CWP-184
May 1995

A study of model covariances in amplitude seismic inversion

Wenceslau Gouveia

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

A study of model covariances in amplitude seismic inversion

Wences Gouveia

ABSTRACT

Bayesian inversion (Tarantola, 1987) provides a concise mathematical framework that formally allows the incorporation of *a priori* information into geophysical data inversion. In this methodology, the general solution of an inverse problem can be regarded as a probability density $\sigma(\mathbf{m})$ over the space of models, that consists of the product of two probability density functions. One, known as the likelihood function $L(\mathbf{m})$, measures the extent that the observed data are fit by model data. The other, $\rho(\mathbf{m})$ quantifies the *a priori* knowledge that is possibly available about the inverse problem. This information, derived for instance from regional geology considerations, well-log data and other types of geophysical data, can be incorporated into the inversion problem, via model covariance matrices of $\rho(\mathbf{m})$. The construction of these matrices from such sources of information is a complicated problem, and in the large majority of cases, *ad hoc* simplifying assumptions are made. As a consequence the significance of the model covariance matrices is lost.

In this work I study the effect of model covariance matrices in a linear, iterative amplitude-inversion algorithm. I illustrate in a simple example some advantages of building covariance matrices from statistical considerations about the underlying model, as opposed to using the Tikhonov regularization method (Tikhonov and Arsenin, 1977) . This method builds covariance matrices generally under the assumption of model smoothness, providing little flexibility to incorporate more realistic information about the inverse problem.

The linear, iterative amplitude-inversion algorithm is proposed in Jin *et al.* (1992). In their work the inversion problem is formulated under the small scatterer, or Born, approximation. The resulting linear system of equations is solved by a minimization of a weighted least-squares norm, with weights derived from ray theory. The solution, i.e., perturbations to a given background velocity model, is obtained by a quasi-Newton optimization method, possibly an expensive approach since it makes use of an approximation to the second derivatives of the objective function. However, Jin *et al.* (1992) showed that the Hessian matrix can be approximated by a diagonal matrix with good results.

INTRODUCTION

Seismic amplitude or travel time inversion methods are a major topic of geophysical research due to their potential capability of extracting detailed lithologic information about the subsurface. Several inversion methodologies are described in the technical literature. Although the procedures differ, it is acknowledged in all of them that the data alone do not constrain all the model features that one aims to estimate. To reduce the ambiguity of the inverse problem it is necessary to incorporate *a priori* information about the underlying model. The Bayesian approach for geophysical data inversion (Tarantola, 1987) paves the way for the incorporation of such knowledge. In this methodology, the general solution of an inverse problem is defined as a probability density $\sigma(\mathbf{m})$ over the space of models, that consists of the product of two probability density functions. One, known as the likelihood function $L(\mathbf{m})$, measures the extent that the observed data are fit by model data. This function accounts for uncertainties in the data, i.e., data features that were not taken into account in the forward modeling step. Examples are noise in the data, multiples in the situation where the forward modeling procedure just generates primaries, and so on. The other probability density function, $\rho(\mathbf{m})$, quantifies the *a priori* knowledge that is possibly available about the inverse problem. In this work I will assume that $\rho(\mathbf{m})$ and $L(\mathbf{m})$ are Gaussian probability distributions, defined by the following expressions

$$\begin{aligned}\rho(\mathbf{m}) &= ((2\pi)^M \det C_M)^{-\frac{1}{2}} exp\left[-\frac{1}{2}(\mathbf{m}-\mathbf{m_0})^T\, {C_M}^{-1}(\mathbf{m}-\mathbf{m_0})\right], \\ L(\mathbf{m}) &= ((2\pi)^N \det C_D)^{-\frac{1}{2}} exp\left[-\frac{1}{2}(g(\mathbf{m})-\mathbf{d_{obs}})^T\, {C_D}^{-1}(g(\mathbf{m})-\mathbf{d_{obs}})\right]. \quad (1)\end{aligned}$$

Here, M is the number of model parameters; N is the number of observations; C_M and C_D are the model and data covariance matrices, respectively; $\mathbf{d_{obs}}$ is the observed data vector; $g(\mathbf{m})$ represents the modeled (synthetic) data for the model $\mathbf{m}$, and $\mathbf{m_0}$ is the mean or most likely model.

In this situation the probability density $\sigma(\mathbf{m})$, also known as *a posteriori* probability density function is also Gaussian and given by

$$\sigma(\mathbf{m}) \propto \exp\left[-\frac{1}{2}(g(\mathbf{m})-\mathbf{d_{obs}})^T C_D^{-1}(g(\mathbf{m})-\mathbf{d_{obs}}) + (\mathbf{m}-\mathbf{m_0})^T C_M^{-1}(\mathbf{m}-\mathbf{m_0})\right]. \quad (2)$$

The covariance matrix C_M of the probability density function $\rho(\mathbf{m})$ is a possible connection between the *a priori* information and the inverse problem. This information can be derived from regional geological considerations, well-logs, interpretative work and so on. To build covariance matrices from those sources is not trivial and is seldom attempted, at least in the published inversion literature. A specific case where model covariance matrices, and higher order statistical moments are derived from well-logs is described in Scales and Tarantola (1994). In view of this difficulty,

ad hoc techniques are commonly used to build the model covariance matrices. Consequently, the significance of these matrices (and also of $\rho(\mathbf{m})$) is lost.

The objective of this work is to illustrate with a simple example how a specific seismic amplitude inversion algorithm can benefit from a model covariance matrix built using statistical considerations about the model one seeks, derived from some source of information (for instance, well logs). I compare the result obtained with this approach with the one obtained when I used the Tikhonov regularization method (Tikhonov and Arsenin, 1977) to construct the model covariance matrix, based on model smoothness assumptions. As will be shown later, both results are equivalent for the case considered here. However, two advantages can be pointed out in favor of the statistical construction of the model covariance matrices. First, the absence of a weighting factor, required by the Tikhonov approach, to incorporate the *a priori* information into the inverse problem. Second, the assessment of the uncertainties of the inversion procedure is probably more accurate when the model covariances are constructed honoring, at least to some extent, the statistics of the model parameters.

The amplitude seismic-inversion algorithm discussed here is based on the work of Jin *et al.* (1992). They linearize the isotropic elastic inversion problem with the Born approximation (Cohen and Bleistein, 1979) yielding a system of equations that is weighted according to ray-theoretic considerations and solved by a quasi-Newton method. They derived a diagonal approximation to the second-derivative matrix, which is a direct consequence of the weighting applied to the system.

This paper is structured as follows. I begin with a brief exposition of the inversion algorithm proposed by Jin *et al.* (1992). Here I restrict this outline to the acoustic approximation. Following that, I review the theoretical aspects of the regularization theory and present an example to illustrate its utility in the situation of inversion of noisy data. I also point out the connection between this theory and the more general Bayesian approach. Next I carry out the comparison, for a given inverse problem, between the results of the asymptotic linear iterative inversion when model covariances derived from the model statistics are used as opposed to Tikhonov regularization matrices. Finally, I present conclusions and future research directions for this work.

ITERATIVE ASYMPTOTIC AMPLITUDE INVERSION

Jin *et al.* (1992) proposed a linearized asymptotic inversion method where the seismic inversion is formulated as the optimization of a data misfit objective function for elastic parameter estimation. Essentially the method consists of solving an overdetermined system of equations obtained from the linearization of the integral solution of the wave equation via the Born approximation. In their work Jin *et al.* (1992) solved this system using a weighted least-squares criterion. This weighting is derived from ray theory considerations. Following is a brief description of the algorithm for the acoustic inverse problem.

Linearization

As in many inversion procedures the velocity $c(\mathbf{r})$ of the medium is characterized by a long-wavelength velocity profile $c_0(\mathbf{r})$ plus small deviations $\delta(\mathbf{r})$ (scatterers) from this background velocity, where $\mathbf{r}$ is the position vector. The ultimate objective of the type of inversion algorithm discussed here is to estimate such deviations, given the background velocity. This approach, derived from perturbation theory, is relatively common in the inversion literature (e.g. Beylkin, 1985, Cohen and Bleistein, 1979, and others).

