Resolution of seismic waveform inversion: Bayes versus Occam

Wences Gouveia and John A. Scales
Department of Geophysics and Center for Wave Phenomena
Colorado School of Mines
Golden, Colorado 80401 USA

ABSTRACT
In Bayesian inference, information about models is posited *a priori*. This information, which may very well include features in the null space of the forward problem, affects both the computed models and the resulting resolution estimates. In Occam's inversion, on the other hand, the goal is to construct the smoothest model consistent with the data. This is not to say that one believes *a priori* that models are really smooth, but rather that the length scale associated with the smoothing is an indirect measure of resolution, since small-scale features not required to fit the data tend to be eliminated. In some cases the mathematical machinery of Bayesian inference resembles that of Occam's inversion, but the goals and interpretations of the two methods are rather different. To understand better the similarities and differences of these two approaches, we show an application of both methods to the problem of inferring the Earth's near surface elastic properties from reflection seismic data. On the one hand we derive *a priori* information about the Earth's layering from fine-scale borehole measurements; coupled with information about the noise in the data and the elastic forward modeling operator, we are able to compute the Bayesian *a posteriori* probability distribution. Pseudo-random realizations of this posterior probability may have features that are implied by the *a priori* information as well as the data, even if the former are not well resolved by the data. Then we solve the Occam's inversion problem by determining the maximum (second difference) regularization that allows for the data to be fit. In this case we estimate the resolution in terms of the degree of model smoothness implied by the data. In the Occam calculation we do not attempt to incorporate *a priori* information about the models, and the resulting averaged Earth model is very similar to the MAP model computed via Bayesian methods. But the error estimates associated with these models are rather different.

Key words: Seismic Waveform Inversion, Occam's Inversion, Bayesian Inference

1 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?

We will describe two fundamentally different strategies for solving inverse problems in the context of reflection seismic waveform inversion. In the Bayesian approach, we encapsulate prior information about layered Earth models in the form of probability distributions; distributions which are independent of the

data and may describe variations in material properties at length scales well below the resolution of the data. When such distributions are combined with probabilistic information about data uncertainties (both random and theoretical), it is possible to derive a final (*a posteriori*) probability distribution assimilating both types of information. All questions of resolution are answered via this posterior probability.

On the other hand, the construction of such prior probabilities is a controversial matter. So an alternative, and we think quite reasonable, approach is to find the most featureless models that are consistent with the data and whatever other deterministic information that is available. This featurelessness is achieved via Tikhonov regularization with a discrete second difference matrix. Resolution is implicitly determined by the length scale over which smoothing is possible while still fitting the data. This approach is called Occam's inversion by Constable *et al.* (1987).

1.0.1 Bayesian Inversion

In the Bayesian approach the three categories of information mentioned above are encapsulated as probability distributions. The net result of this is an equation assimilating all available information into a posterior probability distribution on the space of models $\mathbf{m}$ (Tarantola, 1987)

$$\sigma(\mathbf{m}) = \rho_M(\mathbf{m}) \int_D \frac{\rho_D(\mathbf{d})\Theta(\mathbf{d}|\mathbf{m})}{\mu_D(\mathbf{d})}\, d\mathbf{d} \tag{1}$$

where σ is the posterior probability, ρ_M is a prior distribution on the space of models, ρ_D is the distribution of data uncertainties, Θ is a conditional distribution which incorporates errors in the theory (i.e., in the forward model), and μ_D is a normalizing distribution sometimes taken to be the Jaynes null state of information. Everything on the right side of Equation 1, except for ρ_M can be thought of as a likelihood function, measuring the degree of data fit.

Equation (1) is rather general. For purposes of this paper we will simplify this by assuming that all the uncertainties in the problem can be described by stationary Gaussian distributions. In the Gaussian case the likelihood function reduces to

$$\sqrt{\frac{(2\pi)^{-n}}{\det C_D}} \exp\left[-\frac{1}{2}(g(\mathbf{m}) - \mathbf{d}_{\rm obs})^T C_d^{-1}(g(\mathbf{m}) - \mathbf{d}_{\rm obs})\right] \tag{2}$$

where $\mathbf{d}_{\rm obs}$ is the vector of observed data which dimension is n, C_D is the data covariance matrix and $g(\mathbf{m})$ is the forward operator. Similarly, the Gaussian prior distribution reduces to

$$\sqrt{\frac{(2\pi)^{-m}}{\det C_M}} \exp\left[-\frac{1}{2}(\mathbf{m} - \mathbf{m}_{\rm prior})^T C_M^{-1}(\mathbf{m} - \mathbf{m}_{\rm prior})\right] \tag{3}$$

where m is the number of model parameters and C_M is the covariance matrix describing the distribution model parameters about the *a priori* model $\mathbf{m}_{\rm prior}$. In the event that all uncertainty is Gaussian, the full posterior probability is the normalized product of Equations 3 and 2. If the forward operator is linear, then this distribution is itself a Gaussian.

In the Bayesian approach the information implied by $\rho_M(\mathbf{m})$ could very well be unresolved by the data. But if we really believe that we can encapsulate valid information about the Earth in $\rho_M(\mathbf{m})$, then it should be incorporated into the calculation. Conceptually this is simply a matter of looking at all the models that are reasonable according to $\rho_M(\mathbf{m})$ and seeing which ones fit the data.

1.0.2 Occam's Inversion

Let us accept the exponent of Equation (2) as being a useful measure of data fit irrespective of any Bayesian interpretation of the models. Then the set of models $\mathbf{m}$ such that

$$(g(\mathbf{m}) - \mathbf{d}_{\rm obs})^T C_d^{-1}(g(\mathbf{m}) - \mathbf{d}_{\rm obs}) \tag{4}$$

is less than or equal to some tolerance ϵ, is the set of models that fit the data within that tolerance. The Occam strategy is to find the smoothest model in this set. One way to do this is via the following constrained optimization problem:

$$\min_{\mathbf{m}} \|R\mathbf{m}\| \tag{5}$$

subject to

$$(g(\mathbf{m}) - \mathbf{d}_{\mathbf{obs}})^T C_d^{-1}(g(\mathbf{m}) - \mathbf{d}_{\mathbf{obs}}) \leq \epsilon \tag{6}$$

where R is a discrete second-difference operator. We can implement this practically as a weighted least squares problem with a Lagrange multiplier to control the tradeoff between model smoothness and data fit. By increasing the model smoothness until we can not longer fit the data to the prescribed tolerance, we will have found our model.

1.0.3 Occam versus Bayes

A similar weighted least squares problem arises when one attempts to find the model that maximizes the Bayesian posterior when using Gaussian distributions. Indeed the discrete difference operator R does define some sort correlation among the model parameters: the smoother the

model, the more correlated its parameters. If we were to solve the Occam problem, and thereby determine the maximum degree of smoothness consistent with the data (as defined by Equation (4)), and then incorporate this smoothness into the Bayesian problem as an *a priori* covariance matrix $C_M^{-1} = \lambda R^T R$, where λ is a constant, then the two problems would be almost equivalent. However, this is a rather extreme example because the whole reason for trying to solve the problem from a Bayesian point of view is to have the ability to incorporate data-independent model information. And if such information is available should we not be able to construct a more informative prior covariance matrix C_M than would be implied by the data alone, as in Occam?

How to incorporate realistic deterministic prior information into an Occam style inversion is an interesting issue, but not one we will address in this paper. Here we consider the problem of using elastic reflection seismograms (synthetically generated with known Gaussian errors) to infer P-wave impedance, S-wave impedance and density of the upper crust. We will use 4 seconds of data to image the Earth down to 1.5 km, and typical P-wave wavelengths are on the order of 300 m.

On the one hand, we will use Occam to find the smoothest model that fits the data (achieves a normalized χ^2 of one). For the Bayesian problem, on the other hand, we will construct an *a priori* covariance matrix for layered Earth models using well logs, a form of geologic information commonly available to exploration seismologists. Since well logs are made using much higher frequencies than the reflection seismic data, they have resolution below 1 m. (However, to keep our life simple, we will use a 10 m discretization.) As a result, the *a priori* covariance matrix will contain information about layered Earth models well below the resolution of the surface seismic data. It is our goal to see to what extent this prior information influences the computed models and their resolution, and to make a fair comparison of the results of realistic Bayesian and Occam calculations. Finally, we obtain analytic expressions for the bias of the estimates obtained in the Occam's and Bayesian approaches.

2 FORMULATION OF THE TWO PROBLEMS

For both the Bayesian and Occam calculations we make the following assumptions

- The observed data $\mathbf{d_{obs}}$ are contaminated by Gaussian errors with known covariance C_D.
- The models $\mathbf{m}$ are elements of R^m where m is fixed throughout. In particular $m = 150$, which is three times the number of layers, there being one unknown P-wave impedance, S-wave impedance and density for each of the 50 layers.
- The forward operator $g(\mathbf{m})$ is known and exact. In our case this is an elastic reflectivity-based synthetic seismogram algorithm (Fuchs and Müller, 1971).
- The Bayesian *a priori* model $\mathbf{m_{prior}}$ is taken to be a long wavelength approximation of the true model and is used as the initial model in the Occam calculation.

