4b, which comes from the work of *Deng and Scales* [1999].

[34] At the end of the 1990s, Monte Carlo integration and importance sampling have become firmly established as the technique of choice for Bayesian inversions in nonlinear problems. Debate continues over whether the Bayesian paradigm is appropriate in many cases (see *Scales and Snieder* [1997] for a discussion). However, Monte Carlo (adaptive or nonuniform) sampling of parameter spaces has also remained popular in studies that do not invoke the Bayesian philosophy. (Many of the papers cited above fall into this category.) In these cases the issues of mapping out and characterizing the class of acceptable models remain just as relevant today as when the original hedgehog algorithm was proposed more than 30 years ago [*Keilis-Borok and Yanovskaya*, 1967; *Valius*, 1968]. Characterizing the properties of all acceptable models, or an obtained finite ensemble, has been a central issue for many authors, and a variety of algorithms have been proposed (see *Constable et al*. [1987], *Vasco et al*. [1993], *Lomax and Snieder* [1994, 1995b] *Douma et al*. [1996], and *Sambridge* [1999b] for examples). This is perhaps the factor that most clearly distinguishes a study of inverse problems from that of parameter estimation.

[35] With the growth and spread of high-performance computing, Monte Carlo inversion techniques are no longer restricted to the owners of supercomputers. As their use becomes more widespread, we can expect that direct sampling of a parameter space will become routine for nonlinear problems, and the need for linearization will diminish in many cases. (This is arguably already the case for problems with relatively few unknowns, e.g., earthquake hypocenter location.) Also, one might expect that larger scale problems (i.e., involving many more unknowns) will increasingly be tackled using Monte Carlo techniques, within either a Bayesian or non-Bayesian framework. For the foreseeable future

very large scale nonlinear problems, like 3-D mantle seismic tomography, are likely to remain beyond the range of Monte Carlo techniques; however, it is worth noting that a Monte Carlo technique has already been applied to a 2-D borehole tomography problem (involving fewer unknowns than mantle tomography, but often more nonlinear) [*Vasco et al.*, 1993]. As Monte Carlo sampling is better understood and becomes more accessible, it seems likely that it will become an increasingly useful tool for nonlinear inversion. It is hoped that this paper, and in particular the following tutorial section, will add to that awareness and encourage students and researchers to think about it themselves.

3. MONTE CARLO METHODS: THE TECHNOLOGY OF INFORMATION

[36] In the next section we outline some of the main Monte Carlo approaches that have been used to tackle geophysical inverse problems. We describe some of the basic concepts and provide a source of references for further reading. Some open questions are also highlighted. At all times we assume that we have some criterion, ϕ, which measures the discrepancy between observations and predictions and perhaps includes some other information. Its evaluation for any given model, x, constitutes a solution to the *forward problem*. In some cases we may be interested in optimizing this objective function; in others we may be more interested in sampling it adequately enough to either evaluate Bayesian information measures (as discussed in section 3.3), or estimate properties of the data acceptable models that fit within our chosen (usually finite dimensional) parameter space.

[37] Several of the approaches discussed here are commonly associated with a Bayesian approach to inversion. However, it is worth noting that they can also be employed independently of a Bayesian inversion, that is, simply to perform a direct search of a parameter space.

3.1. Preliminaries

3.1.1. Linearization or Monte Carlo?

[38] The first question one needs to address is whether a Monte Carlo technique (like SA, GA, NA, etc.) or a linearized approach (based on matrix inversion) would be most appropriate for a particular problem. The answer depends on the nature of the data-model relationship, the number of unknowns, and, to a lesser extent, the computational resources available.

[39] As the data-model relationship becomes more complex, the misfit function (or PDF) will also increase in complexity (e.g., multimodal, etc.), and Monte Carlo techniques will be favored over linearized techniques for two reasons. The first is that they are more numerically stable in the optimization/parameter search stage. This is because they do not rely on the convergence of sequence of model perturbations (like a linearized approach) and at the same time avoid the need for matrix inversion.

[40] The second reason for favoring Monte Carlo techniques is that they are usually more reliable in appraising the solution, that is, estimating uncertainty by means of model covariance and resolution matrices (see section 3.3). This is (again) because they avoid derivatives and hence the numerical approximations on which linearized estimates of model covariance and resolution are based [see *Menke*, 1989]. Linearized techniques are prone to underestimate uncertainty when nonlinearity is severe. Also, a direct search of the parameter space may indicate significant trade-offs and even multiple classes of solution, which would not be found using linearization.

[41] Unfortunately, it is not possible to know whether linearized estimates of model covariance and resolution are accurate until a fully nonlinear calculation has been performed. The same is true for the optimization process itself, that is, whether linearized techniques are likely to be unstable or require heavy damping to converge, problems that could well be alleviated using a direct search technique.

[42] In some cases, for example, in many waveform inversion studies encountered in seismic exploration, the data-model relationship can become so complex that fully nonlinear direct search techniques are the only viable approach. At the opposite end of the scale, with discrete linear, or linearized, problems with relatively few unknowns (e.g., 10–50), it is often overlooked that Monte Carlo techniques can be both very convenient and efficient.

[43] It is also worth pointing out that linearization is not always possible or practical in some cases. This is the case when the observables are not differentiable functions of the unknowns. An example is when the forward problem involves a complex calculation such as the numerical modeling of landscape evolution in response to tectonic and erosional processes [*van der Beek and Braun*, 1999; *Braun*, 2002]. In this case the unknowns are the rate of tectonic uplift and parameters that relate rate of surface processes to geometrical features like drainage area and surface slope, while the observables are geochronological constraints on exhumation rate and morphological properties of the landscape. In these problems, there is no analytical relationship between unknowns and observables, and hence linearized techniques are not appropriate. However, direct search techniques can still be applied because they only require the ability to solve the forward problem. Furthermore, all of the direct search algorithms presented in section 3.2 can take advantage of parallel computation because each forward solution can be performed independently. It seems likely that this is an area where Monte Carlo techniques will find more applications in the future.

[44] It is important to stress that Monte Carlo techniques are not a panacea for geophysical inversion. They

are only applicable to discretized problems, that is, ones where a finite parameterization has been chosen, and as with all discrete inversion approaches, the results will inevitably be dependent on the suitability of that choice. Also, it is clear that as the number of unknowns become large (say, greater than a few hundred) then direct search techniques become impractical because of the computation involved. The actual limiting dimension will vary considerably between applications because it depends on the computational cost of solving the forward problem. However, it is also clear that as computing power continues to grow, so will the range of problems that can be addressed with Monte Carlo techniques.

3.1.2. Which Monte Carlo Technique?

[45] The choice between the competing flavors of Monte Carlo technique is much less clear than whether one should be used in the first place. In general, there appears to be no preferred method of choice in all circumstances. In the next few sections we discuss the basic mechanics of different classes of Monte Carlo approach and make some comparisons. Here we make a few general observations which may help in deciding on which Monte Carlo technique to choose.

[46] In cases where the cost of the forward modeling is not excessive and the number of unknowns is small (say <10), a simple deterministic grid search [e.g., *Sambridge and Kennett*, 1986] may be practical (see section 3.2.2). This would have the attraction of being both reliable (guaranteeing a global minimum on the chosen grid) and useful for uncertainty analysis because the samples are uniformly distributed and hence produce an unbiased sample of the parameter space. With modern computing power this most simplistic of techniques can become surprisingly efficient for some problems. Of course, grid search techniques become impractical when either the number of unknowns or the computational cost of the forward problem is high, and one must turn to the more sophisticated irregular parameter space sampling methods.

[47] It is beyond the scope of this paper to enter the (perhaps never-ending) argument between Bayesian and non-Bayesian approaches to inversion [*Scales and Snieder*, 1997]. However, it should be noted that a Bayesian inversion is often implemented with a Markov chain Monte Carlo (MCMC) approach (section 3.2.3.1), for example, using the Metropolis algorithm. This, in turn, is closely related to the optimization technique simulated annealing (section 3.2.3), and so, in general, if a Bayesian approach were preferred, then these would be the natural algorithms of choice for both optimization and uncertainty estimation (see section 3.2.3.1 for a discussion).

[48] Several authors have argued for genetic algorithms (section 3.2.4) as a powerful *parameter space search* technique (see above), although the ensemble of parameter space samples produced by a genetic algorithm does not (in general) follow any prescribed probability distribution and so cannot be used directly for a quantitative uncertainty analysis (within a Bayesian framework) (see section 3.4). The neighbourhood algorithm (section 3.2.5), which is both a search and appraisal technique, offers a potential solution to this problem. Ultimately, the choice between Monte Carlo direct search techniques will often depend as much on issues of practical convenience, like the availability of suitable computer software, as the precise details of the algorithms.

3.1.3. Generating Random Samples

3.1.3.1. Pseudorandom Deviates

[49] All Monte Carlo techniques make use of random number generators of some kind. This is the case even when the ultimate task is to generate multidimensional random deviates distributed according to complicated PDFs, for example, with the *Metropolis-Hastings algorithm* (see section 3.2.3.1). The common approach is to generate *pseudorandom numbers* using a linear or multiplicative congruent method. For a survey of theory and methods, see *Park and Miller* [1988], and for descriptions of "workhorse" techniques, see *Press et al*. [1992]. Figure 5a shows 1000 pairs of pseudorandom numbers plotted as points in the plane.

3.1.3.2. Quasi-Random Sequences

[50] It is a nonobvious point that not all random number generators are equally uniform. One way this can be measured is in how fast the error in a Monte Carlo integration decreases as a function of the number of samples. For example, if one were to calculate the average of the x or y values of the pseudorandom points in Figure 5a, then (assuming the sides of the box were 0 to 1) the rate at which the estimate approaches 0.5 would decrease linearly proportional to $1/\sqrt{N}$, where N is the number of points. However, it is possible to do much better by using quasi-random (or subrandom) sequences. Figure 5b shows an example of $N = 1000$ points generated from two quasi sequences. The quasi sequence generates a set of points that are (in a well-defined sense) "maximally self-avoiding." To the eye this looks like a regular grid with the vertices randomly perturbed. It is clear from Figure 5b that in this case the *quasi random numbers* are more evenly distributed than the pseudorandom deviates. It can be shown that as a consequence they result in more rapid convergence of Monte Carlo integrals. For example, if we calculated the average of the x and y values from the points in Figure 5b, then the numerical error would decrease proportional to $1/N$, which is much faster than for the pseudorandom case. As we show in section 3.3, the evaluation of MC integrals is central to Bayesian inference.

[51] The Sobol-Antonov-Saleev approach [*Sobol*, 1967; *Antonov and Saleev*, 1979] is an efficient way to generate quasi-random sequences and has been imple-

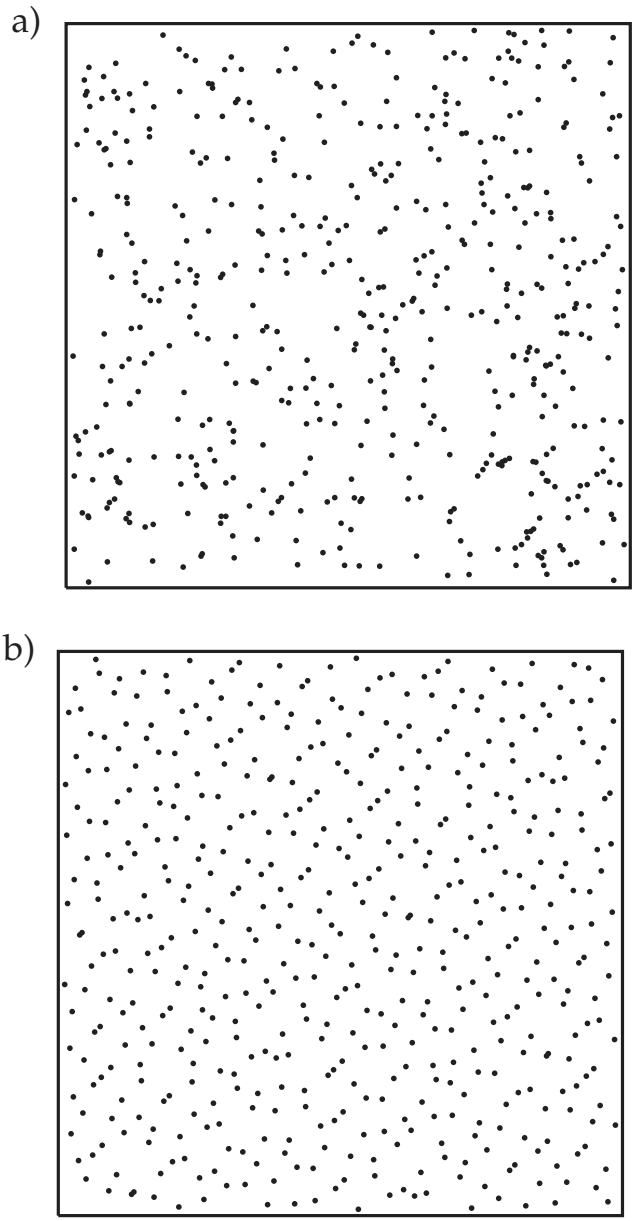

Figure 5. (a) The 1OOO uniform points generated with a pseudorandom number generator, showing typical clustering and uneveness. (b) The 1OOO uniform points generated with two sequences of quasi-random numbers. Note the considerably more uniform appearance and even density compared to Figure 5a.

mented in a convenient form by *Press et al.* [1992]. However, care must be taken in using quasi-random numbers, especially when they form the components of a vector (i.e., Earth model) in a multidimensional space. In this case each component in the (quasi-random) vector must be generated from an independent quasi sequence. The main problem is that some components of multidimensional quasi-random vectors can have a high degree of correlation. It is only when sufficiently many quasi vectors are generated that the correlation disappears. In dimensions as low as 10, it may require many thousands of samples before the overall correlation between all components is negligible, and hence the samples are usable. *Morokoff and Caflisch* [1994] describe these issues in detail. Quasi-random sequences are used in the neighbourhood algorithm of *Sambridge* [1999a] (see also section 3.2.5) and are finding increasing numbers of applications in multidimensional numerical integration [*Morokoff and Caflisch*, 1994].

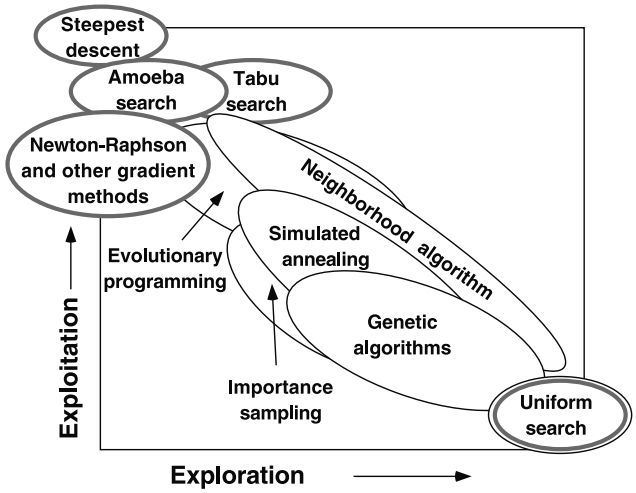

Figure 6. A schematic representation of various search/optimization algorithms in terms of the degrees to which they explore the parameter space and exploit information. Shaded borders indicate a deterministic (non–Monte Carlo) method. Uniform search includes the deterministic grid search.