Leaving the algebraic details to the references (Bleistein *et al.*, 1994), it is possible to show that the recorded (scattered) wave field $u_s(\mathbf{r_g}, \mathbf{r_s}, t)$ and the source (incident) wave field $u_i(\mathbf{r_g}, \mathbf{r_s}, t)$ satisfy the following integral relationship, here expressed in the frequency domain [1]:

$$u_s(\mathbf{r_g}, \mathbf{r_s}, \omega) = \int_D d\mathbf{r}\, g(\mathbf{r}, \mathbf{r_g}, \omega) m(\mathbf{r}) (u_i(\mathbf{r}, \mathbf{r_s}, \omega) + u_s(\mathbf{r}, \mathbf{r_s}, \omega)) \omega^2, \tag{3}$$

where D is the domain of integration over the diffraction points; $m(\mathbf{r}) = \frac{-2\delta c(\mathbf{r})}{c_0(\mathbf{r})^3}$, the unknown perturbation scaled by the background velocity, is the parameter sought in the inversion, and $g(\mathbf{r_1}, \mathbf{r_2}, \omega)$ is the Green's function for an impulsive source at $\mathbf{r_2}$ recorded at $\mathbf{r_1}$, computed in the present work from ray theory. For a 2D medium this Green's function is given by

$$g(\mathbf{r_1}, \mathbf{r_2}, \omega) = A(\mathbf{r_1}, \mathbf{r_2}) e^{i\omega\tau(\mathbf{r_1}, \mathbf{r_2})} \frac{1}{\sqrt{-i\omega}}. \tag{4}$$

Here, $\tau(\mathbf{r_1}, \mathbf{r_2})$ and $A(\mathbf{r_1}, \mathbf{r_2})$ are the ray theoretical traveltime and amplitude that satisfy the eikonal and transport equations, respectively. Equation (3) is nonlinear with respect to $m(\mathbf{r})$ because it contains a product of this unknown quantity and the scattered (observed) field $u_s(\mathbf{r_g}, \mathbf{r_s}, \omega)$. The Born approximation, which basically neglects the scattered wave field in comparison with the incident wavefield under the "small scatterer" assumption, is used in the linearization of Equation (3). Noticing that $u_i(\mathbf{r}, \mathbf{r_s}, \omega) = g(\mathbf{r}, \mathbf{r_s}, \omega)$, this linearization results in the following equation:

$$u_s(\mathbf{r_g}, \mathbf{r_s}, \omega) = \int_D d\mathbf{r}\, g(\mathbf{r}, \mathbf{r_g}, \omega) m(\mathbf{r}) g(\mathbf{r}, \mathbf{r_s}, \omega) \omega^2, \tag{5}$$

or, in matrix form:

$$G\mathbf{m} = \mathbf{u_s}. \tag{6}$$

Here, G is the Born operator matrix; $\mathbf{m}$ is the unknown normalized scattering vector, and $\mathbf{u_s}$ is the recorded field.

In the inversion algorithm, the linear system of equations (6) is solved for the model $\mathbf{m}$ by a quasi-Newton technique described next.

[1] Throughout this paper $\mathbf{r_g}$ and $\mathbf{r_s}$ will represent spatial coordinates of receiver and source respectively, t the traveltime and ω the temporal frequency.

Solution of the linear system

The linear system of equations (6) is usually over-determined, since the number of observations (data points) is larger than the number of unknowns. Therefore it is necessary to define what it is meant by the solution of the system. In their work Jin *et al.* (1992) used the following weighted least-squares criterion for this definition:

$$\min_{\mathbf{m}} S(\mathbf{m}, \mathbf{r_0}) = \min_{\mathbf{m}} \frac{1}{2} \int d\xi \int d\psi \int d\omega \, (\mathbf{u_s} - G\mathbf{m})^T Q (\mathbf{u_s} - G\mathbf{m}). \tag{7}$$

Where $\mathbf{r_0}$ is the output point, i.e., the coordinate of the scatter point to be estimated; Q is a diagonal matrix that implements the weighting and will be described later. In matrix form Equation (7) can be written as:

$$\min_{\mathbf{m}} S(\mathbf{m}, \mathbf{r_0}) = \min_{\mathbf{m}} (\mathbf{u_s} - G\mathbf{m})^T Q (\mathbf{u_s} - G\mathbf{m}). \tag{8}$$

Notice that the sum in Equation (7) is carried out over the angles ψ and ξ defined in Figure 1. Ideally, for the best resolution, it would be desirable to sample the scattering point from all angles, which is not the case for seismic experiments. Moreover, instead of summing over the angles ψ and ξ, a more suitable coordinate system is defined by the source and receiver locations. Considering this coordinate system, Equation (7) can be rewritten as:

$$\min_{\mathbf{m}} S(\mathbf{m}, \mathbf{r_0}) = \min_{\mathbf{m}} \frac{1}{2} \int d\mathbf{r_s} \int d\mathbf{r_g} \int d\omega \, (\mathbf{u_s} - G\mathbf{m})^T Q (\mathbf{u_s} - G\mathbf{m}) J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi). \tag{9}$$

Here, $J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi)$ is the Jacobian of the transformation, that should account for the discretization of the data acquisition. One of the most interesting aspects of the work of Jin *et al.* (1992) is the weighting diagonal matrix Q. The ii^{th} element of the Q matrix relates to a given source-receiver pair and is defined as follows:

$$Q(\mathbf{r_g}, \mathbf{r_s}, \omega, \mathbf{r_0}) = \frac{\|\mathbf{p}(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s})\|^2}{4\pi^2 \omega A^2(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s})}, \tag{10}$$

where (see Figure 1) $\|\mathbf{p}(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s})\|^2$ is the square modulus of the total slowness vector at the scattering point, defined as $\mathbf{p}(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s}) = \nabla\tau(\mathbf{r_0}, \mathbf{r_s}) + \nabla\tau(\mathbf{r_g}, \mathbf{r_0}) = \mathbf{p_s}(\mathbf{r_0}, \mathbf{r_s}) + \mathbf{p_g}(\mathbf{r_g}, \mathbf{r_0})$, and $A(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s})$ is an amplitude factor defined as $A(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s}) = A(\mathbf{r_0}, \mathbf{r_s}) A(\mathbf{r_g}, \mathbf{r_0})$.

The justification for the weighting matrix Q comes from ray-theory. This matrix compensates for geometrical spreading losses and tends to eliminate wide angle reflections from the inversion (small values of $\|\mathbf{p}\|$). Gray (1994) also avoided wide angle reflections in his migration procedure, under the justification that those events might be spatially aliased. Notice that the Q matrix depends on the coordinate $\mathbf{r_0}$. A drawback of this formulation of Q is the presence of the temporal frequency ω in the denominator, weighting down the higher-frequency information. However, as will be shown later, the main benefit of using such weighting is accomplished by a

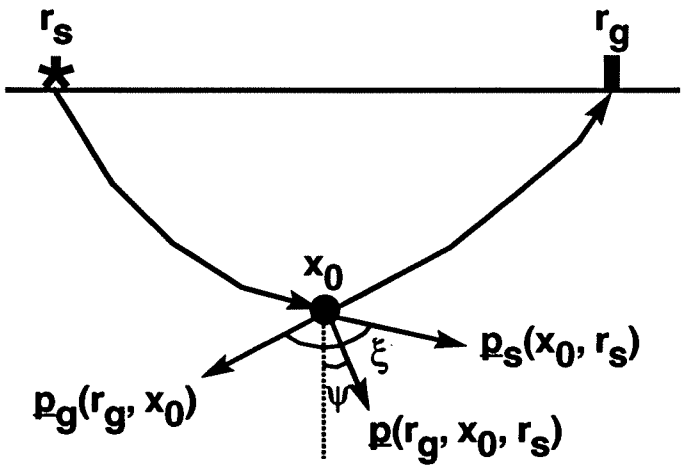

FIG. 1. Geometry for one scattering point.

diagonal approximation of the Hessian matrix (2^{nd} derivative matrix of the objective function with respect to the model parameters) allowing an efficient implementation of a quasi-Newton algorithm for minimizing the objective function defined in Equation (9).

The minimizer of the objective function defined in Equation (9) satisfies the familiar normal system of equations

$$G^H Q G \mathbf{m} = G^H Q \mathbf{u_s}. \tag{11}$$

Where G^H is the Hermitian adjoint of G. It is well known that direct or iterative techniques are available for solving linear system of equations. Jin *et al.* (1992) opted for a quasi-Newton iteration method given by

$$\mathbf{m_{n+1}} = \mathbf{m_n} - {H_a}^{-1}\gamma(\mathbf{m_n}). \tag{12}$$

Here, $\mathbf{m_{n+1}}$ is the updated model; $\mathbf{m_n}$ is the current model; ${H_a}^{-1}$ is an approximation for the inverse of the Hessian matrix evaluated at model $\mathbf{m_n}$, and $\gamma(\mathbf{m_n})$ is the gradient of the objective function evaluated at model $\mathbf{m_n}$. Each iteration performed in Equation (12) has a computational cost equivalent to a pre-stack migration algorithm.

The analytic computation of the derivatives of the objective function in Equation (9) is facilitated by the fact that WKBJ Green's functions and the Born approximation have been used. Therefore it is not complicated to show that the gradient of the objective function with respect to the model parameters is given by

$$\gamma(\mathbf{r}, \mathbf{r_0}) = \frac{1}{2\pi}\int d\mathbf{r_s}\int d\mathbf{r_g}\,\frac{A(\mathbf{r_g}, \mathbf{r}, \mathbf{r_s})}{A^2(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s})}\|\mathbf{p}\|^2 J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi)\mathcal{H}[\delta\mathbf{u_s}(\mathbf{r_g}, \mathbf{r_s}, t = \tau(\mathbf{r_g}, \mathbf{r}, \mathbf{r_s})]. \tag{13}$$

$\mathcal{H}[\delta\mathbf{u_s}(\mathbf{r_g}, \mathbf{r_s}, t = \tau(\mathbf{r_g}, \mathbf{r}, \mathbf{r_s})]$ is the Hilbert transform of the data residual evaluated at the total travel time $\tau(\mathbf{r_g}, \mathbf{r}, \mathbf{r_s}) = \tau(\mathbf{r}, \mathbf{r_s}) + \tau(\mathbf{r_g}, \mathbf{r})$.