Now let us define the specifics of the two calculations. The goal of the Bayesian calculation is to compute the posterior probability distribution:

$$\sigma(\mathbf{m}) \propto \exp -\frac{1}{2}\left[(g(\mathbf{m}) - \mathbf{d_{obs}})^T {C_D}^{-1}(g(\mathbf{m}) - \mathbf{d_{obs}}) + (\mathbf{m} - \mathbf{m_{prior}})^T {C_M}^{-1}(\mathbf{m} - \mathbf{m_{prior}})\right]. \quad (7)$$

Due to the nonlinearity of g, resolution analysis of this probability distribution is computationally expensive. To simplify the calculations we do a nonlinear optimization to find the model ($\mathbf{m_{map}}$) that maximizes $\sigma(\mathbf{m})$. A Gaussian approximation of this density is then made about this model, yielding (Tarantola, 1987):

$$\sigma(\mathbf{m}) \propto \exp\left[(\mathbf{m} - \mathbf{m_{map}})C_{M'}^{-1}(\mathbf{m} - \mathbf{m_{map}})^T\right], \quad (8)$$

where

$$C_{M'} = \left[G^H C_D^{-1} G + C_M^{-1}\right]^{-1}, \quad (9)$$

is the *a posteriori* covariance matrix, a resolution measure that quantifies the uncertainties about $\mathbf{m_{map}}$. G is the Fréchet derivative of g, evaluated at this model. Therefore the Bayesian problem, which we refer to as Problem 1, can be stated as:

Problem 1. Bayesian Inversion

- Compute $\max_{\mathbf{m}} \sigma(\mathbf{m})$ defined in in Equation (7) by solving a nonlinear optimization problem.
- Linearize the forward operator g about the resulting MAP model $\mathbf{m_{map}}$.
- Compute the *a posteriori* covariance matrix.

We consider now the Occam's approach to inversion. Let R be a a Tikhonov second difference regularization matrix. Possible forms for this operator are

$$R = \begin{bmatrix} -2 & 1 & 0 & 0 & \dots & 0 & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -2 & -1 & \dots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & 0 & \dots & 1 & -2 \end{bmatrix}, \quad (10)$$

or

$$R = \begin{bmatrix} 0 & 0 & 0 & 0 & \dots & 0 & 0 \\ 1 & -2 & 1 & 0 & \dots & 0 & 0 \\ 0 & 1 & -2 & 1 & \dots & 0 & 0 \\ & & & \vdots & & & \\ 0 & 0 & 0 & \dots & 1 & -2 & 1 \\ 0 & 0 & 0 & \dots & 0 & 0 & 0 \end{bmatrix}. \quad (11)$$

However, there exists an important difference between these two regularization operators. The latter maps constant and linearly increasing vectors into a zero vector, and therefore has the same null space as the continuous operator. Operator (10) is not singular. It will certainly regularize the inverse problem, no matter what the null space of G. Here we use (10). Next we state the Occam's problem.

Problem 2. Occam's Inversion
For fixed λ solve the weighted, regularized least squares problem

$$\min_{\mathbf{m}} \Gamma(\mathbf{m}) = \frac{1}{2}\left[(g(\mathbf{m}) - \mathbf{d_{obs}})^T {C_D}^{-1}(g(\mathbf{m}) - \mathbf{d_{obs}}) + \lambda(\mathbf{m} - \mathbf{m_{prior}})^T R^T R(\mathbf{m} - \mathbf{m_{prior}})\right].$$

Increase λ until

$$(g(\mathbf{m}) - \mathbf{d_{obs}})^T {C_D}^{-1}(g(\mathbf{m}) - \mathbf{d_{obs}}) = \epsilon$$

can no longer be achieved.

This regularized nonlinear least squares problem can be solved using the same optimization algorithm as used in the Bayesian MAP calculation. Then it's simply a matter of repeating the calculation until we find the optimal regularization parameter λ. The model $\mathbf{m_{occam}}$ associated with the largest value of λ that still fits the data is the solution to this problem.

3 RESOLUTION

In the Bayesian calculation, all questions of resolution are addressed via the posterior probability density $\sigma(\mathbf{m})$. Even assuming Gaussian errors and prior information, for a sufficiently nonlinear forward operator g, we may have to resort to Monte Carlo integration or sampling methods to extract confidence sets or other measures of resolution from $\sigma(\mathbf{m})$. In this paper we are going to sidestep this issue by assuming that we can make a Gaussian approximation to $\sigma(\mathbf{m})$ in a neighborhood of the $\mathbf{m_{map}}$. This is equivalent to linearizing the forward operator about this model. That being the case, then the complete Bayesian picture of resolution is provided by *a posteriori* covariance matrix

$$C_{M'} = \left[G^H C_D^{-1} G + C_M^{-1}\right]^{-1}. \quad (12)$$

In the numerical examples we will extract information from this equation in two different ways. First, by looking at the (square roots) of the diagonal elements, we can put 1-σ error bars on the MAP model. But these error bars can be misleading since they ignore correlations among the parameters. So we also generate a tour of pseudo-randomly simulated models sampled according Equation (12) and $\mathbf{m_{map}}$. By looking at the distribution of models sampled in this way, we get a clear picture of which features in the model are well resolved and which are not.

Now since the prior covariance matrix C_M appears in Equation (12), it is obvious that the information we extract from the fine scale features in the well log influence the resolution; to a greater or lesser extent depending on the relative weights associated with $G^H C_D^{-1} G$ and C_M^{-1}. This is in contrast to the Occam's solution, which seeks a level of model smoothness based solely on data fit. It seems reasonable then that the model produced by Occam would, in general, be smoother than the models sampled from the Bayesian posterior. Which is not to say that the MAP model will be less smooth than the Occam model. The MAP model, being the center of the Gaussian distribution, and thus an average, may very well be significantly smoother than a typical realization of this distribution. This is another argument for sampling models from the Bayesian posterior. If we are Bayesians, then we believe that the Earth models we generate should reflect the fine-scale information contained in the log, even if it is unresolvable by the data. Whereas if we are Occamists, then we try to find the broadest average of the Earth that is consistent with the data. That way we can be confident that the features that are in the model are required to fit the data. So in the Occam calculation such a degree of smoothness is one measure of resolution.

An alternative measure in the Occam's inversion is done by propagating the data uncertainties, represented by C_D, into the model space. That provides a covariance matrix that quantifies the uncertainties on the estimate of the subsurface parameters obtained by solving Problem 2.

As in the Bayesian case the assumption of linear forward modeling is necessary for such an uncertainty analysis. This can be accomplished by linearizing g about $\mathbf{m_{occam}}$, hence reducing Problem 2 to a weighted, regularized linear system of equations that relates model and data perturbations by:

$$\tilde{\mathbf{A}}\delta\mathbf{m} = \tilde{\delta}\mathbf{d}, \quad (13)$$

where,

$$\tilde{\mathbf{A}} = \begin{bmatrix} C_D^{-\frac{1}{2}} G \\ \sqrt{\lambda}\, R \end{bmatrix},$$

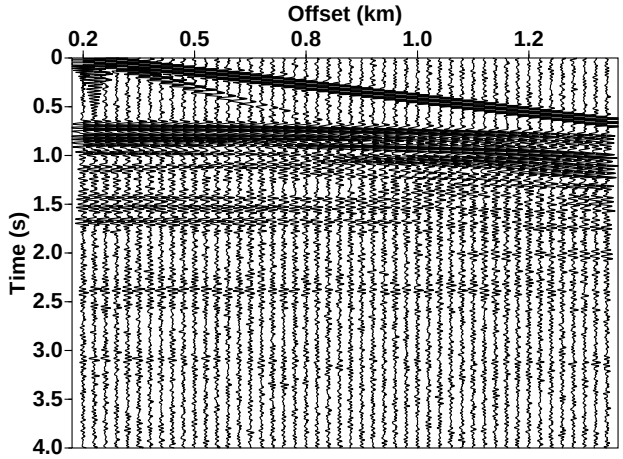

Figure 1. Vertical component of the displacements generated for the elastic model of Figure 2 corrupted by Gaussian noise band-limited to the same frequency as the data.

$$\tilde{\delta}\mathbf{d} = \begin{bmatrix} C_D^{-\frac{1}{2}}\delta\mathbf{d} \\ 0 \end{bmatrix} = \begin{bmatrix} C_D^{-\frac{1}{2}}\left(\mathbf{d_{obs}} - g(\mathbf{m_{prior}})\right) \\ 0 \end{bmatrix}. \quad (14)$$

The uncertainties of the estimate $\mathbf{m_{occam}}$ are given by:

$$\begin{aligned} E\left[\delta\hat{\mathbf{m}}\delta\hat{\mathbf{m}}^T\right] &= E\left[\tilde{\mathbf{A}}^\dagger\tilde{\delta}\mathbf{d}\left(\tilde{\mathbf{A}}^\dagger\tilde{\delta}\mathbf{d}\right)^T\right] \\ &= \tilde{\mathbf{A}}^\dagger E\left[\tilde{\delta}\mathbf{d}\tilde{\delta}\mathbf{d}^T\right]\tilde{\mathbf{A}}^{\dagger T} \\ &= \tilde{\mathcal{A}}^\dagger C_D^{-\frac{1}{2}} E\left[\delta\mathbf{d}\delta\mathbf{d}^T\right] C_D^{-\frac{1}{2}}\tilde{\mathcal{A}}^{\dagger T} \\ &\approx \tilde{\mathcal{A}}^\dagger\tilde{\mathcal{A}}^{\dagger T}. \end{aligned} \quad (15)$$

Here, $\tilde{\mathbf{A}}^\dagger$ is the pseudo-inverse of $\tilde{\mathbf{A}}$, and $\tilde{\mathcal{A}}^\dagger$ is $\tilde{\mathbf{A}}^\dagger$ with the last N columns set to zero, where N is the dimension of the model space. $\hat{\delta\mathbf{m}}$ is the solution obtained from Equation (13). Notice that since G is obtained by linearizing $g(\mathbf{m})$ about $\mathbf{m_{occam}}$, $E\left[\delta\mathbf{d}\delta\mathbf{d}^T\right] \approx C_D$, neglecting linearization errors. Thus, the last equation in (15).