3.2. Searching a Parameter Space With Monte Carlo Methods

3.2.1. Exploration Versus Exploitation

[52] A useful framework for comparing different search algorithms, random or not, is in terms of a tradeoff between exploration and exploitation. By exploration we mean trying to improve the objective function by looking (randomly) in different regions of parameter space, without regard for what we have already learned from previous sampling. Exploitation is the opposite; we make decisions of where to sample next by only using the previous sampling and sometimes just the current best fit model. Most direct search Monte Carlo algorithms fall somewhere in between the extremes. Figure 6 shows our attempt to classify various techniques according to these definitions.

[53] From the point of view of optimization the rule of thumb is that the more explorative an algorithm is, the less likely it will fall into local minima, but the less efficient it will be at converging on a solution. Conversely, the exploitative algorithms will be more efficient at convergence but prone to entrapment, and hence the final result will depend on the starting point. Examples of methods that lie at the extremes would be a uniform random search, which is completely explorative, and a steepest descent algorithm [*Gill et al.*, 1981], which is

purely exploitative. Clearly, the most appropriate technique will depend on the nature of the problem. For smoothly varying near quadratic objective functions we would prefer an exploitative approach, which allows rapid converge, for example, a Newton-type descent method. For highly nonlinear problems with multiple minima/maxima in the objective function a combination of exploration and exploitation would probably suit best. However, controlling the trade-off between the two properties is often quite difficult, as is deciding in advance which approach is best suited to a particular problem.

3.2.2. Uniform Sampling

[54] The simplest form of randomized sampling of a parameter space is uniform sampling. For a problem with d distinct unknowns the ith random sample is the vector $\mathbf{x}_i$,

$$\mathbf{x}_i = \sum_{i=1}^{d} r_i \mathbf{e_i}, \tag{1}$$

where r_i is a (0,1) uniform random deviate and $\mathbf{e_i}$ is the unit vector along the ith axis in parameter space. For each new sample some data fit, or other objective function, ϕ, must be evaluated, and hence forward modeling performed. As discussed above, this was the first fully nonlinear approach used by geophysicists more than 30 years ago. Note that, by definition, uniform sampling is not biased toward any particular region of parameter space, and there is hence no possibility of entrapment in local minima of ϕ. Equally, however, it does not concentrate sampling, and, even with modern supercomputers, it is usually inefficient once the number of unknowns becomes greater than ~10. This is the so-called "curse of dimensionality." For example, if we imagine the parameter space filled by a regular multidimensional Cartesian grid with $(n_k - 1)$ intervals per axis, then the number of distinct nodes (and hence models) in this grid is n_k^d, which can become enormous, even in relatively small dimensions.

[55] In practice, one always undersamples a parameter space. In many Monte Carlo studies the total number of samples generated is much smaller than the number of vertices of a single "unit cube" in a Cartesian grid, that is, 2^d, and in this sense one always tends to undersample parameter spaces in practice. For most problems the only viable approach is for the MC search algorithm to concentrate sampling in particular "promising" regions of parameter space, that is, adapt itself to the objective function ϕ. One area where uniform MC search has continued to be useful is in sampling under hard prior constraints on the unknowns. An example appears in the work of *Wiggins* [1972] [see also *Cary and Chapman*, 1998; *Kennett*, 1988].

3.2.3. Simulated Annealing

[56] The simulated annealing method exploits a statistical mechanical analogy to search for the global minimum of an objective function ϕ possessing a large number of secondary minima. The algorithm simulates the process of chemical annealing in which a melted crystalline material is cooled slowly through its freezing point, thereby approximately settling into its energy ground state. By identifying the objective function with the energy of the crystalline material and by appropriate definition of a temperature parameter for the simulations, it is possible to simulate a "cooling" of the system to be optimized. A sufficiently slow cooling of this system will, by analogy to the chemical annealing, result in convergence to a near-optimal configuration, characterized by a near-minimal value of the objective function.

[57] Simulated annealing is based on the Metropolis-Hastings algorithm or the Gibbs Sampler, and we shall therefore take a closer look at these algorithms here. Later, we shall also see why these algorithms are the workhorses in Bayesian Monte Carlo calculations.

3.2.3.1. Markov Chain Monte Carlo: Metropolis, Hastings, and Gibbs

[58] The idea behind the Metropolis-Hastings algorithm and the Gibbs Sampler is the same. They are both so-called Markov Chain Monte Carlo algorithms designed to generate samples of a probability distribution p over a high-dimensional space $\mathcal{M}$ under the special difficulty that no explicit mathematical expression exists for p. Only an algorithm that allows us to calculate the values of p at a given point in the space is available. This is a typical situation in geophysics where p is a probability density derived from a misfit function ϕ through, for example,

$$p(\mathbf{m}_k) = A \cdot \exp\left(-B\phi(\mathbf{m}_k)\right), \tag{2}$$

where $\mathbf{m}_k$ is a model and A and B are constants. Very often $p\,(\mathbf{m}_k)$ can only be evaluated for a particular Earth model through a very computer-intensive calculation.

[59] Let us consider the mechanics of the Metropolis-Hastings algorithm, for simplicity, we consider a situation where we wish to sample a probability distribution p in a discretized model space $\mathcal{M}$. Sampling from the distribution p means that the probability of visiting model $\mathbf{m}$ is proportional to $p(\mathbf{m})$. To generate a simple algorithm that samples p, we can make the following assumptions:

1. The probability of visiting a point $\mathbf{m}_i$ in model space, given that the algorithm currently is at point $\mathbf{m}_j$, depends only on $\mathbf{m}_j$ and not on previously visited points. This is the so-called *Markov* property. This property means the algorithm is completely described by a transition probability matrix P_{ij} whose ijth component is the conditional probability of going to point $\mathbf{m}_i$, given the algorithm currently visits $\mathbf{m}_j$.

2. For all points $\mathbf{m}_j$ in $\mathcal{M}$, there are exactly N points $\mathbf{m}_i$,

including $\mathbf{m}_j$ itself, for which P_{ij} is nonzero. If this property holds, we say that the algorithm is regular, and the set of N accessible points constitutes what we call the neighborhood $\mathcal{N}_j$ of $\mathbf{m}_j$.

3. It is possible for the algorithm to go from any point $\mathbf{m}_j$ to any other point $\mathbf{m}_i$, given enough steps. An algorithm satisfying this property is called irreducible.

[60] The question now is, which transition probability matrix P_{ij} leads to an algorithm that samples the target distribution p? The answer is that there exists infinitely many such transition probability matrices. We will choose one that is simple, namely, one that satisfies microscopic reversibility:

$$P_{ij}p(\mathbf{m}_j) = P_{ji}p(\mathbf{m}_i)\,. \tag{3}$$

It is easy to see that any choice of P_{ij} satisfying microscopic reversibility will continue to sample p once it has already started doing so. This is a simple consequence of the fact that if the algorithm has probability $p(\mathbf{m}_j)$ of visiting $\mathbf{m}_j$, then the probability that a transition from $\mathbf{m}_j$ to $\mathbf{m}_i$ takes place in a given iteration is $P_{ij}p(\mathbf{m}_j)$. Similarly, the probability that a transition from $\mathbf{m}_i$ to $\mathbf{m}_j$ takes place in a given iteration is $P_{ji}p(\mathbf{m}_i)$. Microscopic reversibility means that the probability of these two transitions is the same at all times, and the fact that each pair of points in $\mathcal{M}$ maintains mutual equilibrium in this way means that there is overall equilibrium sampling at the target distribution p. For more details see, for example, *Mosegaard and Tarantola* [1995] or *Mosegaard and Sambridge* [2002].

[61] In the Metropolis-Hastings algorithm the transition probabilities P_{ij} are given by

$$P_{ij} = \frac{1}{N}\min\left(1, \frac{p(\mathbf{m}_i)}{p(\mathbf{m}_j)}\right), \tag{4}$$

and hence satisfy equation (3). This is, in practice, realized in the following way.

[62] Assume that the current point visited by the algorithm is $\mathbf{m}_j$. We now choose (or, rather, propose) one of its N neighbors $\mathbf{m}_i$ with probability

$$P_{\text{proposal}} = 1/N\,. \tag{5}$$

Finally, we accept $\mathbf{m}_i$ only with probability

$$P_{\text{accept}} = \min\left(1, \frac{p(\mathbf{m}_i)}{p(\mathbf{m}_j)}\right). \tag{6}$$

If $\mathbf{m}_i$ is accepted, the algorithm goes to $\mathbf{m}_i$ in this iteration, but if $\mathbf{m}_i$ is rejected, the algorithm stays in $\mathbf{m}_j$ ($\mathbf{m}_j$ is visited once again). It is clear that the above algorithm has transition probabilities given by equation (4).

[63] The remaining question of how to make the algorithm sample p in the first place is more complicated, but fortunately it can be shown [see, e.g., *Tierney* 1994] that the distribution of samples produced by any algorithm satisfying our requirements will indeed converge toward p when the number of iterations goes to infinity. The problem of estimating the speed of convergence is presently unresolved, but some practical advice on how to run this type of algorithm was found by *Hastings* [1970] and *Mosegaard* [1998].

[64] According to the mechanics of the Gibbs Sampler in a typical implementation, operating in a K-dimensional model space, each iteration consists of K substeps, one for each parameter. The kth substep perturbs only the kth parameter, and it has its own transition probability matrix P_{ij}^k. The neighborhood $\mathcal{N}_j^k$ of a point $\mathbf{m}_j$ consists of all points $\mathbf{m}_i$ that differ from $\mathbf{m}_j$ in only the kth parameter. This means, in particular, that the neighborhoods of two points are either identical or disjoint.

[65] Assume again that the current point visited by the algorithm is $\mathbf{m}_j$. We now run K steps, and in each step, instead of proposing one of its N neighbors uniformly at random, we choose $\mathbf{m}_i$ from the neighborhood $\mathcal{N}_j^k$ according to the conditional probability distribution

$$p(\mathbf{m}_i|\mathcal{N}_j^k) = \frac{p(\mathbf{m}_i)}{\sum_{\mathbf{m}_k \in \mathcal{N}_j^k} p(\mathbf{m}_k)}, \tag{7}$$

without any accept/reject probability (i.e., the chosen $\mathbf{m}_i$ is always accepted). In each step, only one parameter is perturbed (or is possibly left unchanged), so after completion of one iteration (consisting of all K steps), all parameters have been perturbed.

[66] That this algorithm samples p is easy to see. The transition probability matrix for each step is given by

$$P_{ij}^k = p(\mathbf{m}_i|\mathcal{N}_j^k), \tag{8}$$

which clearly satisfies microscopic reversibility, equation (3), so if $\mathbf{m}_i \in \mathcal{N}_j^k$, then

$$\begin{aligned} P_{ij}^k p(\mathbf{m}_j) &= p(\mathbf{m}_i|\mathcal{N}_j^k p(\mathbf{m}_j) \\ &= \frac{p(\mathbf{m}_i)}{\sum_{\mathbf{m}_k \in \mathcal{N}_j^k} p(\mathbf{m}_k)} p(\mathbf{m}_j) \\ &= \frac{\mathrm{p}(\mathbf{m}_j)}{\sum_{\mathbf{m}_k \in \mathcal{N}_i^k} p(\mathbf{m}_k)} p(\mathbf{m}_i) \\ &= p(\mathbf{m}_j|\mathcal{N}_i^k)p(\mathbf{m}_i) \\ &= P_{ji}^k p(\mathbf{m}_i)\,, \end{aligned}$$

where we have used that the neighborhoods $\mathcal{N}_j^k$ and $\mathcal{N}_i^k$ are identical. Since each step of an iteration satisfies microscopic reversibility, so does the entire iteration, and the algorithm samples the target distribution p asymptotically.

[67] Whether the Metropolis-Hastings algorithm or the Gibbs Sampler is the most efficient depends on the problem at hand. The Gibbs Sampler takes much fewer iterations before it samples p correctly, but the many

steps needed to perform an iteration may render it inefficient for problems where evaluation of p is computer-intensive. *Rothman* [1986] gives an example of the use of a Gibbs Sampler in a case where p can be efficiently evaluated for all perturbations of a single model parameter.

3.2.3.2. Simulated Annealing Algorithm

[68] It is an empirical fact that the process of chemical annealing, where a crystalline material is slowly cooled through its melting point, results in formation of highly ordered, low-energy crystals. The slower the cooling, the more perfect is the crystal growth, and the lower is the lattice energy. This process can be viewed as a "physical optimization method" in which the objective function is the lattice energy E. The process can be simulated in large numerical systems by identifying the model parameters of the system with state space variables and the objective function of the optimization problem with the energy E. In each step of the algorithm, thermal fluctuations in the system are simulated by randomly perturbing model parameters, and the fluctuations are controlled by a temperature parameter T. The simulated annealing algorithm [*Kirkpatrick et al.*, 1983] runs as follows: In each step a random perturbation of the model parameters $\mathbf{m}_j$ of the numerical system is attempted. The new set of model parameters $\mathbf{m}_i$ are accepted if the value of the objective function E decreases. However, if E increases, the new parameters may be accepted with probability

$$P_{\text{accept}} = \exp\left(-\frac{\Delta E}{T}\right), \tag{9}$$

where ΔE is the change in the objective function and T is the temperature parameter. If the new model is rejected, a new perturbation is attempted in the next move, and the above process of decision is repeated.

[69] A close inspection of the above algorithm reveals that for constant temperature parameter T it is actually a Metropolis-Hastings algorithm designed to sample the probability distribution [*Metropolis et al.*, 1953],

$$P_B(\mathbf{m}) = \frac{\exp\left(-\frac{E(\mathbf{m})}{T}\right)}{Z(T)}, \tag{10}$$

which is known in statistical physics as the Gibbs-Boltzmann distribution. Here $1/Z(T)$ is a normalization constant. In simulated annealing, however, the temperature parameter is gradually decreased from a high value, allowing large "thermal" fluctuations, down to zero, where only decreasing values of the objective function are allowed. For decreasing temperature T the Gibbs-Boltzmann distribution converges toward a distribution having all its probability mass in the global minimum for E. In other words, as the temperature gradually approaches zero, the probability that our system is close to the global minimum for its objective function approaches 1.

[70] In simulated annealing the Gibbs-Boltzmann distribution can also be sampled by means of a Gibbs sampler. *Rothman* [1986] gives an example where the residual statics problem of reflection seismology is solved in this way.

3.2.3.3. Nulton-Salamon Annealing Schedule

[71] In practice, one has to face the fact that the above theory is only true for an infinitely slow "cooling." The problem is therefore to find a way to decrease the temperature in a finite number of steps, such that the probability that the system is close to the global minimum for E after the annealing simulation is maximum.