Along the same lines it is possible to show that the ij^{th} element of the Hessian matrix $H = G^H Q G$ is given by

$$H[\mathbf{r_i}, \mathbf{r_j}] = \frac{1}{4\pi^2} \int d\mathbf{r_s} \int d\mathbf{r_g} \int d\omega\, \omega \|\mathbf{p}\|^2 \frac{A^2(\mathbf{r_g}, \mathbf{r_j}, \mathbf{r_s})}{A^2(\mathbf{r_g}, \mathbf{r_i}, \mathbf{r_s})} J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi) e^{-i\omega[\tau(\mathbf{r_g}, \mathbf{r_j}, \mathbf{r_s}) - \tau(\mathbf{r_g}, \mathbf{r_i}, \mathbf{r_s})]}. \tag{14}$$

The above equation can be considerably simplified with a sequence of approximations. First notice that in the process of the quasi-Newton iterations the i^{th} row of the Hessian matrix will be dotted with the gradient vector to yield the updated model parameter at the coordinate $\mathbf{r_i}$. Therefore $\mathbf{r_i}$ is the output point $\mathbf{r_0}$ mentioned previously. So, by using the same methodology as in Bleistein *et al.* (1994) and Beylkin (1985) the traveltime $\tau(\mathbf{r_s}, \mathbf{r}, \mathbf{r_g})$ and amplitude $A(\mathbf{r_s}, \mathbf{r}, \mathbf{r_g})$ are expanded about the output point $\mathbf{r_i}$. This procedures results in the following expression, after keeping terms to first order for the travel time and just the zero-th order term for the amplitudes:

$$H_a[\mathbf{r_i}, \mathbf{r_j}] = \frac{1}{4\pi^2} \int d\mathbf{r_s} \int d\mathbf{r_g} \int d\omega\, \omega \|\mathbf{p}\|^2 J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi) e^{-i\omega \mathbf{p}(\mathbf{r_j} - \mathbf{r_i})}. \tag{15}$$

Furthermore assuming that $J \approx 1$, and going to the $[\psi, \omega]$ domain we get:

$$H_a[\mathbf{r_i}, \mathbf{r_j}] = \frac{1}{4\pi^2} \int d\mathbf{r_s} \int d\psi \int d\omega\, \omega \|\mathbf{p}\|^2 e^{-i\omega \mathbf{p}(\mathbf{r_j} - \mathbf{r_i})}. \tag{16}$$

Introducing the variable $\mathbf{K} = \omega \mathbf{p}$ we obtain:

$$H_a[\mathbf{r_i}, \mathbf{r_j}] = \frac{1}{4\pi^2} \int d\mathbf{r_s} \int d\psi \int d\|\mathbf{K}\|\, \|\mathbf{K}\| e^{-i\mathbf{K}.(\mathbf{r_j} - \mathbf{r_i})}. \tag{17}$$

Notice that the integrals over the $[\mathbf{K}, \psi]$ domain represent the integration in cylindrical coordinates of a constant. If $\|\mathbf{K}\|$ ranged from 0 to $+\infty$, and ψ from 0 to 2π, this integral would result in a delta function. This is obviously not the case; nonetheless Jin *et al.* (1992) make this assumption, and the final expression for the approximation of the Hessian is given by:

$$H_a[\mathbf{r_i}, \mathbf{r_j}] = \frac{1}{2\Delta \mathbf{r_g}} \delta(\mathbf{r_j} - \mathbf{r_i}) \int d\mathbf{r_s}. \tag{18}$$

Here, $\Delta \mathbf{r_g}$ is the receiver spacing. Equation (18) is the final approximation to the Hessian that will be used in the quasi-Newton iterations described in Equation (12). Notice that it is invariant from iteration to iteration. Although this approximation worked fine with the examples shown in this work and also in Jin *et al.* (1992) it remains to be seen how it behaves for the situation of more complex data sets.

An Example

Consider the simple velocity model in Figure 2, which consists of just one horizontal interface. I generated the five shot gathers for this model shown in Figure 3.

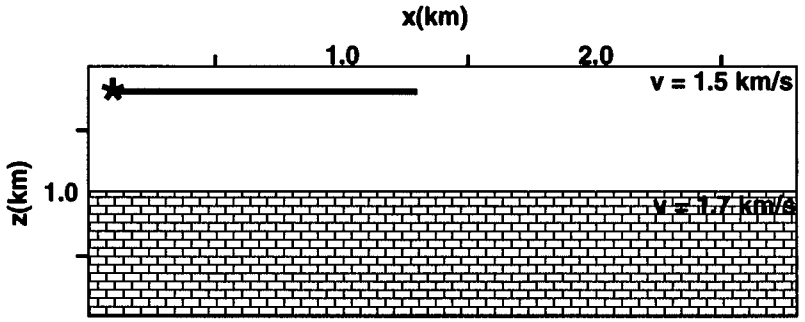

FIG. 2. One-layer velocity model. The length of a shot gather is illustrated by the horizontal line in the figure.

The forward modeling procedure used here is, as the inversion method, based on the Born approximation. To illustrate the action of regularization in amplitude inversion of noisy data, this is acceptable.

The objective of the inversion in this problem is, given a background velocity of 1.5 km/s, to estimate the magnitude of the velocity contrast to the second layer, i.e., .2 km/s. In this noise-free situation the result, obtained in three iterations, is perfect, and is illustrated in Figure 4. This plot illustrates the perturbations to the background for several receiver positions as a function of depth. The spikes peak at the depth of 1.0 km with magnitude equal .2 km/s as expected.

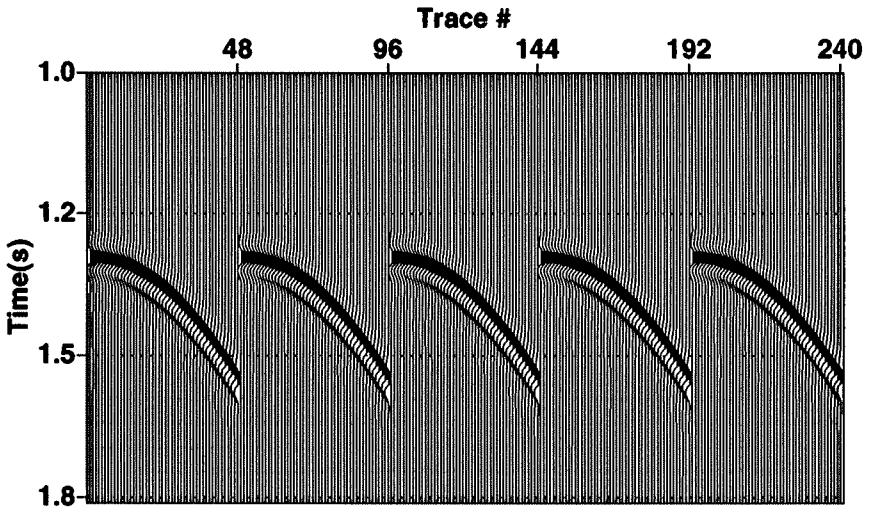

FIG. 3. Five shot gathers generated for Figure 2.

In the next section I illustrate the behavior of this inversion algorithm in the presence of noisy data. I also review how to regularize this procedure using Tikhonov's approach, and present numerical examples that show the lower sensitivity of the regularized inversion with respect to noise, as compared to non-regularized inversion.

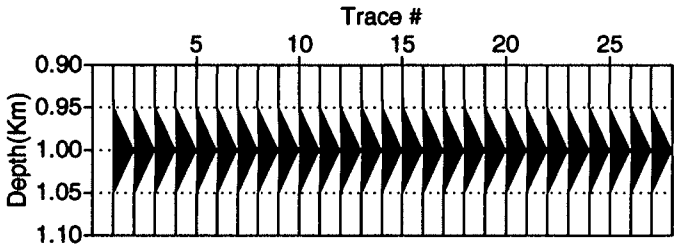

FIG. 4. Result of the inversion for the data in Figure 3.

EFFECTS OF REGULARIZATION ON AMPLITUDE INVERSION

Motivation

Consider the data set shown in Figure 5. It is the same data set illustrated in Figure 3, but with the addition of band-limited random noise such that the signal-to-noise ratio is now 2. Repeating the inversion procedure for this data set, I obtained the result illustrated in Figure 6, which is a considerably degraded version of the noise-free inversion result. The magnitude of the spikes contain errors larger than 50%.

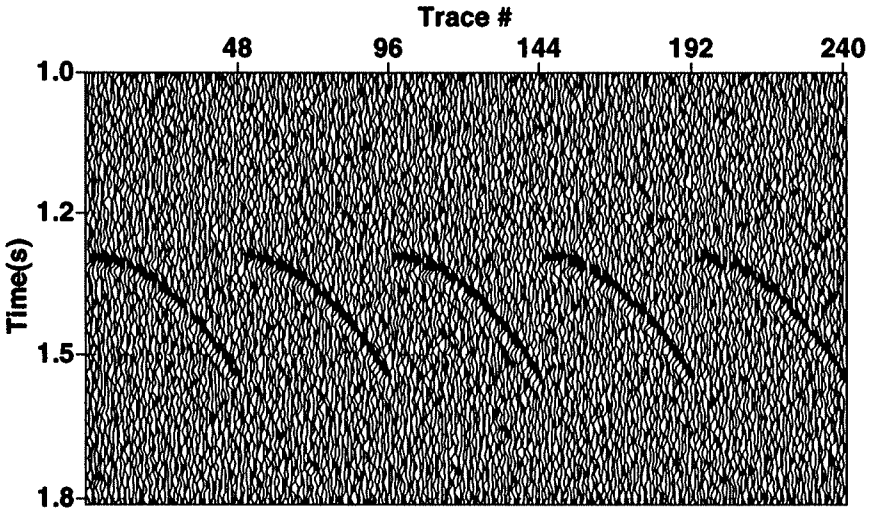

FIG. 5. Data of Figure 3 with band-limited noise. The signal to noise ratio is 2.