NUMERICAL CALCULATIONS

Here, we consider the problem of estimating *P*- and *S*-wave impedance, and density profiles from a synthetic surface seismic data set, generated for the elastic velocity model illustrated in Figure (2). These data, shown in Figure (1), was corrupted by correlated Gaussian noise. Figure (2) plots only the target zone for the inversion. 50 layers of 0.01 km thickness will be considered in this problem. In the following we compare the solutions to problems 1 and 2 when inverting this synthetic data set. We approach those problems in two ways. In the first we

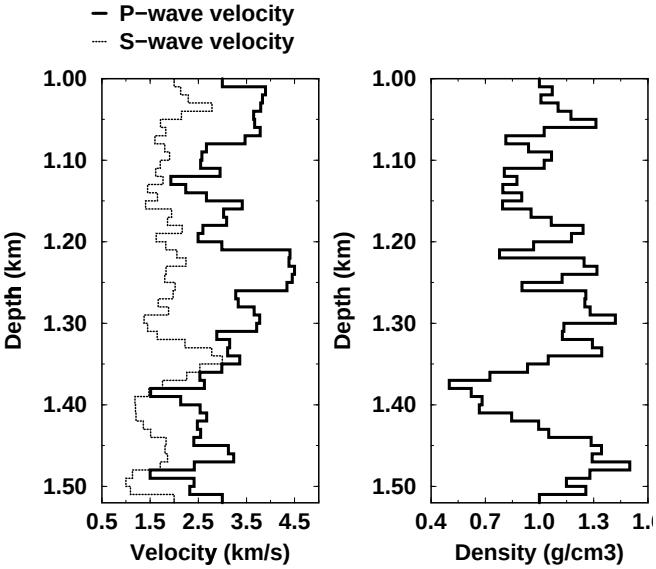

Figure 2. 50-layer elastic model. The medium for depths less than 1 km is homogeneous with *P*-wave velocity, *S*-wave velocity and density equal to 3 km/s, 2 km/s and 1 gm/cm^3, respectively.

don't resort to any kind of linearization, and tackle the problem via nonlinear optimization procedures. In Appendix A we linearize the problem about model $\mathbf{m_{prior}}$, hence obtaining a linear system of equations that is solved iteratively. It is worth mentioning that the criterion to accept a model as a solution is based on data fitness, which is measured according to Equation (4), using the nonlinear forward operator in the first case, and its linearization in the second case.

The nonlinear case

Problem 1

Gouveia (1996), discusses in detail the results obtained when the inversion was carried on under the Bayesian framework, in a nonlinear fashion. In his implementation, no assumptions were made about the structures of the covariance matrices C_D and C_M. Those were built directly from the correlation of the noise in the observed data (assumed known in this experiment) and from the correlation length of the true impedance and density profiles (Figure (2)). Figure (3) illustrates the MAP model, along with the true subsurface parameters and the model $\mathbf{m_{prior}}$. The error bars in the figure are the square-roots of the main diagonal of the *a posteriori* covariance matrix (9), as discussed before. The following conclusions can be drawn from this experiment.

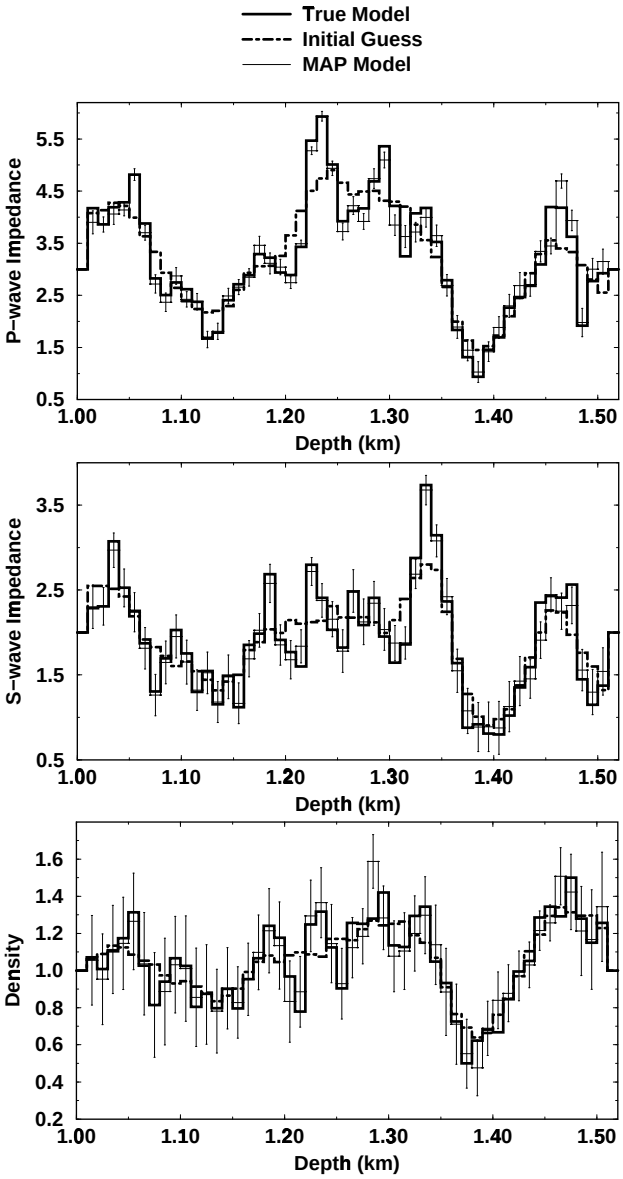

Figure 3. The MAP model estimated from the data set illustrated in Figure 1. The true model and the initial guess used for the nonlinear optimization calculation is also shown.

- The MAP solution reconstructs the subsurface parameters with a good accuracy. Such a model fits the observed data to one standard deviation;
- the error bars include the true model, for most of the depth range within which the inversion was attempted;
- the *P*-wave impedance error bars are the smallest, reflecting the fact that they are the best resolved parameters in this experiment, and
- the density error bars are the largest, indicative of the poor resolution of this parameter available in seismic amplitudes. Moreover, since the initial guess for the density profile is also within the error bars, it is as good a model as the MAP model. Therefore, the only information on density provided by the inversion was the error bars, representing the range within each the true density might lie.

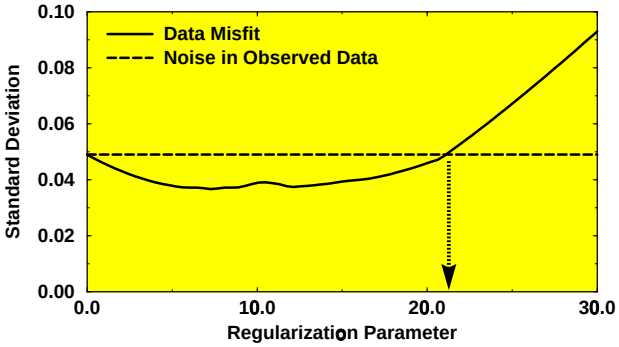

Figure 4. Regularization parameter λ plotted as a function of the standard deviation of the data misfit associated with the optimum model obtained from the regularized nonlinear inversion. The dotted line corresponds to one standard deviation of the noise. Therefore the intersection of the solid and dotted curves corresponds to having fit the data to one standard deviation, on average. This graph was smoothed over a λ interval of length five.

Problem 2

In order to find the model associated with the largest regularization parameter λ, a search procedure should be implemented. The one used here, was simply to increase this parameter to a value at which the data could be no longer fit. Figure (4) shows the standard deviation of the data misfit associated with the optimum model, as a function of the regularization parameter. The model $\mathbf{m}_{\mathbf{occam}}$ associated with the optimal regularization parameter (indicated in the figure by the arrow) is shown in Figure (4). The figure also displays the true model and the initial guess. This result should be compared with the model displayed in Figure (3). Figure (6) plots $\mathbf{m}_{\mathbf{map}}$ and $\mathbf{m}_{\mathbf{occam}}$ in the same graph to make the comparison easier. Although both models are similar, it is not difficult to see the higher accuracy of $\mathbf{m}_{\mathbf{map}}$ compared to $\mathbf{m}_{\mathbf{occam}}$ (Compare *P*-wave impedances at the depths of 1.25 and 1.30 km, and *S*-wave impedances at depth 1.34 km, for instance). This illustrates, at least for this specific example, that the use of *a priori* information consistent with the true subsurface parameters (in this case, the correlation length of the medium) can emphasize true features in the inverse problem solution. Such features might be smoothed out by the Occam's procedure since they might not be required to fit the data.

