

Probabilistic characterization of geologic and geophysical information

John A. Scales

ABSTRACT

It is useful to think of inverse calculations as the process of quantifying states of information about physical systems from data. The more successful the calculation, the more the uncertainty about the system is reduced relative to the state of information that existed before the data were gathered. It is impossible to do such a calculation without making *a priori* statistical assumptions about models and data. Typically, it is assumed that these *a priori* distributions are normally distributed, whether explicitly through the use of *a priori* covariances, or implicitly, with regularized least squares. Here we look at some examples of how one might use probabilistic ideas to quantify geophysical *a priori* information. In one case we compute covariance matrices associated with ambient fluctuations in seismic data, and in the other we estimate nonparametrically the *a priori* distribution of layered earth models directly from a well log. From the standpoint of inversion, the net result, whether the distributions are Gaussian or not, is a quantitative comparison of the posterior versus prior uncertainty of each parameter in the calculation. This we regard as a complete solution to the inverse calculation.

THE USE OF PROBABILITY IN INVERSE THEORY

It is possible to give different interpretations to any abstract theory. Probability theory, for instance, can be interpreted in terms of the frequencies of outcomes of random events. But we also use probabilistic concepts to quantify states of information about completely deterministic events. Suppose we ask, for example, the probability of encountering hydrocarbons upon drilling to a certain depth at some location. Presumably this is a deterministic event: either we will or we will not find the target. But we use probability to quantify the likelihood of various outcomes. Assigning a probability to such questions must take into account pre-existing geological and geophysical information: what is known about the geology of the area; its past production history; what other data are available in the area; what experts say. Such information is not uniquely specifiable. In the first place, some of it subjective, such as expertise. But different people will have different data and information at their

disposal. Thus, the ultimate assessment of the likelihood of success in the experiment is not uniquely determined either. But given certain prior information, we should be able to work out the posterior probability of any given hypothesis concerning the experiment.

Throughout this paper the terms prior and posterior are used to denote the states of information, respectively, before and after any inferences are made. Prior information is therefore any information which is independent of the data. The information content of a distribution is related to how broad it is. A flat distribution is relatively uninformative, while a delta function is perfectly informative. These qualitative ideas can be made quantitative by introducing more rigorous definitions of information in terms of entropy.

UNCERTAINTIES IN INVERSE CALCULATIONS

It is impossible to do an inverse calculation without making *a priori* assumptions. These assumptions may have to do with the distribution of possible models, the uncertainties in the data, or the nature of the mapping from the space of model vectors to the space of data vectors. The most common formulations are in terms of least-squares misfit functions. Least-squares presupposes that all uncertainties in the problem are Gaussian: data uncertainties, theoretical uncertainties (i.e., uncertainties in the forward modeling), and uncertainties in the distribution of feasible models. In practice, even stronger assumptions are often made, namely that the uncertainties are uncorrelated; an example would be damped least squares. Even if these assumptions are regarded as mere convenience, they have important implications for the reliability of the resulting calculations.

Consider a simple example. Let's suppose we are doing a damped least-squares calculation for a reflectivity series and are assuming, therefore, that the set of possible solutions is distributed in an uncorrelated Gaussian fashion about the starting model. Imagine that we have a black box that will select models from this Gaussian *a priori* distribution. We then evaluate the response of each model (synthetic data) and see how well it fits the observed data. We do as many iterations of this procedure as is necessary to find a model that agrees with the data to our satisfaction. Perhaps this solution looks like the top left plot in Figure 1. There's only one problem: the top right plot in this figure has the same mean, standard deviation, and 1-D probability distribution. This means that with the same prior information of a Gaussian distribution, the black box could have been programmed to generate models that look like that on the left of Figure 1 rather than the right.

The rather different look of these two pseudo-random models lies in their different correlation length. The model on the left is uncorrelated (white) and the model on the right has a correlation length of 10 samples. This can be seen from the autocorrelation functions shown in the middle of the figure. So both models are consistent with the *a priori* assumptions of a Gaussian distribution, yet they look completely different. We could carry this argument to extremes by making the pseudo-random models more

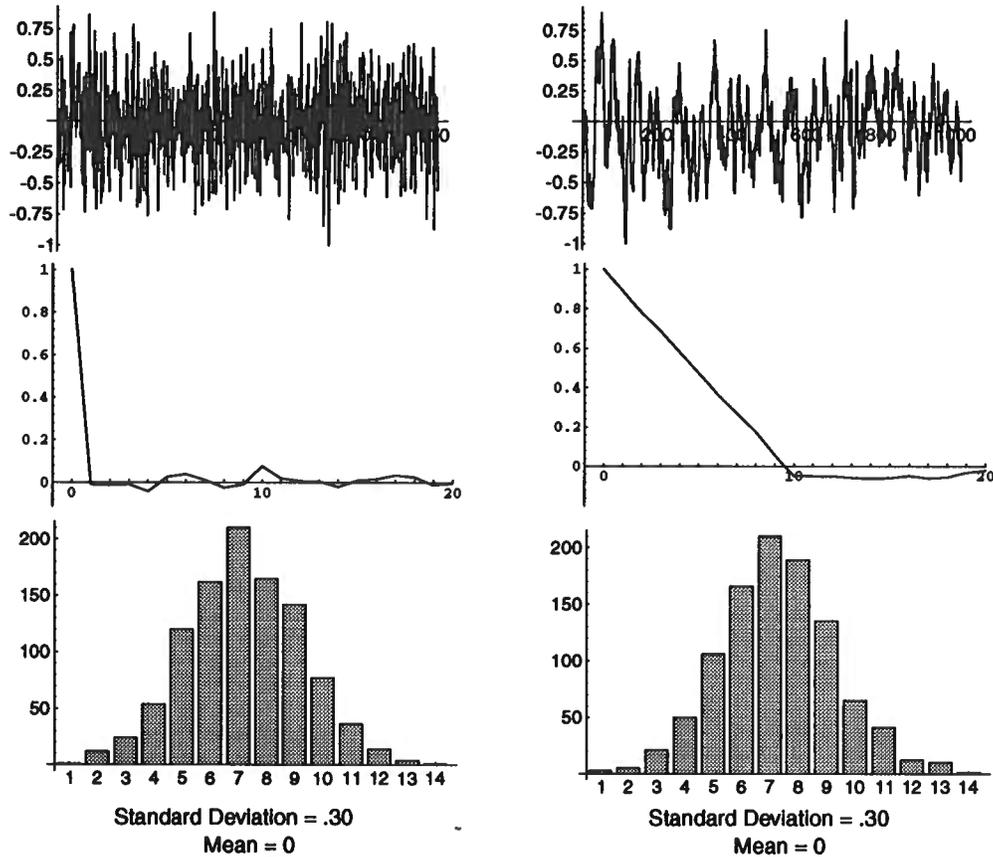


FIG. 1. Two pseudo-random Gaussian simulations with approximately the same mean, standard deviation and one-dimensional marginals, but different correlation lengths. The middle figures shown the correlation function (autocorrelation).

and more correlated while still maintaining the same 1-D distribution. Of course, we cannot say that models with different correlation lengths would fit the data equally well, but they would be equally consistent with the *a priori* assumptions behind damped least squares.

