

Delivery: an open-source model-based Bayesian seismic inversion program[☆]

James Gunning^{a,*}, Michael E. Glinsky^b

^a Australian Petroleum Cooperative Research Centre, CSIRO Division of Petroleum Resources, Glen Waverley, Victoria, 3150, Australia

^b BHP Billiton Petroleum, 1360 Post Oak Boulevard Suite 150, Houston, TX 77056, USA

Received 13 February 2003; received in revised form 17 October 2003; accepted 30 October 2003

Abstract

We introduce a new open-source toolkit for model-based Bayesian seismic inversion called *Delivery*. The prior model in *Delivery* is a trace-local layer stack, with rock physics information taken from log analysis and layer times initialised from picks. We allow for uncertainty in both the fluid type and saturation in reservoir layers: variation in seismic responses due to fluid effects are taken into account via Gassman's equation. Multiple stacks are supported, so the software implicitly performs a full AVO inversion using approximate Zoeppritz equations. The likelihood function is formed from a convolutional model with specified wavelet(s) and noise level(s). Uncertainties and irresolvabilities in the inverted models are captured by the generation of multiple stochastic models from the Bayesian posterior (using Markov Chain Monte Carlo methods), all of which acceptably match the seismic data, log data, and rough initial picks of the horizons. Post-inversion analysis of the inverted stochastic models then facilitates the answering of commercially useful questions, e.g. the probability of hydrocarbons, the expected reservoir volume and its uncertainty, and the distribution of net sand. **Delivery** is written in java, and thus platform independent, but the SU data backbone makes the inversion particularly suited to Unix/Linux environments and cluster systems.

© 2004 Elsevier Ltd. All rights reserved.

Keywords: Seismic; Inversion; AVO; Bayesian; Stochastic; Geostatistics; Markov Chain Monte Carlo; Open source

1. Introduction

The basic purpose of the seismic experiment in the oil exploration business has always been the extraction of information regarding the location, size and nature of hydrocarbon reserves in the subsurface. To say this is to also grant that the analysis of seismic data is necessarily and always an *inversion* problem: we do not measure reservoir locations and sizes; we measure reflected

waveforms at the surface, from which information about potential reservoir zones is extracted by a sequence of processing steps (stacking, migration, etc.) that are fundamentally inversion calculations designed to put reflection responses at their correct positions in time and space.

Such inversion calculations invariably depend on assumptions about the character of the geological heterogeneity that are usually described informally ("gentle dips", "weak impedance contrasts", etc), but could equally well be couched in formal probabilistic language. Further, such assumptions often hold well across a range of geological environments, and are of a nature that leads to generic processing formulae (e.g. Kirchoff migration) that may be applied by the practitioner with merely formal assent to the assumptions

[☆] Electronic appendixes available from server at <http://www.iamg.org/CGEditor/index.html>, and code from <http://www.petroleum.csiro.au> (follow links to *Delivery*).

*Corresponding author. Tel.: +61-3-9259-6896; fax: +61-3-9803-2052.

E-mail address: james.gunning@csiro.au (J. Gunning).

underlying their validity. From this point of view, we can say that traditional inversion is fundamentally based on a *model* of the subsurface, but no particular details of that model appear explicitly in the resulting formulae. These inversions can also be claimed to have integrated certain portions of our geological knowledge, in that the empowering assumptions have been made plausible by observation and experience of typical geological structures.

By this line of argument, traditional inversion can be seen as an attempt to explain the observed surface data by appeal to a broad (but definite) model of the subsurface, which yields congenial formulae when we incorporate certain geological or rock-physics facts (“weak reflections”, etc.). It is then no great step of faith to assert that inversion techniques ought to incorporate more definite forms of knowledge about the subsurface structure, such as explicit surfaces or bodies comprised of certain known rock types whose loading behaviour may be well characterised.

Such a step is in fact no longer contentious: it has long been agreed in the geophysical community that seismic inversion tools ought to use whatever rock-physics knowledge is regionally available in order to constrain the considerable non-uniqueness encountered in traditional inversion methodologies. Further, in an exploration or early-appraisal context, some limited log or core data are usually available, from which additional constraints on likely fluid types, time and depth horizons, etc. can be constructed.

In practice, such knowledge is difficult to incorporate into “model-free” methods like sparse-spike inversion, even though these methods have a sound geophysical basis. Conversely, detailed multi-property models of the reservoir—such as geostatistical earth models—are often weak in their connection to geophysical principles: the relationship to seismic data is often embodied in arbitrary regression or neural-network mappings that implicitly hope that the correct rock physics has been divined in the calibration process.

Another long-settled consensus is that the inversion process is incurably non-unique: the question of interest is no longer “what is such and such a quantity”, but “what is the multi-component distribution of the quantities of interest” (fluids, pay zone rock volume, etc.) and the implications of this distribution for development decisions. The development of this distribution by the integration of information from disparate sources and disciplines with the seismic data is now the central problem of seismic inversion. An interesting aspect of this consensus is that most workers in seismic inversion (and closely related problems like reservoir history matching) now concede that the most satisfactory way to approach the non-uniqueness and data-integration problems is via a Bayesian formalism (other approaches to stabilising the inverse problem via

regularisation or parameter elimination are perhaps best seen as varieties of Bayesian priors, perhaps augmented with parsimony-inducing devices like the Bayes information criterion). These approaches use explicit models for the quantities of interest: typically a suite of layers or facies, whose location and internal properties are the properties we wish to invert for. Examples of such work are Omre and Tjelmeland (1997), Eide et al. (2002), Buland and Omre (2003), Buland et al. (2003), Eidsvik et al. (2002), Leguijt (2001) and Gunning (2000) in the context of seismic problems, and the many papers of Oliver and his school in the context of reservoir dynamics problems, e.g. Chu et al. (1995), Oliver (1996, 1997). Newer kinds of Bayesian inversion are also appearing, where the model uncertainty itself (e.g. the number of layers) is taken into account, e.g. Malinverno (2002). It is recognised also that the posterior distribution of interest is usually quite complicated and impossible to extract analytically; its impact on decisions will have to be made from Monte Carlo-type studies based on samples drawn from the posterior.

Nevertheless, the use of model-based Bayesian techniques in seismic inversion at this point in time is still novel or unusual, for a mixture of reasons. The main reason is undoubtedly the lack of accessible software for performing such inversions, and the associated lack of fine control of such systems even when they are available as commercial products. Bayesian inversion methodologies are unlikely to ever become “black-box” routines which practitioners can apply blindly, and successful inversions will usually be the result of some iterative process involving some adjustment of the model-prior parameters and choice of algorithms. Such flexibility is hard to achieve in “black-box” algorithms. A second reason is that the amount of effort required to construct a suitable prior model of the rock physics and geological surfaces of interest will always be non-trivial, though repeated experience of such an exercise will reduce the required effort with time. This effort is justified by the fact that a Bayesian inversion constructed around an appropriate prior will always produce more reliable predictions than an inversion technique which does not integrate the regional rock physics or geological knowledge. This fact will be obvious if we do the thought experiment of asking what happens when the seismic data have poor signal to noise ratio. We assert that the use of Bayesian model-based inversion techniques should become far more widespread once the first mentioned obstacle above is overcome.

We do not presume in this paper to offer even a partial critique of non-model based inversion techniques from a Bayesian point of view: the reader will be able to do this for themselves after consideration of the methods and principles outlined later. The aim of this paper is rather to introduce a new open-source software tool

Delivery for Bayesian seismic inversion, and demonstrate how this code implements a model-based Bayesian approach to the inversion problem. The software described in this paper is a trace-based inversion routine, and is designed to draw stochastic samples from the posterior distribution of a set of reservoir parameters that are salient for both reservoir volumetrics and geophysical modelling. The exposition will cover the choice of the model parameters, construction of the prior model, the development of the forward seismic model and the associated likelihood functions, discuss the mapping of the posterior density, and sketch the methods used for sampling stochastic realisations from the posterior density. The latter involves use of sophisticated Markov Chain Monte Carlo (MCMC) techniques for multiple models that are relatively new in the petroleum industry.

The main content of the paper is laid out as follows; in Section 2 the overall framework and design of the inversion problem is outlined. Section 2.1 describes the basic model and suitable notation, Section 2.2 outlines the construction of the prior model, and Section 3 describes the forward model and associated likelihood. Section 4 covers the problems of mode mapping and sampling from the posterior. An outline of the software is provided in Section 5: it is released under a generic open-source licence rather like the popular GNU and open-BSD style licenses. A discussion of a suite of example/test cases is given in Section 6, and conclusions are offered in Section 7.

2. Outline of the model

The inversion routine described in this paper is a *trace-based* algorithm, designed to operate in a computing environment where a local seismic trace (usually post-stack, post-migration) in *Seismic Unix* (SU) format Cohen and Stockwell, 1998¹ is piped to the routine in conjunction with a set of parameters describing the local prior model, also in SU format. Stochastic realisations are then drawn from the posterior and written out, also in SU format. The details of these formats are discussed in Section 5. Inverting on a trace-by-trace basis amounts to an assumption that the traces are independent, and this can be guaranteed by decimating the seismic data to a transverse sampling scale equal to the longest transverse correlation appropriate for the local geology. Spacings of a few hundred metres may be appropriate. Working with independent traces has the great advantage that the inversion calculation is massively parallel,

and the computation may be farmed out by scatter-gather operations on cluster systems. Finer scale models may then be reconstructed by interpolation if desired.

Inversion calculations on systems with inter-trace correlations are difficult to sample from rigorously. Sketches of a suitable theory are contained in Eide (1997), Eide et al. (2002), Abrahamsen et al. (1997), Huage et al. (1998) and Gunning (2000), from which we offer the following summary. If the variables to be inverted are jointly multivariate Gaussian, some analytical work can be done which yields a sequential trace-based algorithm, but the matrix sizes required to account for correlations from adjacent traces are very large. Models which use non-Gaussian models for distributing facies (e.g. the indicator model used in Gunning (2000)) require methods that involve multiple inversions over the whole field in order to develop certain necessary marginal distributions. These calculations are very demanding, even for purely multi-Gaussian models.

From another point of view, the sheer difficulty of rigorously sampling from inter-trace correlated inversion problems is the price of modelling at a scale finer than the transverse correlation length of the sediments (and/or surfaces) of interest, which is commonly several hundred metres or more. We know from the Nyquist theorem that any random signal can be largely reconstructed by sampling at the Nyquist rate corresponding to this correlation length, and intermediate values can be recovered by smooth interpolation. This is a strong argument for performing inversion studies at a coarser scale than the fine scale (say 10–30 m) associated with the acquisition geometry. The choice of the transverse sampling rate depends also on the form of the transverse correlation dependence, since, e.g. smooth surfaces are better reconstructed by interpolation than noisy ones. This requires some judgement. Possible correlation lengths appropriate to various geological environments are discussed in Deutsch (2002).