Although $\mathbf{m}_{\mathbf{occam}}$ is smoother than $\mathbf{m}_{\mathbf{map}}$ by con-

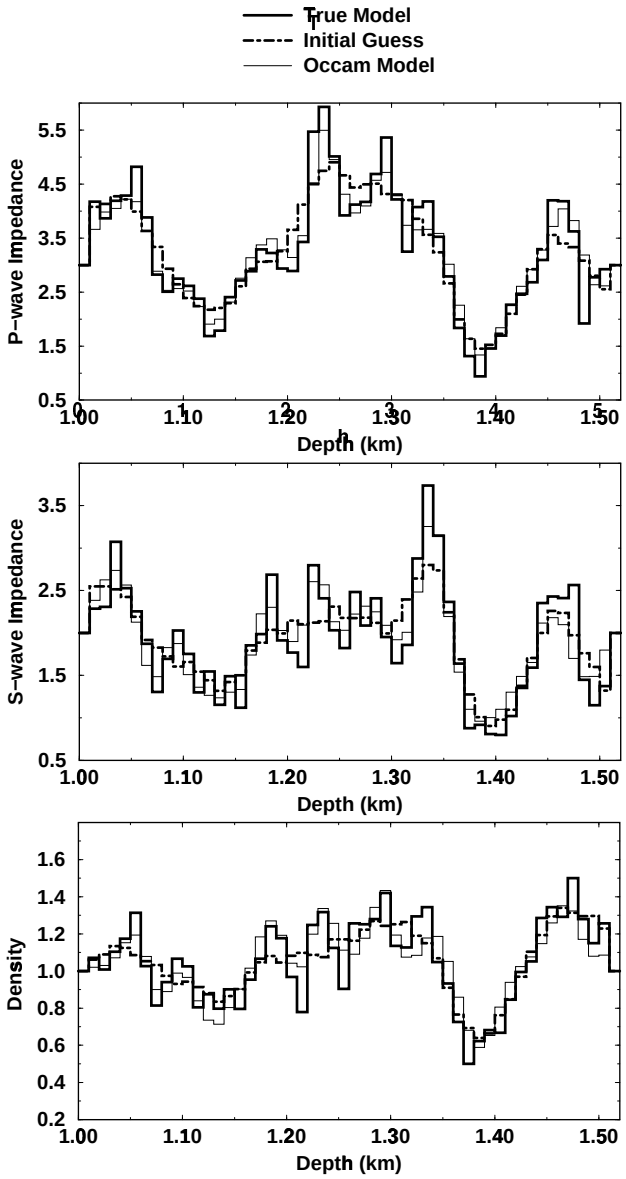

Figure 5. The Occam's model estimated from the data in Figure 1 using a non-linear, regularized optimization procedure. The true model and the initial guess are also displayed.

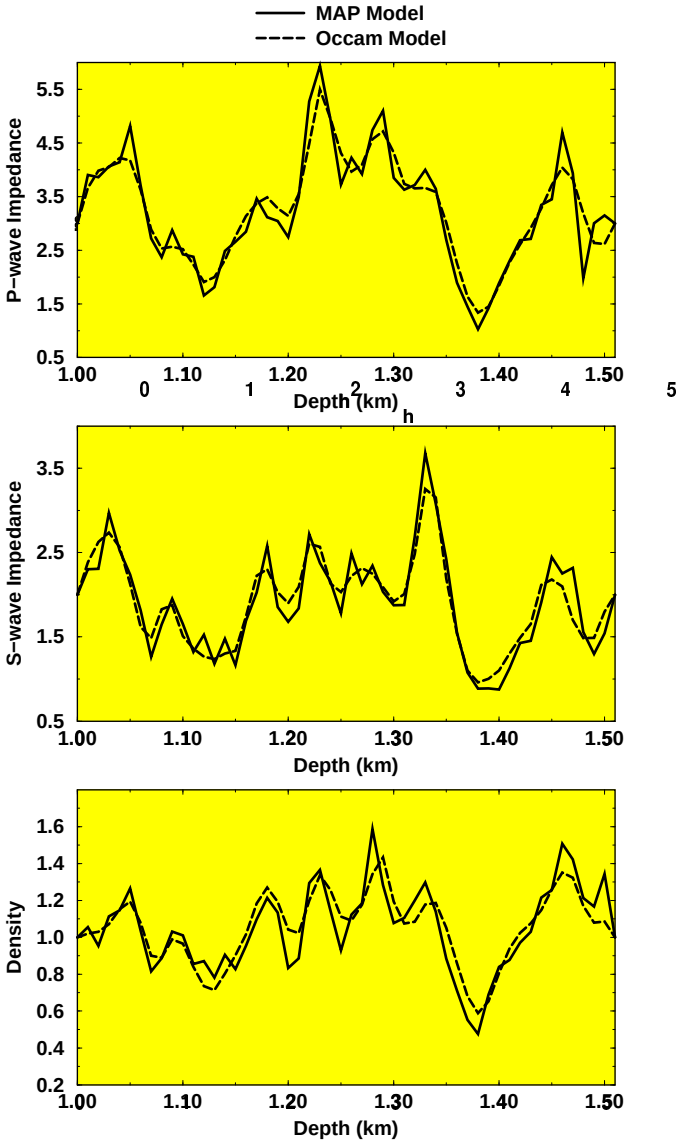

Figure 6. Comparison of the models obtained from Bayesian and Occam inversions.

struction, it is not much smoother, especially for the *P*-wave impedance profile. This behavior is not complicated to explain. If one thinks of regularized least-squares as an averaging like procedure, and considering that the approach followed to compute $\mathbf{m}_{\mathbf{map}}$ involves solving a sequence of regularized least-squares systems, this model should be smooth.

On the Resolution of the Estimates

In this section we discuss the resolution provided by the Bayesian and Occam techniques by examining the covariance matrices about $\mathbf{m}_{\mathbf{map}}$ and $\mathbf{m}_{\mathbf{occam}}$ for the nonlinear problem. In particular, with respect to the Bayesian technique, we sample the *a posteriori* probability distribution $\sigma(\mathbf{m})$, thus generating pseudo-realizations of subsurface models that are consistent with the *a priori* information and observed data.

The resolution of the the forward modeling operator $g(\mathbf{m})$ is discussed in Appendix B. There we discuss which features of the model space (long- or short-wavelength components of impedance and density) can be resolved

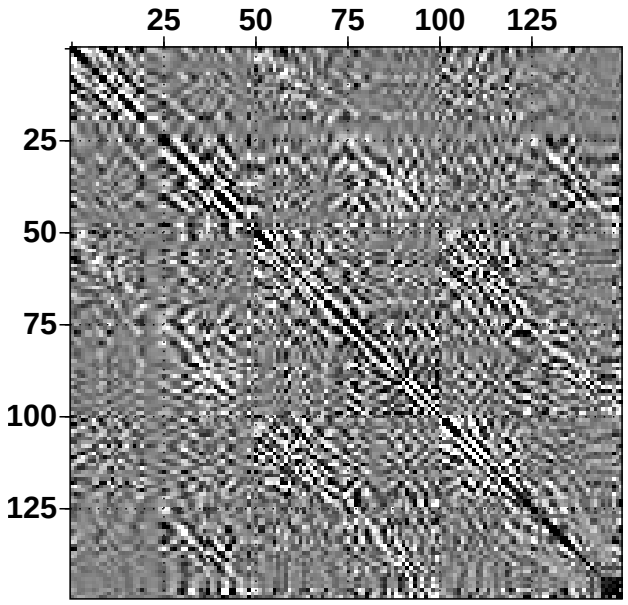

Figure 7. *A Posteriori* covariance matrix for the elastic waveform inversion Problem.

by the inversion procedure just considering the observed data. We do so via the Singular Value Decomposition (SVD) of this operator linearized about a given model. Although any model would do in this analysis, we chose the model $\mathbf{m_{map}}$ displayed in Figure (3).

Resolution Analysis in Bayesian Inference

The *a posteriori* covariance matrix is displayed in Figure (7). As explained in (Gouveia, 1996) the direct inversion of the quantity $G^H C_D^{-1} G + C_M^{-1}$ can be numerically unstable due the presence of small eigenvalues. What Figure (7) actually shows is the pseudo-inverse of $G^H C_D^{-1} G + C_M^{-1}$, in which the smallest eigenvalue allowed to enter in the calculation was 0.001. Once $\mathbf{m_{map}}$ and $C_{M'}$ are computed it is possible to sample the *a posteriori* distribution and generate pseudo-random realizations of the subsurface that are consistent with the observed data and the *a priori* information. This is done by taking the inner product of the lower triangular part of the LU decomposition of $C_{M'}$, with an uncorrelated Gaussian pseudo-random sequence. Typical realizations are displayed in Figure (8). As expected the P-wave impedance show the smaller variance. The density realizations show the largest variance, emphasizing that small information on density is available from seismic amplitudes. The S-wave impedance realizations are somewhere in between. Another way to display the pseudo-random models is shown in Figure (9). In this figure the realizations are shown side by side. Thus, larger lateral continuity implies a higher degree of confidence in the estimates of the subsurface parameter at a particular depth. Figure (9) illustrates the main purpose of the Bayesian inversion approach. With this procedure, we aim at assessing the non uniqueness of the inverse problem, due to noise in the observed data and to the null-space of the forward modeling operator. By generating realizations of the subsurface that are consistent with the available information, we are able to see a spectrum of plausible models. Those simulations bear some similarity with stochastic simulations done in geostatistics (Isaaks and Srivastava, 1989). However, to our knowledge, the forward modeling operator is not directly considered in geostatistical simulations.

Resolution Analysis in Occam's Inversion

The regularization parameter λ is directly related to resolution. The larger the regularization parameter associated with a given model, the larger its correlation length, and hence more limited its resolution. The covariance matrix $E[\delta\hat{\mathbf{m}}\delta\hat{\mathbf{m}}^T]$ is of course function of this parameter, as well as the regularization operator R and the modeling operator. The Occam's covariance matrix of the model perturbations is shown in Figure (10). This quantity provides a measure of resolution that is spatial and parameter-dependent. The wider diagonals of $E[\delta\hat{\mathbf{m}}\delta\hat{\mathbf{m}}^T]$, as compared to the model covariance obtained in the Bayesian case (Figure (7)), imply a larger correlation length (i.e., smoothness) for the model perturbations. As a final remark, note that the diagonal of this covariance matrix becomes progressive wider as we consider P-wave impedance (upper right-hand quadrant), S-wave impedance (central quadrant) and density (lower left-hand quadrant). This reflects the difference in resolution for each one of the parameters.