If just one shot gather of Figure 5 is used in the inversion, the final image, illustrated in Figure 7, is even a worse one. This is an expected result since the noise is attenuated when five shot gathers are used due to the larger data redundancy.

I applied Tikhonov regularization in the inversion algorithm discussed here, aiming at reducing its sensitivity with respect to perturbations (noise) in the data. I show the results later in this section, but first I briefly discuss this procedure and illustrate, with examples, the behavior of the regularized inversion algorithm for noisy data.

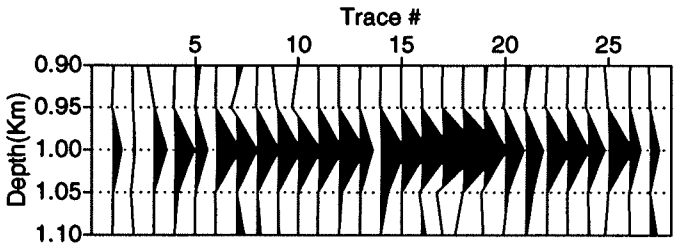

FIG. 6. Result of the inversion for the data in Figure 5.

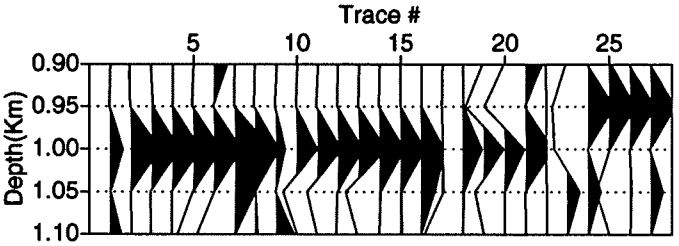

FIG. 7. Result of the inversion for only the first gather in Figure 5.

Basics of regularization

Consider the linear system

$$A\mathbf{x} = \mathbf{b}, \tag{19}$$

where A is an operator of the forward problem that computes the data $\mathbf{b}$ for a given model $\mathbf{x}$. The solution for $\mathbf{x}$ in Equation (19) is said to be ill-conditioned if it is non-unique and/or if a small perturbation on the data $\mathbf{b}$ corresponds to a large perturbation in the model $\mathbf{x}$. The fundamental idea of Tikhonov's regularization method (Tikhonov and Arsenin, 1977) is to replace the operator A by a family of approximate operators, functions of the so-called regularization parameter α, such that the solution ${x_\alpha}^*$ for each one of those parameters is well-conditioned, but, in some sense tends to $\mathbf{x}$ as α goes to zero. The approximated solution ${\mathbf{x}_\alpha}^*$ can be defined as the minimizer of the quadratic functional:

$$\|A\mathbf{x} - \mathbf{b}\|^2 + \alpha\|R\mathbf{x}\|^2. \tag{20}$$

in the domain of R. The matrix R defines the correlation between different elements of model space according to some criterion, usually related to model smoothness, As it will be shown latter, the regularization matrix R and the model covariance of the probability density function $\rho(\mathbf{m})$ are closely related.

Similarly it is possible to introduce ${\mathbf{x}^*}_\alpha$ as the solution to the regularized normal equations

$$(A^TA + \alpha R^TR)\mathbf{x} = A^T\mathbf{b}. \tag{21}$$

Note that R^TR is positive semidefinite, so a direct consequence of Tikhonov's method is to shift the spectrum of singular values of A^TA in the positive direction. Generally

this implies that the solution of the regularized normal system of equations should be less susceptible to perturbations in the data vector **b** (Karlsruhe and Lyngby, 1993). However, that is not always the case. As shown in Scales *et al.* (1990), a singular value decomposition of the matrix A is needed to understand what is actually being accomplished with regularization.

The regularization parameter α controls the influence of the penalty term on the optimization problem described in Equation (20). If it is chosen too small, Equation (20) is close to the original ill-posed problem, and the regularization would be of no or little effect. It α is too large, the problem solved would have little connection with the original Equation (19). Choosing the "optimum" value of α is a complicated matter in practice. Algorithms do exist with this intent (e.g., Hansen, 1992), and one of them is described later in this paper. A drawback is that the computational cost of those procedures is sometimes too high to make this method reasonable in problems such as amplitude seismic inversion.

Regularization of Amplitude Inversion

In the specific problem of seismic inversion, a possible approach would be to use R to add *a priori* knowledge about the model one seeks. For example, if lateral velocity variations are negligible, R can be constructed such that the scatterers in the horizontal direction are correlated with (or imposed to be similar to) each other. In this case just for illustrative purposes, assume that we have five scatterers per layer. The matrix R that correlates those scatterers is:

$$R = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 1 & 0 & -1 & 0 & 0 \\ 1 & 0 & 0 & -1 & 0 \\ 1 & 0 & 0 & 0 & -1 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & 0 & -1 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 1 & 0 & -1 \\ 0 & 0 & 0 & 1 & -1 \end{bmatrix}. \tag{22}$$

Other regularization schemes are available. It is also possible to regularize the inverse problem by attenuating the roughness of the final solution. This is accomplished by imposing small variations on the first or second derivatives of the model parameters. Schematically those two regularization procedures are represented for the situation of five scatterers by the matrix R in Equation (23) and (24), respectively.

$$R = \begin{bmatrix} 1 & -1 & 0 & 0 & 0 \\ 0 & 1 & -1 & 0 & 0 \\ 0 & 0 & 1 & -1 & 0 \\ 0 & 0 & 0 & 1 & -1 \\ 0 & 0 & 0 & 0 & 1 \end{bmatrix}, \tag{23}$$

$$R = \begin{bmatrix} -2 & 1 & 0 & 0 & 0 \\ 1 & -2 & 1 & 0 & 0 \\ 0 & 1 & -2 & 1 & 0 \\ 0 & 0 & 1 & -2 & 1 \\ 0 & 0 & 0 & 1 & -2 \end{bmatrix}. \tag{24}$$

To incorporate Tikhonov's regularization (for example the one described in Equation (22)) in the iterative asymptotic inversion presented in the last section, the following steps have to be undertaken. The regularization term should be added to the objective function in Equation (9). This results in the following expression:

$$\begin{aligned} \min_{\mathbf{m}} S(\mathbf{m}, \mathbf{r_0}) &= \min_{\mathbf{m}} \frac{1}{2} \int d\mathbf{r_s} \int d\mathbf{r_g} \int d\omega \, (\mathbf{u_s} - G\mathbf{m})^T Q(\mathbf{u_s} - G\mathbf{m}) J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi) \\ &\quad + \alpha \sum_i \sum_j \lambda(\mathbf{r_i}, \mathbf{r_j})(m(\mathbf{r_i}) - m(\mathbf{r_j}))^2. \end{aligned} \tag{25}$$

Here, α is the regularization parameter. The matrix R is constructed with the parameter $\lambda(\mathbf{r_i}, \mathbf{r_j})$, which has the value of 1 if the scatterers $\mathbf{r_j}$ and $\mathbf{r_i}$ are to be correlated, or 0 otherwise. $m(\mathbf{r})$ is the magnitude of the scatterer at $\mathbf{r_i}$. As in Equation (8) the corresponding matrix form of Equation (25) is:

$$\min_{\mathbf{m}} S(\mathbf{m}, \mathbf{r_0}) = \min_{\mathbf{m}} (\mathbf{u_s} - G\mathbf{m})^T Q(\mathbf{u_s} - G\mathbf{m}) + \alpha \, \mathbf{m}^T R^T R \, \mathbf{m}, \tag{26}$$

where R is the regularization operator that couples the model parameters according to $\lambda(\mathbf{r_i}, \mathbf{r_j})$.

The differentiation of the regularization term with respect to the model parameters is incorporated in the gradient (13), yielding the following expression:

$$\begin{aligned} \gamma(\mathbf{r}, \mathbf{r_0}) &= \frac{1}{2\pi} \int d\mathbf{r_s} \int d\mathbf{r_g} \, \frac{A(\mathbf{r_g}, \mathbf{r}, \mathbf{r_s})}{A^2(\mathbf{r_g}, \mathbf{r_0}, \mathbf{r_s})} \|\underline{p}\|^2 J(\mathbf{r_g}, \mathbf{r_s}, \xi, \psi) \mathcal{H}[\delta \mathbf{u_s}(\mathbf{r_g}, \mathbf{r_s}, t = \tau(\mathbf{r_g}, \mathbf{r}, \mathbf{r_s})] \\ &\quad + \alpha \sum_i \sum_j \lambda(\mathbf{r_i}, \mathbf{r_j})(m(\mathbf{r_i}) - m(\mathbf{r_j})). \end{aligned} \tag{27}$$

The approximation used in Equation (18) is used in the regularized version of the algorithm.

Finally the search direction in the Newton's iterations (Equation (12)) should take into account the regularization operator R, resulting in the following update scheme:

$$\mathbf{m_{n+1}} = \mathbf{m_n} - [H_a + R^T R]^{-1} \gamma(\mathbf{m_n}). \tag{28}$$

In the next section I assess the performance of the regularized asymptotic inversion with the data set of Figure 5.