Data Uncertainties

To see whether this is an issue in practice Figure 2 shows 25 unstacked, uncorrelated vibroseis traces. For better or worse, these will be assumed to be “raw” data. These data were recorded with a 3 second 40-60 Hz sweep at far offset using a horizontally polarized source. To see the ambient noise, we have windowed the first second of data. Figure 3 shows the autocorrelation and power spectra of the first 6 traces windowed to the first second of data. Clearly these data are correlated vertically. Typical correlation lengths are on the order of 15-20 samples. Since we have 25 realizations of each sweep, we can estimate the sample covariance matrix by summing over this ensemble. We call this C_d and it represents what we know about observational errors

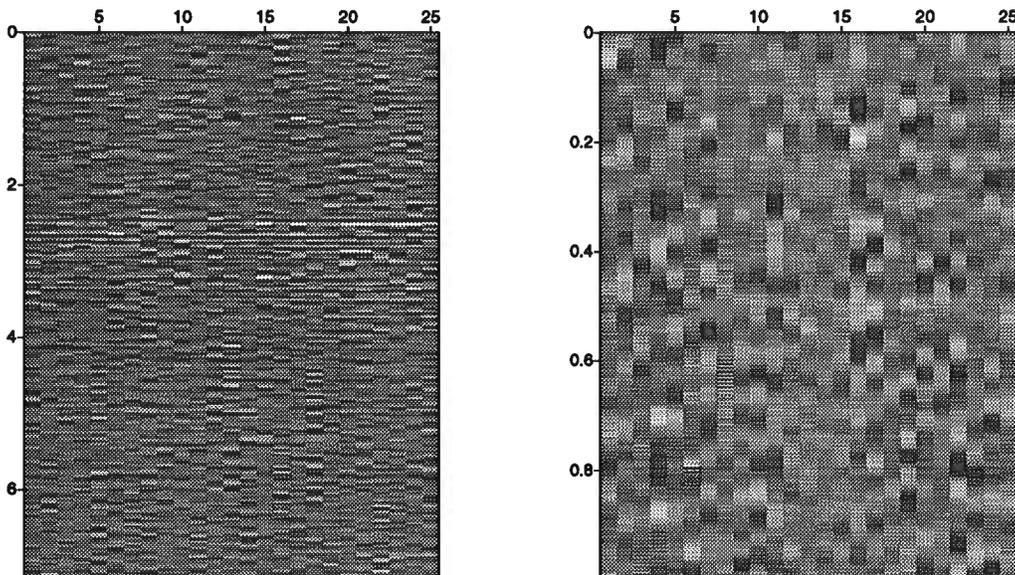


FIG. 2. Unstacked vibroseis traces. Can we regard these as independent realizations of the experiment? On the right, is a zoom of the 0-1 second time window.

after the experiment (within the Gaussian approximation). Figure 4 shows a stack of the noise traces (in the 0-1 second window), the power spectrum and autocorrelation of the stacked trace, and the first trace in the ensemble. Clearly stacking has attenuated the high frequency noise somewhat, but the stacked trace still shows a correlation length of around 15 samples.

So far we have refrained from defining “noise”, for the very good reason that there is no precise definition. By windowing the vibroseis traces to a time interval before any energy from the source could have arrived, we hope that the resulting data gives a good picture of the background signal. But clearly there is coherent energy even in these early samples. Do we call this noise? Unfortunately there is no completely satisfying answer to this question. Essentially we call it noise if we have no interest in inferring its cause or using it directly in the inversion procedure. If the high frequency environmental contamination we are seeing in the traces is a nuisance, to be filtered out perhaps, then we shall call it noise. But who is to say that someone might not find interesting “signal” in this noise after all, in which case this physics would have to be modeled in the inversion process also. Thus any covariance matrix associated with unmodeled physics should add to the covariance matrix associated with the random fluctuations in the data themselves. We can lump these together into a data covariance $C_D = C_d + C_T$, where we have used T to denote theoretical errors.

These data errors enter into the theory via the *Likelihood* function L , which measures the degree to which models fit data. Tarantola (1987) shows (in exercise 1.21)

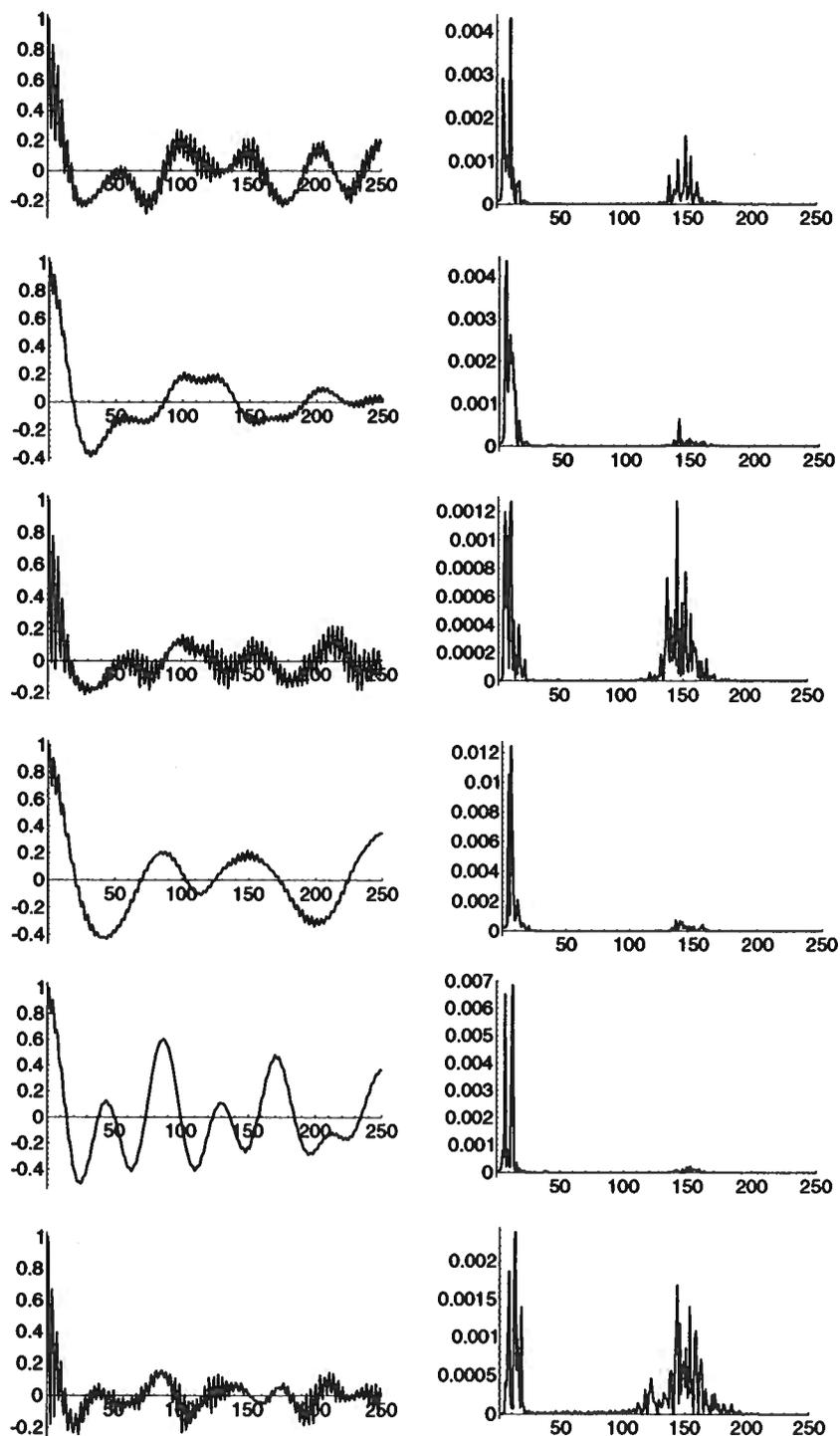


FIG. 3. Normalized autocorrelation functions of the first 6 traces windowed to the 0-1 second interval. The correlation length is typically 10-20 samples with some almost monochromatic high frequency noise. To the right of each autocorrelation is the power spectrum, with abscissa units in samples (.002 s/sample).