Bias of the Estimates

Let Ω be an estimator of some parameter θ. The estimator is unbiased if

$$E\left[\theta - \Omega\right] = 0. \tag{16}$$

Next we compute the bias for the Occam and Bayesian inverse problems. To keep the discussion as simple as possible we will assume in this section that the inverse problem is linear, that the forward operator is known exactly and that the discretization errors are negligible.

In statistics one of the main motivations for the computation of bias is to obtain an unbiased estimator from a biased one. However this will not be possible here. The reason is that the amount of bias depends on the true

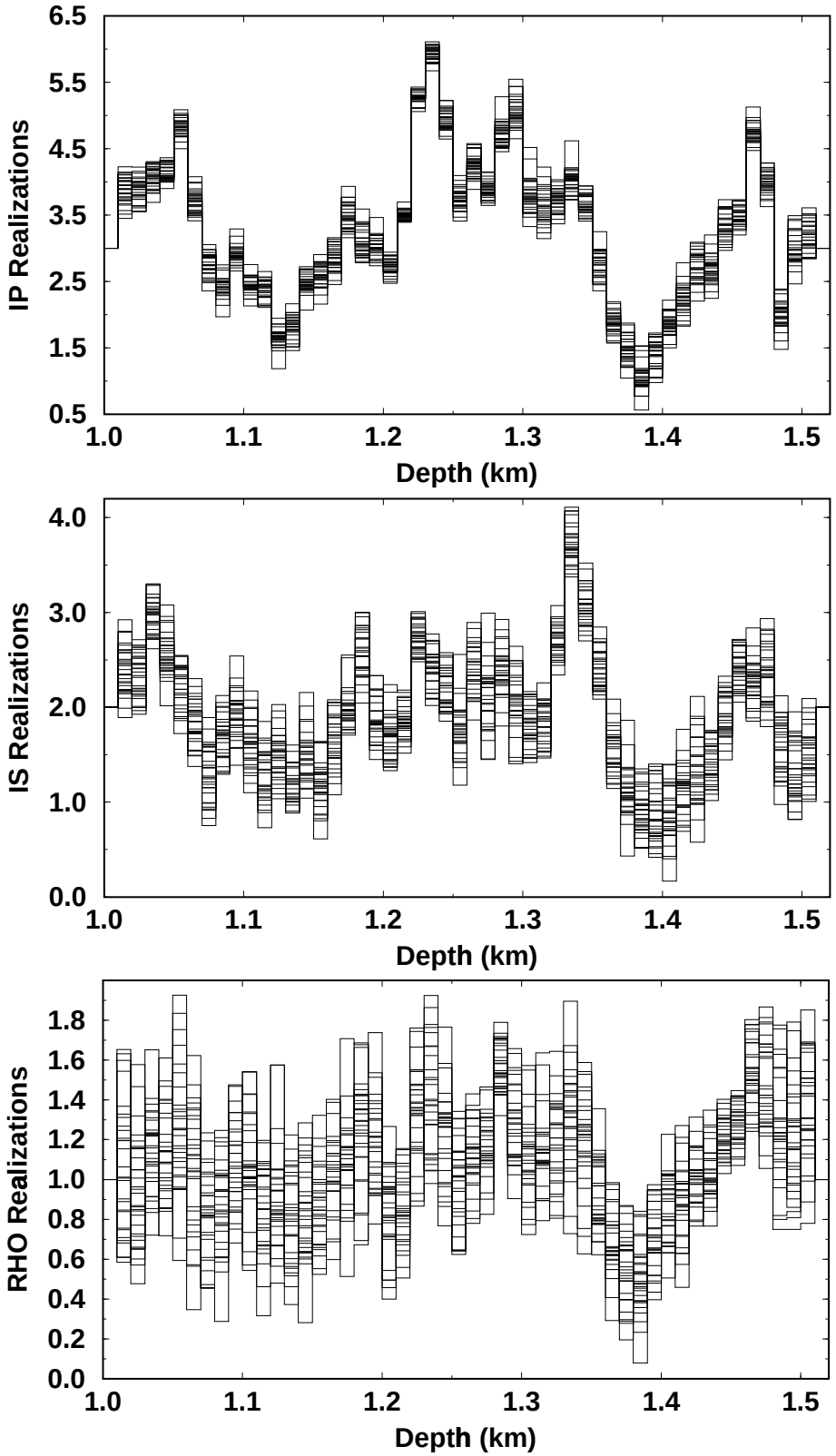

Figure 8. Pseudo-random realizations of the *a posteriori* probability distribution. Shown are *P*-wave impedance (top), *S*-wave impedance (middle) and density (bottom).

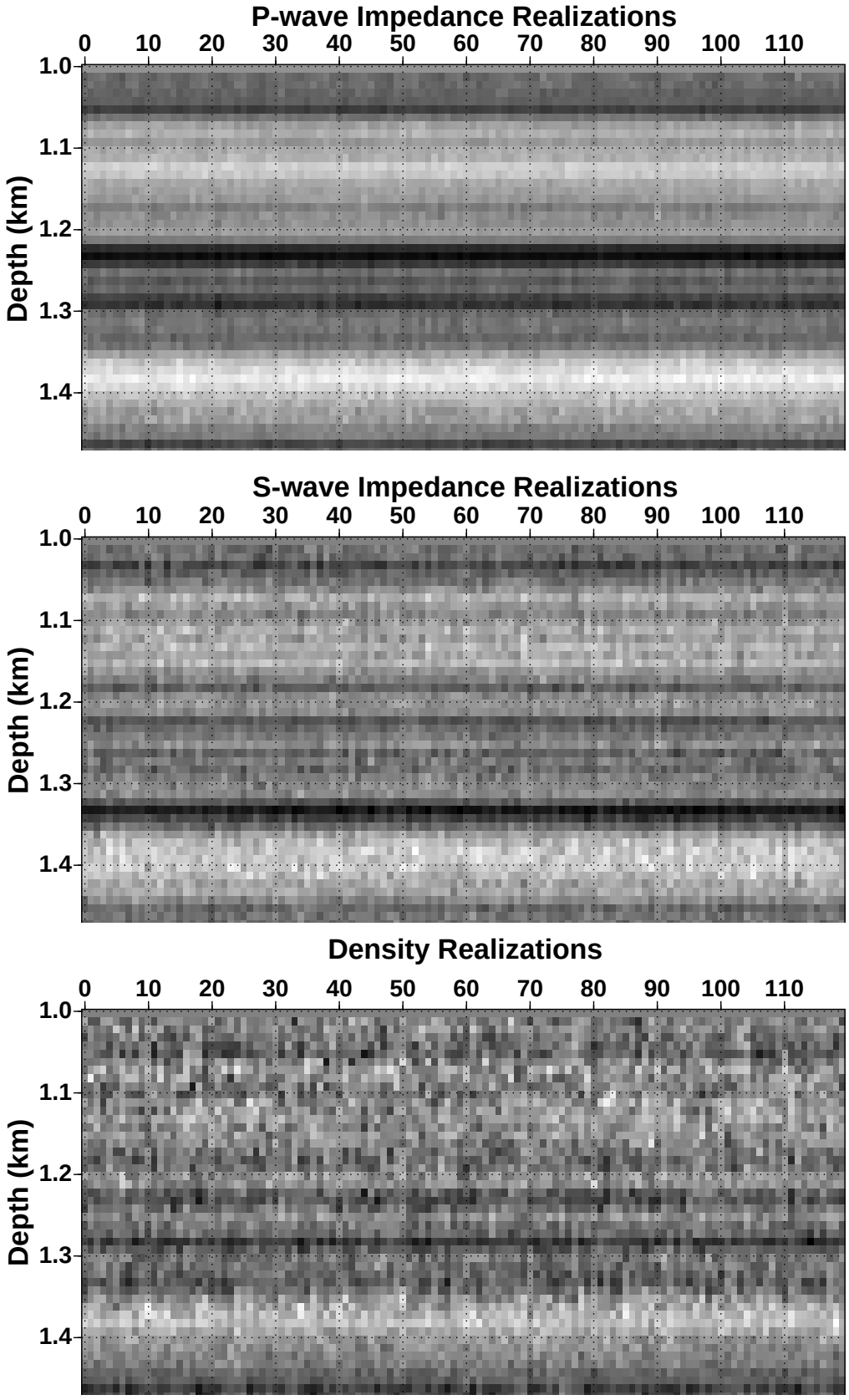

Figure 9. Random realizations of the *a posteriori* probability distribution. Lateral continuity implies on a high confidence of the estimates.

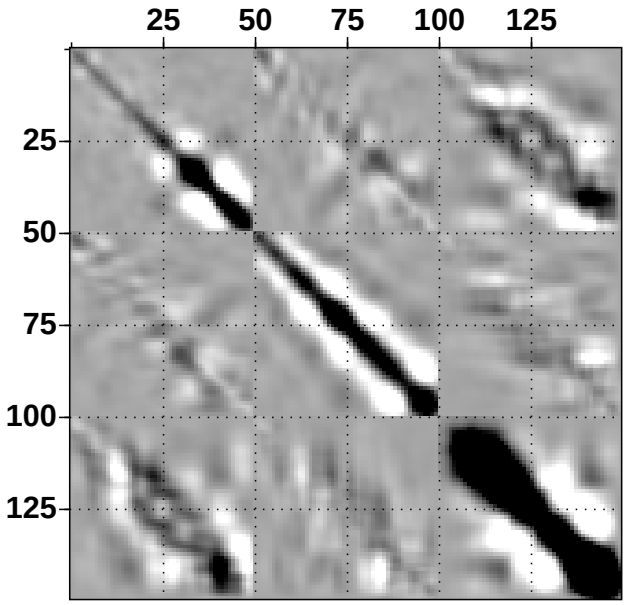

Figure 10. Covariance matrix obtained for the Occam's inversion.

subsurface. Nonetheless, the bias calculations that follow, provide further understanding on the inverse procedures studied in this work.