[72] *Nulton and Salamon* [1988] devised a near-optimal annealing method that keeps the numerical system close to "thermal equilibrium" at all times. The actual mean value $\langle E \rangle$ of the objective function is kept at a constant distance

$$v = \frac{\langle E \rangle - \langle E \rangle_{eq}}{\sigma_E(T)} \tag{11}$$

from the never realized equilibrium mean value $\langle E \rangle_{eq}$. Here $\sigma_E(T)$ is the standard deviation of the fluctuating objective function, and the distance v is known as the "thermodynamic speed." The authors arrive at the following differential equation for the annealing temperature schedule $T(t)$:

$$\frac{dT}{dt} = -\frac{vT}{\epsilon(T)\sqrt{C(T)}}, \tag{12}$$

where $C(T)$ and $\epsilon(T)$ are the heat capacity and the relaxation time of the system, respectively. Estimation of $C(T)$ and $\epsilon(T)$ from statistical information about the system, collected during the annealing process, is described by *Andresen et al.* [1988]. An approximate, temperature-dependent transition probability matrix $P_E(T)$ for transitions between "energy levels" is formed during the annealing process. For each temperature the heat capacity can be evaluated from the eigenvector of P_E $(T \to \infty)$ with eigenvalue 1, and the relaxation time can be calculated from the second largest eigenvalue of $P_E(T)$.

[73] In practice, the thermodynamic speed v in equation (12) is adjusted such that the annealing temperature is close to zero after a predetermined number of iterations, given by the computer resources available to the optimization.

[74] Figure 7 shows an example where a reflection seismic data set was inverted through simulated annealing optimization. A common-depth-point gather from the data set was transformed into the $\tau - p$ domain (essentially a plane-wave decomposition), and two traces, representing ray parameters $p = 0.000025$ *s/m* and $p = 0.000185$ *s/m*, respectively, were inverted to obtain a horizontally stratified nine-layer model. Each

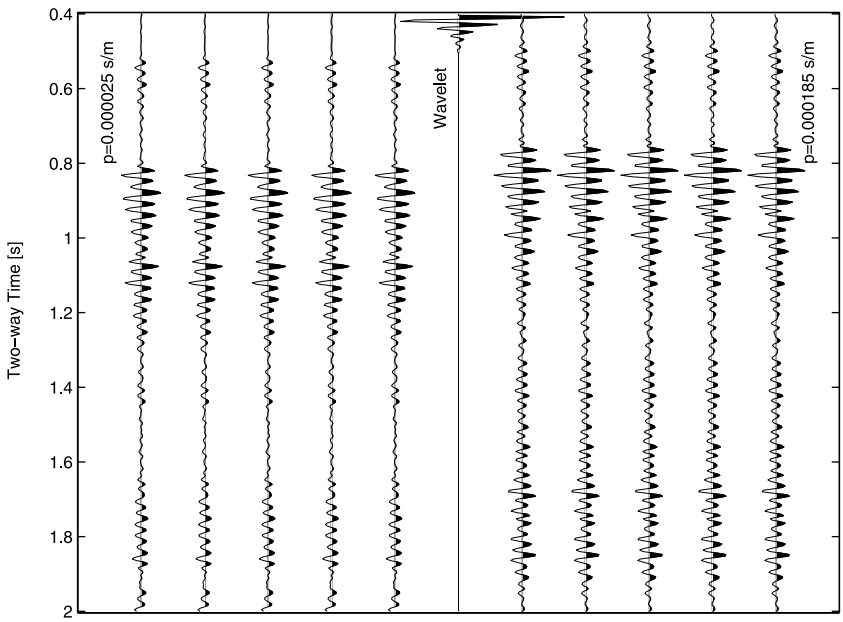

Figure 7. The two seismic $\tau-p$ data traces inverted by a simulated annealing procedure (each trace is repeated 5 times to facilitate inspection). A nine-layer, horizontally stratified *P* wave and density model was fitted to this data set using waveform modeling and the wavelet displayed in the middle of Figure 7.

layer was characterized by a *P* velocity, a density, and an attenuation. Figure 7 shows the two traces, each repeated 5 times, and the wavelet. Misfit ("energy" in simulated annealing terminology) was calculated by generating full waveform acoustic $\tau - p$ seismograms from the subsurface model and computing the L_2-norm of the difference between computed and observed ($\tau - p$ transformed) data.

[75] Figure 8 shows the "thermodynamic portrait" of the problem, generated by the method described by *Andresen et al.* [1988]. The top left graph displays the approximate "density of states" for the problem, showing the fraction of models having misfit (energy) between E and $E + dE$. There are several reasons for the small values of the density of states at small energies. First of all, the fractional volume of the model space with near-optimal models is extremely small, a typical property of inverse problems with high-dimensional model spaces. Second, the fact that noise is present in the data means that zero energy cannot be obtained (unless we overparameterize the problem). A third reason is that the method used to estimate density of states is imperfect. Thermodynamic properties of the problem are estimated through an initial exploratory sampling of the model space, and this sampling visits very few low-energy models.

[76] From the density of states the temperature-dependent "heat capacity" for the problem can be calculated (top right graph of Figure 8). The heat capacity for this problem has two distinct peaks, and these two peaks indicate where annealing must "slow down" if the system is to remain close to equilibrium.

[77] The temperature-dependent relaxation time (bottom left graph) shows a typical high value at low temperatures. This is the result of slow movement in an "energy landscape" with many secondary minima.

[78] The bottom right graph shows the resulting temperature schedule satisfying approximate constant thermodynamic speed. Practical experience shows that this kind of temperature schedule is far superior to ad hoc schedules not reflecting the thermodynamic properties of the problem [*Jakobsen et al.*, 1988; *Andresen et al.*, 1988; *Koren et al.*, 1991; *Mosegaard and Vestergaard*, 1991].

3.2.4. Genetic Algorithms

[79] It has long been observed that biological evolution is a form of optimization. In the natural world, fauna compete for survival and adapt to their surroundings. A natural system evolves by using a large population to explore many options in parallel rather than concentrating on trying many changes around a simple design [*Gershenfeld*, 1999]. Genetic algorithms, or GAs, as they have become known, are the class of numerical method that try to do the same thing. As noted above, GAs largely grew out of the work of *Holland* [1975], although independent earlier work of *Fogel et al.* [1966] first established the concept of evolutionary computation.

[80] In contrast to the basic form of simulated annealing, which keeps one set of parameters that are continually updated, GAs work on an ensemble of sets of parameters, with less emphasis placed on any particular member. (Note that there also exist more complex forms of simulated annealing where an ensemble of *random walkers* share or exchange statistical information about

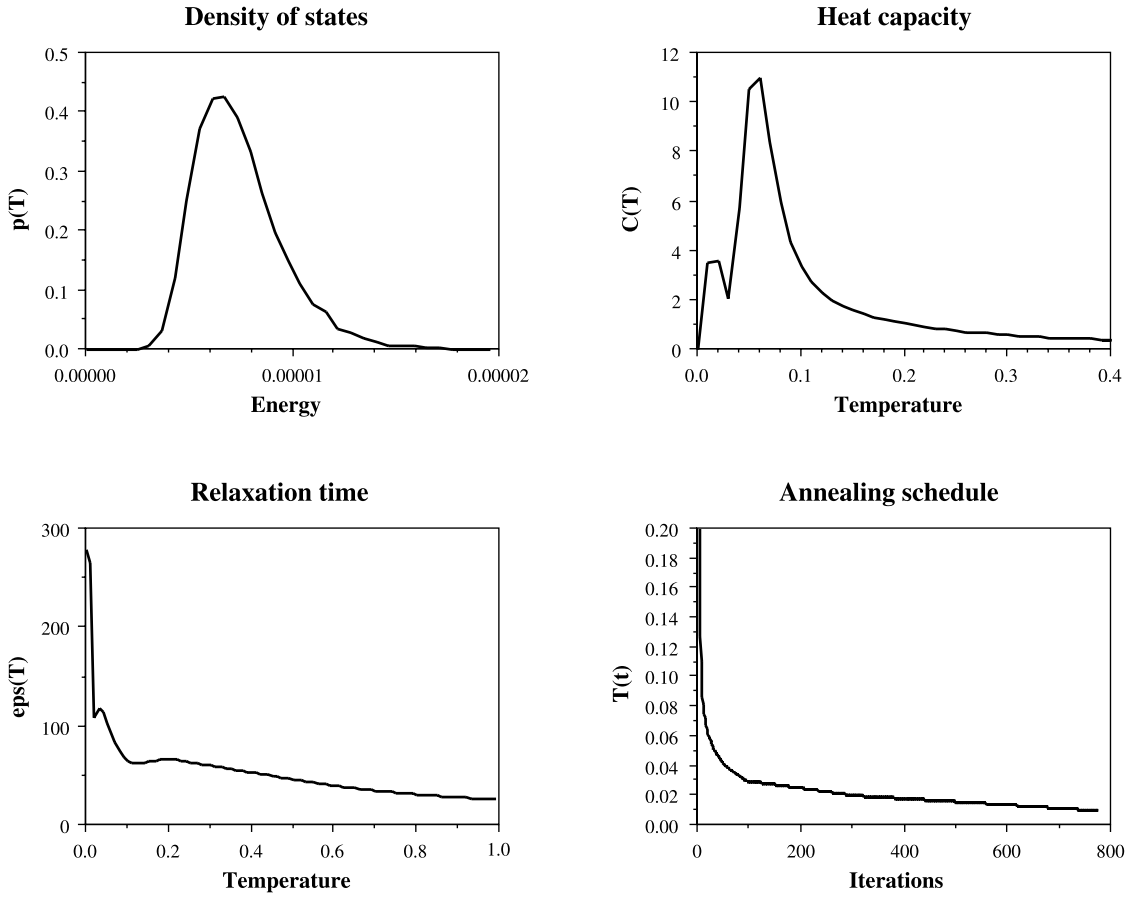

Figure 8. The "thermodynamic portrait" of the simulated annealing problem arising from inversion of the seismic data. The heat capacity can be calculated from the density of states. Combined with the relaxation time, it allows calculation of an annealing temperature schedule satisfying constant thermodynamic speed.

the misfit surface. See, for example, *Harland and Salamon* [1988], *Hoffmann et al*. [1990], *Mosegaard and Vestergaard* [1991], and *Ruppeiner et al*., [1991]). The early history of GAs in the Earth sciences has been traced above. Here we provide a brief description of how they can be used to search a parameter space. For more information the reader is referred to the comprehensive summaries that have appeared [*Davis*, 1987; *Goldberg*, 1989; *Rawlins*, 1991; *Whitley*, 1994; *Gallagher and Sambridge*, 1994; *Winter et al*., 1995] and also to the web resources in Table 1.

[81] A key aspect of a genetic algorithm is the representation of complicated models with a simple encoding. The most common is that used by *Holland* [1975], and it consists of a simple bit string. For example, in a problem involving d unknowns, each could be allowed to take 2^n values and be represented by a bit string of length $d \times n$. This is, in effect, a parameterization of the problem. The overall effectiveness of a GA can depend crucially on how it is encoded. There are many other possibilities besides the simple bit string, some of which do not even involve direct representation of parameter values [e.g., *Mathias and Whitley*, 1992]. There is no doubt the most suitable will depend on the particular problem in hand.

[82] The genetic algorithm works with a population of Q models simultaneously. Usually, the population is initially generated randomly, but at each iteration it is altered by the action of three operators. The fitness (objective function) for each model in the starting population is evaluated by solving the forward problem for each of the Q models. The purpose of the GA is then to seek out fitter models in parameter space. The three operators are known as selection, crossover, and mutation, and they are illustrated for the simple bit string encoding in Figure 9.

3.2.4.1. Selection

[83] From the initial population of Q bit-strings an interim population of Q parents is generated by selecting models from the original group with the probability of selection determined by the fitness value. For example,

$$P(\mathbf{m}_k) = A \exp\left[-B\phi(\mathbf{m}_k)\right], \qquad (13)$$

where A and B are problem specific constants. (Note the similarity with probability function used in the Metropolis-Hastings method, equation (2).) The higher the fitness of each model, the more likely that it will pass

Table 1. World Wide Web Addresses Where Papers, Tutorials, and Software on Monte Carlo and Global Optimization Methods Can Be Found.[a]

Subject	*Title*	*Address*
Genetic algorithms	IlliGAL home page	www-illigal.ge.uiuc.edu
	genitor group (Colorado state)	www.cs.colostate.edu/~em genitor
	genetic algorithm group	www.cs.gmu.edu/research/gag
	genetic algorithms archive	www.aic.nrl.navy.mil/galist
	genetic programming notebook	www.geneticprogramming.com
Simulated annealing and global optimisation	ensemble-based simulated annealing	www-rohan.sdsu.edu/~frostr/Ebsa
	Optimization Technology Center	www.ece.nwu.edu/OTC
	adaptive simulated annealing	www.ingber.com
	global optimization survey	www.cs.sandia.gov/opt/survey
	global optimization archive	www.mat.univie.ac.at/~neum/glopt.html
Markov chain Monte Carlo and other	MCMC reprint service	www.statslab.cam.ac.uk/~mcmc
	neighbourhood algorithm page	rses.anu.edu.au/~malcolm/na
	MCMC and SA in inverse problems	www.gfy.ku.dk/~klaus/mcinversion
	BUGS website	www.mrc-bsu.cam.ac.uk/bugs
	quasi Monte Carlo	www.mcqmc.org
Monte Carlo in geophysical inversion	a website containing material and links from this article	rses.anu.edu.au/~malcolm/mcgi

[a]All links were active at the time of writing.

into the next population. Since the population size does not change, multiple copies of the fitter models will survive at the expense of the less fit models, which may be extinguished completely. This operator introduces the element of survival of the fittest into the algorithm.

3.2.4.2. Crossover

[84] This operator cuts and mixes pairs of randomly chosen bit strings together. All Q parent strings are randomly paired to produce $Q/2$ couples. A crossover probability P_c is assigned, and if a random number between 0 and 1 is less than P_c, parts of the two strings are interchanged (see Figure 9). If a crossover is selected, the location at which the strings are cut is determined randomly, otherwise the two parent strings are passed unscathed to the next generation (see Figure 9). In this way it is hoped that information is passed on to subsequent generations.