Choosing the Regularization Parameter

Equation (20) defines an objective functional formed by two terms. The first one relates to the data misfit while the second incorporates the regularization scheme. An "optimum" regularization parameter α would provide an ideal balance between those two components, minimizing the regularization error and the perturbation error in the solution $\mathbf{x}_\alpha{}^*$. Several algorithms are available in the literature (Karlsruhe and Lyngby, 1993) for this purpose. Those methods are usually subdivided into two main categories, according to the assumption as to whether the magnitude of the perturbation on the data vector is known or not.

In this section I present the algorithm described in Scales *et al.* (1990) and used in this work. This procedure relies on the definition of a data misfit threshold as a stopping criterion for the iterative inversion. Here, this stopping criterion is satisfied when the root-mean-square (RMS) of the residual is less than the RMS amplitude of the noise for a given time window. In this case convergence is assumed.

To find an "optimum" α, the method proceeds as follows. The inverse problem is solved for several values of the regularization parameter. The data misfit for each one of the solutions $\mathbf{x}_\alpha{}^*$ is plotted as a function of α, as shown schematically in Figure 8.

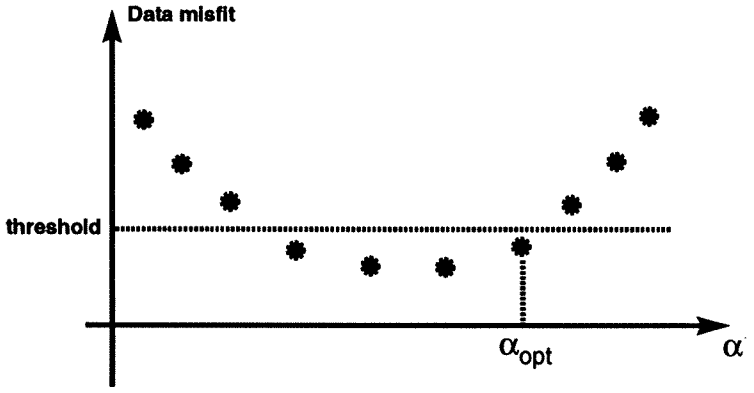

FIG. 8. Data misfit as a function of the regularization parameter α.

The "optimum" value of α, α^{opt} in Figure 8, is postulated as the largest regularization parameter for which the pre-specified data fit is achieved. Therefore the solution $\mathbf{x}_\alpha^{\mathbf{opt}}{}^*$ would have the desired data fitness and also be the most consistent with the *a priori* information used to build the operator R in Equation (20).

This can be expensive procedure to carry on in practice, since the full inverse problem is solved several times for different regularization parameters. Nonetheless, this procedure has been successfully applied in some situations as described in Scales *et al.* (1990) and Pratt *et al.* (1993).

An Example

Carrying out the inversion procedure for the data set shown in Figure 5 under the assumption that the medium is laterally homogeneous, the regularization described in Equation (22), leads to the result shown in Figure 9. This should be compared with Figures 4 and 6. The regularization was effective in reducing the sensitivity of the inversion procedure to the noise in the data, yielding scatterers shown in Figure 9 with the correct magnitude.

Figure 10 shows the convergence of a particular scattering point as a function of the number of iterations for the noise-free inversion and the regularized inversion. The smaller number of iterations needed to obtain convergence in the latter case is an indication that the condition number of the problem was reduced in comparison with the non-regularized approach. Note that convergence to 0% error in the presence of noise is not possible since the amplitudes of the data are corrupted.

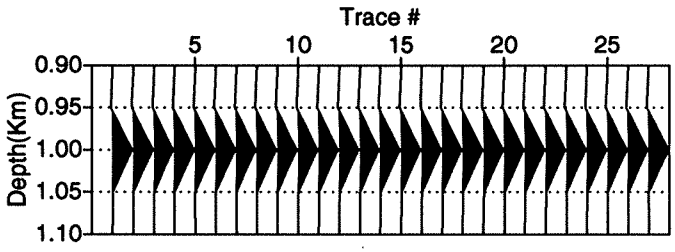

FIG. 9. Result of the inversion for in Figure 5 using regularization.

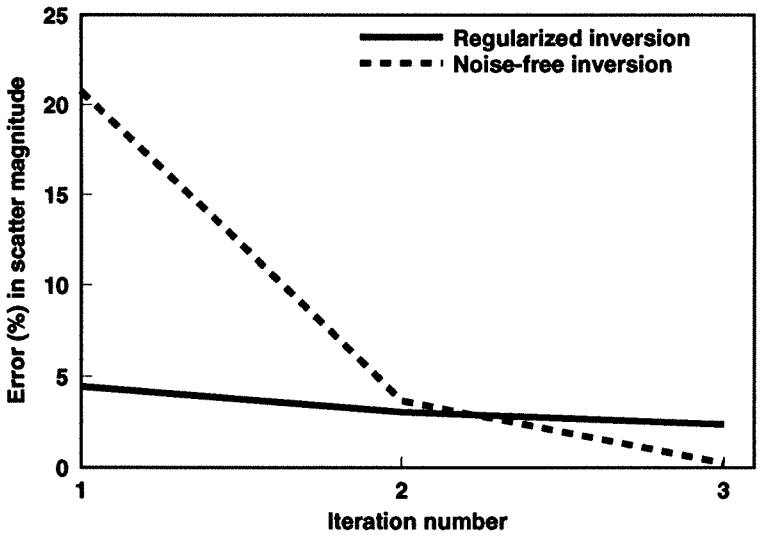

FIG. 10. Convergence plots for the noise-free and regularized inversions.

Using the first- and second-order derivative schemes described in Equations (23) and (24) to invert the data set of Figure 5, I obtained the results shown in Figures 11 and 12, respectively. The first-order derivative scheme provided comparable results

to the one shown in Figure 9. Such is not the case for the second-order derivative regularization. This scheme yields a smooth result, as illustrated in Figure 13. However, due to the limited lateral extent of the velocity model, and the fact that the smoothing operator R in Equation (24) does not allow rapid changes in the model parameters, this procedure does not produce satisfactory results.

Figure 14 shows a plot of the regularization parameter as a function of data misfit for the three regularization schemes. Notice that since the model is indeed laterally homogeneous, once the data misfit of the optimum solution is below the threshold, it becomes independent of the regularization parameter. According to this curve, a value for α of 5 for the regularization scheme described in Equation (22) and of 25 for the first-order derivative scheme were chosen. In the case of the second-order derivative regularization the value was arbitrary, since the curve never drops below the data misfit threshold.

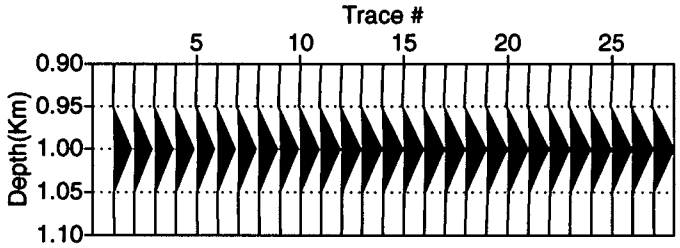

FIG. 11. Result with first-order derivative regularization.

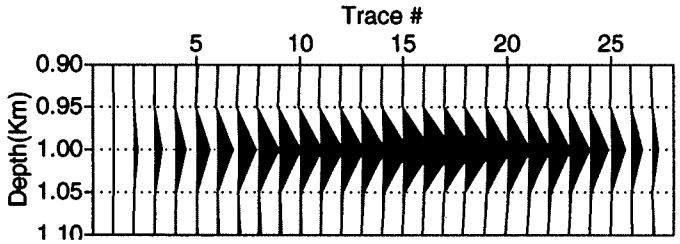

FIG. 12. Result with second-order derivative regularization.

RELATING BAYESIAN INVERSION AND REGULARIZATION

In the inversion procedure proposed by Jin *et al.* (1992) the solution of the inversion problem was defined as the minimizer of the weighted least-squares norm defined by Equation (7), or by Equation (26) when regularization is used. The simple statement of these equations implies important assumptions about the statistical nature of the noise in the observed data and on the correlation between model parameters. The purpose of this section is to describe what those assumptions are.

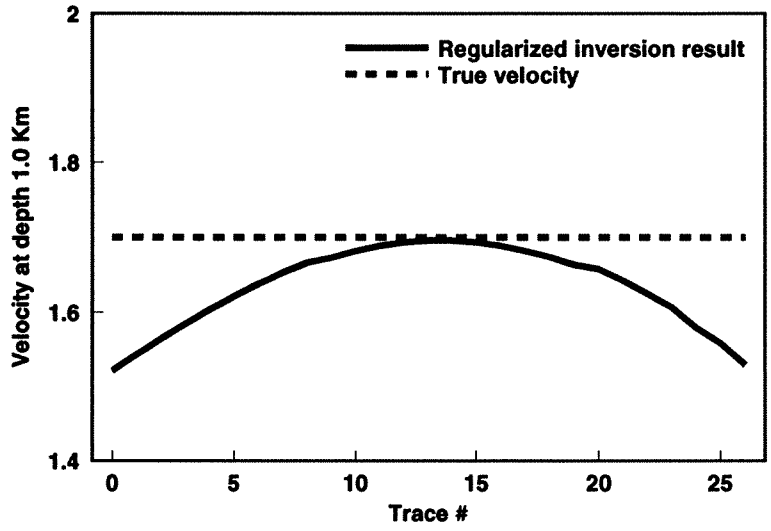

FIG. 13. Velocity at 1.0 km depth obtained from second-order regularization.