Biased of an Exactly Linear Inverse Problem

For a linear inverse problem with exact forward operator, the *true* model and observed data are related by:

$$G\,\mathbf{m}^{\mathbf{true}} = \mathbf{d} + \mathbf{n}, \tag{17}$$

where $\mathbf{n}$ is a noise term. The least-squares estimator $\hat{\mathbf{m}}$ of the true model is

$$\hat{\mathbf{m}} = G^{\dagger}\,\mathbf{d}, \tag{18}$$

where $G^{\dagger}$ is the pseudo-inverse of G. Therefore the bias is:

$$\begin{aligned} E\left[\mathbf{m}^{\mathbf{true}} - \hat{\mathbf{m}}\right] &= E\left[\mathbf{m}^{\mathbf{true}} - G^{\dagger}\,\mathbf{d}\right] \\ &= E\left[\mathbf{m}^{\mathbf{true}} - G^{\dagger}\left(G\,\mathbf{m}^{\mathbf{true}} - \mathbf{n}\right)\right] \\ &= \left(I - G^{\dagger}G\right)\mathbf{m}^{\mathbf{true}} + G^{\dagger}E\left[\mathbf{n}\right] \\ &= \left(I - G^{\dagger}G\right)\mathbf{m}^{\mathbf{true}}, \end{aligned} \tag{19}$$

assuming zero mean noise. Since the matrix $I - G^{\dagger}G$ is a projector onto the null space of G, in this case the bias of the generalized inverse model is the component of the true model in the null space of the forward problem. Thus the bias will zero if:

- The operator G is invertible and so $G^{\dagger} = G^{-1}$, or
- $\mathbf{m}^{\mathbf{true}}$ is orthogonal to the null space of G, so that the true model has no component in the null space of G.

Now, if the true earth model were constant, then the second condition would imply that the row sums of the resolution matrix $G^{\dagger}G$ must be zero. But, of course, in general this is not the case.

Bias of the Bayesian estimate

Now in the Bayesian calculation we do not really compute estimators of the true model; we compute probabilities associated with the true model. So while we may certainly compute the expected difference between the unknown true model and any particular model, it is not clear whether this can be interpreted as a bias. Nevertheless, it is an interesting exercise to compute this difference for the MAP model and for now we will think of this as the analogue of bias in the Bayesian calculation.

For a linear problem, the MAP model is the least squares solution of

$$\left(G^T C_D^{-1} G + C_M^{-1}\right)\mathbf{m} = G^T C_D^{-1}\mathbf{d}. \tag{20}$$

Thus

$$\mathbf{m}_{\mathbf{map}} = A^{\dagger} G^T C_D^{-1}\mathbf{d}, \tag{21}$$

where $A^{\dagger}$ is the pseudo-inverse of $G^T C_D^{-1} G + C_M^{-1}$. The expected difference between this model and the true earth model is:

$$\begin{aligned} &E\left[\mathbf{m}^{\mathbf{true}} - \mathbf{m}_{\mathbf{map}}\right] = \\ &E\left[\mathbf{m}^{\mathbf{true}} - A^{\dagger} G^T C_D^{-1}\mathbf{d}\right] = \\ &E\left[\mathbf{m}^{\mathbf{true}} - A^{\dagger} G^T C_D^{-1}\left(G\,\mathbf{m}^{\mathbf{true}} - \mathbf{n}\right)\right] = \\ &\left[I - \left(A^{\dagger} G^T C_D^{-1} G\right)\right]\mathbf{m}^{\mathbf{true}} + A^{\dagger} G^T C_D^{-1} E[\mathbf{n}] = \\ &\left[I - \left(A^{\dagger} G^T C_D^{-1} G\right)\right]\mathbf{m}^{\mathbf{true}}. \end{aligned} \tag{22}$$

We will now consider two cases. First assume that the matrix $G^T C_D^{-1} G + C_M^{-1}$ is invertible. That being the case, we have:

$$A^{\dagger} = \left(G^T C_D^{-1} G + C_M^{-1}\right)^{-1},$$

and thus,

$$G^T C_D^{-1} G = A^{\dagger\,-1} - C_M^{-1}. \tag{23}$$

Using this last expression in (22) it is easy to show that:

$$E\left[\mathbf{m}^{\mathbf{true}} - \hat{\mathbf{m}}\right] = C_{M'} C_M^{-1}\mathbf{m}^{\mathbf{true}}, \tag{24}$$

where $C_{M'}$ is the *a posteriori* model covariance matrix defined in Equation (9). Equation (24) has an interesting interpretation. Note that the bias is proportional to the product of the *a posteriori* and *a priori* model covariance matrices. In the situation that $C_{M'} = C_M$, i.e., no new information was provided by the observed data, the bias is equal to the true model $\mathbf{m}^{\mathbf{true}}$, regardless of the *a priori* model covariance C_M. As the inversion becomes

more informative, the elements of the matrix $C_{M'}$ get smaller and so does the bias.

In the situation where the matrix $G^T C_D^{-1} G + C_M^{-1}$ is not invertible, the interpretation is less clear. In this case $A^\dagger \neq \left(G^T C_D^{-1} G + C_M^{-1}\right)^{-1}$, and the expected difference between the true and the MAP model is given by Equation (22), which can be rewritten as:

$$E\left[\mathbf{m}^{\mathbf{true}} - \hat{\mathbf{m}}\right] = \mathbf{m}^{\mathbf{true}} - \left(A^\dagger G^T C_D^{-1} G\right) \mathbf{m}^{\mathbf{true}}. \quad (25)$$

Note that if $\mathbf{m}^{\mathbf{true}}$ lies in the null space of G, then $G\ \mathbf{m}^{\mathbf{true}} = 0$ and the bias is equal to $\mathbf{m}^{\mathbf{true}}$. If that is not the case, the bias will be a complicated function of the forward modeling operator and covariance matrices.

Bias of the Occam Estimator

For a linear Occam inversion, model and observed data are related by:

$$\left(G^T C_D^{-1} G + \lambda R^T R\right) \mathbf{m} = G^T C_D^{-1} \mathbf{d}. \quad (26)$$

In this case we seek an estimate of an *average* of the true model ($\bar{\mathbf{m}}^{\mathbf{true}}$), as opposed to an estimate of the *true* model itself. Hence, a more appropriate definition of bias is $E\left[\bar{\mathbf{m}}^{\mathbf{true}} - \mathbf{m}_{\mathbf{occam}}\right]$. A possible description (certainly not the only one) of the true subsurface average $\bar{\mathbf{m}}^{\mathbf{true}}$ is:

$$\bar{\mathbf{m}}^{\mathbf{true}} = \mathbf{m}^{\mathbf{true}} - \mathcal{R}\ \mathbf{m}^{\mathbf{true}} = \left(\mathcal{I} - \mathcal{R}\right) \mathbf{m}^{\mathbf{true}}, \quad (27)$$

where $\mathcal{R}$ is a roughness operator. This operator can be directly related to the regularization operator R used in the Occam's procedure, by $\mathcal{R} = \lambda R^T R$. Under this definition, one can show that the bias is given by:

$$E\left[\bar{\mathbf{m}}^{\mathbf{true}} - \mathbf{m}_{\mathbf{occam}}\right] = \left[\left(G^T C_D^{-1} G + \lambda R^T R\right)^{-1} - I\right] \cdot \left(I - \lambda R^T R\right)^{-1} \lambda R^T R\ \bar{\mathbf{m}}^{\mathbf{true}}, \quad (28)$$

assuming that the matrix $G^T C_D^{-1} G + \lambda R^T R$ is invertible. Equation (28) indicates that the Occam's estimate has little bias since $\bar{\mathbf{m}}^{\mathbf{true}}$ is an average, thus smooth by construction, and $R^T R$ is a differentiation operator.

Conclusions

If we use a parameterization of model space which is sufficiently fine to avoid discretization errors, then there will inevitably be features described by these models that are not well constrained by data. In Occam's inversion we compute the smoothest model that is consistent with the data and interpret the degree of smoothness as a direct measure of resolution. In Bayesian inversion we compute a probability distribution of models that is consistent with the data and use error bars derived from this distribution to assess the resolution provided by the inverse procedure. Bayesian inversion, however, allows one to incorporate probabilistic statements of data-independent information, if such information is available. In reflection seismology we believe such information is often available from *in situ* measurements of rock properties, geologic models, or other data. Since some of this information is clearly in the null space of the reflection seismic data, its incorporation into the calculation affords a clear illustration of the differences between the Occamist and Bayesian points of view.

We have shown a synthetic example of a reflection seismic inverse problem carried out from both a Bayesian and an Occamist point of view. On the one hand, we have computed a Gaussian *a priori* distribution of layered Earth models by estimating a covariance matrix from hypothetical well log measurements. These measurements are performed at a much finer resolution than the seismic wavelength. With this information we computed the maximum *a posteriori* model via a nonlinear optimization procedure. Making a Gaussian approximation to the posterior distribution about this MAP model allows us to compute the *a posteriori* covariance, from which we derive both 1-σ error bars and a tour of typical models pseudo-randomly sampled from this posterior. For the Occam calculation we optimized the nonlinear least squares data misfit function, regularized with a discrete second difference operator, for a sequence of increasingly large values of the smoothness. This gives the smoothest model that fits the data. And while the length scale of the smoothing is a measure of resolution, we derived a generalized covariance matrix from the augmented linear system. The bandwidth of this matrix, being essentially a depth-dependent correlation length, offers a more detailed measure of the resolution provided by the data alone.