By assuming that the inter-trace correlation is negligible, the inversion can proceed on an independent trace basis, and the incorporation of non-linear effects like fluid substitution, and discrete components of the parameter space (what type of fluid, presence or absence of a layer, etc.) become computationally feasible. In short, the correct sampling from systems with inter-trace correlations is probably only possible in systems with *fully* multi-Gaussian distributions of properties, but such a restriction is too great when we wish to study systems with strong non-linear effect like fluid substitution and discrete components like layer pinchouts or uncertain fluids. The latter, more interesting problems only become possible if we reduce the transverse sampling rate.

¹ Cohen, J.K., Stockwell Jr., J., 1998. CWP/SU: Seismic Unix Release 35: a free package for seismic research and processing. Center for Wave Phenomena, Colorado School of Mines, <http://timna.mines.edu/cwpcodes>.

The model used in this inversion is somewhat coarser than that commonly used in geocellular models. At each trace location, the time-region of interest is regarded as a stack of layers, typically a metre to several tens of metres in size. Each layer is generically a mixture of two “end-member” rock types: a permeable reservoir member, and an impermeable non-reservoir member. The balance of these two is determined by a layer net-to-gross (N_G), and the internal structure of mixed layers ($0 < N_G < 1$) is assumed to be finely horizontally laminated, so acoustic properties can be computed using an effective medium theory appropriate for this assumption. Seismic energy reflects at the boundaries between layers, producing a surface signal that may be synthetically computed, given an incident wavelet and all the requisite rock properties.

2.1. Description and notation of the local layer model

At the current trace location, the set of rock layers in the inversion region is fundamentally modelled in time rather than depth. The depth d enters as a relatively weakly controlling parameter of the rock properties, but time is a fundamental variable in terms of computing seismic responses, so is the better choice of variable for the basic parameterisation.

Models in depth are more fundamentally useful for reservoir decision making, but we adopt a flow wherein depth models can be created from any time model by a simple post-processing step. Depth constraints can take the form of thickness constraints or absolute depth constraints, and both of these can be applied either through the prior parameters or the isopach criterion we discuss later. Either way, the generation of depth models will require the specification of a *reference layer*, from which all layer depths will be hung or supported as required. This scheme ensures that depth and time models are mutually consistent.

The model consists of N_ℓ layers, with t_i the top of layer i . Layer i is bounded by the times (t_i, t_{i+1}) , $i = 1 \dots N_\ell$. An additional parameter, t_{base} , is required to

specified the bottom of the model. Fig. 1 shows a cartoon of the model.

Each layer is modelled as a mixture of two finely-laminated end-member rock types; a permeable member like sand or carbonate, and an impermeable member, such as shale or mudstone. The subscript f is used generically to denote these facies, but also s for the permeable member (think “sand”) and m for the impermeable (“mud”). The net-to-gross N_G specifies the ratio of permeable to impermeable rock by volume. Pure shales or other impermeable rocks can be modelled by $N_G = 0$. Hydrocarbons may be present only in the permeable members of the laminated mixture.

The properties of permeable rocks that we explicitly model are the p-wave velocity $v_{p,s}$, the shear velocity $v_{s,s}$, and porosity ϕ_s , but for impermeable members we use $v_{p,m}$, $v_{s,m}$, and density ρ_m . Facies are assumed isotropic. These rock properties are in general controlled by a loading curve which depends primarily on depth but also possibly a low-frequency interval velocity (LFIV) (derived perhaps from the migration), as explained in Section 2.2.3.

Permeable members that are susceptible of fluid substitution will also require knowledge of the dry matrix grain properties, and saturations, densities and p-wave velocities of the fluids undergoing substitution. For a particular rock type, the grain properties are taken as known, but the fluid saturations, densities and velocities can form part of the stochastic model.

The set of parameters describing the full acoustic properties for layer i , bounded by times t_{i-1}, t_i , with hydrocarbon h present, is then

$$\mathbf{m} = \{d, \text{LFIV}, N_G, \phi_s, v_{p,s}, v_{s,s}, \rho_m, v_{p,m}, v_{s,m}, \rho_b, v_{p,b}, \rho_h, v_{p,h}, S_h\}, \quad f = s, m, \tag{1}$$

at each trace. An i subscript is implicit for all quantities. If the layer is a pure impermeable rock ($N_G = 0$), this simplifies to

$$\mathbf{m} = \{d, \text{LFIV}, N_G, \rho_m, v_{p,m}, v_{s,m}\}. \tag{2}$$

Model parameters per layer:

$$\mathbf{m} = \{t_{top}, N_G, \phi_R, v_{p,R}, v_{s,R}, \rho_{NR}, v_{p,NR}, v_{s,NR}, \rho_b, v_{p,b}, \rho_h, v_{p,h}, S_h\}$$

R = reservoir rocks
 NR = impermeable rocks
 b = brine
 h = hydrocarbon

Full suite of model parameters:

$$\mathbf{M} = \{m_1, m_2, m_3, \dots\}$$

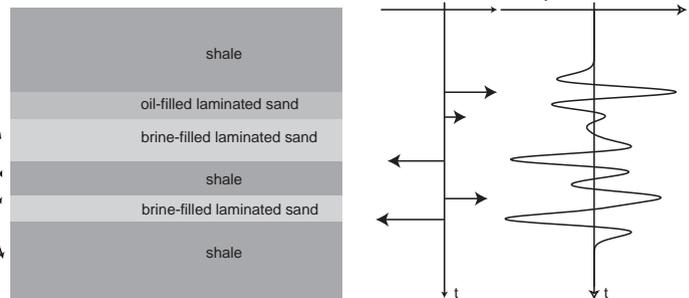


Fig. 1. Schematic of local layer-based model and its notation. For description of model parameters see main text. Reflection coefficient sequence and synthetic seismic are part of forward model of Section 3.

The union of these parameters with the layer times t_i (and the final bottom-layer base time t_{base}) then forms the full model vector of interest. Since the times are the dominant variables, it is convenient to arrange them to occur first in the full model vector, so this is assembled as

$$\mathbf{m} = \{t_1, t_2, \dots, t_{N_\ell}, t_{N_\ell, \text{base}}, \mathbf{m}_{\text{layer-1}}, \mathbf{m}_{\text{layer-2}}, \dots, \mathbf{m}_{\text{layer-}N_\ell}, A, B\}. \quad (3)$$

The extra stochastic factors A, B affect the synthetic seismic and are explained in Section 3.1.4.

An additional feature is that some parameters may be common to several layers (e.g. in many cases the LFIV is an average across many layers), so the underlying model vector will map all duplicate instances to the one variable (so, e.g. there may be only one LFIV in the overall model vector).

Also, it is sometimes useful to “synchronise” the rock-matrix properties between two layers, so that if they have the same end-members, acoustic contrasts at the interface must be entirely due to fluid effects. In this case, the underlying parametrisation will map all rock-matrix affecting parameters in the lower layer to those of the layer above, and remove all redundancies in the model vector.

2.2. Construction of the prior

2.2.1. Layer times

Before inversion, the times t_i are not known precisely, but are modelled as stochastic variables with prior distributions $N(\bar{t}_i, \sigma_{t_i}^2)$. The mean and standard deviation are estimated from horizon picking and approximate uncertainty estimates (e.g. a half-loop or so may be chosen for the σ_{t_i}). The prior distribution on layer times is supplemented by an ordering criterion which says that layers cannot swap:

$$t_i \geq t_{i-1}, \quad (4)$$

so the prior for the layer times is actually a product of truncated Gaussians with delta functions at the end-points absorbing the mis-ordered configurations. A typical scenario is that the top and bottom of a large packet of layers have been picked reasonably accurately, but the time location of the intermediate surfaces is not well known.

The truncation rules allow the useful possibility of pinchouts. For example, layer 2 may disappear if we have $t_3 = t_2$, in which case the interface at this time comprises a contrast between layer 1 and layer 3 properties, providing $t_3 > t_1$. Modelling pinchouts in this way then requires a specific algorithm for determining where the pinchouts occur from a given set of (untruncated) layer times t_i . We have used this

scheme:

- (1) Form a sort order based on the layer prior time distributions uncertainties σ_{t_i} (sorting in increasing order), and from the top layer down as a secondary sort order if some σ_{t_i} 's are identical.
- (2) From the set of candidate time samples $\{t_i\}$ (which will not in general satisfy (4)), proceed to fix the times in the sort order above, truncating values to preserve the ordering criterion (4) as we proceed. For example, if the sort order is $\{1, 4, 2, 3\}$, and we have a set $t_1 < t_4 < t_3 < t_2$, this will end up truncated at $t_1, (t_2 = t_4), (t_3 = t_4), t_4$.

This recipe is designed to allow the “better picked” horizons higher priority in setting the layer boundary sequence.

2.2.2. Prior beliefs about hydrocarbons

The modeller will have formed beliefs about the probability of certain kinds of hydrocarbons in each layer, informed by non-seismic sources such as pressure data or resistivity logs. A relative prior probability is assigned to each hydrocarbon type in each layer on this basis. Specifically, each layer i may bear fluids; oil (o), gas (g), brine (b), or low-saturation gas (l). The modeller must specify the prior probabilities of each of these phases, on a layer basis, as $F_{io}, F_{ig}, F_{ib}, F_{il}$, respectively, with $F_{io} + F_{ig} + F_{ib} + F_{il} = 1$.

Depending on the likely hydraulic communication between layers, the hydrocarbons allowed to be present in the permeable layers may be subjected to a density-ordering criterion, e.g. oil is not permitted above gas in two adjacent permeable layers. At least three types of density ordering rule can be envisaged:

- (1) *None*: any fluids are allowed in any permeable layer.
- (2) *Partial*: fluids are density ordered for all adjacent permeable layers not separated by an impermeable layer.
- (3) *Full*: fluids are density ordered across the entire reservoir model, regardless of whether there are impermeable layers separating permeable ones.