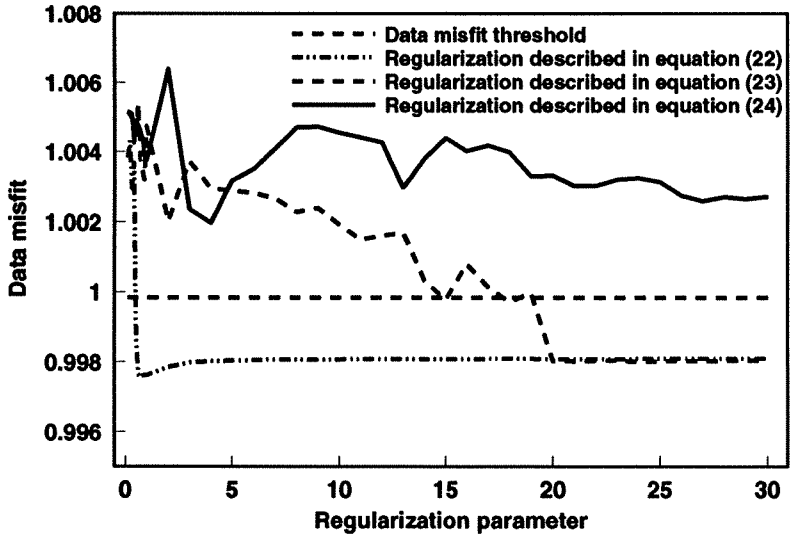

FIG. 14. Regularization curves for the data of Figure 5.

As mentioned before, for Gaussian statistics the *a posteriori* probability density function $\sigma(\mathbf{m})$ is given by Equation (2), repeated here for convenience.

$$\sigma(\mathbf{m}) \propto \exp\left[-\frac{1}{2}(g(\mathbf{m}) - \mathbf{d_{obs}})^T C_D^{-1}(g(\mathbf{m}) - \mathbf{d_{obs}}) + (\mathbf{m} - \mathbf{m_0})^T C_M^{-1}(\mathbf{m} - \mathbf{m_0})\right]. \tag{29}$$

When approaching an inverse problem using the Bayesian framework, we intend to determine which models, if any, are associated with large values of $\sigma(\mathbf{m})$. In other words we are interested in finding the maximizers of Equation (29). Under the Gaussian hypothesis this corresponds to minimizing the following quantity:

$$\min_{\mathbf{m}} S(\mathbf{m}) = \min_{\mathbf{m}} (g(\mathbf{m}) - \mathbf{d_{obs}})^T C_D^{-1}(g(\mathbf{m}) - \mathbf{d_{obs}}) + (\mathbf{m} - \mathbf{m_0})^T C_M^{-1}(\mathbf{m} - \mathbf{m_0}). \tag{30}$$

If m_0 is a null vector, Equations (30) and (26) are completely equivalent. The inverse of the weighting matrix Q^{-1} plays the role of a data covariance matrix C_D, and the product $\left[\alpha R^T R\right]^{-1}$ implements the model covariance matrix C_M. Therefore, the least-squares formulation of the inverse problem expressed in Equation (26) implies the assumption of uncorrelated Gaussian noise in the data and that the model parameters are also described by a Gaussian probability density function with zero mean. The model covariance matrix is defined by the regularization scheme.

In the next section, I consider the situation where the model covariance matrix is built by assuming a given correlation length (Tarantola, 1987) between the scatterers that form the interface between two acoustic layers. The results will be compared with those obtained by the regularized inversion procedure.

MODEL COVARIANCE ESTIMATION

The random sequence illustrated in Figure 15 was generated by filtering a random white process with a filter of a given correlation length. Figure 16 illustrates the autocorrelation of this series and a possible exponential fit that will be used later in building the covariance matrix. A portion of this sequence was used to construct the lateral variation in velocity of the second layer as illustrated in the model shown in Figure 17.

Five shot gathers generated for the velocity model of Figure 17, are illustrated in Figure 18. Notice the change in polarity of the reflection from the interface, caused by the lateral inhomogeneity of the second layer.

As described in Tarantola (1987), a sequence such as that in Figure 15 can be approximately modeled as a Gaussian process with covariance function given by:

$$C[i,j] = \sigma^2 e^{\left(\frac{|i-j|}{\Delta}\right)}, \tag{31}$$

where Δ is the correlation length of the sequence, and σ its standard deviation. A plot of this covariance matrix for a small problem consisting of 27 scatterers is shown

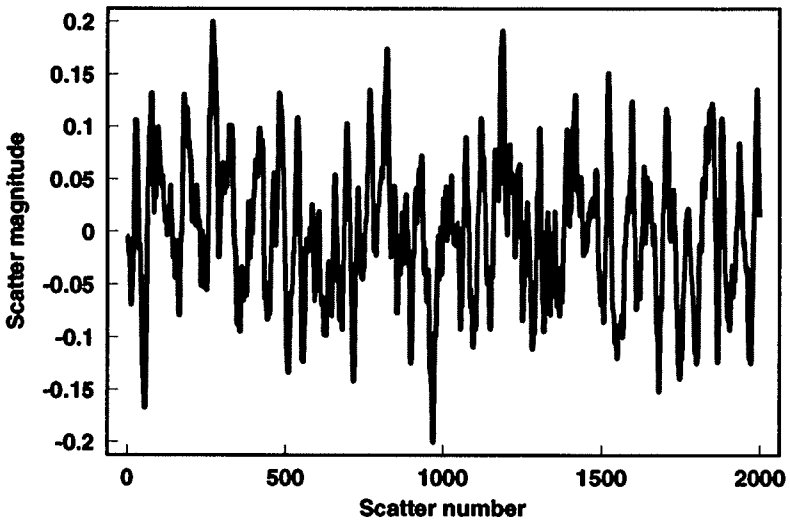

FIG. 15. Sequence with a given correlation length.

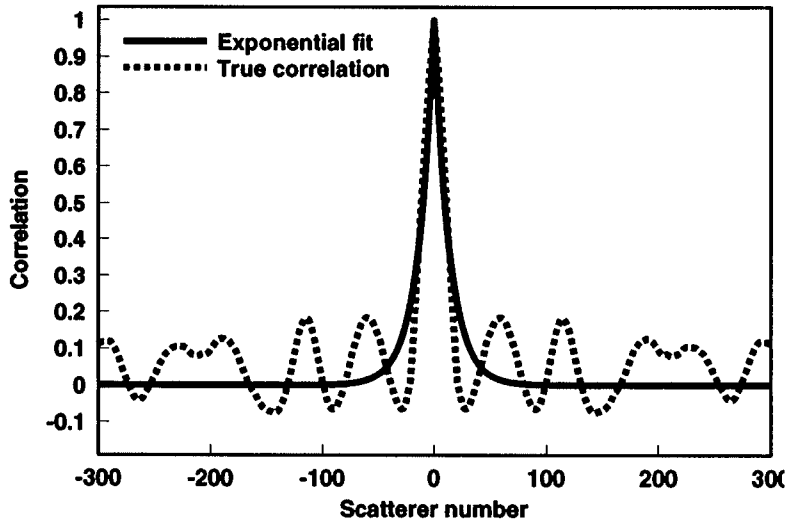

FIG. 16. Correlation and exponential fit for the sequence of Figure 15.

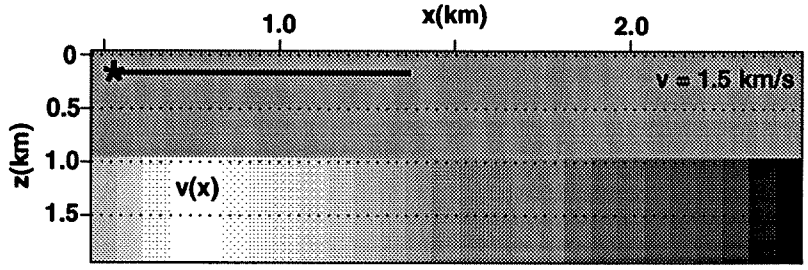

FIG. 17. Laterally inhomogeneous velocity model.

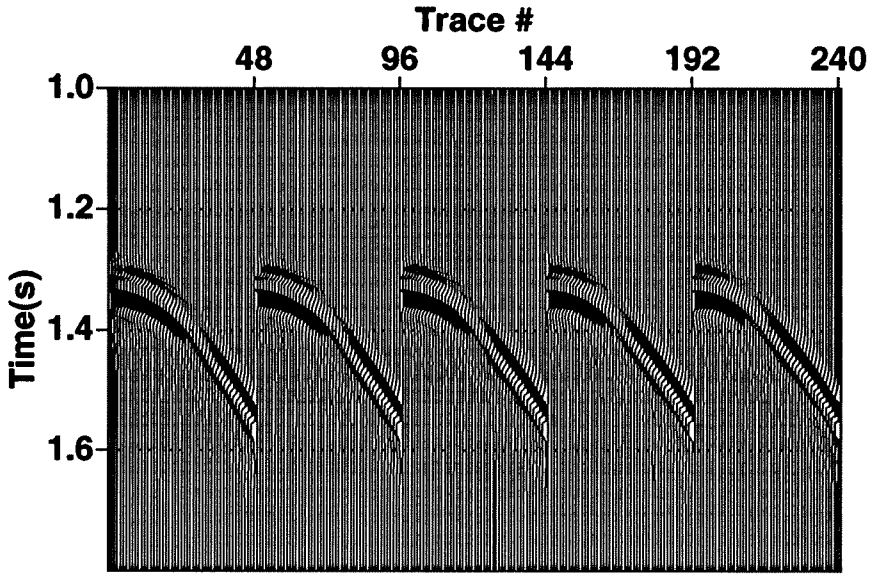

FIG. 18. Five shot gathers generated for the model of Figure 17.

in Figure 19. Here, the correlation length is determined by fitting an exponential of the form $e^{\frac{x}{\Delta}}$ to the autocorrelation of the scatterers (Figure 16). Although the fitting is not very accurate, it provides an initial estimation of the correlation matrix that is likely to be superior to the assumption that the model parameters are not correlated.