3.2.4.3. Mutation

[85] The purpose of the mutation operator is to maintain a degree of diversity in the population. Note that the

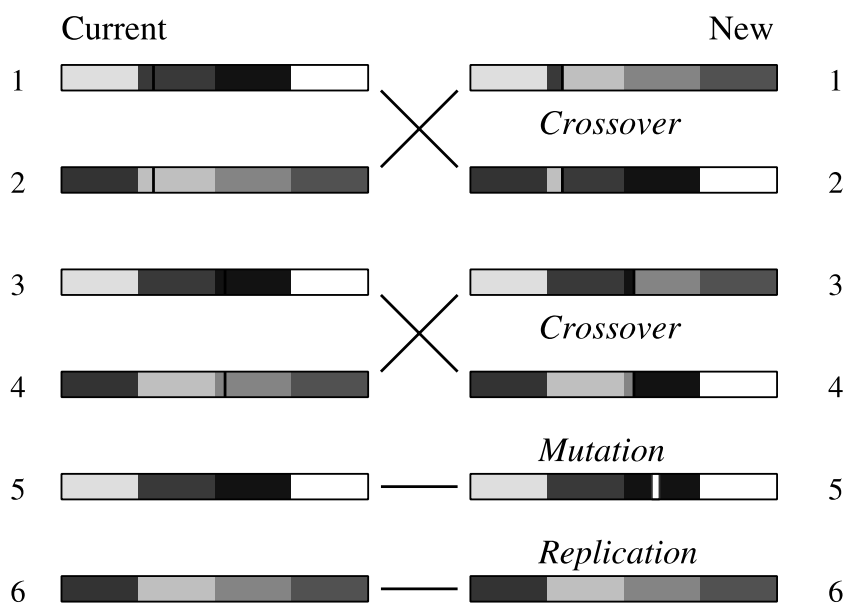

Figure 9. The effect of the three bit string operators in a genetic algorithm. In cases 1–4, the crossover operator cuts, swaps, and rejoins pairs of strings. In case 5, mutation changes a single bit. In case 6 a whole string is replicated. These operators act sequentially on a population of strings. The shading represents the values of four different variables encoded into the bit string.

selection operator acts to remove diversity of bit patterns from the population. In mutation any bit in an individual string is allowed to flip between 0 and 1. This is usually performed with a relatively low probability P_m.

[86] Overall, the action of the three operators is to produce a new population of models for which the forward problem must be solved and the fitness evaluated. After many iterations the population has the potential to evolve toward a fitter on-average state. Each stage involves randomized decisions which are influenced by the "control" parameters. Even in this basic description, we have five control parameters (Q, P_c, P_m, A, and B), which usually must be chosen (tuned) for each application. Although a GA is often thought of as a direct search method of global optimization (i.e., one not using gradient information), it can still become trapped in local minima or be stalled by complicated fitness landscapes. (The same is true of almost all other search and optimization methods.) Careful tuning of the control parameters may be required in these circumstances, and there is never any guarantee of convergence to a global minimum. Many variants of the basic genetic algorithm have been developed, several of them in geoscience applications (for discussions see *Gallagher and Sambridge* [1994], *Sen and Stoffa* [1995], *Koper et al.* [1999], and the references above).

[87] The field of research into genetic algorithms has grown significantly over the past 10 years. A large literature has now developed, much of which can be found either through the papers cited above or the web resources in Table 1. Outside of the Earth sciences genetic algorithms have found applications in a range of areas and not just in function or combinatorial optimization problems. For example, in the original book by *Holland* [1975] the idea that GAs could be used to train neural networks is presented. As a Monte Carlo inversion method, GAs are perhaps best thought of as a direct search method in a parameter space. The result is an ensemble of models that are geared toward the better/fitter regions of parameter space. Information still needs to be extracted from the resulting ensemble, which is a focus of current research and is discussed further in section 3.4.

3.2.5. Other Approaches

[88] A major field within computational statistics is the development of techniques for Monte Carlo integration. Essentially, this means numerically evaluating multidimensional integrals of the form

$$I = \int_{\mathcal{M}} f(\mathbf{x}) d\mathbf{x} \,. \tag{14}$$

This type of integral occurs repeatedly in Bayesian inference studies. The main issue is to sample the multidimensional function $f(\mathbf{x})$ in some efficient manner so that numerical estimates of the integral can be made. A large literature exists on the subject [see *Flournay and Tsutakawa*, 1989, and references therein]. The Metropolis-Hastings algorithm and Gibbs Sampler (discussed above) are examples of statistical importance sampling techniques that can be used for this task. Hence there is a direct connection between Monte Carlo integration and Monte Carlo (adaptive) sampling of a parameter space. In principle then, other types of importance sampling methods specifically designed for multidimensional integration might be useful to adaptively sample a parameter space.

[89] One such technique, widely used in the field of particle physics, is the Adaptive Monte Carlo algorithm of *Lepage* [1978]. This procedure iteratively generates its own sampling density function, which approximates the integrand in equation (14). The sampling density is chosen to be separable, that is, a function of M variables that is a product of M 1-D functions, one for each axis in parameter space. By drawing random samples according to this density function the algorithm samples in the significant regions of $f(\mathbf{x})$. If $f(\mathbf{x})$ were taken as the data misfit/objective function, then the algorithm could be used to successively concentrate sampling where the function was high and hence act as an MC search algorithm. For further details see *Lepage* [1978] and *Press et al.* [1992]. Given the wide use of this approach for multidimensional integration, it is surprising that to date it does not seem to have been used in geophysical inversion. One potential problem is that a separable sampling density may poorly represent multidimensional distributions when variables are highly correlated, which is likely in many inverse problems. In this case, narrow valleys can occur in the multidimensional fitness function, which is inclined with all of the parameter axes (A. Lomax personal communication, 1996).

[90] Another recent technique is the neighbourhood algorithm [*Sambridge*, 1998, 1999a, 1999b]. In this case the main idea is to generate a set of samples, at each generation, whose sampling density function is constructed from all previous models using the neighbourhood approximation. This is a partition of parameter space into Voronoi (nearest-neighbor) cells about each of the previous models. In this way the information in the previous samples drives the search for new models. At regular intervals the approximation is updated, and sampling can concentrate in multiple regions.

[91] As *Sambridge* [1999a] points out, the number of vertices of Voronoi cells grows exponentially as the dimension of the parameter space increases, because of the curse of dimensionality. However, the NA sampling algorithm does not require their explicit calculation and hence remains practical in problems with hundreds of unknowns. Figure 10 shows an example of the neighbourhood sampling algorithm on a simple test problem from *Sambridge* [1998]. At each iteration a choice must be made on how to sample from the current approximate fitness function. *Sambridge* [1998] used a Gibbs Sampler, while *Sambridge* [1999a] resampled particular

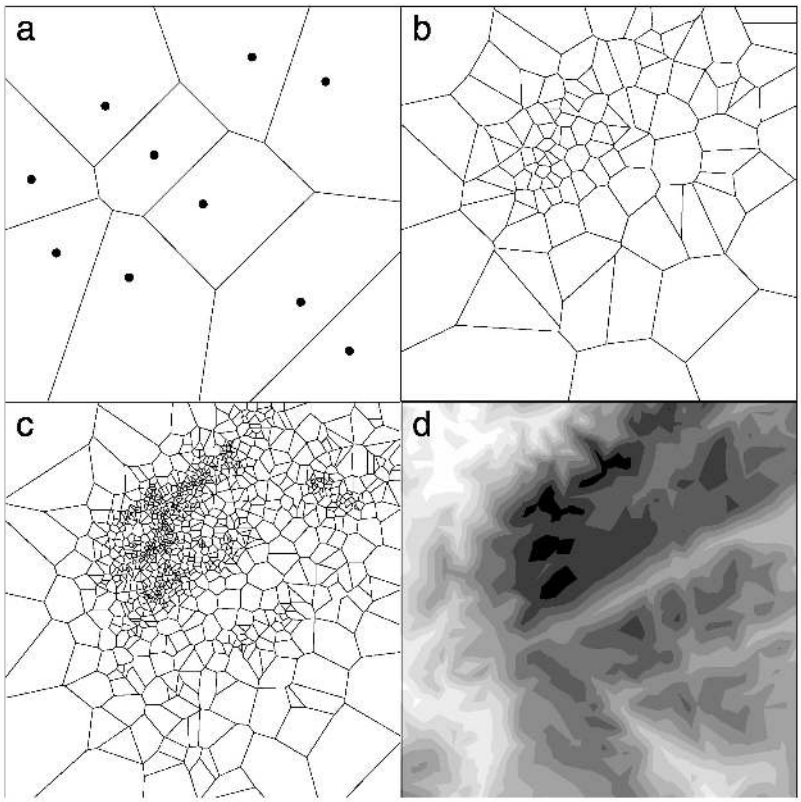

Figure 10. Voronoi cells drawn around the sampling produced by a neighbourhood algorithm [*Sambridge*, 1999a] for a simple 2-D problem. (a–c) Voronoi cells of about 10, 100, and 1000 generated samples, respectively. Figure 10c shows the true fitness surface. As the algorithm proceeds, each Voronoi diagram is used to guide the sampling. Note how the Voronoi cells are more concentrated in the darker regions where the fit is high. (From *Sambridge* [1999a].)

subsets of Voronoi cells. Like a genetic algorithm, the neighbourhood approach updates a population of models at each iteration but does so using the Voronoi cell concept to identify "promising regions" of parameter space. Initial results with this technique are quite promising, but more experience is required.

3.3. Bayesian Inference

[92] In a Bayesian formulation the solution to an inverse problem is given by the so-called posterior probability distribution $\sigma(\mathbf{m})$ over the model space $\mathcal{M}$. $\sigma(\mathbf{m})$ carries all information available on models originating from two sources:

1. The data are one source. This information is given by a likelihood function $L(\mathbf{m})$, measuring the likelihood of a given model $\mathbf{m}$ through its misfit $\phi(\mathbf{m})$. A typical likelihood function can be expressed as

$$L(\mathbf{m}) = \text{const} \cdot \exp\left(-\phi(\mathbf{m})\right). \tag{15}$$

2. Data-independent prior information, expressed through a *prior probability density* $\rho(\mathbf{m})$, is another source. *Prior probability densities* may be simple Gaussian PDFs [see, e.g., *Mosegaard*, 1998; *Mosegaard and Rygaard-Hjalsted*, 1999], or they may be more complicated PDFs carrying information from earlier observations or measurements [see, e.g., *Mosegaard et al.*, 1997].

[93] The posterior PDF is related to data and prior information through the relation

$$\sigma(\mathbf{m}) = \text{const} \cdot L(\mathbf{m})\rho(\mathbf{m}). \tag{16}$$

In cases where no explicit mathematical expression for $L(\mathbf{m})$ and/or $\rho(\mathbf{m})$ is available, Monte Carlo sampling is the only efficient way to explore $\sigma(\mathbf{m})$. In Bayesian analysis the Metropolis-Hastings algorithm or the Gibbs Sampler is used to generate samples from $\sigma\mathbf{m}$), thereby allowing us to estimate averages over the model space. The algorithm will sample a large number of models $\mathbf{m}_1, \ldots, \mathbf{m}_N$, according to $\sigma(\mathbf{m})$, after which any average of a function $f(\mathbf{m})$ over the model space $\mathcal{M}$ (e.g., a covariance or an expectation) can be approximated by the simple average:

$$E_f = \int_{\mathcal{M}} \sigma(\mathbf{m}) f(\mathbf{m}) d\mathbf{m} \approx \frac{1}{N} \sum_{n=1}^{N} f(\mathbf{m}_n). \tag{17}$$

The probability of an event $\in \subseteq \mathcal{M}$, containing all models in model space with a given feature, is found by putting $f(\mathbf{m})$ equal to the indicator function

$$f(\mathbf{m}) = \begin{cases} 1 & \text{if } \mathbf{m} \in \varepsilon \\ 0 & \text{otherwise.} \end{cases} \tag{18}$$

ε may, for instance, contain all Earth models that are "approximately" constant in a given area of the subsurface (appropriately defined) or models containing a sharp boundary (again appropriately defined) in a certain depth interval. In the special case when $\varepsilon = \mathcal{M}$ and $f(\mathbf{m}) = m_i$, the integral E_f is simply the mean $\langle m_i \rangle$ of the ith model parameter m_i. If $f(\mathbf{m}) = (m_i - \langle m_i \rangle)(m_j - \langle m_j \rangle)$, E_f becomes the posterior covariance between the ith and jth model parameters.

[94] In the general inverse problem, direct evaluation of E_f may be impossible because no analytical expression for $\sigma(\mathbf{m})$ is available. Even for a linear inverse problem with known $\sigma(\mathbf{m})$, we may not be able to evaluate E_f directly because of the complexity of f support ε. In all these cases the Metropolis-Hastings algorithm or the Gibbs Sampler allows us to perform an approximate, numerical evaluation of E_f.

[95] A large collection of samples from $\sigma(\mathbf{m})$ also allows us to compute approximate marginal distributions. For instance, samples from the 1-D marginal $\sigma(m_k)$ are obtained simply by collecting values of m_k from samples $\mathbf{m} = (m_1, \ldots, m_k, \ldots, m_M)$ of the posterior $\sigma(\mathbf{m})$.

[96] It is often very difficult to obtain good estimates of the averages E_f because the posterior $\sigma(\mathbf{m})$ may be multimodal and typically undersampled. The 1-D marginals, however, are usually quite well sampled because all the samples from the high-dimensional model space contribute to the 1-D histogram of m_k. For this reason, posterior 1-D marginals play an important role in applications.

[97] Although the above procedure looks very general and simple, the practical problem is often how to discover which features of the model have a high probability (in traditional non-Bayesian terminology: which features of the model are "well resolved"). It turns out that inspection of the output from the Monte Carlo sampling algorithm is often the most efficient way to discover well-resolved structure. An example of this is seen in Figures 11 and 12, taken from *Mosegaard et al.* [1997]. Seismic vertical-incidence reflection data d_{obs} from the base of the crust (Figure 11) were analyzed using a Metropolis-Hastings algorithm, generating posterior acoustic impedance models $\mathbf{m}_n$ of the subsurface. A wavelet was estimated, and a full waveform forward algorithm was used to generate synthetic data $d(\mathbf{m})$ in the two-way time interval between 8.0 and 10.0 s. A likelihood function

$$L(\mathbf{m}) = \exp\left(\frac{\|d_{\text{obs}} - d(\mathbf{m})\|^2}{\sigma^2}\right)$$

measured the degree of fit between observed and computed data, where σ is the estimated standard deviation of the noise in the data. The prior probability density $\rho(\mathbf{m})$ for the acoustic impedance models was derived

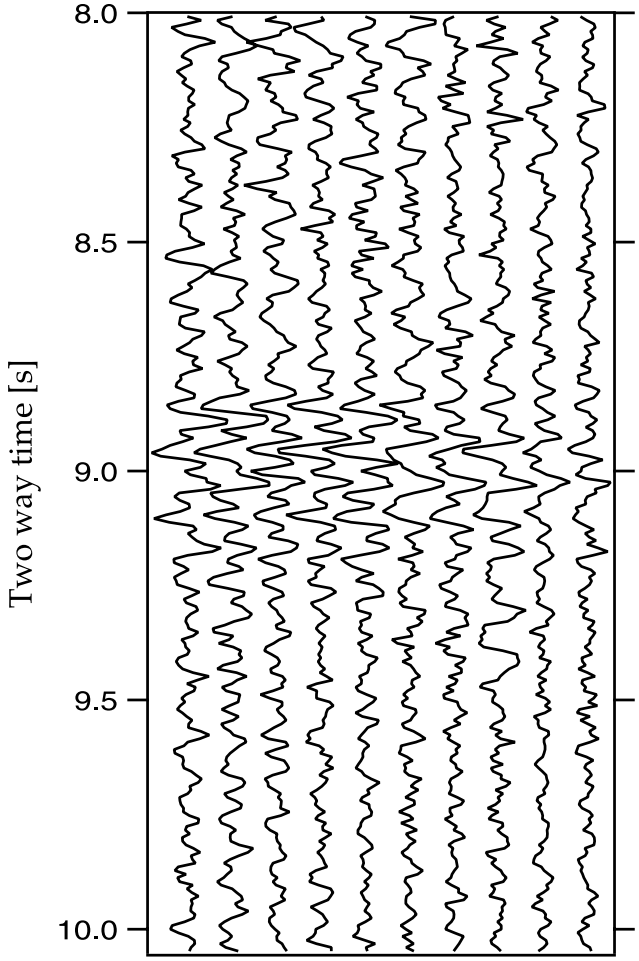

Figure 11. The seismic data used for Bayesian inversion, covering the time range of 8–10 s that contains the Moho reflection at ~8.9 s. (From *Mosegaard et al.* [1997].)

from histograms of reflection coefficients and layer thicknesses (Figure 13), obtained from laboratory measurements on near-surface rocks similar to those expected at the base of the crust (and corrected for pressure differences between surface and Moho). Samples from the posterior distribution $\sigma(\mathbf{m})$ are shown in Figure 12. Well-resolved structure is seen as structure that is persistent (occurs very often) in the output. In this case the simplest way to discover persistent structure is to plot all output models side by side or on top of each other. Figure 12 shows that impedance structure between 4.4 and 4.7 s two-way time is persistent and therefore well resolved. In contrast to this the main trend of the impedance and impedance fluctuations below 4.7 s are transient (nonpersistent) and therefore are poorly resolved.