The set of possible fluids in each layer (as implied by the prior probabilities) are combined with a rule for density ordering to then enumerate a discrete set of possible fluid combinations $k = 1 \dots N_F$. For example, a two-layer system under ordering rule Eq. (3), where it is known that gas (and low-saturation gas) cannot occur ($F_{ig} = F_{il} = 0$), may have the allowable set $\{(\text{brine, brine}):(\text{oil, oil})\}$, so $N_F = 3$. Suppose the fluid-combination k corresponds to the set of fluid labels $f_{ik} \in \{b, l, g, o\}$, $i = 1 \dots N_\ell$. Then the prior probability of this fluid-combination is taken to be

$$p_k = \frac{\prod_i F_{i,f_{ik}}}{\sum_{k'=1}^{N_F} \prod_i F_{i,f_{ik'}}}. \quad (5)$$

Note that in multi-layer systems, this makes the marginal prior probability of obtaining a certain fluid in a given layer quite different to the prior probability specified on a layer basis, simply because of the ordering criterion. For example, in the two-layer problem described above, if the prior probability of oil in each layer had been specified as, say, 0.6, the three combinations would have probabilities proportional to 0.4^2 , 0.6×0.4 , 0.6^2 , respectively, or 0.21, 0.316, 0.474 after normalisation, so the *post-ordering* prior probability of oil in layer 2 is 0.474, and in layer 1 is $0.316 + 0.474 = 0.79$. In subsequent discussions, the *prior* probability of each distinct fluid combination k enumerated in this way is denoted p_k .

Supplementing this fluid categorisation, it is desirable also to model the fluid saturations S_{if} , $f = \text{o, g, b, l}$ for each fluid type. The two phases present are taken as brine (b) and hydrocarbon (o, g, l). The petrophysicist can assign Gaussian stochastic priors $S_{if} \sim N(\overline{S_{if}}, \sigma_{S_{if}})$, truncated outside $[0, 1]$ to these saturation parameters, based on regional knowledge. (The algorithms in this code treat truncated Gaussian distributions as a mixture of a true truncated Gaussian plus delta functions at the truncation points, of weight equal to the truncated probability. This has the advantage of giving non-vanishing probability to certain physically reasonable scenarios, like $N_G = 0$, or a layer pinchout).

2.2.3. Prior information about rock properties

Net-to-gross: The net-to-gross prior distribution is taken as $N(\overline{N_G}, \sigma_{N_G}^2)$, truncated within $[0, 1]$. The mean and standard deviation of this distribution can be determined by consultation with the geologist. A broad prior may be used to reflect uncertain knowledge.

Trend curves: From logging information in wells drilled through intervals deemed to be representative of the rock behaviour in the region to be inverted, a set of regression relationships for useful acoustic properties in all the facies is developed. The points used in these local “trend curves” are computed using a known reference fluid (typically, a brine) in place of the in situ fluid, and the basic acoustic properties (ρ, v_p, v_s) of these reference fluids have to be established (the reference fluid may vary with facies for various reasons). These “reference-fluid” properties may also be slightly uncertain, with normal distributions modelling their variation about a mean.

The trend curves are a set of rock-matrix velocities (p-wave, s-wave) v_{pf} , v_{sf} , and density ρ_f (or porosity ϕ_f) regressions for all end-member rocks. For reservoir-members, the porosity is used in preference to the

density, and the set of regressions is

$$\begin{aligned}\phi_f &= (A_\phi + B_\phi v_{pf}) \pm \sigma_{\phi f}, \\ v_{sf} &= (A_{v_s} + B_{v_s} v_{pf}) \pm \sigma_{sf}, \\ v_{pf} &= (A_{v_p} + B_{v_p} d + C_{v_p} \text{LFIV}) \pm \sigma_{pf},\end{aligned}\quad (6)$$

where $d = \text{depth}$. Density is computed from

$$\rho_{\text{sat}} = (1 - \phi)\rho_g + \phi\rho_{\text{fluid}},\quad (7)$$

for permeable members. The last equation models v_p as linear in $d = \text{depth}$ (compaction, etc.) and linear in the low-frequency interval velocity (LFIV); a local mean vertical p-wave velocity obtained from pure seismic data like VSP, moveout or stacking considerations. Such a regression can capture most stratigraphic, compaction and overpressure effects.

For impermeable rocks, the porosity is of no direct interest, and the density is regressed directly on v_{pf} using a Gardner–Gardner–Gregory (GGG) type relationship

$$\begin{aligned}\log \rho_f &= (\log A_\rho + B_\rho \log v_{pf}) \pm \sigma_{\rho f} \quad \text{or} \\ \rho_f &= A_\rho v_{pf}^{B_\rho} \pm \sigma_{\rho f}, \\ v_{sf} &= (A_{v_s} + B_{v_s} v_{pf}) \pm \sigma_{sf}, \\ v_{pf} &= (A_{v_p} + B_{v_p} \times \text{depth} + C_{v_p} \times \text{LFIV}) \pm \sigma_{pf}.\end{aligned}\quad (8)$$

Typically, $B_\rho \approx 0.25$. Since the range of densities and velocities in a single rock type is not large, this GGG-type regression can also be cast as a linear regression over a suitable range.

The regression errors $(\sigma_{pf}, \sigma_{sf}, \dots)$ used in these relationships are the prediction errors formed from linear regression studies, which yield t -distributions for the predictive distribution. We approximate this result by taking the prior to be of Normal form, with variance set to the regression variance. For example, the prior for v_{pf} is $N(A_{v_p} + B_{v_p} d + C_{v_p} \text{LFIV}, \sigma_{pf}^2)$. This approximation is exact in the limit of large data.

2.2.4. Fluid properties

We have also, from measurements or prior information, Gaussian prior distributions for the fluid p-wave velocities $N(v_p, \sigma_{v_p})$ and densities $N(\rho, \sigma_\rho)$ (from which the bulk moduli distribution can be computed) for the reference brines and any possible hydrocarbons.

3. The forward model

The Bayesian paradigm requires a likelihood function which specifies how probable the data are, *given* a particular model. This requires calculation of a synthetic seismic trace from the suite of layers and their properties, and forming the likelihood by comparing the seismic data and the synthetic seismic. The forward seismic model is a simple convolutional model, which treats layers as isotropic homogeneous entities with

effective properties computed from the successive application of Gassman fluid substitution in the permeable rocks and Backus averaging with the impermeable rocks.

There may be additional constraints in the form of isopach specifications, where a particular layer-thickness is known to within some error by direct observation, or other source of knowledge. The isopach constraint, being independent from the seismic, forms its own likelihood, and the product of the synthetic seismic likelihood and the isopach likelihood form the overall likelihood for the problem.

3.1. Computing the synthetic seismic

3.1.1. Rock mixtures

When two different isotropic rocks are mixed at a fine (sub-seismic resolution) scale with strong horizontal layering, it is well known that the effective acoustic properties of the mixture can be computed using the Backus average (Mavko p. 92). This assumes that the wavelengths are long compared to the fine-scale laminations represented by the net-to-gross measure. The standard formulae are

$$\frac{1}{M_{\text{eff}}} = \frac{N_G}{M_{\text{permeable}}} + \frac{1 - N_G}{M_{\text{impermeable}}}, \quad (9)$$

where M can stand for either the p-wave (M) or shear (μ) modulus ($\mu_{\text{eff}} = \rho_{\text{eff}} v_{\text{s,eff}}^2$, $M_{\text{eff}} = K_{\text{eff}} + \frac{4}{3}\mu_{\text{eff}} = \rho_{\text{eff}} v_{\text{p,eff}}^2$), and the effective density is

$$\rho_{\text{eff}} = N_G \rho_{\text{permeable}} + (1 - N_G) \rho_{\text{impermeable}}. \quad (10)$$

3.1.2. Fluid substitution in permeable rocks

When the saturation of a particular hydrocarbon is not trivially 0 or 1, we take the hydrocarbon and brine phases to be well mixed on the finest scale, so the fluid can be treated as an effective fluid (see Mavko et al., 1998, p. 205) whose properties are computed using the Reuss average

$$K_{\text{fluid}}^{-1} = \frac{S_{ix}}{K_{ix}} + \frac{1 - S_{ix}}{K_{ib}}. \quad (11)$$

When the effective fluid replaces brine in the pore space, the saturated rock has effective elastic parameters that are computed from the usual low-frequency Gassman rules (Mavko et al., 1998): viz

$$\frac{K_{\text{sat}}}{K_g - K_{\text{sat}}} = \frac{K_{\text{dry}}}{K_g - K_{\text{dry}}} + \frac{K_{\text{fluid}}}{\phi_s (K_g - K_{\text{fluid}})}. \quad (12)$$

The $M_{\text{permeable}}$ in the Backus relation (9) will be the p-wave modulus for permeable rock after fluid substitution via Gassman (only the permeable end-member will undergo fluid substitution). Here K_{dry} is the dry rock bulk modulus, K_g is the bulk modulus of rock mineral grains, K_{fluid} the bulk modulus of the substituted fluid (gas/oil/brine/low-saturation gas), and K_{sat} the effective

bulk modulus of the saturated rock. The shear modulus μ is unchanged by fluid substitution.

Under replacement of the reference fluid b by a fluid fl (a two-phase mix of brine and hydrocarbon h), the Gassman law, assuming no pressure effects on the dry modulus K_{dry} , can be written as

$$\begin{aligned} \frac{K_{\text{sat,fl}}}{K_g - K_{\text{sat,fl}}} - \frac{K_{\text{sat,b}}}{K_g - K_{\text{sat,b}}} \\ = \frac{K_{\text{fl}}}{\phi_s (K_g - K_{\text{fl}})} - \frac{K_{\text{b}}}{\phi_s (K_g - K_{\text{b}})}. \end{aligned} \quad (13)$$

3.1.3. Typical computation sequence

A typical computation sequence for computing the set of effective properties for a laminated, fluid-substituted rock layer would run as described in Appendix 1 on the IAMG server (<http://www.iapg.org/CGEditor/index.html>). Prior to computing the synthetic seismic, the effective properties of all the layers must be computed following this recipe.

3.1.4. The synthetic seismic and seismic-likelihood

Given a wavelet \mathbf{w} , an observed seismic \mathbf{S} , and an estimate of the seismic noise power σ_s^2 , we can use the reflectivities \mathbf{R} associated with effective-property contrasts between layers to construct the synthetic seismic appropriate to any particular stack. The synthetic is taken to be

$$\mathbf{S}_{\text{syn}} \equiv \mathbf{w} * \mathbf{R}, \quad (14)$$

where we use an FFT for the convolution, and \mathbf{w} and \mathbf{R} will be discretised at the same sampling rate Δt as the seismic data set \mathbf{S} for the trace. The set of delta functions in \mathbf{R} are projected onto the discretised time-grid using a 4-point Lagrange interpolation scheme (Abramowitz and Stegun, 1965) based on the nearest four samples to the time of a spike. This ensures that the synthetic seismic has smooth derivatives with respect to the layer times, a crucial property in the minimisation routines that are described in Section 4.1.