The result is that the MAP model is similar, although slightly superior, to the Occam model. Moreover, models pseudo-randomly simulated from the Bayesian posterior distribution have features that are not required to fit the data since they are influenced by the data independent prior information. Whereas, by construction, the Occam inversion does not generate an Earth model *per se*, but rather an estimate of the coarsest average of the Earth that is consistent with the data. We also compute the bias of the solutions in the two approaches. Strictly speaking the meaning of bias is not clear in the Bayesian case, but it is easy to compute the expected difference between the true model and any particular model; we do this for the MAP model and show that the result depends only on the ratio of the posterior to prior model covari-

ance. For the Occam's estimate we compute the bias with respect to the average of the true subsurface, instead of the true subsurface itself.

Acknowledgments

This work was partially supported by the sponsors of the Consortium Project on Seismic Inverse Methods for Complex Structures at the Center for Wave Phenomena, Colorado School of Mines, the Army Research Office, the Shell Foundation, and the National Science Foundation under grant DMS-9506603.

References

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

Fuchs, K. and Müller, G. 1971. Computation of synthetic seismograms with the reflectivity method and comparison with observations *Geophys. J.R. Astron. Soc*, **11**, 417–433.

Gouveia, W. 1996. Reflectivity-based nonlinear amplitude inversion *CWP Project Review*, **203**.

Isaaks, E. and Srivastava, R. 1989. *An introduction to applied geostatistics* Oxford University Press.

Tarantola, A. 1987. *Inverse Problem Theory - Methods for Data Fitting and Model Parameter Estimation.* Elsevier.

APPENDIX A: The Linearized inversion

In this appendix we solve Problems 1 and 2 by linearizing the forward modeling operator about $\mathbf{m_{prior}}$.

A1 Problem 1

The linearized version of Problem 1 is

$$(G^T C_D^{-1} G + C_M^{-1})\,\delta\mathbf{m} = G^T C_D^{-1}(g(\mathbf{m_{prior}}) - \mathbf{d_{obs}}),$$

$$\mathbf{m_{map}} = \mathbf{m_{prior}} + \delta\mathbf{m}. \qquad \text{(A1)}$$

G is the forward modeling operator $g(\mathbf{m})$ linearized about $\mathbf{m_{prior}}$. This weighted, regularized linear system of equations was solved by a standard conjugate gradient procedure. The result, initial guess and true model are illustrated in Figure (A1). Comparison with the model obtained in the nonlinear approach (Figure (3)), shows that the linearized result is considerably poorer in accuracy. A direct comparison is obviously not fair due to the large differences in computational cost of both procedures. In this particular problem the nonlinear implementation costs approximately ten times more than the linearized version. The main objective here is to illustrate the higher accuracy in the results provided by the nonlinear approach to the problem of extracting impedance and density information from seismic amplitudes.

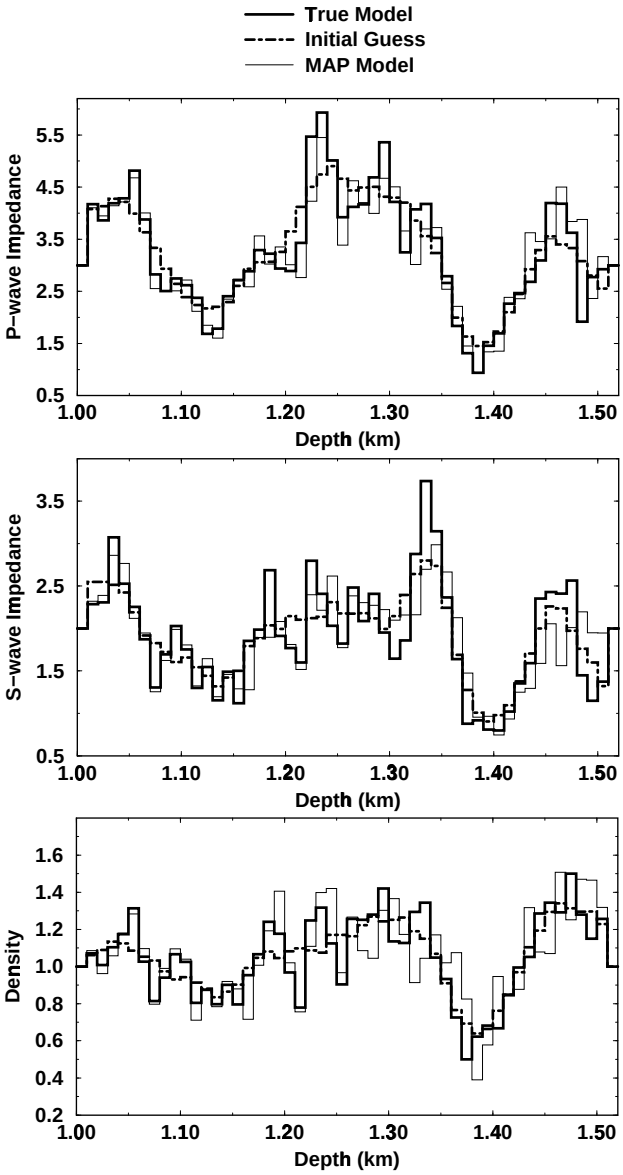

Figure A1. The MAP model estimated from the data in Figure 1 when linearization about the model $\mathbf{m}_{prior}$ is performed. The true model and the initial guess are also displayed.

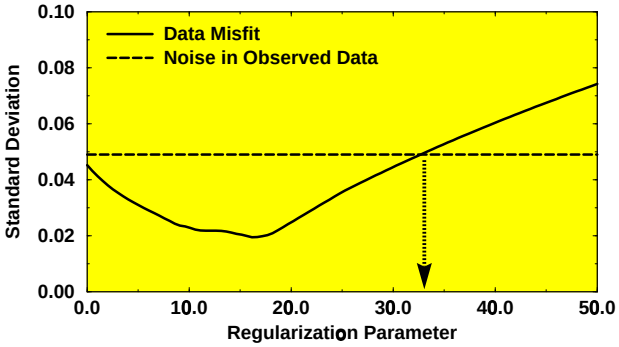

Figure A2. Regularization parameter λ plotted as a function of the standard deviation of the misfit of the linear system (A2) associated with the optimum model obtained from the regularized, linearized inversion. The dotted line corresponds to one standard deviation of the noise. Therefore the intersection of the solid and dotted curves corresponds to having fit the data to one standard deviation, on average. This graph was smoothed over a λ interval of length five.

A2 Problem 2

As expected the linear system obtained here has a similar form the the one obtained in the Bayesian case. It can be written as:

$$(G^T C_D^{-1} G + \lambda R^T R)\, \delta\mathbf{m} = G^T C_D^{-1} (g(\mathbf{m_{prior}}) - \mathbf{d_{obs}}).$$

$$\mathbf{m_{occam}} = \mathbf{m_{prior}} + \delta\mathbf{m}. \qquad \text{(A2)}$$

As in the nonlinear case, Figure (A2) shows the regularization parameter as a function of the data misfit. The model $\mathbf{m_{occam}}$ associated with the arrow in the figure is shown in Figure (A3). If this result is compared with the one obtained in the nonlinear version of the Occam's procedure (Figure (5)), we see that the linearization errors are not negligible. As in the Bayesian case, the model obtained with the nonlinear optimization (Figure (5)) is far superior than the one obtained in the linearized inversion. Figure (A4) shows $\mathbf{m_{map}}$ and $\mathbf{m_{occam}}$ in the same graph. A few comments can be made.

- $\mathbf{m_{occam}}$ is considerably smoother than $\mathbf{m_{map}}$. This model is obtained by solving a regularized least-squares linear system just once, as opposed to the nonlinear case, where a sequence of such systems are solved. This implies a more extensive averaging that will make $\mathbf{m_{map}}$ obtained in the nonlinear case smoother.
- Both solutions are relatively poor in accuracy, which can be verified by inspection of Figures (A1) and (A3).
- $\mathbf{m_{map}}$ preserves the general trend of the underlying model to a larger extent than the regularized inversion does. That can be attributed to the use of *a priori* information that is consistent to the true subsurface parameters.

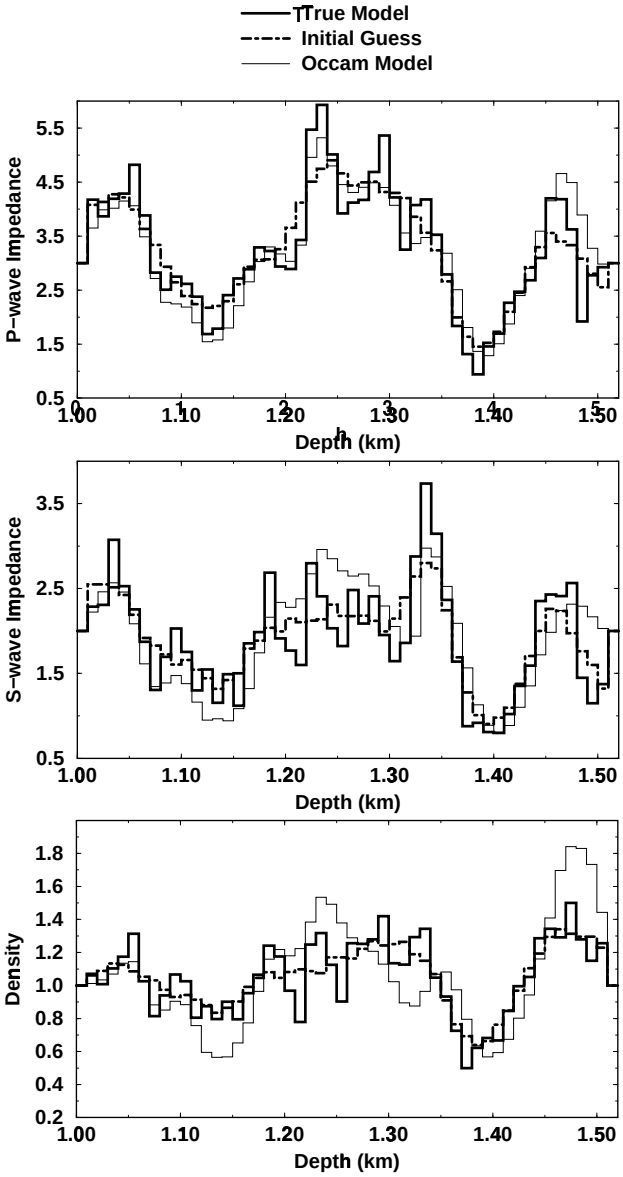

Figure A3. The Occam's model estimated from the data in Figure 1, using a linearized, regularized inversion procedure. The true model and the initial guess are also displayed.