[98] However, the above method is not useful for models which are more naturally plotted as a function of two or three spatial dimensions. A more generally valid method is to display all output models sequentially as pictures in a movie [*Mosegaard and Tarantola*, 1995]. Often, the human eye will be able to discover which

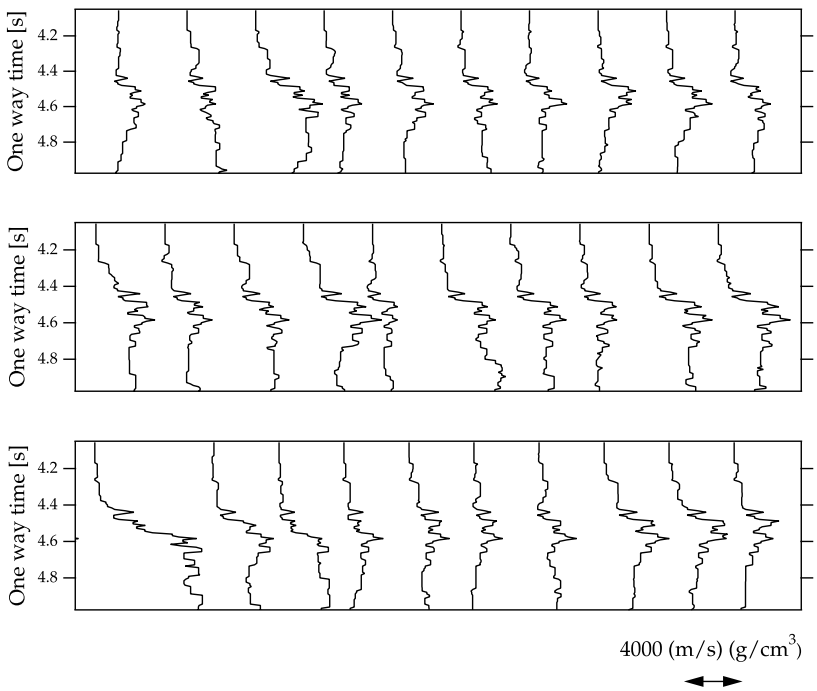

Figure 12. The results of the inversion on the seismic data set. A selection of a posteriori models are plotted side by side to facilitate the discovery of well-resolved structure. Such structure is characterized by its presence in most (or all) of the posterior models. (From *Mosegaard et al*. [1997].)

structure is well resolved by the data and the prior information. This method can often be facilitated by viewing movies of models with different degrees of smoothing.

3.4. Ensemble Inference

[99] All of the Monte Carlo direct search algorithms discussed above require the solving of the forward problem many times. In each case the result is an ensemble of models, and the task is to try and use their collective sampling to extract some meaningful information about the Earth. In most cases the distribution of the complete ensemble (i.e., all models for which the forward problem has been solved) is arbitrary and follows no prescribed distribution. However, as discussed above, when Monte Carlo sampling is applied within a Bayesian context, the aim is to generate samples that follow a particular statistical distribution, usually the posterior PDF. In that case, "Bayesian information measures" (e.g., model covariance estimates, etc.) can be derived from a subset of the complete distribution. For the case when one has an arbitrary ensemble of models with a range of fits to the data, the question remains of how to draw inferences.

[100] A simple answer would be to examine the features present in only the best data fitting model. However, as discussed above, in almost all cases this is insufficient because of nonuniqueness in the problem and noise in the data. A useful thing to do may be to filter the complete ensemble into only those which have acceptable fit to the data (given a prior understanding of noise processes, etc.), as done by *Press* [1968, 1970a]. One could also include other criteria like prior constraints on the form of acceptable solutions, for example, the velocity gradient constraints used by *Wiggins* [1969, 1972]. The task then would be to try and determine properties that all of these acceptable models share in the hope that the real Earth shares it also.

[101] Several authors have proposed methods for characterizing an ensemble of data acceptable models. A summary is given by *Sen and Stoffa* [1995]. The earliest approach was simply to compare the models directly, usually by plotting them on top of one another [e.g., *Press*, 1968]. As more modern search algorithms were used, and many more acceptable models produced, more quantitative analyses were performed. Figure 14 shows an example from *Lomax and Sneider* [1995a] where upper mantle seismic *S* velocity models are sought that satisfy synthetic Rayleigh wave group-velocity dispersion data contaminated by noise. Figure 14a shows the ensemble of data-acceptable velocity models, and Figure 14b shows the ± 1 and $\pm 2\sigma$ spread, calculated for each depth. Although strictly speaking these intervals are not formal confidence limits on the velocity at each depth, they do provide an effective means at characterizing the ensemble. Note that in this synthetic problem the true solution is the IASP91 Earth model [*Kennett and Engdahl*, 1991], which plots neatly within the $\pm 1\sigma$ bounds. Other authors have made similar attempts at a graphical representation of the acceptable ensemble. Variations include plots of the density of acceptable models or all models weighted by data fit. For examples,

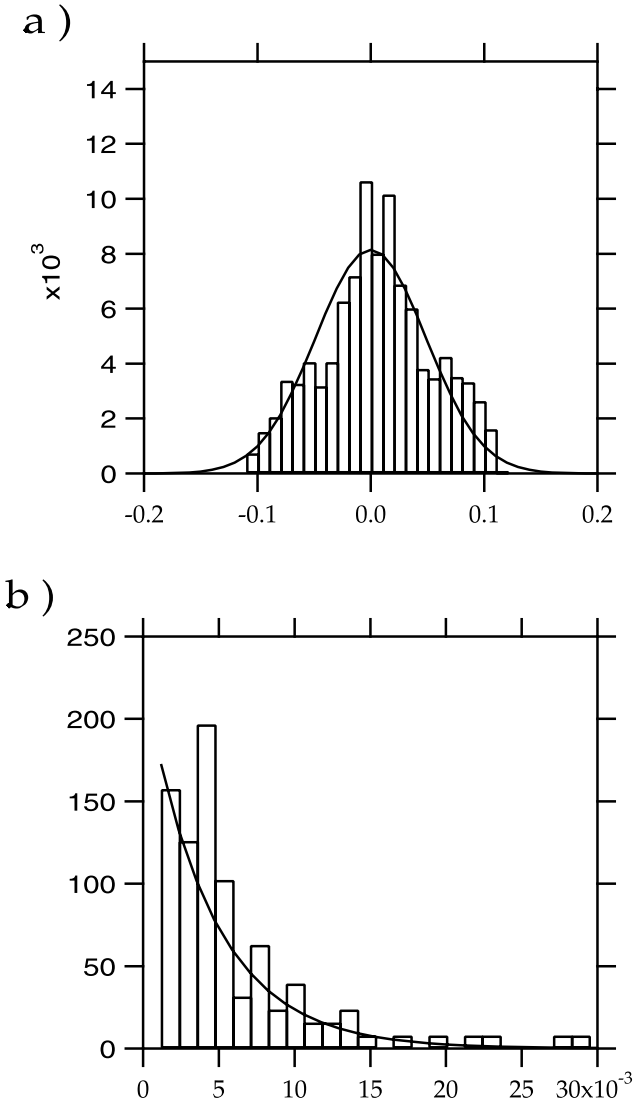

Figure 13. A priori information used in the Bayesian inversion scheme. (a) Reflection coefficients distribution obtained from studies of the igneous intrusions of Rum, Scotland, and the Great Dyke, Zimbabwe. The solid curve is a Gaussian distribution with standard deviation $\sigma = 0.047$. (b) Layer thickness distribution as a function of one-way travel time derived from observations of the Rum and Great Dyke intrusions. The solid curve is an exponential distribution with $\lambda = 225.0\ s^{-1}$. (From *Mosegaard et al.* [1997].)

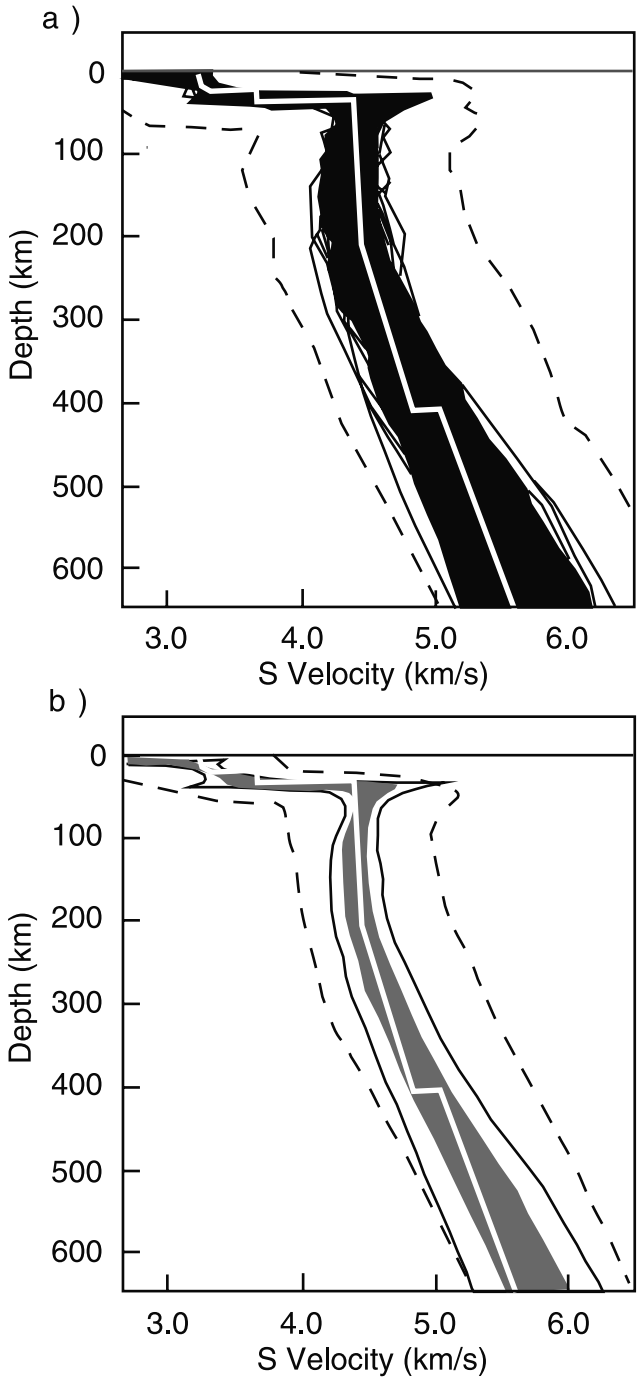

Figure 14. (a) Ensemble of upper mantle *S* velocity models generated by *Lomax and Snieder* [1995a], which have an acceptable level of data fit. The data are synthetically generated Rayleigh wave group velocity dispersion measurements. The dashed lines show the boundaries of the parameter space. (b) The $\pm 1\sigma$ (gray shaded region) and $\pm 2\sigma$ (solid line) in the distribution of the acceptable models, calculated at each depth. The outer dashed lines are the $\pm 2\sigma$ spread of a uniform distribution of models. Note how the spread measures are smooth, even though the individual acceptable models are not. The true model (thick white line) lies, for the most part, in the center of the $\pm 1\sigma$ region. (After *Lomax and Snieder* [1995a].)

see *Basu and Frazer* [1990], *Nolte and Frazer* [1994], *Shibutani et al.* [1996], and *Kennett* [1998].

[102] An alternative to graphical methods is the cluster analysis method proposed by *Vasco et al.* [1993]. They used statistical techniques to characterize the acceptable ensemble and make inferences about properties that all models shared. They illustrated their approach with some impressive applications to gravity inversion and cross-borehole seismic imaging. A related approach is that proposed by *Douma et al.* 1996. In their method the ensemble of acceptable models is projected onto a set of orthogonal functions. In this way they hope to determine

the well constrained information hidden in the ensemble. As with the method of *Vasco et al*. [1993] the technique is appropriate if the ensemble forms a single cluster and not multiple disjointed clusters, as depicted in Figure 1. If multiple clusters exist, then they must be identified, and the approach applied separately to each.

[103] Recently, *Sambridge* [1999b] has proposed an approach to estimate Bayesian information measures from an arbitrarily distributed ensemble. In this case the neighborhood approximation (mentioned above in connection with a search algorithm) is applied to the complete ensemble, that is, all models for which the forward model has been solved. A standard MCMC integration procedure is then used to resample according to the neighborhood approximation and to produce unbiased estimates of Bayesian integrals. The advantage of this approach is that it can, in principle, extract Bayesian estimates from ensembles generated by any method, for example, a genetic algorithm. A caveat, which applies to all of these techniques, is that at best they can only extract what information exists in a given ensemble of models. If the available ensemble does not adequately represent (or sample) the acceptable region in parameter space, then the result will be biased.

[104] In the past few years, geophysicists have begun to develop techniques for ensemble inference. The underlying idea is that basing inferences on an ensemble of potential solutions is more useful than considering just one (usually best fit) model. The current range of approaches available are useful but by no means complete or applicable in all circumstances. It seems likely that this will be an area of further research in the next decade.

3.5. Resources on the Net

[105] It may be useful to the reader to have a list of web site addresses where more information can be found on various topics discussed in this paper. At some sites a computer code is also available. Table 1 shows a collection that we are aware of. Many other sites can be found by following the links on these pages.

4. CONCLUSIONS

[106] In this paper we have summarized the current state of the Monte Carlo methodology, as applied in geophysics. Some open questions and directions for future research have been discussed. The Earth sciences are a rich source of large and complicated inverse problems. Often they can be posed in a way suitable for attack with MC methods. In doing so, the Earth scientist must remember that the development of Monte Carlo techniques is a multidisciplinary effort, and it can be beneficial to look at what has been done elsewhere. It is hoped that this paper will facilitate that process by drawing the attention of geophysicists to relevant reference material in computational statistics and other fields.