Available data traces \mathbf{S} may be near or far-offset (or both), and an appropriate wavelet \mathbf{w} will be provided for each. The P–P reflection coefficient for small layer contrasts and incident angles θ is (from the small-contrast Zoeppritz equation for $R_{\text{pp}}(\theta)$, expanded to $O(\theta^2)$ (Mavko p. 63))

$$\begin{aligned} R_{\text{pp}} = \frac{1}{2} \left(\frac{\Delta \rho}{\rho} + \frac{\Delta v_p}{v_p} \right) \\ + \theta^2 \left(\frac{\Delta v_p}{2v_p} - \frac{2v_s^2 (\Delta \rho / \rho + 2\Delta v_s / v_s)}{v_p^2} \right) \end{aligned} \quad (15)$$

with

$$\rho = (\rho_1 + \rho_2) / 2, \quad (16)$$

$$v_p = (v_{p,1} + v_{p,2}) / 2, \quad (17)$$

$$v_s = (v_{s,1} + v_{s,2})/2, \quad (18)$$

$$\Delta\rho = \rho_2 - \rho_1, \quad (19)$$

$$\Delta v_p = v_{p,2} - v_{p,1}, \quad (20)$$

$$\Delta v_s = v_{s,2} - v_{s,1} \quad (21)$$

and layer 2 below layer 1. All properties in these formulae are effective properties obtained as per electronic (electronic appendix 1). The coefficient of θ^2 is usually called the AVO gradient.

Due to anisotropy and other effects related to background AVO rotation (Castagna and Backus, 1993), some corrections to this expression may be required, so the reflection coefficient will take the form

$$R_{pp}(A, B) = \frac{A}{2} \left(\frac{\Delta\rho}{\rho} + \frac{\Delta v_p}{v_p} \right) + B\theta^2 \left(\frac{\Delta v_p}{2v_p} - \frac{2v_s^2(\Delta\rho/\rho + 2\Delta v_s/v_s)}{v_p^2} \right), \quad (22)$$

where the factors A and B may be stochastic, typically Gaussian with means close to unity and small variances: $A \sim N(\bar{A}, \sigma_A)$, $B \sim N(\bar{B}, \sigma_B)$.

Explicitly, θ^2 for a given stack is obtained from the Dix equation. The stack has average stack velocity V_{st} , event-time T_{st} and mean-square stack offset

$$\langle X_{st}^2 \rangle = \frac{(X_{st,max}^3 - X_{st,min}^3)}{3(X_{st,max} - X_{st,min})}, \quad (23)$$

from which the θ^2 at a given interface is computed as

$$\theta^2 = \frac{v_{p,1}^2}{V_{st}^4 T_{st}^2 / \langle X_{st}^2 \rangle}. \quad (24)$$

The likelihood function associated with the synthetic seismic mismatch is constructed as

$$L_{seis} = \exp \left(-f_{error} \sum_{\text{error-sampling points}} (\mathbf{S}_{syn} - \mathbf{S})^2 / 2\sigma_s^2 \right), \quad (25)$$

where the *range* of the error-sampling points is computed from the prior mean-layer time range and the wavelet characteristics as per Fig. 2. Within this range, the points used in the error sum are spaced at the maximum multiple of the seismic sampling time $p\Delta t$ which is smaller than the optimal error sampling time ΔT_s derived in electronic Appendix 2 (i.e. $p\Delta t \leq \Delta T_s$). The correction factor $f_{error} \equiv p\Delta t / \Delta T_s$, is designed to recover the error that would be obtained if the error trace were discretised at exactly ΔT_s . The signal-to-noise ratio S_N is implied by the noise level σ_s supplied by the user, with typically $S_N \approx 3$.

If both normal-incidence (small θ) and far-offset seismic data are available (“large” θ), we model the

overall likelihood as a product of likelihoods for normal-incidence and far-offset angles.

The stack information required by the likelihood model thus comprises the parameters V_{st} (stack velocity to event), $X_{st,max}$ (far offset), $X_{st,min}$ (near offset), and T_{st} (event time), a wavelet appropriate for the stack, and the noise level σ_s .

3.2. Isopach constraints

For certain layers there may be a constraint on the thickness which is obtained from well data, or kriged maps of layer thicknesses constrained to well data. Let $j = 1, \dots, N_l$ label the layers on which such a constraint is imposed. The thickness of the layer must match a known thickness ΔZ_j , within an error $\sigma_{\Delta Z_j}$ specified. This can be represented in a likelihood function

$$L_{iso} = \exp \left(- \sum_j \frac{(v_{j,p,eff}(t_j - t_{j-1}) - \Delta Z_j)^2}{2\sigma_{\Delta Z_j}^2} \right). \quad (26)$$

Similarly, constraints on the net-to-gross may be imposed by kriging net-to-gross estimates from well data. These will then be used to populate the prior-model traces with a spatially variable net-to-gross.

4. Sampling from the posterior

The posterior distribution for the inversion problem is a simultaneous model selection and model sampling problem. The selection problem is primarily that of choosing a model from the suite of models associated with the fluid combinations described in Section 2.2.2. For a particular combination k of fluids in the permeable layers, the model vector \mathbf{m}_k can be constructed from the union of all relevant parameters in Eq. (3). Hydrocarbon terms are omitted if the layer contains brine. Any fixed or irrelevant parameters are discarded. Different possible fluid combination will then yield model vectors \mathbf{m}_k of different dimension.

A further model-selection problem can also emerge from strong multimodality that may appear in the posterior for any particular fluid combination. If, for a fixed fluid combination, the posterior contains strongly isolated modes, sampling from this density is best approached by regarding the modes as separate models and incorporating them into the model-selection problem.

Model-selection problems can proceed in two ways. The first is to seek the marginal probability of the k th model $P(k|d)$ by integrating out all the continuous variables \mathbf{m}_k in the problem and then drawing samples via the decomposition $P(k, \mathbf{m}_k|d) = P(\mathbf{m}_k|k, d)P(k|d)$. The problem is to find reliable estimates of $P(k|d)$ when the necessary integrals cannot be done analytically.

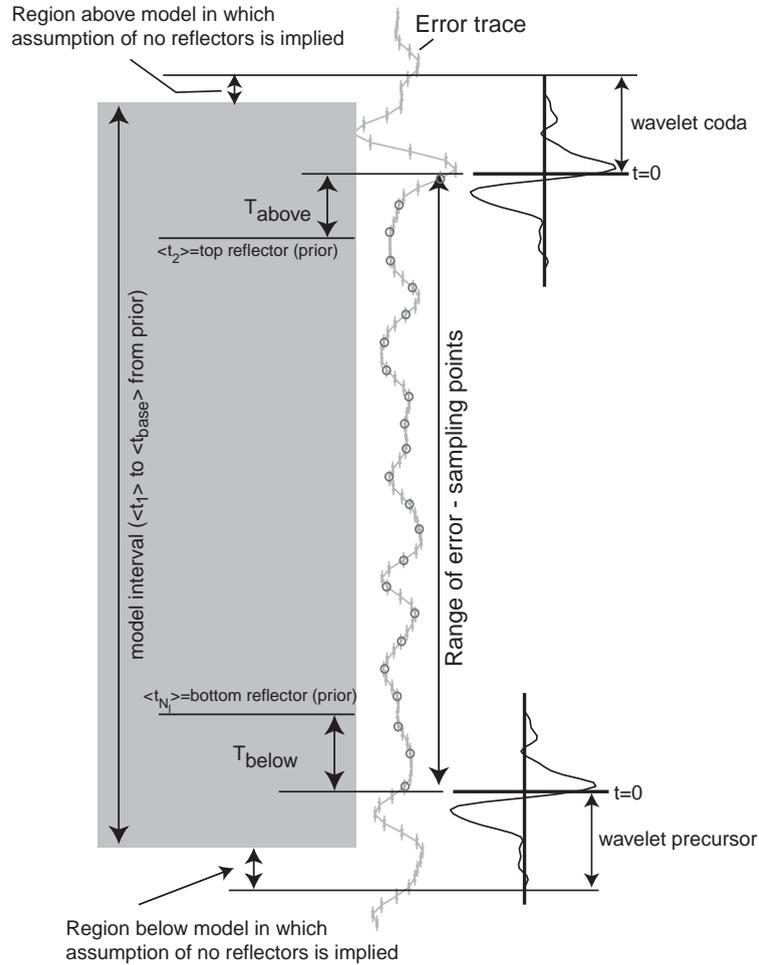


Fig. 2. Diagram indicating range over which error samples are computed. In general $T_{above,below}$ are free parameters, but we take T_{above} = wavelet-coda and T_{below} = wavelet-precursor as a sensible choice, since this allows full reflection energy from top and bottom reflections to enter error count for models not too different to mean prior model. Error is usually acquired by a sum over a subsampled range (circled samples), sub-sampling rate computed as per main text.

Methods to form such estimates using MCMC methods are described in Raftery (1996), and tend to focus on harmonic means of the likelihood computed with MCMC samples drawn from the conditional $P(\mathbf{m}_k|k, d)$. Preliminary experiments have shown some difficulty in stabilising the various estimators described by Raftery. The best estimator we have found is the full Laplace estimator, based on the posterior covariance matrix obtained by Newton updates at the mode. This forms the basis of an “independence sampler”, whose implementation details are described in electronic Appendix 4.

The second method samples for the full posterior $P(k, \mathbf{m}_k|d)$ by constructing a hybrid MCMC scheme that make jumps between the various models as well as jumps within the continuous parameter space of each model. Such methods are variously described as model

jumping methods, e.g. Andrieu et al. (2001), Phillips and Smith (1996); we implement the methods described by Andrieu.

For a fixed fluid combination k (of prior probability p_k), the posterior distribution of the model parameters \mathbf{m}_k should clearly be proportional to the product of the prior distributions with all the applicable likelihoods:

$$\Pi(\mathbf{m}_k|\mathbf{S}) \sim L_{seis}(\mathbf{m}'_k|\mathbf{S})L_{iso}(\mathbf{m}'_k)P(\mathbf{m}_k), \quad (27)$$

where the full prior is

$$P(\mathbf{m}_k) = \frac{p_k}{(2\pi)^{d_k/2} |C_k|^{1/2}} \exp\left(-\frac{1}{2}(\mathbf{m}_k - \bar{\mathbf{m}}_k)^T \times C_k^{-1}(\mathbf{m}_k - \bar{\mathbf{m}}_k)\right). \quad (28)$$

Note that the likelihood functions are evaluated for the model \mathbf{m}'_k obtained from \mathbf{m}_k after applying time ordering

(Section 2.2.1) and truncation of any values to their appropriate physical range of validity (e.g. $0 \leq N_G \leq 1$). In a great majority of cases in realistic models, $\mathbf{m}'_k = \mathbf{m}_k$. The prior mean vector $\bar{\mathbf{m}}_k$ and inverse covariance C_k^{-1} can be built directly from the description of the prior in Section 2.2.

Efficient sampling from the posterior will require location of all the feasible modes for a particular fluid combination, plus an estimate of the dispersion at each mode. The latter is usually taken as the local covariance associated with a Gaussian approximation to the local density.