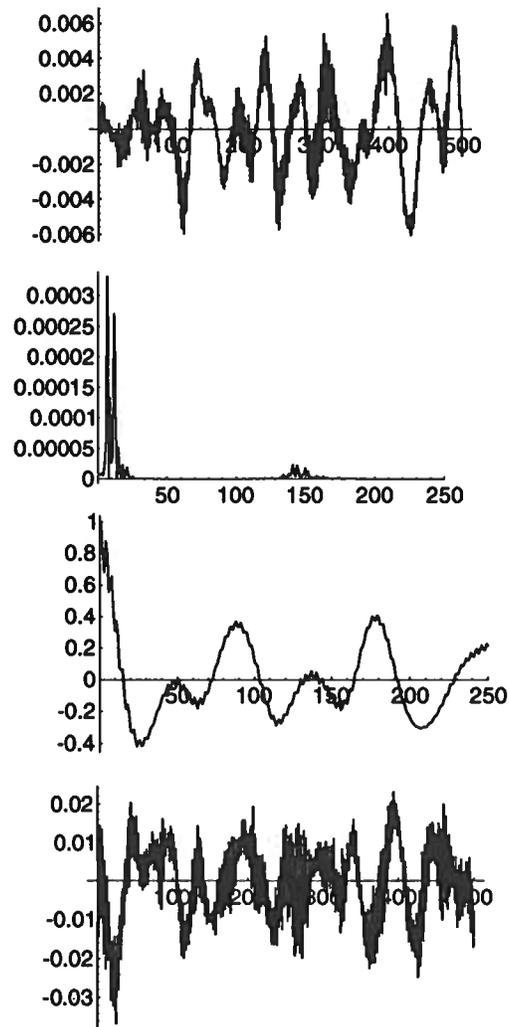


FIG. 4. Stack of the 25 noise traces (0-1 second window), the normalized autocorrelation of the stacked trace and the power spectrum (abscissa in samples, 2 ms per sample). Trace number 1 is at the bottom for comparison. Stacking attenuates about 80% of the energy in the 0-1 second time window. The average vertical correlation is 15 samples.

that the complete likelihood function, assuming Gaussian errors on both data and modeling is:

$$L(\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^N \det C_D}} \exp \left[-\frac{1}{2} (g(\mathbf{m}) - \mathbf{d}_{\text{obs}})^T C_D^{-1} (g(\mathbf{m}) - \mathbf{d}_{\text{obs}}) \right] \quad (1)$$

where g is the *forward problem*, mapping elements of the model space into data space, \mathbf{d}_{obs} is a vector of observations and N is the number of observations. Notice that the covariance matrices C_d and C_T add together to achieve the final result even though we have made no assumptions about the linearity of g . In the special case that g is linear (i.e., a matrix A) and C_D is a constant times the identity matrix, then the maximum likelihood model is the solution of the ordinary least squares problems $\min \|\mathbf{A}\mathbf{m} - \mathbf{d}\|^2$. If C_D is more complicated, we get a weighted least squares problem.

The point in all of this is not to try to suppress noise, which requires some understanding of its sources, but to quantify the noise. It's not as if there are good data and bad data, just data with different degrees of certainty.

Model Uncertainties

As we observed above, doing conventional least-squares makes assumptions not only about the distribution of data uncertainties, but also about the uncertainties in the models themselves. When we “damp” a calculation, we are actually requiring that the computed model lie in a Gaussian neighborhood of the starting model. And while it may be plausible that data uncertainties (at least the “random” ones) can be described by a Gaussian distribution this seems harder to justify for the models themselves.

Let us consider an example of how we might do an inverse calculation without assuming a Gaussian distribution of models. Suppose we wish to invert reflection seismic data with a nearby sonic log as *a priori* information about the distribution of feasible 1-D earth models. Figure 5 shows about 600 m of P-wave sonic log recorded in the North Sea (1 sample every 30 cm). The middle plot is obtained by applying a running average to the data. The bottom plot shows the fluctuations obtained by subtracting the trend from the data. We regard the fluctuations as representing a single realization of a random vector field. The trend will be assumed to be known. Our job now is to characterize the statistical properties of the fluctuations as generally as possible.

Whereas for Gaussian or Markovian processes, marginal distributions of first and second order are sufficient to characterize the process completely, for a general random vector field one must specify the joint n -dimensional distribution function. While it is clearly impossible to construct the joint distribution function for anything as complicated as a well log, it is possible to construct enough of it to fool a standard statistical hypothesis test into thinking that you have computed it all. In Scales and Tarantola (1994) a technique is developed for computing low order marginal

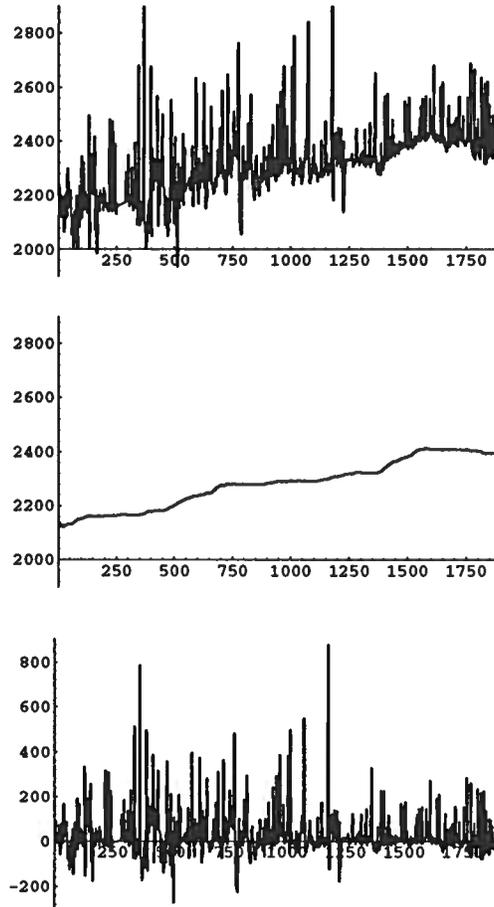


FIG. 5. North Sea P-wave sonic log, trend and fluctuations.

distributions from data by making histograms. We then generate pseudo-random realizations of the process from the histograms and compare them to the data using the Kolmogorov-Smirnov two-sample test, the null-hypothesis being that the two samples are drawn from the same distribution. When we have computed enough of the marginals to satisfy K-S with a high degree of confidence, then we stop. The resulting marginals represent a non-parametric description of the statistics of the underlying process. Some examples of pseudo-random logs generated according to this procedure, as well as the original data, are shown in Figure 6.

So, we have a method for approximating the joint distribution function of the well log, but how can we use this to actually solve an inverse problem? To answer this we need a few more tools.