[107] We have highlighted the connection between the computational approaches commonly used for data fitting and model appraisal with those in the fields of optimization and statistical inference. Techniques such as simulated annealing, genetic algorithms, and other statistical importance sampling methodologies have all started life elsewhere but become useful additions to the geophysicists tool bag. One issue that was prevalent at the outset (>30 years ago) was how to make robust inferences from an ensemble of data acceptable models. In recent times this question has come back into focus. Geophysicists are good at generating models, and even at finding some that fit data, but quantitative and reliable inferences are needed for drawing conclusions. The statistical and other approaches discussed in this paper have begun to answer these questions in a quantitative manner.

GLOSSARY

[108] **Acceptable models**: A set of Earth models (see definition) that are consistent with observations, taking into account the presence of noise and errors in the observables. Usually, the acceptable models are those for which the objective (or data misfit) function is less than, or equal to, a predetermined value.

[109] **Bayesian inference**: A (purely) probabilistic procedure for information updating. In the typical scenario prior information, represented by a prior probability distribution, is updated by information from experimental data, represented by a likelihood function. The resulting state of information is represented by the (normalized) product of the prior distribution and the likelihood function, the so-called posterior probability distribution. The terms "prior" and "posterior" refer to "before" and "after" data collection, respectively.

[110] **Direct search**: A parameter space search method (see definition) that does not use Fréchet derivatives or gradient information of the objective function. It only uses evaluations of the objective function at points in parameter space.

[111] **Earth model**: A representation of some physical properties of the Earth, usually in a discrete form where the earth property, for example, Earth's density variation, is described by a finite number of parameters, for example, the average density of rocks in a series of depth intervals from the surface of the Earth.

[112] **Ensemble inference**: The procedure whereby information is extracted or conclusions are drawn from a collection of inputs, or Earth models, rather than a single model. Genetic Algorithms, Simulated Annealing, and the Neighbourhood algorithm are all examples of ensemble inference (direct search) techniques.

[113] **Forward problem**: The calculation of predictions from an Earth model, to be compared with the observa-

tions. The solution of the forward problem is a necessary step in determining the fit to the data. The degree of nonlinearity in the forward problem can be a major cause of difficulty in finding Earth models that fit data.

[114] **Genetic algorithm**: A computational method popular for global optimization problems. This is a fully nonlinear direct search method, which means that it avoids linearization and does not use derivatives of the objective function being optimized. The procedure is based on an analogy with principles of evolution and "survival of the fittest." Originally developed as a model for adaptive behavior in an artificial system, it has been applied to complex optimization problems across the physical sciences.

[115] **Global optimization**: A numerical procedure where the values of some variables are sought that give an optimum value of some function, that is, maximum or minimum over all possibilities.

[116] **Importance sampling**: A Monte Carlo algorithm designed to sample a given function, or probability density, with a sampling density proportional to that function or probability density.

[117] **Inverse problem**: The term used to describe the situation where inversion is applied. (See definition of inversion).

[118] **Inversion**: The process by which information is obtained about a physical quantity from indirect information, for example, on Earth's seismic structure from the travel times of elastic waves generated by earthquakes.

[119] **Likelihood function**: A function describing the probability that a given parameterized model is consistent with observed data. The likelihood function is a function of the model and measures the fit between model predictions and observations.

[120] **Markov Chain**: See random walk.

[121] **MCI** (Monte Carlo Inversion): The term used to describe the earliest forms of uniform random parameter space searching, as applied to seismological problems in the work of *Press* [1968, 1970a].

[122] **MCMC** (Markov Chain Monte Carlo): A multidimensional random sampling procedure where the next sample only depends on the location of the current sampling point.

[123] **Metropolis-Hastings algorithm**: A simple importance sampling algorithm (see definition of importance sampling) based on a random walk.

[124] **Misfit function**: The same as objective function, except that one always minimizes a misfit function.

[125] **Monte Carlo**: A prefix indicating that the method or approach makes use of repeated trials, or sampling, generated with the use of random numbers, named after the famous French city associated with casinos.

[126] **Monte Carlo integration**: A numerical integration procedure for functions, which is efficient in high-dimensional spaces. Instead of using functional values in a uniform grid to produce a numerical approximation of an integral, they are evaluated at randomly selected points (often generated by an importance sampling algorithm (see above). The integral can then be approximated by an average of these values, and the chance of missing areas of high functional values is strongly reduced.

[127] **Objective function**: A function calculated from one or more variables which is usually the subject of an optimization process, that is, the variables are sought, which optimize the objective function. The objective function can be a measure of discrepancy with observed data (data misfit) or a combination of data misfit and other criteria.

[128] **Parameter space**: An abstract multidimensional region describing the set of all possible values that Earth models can take. Each point in the parameter space represents an Earth model.

[129] **Parameter space search**: The process of finding Earth models that have an acceptable or optimal value of the objective function.

[130] **Posterior PDF**: A probability density representing, in a Bayesian computation, the combination of prior information and data information.

[131] **Prior probability density**: A probability density representing, in a Bayesian computation, the information available before data is acquired (or considered).

[132] **Probability density function (PDF)**: A statistical term used to describe the probability distribution associated with a random variable. For example, the well-known "bell-shaped" Gaussian distribution is a PDF of a one-dimensional normal random variable. The area under the curve between any two given values represents the probability that a realization of the variable will fall between these values. Multidimensional PDFs are used to describe the joint behavior of a random vector (whose components are all random variables).

[133] **Pseudorandom number**: The result of a numerical procedure (a generator), which has certain statistical properties. For example, the histogram of many pseudorandom numbers will tend toward a uniform distribution. Most computing systems provide pseudorandom number generators that use a linear congruent method [see *Park and Miller*, 1988].

[134] **Quasi-random number**: Sometimes called "subrandom" numbers, similar definition as for pseudorandom numbers. Often it is based on number theoretic techniques and usually converges much faster than pseudorandom numbers to a uniform distribution.

[135] **Random Walk**: A perturbative sequence of random changes to a point in a multidimensional space. A uniform random walk is one where the asymptotic convergence is to a uniform density of samples in parameter space. See also Markov Chain Monte Carlo (MCMC).

[136] **Simulated annealing**: A Metropolis-Hastings algorithm designed to simulate certain aspects of a statistical mechanical system. The simulated annealing (SA) algorithm samples a Gibbs-Boltzmann distribution in the model space, and a characteristic feature of this

distribution is its temperature parameter. When this parameter is gradually lowered, it is possible to simulate an annealing process in which the system is taken from a disordered high-temperature state to a well-ordered low temperature state.

[137] **ACKNOWLEDGMENTS.** We would like to acknowledge Steve Billings, Jean Braun, Kerry Gallagher, Brian Kennett, Anthony Lomax, John Scales, and Roel Snieder for their assistance with this manuscript and also these and many other colleagues who have contributed to our understanding of the subject. Thanks also go to Craig Bina, who suggested this paper in the first place. The constructive criticisms of two anonymous reviewers helped improve an earlier draft of this manuscript.

[138] Louise Kellogg was the Editor responsible for this paper. She thanks technical reviewer Anthony Lomax and another anonymous technical reviewer.

REFERENCES

Aarts, E., and J. Korst, *Simulated Annealing and Boltzmann Machines*, John Wiley, New York, 1989.

Anderssen, R. S., The character of non-uniqueness in the conductivity modelling problem for the Earth, *Pure Appl. Geophys., 80*, 238–259, 1970.

Anderssen, R. S., and E. Senata, A simple statistical estimation procedure for Monte Carlo Inversion in geophysics, *Pure Appl. Geophys., 91*, 5–13, 1971.

Anderssen, R. S., and E. Senata, A simple statistical estimation procedure for Monte Carlo inversion in geophysics: Efficiency and Hempel's paradox, *Pure Appl. Geophys., 96*, 5–14, 1972.

Anderssen, R. S., M. H. Worthington, and J. R. Cleary, Density modelling by Monte Carlo Inversion, I, Methodology, *Geophys. J. R. Astron. Soc., 29*, 433–444, 1972.

Andresen, B., K. H. Hoffman, K. Mosegaard, J. Nulton, J. M. Pedersen, and P. Salamon, *J. Phys., 49*, 1485, 1988.

Antonov, I. A., and V. M. Saleev, *USSR Comput. Math. Math. Phys., 19*(1), 252–256, 1979.

Backus, G. E., Bayesian inference in geomagnetism, *Geophys. J., 92*, 125–142, 1988.

Backus, G. E., and J. F. Gilbert, Numerical applications of a formalism for geophysical inverse problems, *Geophys. J. R. Astron. Soc., 13*, 247–276, 1967.

Backus, G. E., and J. F. Gilbert, The resolving power of gross Earth data, *Geophys. J. R. Astron. Soc., 16*, 169–205, 1968.

Backus, G. E., and J. F. Gilbert, Uniqueness in the inversion of inaccurate gross Earth data, *Philos. Trans. R. Soc. London, Ser. A., 266*, 123–192, 1970.

Basu, A., and L. N. Frazer, Rapid determination of critical temperature in simulated annealing inversion, *Science, 249*, 1409–1412, 1990.

Bayes, T., An essay towards solving a problem in the doctrine of chances, *Philos. Trans. R. Soc. London, 53*, 370–418, 1763 (Reprinted, with biographical note by G. A. Barnard, in *Biometrika, 45*, 293–315, 1958).

Billings, S. D., Simulated annealing for earthquake location, *Geophys. J. Int., 118*, 680–692, 1994.

Billings, S. D., B. L. N. Kennett, and M. S. Sambridge, Hypocentre location: Genetic algorithms incorporating problem specific information, *Geophys. J. R. Astron. Soc., 118*, 693–706, 1994.

Bina, C. R., Free energy minimization by simulated annealing with applications to lithospheric slabs and mantle plumes, in *Geodynamics of Lithosphere and Earth's Mantle; Seismic Anisotropy as a Record of the Past and Present Dynamic Processes*, edited by J. Plomerova, R. C. Liebermann, and V. Babuska, *Pure Appl. Geophys., 151*, 605–618, 1998.

Biswas, N. N., and L. Knopoff, The structure of the upper mantle under the United States from dispersion of Rayleigh waves, *Geophys. J. R. Astron. Soc., 36*, 515–539, 1974.

Boschetti, F., M. C. Dentith, and R. D. List, Inversion of seismic refraction data using genetic algorithms, *Geophysics, 61*, 1715–1727, 1996.

Boschetti, F., M. C. Dentith, and R. D. List, Inversion of potential field data by genetic algorithms, *Geophys. Prospect., 45*(3), 461–478, 1997.

Box, G. E. P., and G. C. Tiao, *Bayesian Inference in Statistical Analysis*, Addison-Wesley, Reading, Mass., 1973.

Braun, J., Estimating exhumation rate and relief evolution by spectral analysis of age-elevation profiles, *Terra Nova, 14*(3), 205–209, 2002.

Buland, R., The mechanics of locating earthquakes, *Bull. Seismol. Soc. Am., 66*, 173–187, 1976.

Burton, P. W., Inversions of high frequency $Q_\gamma^{-1}(f)$, *Geophys. J. R. Astron. Soc., 48*, 29–51, 1977.

Burton, P. W., and B. L. N. Kennett, Upper Mantle Zone of low Q, *Nature, 238*, 84, 87–90, 1972.

Cadek, O., D. A. Yuen, and H. Cizkova, Mantle viscosity inferred from geoid and seismic tomography by genetic algorithms; results for layered mantle flow, in *Joint Inversion as a General Problem in Earth Sciences*, edited by L. Ballani, *Phys. Chem. Earth, 23*, 9–10, 865–872, 1998.

Calderon, M. C., M. K. Sen, and P. L. Stoffa, Hopfield neural networks, and mean field annealing for seismic deconvolution and multiple attenuation, *Geophysics, 62*, 992–1002, 1997.

Cary, P. W., and C. H. Chapman, Automatic 1-D waveform inversion of marine seismic refraction data, *Geophys. J., 93*, 527–546, 1988.

Chunduru, R. K., M. K. Sen, and P. L. Stoffa, 2-D resistivity inversion using spline parameterization and simulated annealing, *Geophysics, 61*, 151–161, 1996.

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

Constable, S. C., R. L. Parker, and C. G. Constable, Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics, 52*, 289–300, 1987.

Conway, A. J., and P. A. Bland, A genetic algorithm scheme for pairing meteorite finds, *Meteorol. Planet. Sci., 33*(3), 491–499, 1998.

Courboulex, F., J. Virieux, and D. Gibert, On the use of simulated annealing method and cross-validation theory for deconvolution of seismograms, *Bull. Seismol. Soc. Am., 86*(4), 1187–1193, 1996a.

Courboulex, F., J. Virieux, A. Deschamps, D. Gibert, and A. Zollo, Source investigation of a small event using empirical Green's functions and simulated annealing, *Geophys. J. Int., 125*, 768–780, 1996b.

Curtis, A., and R. Snieder, Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics, 62*, 1524–1532, 1997.

Cvijović, D., and J. Klinowski, Taboo Search: An Approach to the Multiple Minima Problem, *Science, 267*, 664–666, 1995.

Datta, G. A., L. W. Lake, and G. A. Pope, Stochastic simulation of categorical variables using a classification algorithm and simulated annealing, *Math. Geol., 27*(6), 763–787, 1995.

Davis, L., *Genetic Algorithms and Simulated Annealing, Research Notes in Artificial Intelligence*, Pitman, London, 1987.

Davis, T. E., and J. C. Principe, A simulated annealing like

convergence theory for the simple genetic algorithm, in *Proc. of the Fourth International Conference on Genetic Algorithms*, Morgan Kaufmann, San Mateo, Calif., 1991.

Deng, L. H., and J. A. Scales, Estimating the topography of multi-dimensional fitness functions, *Cent. for Wave Phenomena Tech. Rep. 208*, Colo. Sch. of Mines, Golden, 1999.

Dittmer, J. K., and J. E. Szymanski, The stochastic inversion of magnetics and resistivity data using the simulated annealing algorithm, *Geophys. Prospect., 43*(3), 397–416, 1995.

Docherty, P., R. Silva, S. Singh, Z. M. Song, and M. Wood, Migration velocity analysis using a genetic algorithm, *Geophys. Prospect., 45*(5), 865–878, 1997.

Dosso, S. E., and D. W. Oldenburg, Magnetotelluric appraisal using simulated annealing, *Geophys. J. Int., 106*, 370–385, 1991.

Dougherty, D. E., and R. A. Marryott, Optimal groundwater management, 1, Simulated annealing, *Water Resour. Res., 27*(10), 2493–2508, 1991.

Douma, H., R. Snieder, and A. Lomax, Ensemble inference in terms of Empirical Orthogonal Functions, *Geophys. J. Int., 127*, 363–378, 1996.

Drijkoningen, G. G., and R. S. White, Seismic velocity structure of oceanic crust by inversion using genetic algorithms, *Geophys. J. Int., 123*, 653–664, 1995.