4.1. Mode location and local covariance

Location of the modes is usually performed by searching for minima of the negative log-posterior starting from strategic points. For each fluid combination, the full model vector \mathbf{m} can easily have dimension of $d = O(50)$ or more, so this minimisation is quite demanding. The log-posterior surface is smooth for all parameters except where a layer pinches out, since the disappearing layer causes an abrupt discontinuity in the reflection coefficient. Such problems can be tackled with variable metric methods like BFGS minimisers specially adapted to detect solutions that lie along the lines of discontinuity (typically at some $t_{i+1} = t_i$). Still, very best-case d -dimension minimisation usually requires $O(d^2)$ function evaluations, with the coefficient being in the order of 10–100. Observe also that the function evaluation is a synthetic seismic (over N_t samples), plus prior, plus isopach, so the work count for this is $O(d^2 + N_t \log(N_t))$, which is typically $O(10^3)$ flops. The code actually uses an adaptation of Verrill's java UNCMIN routines (Koontz and Weiss, 1982), which implement both BFGS and Hessian-based Newton methods.

This accounting shows that the overall minimisation, performed over all d parameters will be quite slow. In practice, some of the parameters affect the posterior much more strongly than others. Layer times, p-velocities and densities have strong effects, whereas, e.g. hydrocarbon saturations have rather weak ones. This observation leads to a scheme wherein the minimisation is carried out over a subset of important parameters (say 2 or 3 per layer), and the remainder are estimated after the minimisation terminates using a sequence of Newton iterates as follows.

4.1.1. Newton updates

If the prior distribution for \mathbf{m}_k is multi-Gaussian, with negative log-likelihood $-\log(P(\mathbf{m}_k)) \sim (\mathbf{m}_k - \bar{\mathbf{m}}) C_m^{-1} (\mathbf{m}_k - \bar{\mathbf{m}})$, and the likelihood function is of form

$$\log(L(m|D)) \sim -\frac{1}{2} (f(\mathbf{m}_k) - y)^T \cdot \text{diag}(\sigma_i^{-2}) \cdot (f(\mathbf{m}_k) - y), \quad (29)$$

then the posterior mode $\tilde{\mathbf{m}}$ can be estimated by linearising $f(\mathbf{m})$ about an initial guess \mathbf{m}_0 using the standard inverse theory result (Tarantola, 1987)

$$X = \nabla f(\mathbf{m}_0), \quad (30)$$

$$\tilde{C}_m = (X^T C_D^{-1} X + C_m^{-1})^{-1}, \quad (31)$$

$$\tilde{\mathbf{m}} = \tilde{C}_m (X^T C_D^{-1} (y + X\mathbf{m}_0 - f(\mathbf{m}_0)) + C_m^{-1} \bar{\mathbf{m}}), \quad (32)$$

where $C_D = \text{diag}(\sigma_i^2)$. Here the $f(\mathbf{m})$ will be read from seismic (Eq. (25)) or isopach likelihoods (Eq. (26)).

The updates encountered in the code are all cases where the error covariance is diagonal, so the formulae are more simply stated in terms of the scaled residuals \mathbf{e} :

$$\mathbf{e}(\mathbf{m}_0) = C_D^{-1/2} (f(\mathbf{m}_0) - y), \quad (33)$$

$$\tilde{X} \equiv \nabla \mathbf{e}(\mathbf{m}_0), \quad (34)$$

$$\mathbf{d} = \tilde{X} \mathbf{m}_0 - \mathbf{e}(\mathbf{m}_0), \quad (35)$$

$$\tilde{C}_m = (\tilde{X}^T \tilde{X} + C_m^{-1})^{-1}, \quad (36)$$

$$\tilde{\mathbf{m}} = \tilde{C}_m (\tilde{X}^T \mathbf{d} + C_m^{-1} \bar{\mathbf{m}}). \quad (37)$$

The gradient \tilde{X} is evaluated by finite differences, and the scheme can be iterated by replacing $\mathbf{m}_0 \leftarrow \tilde{\mathbf{m}}$ at the end. It converges quadratically to the true mode if the path crosses no layer pinchouts. Typically, after minimisation using, say, BFGS methods over the dominant parameters in \mathbf{m} , only a few Newton iterates are required to get decent estimates of the remaining parameters, and the posterior covariance matrix \tilde{C}_m comes along for free.

4.2. Mode enumeration

Incorporation of the isopach likelihood alone will likely yield an approximately quadratic log-posterior function, which has a single mode. This mode may also be quite tight, especially at the wells. Conversely, the synthetic seismic likelihood is likely to be strongly multimodal, especially in the time parameters.

For this reason, we always perform a small set of Newton updates (say 5) to the prior based on the isopach likelihood before commencing a minimisation step for the entire posterior. The “partial-posterior” formed from the product of the prior and isopach constraint will be approximately multi-Gaussian, and the mode and covariance of this distribution are then used to choose starting points for a set of subsequent minimisation calls. The subset of parameters used in the modified BFGS minimisation routines are typically the layer times t_i , the impermeable layer p-velocities v_p and the net-to-gross N_G . Other combinations are conceivable and perhaps better. The initial starting values of the non-time parameters are taken from the “partial-posterior”

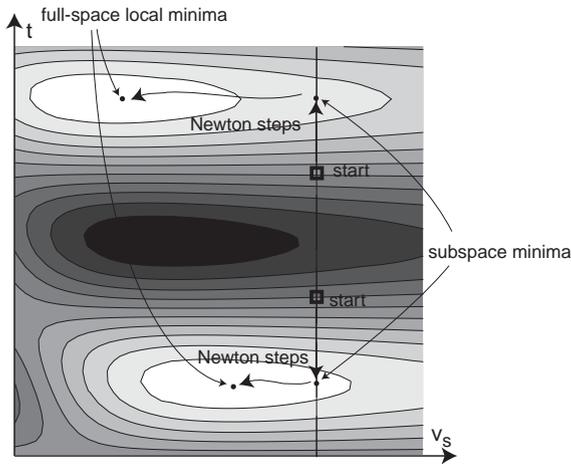


Fig. 3. Minimisation scheme consisting of subspace minimisations (subspace represented by t) starting from dispersed points, followed by global Newton steps, yielding a set of local minima in full parameter space.

mean, and a set of starting times are assembled as per the description in electronic Appendix 3. A run of subspace minimisations looping over this set of starting configurations will yield a set of points hopefully in the vicinity of global modes. Duplicates are eliminated. Fig. 3 indicates how the sequence of dispersed starting points, subspace minimisation, and global Newton methods converge to the desired separate modes.

The remaining candidate points then undergo a set of Newton updates (about 10) for the full model. Subsequently, any modes that look wildly implausible are eliminated (the remaining freedom in the frozen parameters is unlikely to redeem a bad loop-skip). If a mode is acceptable, the Newton-updated mean and covariance are stored, along with a measure of the mode weight such as the likelihood at the peak of the mode, and the Laplace-estimator weight (Eq. (59)) described in Appendix 4. Diagnostic routines capable of drawing graphs of the log posterior centred at the mode are useful in checking the character of the mode and the adequacy of the Gaussian approximation.

In cases where the seismic likelihood value obtained after these local optimisations looks poor (or for other reasons), a global optimisation step may be initiated operating on the subvector of layer times, with non-time properties fixed at their prior + isopach means. The initiation conditions for this step are designed to detect trapping of the BFGS/Newton optimisers in poor local solutions. The globally optimal t values are then used as starting points for the Newton optimiser. The global optimiser uses the Differential Evolution genetic algorithm of Storn and Price (1997).

4.3. Sampling

The overall Markov chain for the sampling is generated by a kernel that flips rapidly between two kinds of proposals: (a) jumps to different models, and (b) jumps within the current model. Currently the scheme chooses one of these moves with a 50% probability at each step, and performs a within-model jump or model-jump according to the schemes described below. This is the mixture hybrid-kernel of Section 2.3 in Brooks (1998). Special “moves” for unusual situations can also be devised, and are generally safe to use providing they are drawn independently of the current state of the chain.

4.3.1. Within-model jumps

For single-model d -dimensional problems that have smooth densities, a Metropolis–Hastings sampler that is both efficient and robust is the scaled reversible-jump random-walk sampler (RWM) described in Chapter 11 of Gelman et al. (1995). In this sampler, the jumps are drawn from the distribution $q(\mathbf{m}_{\text{new}}|\mathbf{m}_{\text{old}}) = N(\mathbf{m}_{\text{old}}, s^2C)$, where C is the estimated covariance at the mode, and s is a scale factor set to $s = 2.4/\sqrt{d}$. For multi-Gaussian distributions, this scale factor leads to optimal sampling efficiency for this class of samplers (acceptance rates near 0.23), and the sampling efficiency is then about $0.3/d$. The proposals \mathbf{m}_{new} drawn from $q(\mathbf{m}_{\text{new}}|\mathbf{m}_{\text{old}})$ are then time-ordered and truncated if necessary (as per Section 4) to produce \mathbf{m}'_{new} and then used to compute the full posterior density

$$\Pi(\mathbf{m}_{\text{new}}|\mathbf{S}) = L_{\text{seis}}(\mathbf{m}'_{\text{new}}|\mathbf{S})L_{\text{iso}}(\mathbf{m}'_{\text{new}})P(\mathbf{m}_{\text{new}}). \quad (38)$$

If the posterior were perfectly multi-Gaussian, samples could be drawn from an independence sampler (Brooks, 1998) using a Gaussian proposal density, which would be 100% efficient since successive samples are independent, but the assumption of a compact Gaussian distribution for the proposal density in a Metropolis technique will force undersampling of regions of the posterior that may have thick tails. Such regions appear reasonably likely when cross-sections of the log-posterior are plotted as diagnostic output: significant non-parabolicity is usually evident in the time-parameters (see Fig. 4).