Again, the objective of the inversion is to determine the scatterer distribution given the velocity of the first layer. In this example, I compare the results obtained for the inversion of the model shown in Figure 17 for the target depth of 1 km obtained for the following cases: 1) non-regularized inversion; 2) regularized inversion using the operator R described in Equations (22), (23), and (24), and 3) Bayesian inversion with the covariance matrix is given by Equation (31).

Figure 20 shows the result obtained with the non-regularized inversion. The inverse result is reasonably close to the true scatterer distribution, represented by the dashed curve in the Figure. The residual of this final solution is illustrated in Figure 21.

Figures 22, 23 and 24 show the results of the inversion using the three regularization schemes described in Equations (22), (23) and (24), respectively. As expected, a poor result was obtained with the regularization implemented by Equation (22) (Figure 22), since the assumption of lateral homogeneity is a bad one for this situation. The other two approaches smoothed the final solution to a some degree, defined by the parameter α described earlier. I experimented several values for α, and the best results are illustrated by Figures 23 and 24, obtained with $\alpha = 1$ in both cases. The first and second-derivative regularizations provided superior solutions than the one obtained with the non-regularized inversion.

Finally, Figure 24 shows the inverse result when I used the model covariance

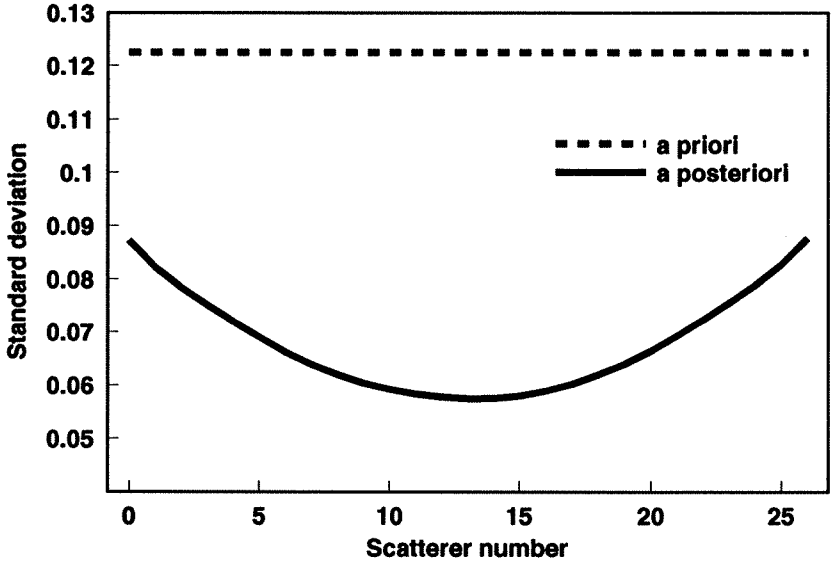

FIG. 19. Model covariance matrix as define in Equation (31).

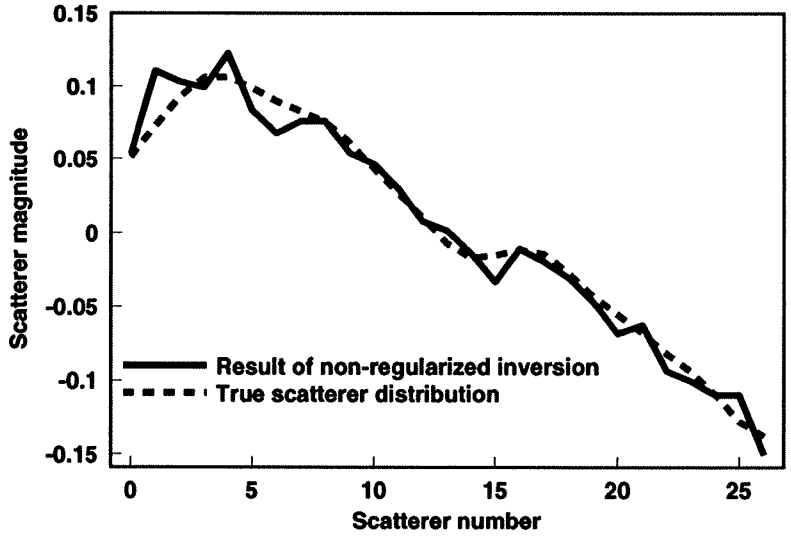

FIG. 20. Result of the non-regularized inversion.

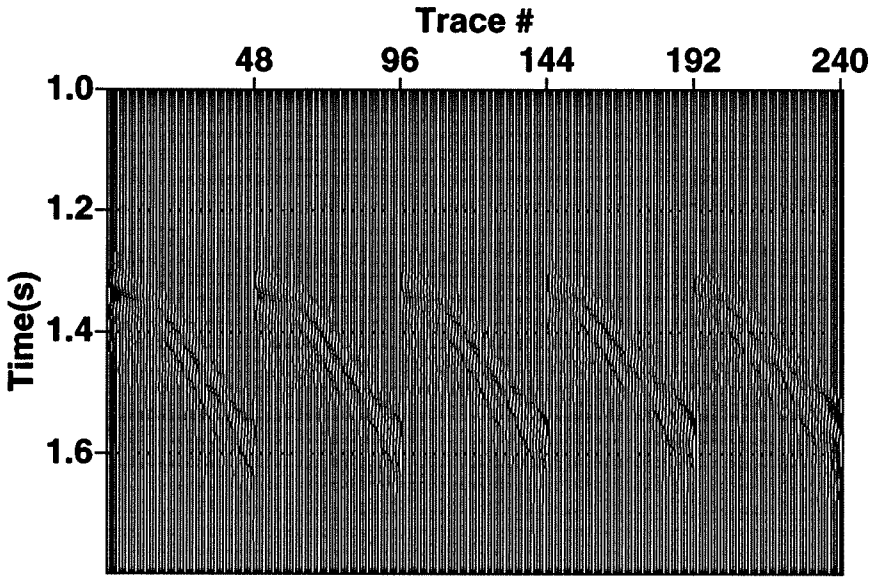

FIG. 21. Residual of the non-regularized inversion result.

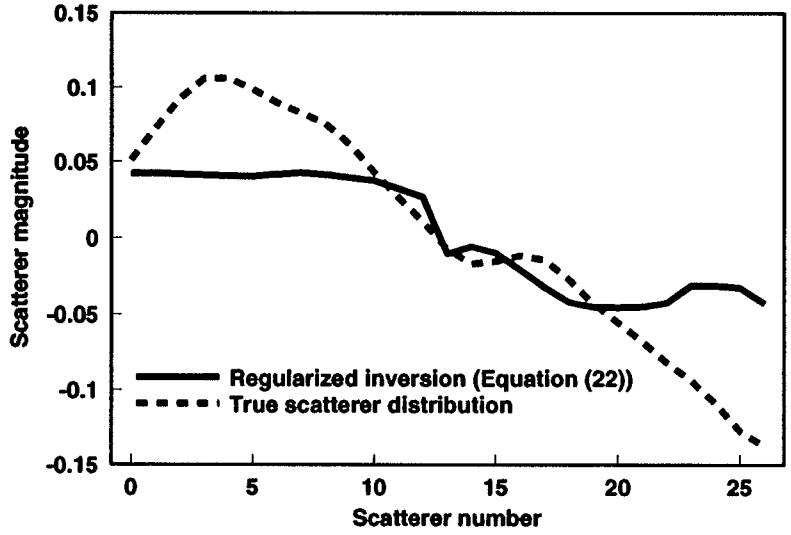

FIG. 22. Result of the regularized inversion using Equation (22).

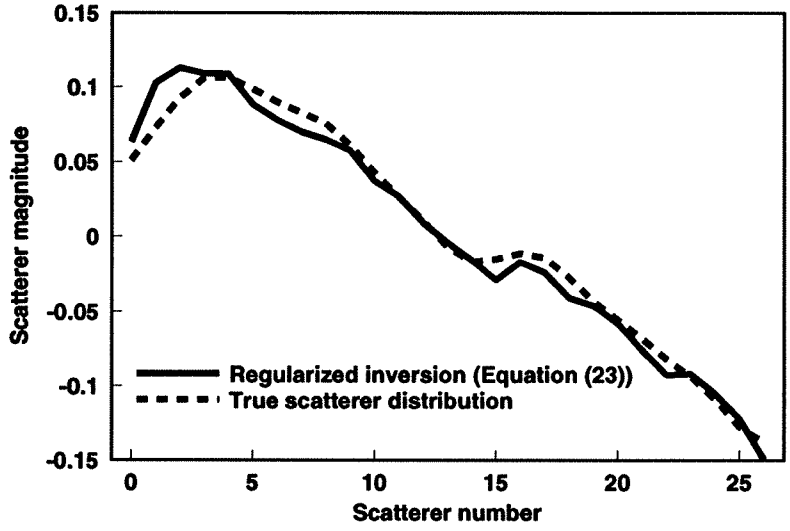

FIG. 23. Result of the regularized inversion using Equation (23).