Duijndam, A. J. W., Bayesian estimation in seismic inversion part I: Principles, *Geophys. Prospect., 36*, 878–898, 1988a.

Duijndam, A. J. W., Bayesian estimation in seismic inversion part II: Uncertainty analysis, *Geophys. Prospect., 36*, 899–918, 1988b.

Everett, M. E., and A. Schultz, Two-dimensional nonlinear magneto-telluric inversion using a genetic algorithm, *J. Geomagn. Geoelectr., 45*, 1013–1026, 1993.

Flournay, N., and R. K. Tsutakawa (Eds.), Statistical Multiple Integration, in *Proceedings of the AMS-IMS-SIAM Summer Research Conference on Statistical Multiple Integration*, Am. Math. Soc., Providence, R. I., 1989.

Fogel, L. J., Autonomous Automata, *Ind. Res., 4*, 14–19, 1962.

Fogel, L. J., A. J. Owens, and M. J. Walsh, *Artificial Intelligence Through Simulated Evolution*, John Wiley, New York, 1966.

Franklin, J. N., Well posed stochastic extensions of ill posed linear problems, *J. Math. Anal. Appl., 31*, 682–716, 1970.

Gallagher, K., Evolving thermal histories from fission track data, *Earth Planet. Sci. Lett., 136*, 421–435, 1995.

Gallagher, K., and M. Sambridge, Genetic algorithms: A powerful tool for large-scale non-linear optimization problems, *Comput. Geosci., 20*(7/8), 1229–1236, 1994.

Gallagher, K., M. S. Sambridge, and G. G. Drijkoningen, Genetic algorithms: An evolution on Monte Carlo methods in strongly nonlinear geophysical optimization problems, *Geophys. Res. Lett., 18*, 2177–2180, 1991.

Gallagher, K., M. Ramsdale, L. Lonergan, and D. Morrow, The role of thermal conductivity measurements in modelling thermal histories in sedimentary basins, *Mar. Pet. Geol., 14*, 201–214, 1997.

Gelfand, A. E., and A. F. M. Smith, Sampling based approaches to calculating marginal densities, *J. Am. Stat. Assoc., 85*, 398–409, 1990.

Gell-Mann, M., *The Quark and the Jaguar: Adventures in the Simple and the Complex*, W. H. Freeman, New York, 1994.

Gelman, A., J. B. Carlin, H. S. Stern, and D. B. Rubin, *Bayesian Data Analysis*, Chapman and Hall, New York, 1995.

Geman, S., and D. Geman, Stochastic relaxation, Gibbs distributions and the Bayesian restoration of images, *IEEE Trans. Patt. Anal. Mach. Int., 6*, 721–741, 1984.

Gershenfeld, N., *The Nature of Mathematical Modeling*, Cambridge Univ. Press, New York, 1999.

Gilbert, D., and J. Virieux, Electromagnetic imaging and simulated annealing, *J. Geophys. Res., 96*(5), 8057–8067, 1991.

Gill, P. E., W. Murray, and M. H. Wright, *Practical Optimization*, Academic San Diego, Calif., 1981.

Goldberg, D. E., *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading, Mass., 1989.

Goncz, J. H., and J. R. Cleary, Variations in the Structure of the Upper Mantle Beneath Australia, from Rayleigh Wave Observations, *Geophys. J. R. Astron. Soc., 44*, 507–516, 1976.

Goovaerts, P., Stochastic simulation of categorical variables using a classification algorithm and simulated annealing, *Math. Geol., 28*(7), 909–921, 1996.

Gouveia, W. P., and J. A. Scales, Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis, *J. Geophys. Res., 103*(B2), 2759–2779, 1998.

Haddon, R. A. W., and K. E. Bullen, An Earth model incorporating free Earth oscillation data, *Phys. Earth Planet. Inter., 2*, 35–49, 1969.

Hajek, B., Cooling schedules for optimal annealing, *Math. Oper. Res., 13*, 311–329, 1988,

Hall, A., On an experimental determination of π, *Messeng. Math., 2*, 113–114, 1873.

Hammersley, J. M., and D. C. Handscomb, *Monte Carlo Methods*, Chapman and Hall, New York, 1964.

Harland, J. R., and P. Salamon, Simulated annealing: A review of the thermodynamic approach, *Proc. Suppl., 5A*, 109–115, 1988.

Hartzell, S., and P. Liu, Calculation of earthquake rupture histories using a hybrid global search algorithm; application to the 1992 Landers, California, earthquake, *Phys. Earth Planet. Inter., 95*(1–2), 79–99, 1996.

Hastings, W. K., Monte Carlo sampling methods using Markov Chain and their applications, *Biometrika, 57*, 97–109, 1970.

Hermance, J. F., and L. R. Grillot, Constraints of temperatures beneath Iceland from magneto data, *Phys. Earth Planet. Inter., 8*, 1–12, 1974.

Hoffmann, K. H., P. Sibani, J. M. Pedersen, and P. Salamon, Optimal Ensemble Size for Parallel Implementations of Simulated Annealing, *Appl. Math. Lett., 3*, 53–56, 1990.

Holland, J. H., *Adaptation in Natural and Artificial Systems*, Univ. of Mich. Press, Ann Arbor, Mich., 1975.

Horne, S., and C. Macbeth, Inversion for seismic anisotropy using genetic algorithms, *Geophys. Prospect., 42*(8), 953–974, 1994.

Horne, S., and C. Macbeth, A comparison of global optimisation methods for near-offset VSP inversion, *Comput. Geosci., 24*(6), 563–572, 1998.

Ihmle, P., and J. C. Ruegg, Source tomography by simulated annealing using broad-band surface waves and geodetic data; application to the M_w = 8.1 Chile 1995 event, *Geophys. J. Int., 131*, 146–158, 1997.

Ingber, L., Very fast simulated reannealing, *Math. Comput. Model., 12*, 967–993, 1989.

Ingber, L., and B. Rosen, Genetic algorithms and simulated reannealing: A comparison, *Math. Comput. Model., 16*, 87–100, 1992.

Jackson, D. D., Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. Astron. Soc., 28*, 97–110, 1972.

Jackson, D. D., Marginal solutions to quasi-linear inverse problems in geophysics, *Geophys. J. R. Astron. Soc., 12*, 165–180, 1973.

Jackson, D. D., Most squares inversion, *J. Geophys. Res., 81*, 1027–1030, 1976.

Jackson, D. D., The use of a priori data to resolve nonuniqueness in linear inversion, *Geophys. J. R. Astron. Soc., 57*, 137–157, 1979.

Jackson, D. D., and M. Matsu'ura, A Bayesian approach to nonlinear inversion, *J. Geophys. Res., 90*, 581–591, 1985.

Jakobsen, M. O., M. Mosegaard, and J. M. Pedersen, Global

model optimization in reflection seismology by simulated annealing, in *Model Optimization in Exploration Geophysics, 2, Proceedings of the 5th International Mathematical Geophysics Seminar*, vol. 2, edited by A. Vogel, Friedr. Vieweg and Sohn, pp 361–381, Braunschweig, Germany, 1988.

Jervis, M., P. L. Stoffa, and M. K. Sen, 2-D migration velocity estimation using a genetic algorithm, *Geophys. Res. Lett., 20*(14), 1495–1498, 1993.

Jestin, F., P. Huchon, and J. M. Gaulier, The Somalia Plate and the East African rift system: Present-day kinematics, *Geophys. J. Int., 116*, 637–654, 1994.

Jin, S., and R. Madariaga, Background velocity inversion with a genetic algorithm, *Geophys. Res. Lett., 20*(2), 93–96, 1993.

Jin, S., and R. Madariaga, Nonlinear velocity inversion by a two-step Monte Carlo method, *Geophysics, 59*(4), 577–590, 1994.

Jones, A. G., and R. Hutton, A multi-station magnetotelluric study in southern Scotland, II, Monte-Carlo inversion of the data and its geophysical and tectonic implications, *Geophys. J. R. Astron. Soc., 56*, 351–368, 1979.

Jones, R. H., C. M. Rayne, and U. Lindblom, The use of a genetic algorithm for the optimal design of microseismic monitoring networks, in *Rock Mechanics in Petroleum Engineering Proceedings*, A. A. Balkema, Brookfield, Vt., pp. 615–619, 1994.

Jordan, T. H., and J. Franklin, Optimal solutions to a linear inverse problem in geophysics, *Proc. Natl. Acad. Sci., 70*, 1410–1413, 1971.

Keilis-Borok, V. I., and T. B. Yanovskaya, Inverse problems of seismology, *Geophys. J., 13*, 223–234, 1967.

Kelvin, Lord, Nineteenth century clouds over the dynamical theory of heat and light, *Philos. Mag., 2*(6), 1–40, 1901.

Kennett, B. L. N., Some aspects of non-linearity in inversion, *Geophys. J. R. Astron. Soc., 55*, 373–391, 1978.

Kennett, B. L. N., and G. Nolet, Resolution analysis for discrete systems, *Geophys. J. R. Astron. Soc., 53*, 413–425, 1978.

Kennett, B. L. N., On the density distribution within the Earth, *Geophysics. J. Int., 132*, 374–382, 1998.

Kennett, B. L. N., and E. R. Engdahl, Travel times for global earthquake location and phase identification, *Geophysics. J. Int., 105*, 429–466, 1991.

Kennett, B. L. N., and M. Sambridge, Earthquake location: Genetic algorithms for teleseisms, *Phys. Earth Planet. Inter., 75*, 103–110, 1992.

Kido, M., D. A. Yuen, O. Cadek, Mantle viscosity derived by genetic algorithm using oceanic geoid and seismic tomography for whole-mantle versus blocked-flow situations, *Phys. Earth Planet. Inter., 107*, 307–326, 1998.

King, S. D., Radial models of mantle viscosity: Results from a genetic algorithm, *Geophys. J. Int., 122*, 725–734, 1995.

Kirkpatrick, S. C., D. Gelatt, and M. P. Vecchi, Optimization by simulated annealing, *Science, 220*, 671–680, 1983.

Kobayashi, R., and I. Nakanishi, Application of genetic algorithms to focal mechanism determination, *Geophys. Res. Lett., 21*(8), 729–732, 1994.

Koper, K. D., M. E. Wysession, and D. A. Wiens, Multimodal function optimization with a Niching genetic algorithm: A seismological example, *Bull. Seismol. Soc. Am., 89*, 978–988, 1999.

Koren, Z., K. Mosegaard, E. Landa, P. Thore, and A. Tarantola, Monte Carlo estimation and resolution analysis of seismic background velocities, *J. Geophys. Res., 96*(B12), 20,289–20,299, 1991.

Kuo, C. H., A. N. Michel, and W. G. Gray, Design of optimal pump-and-treat strategies for contaminated ground water remediation using the simulated annealing algorithm, *Adv. Water Resour., 15*(2), 95–105, 1992.

Landa, E., W. Beydoun, and A. Tarantola, Reference velocity model estimation from prestack waveforms: Coherency optimization by simulated annealing, *Geophys., 54*(8), 984–990, 1989.

Lepage, G. P., A new algorithm for adaptive multi-dimensional integration, *J. Comput. Phys., 27*, 192–203, 1978.

Levin, V., and J. Park, P-SH conversions in a flat-layered medium with anisotropy of arbitrary orientation, *Geophys. J. Int., 131*, 253–266, 1997.

Lomax, A., and R. Snieder, Finding sets of acceptable solutions with a genetic algorithm with application to surface wave group dispersion in Europe, *Geophys. Res. Lett., 21*(24), 2617–2620, 1994.

Lomax, A., and R. Snieder, Identifying sets of acceptable solutions to nonlinear geophysical inverse problems which have complicated misfit functions, *Nonlinear Proc. Geophys., 2*, 222–227, 1995a.

Lomax, A., and R. Snieder, The contrast in upper mantle shear-wave velocity between the East European Platform and tectonic Europe obtained with genetic algorithm inversion of Rayleigh-wave group dispersion, *Geophys. J. Int., 123*, 169–182, 1995b.

Lomax, A., J. Virieux, P. Volant, and C. Berge, Probabilistic earthquake location in 3D and layered models: Introduction of a Metropolis-Gibbs method and comparison with linear locations, in *Advances in Seismic Event Location*, edited by C. Thurber, and N. Rabinowitz, Kluwer, Acad. Norwell, Mass., 2000.

Mathias, K., and D. L. Whitley, Genetic Operators, the Fitness Landscape and the Traveling Salesman Problem, in *Parallel Problem Solving from Nature-PPSN 2*, edited by R. Manner and B. Manderick, pp. 219–228, Elsevier Sci., New York, 1992.

McKinney, D. C., and M. D. Lin, Genetic algorithm solution of groundwater management models, *Water Resour. Res., 30*(6), 1897–1906, 1994.

Mellman, G. R., A method of body-wave waveform inversion for determination of Earth structure, *Geophys. J. R. Astron. Soc., 62*, 481–504, 1980.

Menke, W., *Geophysical Data Analysis: Discrete Inverse Theory*, rev. ed., Academic, San Diego, Calif., 1989.

Metropolis, N., and S. M. Ulam, The Monte Carlo method, *J. Am. Stat. Assoc., 44*, 335–341, 1949.

Metropolis, N., M. N. Rosenbluth, A. W. Rosenbluth, A. H. Teller, and E. Teller, Equation of state calculations by fast computing machines, *J. Chem. Phys., 21*, 1087–1092, 1953.

Mills, J. M., and T. J. Fitch, Thrust faulting and crust-upper mantle structure in East Australia, *Geophys. J. R. Astron. Soc., 48*, 351–384, 1977.

Mills, J. M., and A. L. Hales, Great-circle Rayleigh wave attenuation and group velocity, Part III, Inversion of global average group velocities and attenuation coefficients, *Phys. Earth Planet. Inter., 17*(4), 307–322, 1978.

Minster, J.-B. H., N. P. Williams, T. G. Masters, J. F. Gilbert, and J. S. Haase, Application of evolutionary programming to earthquake hypocenter determination, in *Proceedings of 4th Annual Conference of Evol. Prog.*, edited by J. R. McDonnell, R. C. Reynolds, and D. B. Fogel, pp. 3–17, MIT Press, 1995.

Morokoff, W. J., and R. E. Caflisch, Quasi-random sequences and their discrepancies, *SIAM J. Sci. Comput., 15*(6), 1251–1279, 1994.

Morshed, J., and J. J. Kaluarachchi, Parameter estimation using artificial neural network and genetic algorithm for free-product migration and recovery, *Water Resour. Res., 34*(5), 1101–1113, 1998.

Mosegaard, K., Resolution analysis of general inverse problems through inverse Monte Carlo sampling, *Inverse Probl., 14*, 405–426, 1998.

Mosegaard, K., and C. Rygaard-Hjalsted, Bayesian analysis of implicit inverse problems, *Inverse Probl., 15*, 573–583, 1999.

Mosegaard, K., M. Sambridge, Monte Carlo analysis of inverse problems, *Inverse Probl., 18*, R29–R54, 2002.

Mosegaard, K., and A. Tarantola, Monte Carlo sampling of solutions to inverse problems, *J. Geophys. Res., 100*(B7), 12,431–12,447, 1995.