Hence, for random walks within a fixed model, we use the RWM sampler where the covariance used is that produced by the sequence on Newton updates at the mode. The initial state is taken at the mode peak. Jumps that produce an unphysical state are treated as having zero likelihood. The acceptance probability for a jump $\mathbf{m}_{\text{old}} \rightarrow \mathbf{m}_{\text{new}}$ is the usual rule

$$\alpha = \min\left(1, \frac{\Pi(\mathbf{m}_{\text{new}}|\mathbf{S})}{\Pi(\mathbf{m}_{\text{old}}|\mathbf{S})}\right). \quad (39)$$

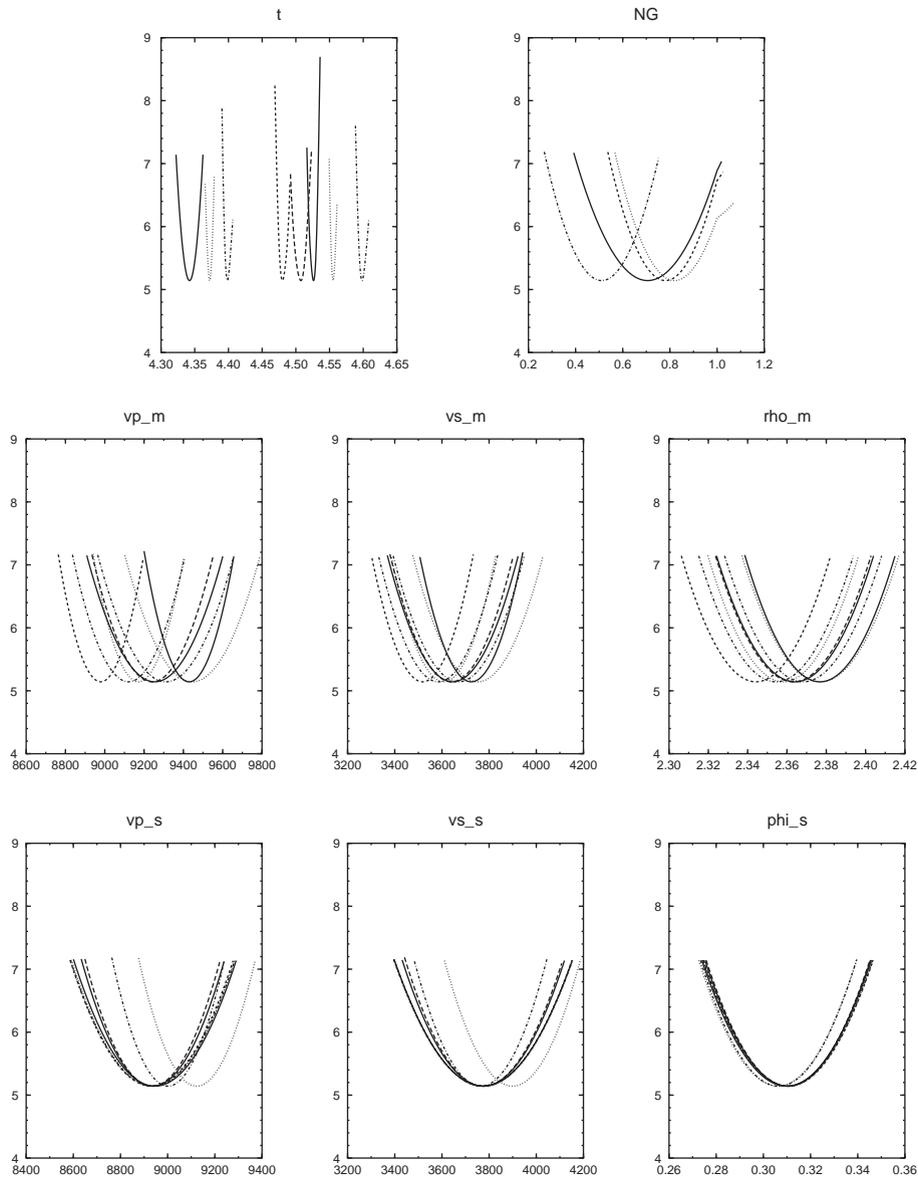


Fig. 4. Series of log-posterior cross-sections taken through maximum-likelihood point found from optimisation step and sequence of Newton updates. Cross-sections are shown for two parameters shown, in all relevant layers, over a range of two standard deviations as computed from local covariance matrix. Near-Gaussian parameters will show symmetric parabolas with a rise of +2 at endpoints, but non-Gaussianity is evidenced by skewed or a-symmetric profiles, most obvious in case of time parameters in this instance.

Note that the actual proposal probability q does not need to be evaluated, but the appropriate scaling for it is crucial.

4.3.2. Model jumping

Jumps between models have to be made with care so as to preserve reversibility. A notable feature of the set of models to be sampled from is that a great many parameters are common to all models (times, rock parameters, etc.), and only the fluid characteristics

distinguish the models corresponding to different fluid states. For example, if a jump involves introducing gas into a layer, the model vector is augmented by the gas velocity, saturation, and density (if they are not fixed). The models are thus good examples of the nested structures discussed in Andrieu et al. (2001), and the algorithms below are simple instances of the methods discussed in Section 3 of that paper.

Suppose the models \mathbf{m}_1 and \mathbf{m}_2 are partitioned into components as $\mathbf{m}_1 = \{\mathbf{m}_{12}^*, \mathbf{m}'_{12}\}$, $\mathbf{m}_2 = \{\mathbf{m}_{21}^*, \mathbf{m}'_{21}\}$, the

asterisk denoting common elements and the prime non-common parameters (so $\mathbf{m}_{12}^* = \mathbf{m}_{21}^*$ is the same set). We have estimates of the maximum likelihood model values $\bar{\mathbf{m}}_1 = \{\bar{\mathbf{m}}_{12}^*, \bar{\mathbf{m}}'_{12}\}$ and $\bar{\mathbf{m}}_2 = \{\bar{\mathbf{m}}_{21}^*, \bar{\mathbf{m}}'_{21}\}$ from the sequence of minimisation steps for each mode, and it is invariably the case that the shared parameters will have different values at each of these maximum-likelihood points. A jump from model 1 to 2 is then performed by the mapping

$$\mathbf{m}_{2,\text{new}} = \{\mathbf{m}_{12,\text{old}}^* + (\bar{\mathbf{m}}_{21}^* - \bar{\mathbf{m}}_{12}^*), \mathbf{m}'_{21}\}, \quad (40)$$

where \mathbf{m}'_{21} is drawn from a suitable proposal distribution $q_{12}(\mathbf{m}'_{21})$, which we usually take to be the prior for those components, since the fluid parameters are uncorrelated with the rest of the model in the prior:

$$q_{12}(\mathbf{m}'_{21}) = \prod_{\substack{\text{new fluid components} \\ \text{parameters } j \text{ in model}}} p_j(m'_{21,j}). \quad (41)$$

A reverse jump, from 2 to 1, would have invoked the proposal probability $q_{21}(\mathbf{m}'_{12})$ for the components of model 1 not present in model 2. The jump from model 1 to 2 is then accepted with probability

$$\alpha = \min\left(1, \frac{\pi(\mathbf{m}_{21})q_{21}(\mathbf{m}'_{12})}{\pi(\mathbf{m}_{12})q_{12}(\mathbf{m}'_{21})}\right), \quad (42)$$

where $\pi(\cdot)$ is the full posterior density of Eq. (27), including the dimension-varying terms in the prior.

In the thought-experiment limit that the likelihood functions are trivially unity (we “switch” them off) this jumping scheme will result in certain acceptance, as the $q_{ij}(\cdot)$ densities will cancel exactly with corresponding terms in the prior. The likelihoods are expected to be moderately weak functions of the fluid properties, so this seems a reasonable choice. Further, in many practical scenarios, the fluid constants may be sufficiently well known that they are fixed for the inversion, in which case the models share all the parameters, and the q terms disappear from Eq. (42).

The linear translation embodied in Eq. (40) is based on a plausible assumption that the “shape” of the posterior density for the common parameters remains roughly the same across all the models, but the position may vary. Corrections to this assumption can be attempted by rescaling the mapping using scale-factors estimated from covariance estimates—which introduces a constant Jacobian term into Eq. (42)—but numerical experiments have not shown that this embellishment significantly improves the acceptance rates.

Mixing has generally been found to be very good when jumps are performed between various fluid states using this scheme. When multimodality in the posterior due to loop-skipping occurs, mixing can be more problematic, and the posterior samples should be checked carefully to ensure mixing is adequate. Various

options for detecting and auto-correcting for this problem, using decimation estimates computed from the time-series in the Markov chain (Gelman et al., 1995, Chapter 7), have been added to the code.

5. The software

The Delivery inversion software is written in java, a design decision rather unusual in the context of numerically intensive scientific software. It should run on java implementations from 1.2 on. The advent of hotspot and just-in-time (JIT) compiler technology has made java a much more efficient language than in its early days as a purely interpreted language (see the Java Numerics website as an entry point Boisvert and Pozo, 2003.² The core cpu demands in the code are (a) linear algebra calls, for which we use the efficient CERN colt library Hoschek, 2003³ and (b) FFTs, for which we provide a library operating on primitive double[] or colt DenseDoubleMatrix1D arrays, and the former have been demonstrated to be as fast as C on the platforms we use.

On platforms with modern hotspot or JIT compilers, the code is estimated to run somewhere within a factor of 2–5 of the speed of a C++ implementation, and has the advantage of being platform independent, free of memory-management bugs, and has been simpler to develop using the large suite of libraries now standard in java 1.2 and its successors.

In practice, the inversion code will likely be run in an operating system that allows piping of input and output data in the form of SU streams, and the attractive possibility of clustering the calculation means that it will likely be run on some flavour of unix or Linux.

Some explanation of the files required for input/output and usage is given in Appendix 7. The inversion is chiefly driven by an XML file that specifies the necessary rock-physics, layer descriptions, and information about the seismic. A quality schema-driven GUI editor expressly developed for the construction of this XML file is also available at the website: further details are in the appendix.

The code is available for download Gunning, 2003⁴ under a generic open-source agreement. Improvements to the code are welcome to be submitted to the author. A set of simple examples is available in the distribution.

²Boisvert, R., Pozo, R., 2003. The Java Numerics website, <http://math.nist.gov/javanumerics/>.

³Hoschek, W., 2003. The CERN colt java library, <http://tilde-hoschek.home.cern.ch/~hoschek/colt/index.htm>.

⁴Gunning, J., 2003. Delivery website: follow links from <http://www.petroleum.csiro.au>.