Above, we showed an expression for the Gaussian likelihood function. The general form of the likelihood function is

$$L(\mathbf{m}) = \int_D \frac{\rho_D(\mathbf{d})\Theta(\mathbf{d}|\mathbf{m})}{\mu_D(\mathbf{d})} d\mathbf{d} \quad (2)$$

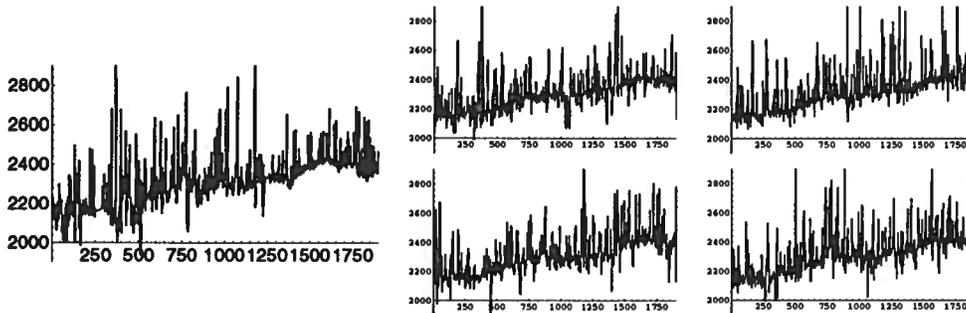


FIG. 6. P-wave sonic log (left) and pseudo-random simulations which have the same marginal distributions.

where $\rho_D(\mathbf{d})$ is the probability distribution associated with the data, $\mu_D(\mathbf{d})$ is Jaynes' non-informative distribution, which acts as a normalization, and $\Theta(\mathbf{d}|\mathbf{m})$ is the conditional distribution telling us the probability that a certain data vector could have been generated by a given model. Theoretical (modeling) uncertainties are described by Θ .

Let us introduce the notation $\rho_M(\mathbf{m})$ to denote the *a priori* probability distribution quantifying the current state of information about the model space. For example, if we know with certainty that a parameter m has some value m_0 , then $\rho_M(m) = \delta(m - m_0)$. Or, if we know that all values of m between m_{\min} and m_{\max} are equally likely, then $\rho_M(m)$ is a uniform distribution on the interval $[m_{\min}, m_{\max}]$. *A priori* refers to the fact that this is information which is independent of the experiment.

Combining the *a priori* knowledge of models $\rho_M(\mathbf{m})$, with the likelihood function $L(\mathbf{m})$, we arrive at the *a posteriori* probability distribution on the space of models

$$\sigma(\mathbf{m}) \propto \rho_M(\mathbf{m}) \int_D \frac{\rho_D(\mathbf{d})\Theta(\mathbf{d}|\mathbf{m})}{\mu_D(\mathbf{d})} d\mathbf{d}. \quad (3)$$

A posteriori here means that $\sigma(\mathbf{m})$ results from the analysis of data obtained in the experiment and combines all available information about the models. Comparing $\sigma(\mathbf{m})$ to $\rho_M(\mathbf{m})$ tells us precisely what we have learned about \mathbf{m} from the inversion. To the extent that σ is more informative (narrower, smaller entropy), for a given parameter, than ρ_M , then the inversion has succeeded in resolving the parameter to a readily quantifiable extent. A direct derivation of Equation 3 from the classical form of Bayes' theorem is given in Scales and Tarantola (1994).

Since we can generate models at will according to the prior information, we could imagine simply ranking them in order according to how well they fit the data. But if we did this we would have no way of making *a posteriori* estimates of the uncertainty of computed parameters. I.e., all we would have would be a best-fitting model, but no "error bars." The solution is to produce an importance sampling of the posterior itself. In other words we need a collection of models which has the following property:

if model x is twice as likely, *a posteriori*, as model y , than model x should appear, on average, twice as often as model y in the collection. Since we cannot possibly sample all of the points in model space, this sampling property is a *sine qua non* for any inverse method which does not involve analytic specification of the uncertainties.

Fortunately, we can sample the posterior in a conceptually straightforward fashion. The basic result is due to Mosegaard and Tarantola (1994) and represents a generalization of the Metropolis technique (Metropolis *et al.*, 1953). Let us assume that we are able to obtain samples $\mathbf{m}_1, \mathbf{m}_2, \dots$, of the prior probability distribution $\rho(\mathbf{m})$. Then, if we wish to obtain samples of the posterior probability distribution $\sigma(\mathbf{m}) = k \rho(\mathbf{m}) L(\mathbf{m})$ all that we need to do is to iterate the following procedure.

- Let \mathbf{m}_i be the “current model.”
- Use the rules that allow one to sample the prior probability distribution $\rho(\mathbf{m})$ to obtain a new model, say \mathbf{m}'_i .
- If $L(\mathbf{m}'_i) \geq L(\mathbf{m}_i)$, take \mathbf{m}'_i as new current model, i.e., make $\mathbf{m}_{i+1} = \mathbf{m}'_i$.
- If $L(\mathbf{m}'_i) < L(\mathbf{m}_i)$, then decide randomly to take the model \mathbf{m}'_i as new current model, or to destroy it, with a probability of taking it as new current model equal to the ratio $L(\mathbf{m}'_i)/L(\mathbf{m}_i)$.

Mosegaard and Tarantola (1994) show that this procedure converges to the true *a posteriori* distribution. To estimate the the degree of convergence, we look at the fluctuations in the “internal energy”, i.e., the sample average of the likelihood functions (Scales *et al.* (1991); Scales *et al.* (1992)).

The real power of this method cannot be over-emphasized: Since we are sampling from the prior distribution, we are only computing synthetic data for models which are *a priori* likely. The fatal flaw of most conventional Monte Carlo methods is that they waste too much time evaluating the response for models which are unlikely.

Let us now consider a simple experiment to test these ideas. We generate a synthetic seismogram from the North Sea sonic log and use this as the “observed data” for a zero-offset 1-D seismic inverse calculation. To make the problem slightly more interesting, we add Gaussian pseudo-random noise (signal to noise ratio 4), bandlimited to the same frequency spectrum as the seismogram. Figure 7 shows pseudo-random models from the posterior tour. I.e., these models have been pseudo-randomly generated according to the marginals of the well log, and then selected according to the Metropolis procedure. One way to make sense of this mass of information is to make histograms of the values of relevant parameters as they appear in the prior and posterior tours. This gives a quantitative measure of the values a parameter is most likely to take as well as the extent to which it has been resolved.