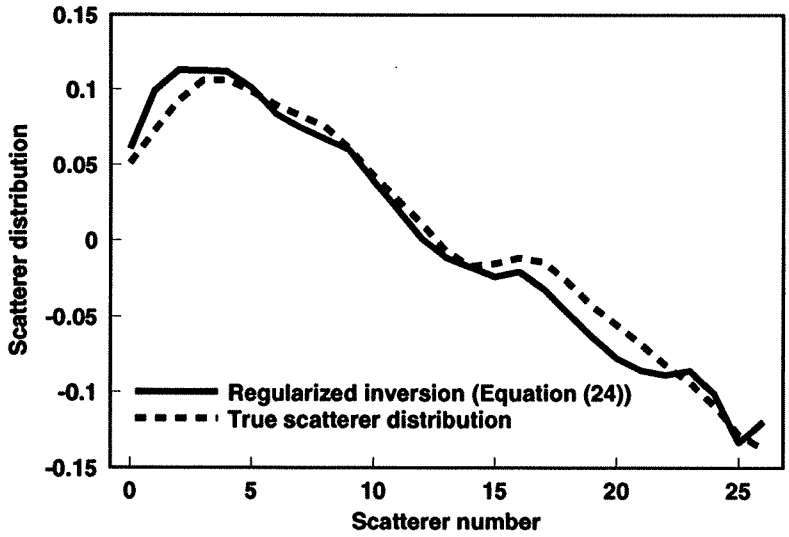

FIG. 24. Result of the regularized inversion using Equation (24).

shown in Figure 19. This solution is comparable to the ones obtained by the first and second-order regularization schemes. An appealing advantage of this approach is the absence of the parameter α, since the covariance matrix is built based on statistical considerations.

The fact that the results obtained with the derivative-based regularizations (Figures 23 and 24) and with the model covariance defined by Equation (31) are equivalent, should not be surprising since the true scatterer distribution is smooth. Examples dealing with more complex models are required to carry out a more thorough comparison between the two procedures.

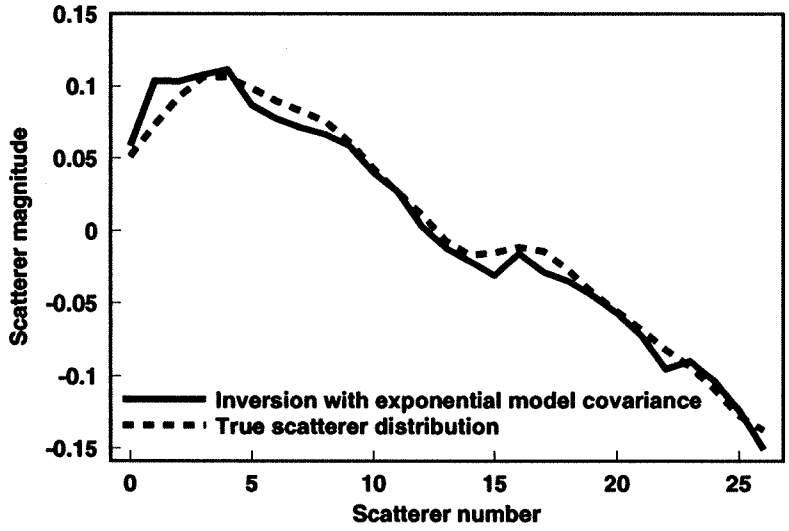

FIG. 25. Result of the exponential model covariance inversion.

Another important aspect of the Bayesian methodology is the possibility of computing the *a posteriori* model covariance, which provides insights on the resolution of the inverse problem solution. The *a posteriori* covariance $C_{M'}$ (Tarantola, 1987) is given by:

$$C_{M'} = \left[G^H Q G + {C_M}^{-1}\right]^{-1}. \tag{32}$$

Notice that, since the *a posteriori* covariance includes the model covariance C_M, this analysis would be of little significance if this matrix is built without resorting to the statistics of the model parameters, as it is done in the Tikhonov regularization.

The *a posteriori* and *a priori* standard deviations (square-root of the main diagonal of the *a posteriori* and *a priori* covariance matrices, respectively) are plotted in Figure 26. As expected, the *a posteriori* are smaller than the *a priori* standard deviations indicating that the inversion succeeded in reducing the uncertainties of the model parameters. Also, the deviations reduce towards the center of the model, which corresponds to a larger data redundancy available at this location.

The results presented in this work, including the solution of the inverse problem shown in Figure 25 are still preliminary and require further research. The likely next

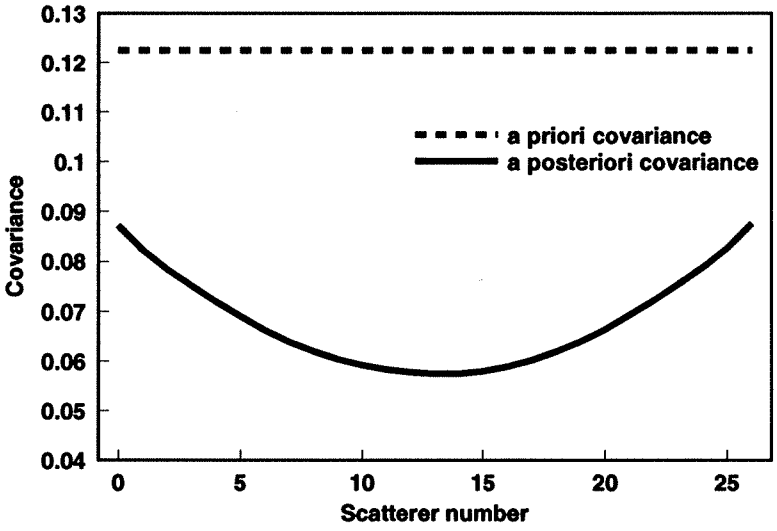

FIG. 26. Comparison between the *a priori* and *a posteriori* standard deviations.

step is to estimate the model covariance matrix directly from the data (Figure 15), without resorting to exponential models.

CONCLUSIONS

Here, I presented a study on model covariances under the framework of amplitude seismic inversion. For simple examples I compared inversion results when the model covariance was built taking into account statistical considerations about the underlying model with the case in which model covariances were derived from Tikhonov regularization. As indicated in this paper, the advantages of the former approach is the absence of a regularization parameter (α) and a more reliable *a posteriori* uncertainty analysis of the inverse problem solution.

Tikhonov regularization, although an effective procedure for reducing the sensitivity of the inversion method to perturbations in the data (noise), might not provide the necessary flexibility for incorporating more general information about the inverse problem. In a more complex situation, not only in terms of the difficulty posed by the inverse problem, but also in the presence of different levels of information one would like to consider in the inverse problem, Tikhonov regularization is probably a limited approach for constructing covariance matrices.

The simple results discussed in this paper motivate the use of realistic model covariance matrices in geophysical parameter estimation. For instance, one of the objectives to be accomplished in future work is to estimate model covariances directly from some source of information (for instance well logs), probably without resorting to exponential models as done in this paper.

Closely related to the estimation of model covariance matrices is the quantification of the uncertainties of the inverse problem solution. This is an very important subject

that will be addressed in this research.

ACKNOWLEDGMENTS

I thank John Scales, Norman Bleistein and Ken Larner for their critical review.

REFERENCES

Beylkin, G., 1985, Imaging of discontinuities in the inverse scattering problem by inversion of a causal generalized Radon transform: J. Math Phys., **26**, 99-108.

Bleistein, N., Cohen, J.K. and Stockwell, J., 1994, Mathematics of multidimensional inversion: To be published.

Cohen J., and Bleistein, N., 1979, Velocity inversion procedure for acoustic waves: Geophysics, **50**, 1077-1085.

Gray, S., 1994, Kirchhoff migration using eikonal equation traveltimes: Geophysics, **5**, 810-833.

Hansen, P., 1992, Analysis of discrete ill-posed problems by means of the L-curve: SIAM Rev. **34**, 561-580.

Jin, S., Mandariaga, R., Virieux, J. and Lambare, G., 1992, Two-dimensional asymptotic iterative elastic inversion: Geophys. J. Int., **108**, 575-588.

Karlsruhe, M. and Lyngby, P., 1993, Regularization methods for large-scale problems: Surv. Math. Ind., **3**, 253-315.

Pratt, R., McGaughey, W. and Chapman, C., 1993, Anisotropic velocity tomography: A case study in a near-surface rock mass: Geophysics, **58**, 1748-1763.

Scales, J., Docherty, P. and Gersztenkorn, A., 1990, Regularization of nonlinear inverse problems: imaging the near-surface weathering layer: Inverse Problems, **6**, 115-130.

Scales, J. and Tarantola, A., 1994, An example of geologic prior information in a Bayesian seismic inversion calculation: Center for Wave Phenomena Report, *CWP-159.*

Tarantola, A., 1987, Inverse problem theory: Elsevier.

Tikhonov, A.N. and Arsenin, V.Y., 1977, Solution of ill-posed problems: Wiley, NY.