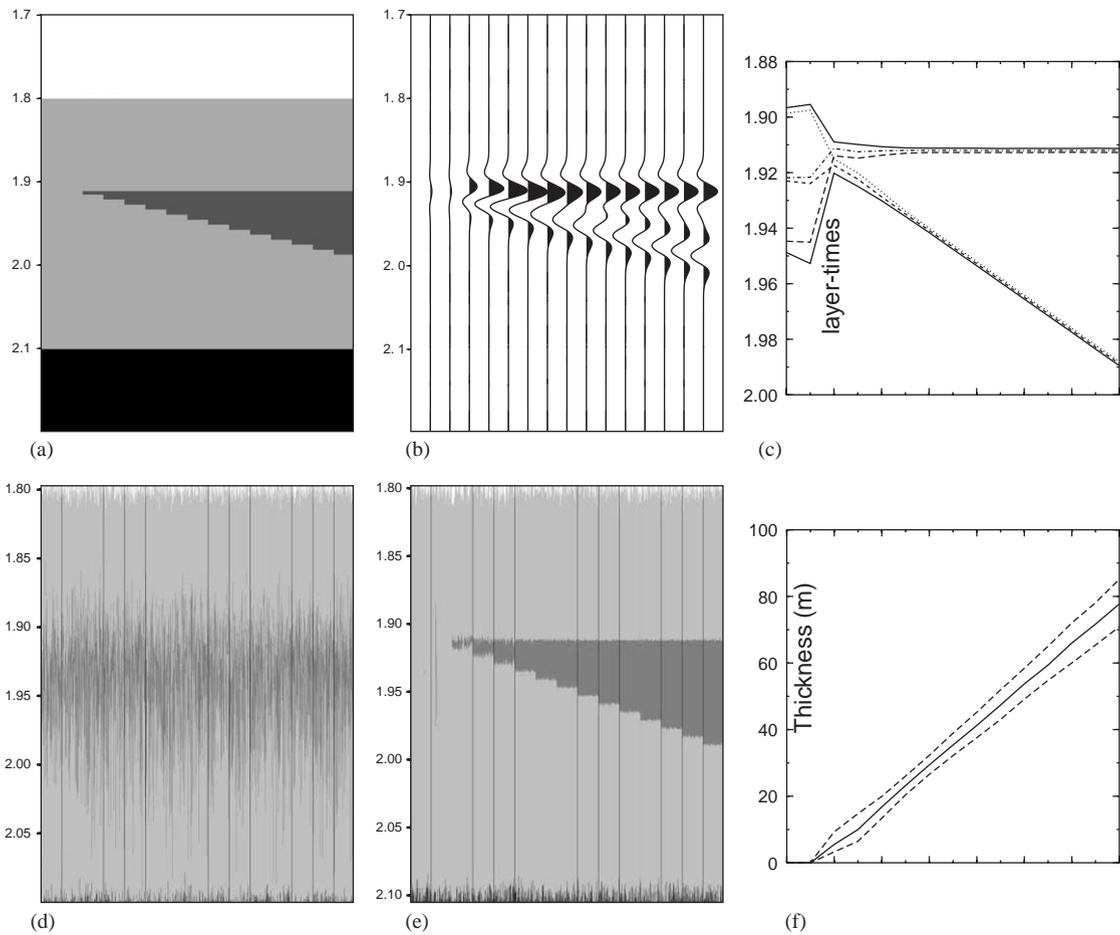


Fig. 5. (a) Truth case wedge model, (b) truth case seismic, (d) pre-inversion stochastic samples (i.e. drawn from prior distribution only), (e) post-inversion samples (100 per trace), (c) p_{10}, p_{50}, p_{90} quantiles for layer times of wedge, and (f) p_{10}, p_{50}, p_{90} quantiles of wedge thickness. Note how wedge layer time uncertainty increases to prior when wedge pinches out, as there is no longer a significant reflection event to facilitate detection of this horizon: inverter does not care where horizon is, as long as it pinches out. Wedge thickness (zero) is still well constrained though.

6. Examples

6.1. Sand wedge

A simple but useful test problem is one where a wedge of sand is pinched out by surrounding shale layers. This is modelled as a three layer problem, where we expect the inversion to force pinching-out in the region-where no sand exists, if the signal-to-noise ratio is favourable and the sand-shale acoustic contrast is adequate.

In this example, the prior time parameters are independent of the trace, so there is no information in the prior to help detect the wedge shape. The internal layer times have considerable prior uncertainty (20, 50 ms, respectively). Fig. 5 illustrates pre- and post-inversion stochastic samples of the layers, displayed with

many realisations per trace. Here the noise strength is about $\frac{1}{4}$ of the peak reflection response when the wedge is thick (i.e. away from tuning effects).

This test problem shows how the inverter can readily unscramble tuning effects from rock-physics to produce an unbiased inversion of a layer pinchout.

6.2. Net-to-gross wedge

Another useful test problem is where a slab of shaly sand is gradually made cleaner from left to right across a seismic section, and embedded in the mixing shale. As the net-to-gross increases, the reflection strength improves, and the posterior distribution of the net-to-gross and layer thickness is of interest.

In this example, the only parameters that varies areally is the mean net-to-gross: this is fixed to be the

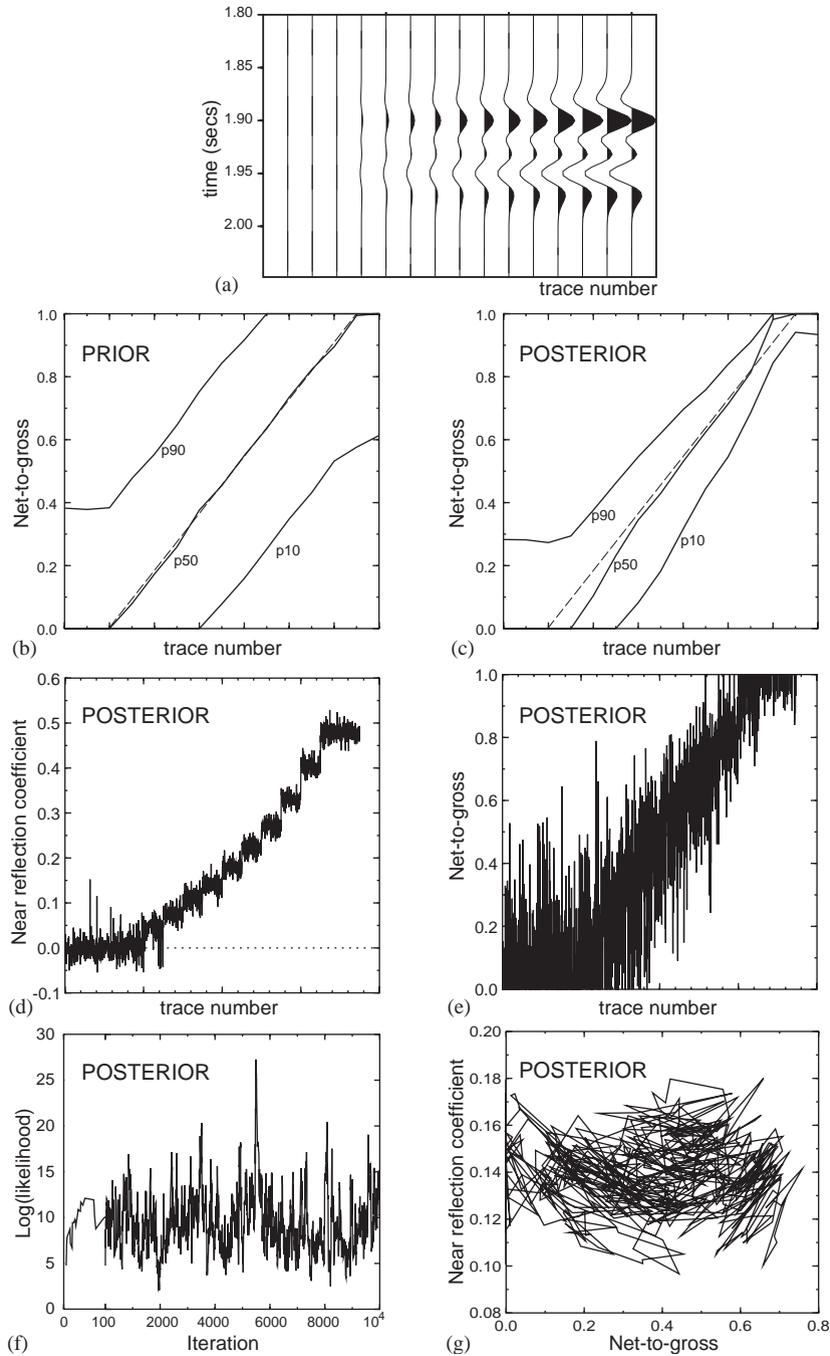


Fig. 6. (a) Truth case seismic for net-to-gross (N_G) wedge model, (b) pre-inversion (prior) spread of N_G as show by p10, p50, p90 quantiles; truth- case shown dashed, (c) as per (b), but post-inversion. Inset (d) shows near-stack reflection coefficient variations, and (e) associated samples of N_G . Signal-to-noise ratio here is very favourable, with noise 1/10th of peak seismic response. Residual uncertainty in N_G is due to fact that there is also uncertainty in layer velocity variations. Seismic constrains primarily reflection coefficient, but system indeterminacy will still allow appreciable variation in N_G . Convergence and mixing of MCMC scheme is illustrated for a highly non-informative prior $N_G \sim N(0.45, 0.7^2)$ on trace 8: (f) trace plot of log-likelihood, with early iterations magnified, (g) trajectory of Markov chain walk for N_G of layer 2 vs. its reflection coefficient at top.

same as the “truth-case” model, but has a (very broad) prior standard deviation of $\sigma_{NG} = 0.3$. Fig. 6 illustrates pre- and post-inversion estimates of the net-to-gross distribution parameters, superposed on the truth case. This example is also used to demonstrate the convergence of the MCMC sampler, for which we show some characteristic plots of the random walk for the inversion on trace 8. Here the prior was loosened considerably to $N_G \sim N(0.45, 0.7^2)$. The chains are always started at the maximum likelihood point, so ‘burn-in’ times are generally very short.

This test problem shows that the inverter produces a relatively unbiased inversion for the net-to-gross, but subject to increasing uncertainty as the reflecting layer dims out.

6.3. Simple single-trace fluid detection problem

In this simple test a slab of soft sand surrounded by shale layers has oil present, and the prior model is set to be uninformative with respect to fluid type (so the reservoir has 50% prior chance of brine, 50% oil). The rest of the story is told in the caption of Fig. 7. In summary, the inverter will detect fluid contrasts reliably if the signal quality is sufficient, the rocks sufficiently “soft” and porous, and the wavelet well calibrated.