Figure 8 shows the prior and posterior histograms of the P-wave velocity at three points (in time) along the log. The totality of these histograms is the complete

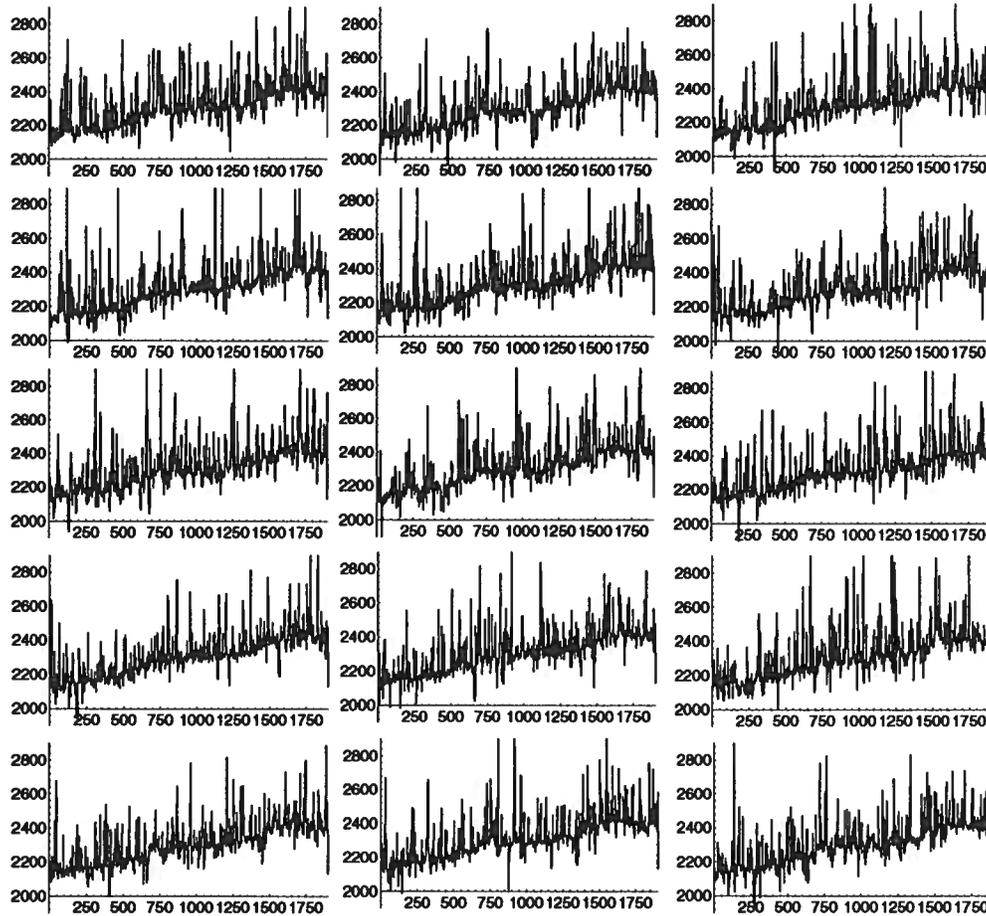


FIG. 7. Pseudo-random models sampled from the prior distribution and selected according to the Metropolis criterion.

solution of the problem. In this particular example, we have estimated the prior distribution directly from the log, but generated a synthetic trace (acoustic, normal-incidence, including multiples) from the log to represent a zero-offset experiment. In practice we would use a log recorded in a nearby seismic survey.

Figure 9 shows an individual trace from the posterior tour compared with the data. Figure 10 shows a stack of all the data traces generated by the posterior models compared to the data. None of the models in the posterior tour fit within all of the error bars of the data. Therefore the procedure has not yet managed to find the region of high probability in the model space. Whether this is a limitation of the implementation or a basic feature of Monte Carlo applied to such large problems (nearly 2000 parameters) remains to be seen.

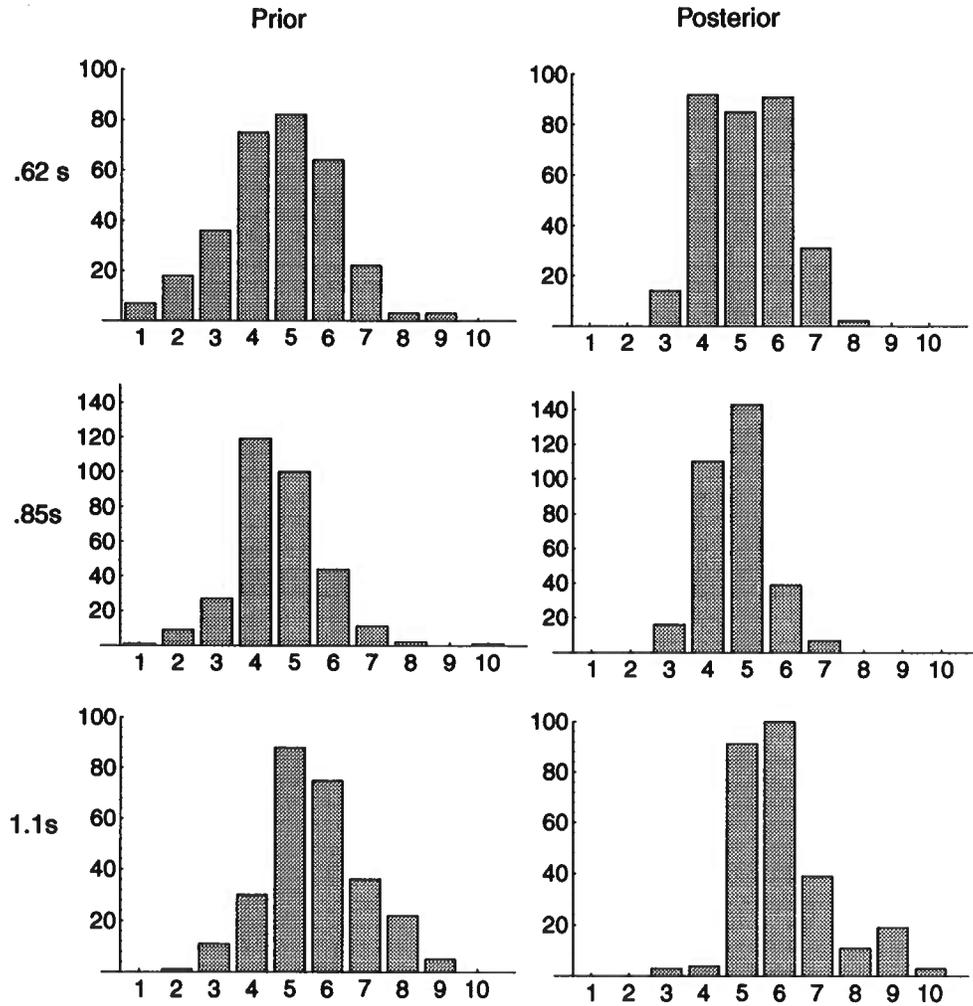


FIG. 8. Histograms of parameter values appearing in the prior tour and the posterior tour. This is a quantitative measure of resolution taking into account all the prior information, both on models and data.