APPENDIX B: Appendix B: Resolution of the forward modeling operator

In this appendix we present an SVD analysis of the forward modeling operator $g(\mathbf{m})$. Let $G = Q\Sigma V^T$ be the SVD decomposition of the linearized operator G. Considering that we have a linear system $G\mathbf{m} = \mathbf{d}$, the solution model $\mathbf{m}$ can be represented as:

$$\mathbf{m} = \sum_{i=1}^{N} \frac{\mathbf{q}_i.\mathbf{d}}{\sigma_i} \mathbf{v}_i, \qquad \text{(B1)}$$

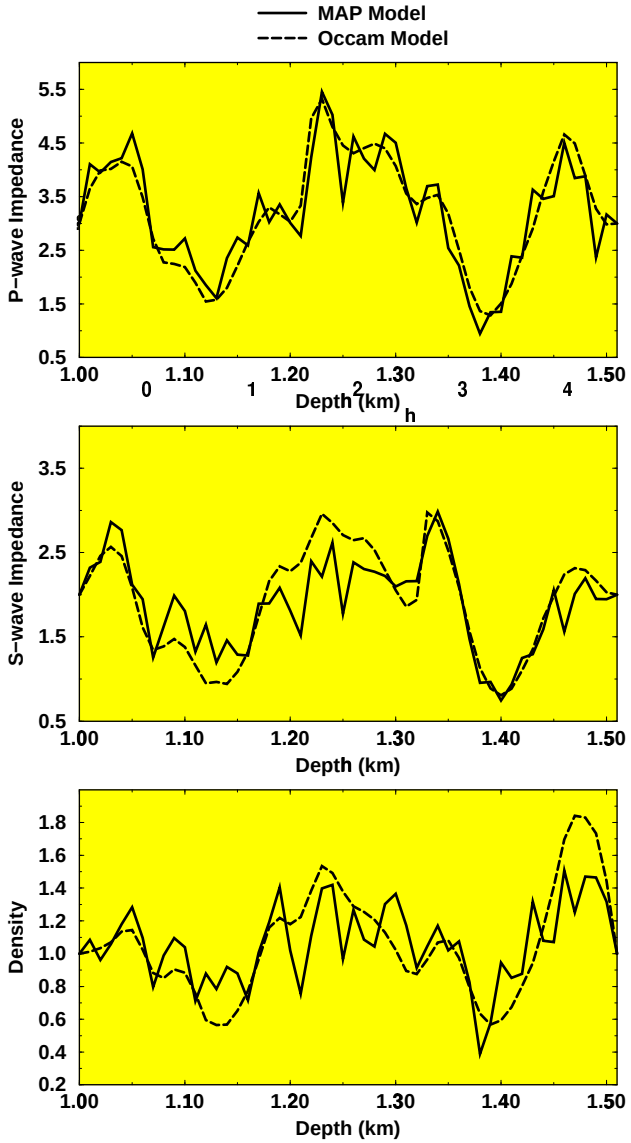

Figure A4. Comparison of the models obtained from Bayesian and regularized linearized inversions.

where $\mathbf{q}_i$ and $\mathbf{v}_i$ are the i-th columns of the matrices Q and V, respectively. σ_i is the i-th singular value of G. Model eigenvectors $\mathbf{v}_i$ associated with large values of σ_i are well resolved. Figure (B1) displays the model eigenvectors of G. In the upper part of the figure we show the normalized singular values. In the lower part we show the coefficients $\frac{\mathbf{q}_i \cdot \mathbf{d}}{\sigma_i}$ of the expansion (B1). In order to form the solution model $\mathbf{m}$ in Equation (B1) we add model eigenvectors with their respective weight until the superposition fits the data to one standard deviation. The cutoff is indicated by the arrow in Figure (B1).

Studying the spectrum of model eigenvectors we conclude that for both impedance and density profiles, long-wavelength (smoother) components can be better resolved than short-wavelength (rougher) components from seismic amplitudes. This analysis should, however, be coupled with the coefficients displayed in the bottom part of Figure (B1). Notice that the longest-wavelength components of both impedance and density make a very small contribution (due to the small value of the coefficient) to the expansion (B1). Therefore, this portion of the parameter spectrum is not well resolved by the data, though model eigenvectors are associated with large singular values. In essence, what can be learned from this SVD analysis is that long- to middle-wavelengths of P-wave impedance and middle- to small-wavelength of both S-wave impedance and density could be resolved by the forward modeling operator G.

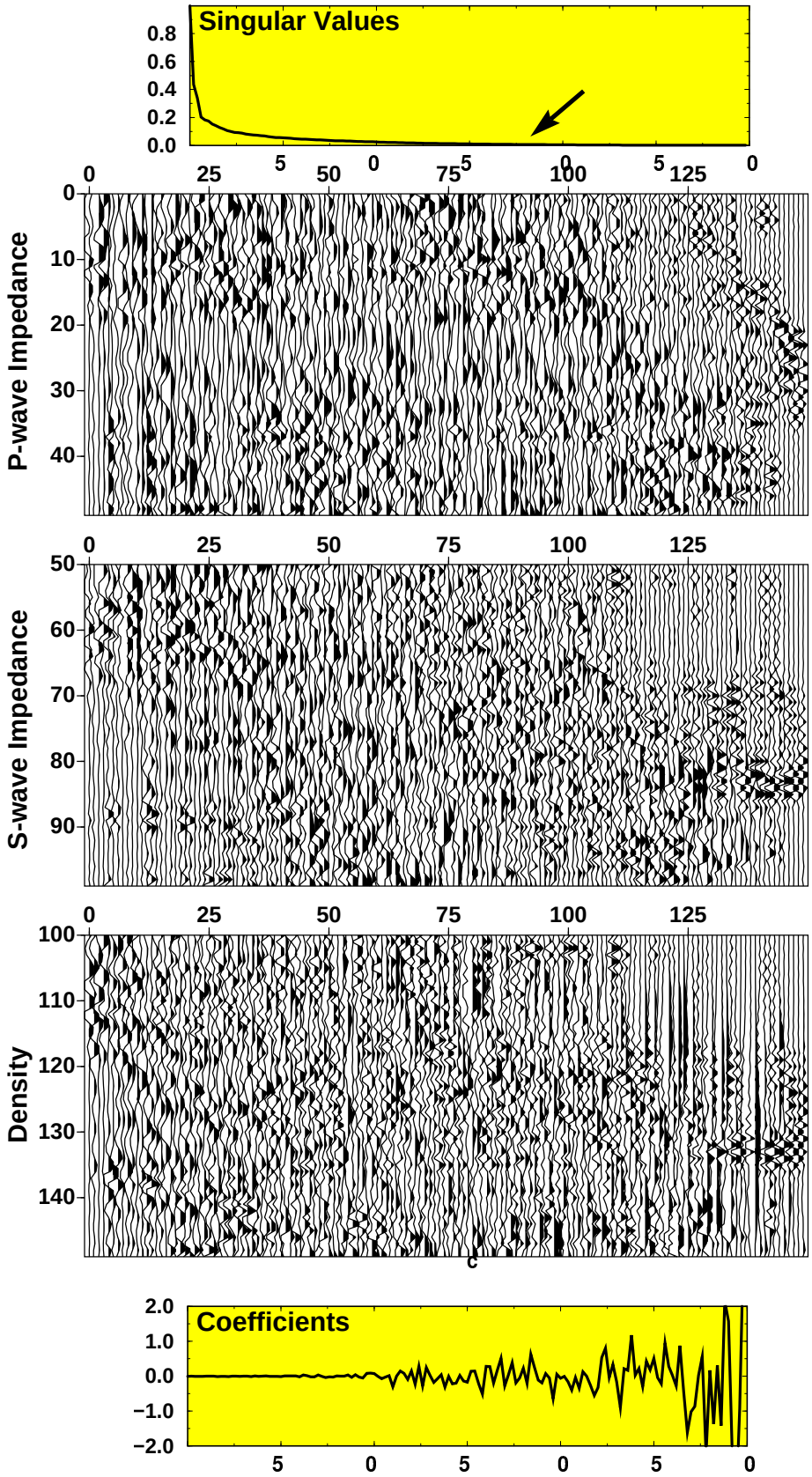

Figure B1. Singular Value Decomposition of the elastic waveform inversion. The arrow indicates the last model eigenvectors incorporated in the solution such that the data is fit up to one standard deviation of the noise. For each of the parameters the model eigenvectors are normalized to one.