6.4. Field example

We expect in the near future to publish several extended papers focussed on the inversion of some real

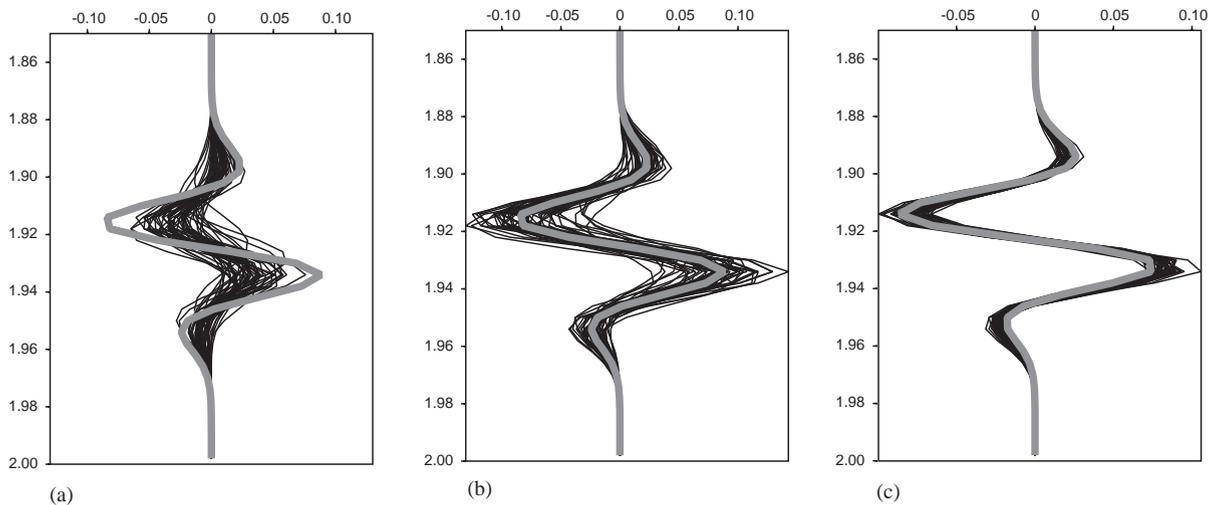


Fig. 7. Three “spaghetti” graphs showing synthetic seismic traces drawn from (a) the prior models with brine in reservoir (b) prior models with oil, and (c) posterior distribution, which puts oil probability at $\approx 97\%$. Noise level is 0.01. “Truth case” trace is shown in gray.

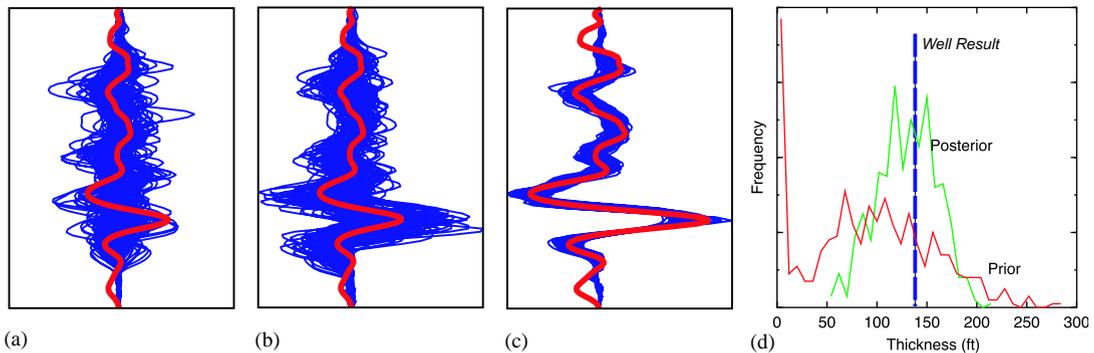


Fig. 8. Synthetic seismic traces overlaid by observed seismic trace at a particular location in survey. (a) Sample models drawn from prior with brine filling reservoir layer. (b) Same, but with oil in reservoir layer. (c) Traces computed from models drawn from posterior distribution, conditional on oil present in reservoir layer. (d) Histogram of reservoir thickness drawn from prior model and after seismic inversion (posterior). Pinchouts and very thin reserves are obviously precluded. Posterior distribution of thickness is also consistent with observed thickness.

field data using Delivery and the workflow of developing the prior. For the moment, we will confine our discussion to a simple test for an actual field, shown in Fig. 8. This is a ‘test’ inversion at the well location. An 8-layer model for a set of stacked turbidite sands has been built with proven hydrocarbons in the second-bottom layer. The sands are quite clean and have high porosities ($\approx 30\%$), so the effects of Gassman substitution are very strong in the reservoir layers. The layers are constructed from log analysis, but their boundaries are set to have a broad prior uncertainty around the 10–15 ms range. Low net-to-gross layers ($N_G < 0.1$) are often set as pure shales ($N_G = 0$) to keep the model dimensionality down. The prior for fluid type (brine:oil) is set as 50:50, as in the previous example.

The inversion at the well location not only confirms the presence of oil (>80% probability) but also demonstrates that the posterior distribution of layer thickness for the reservoir is consistent with the well observation.

7. Conclusions

We have introduced a new open-source tool for model-based Bayesian seismic inversion called Delivery. The inverter combines prior knowledge of the reservoir model with information from the seismic data to yield a posterior uncertainty distribution for the reservoir model, which is sampled to produce a suite of stochastic inversions. The Monte Carlo samples produced by the inverter summarise the full state of knowledge about the reservoir model after incorporating the seismic data, and can be used to generate predictive distributions of most commercial or engineering quantities of interest (net-pay, column-thickness, etc.).

The inverter is driven from a combination of XML files and SU/BHP.SU data, and outputs are in SU/BHP.SU form. The SU backbone means the inverter interfaces nicely with other free seismic software, such as the INT viewer INT, 2003⁵ and BHP’s SU extensions Miller and Glinsky, 2003.⁶ We believe the “small-is-beautiful” philosophy associated with backbone designs improves the flexibility and maintainability of the software.

The authors hope that this tool will prove useful to reservoir modellers working with the problem of seismic data integration, and encourage users to help improve the software or submit suggestions for improvements. We hope that the newer ideas on probabilistic model-comparison and sampling (vis-à-vis the petroleum

community) prove useful and applicable to related problems in uncertainty and risk management.

Acknowledgements

This work has been supported by the BHP Billiton Technology Program.

Appendix. 1–7

There are 7 Appendixes in electronic form on the IAMG server, <http://www.iapg.org/CGEditor/index.html>.

References

- Abrahamsen, P., Hektoen, A.-L., Holden, L., Munthe, K., 1997. Seismic impedance and porosity: support effects. In: Geostatistics Wollongong '96. Kluwer Academic, Dordrecht, The Netherlands, pp. 489–500.
- Abramowitz, M., Stegun, I.A., 1965. Handbook of Mathematical Functions. Dover, New York, 1046pp.
- Andrieu, C.A., Djurić, P.M., Doucet, A., 2001. Model selection by MCMC computation. Signal Processing 81, 19–37.
- Brooks, S.P., 1998. Markov chain Monte Carlo and its application. The Statistician 47, 69–100.
- Buland, A., Omre, H., 2003. Joint AVO inversion, wavelet estimation and noise-level estimation using a spatially coupled hierarchical Bayesian model. Geophysical Prospecting 51 (6), 531–550.
- Buland, A., Kolbjørnsen, A., Omre, H., 2003. Rapid spatially coupled AVO inversion in the Fourier domain. Geophysics 68 (3), 824–836.
- Castagna, J.P., Backus, M.M. (Eds.), 1993. Offset-dependent Reflectivity—Theory and Practice of AVO Analysis. Society of Exploration Geophysicists, Tulsa, OK, 348pp.
- Chu, L., Reynolds, A.C., Oliver, D.S., 1995. Reservoir description from static and well-test data using efficient gradient methods. In: Proceedings of the International Meeting on Petroleum Engineering, Beijing, China. Paper SPE 29999, Society of Petroleum Engineers, Richardson, TX, pp. 467–482.
- Deutsch, C.V., 2002. Geostatistical Reservoir Modeling. Oxford University Press, New York, 386pp.
- Eide, A.L., 1997. Stochastic reservoir characterization conditioned on seismic data. In: Geostatistics Wollongong '96. Kluwer Academic, Dordrecht, The Netherlands, pp. 442–453.
- Eide, A.L., Omre, H., Ursin, B., 2002. Prediction of reservoir variables based on seismic data and well observations. Journal of the American Statistical Association 97 (457), 18–28.
- Eidsvik, J., Omre, H., Mukerji, T., Mavko, G., Avseth, P., 2002. Seismic reservoir prediction using Bayesian integration of rock physics and Markov random fields: a North Sea example. The Leading Edge 21 (3), 290–294.
- Gelman, A., Carlin, J.B., Stern, H.S., Rubin, D.B., 1995. Bayesian Data Analysis. Chapman and Hall, Boca Raton, FL, 526 pp.

⁵INT, 2003. BHP viewer from INT, see link at <http://timna.mines.edu/cwpcodes>.

⁶Miller, R., Glinsky, M., 2003. BHP.SU: BHP extensions to Seismic Unix, see link at <http://timna.mines.edu/cwpcodes>.

- Gunning, J., 2000. Constraining random field models to seismic data: getting the scale and the physics right. In: *ECMOR VII: Proceedings, Seventh European Conference on the Mathematics of Oil Recovery*, Baveno, Italy, p. M-20.
- Huage, R., Skorstad, A., Holden, L., 1998. Conditioning a fluvial reservoir on stacked seismic amplitudes. In: *Proceedings of the Fourth Annual Conference of the IAMG*. <http://www.nr.no/research/sand/articles.html>.
- Koontz, R.S.J., Weiss, B., 1982. A modular system of algorithms for unconstrained minimization. Technical Report, Computers Science Department, University of Colorado at Boulder, Steve Verrill's Java translation is at <http://www1.fpl.fs.fed.us/optimization.html>.
- Leguijt, J., 2001. A promising approach to subsurface information integration. In: *Extended Abstracts, 63rd EAGE Conference and Technical Exhibition*, pp. L-35.
- Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem. *Geophysical Journal International* 151, 675–688.
- Mavko, G., Mukerji, T., Dvorkin, J., 1998. *The Rock Physics Handbook*. Cambridge University Press, Cambridge, UK, 329pp.
- Oliver, D.S., 1996. On conditional simulation to inaccurate data. *Mathematical Geology* 28, 811–817.
- Oliver, D.S., 1997. Markov Chain Monte Carlo Methods for conditioning a permeability field to pressure data. *Mathematical Geology* 29, 61–91.
- Omre, H., Tjelmeland, H., 1997. *Petroleum Geostatistics*. In: *Geostatistics Wollongong '96*. Kluwer Academic, Dordrecht, The Netherlands, pp. 41–52.
- Phillips, D.B., Smith, A.F.M., 1996. Bayesian model comparison via jump diffusions. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*. Chapman & Hall, Boca Raton, FL, pp. 215–239.
- Raftery, A.E., 1996. Hypothesis testing and model selection. In: Gilks, W.R., Richardson, S., Spiegelhalter, D.J. (Eds.), *Markov Chain Monte Carlo in Practice*, Chapman & Hall, pp. 163–187.
- Storn, R., Price, K., 1997. Differential evolution—a simple and efficient heuristic for global optimisation over continuous spaces. *Journal of Global Optimization* 11, 341–359.
- Tarantola, A., 1987. *Inverse Problem Theory, Methods for Data Fitting and Model Parameter Estimation*. Elsevier Science Publishers, Amsterdam, 613pp.