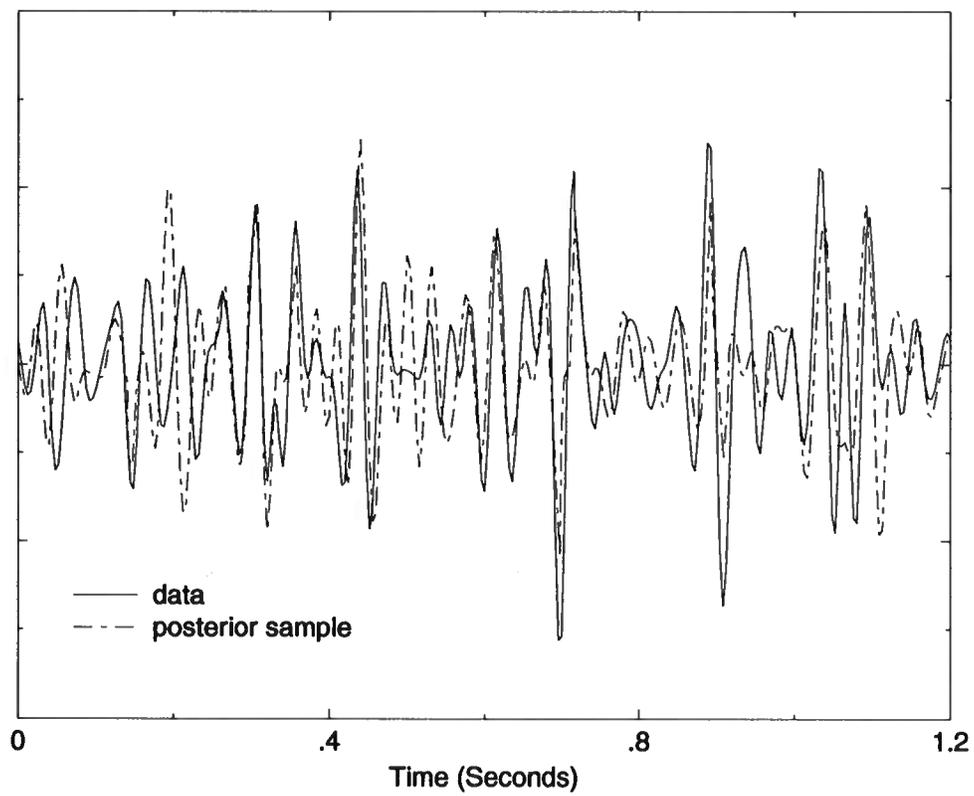


FIG. 9. A comparison of the data trace with a single sample from the posterior tour.

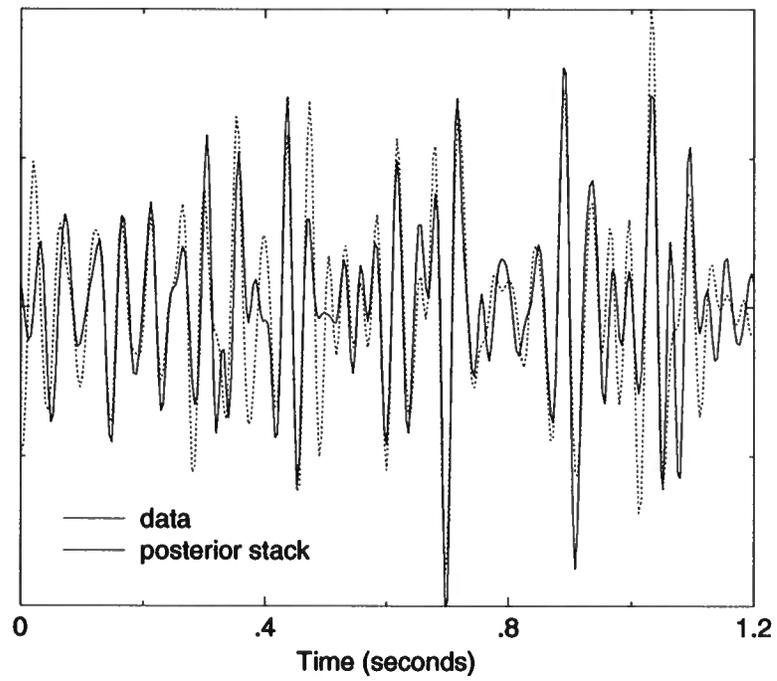


FIG. 10. A comparison of the stack trace from the posterior tour and the data.

CONCLUSIONS AND FUTURE DIRECTIONS

All inverse calculations require the specification of *a priori* information, on the data, the space of feasible models, and the accuracy of the forward modeling error. The goal of this research is to quantify this prior information so that posterior estimates of the uncertainty of computed models can be made. When dealing with real data it is never sufficient to simply produce an “optimal” model. It is only by producing distributions of models consistent with the prior information that we can hope to assign quantitative confidence values to the inferences we make.

Some progress has been made in estimating these uncertainties nonparametrically from real data. But these seem to require expensive Monte Carlo importance sampling techniques. On the other hand, it is not clear to what extent nature can be well approximated by simpler parametric models, which are amenable to gradient-based methods of optimization.

REFERENCES

- Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A., & Teller, E. 1953. Equation of state calculations by fast computing machines. *Journal of Chemical Physics*, **21**, 1087–1092.
- Mosegaard, K., & Tarantola, A. 1994. Monte Carlo sampling of solutions to inverse problems. *JGR: submitted*.
- Scales, J. A., & Tarantola, A. 1994. *Bayesian inversion with realistic a priori information*. Tech. rept. 159. Center for Wave Phenomena.
- Scales, J. A., Smith, M. L., & Fischer, T. L. 1991. Global optimization methods for highly nonlinear inverse problems. In: Gary, C., Halpern, L., & Joly, P. (eds), *Mathematical and Numerical Aspects of Wave Propagation Phenomena*. SIAM.
- Scales, J. A., Smith, M. L., & Fischer, T. L. 1992. Global optimization methods for multimodal inverse problems. *Journal of Computational Physics*, 258–268.
- Tarantola, Albert. 1987. *Inverse Problem Theory*. New York: Elsevier.

CWP-182P
September 1995



Uncertainties in Seismic Inverse Calculations

John A. Scales

To be included in the proceedings volume of the Interdisciplinary Conference on
Methodology Computation and Integrated Applications
University of Aarhus, Denmark, May 1995.

Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401
303/273-3557

Uncertainties in Seismic Inverse Calculations

John A. Scales

Center for Wave Phenomena, Department of Geophysics
Colorado School of Mines, Golden USA

Introduction

Solving an inverse problem means making inferences about physical systems from real data. This requires taking into account three different kinds of information.

- What is known about the parameters independently of the data? In other words, what does it mean for a model to be reasonable or unreasonable?
- How accurately are the data known? That is, what does it mean to “fit the data”?
- How accurately is the physical system modeled? Does the model include all the physical effects that contribute significantly to the data?

Exploration seismologists have a large amount of information available that could be used to refine inferences about the subsurface. In this paper I will describe preliminary work aimed at quantifying this information in a form amenable to calculation.

A Simple Example

To motivate the discussion, let us consider first a problem of linear tomographic inversion.

The Forward Problem Acoustic sources and receivers are lowered into two vertical boreholes as shown in Figure 1. Sources are excited, seismograms are recorded, and the goal is to infer the elastic properties of the medium between the boreholes. To some approximation the wave propagation can be treated kinematically, so the travel time of a wave (e.g., the direct arrival) is related to the unknown slowness between the boreholes by

$$t(s(x, y, z)) = \int_{\text{ray}(s(x, y, z))} s(x, y, z) d\ell. \quad (1)$$

slowness model, and σ_i is the standard deviation (assumed known) of the i -th datum.

Let's see how this works in practice. Figure 2 shows an example of such a tomography problem. The model consists of a homogeneous background with slowness equal to 3, and an embedded inclusion shaped like an "E" having slowness equal to 1.5. At the right we see the "illumination" of the rays. There are 10 sources, 20 receivers and 20^2 cells within which the slowness is assumed to be constant.

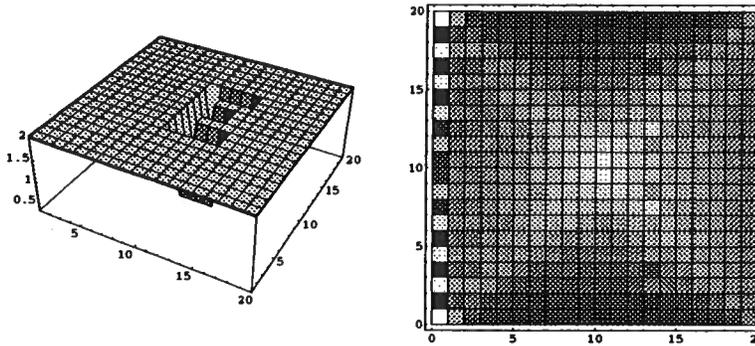


Fig. 2. Straight rays are traced between 10 sources and 20 receivers. The goal is to be able to characterize the unknown inclusion (left). On the right, the ray "illumination" plot is shown.