Mosegaard, K., and P. Vestergaard, A simulated annealing approach to seismic model optimization with sparse prior information, *Geophys. Prospect., 39*(5), 599–611, 1991.

Mosegaard, K., S. C. Singh, D. Snyder, and H. Wagner, Monte Carlo analysis of seismic reflections from Moho and the W-reflector, *J. Geophys. Res., 102*, 2,969–2,981, 1997.

Muramatsu, Y., and I. Nakanishi, Hypocenter determination using a genetic algorithm, *J. Natl. Disaster Sci., 19*(1), 1–7, 1997.

Muttiah, R. S., B. A. Engel, and D. D. Jones, Waste disposal site selection using GIS-based simulated annealing, in *Neural Network Applications in the Geosciences*, edited by B. G. Lees, *Comput. Geosci., 22*(9), 1013–1017, 1996.

Neves, F. A., S. C. Singh, and K. F. Priestley, Velocity structure of upper-mantle transition zones beneath central Eurasia from seismic inversion using genetic algorithms, *Geophys. J. Int., 125*, 869–878, 1996.

Nolte, B., and L. N. Frazer, Vertical seismic profile inversion with genetic algorithms, *Geophys. J. Int., 117*, 162–179, 1994.

Normark, E., and K. Mosegaard, Residual statics estimation: Scaling temperature schedules using simulated annealing, *Geophys. Prospect., 41*(5), 565–578, 1993.

Nulton, J. D., and P. Salamon, Statistical mechanics of combinatorial optimization, *Phys. Rev., A37*, 1351–1356, 1988.

Pardo, I. E., Optimal selection of number and location of rainfall gauges for areal rainfall estimation using geostatistics and simulated annealing, *J. Hydrol., 210*, 206–220, 1998.

Park, S. K., and K. W. Miller, Random number generators: Good ones are hard to find, *Commun. ACM, 31*, 1192–1201, 1988.

Parker, R. L., Understanding inverse theory, *Annu. Rev. Earth Planet. Sci., 5*, 35–64, 1977.

Parker, R. L., *Geophysical Inverse Theory*, Princeton Univ. Press, Princeton, N. J., 1994.

Press, F., Earth models obtained by Monte Carlo inversion, *J. Geophys. Res., 73*, 5223–5234, 1968.

Press, F., Earth models consistent with geophysical data, *Phys. Earth Planet. Inter., 3*, 3–22, 1970a.

Press, F., Regionalized Earth models, *J. Geophys. Res., 75*, 6575–6581, 1970b.

Press, W. H., B. P. Flannery, A. T. Saul, and W. T. Vetterling, *Numerical Recipes*, 2nd ed., Cambridge Univ. Press, New York, 1992.

Pullammanappallil, S. K., and J. N. Louie, Inversion of seismic reflection traveltimes using a nonlinear optimization scheme, *Geophys., 58*(11), 1607–1620, 1993.

Pullammanappallil, S. K., and J. N. Louie, A generalized simulated-annealing optimization for inversion of first-arrival times, *Bull. Seismol. Soc. Am., 84*(5), 1397–1409, 1994.

Rawlins, G. J. E., (Ed.), *Foundations of Genetics Algorithms*, Morgan Kaufmann Burlington, Mass., 1991.

Ricard, Y., C. Vigny, and C. Froidevaux, Mantle heterogeneities, geoid, and plate motion: A Monte Carlo inversion, *J. Geophys. Res., 94*, 13,739–13,754, 1989.

Ritzel, B. J., J. W. Eheart, and S. Ranjithan, Using genetic algorithms to solve a multiple objective groundwater pollution containment problem, *Water Resour. Res., 30*(5), 1589–1603, 1994.

Rizzo, D. M., and D. E. Dougherty, Design optimization for multiple management period groundwater remediation, *Water Resour. Res., 31*(8), 2549–2561, 1996.

Rogers, L. L., F. U. Dowla, and V. M. Johnson, Optimal field-scale groundwater remediation using neural networks and the genetic algorithm, *Environ. Sci. Tech., 29*(5), 1145–1155, 1995.

Rogers, L. L., V. M. Johnson, and R. B. Knapp, Remediation tradeoffs addressed with simulated annealing optimization, in *Computational Methods in Contamination and Remediation of Water Resources*, edited by V. N. Burganos, et al., *Comput. Methods Water Resour., 12*(1), 59–66, Comput. Mech. Publ., Billerica, Mass., 1998.

Rothman, D. H., Nonlinear inversion statistical mechanics, and residual statics corrections, *Geophysics, 50*, 2784–2796, 1985.

Rothman, D. H., Automatic estimation of large residual statics corrections, *Geophysics, 51*, 332–346, 1986.

Ruppeiner, G., J. M. Pedersen, and P. Salamon, Ensemble approach to simulated annealing, *J. Phys., 1*(1), 455–470, 1991.

Sabatier, P. C., On geophysical inverse problems and constraints, *J. Geophys., 43*, 115–137, 1977a.

Sabatier, P. C., Positivity constraints in linear inverse problems, I, General theory, *Geophys. J. R. Astron. Soc., 48*, 415–441, 1977b.

Sabatier, P. C., Positivity constraints in linear inverse problems, II, Applications, *Geophys. J. R. Astron. Soc., 48*, 443–459, 1977c.

Sadeghi, H., S. Suzuki, H. Takenaka, A two-point, three-dimensional seismic ray tracing using genetic algorithms, *Phys. Earth Planet. Inter., 113*, 355–365, 1999.

Sambridge, M., Exploring multi-dimensional landscapes without a map, *Inverse Probl., 14*(3), 427–440, 1998.

Sambridge, M., Geophysical inversion with a Neighbourhood algorithm, I, Searching a parameter space, *Geophys. J. Int., 138*, 479–494, 1999a.

Sambridge, M., Geophysical inversion with a neighbourhood algorithm, II, Appraising the ensemble, *Geophys. J. Int., 138*, 727–746, 1999b.

Sambridge, M., and G. G. Drijkoningen, Genetic algorithms in seismic waveform inversion, *Geophys. J. Int., 109*, 323–342, 1992.

Sambridge, M., and K. Gallagher, Hypocentre location using genetic algorithms, *Bull. Seismol. Soc. Am., 83*(5), 1467–1491, 1993.

Sambridge, M., and B. L. N. Kennett, A novel method of hypocentre location, *Geophys. J. R. Astron. Soc., 87*, 679–697, 1986.

Scales, J. A., and R. Snieder, To Bayes or not to Bayes, *Geophysics, 62*, 1045–1046, 1997.

Scales, J. A., M. L. Smith, and T. L. Fischer, Global optimization methods for multimodal inverse problems, *J. Comput. Phys., 103*(2), 258–268, 1992.

Scherbaum, F., C. Palme, and H. Langer, Model parameter optimization for site-dependent simulation of ground motion by simulated annealing: Re-evaluation of the Ashigara Valley prediction experiment, *Nat. Hazards, 10*(3), 275–296, 1994.

Sen, M. K., and P. L. Stoffa, Nonlinear one-dimensional seismic waveform inversion using simulated annealing, *Geophysics, 56*, 1624–1638, 1991.

Sen, M. K., and P. L. Stoffa, Rapid sampling of model space using genetic algorithms, *Geophys. J. Int., 108*, 281–292, 1992.

Sen, M. K., and P. L. Stoffa, *Global Optimization Methods in Geophysical Inversion*, Adv. Explor. Geophys., vol. 4, Elsevier Sci., New York, 1995.

Sharma, S. P., and P. Kaikkonen, Two-dimensional non-linear inversion of VLF-R data using simulated annealing, *Geophys. J. Int., 133*, 649–668, 1998.

Shibutani, T., M. Sambridge, and B. L. N. Kennett, Genetic algorithm inversion for receiver functions with application

to crust and uppermost mantle structure beneath eastern Australia, *Geophys. Res. Lett., 23*, 1829–1832, 1996.

Sileny, J., Earthquake source parameters and their confidence regions by a genetic algorithm with a "memory," *Geophys. J. Int., 134*, 228–242, 1998.

Simpson, A. R., and S. D. Priest, The application of genetic algorithms to optimisation problems in geotechnics, *Comput. Geotech., 15*(1), 1–19, 1993.

Smith, A. F. M., Bayesian computational methods, *Philos. Trans. R. Soc. London Ser. A, 337*, 369–386, 1991.

Smith, A. F. M., and G. O. Roberts, Bayesian Computation via the Gibbs Sampler and Related Markov Chain Monte Carlo Methods, *J. R. Stat. Soc, Ser. B., 55*, 3–23, 1993.

Smith, M. L., J. A. Scales, and T. L. Fischer, Global search and genetic algorithms, *Leading Edge, 11*, 22–26, 1992.

Snieder, R., and J. Trampert, Inverse problems in geophysics, in *Wavefield Inversion*, edited by A. Wirgin, pp. 119–190, Springer-Verlag, New York, 1999.

Sobol, I. M., On the distribution of points in a cube and the approximate evaluation of integrals, *USSR Comput. Math. Math. Phys., 7*(4), 86–112, 1967.

Stark, P. B., Minimax confidence intervals in geomagnetism, *Geophys. J. Int., 108*, 329–338, 1992.

Steck, L. K., Simulated annealing inversion of teleseismic P-wave slowness and azimuth for crustal velocity structure at Long Valley Caldera, *Geophys. J. Int., 22*, 497–500, 1995.

Stoffa, P. L., and M. K. Sen, Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms, *Geophys, 56*, 1794–1810, 1991.

Szu, H., and R. Hartley, Fast simulated annealing, *Phys. Lett., Ser., A. 122*, 157–162, 1987.

Tang, A., and L. W. Mays, Genetic algorithms for optimal operation of soil aquifer treatment systems, *Water Resour. Manage., 12*(5), 375–396, 1998.

Tanner, M. A., *Tools for Statistical Inference: Methods for the Exploration of Posterior Distributions and Likelihood Functions*, 3rd ed., Springer-Verlag, New York, 1996.

Tarantola, A., *Inverse Problem Theory*, Elsevier Sci., New York, 1987.

Tarantola, A., and B. Valette, Inverse problems = quest for information, *J. Geophys., 50*, 159–170, 1982a.

Tarantola, A., and B. Valette, Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys., 20*(2), 219–232, 1982b.

Thomson, W. E., Analogue computers for stochastic processes, in *Abbrev. Processes of the Oxford Math, Conference for Schoolteachers and Industrialists*, pp. 54–57, The Times Publ., London, 1957.

Tierney, L., Markov chains for exploring posterior distributions, *Ann. Stat., 22*, 1701–1762, 1994.

Valius, V. P., Opredelenie seismicheskikh razrezov po sovokupnosti nabludenii, (Determination of seismological sections through ensemble control), *Vichislitelnaya Seismol., 4*, 3–14, 1968.

van der Beek, P., and J. Braun, Control on post-mid Cretaceous landscape evolution in the southeastern highlands of Australia: Insights from numerical surface process models, *J. Geophys. Res., 104*, 4945–4966, 1999.

Varela, C. L., P. L. Stoffa, and M. K. Sen, Background velocity estimation using non-linear optimization for reflection tomography and migration misfit, *Geophys. Prospect., 46*(1), 51–78, 1998.

Vasco, D. W., L. R. Johnson, and E. L. Majer, Ensemble inference in geophysical inverse problems, *Geophys. J. Int., 117*, 711–728, 1993.

Vasudevan, K., and W. G. Wilson, Simulated annealing statics computation using an order-based energy function, *Geophys., 56*(11), 1831–1839, 1991.

Velis, D. R., and T. J. Ulrych, Simulated annealing two-point ray tracing, *Geophys. Res. Lett., 32*(2), 201–204, 1996.

Vestergaard, P., and K. Mosegaard, Inversion of post-stack seismic data using simulated annealing, *Geophys. Prospect., 39*(5), 613–624, 1991.

Vinther, R., and K. Mosegaard, Seismic inversion through Tabu Search, *Geophys. Prospect., 44*(4), 555–570, 1996.

Wan, Y., R. Liu, and H. Li, The inversion of 3-D crustal structure and hypocenter location in the Beijing-Tianjin-Tangshan-Zhangjiakou area by genetic algorithm, *Acta Seismol. Sin., 10*(6), 769–781, 1997.

Whitley, D. L., A genetic algorithm tutorial, *Stat. Comput., 4*, 65–85, 1994.

Wiggins, R. A., Monte Carlo inversion of body wave observations, *J. Geophys. Res., 74*, 3171–3181, 1969.

Wiggins, R. A., The general inverse problem: Implication of surface waves and free oscillations for Earth Structure, *Rev. Geophys., 10*, 251–285, 1972.

Wilson, W. G., K. Vasudevan, Application of the genetic algorithm to residual statics estimation, *Geophys. Res. Lett., 18*, 2181–2184, 1991.

Winter, G., J. Periaux, M. Galan, and P. Cuesta, (Eds.), *Genetic Algorithms in Engineering and Computer Science*, John Wiley, New York, 1995.

Worthington, M. H., J. R. Cleary, and R. S., Anderssen, Density modelling by Monte Carlo inversion, II, Comparison of recent Earth models, *Geophys. J. R. Astron. Soc., 29*, 445–457, 1972.

Worthington, M. H., J. R. Cleary, and R. S., Anderssen, Upper and lower mantle shear velocity modelling by Monte Carlo inversion, *Geophys. J. R. Astron. Soc., 36*, 91–103, 1974.

Yamanaka, H., and H. Ishida, Applicaton of genetic algorithms to an inversion of surface-wave dispersion data, *Bull. Seismol. Soc. Am., 86*(2), 436–444, 1996.

Yu, T. T., J. Fernandez, and J. B. Rundle, Inverting the parameters of an earthquake-ruptured fault with a genetic algorithm, *Comput. and Geosci., 24*(2), 173–182, 1998.

Zhao, L. S., M. K. Sen, P. L. Stoffa, and C. Frohlich, Application of very fast simulated annealing to the determination of the crustal structure beneath Tibet, *Geophys. J. Int., 125*, 355–370, 1996.

Zheng, C., P. Wang, Parameter structure identification using tabu search and simulated annealing, *Adv. Water Resour., 19*(4), 215–224, 1996.

Zhou, R., F. Tajima, and P. L. Stoffa, Earthquake source parameter determination using genetic algorithms, *Geophys. Res. Lett., 22*(4), 517–520, 1995a.

Zhou, R., F. Tajima, and P. L. Stoffa, Application of genetic algorithms to constrain near-source velocity structure for the 1989 Sichuan earthquakes, *Bull. Seismol. Soc. Am., 85*(2), 590–605, 1995b.

K. Mosegaard, Neils Bohr Institute for Astronomy, Physics, and Geophysics, Department of Geophysics, University of Copenhagen, Juliane Maries Vej 30, DK-2100 Copenhagen 0, Denmark. (Klaus@gfy.ku.dk)

M. Sambridge, Research School of Earth Sciences, Institute of Advanced Studies, Australian National University, Canberra ACT 0200, Australia. (malcolm@rses.anu.edu.au)