Then add uncorrelated Gaussian noise, so that the standard deviations σ_i are constant and known. The "textbook" solution to the problem is do a singular value decomposition (SVD) of the Jacobian matrix. Then, build up the solution one singular vector at a time, starting with the largest singular values and moving down the spectrum, until the solution *just* manages to fit the data.

Figure 3 shows this truncated SVD solution as both a perspective (left) and contour (right) plot. Although we can just see the inclusion in the contour plot, one would perhaps be justified in saying this solution is rather unrealistic looking. It fits the data, undeniably, but surely many of the complicated looking features are artifacts in the sense that they are not required to fit the data. Further, there is almost certainly information available that would allow one to say *a priori* whether this was a realistic looking model or not.

What's the Point? One conclusion to be drawn from this simple experiment is that data fit alone is *never* enough to judge the solution of an inverse calculation. Since unrealistic models can fit the data too, one must balance the extent to which a model fits the data against the extent to which it conforms to our knowledge about what makes a model realistic or not in a particular case. This knowledge takes many different forms. It might be as simple as saying that a particular

Information About Models

In exploration seismology there is a large amount of *a priori* information that could be used to influence inverse calculations. Plausible geologic models can be based on rock outcrops, models of sedimentological deposition, tectonics, etc. There are often *in situ* measurements of the elastic properties of rocks from nearby locations (well-logs, for instance). There are laboratory measurements of rock properties. There are other, less quantitative, forms of information as well, the knowledge of experts for instance.

There is a simple conceptual model that can be used to visualize the application of this diverse *a priori* information. Suppose we have a black box into which we put information about our problem. We can turn a crank or push a button and out pops a model that is consistent with whatever information that is put in. One may repeat the process indefinitely, producing a collection of models that have one thing in common: they are all consistent with the information put into the black box. Suppose we know, for example from *in situ* measurements, that the reflection coefficient r in a particular sedimentary basin almost never exceeds .1. What does it mean for a model to be consistent with this information? We can push the button on the black box and generate models which satisfy this requirement. Figure 4 shows some examples.

The first model is uncorrelated with a uniform distribution on $[-.1,.1]$. It is consistent with the prior information but is rather unrealistic looking. Surely reflection coefficients near .1 are much less likely than those near, say .01. So, while a hard constraint $|r| < .1$ on reflection coefficients seems reasonable, we almost certainly have more information than just this. Further, taking a hard constraint and representing it as a uniform probability distribution may not be a sensible thing to do. The next simplest thing to try is a Gaussian distribution. The middle plot in Figure 4 shows an uncorrelated Gaussian model with a standard deviation chosen such that the constraint on $|r|$ is satisfied. This looks somewhat more reasonable, yet it is difficult to imagine that the underlying geologic process of sedimentation would give rise to a completely uncorrelated measurement such as this. Further, we know from having looked at many logs, that finite correlation lengths are ubiquitous and represent a measure of the feature size of the rocks. In sedimentary basins vertical correlation lengths of a few meters to a few tens of meters are common. So let us generate a model with the same mean and standard deviation, but with a finite correlation length, in this example 15 samples. This is shown in the bottom of Figure 4. We could continue in this fashion, building ever more realistic models.

If we were able to build such a black box in practice, then a plausible strategy for solving the inverse problem would be to generate a sequence of models according to the prior information and see which ones fit the data. In the case of the reflectivity sequence, imagine that we have surface seismic data to be inverted. So for each model generated by the black box, compute synthetic seismograms, compare them to the data and decide whether they fit the data well enough to be acceptable. If so, the models are saved, if not, rejected. Repeating this procedure many times results in a collection of models that are by definition consistent

long-wavelength trend, which is likely the result of more gradual, deterministic changes such as compaction. The trend and the resulting fluctuations are shown in Figures 6 and 7.

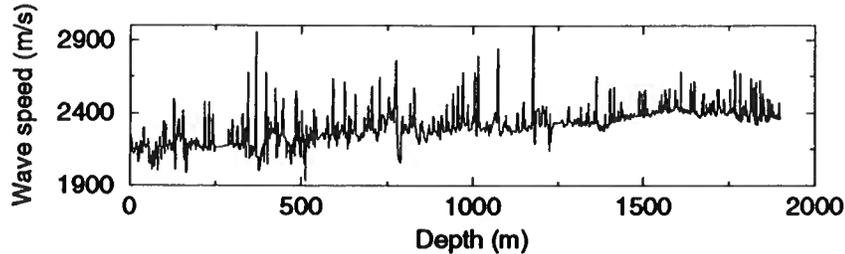


Fig. 5. Estimates of P and S wave velocity are obtained from the travel times of waves propagating through the formation between the source and receiver on a tool lowered into the borehole.

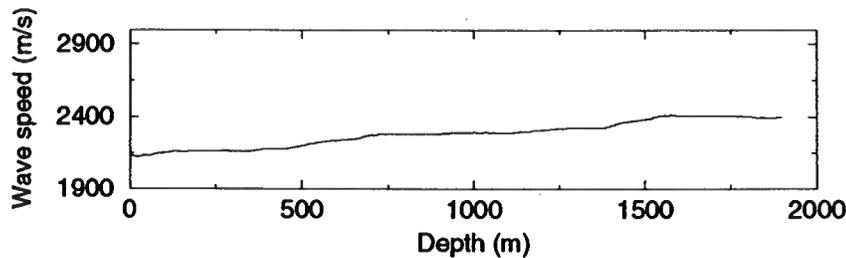


Fig. 6. Trend of Figure 5 obtained with a 150 sample running average.

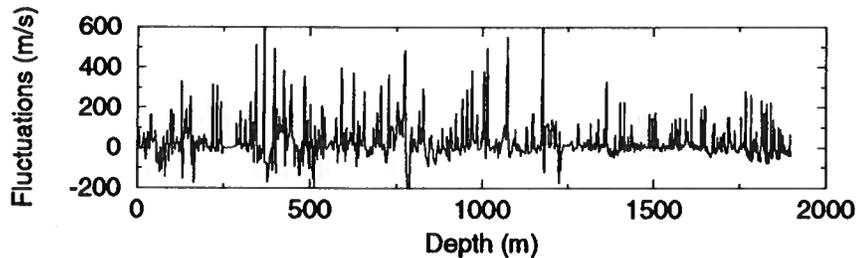


Fig. 7. Fluctuating part of the log obtained by subtracting the trend from the log itself.

If we were willing to assume *a priori* that the properties of this log are stationary and Gaussian, then it would be relatively straightforward to compute a Gaussian covariance matrix. For example, Figure 8 shows the first 100 lags (1 meter per lag) of the correlation function for the fluctuating part of Figure 7, as

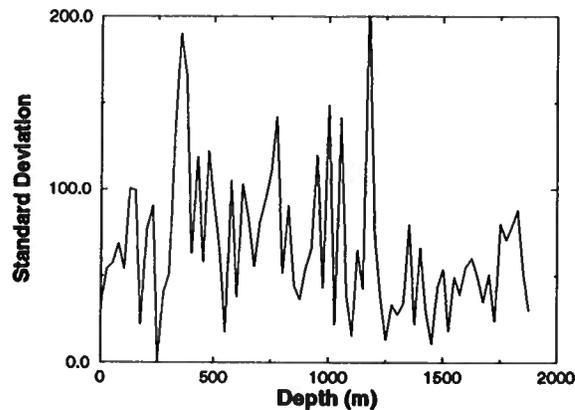


Fig. 9. The standard deviation as a function of depth of Figure 7.

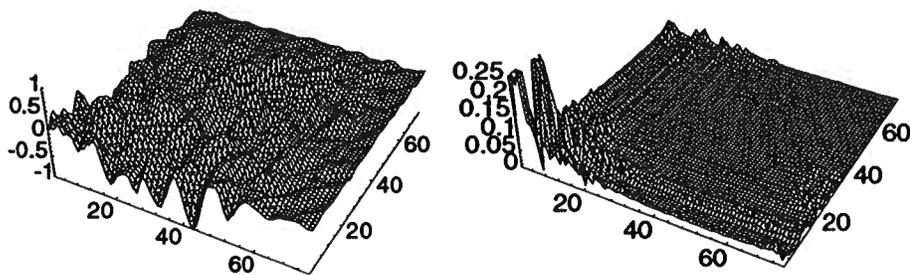


Fig. 10. The normalized third order cumulant and bispectrum of the fluctuating part of the data in Figure 7, computed for a short window of 75 lags. To account for non-stationarity, we can slide this window down the full length, repeating the estimates in each window.

Uncertainties in the data

In the last section we looked at characterizations of models via covariance matrices (or their higher order generalizations) obtained from commonly available *in situ* well log measurements. Next we look at estimating uncertainties in the seismic data themselves. We are used to thinking of seismic experiments as being non-repeatable. While this is true in earthquake seismology, it is not necessarily true in controlled-source seismology. Most land reflection data recorded nowadays uses a vibroseis source, i.e., a swept continuous wave signal. By careful stacking and cross-correlation with the known input signal, such data can approximate the results of explosive sources with high signal to noise ratio. But usually the raw, unstacked data are not saved; once the stacking and correlation have been done, the raw data are usually discarded. Figure 11 shows an example in which the raw traces have been saved. There are 25 unstacked, uncorrelated vibroseis

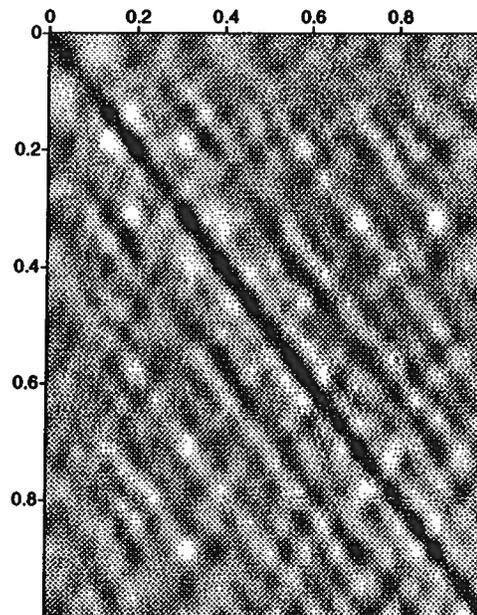


Fig. 12. Sample covariance matrix for the ambient noise present in a reflection seismic experiment. The ensemble consists of the repeated sweeps of the vibrator (before stacking and correlation) and has been windowed to the first second of data, before any source generated energy has arrived at the receivers.

sources and receivers in exactly the right places, so small shifts in the assumed known-source and receiver locations will introduce errors into the problem. Do we regard these as being essentially random fluctuations which broaden the distribution of the data, or do we look at the shifts in source and receiver locations as being model parameters (like station corrections) to be inferred from the data?

Tarantola (1987) shows that as long as the uncertainties are Gaussian, we can still use Equation 12 to describe the likelihood function, provided we replace the data covariance matrix C_d , with a combined covariance matrix

$$C_D = C_d + C_T$$

where C_T represents the contribution due to theoretical errors.

More often than not, theoretical errors are neglected in real applications. And the reason is plain, for it is not usually possible to account for a particular kind of unmodeled physics without incurring the additional expense that led to the neglect of this physics in the first place. How can one quantify the effects of neglecting anisotropy without doing the complete calculation anisotropically

- Gouveia, W. 1995. A study of model covariances in amplitude seismic inversion. In: *INVERSE METHODS: Interdisciplinary elements of Methodology, Computation and Application*.
- Mosegaard, K., & Tarantola, A. 1995. Monte Carlo sampling of solutions to inverse problems. *JGR: in press*.
- Scales, J. A. 1993. On the use of localization theory to characterize elastic wave propagation in randomly stratified 1-D media. *Geophysics*, **58**, 177–179.
- Scales, J.A., Docherty, P., & Gersztenkorn, A. 1990. Regularization of nonlinear inverse problems: imaging the near-surface weathering layer. *Inverse Problems*, **6**, 115–131.
- Slot, R. E., & Wapenaar, C. P. A. 1994. Extending the macro model by incorporating fine-layering propagation effects. *Page Session:G009 of: 56th Mtg. Eur. Assoc. Expl Geophys., Extended Abstracts*. Eur. Assoc. Expl. Geophys.
- Subba Rao, T., & Gabr, M.M. 1984. *An introduction to bispectral analysis and bilinear time series models*. Springer-Verlag.
- Tarantola, A. 1987. *Inverse Problem Theory*. New York: Elsevier.

A The Bayesian Posterior Probability

The Bayesian posterior probability density on the space of models must be the product of two terms: a term which involves the *a priori* probability on the space of models and a term which measures the extent of data fit

$$\sigma(\mathbf{d}) = k\rho(\mathbf{m}) L(\mathbf{m}) \quad (6)$$

where k is the normalization constant and L is the likelihood function, which depends implicitly on the data.

We now show how Equation (6) follows logically from Bayes' theorem provided we generalize our notion of "data" to allow for the possibility that the data might be specified by probability distributions (Tarantola, 1987). Following Duijndam (1987), we begin by using the notation common amongst Bayesians, then we show how this relates to the more standard inverse-theoretic notation in Tarantola (1987).

In this approach we assume that we have some prior joint distribution $p_0(\mathbf{m}, \mathbf{d})$. Further, we suppose that as the result of some observation, the marginal pdf of \mathbf{d} changes to $p_1(\mathbf{d})$. We regard $p_1(\mathbf{d})$ as being the "data" in the sense that we often know the data only as a distribution, not exact numbers. In the special case where the data are exactly known, $p_1(\mathbf{d})$ reduces to a delta function $\delta(\mathbf{d} - \mathbf{d}_{obs})$.

How do we use this new information in the solution of the inverse problem? The answer is based upon the following assumption: whereas the information on \mathbf{d} has changed as a result of the experiment, there is no reason to think that the conditional degree of belief of \mathbf{m} on \mathbf{d} has. I.e.,

$$p_1(\mathbf{m}|\mathbf{d}) = p_0(\mathbf{m}|\mathbf{d}). \quad (7)$$

From this one can derive the posterior marginal $p_1(\mathbf{m})$: