

Delivery: an open–source model–based Bayesian seismic inversion program [★]

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

^a*CSIRO Division of Petroleum Resources, Ian Wark Laboratory, Bayview Ave., Clayton, Victoria, 3168, Australia, ph. +61 3 9545 8396, fax +61 3 9545 8380*

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

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.

Key words: seismic, inversion, AVO, Bayesian, stochastic, geostatistics, Markov Chain Monte Carlo, open–source,

[★] Updated manuscript with additional Appendices from citation (Gunning and Glinsky, 2004). Please cite as both (Gunning and Glinsky, 2004) and the website.

* Corresponding author.

Email address: James.Gunning@csiro.au (James Gunning).

Contents

1	Introduction	4
2	Outline of the model	6
2.1	Description and notation of the local layer model	8
2.2	Construction of the prior	10
3	The forward model	14
3.1	Computing the synthetic seismic	14
3.2	Isopach constraints	19
4	Sampling from the posterior	19
4.1	Mode location and local covariance	20
4.2	Mode enumeration	22
4.3	Sampling	23
5	The software	27
6	Examples	28
6.1	Sand Wedge	28
6.2	Net-to-gross wedge	28
6.3	Simple single-trace fluid detection problem	31
6.4	Field Example	31
7	Conclusions	32
	Appendix 1: Typical computation of effective rock properties	36
	Appendix 2: Error sampling rates	37
	Appendix 3: Mode-location starting points	37
	Appendix 4: An Independence Sampler	39
	Appendix 5: Modified SU trace formats for properties	40
5.1	Model-prior trace formats	40

5.2	Stochastic trace outputs	40
	Appendix 6: Wavelet format	41
	Appendix 7: Usage of the code	41
7.1	Inputs	41
7.2	Outputs	43
	Appendix 8: Incorporation of Floating Grain models into <i>Delivery</i>	44
8.1	Floating grain effects	44
8.2	Recommended methodology	46
8.3	Conversion to <i>Delivery</i> formats	46
8.4	Rock trends	47
8.5	Model system	48
8.6	Inversion Analysis of Posteriors	50
8.7	Delivery Implementation details	52
8.8	Inputs/Outputs	54
	Appendix 9: Vertical model–stacking functionality	54
	Appendix 10: Internal fluid–contact modelling	58
10.1	Upscaling rules	60
10.2	New outputs in the realisations SU file	60
10.3	Example Inversion	61
10.4	Likely Issues	63
	Appendix 11: Model selection facilities	63
	Appendix 11: MAP point estimates and fluid model probabilities in noisy problems	68
12.1	An illustrative toy problem	68
12.2	Demonstration on the Gas–Oil–Flank toy Delivery problem	73
12.3	Recommendations	74

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 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” meth-

ods 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 ¹. 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 (2000), 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 (1997), Oliver (1996). 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

¹ 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.

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 has 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 below. 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,

1998) 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 parallelizable, 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 very 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 potted 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 10-30m) associated with the acquisition geometry²

By assuming 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

² 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).

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 nonlinear 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.

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}) , $i = 1 \dots N_l$. An additional parameter, t_{base} , is required to specified the bottom of the model. Figure 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

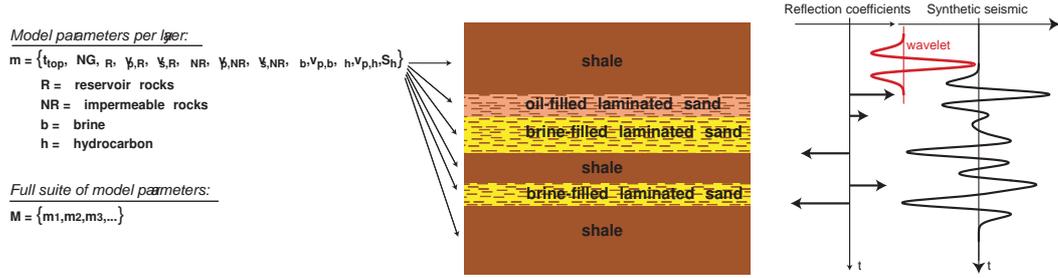


Fig. 1. Schematic of the local layer-based model and its notation. For description of the model parameters see the main text. The reflection coefficient sequence and synthetic seismic are part of the forward model of section 3.

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

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

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

$$\mathbf{m} = \{d, \text{LFIV}, \rho_m, v_{p,m}, v_{s,m}\}. \quad (2)$$

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

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

The extra stochastic factors A, B affect the synthetic seismic and are explained in section 3.1.4. Hereon we use d_m for the dimensionality of the model \mathbf{m} .

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 endpoints 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 properties 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 (un-truncated) 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.

Sometimes it is desirable to force a layer to permanently pinchout in some region of space, and thus effectively drop out of the model: see Appendix 5.1 for a discussion.

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 default density ordering is $\{b, l, o, g\}$, with low-saturation gas placed between brine and oil because it usually arises as a residual saturation trail behind migrating hydrocarbons.

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 $k = 1 \dots N_F$.³ Suppose the fluid–combination k corresponds to the set of fluid labels $f_{ik} \in \{b, l, g, o\}$, $i = 1 \dots N_l$. 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 criterion⁴. 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} , $f = o, g, b, l$ for each fluid type. The two phases present are taken as brine (b) & hydrocarbon (o,g,l). The petrophysicist can assign Gaussian stochastic priors $S_{if} \sim N(\overline{S_{if}}, \sigma_{S_{if}})$, truncated⁵ outside $[0, 1]$ to these saturation parameters, based on regional knowledge.

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.

³ E.g. a two layer system under ordering rule (3), where it is known that gas (and low sat. gas) cannot occur ($F_{ig} = F_{il} = 0$), may have the allowable set {(brine, brine) : (oil, brine) : (oil, oil)}, so $N_F = 3$.

⁴ For example, in the two–layer problem of the preceding footnote, if the prior probability of oil in each layer had been specified as, say, 0.6, the 3 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 $0.316 + 0.474 = 0.79$.

⁵ 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 pinch-out.

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 density⁶, and the set of regressions is

$$\begin{aligned}\phi_f &= (A_\phi + B_\phi v_{pf}) \pm \sigma_{\phi f} \\ v_{sf} &= (A_{v_s} + B_{v_s} v_{pf}) \pm \sigma_{sf} \\ v_{pf} &= (A_{v_p} + B_{v_p} d + C_{v_p} \text{LFIV}) \pm \sigma_{pf},\end{aligned}\tag{7}$$

where d =depth. The last equation models v_p as linear in d = 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:

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

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.

⁶ Density is computed from

$$\rho_{\text{sat}} = (1 - \phi)\rho_g + \phi\rho_{\text{fluid}}\tag{6}$$

for permeable members.

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 $v_{p,f}$ 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 et al. (1998), p. 92). This assumes that the wavelengths are long com-

pared 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 p. 205, Mavko et al. (1998)) 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-sat. 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 can be written ⁷

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

⁷ Assuming no pressure effects on the dry modulus K_{dry} .

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. 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 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 4 samples to the time of a spike. This ensures 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 ⁸

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

with

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

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

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

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

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

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

⁸ From the small-contrast Zoeppritz equation for $R_{pp}(\theta)$, expanded to $O(\theta^2)$ (top of p.63, Mavko (1998))

and layer 2 below layer 1. All properties in these formulae are effective properties obtained as per 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

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

where the factors A and B may be stochastic, typically Gaussian with means close to unity and small variances: $A \sim N(\bar{A}, \sigma_A)$, $B \sim N(\bar{B}, \sigma_B)$ ⁹.

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

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

where the *range* of the error-sampling points is computed from the prior mean-layer time range and the wavelet characteristics as per figure 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 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 recovery the error that would obtain 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 $S_N \approx 3$.

⁹ In some cases, there are uncertainties with the overall wavelet scale. The facility to lock A and B also exists, so $B = A \sim N(\bar{A}, \sigma_A)$ is then used in the prior.

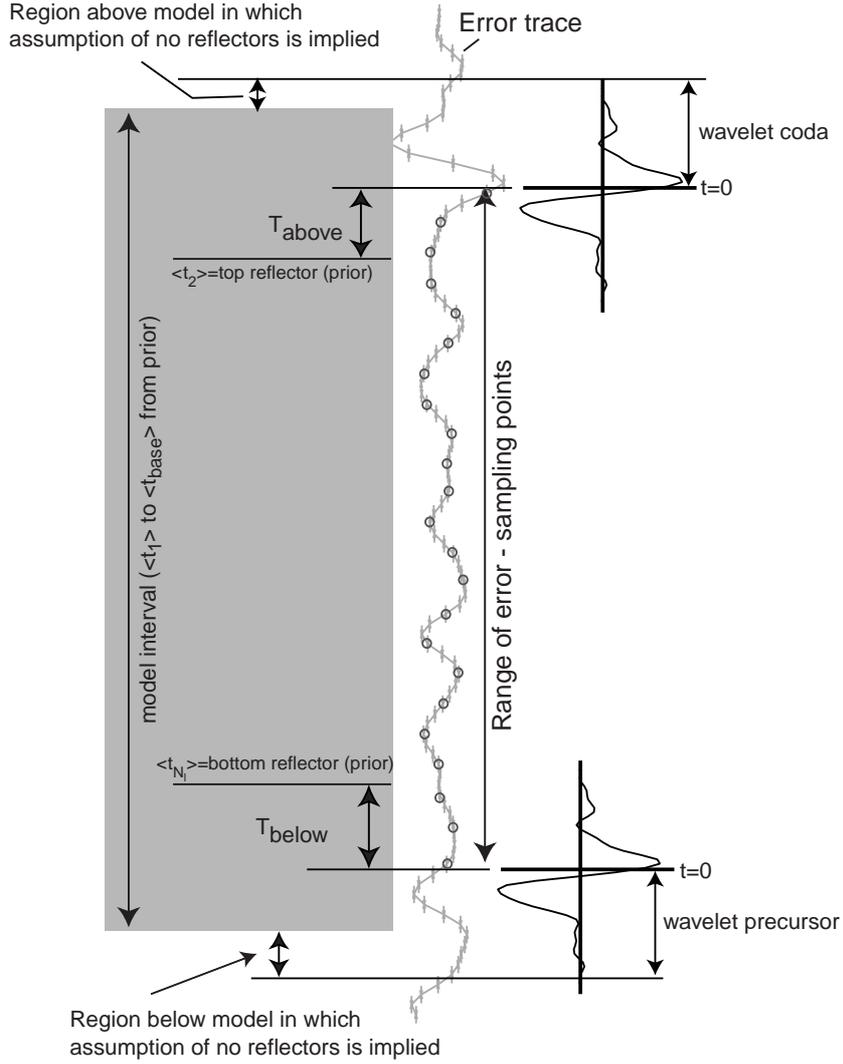


Fig. 2. Diagram indicating range over which the error samples are computed. In general $T_{above, below}$ are free parameters, but we take $T_{above} = \text{wavelet-coda}$ and $T_{below} = \text{wavelet-precursor}$ as a sensible choice, since this allows the full reflection energy from the top and bottom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

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 $j = 1, \dots, N_t$ 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

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

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 equation (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. 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 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 Adrieu.

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:

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

where the full prior is

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

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 minimizers ¹⁰ 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 -dimensional 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.

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 three 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 the seismic (25) or isopach likelihoods (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} :

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

¹⁰The code actually uses an adaptation of Verrill's java UNCMIN routines (Koontz and Weiss, 1982), which implement both BFGS and Hessian-based Newton methods.

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

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

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

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

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” mean, and a set of starting times are assembled as per the description in 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. Figure 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 ¹¹. If a mode is acceptable, the Newton-updated mean and

¹¹ the remaining freedom in the frozen parameters is unlikely to redeem a bad loop-skip

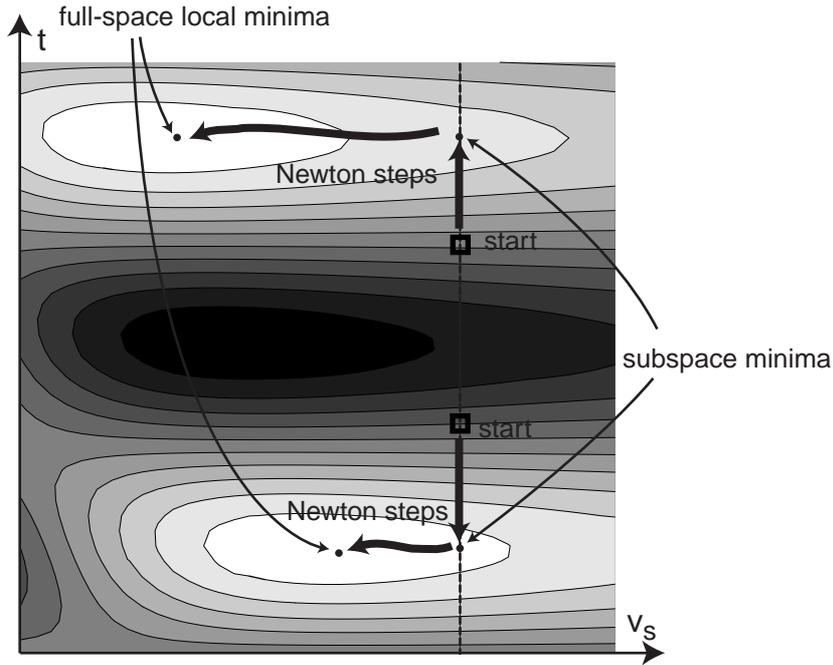


Fig. 3. Minimisation scheme consisting of subspace minimisations (the subspace represented by t) starting from dispersed points, followed by global Newton steps, yielding a set of local minima in the full parameter space.

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, equation (59), described in Appendix 4. Diagnostic routines capable of drawing graphs of the log posterior centered 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.

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 Ch. 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 figure 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

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

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

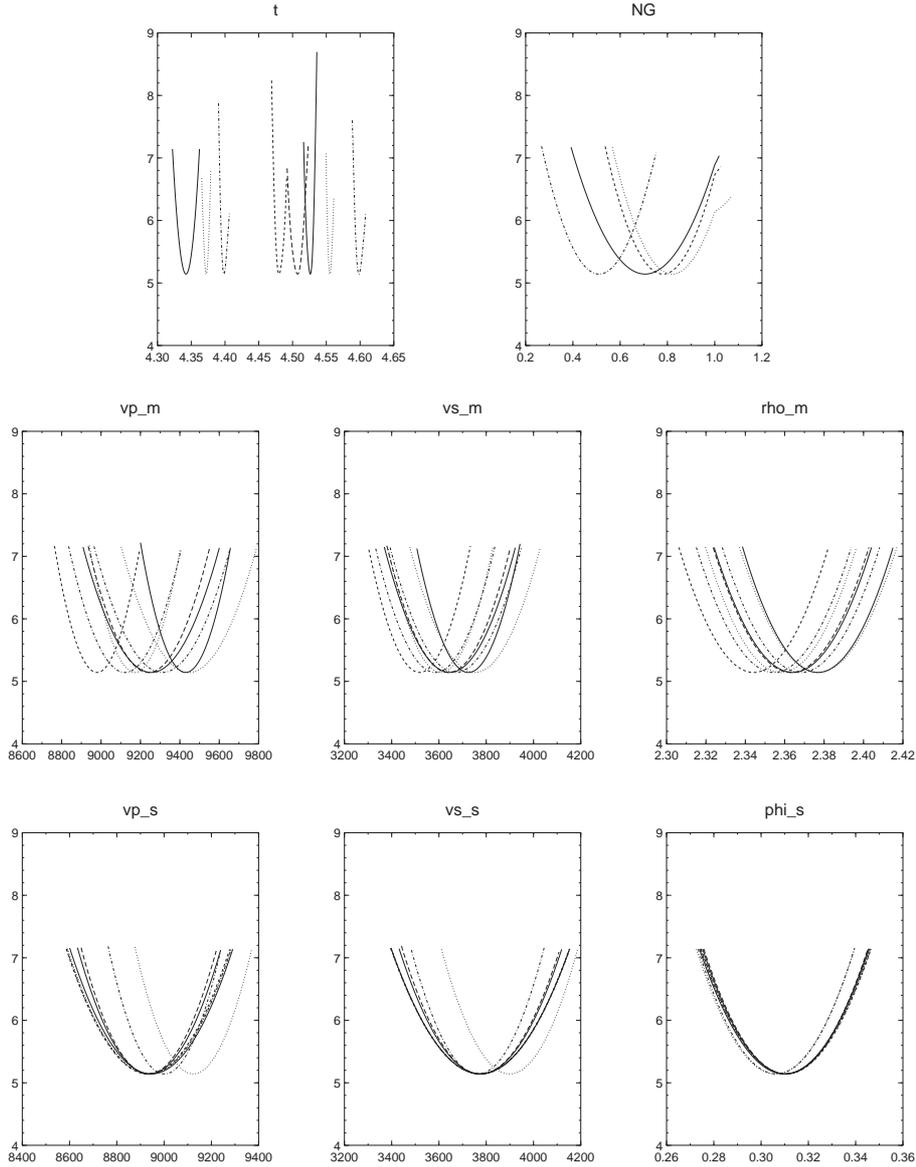


Fig. 4. Series of log-posterior cross-sections taken through the maximum-likelihood point found from the optimisation step and the sequence of Newton updates. Cross-sections are shown for the parameters shown, in all relevant layers, over a range of 2 standard deviations as computed from the local covariance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

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 equation (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 equation (42).

The linear translation embodied in equation (40) is based on an 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 equation (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 (Ch. 7, Gelman et al. (1995)), 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 very much more efficient language than in its early days as a purely interpreted language (see the Java Numerics website as an entry point ¹². The core cpu demands in the code are (a) linear algebra calls, for which we use the efficient CERN colt library ¹³ 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

¹²Boisvert, R., Pozo, R., 2003. The Java Numerics website, <http://math.nist.gov/javanumerics/>.

¹³Hoschek, W., 2003. The CERN colt java library, <http://dsd.lbl.gov/~hoschek/colt>.

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, 2003), 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.

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 (20ms, 50ms respectively). Figure 5 illustrates pre and post inversion stochastic samples of the layers, displayed with many realisations per trace. Here the noise strength is about 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 same as the “truth–case” model, but has a (very broad) prior standard deviation of $\sigma_{\text{NG}} = 0.3$. Figure 6 illustrates pre and post

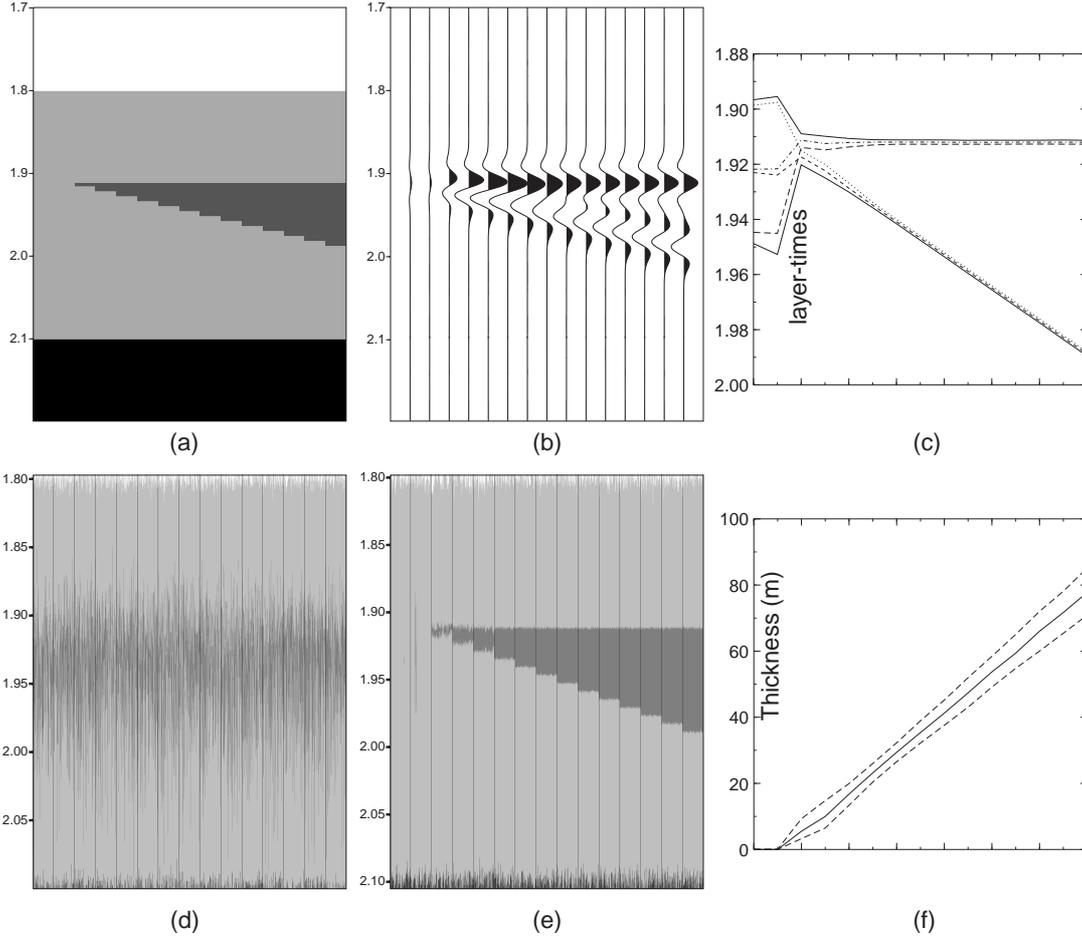


Fig. 5. (a) Truth case wedge model, (b) truth case seismic, (d) pre-inversion stochastic samples (i.e. drawn from prior distribution only), (e) post inversion samples (100 per trace), (c) p_{10}, p_{50}, p_{90} quantiles for the layer times of the wedge, and (f) p_{10}, p_{50}, p_{90} quantiles of the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

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.

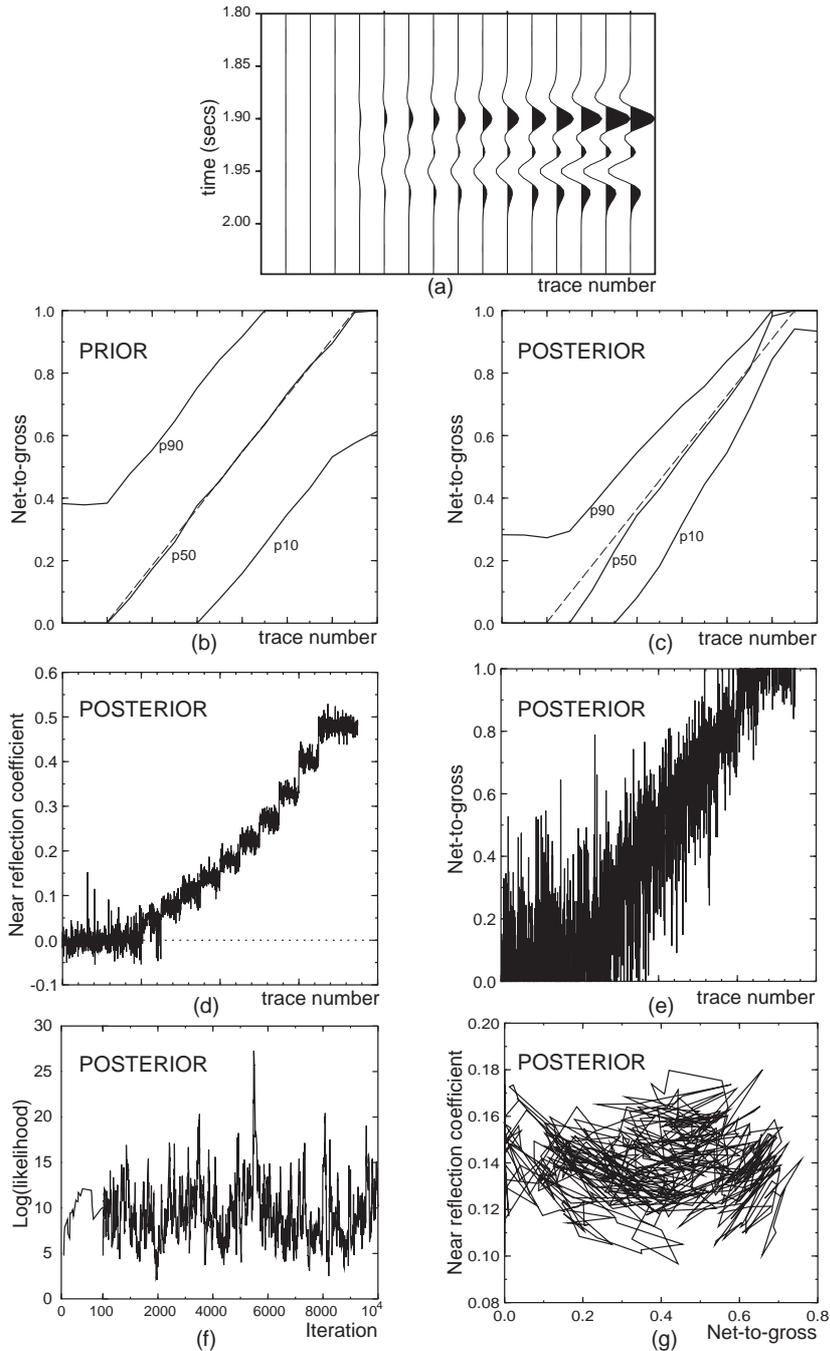


Fig. 6. (a) Truth case seismic for net-to-gross (N_G) wedge model, (b) pre-inversion (prior) spread of N_G as show by p10, p50, p90 quantiles; truth-case shown dashed, (c) as per (b), but post-inversion. Inset (d) shows the near-stack reflection coefficient variations, and (e) the associated samples of N_G . The signal-to-noise ratio here is very favourable, with noise 1/10th of the peak seismic response. The residual uncertainty in the N_G is due to the fact that there is also uncertainty in the layer velocity variations. The seismic constrains primarily the reflection coefficient, but the system indeterminacy will still allow appreciable variation in the N_G . Convergence & mixing of the 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 the N_G of layer 2 vs. its reflection coefficient at the top.

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 figure 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.

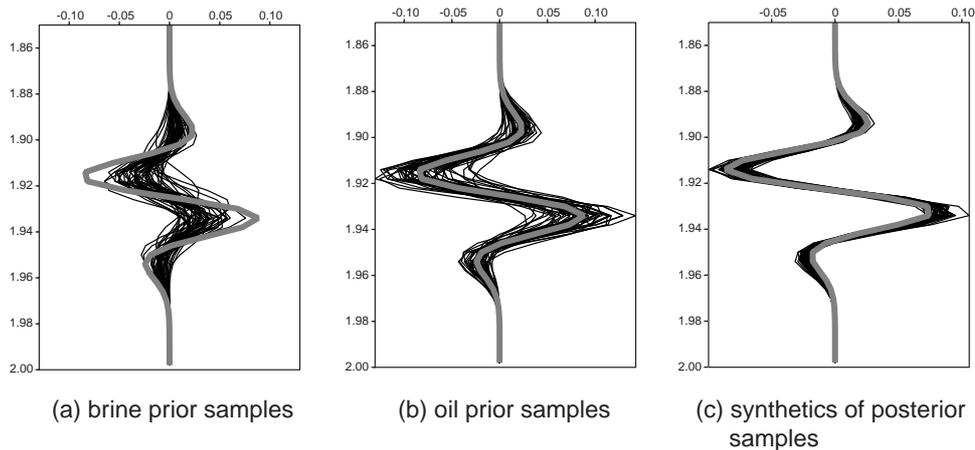


Fig. 7. Three “spaghetti” graphs showing synthetic seismic traces drawn from (a) the prior models with brine in the reservoir (b) prior models with oil, and (c) the posterior distribution, which puts the oil probability at $\approx 97\%$. the noise level is 0.01. The “truth case” trace is shown in gray.

6.4 Field Example

We expect in the near future to publish several extended papers focussed on the inversion of some real 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 figure 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 10-15ms 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

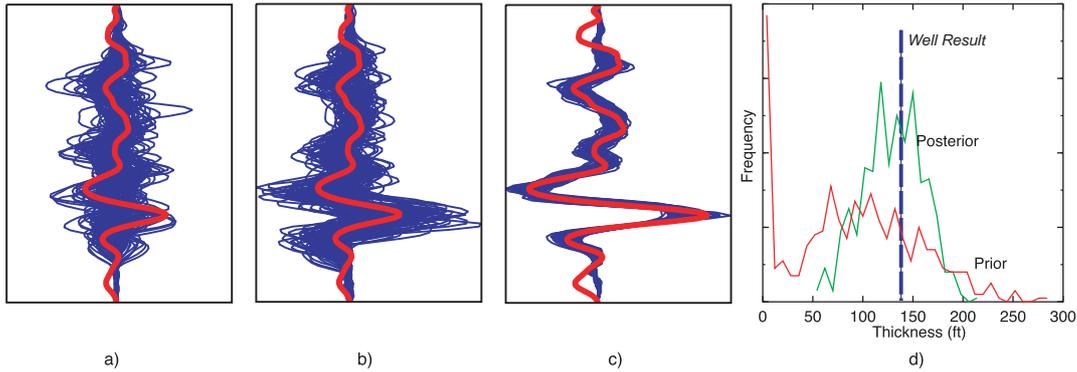


Fig. 8. Synthetic seismic traces overlaid by the observed seismic trace at a particular location in the survey. (a) Sample models drawn from the prior with brine filling the reservoir layer. (b) The same, but with oil in the reservoir layer. (c) Traces computed from models drawn from the posterior distribution,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

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 BHP viewer ¹⁴ and BHP’s SU extensions ¹⁵. We believe the “small-is-beautiful” philosophy associated with backbone designs improves the flexibility and maintainability of the software.

¹⁴ see link at <http://timna.mines.edu/cwpcodes>

¹⁵ *BHP* (Miller and Glinsky) extensions to Seismic Unix, <http://timna.mines.edu/cwpcodes>

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.

Acknowledgement

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

References

- Abrahamsen, P., et al., 1997. Seismic impedance and porosity: support effects. In: *Geostatistics Wollongong '96*. Kluwer Academic, pp. 489–500.
- Abramowitz, M., Stegun, I. A., 1965. *Handbook of Mathematical Functions*. Dover.
- 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., Kolbjornsen, A., Omre, H., 2003. Rapid spatially coupled AVO inversion in the Fourier domain. *Geophysics* 68 (3), 824–83.
- Buland, A., Omre, H., 2000. Bayesian AVO inversion. Abstracts, 62nd EAGE Conference and Technical Exhibition, 18–28 See extended paper at <http://www.math.ntnu.no/~omre/BuOmre00Bpaper.ps>.
- Castagna, J. P., Backus, M. M. (Eds.), 1993. *Offset-dependent reflectivity – Theory and practice of AVO Analysis*. Society of Exploration Geophysicists.
- Chu, L., Reynolds, A. C., Oliver, D. S., Nov. 1995. Reservoir Description from Static and Well-Test Data using Efficient Gradient Methods. In: *Proceedings of the international meeting on Petroleum Engineering*, Beijing, China. Society of Petroleum Engineers, Richardson, Texas, paper SPE 29999.
- 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>.
- Denison, D. G. T., et al., 2002. *Bayesian Methods for Nonlinear Classification and Regression*. Wiley.
- Deutsch, C. V., 2002. *Geostatistical Reservoir Modelling*. Oxford University Press.
- Eide, A. L., 1997. Stochastic reservoir characterization conditioned on seismic data. In: *Geostatistics Wollongong '96*. Kluwer Academic.

- 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., et al., 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.
- Gilks, W., Richardson, S., Spiegelhalter, D., 1996. *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- Gunning, J., 2000. Constraining random field models to seismic data: getting the scale and the physics right. In: *ECMOR VII: Proceedings, 7th European conference on the mathematics of oil recovery*, Baveno, Italy.
- Gunning, J., 2003. **Delivery** website: <http://www.csiro.au/products/Delivery.html>, accessed 2 Jan 2010.
- Gunning, J., Glinsky, M., 2004. Delivery: an open-source model-based Bayesian seismic inversion program. *Computers and Geosciences* 30 (6), 619–636.
- Gunning, J., Glinsky, M., 2006. WaveletExtractor: A Bayesian well-tie and wavelet extraction program. *Computers and Geosciences* 32, 681–695.
- Huuge, R., Skorstad, A., Holden, L., 1998. Conditioning a fluvial reservoir on stacked seismic amplitudes. In: *Proc. 4th 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. Tech. rep., Comp. Sci. Dept., University of Colorado at Boulder.
- Leguijt, J., 2001. A promising approach to subsurface information integration. In: *Extended abstracts, 63rd EAGE Conference and Technical Exhibition*.
- Mackay, D. J. C., June 2002. *Information Theory, Inference & Learning Algorithms*, 1st Edition. Cambridge University Press.
- 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.
- 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.
- Phillips, D. B., Smith, A. F. M., 1996. Bayesian model comparison via jump diffusions. In: *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- Raftery, A. E., 1996. Hypothesis testing and model selection. In: *Markov Chain Monte Carlo in Practice*. Chapman and Hall.
- Storn, R., Price, K., 1997. Differential evolution – a simple and efficient heuris-

tic 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.

Appendices

Appendix 1: Typical computation of effective rock properties

Here we illustrate how the effective rock properties for a single layer might be computed from a full suite of fundamental model parameters (equation (1)) for the layer.

- Compute the brine+matrix density

$$\rho_{\text{sat,b}} = \phi_s \rho_b + (1 - \phi_s) \rho_g. \quad (43)$$

- Compute the associated moduli

$$\mu_{\text{sat,b}} = \rho_{\text{sat,b}} v_{s,s}^2 \quad (44)$$

$$M_{\text{sat,b}} = \rho_{\text{sat,b}} v_{p,s}^2 \quad (45)$$

$$K_{\text{sat,b}} = M_{\text{sat,b}} - \frac{4}{3} \mu_{\text{sat,b}} \quad (46)$$

$$\mu_{\text{sat,fl}} = \mu_{\text{sat,b}} \quad (47)$$

- Use the fluid properties to compute $K_b = \rho_b v_{p,b}^2$ and $K_h = \rho_h v_{p,h}^2$, and then

$$K_{\text{fl}} = \left(\frac{S_h}{K_h} + \frac{1 - S_h}{K_b} \right)^{-1}$$

- Using literature values of K_g , compute $K_{\text{sat,fl}}$ using equation (13). This comes out to be

$$K_{\text{sat,fl}} = K_g / \left\{ 1 + \left[\frac{1}{\phi_s} \left(\frac{1}{K_g/K_{\text{fl}} - 1} - \frac{1}{K_g/K_b - 1} \right) + \frac{1}{K_g/K_{\text{sat,b}} - 1} \right]^{-1} \right\} \quad (48)$$

This equation has regions of parameter space with no positive solution (often associated with, say, small ϕ). The forward model must flag any set of parameters that yield a negative $K_{\text{sat,fl}}$ and furnish a means of coping sanely with such exceptions. The same remark applies to equations (46) and (51) below.

- Compute the new effective density

$$\rho_{\text{sat,fl}} = (1 - \phi_s) \rho_g + \phi_s (S_h \rho_h + (1 - S_h) \rho_b),$$

- Compute also the new effective p-wave modulus from

$$M_{\text{sat,fl}} = K_{\text{sat,fl}} + \frac{4}{3} \mu_{\text{sat,fl}}$$

- Compute the impermeable–rock moduli

$$\mu_m = \rho_m v_{s,m}^2 \quad (49)$$

$$M_m = \rho_m v_{p,m}^2 \quad (50)$$

$$K_m = M_m - \frac{4}{3}\mu_m \quad (51)$$

- Mix the f -substituted permeable rock with the shale properties using the rock–mixing Backus averaging formula (9). Compute also the mixed density from (10).
- From the mixed rock, back out the velocities $v_{p,\text{eff}} = (M_{\text{eff}}/\rho_{\text{eff}})^{1/2}$, and $v_{s,\text{eff}} = (\mu_{\text{eff}}/\rho_{\text{eff}})^{1/2}$.

Appendix 2: Error sampling rates

When comparing synthetic seismic traces to co–located actual seismic, it is a question of some nicety as to what rate of sampling of the difference trace must be applied to quantify the error between the two. It is trivial to show that if a random process has, eg. a Ricker-2 power spectrum (i.e. $\sim f^2 \exp(-(f/f_{\text{peak}})^2)$), then 95% of the spectral energy in the process can be captured by sampling at times

$$\Delta T_s = 0.253/f_{\text{peak}}, \quad (52)$$

where f_{peak} is the peak energy in the spectrum (and often about half of the bandwidth). Most practical seismic spectra will yield similar results.

The sampling rate in equation (52) is about 2 times the Nyquist rate associated with the peak frequency. If the error is acquired with a faster sampling rate than this, adjacent error samples will be strongly correlated and the error term will become large. Since the error covariance is unknown, and bearing in mind that we seek an error measure corresponding to a maximum number of “independent” data points, we use the rate given by equation (52) and model the error points as i.i.d. values. The peak frequency f_{peak} is estimated at initialisation from the FFT spectrum of the wavelet.

Appendix 3: Mode–location starting points

The location, characterisation, and enumeration of local modes in the posterior distribution is performed by a loop over a set of strategically chosen starting

points in the model parameter space. Since one of the major functions of the prior distribution is to exclude improbable posterior modes, it makes sense to base the choice of starting points on the prior distribution.

The fundamental variables associated with multi-modality are the layer times, in that the posterior likelihood surface is frequently oscillatory in the layer times. Convergence to a secondary or minor solution is usually called loop-skipping, and may or may not be desirable. In situations where loop skipping may be permissible or even desirable, the prior uncertainty on the layer times is usually set to be quite broad, as these secondary solutions may be feasible. Convergence to these solutions can then be prompted by setting the initial layer times fed to the minimisation routine at values likely to be within the basin of attraction of the desired solution. The heuristic method used for defining these starting points is as follows:

Form a set of starting configurations, for each fluid state, by:

- (1) Forming an approximate multi-Gaussian posterior distribution for the properties by updating the prior to account for any isopach constraints and the depth consistency requirement. This posterior is characterised by the mean $\tilde{\mathbf{m}}$ and covariance \tilde{C} . Recall that the layer times are the first $i = 1 \dots (n + 1)$ elements in the vector $\tilde{\mathbf{m}}$. Fix all non-time parameters $i = n + 2 \dots$ at the values in $\tilde{\mathbf{m}}$.
- (2) From this posterior, compute the expected layer time thickness $\overline{\Delta t_i}$, layer time uncertainties σ_{t_i} and layer time thickness uncertainty $\sigma_{\Delta t_i}$, from

$$\overline{\Delta t_i} = \tilde{\mathbf{m}}_{i+1} - \tilde{\mathbf{m}}_i \quad (53)$$

$$\sigma_{t_i}^2 = \tilde{C}_{i,i} \quad (54)$$

$$\sigma_{t_{i+1}}^2 = \tilde{C}_{i+1,i+1} \quad (55)$$

$$\sigma_{\Delta t_i}^2 = \tilde{C}_{i,i} + \tilde{C}_{i+1,i+1} - 2\tilde{C}_{i,i+1}. \quad (56)$$

- (3) Now, if $\sigma_{\Delta t_i} / \overline{\Delta t_i} > 0.5$, this amounts to a $\approx 5\%$ chance of the layer pinching out ($\Delta t_i < 0$) before taking into account the seismic, so this is likely enough to warrant investigation. The starting points

$$t'_i = \frac{\sigma_{t_{i+1}} \tilde{\mathbf{m}}_i + \sigma_{t_i} \tilde{\mathbf{m}}_{i+1}}{\sigma_{t_{i+1}} + \sigma_{t_i}} \quad (57)$$

$$t'_{i+1} = t'_i - 2\epsilon_{FD} \quad (58)$$

are then used as suitable starting times for the times t_i, t_{i+1} . The remaining times $t_j, j \neq i, i + 1$ are set at the posterior means $t_j = \tilde{\mathbf{m}}_j$. If more than one layer is potentially pinched out, we form a set of starting points by looping over the set of potentially-pinched-out layers, and setting only one of these layers at a time to the pinched-out starting configuration just described, with remaining times at the values in $\tilde{\mathbf{m}}$.

Appendix 4: An Independence Sampler

For each fixed fluid combination k , a successful location of a mode in the full parameter space will produce a “maximum likelihood” vector $\tilde{\mathbf{m}}_{k_j}$ and local covariance \tilde{C}_{k_j} , by successive Newton iterates (equations (33)–(37)). Here the index j runs over the set of distinct modes we have found for each fluid combination by starting from separated starting points, as per Appendix 3 (in many cases there will be only one). To save notational baggage, we will use k henceforth to mean k_j - the context makes the meaning clear.

The Laplace approximation to the mode posterior marginal likelihood is formed by assuming the posterior surface is fully Gaussian, centred at the maximum-likelihood point, and characterised by the covariance \tilde{C}_k . Integrating over all the model parameters then gives the marginal probability to be proportional to

$$q_k \sim (2\pi)^{d_{m_k}} |\tilde{C}_k|^{1/2} \Pi(\tilde{\mathbf{m}}_k | \mathbf{S}), \quad (59)$$

where the posterior Π is evaluated as per (27,28), and the determinant and dimension-varying terms in (28) are carefully retained.

A full independence sampler can then be implemented using a Metropolis Hastings method. With the chain in mode k , with model vector \mathbf{m}_k , we propose a new mode k' with probability $\sim p'_k$, whose new model vector \mathbf{m}'_k is drawn from $q(\mathbf{m}'_k) = N(\mathbf{m}'_k | \tilde{\mathbf{m}}'_k, \tilde{C}'_k)$. We accept this candidate with the usual probability

$$\alpha = \min\left(1, \frac{\Pi(\mathbf{m}'_k | \mathbf{S}) p_k q(\mathbf{m}_k)}{\Pi(\mathbf{m}_k | \mathbf{S}) p_{k'} q(\mathbf{m}'_k)}\right), \quad (60)$$

or leave the state of the chain unchanged. This method has had somewhat mixed success - it is perfect when drawing from the prior (since the prior is multi-Gaussian), but is susceptible of being trapped in particularly favourable states for long periods when the posterior departs sharply from Gaussianity. Further, any thick tails in the posterior density are not likely to be explored by the proposal density, since it has compact Gaussian tails. The proposal overdispersion recommended by Gelman (Gelman et al., 1995, Ch. 11) does not work well, since the dimensionality is so high that even mildly overdispersed t -distributions dramatically reduce the acceptance rates, even when sampling from the prior.

It has proven useful to implement an approximate sampler where the modes are proposed from the posterior marginal ($\sim q_{k'}$) with normal approximations at each mode ($q(\mathbf{m}'_k) = N(\mathbf{m}'_k | \tilde{\mathbf{m}}'_k, \tilde{C}'_k)$), and mandatory acceptance ($\alpha = 1$).

Such candidates appear to very rarely produce unlikely χ^2 values for the seismic and isopach likelihoods.

Appendix 5: Modified SU trace formats for properties

5.1 Model-prior trace formats

The local prior information for the inversion at each trace is communicated to the inversion code by a SU data set with a set of traces on a pattern and ordering that matches the SU seismic data set exactly. The xml *Modelparameters.xml* file which sets up the inversion will specify which prior quantities vary on a trace-by-trace basis, and the block indices k at which they are located. Suppose there are N_p properties which vary, and there are N_l layers. The value for property k in layer l is then the $(l + (k - 1)N_l)$ th float value in the current trace of prior data. There are some special rules as well

- The layer times must always be present in the prior trace, and they must always be the last block ($k = N_k$). Further, one additional float value is appended for the base time of the last layer t_{base} .
- All values are expected to be positive, and are converted to positive numbers if not. The value -999.0 signifies a “bad value”, and will be ignored (the default prior will then apply)
- All times are in milliseconds.
- Layers can be effectively suppressed from the model by providing the isopach constraint $\Delta Z = N(0, 0)$ (see eqn (26)), either from the XML or the model traces file for some subset of traces. If this is supplied, the layer is deemed *inactive*. This is advantageous when a layer is known to permanently pinchout in some region. Any model reduction help in speeding up the inversion.

5.2 Stochastic trace outputs

Delivery will write stochastic samples from the posterior to the requested output file in a very similar format to the model-prior traces. A sequence of blocks is written, each of size N_l , and the last block of $N_l + 1$ floats is either time or depth, as specified by the obvious entry in the *ModelDescription.xml* file. The names of the properties can be written to a single-line ascii file by supplying **-BHPcommand filename** to the *delivery* command line.

Similarly, the *deliveryAnalyser* program can collate the large statistical ensemble produced by the *delivery* output into a compact set of traces representing

either means, standard deviations or quantiles of all the salient quantities of interest. The set of property names of the blocks in these summary traces is obtained by adding **-BHPcommand filename** to the *deliveryAnalyser* command line.

Appendix 6: Wavelet format

Wavelets are expected to be single trace SU files, with the fields `ns`, `dt` and `f1` set to reflect the number of samples, sampling interval, and starting time of the first sample (beginning of precursor). The `ns` and `f1` fields are set so as to make the wavelet as compact as possible with respect to the tapering, so excessive samples are discouraged, but the tapering must also be smooth to the first and last sample. Wavelets generated by the *WaveletExtractor* code (Gunning and Glinsky, 2006) distributed with *Delivery* are automatically properly initialized.

Appendix 7: Usage of the code

7.1 Inputs

Typical usage of the **delivery** inversion code demands 3 input files:

- An XML file **ModelDescription.xml** specifying the disposition of the layers, the rock and fluid physics, and so meta-information about the su traces containing the variable prior information.
- One or two SU files for the seismic data. These will represent near (and possibly far) offset data.
- An SU file containing the variable prior information.

In distributed systems, the latter two items may refer to named pipes on a unix/Linux system.

A typical inversion would be run with the command

```
% delivery ModelDescription.xml -m prior_traces.su -v 3 -PD -o realisations.su -N 1000
```

which will generate (-N) 1000 realisations per trace and write them (-o) to the file **realisations.su**, using the pinchout detection (-PD) BFGS methods described in section 4.1. The prior model (-m) is read from the SU file **prior_traces.su**. Many other commandline options are available, and can be perused in the code's SU-style self documentaion.

7.1.1 The ModelDescription.xml file

The inversion is chiefly driven by an XML file that specifies the necessary rock-physics, layer descriptions, and information about the seismic. A well formed XML file is crucial to this end, so the file can be created using a schema-driven GUI editor expressly built for the purpose, which is also available at the website. “Help” documentation for the inverter is available through standard menus. The editor is also capable of creating, running and controlling complex scripts on distributed systems, using a schema for common unix, SU, BHP_SU and **delivery** commands.

The structure of the XML document is controlled by the schema, and the meaning of the various entries is readily inferred by reference to the examples in the distribution.

Some particular entries in the XML file are worth further explanation

- The attribute “varies_areally” may be set to true or false for a certain *property*. If true, the quantity in question will vary on a trace-by-trace basis, and is expected to be read in from the SU model file. The “<property_indices>” block then will contain an entry “*property_link*” which defines the block of the SU model file containing the property (see Appendix 5).
- Various properties may contain the attribute “same_as_layer_above”. This is set to true when the property in question should be a replica of the same property in the layer above. For example, the low-frequency-interval-velocity (LFIV) is usually a coarse scale measure representing an average across many layers, so it is best fixed at the same value (or mapped to the same stochastic variable) for all of the relevant layers. Likewise, the “depth” input is an approximation used to sample from the trend curves, and may safely be set to the same value for many adjacent thin layers. Prudent use of this attribute helps to reduce the dimensionality of the final model.
- Each “<layer>” entry may have the attribute “synchronise_rock_matrix” set to true. This is a flag to indicate that any stochastic rock properties of a given rock type in this layer must be matched to those of the matching rock type in the layer above (for example, the v_p , v_s and ϕ_s of the permeable member in the two layers is mapped to the same underlying stochastic variable. By this means, two adjacent layers with matching end-member rocks are made acoustically indistinguishable but for fluid property differences. Any internal reflections are then due to fluid contrasts, and this feature can then be used to make the layer boundaries track a fluid interface.
- Linear trend curves are of form $v_s = \text{slope} \times v_p + \text{intercept} \pm \text{sigma}$. The Gardner-Gardner-Gregory regression curves for impermeable rocks are of form $\rho = \text{factor} \times v_p^{\text{exponent}} \pm \text{sigma}$.
- The “<output>” block has the following useful flags.
 - “<density_ordering>” is set to one of full/partial/none, with the meanings

as described in section 2.2.2.

- “<realisation_number_header_field>” denoted the SU header word used to store the realisation number.
- “<master_depth_layer_number>” denotes the layer number from which depths of the other layers will be computed using summations of layer thicknesses ($v_{p,\text{eff}} \times \text{one-way time thickness}$). The actual depth of the master layer is the depth variable d in the model vector \mathbf{m} .

7.2 Outputs

The inversion code dumps stochastic realisation of the posterior model to a series of SU traces, so the fastest scan direction in the resulting (large) SU file is that of the realisation number, which defaults to the **mark** header word. Posterior analysis of the simulations is performed with a second program **deliveryAnalyser**, which can form various statistical quantities of interest, synthetic seismic traces, and blocked layer maps of the simulations. Some of the salient quantities of interest include

- All the basic model variables in \mathbf{m} of equation (3).
- Fluid type.
- Effective layer properties such as $v_{p,\text{eff}}$, ρ_{eff} , Z_{eff} (impedance).
- Effective reflection coefficients for each stack at the layer top.
- Both a normalised and full χ^2 variable indicating the degree of match with the seismic (*summing over all samples and stacks*), from equation (25)).

$$\chi_{\text{seismic}}^2 \equiv f_{\text{error}} \sum_{\text{error-sampling points}} (\mathbf{S}_{\text{syn}} - \mathbf{S})^2 / \sigma_s^2 \quad (61)$$

$$\chi_{\text{seismic,normalised}}^2 \equiv \frac{\chi_{\text{seismic}}^2}{N_{\text{error-sampling-points}}} \quad (62)$$

- Thickness (one-way time-thickness $\times v_{p,\text{eff}}$).
- Net sand (net sand = thickness $\times N_G$).
- Net hydrocarbon (Net sand \times porosity \times saturation).

Like **delivery** and SU programs, the program has self documentation, but some typical analysis of a suite of realisations **realisation.su** might be

- (1) **deliveryAnalyser -i realisations.su -p-ascii t 6 | ascii-graphing-program**
Print the time of layer 6 from successive realisations out as an ascii stream and graph it.
- (2) **cat realisations.su |deliveryAnalyser --mean NG 6**
Print a series of realisation averages of N_G in layer 6, one for each trace location.

- (3) **cat realisations.su |deliveryAnalyser --filter fluid_type = 1 4 --trace--filter 'fdr>1234' --mean NG 6**
 Print a series of realisation averages of N_G in layer 6, one for each trace location, but only for realisations with oil (fluid_type=1) in layer 4, and for the traces with fdr>1234 (note the shell protection).
- (4) **deliveryAnalyser -i realisations.su --histogram net-sand 6 --trace-filter 'fdr=34,trafc=12' | ascii-graphing-program**
 Print a histogram of net-sand ($N_G \times \text{thickness}$) in layer 6 at the specified location.
- (5) **deliveryAnalyser -i realisations.su -s 4 5 R_near wavelet.su | suswapbytes | suximage**
 Produce a synthetic near-stack seismic from the model realisations over the time interval 4-5 seconds from the wavelet **wavelet.su** and display this on a little endian machine (all the java codes wrote big-endian SU files).

Appendix 8: Incorporation of Floating Grain models into *Delivery*

A lithological phenomenon that affects some kinds of reservoir rocks is the presence of granular material in the pore space that does not support the rock matrix, though it does reduce the porosity and permeability. We assume adequate rock-physics regressions for the relevant facies can be built from log data samples unaffected by such material. *Delivery* has been extended to model these effects along the following lines.

8.1 Floating grain effects

The effects of floating grains in the rock matrix can be taken into account using regressions curves of the form

$$\phi = \frac{-\phi_{\text{flt}}}{1 - f_c} + a_\phi + b_\phi Z_{\text{eff}} + \epsilon_\phi \quad (63)$$

$$v_p = a_p + b_p(\phi_{\text{flt}} + (1 + g)\phi) + \epsilon_p \quad (64)$$

where f_c is the capture fraction (volume of second material taken into supporting matrix divided by total secondary material). The effective stress dependence appears is some well chosen function Z_{eff} , with available data, which may be, e.g. of form $Z_{\text{eff}} = (1 - \exp(-P/P_0))$. The ϵ terms are noise terms.

These forms are chosen as linearisations of the nonlinear Gassman model

$$v_p(\phi, \phi_{\text{ft}}) = \left\{ \frac{K_g}{\rho_g + (\rho_b - \rho_g)\phi} \left(\frac{3(1 - \nu_m)}{1 + \nu_m} \beta(\phi, \phi_{\text{ft}}) + \frac{(1 - \beta(\phi, \phi_{\text{ft}}))^2}{(\phi + \phi_{\text{ft}})((K_g/K_f(\phi, \phi_{\text{ft}}) - 1) + 1 - \beta(\phi, \phi_{\text{ft}}))} \right) \right\}^{1/2} \quad (65)$$

about a mean porosity $\bar{\phi}$ and $\phi_{\text{ft}} = 0$, which arises from the following strong assumptions. The grains are assumed to have the same properties as the matrix (same material – subscript g), and variations in the Poisson's ratio ν_m with ϕ and ϕ_{ft} are neglected. We model the rock beta behaviour as a function of structural porosity $\phi_s = \phi + \phi_{\text{ft}}$ as

$$\beta(\phi, \phi_{\text{ft}}) = (1 - (\phi + \phi_{\text{ft}})/\hat{\phi}_0)^\lambda, \quad (66)$$

and the 'effective fluid' modulus by the Reuss average of floating grains and brine moduli

$$K_f(\phi, \phi_{\text{ft}}) = \left(\frac{\phi/K_b + \phi_{\text{ft}}/K_g}{\phi + \phi_{\text{ft}}} \right)^{-1}. \quad (67)$$

The parameters λ , $\hat{\phi}_0$ can be obtained by a nonlinear regression of the clean-rock $\{v_p, \phi\}$ data using $\phi_{\text{ft}} = 0$ in the above formulation (equation 65), using known values for K_g, ρ_g, ρ_b etc. The critical porosity $\hat{\phi}_0$ is very likely to be in the range 0.4 – 0.45. If a value for Poisson's ratio representative of the average porosity is chosen (e.g. $\nu_m = 0.15$), the dominant unknown in equation (65) is λ , which can be simply regressed for. Note that using an assumed value for ν_m (e.g. $\nu_m = 0.15$), appropriate for the average of data porosity will not yield the correct pure quartz velocity $v_p(0, 0)$ since a pure quartz Poisson's ratio is closer to 0.07.

The slope and intercept terms in eqn (64) emerge naturally from the linearisation of the nonlinear fit to (63) of clean rock data.

We *define* $(1 + g)$ by

$$(1 + g) \equiv \left. \frac{\partial v_p(\phi, \phi_{\text{ft}})}{\partial \phi} \right|_{\phi=\bar{\phi}, \phi_{\text{ft}}=0} / \left. \frac{\partial v_p(\phi, \phi_{\text{ft}})}{\partial \phi_{\text{ft}}} \right|_{\phi=\bar{\phi}, \phi_{\text{ft}}=0} \quad (68)$$

and sufficiently accurate numerical values for the derivatives are most simply obtained by finite differences.

8.2 Recommended methodology

A possible regression methodology is as follows. For clean rock data,

- Assume $K_g, K_b, \rho_g, \rho_b, \nu_m$
- Compute the mean porosity $\bar{\phi}$
- Fit the v_p data for the non-linear parameters $\lambda, \hat{\phi}_0$
- Compute the $b_p, b_p(1+g)$ coefficients of eqn (64) from the derivative expressions in the Appendix at the mean porosity. Then $a_p = v_p(\bar{\phi}, 0) - b_p(1+g)\bar{\phi}$.
- For this clean rock data, here are no ϕ_{flt} terms in (63) and (64), so fit the loading-behaviour data to get the coefficients a_ϕ, b_ϕ and residuals RMS $\sigma_{\epsilon_\phi} = \sqrt{\langle \epsilon_\phi^2 \rangle}$
- Fit the linearised velocity trend data to get the coefficients a_p, b_p , and residuals RMS $\sigma_{\epsilon_p} = \sqrt{\langle \epsilon_p^2 \rangle}$

Now, for each group of float-polluted well data¹⁶ (i.e. which can be modelled with an common effective ϕ_{flt})

- Refit the two linearised trends, using different intercepts:

$$\phi = a'_\phi + b_\phi Z_{\text{eff}} + \epsilon_\phi \quad (69)$$

$$v_p = a'_p + b_p(1+g)\phi + \epsilon_p \quad (70)$$

We then compute the common float fraction and capture ratio:

$$\phi_{\text{flt}} = (a'_p - a_p)/b_p \quad (71)$$

$$f_c = 1 - \phi_{\text{flt}}/(a_\phi - a'_\phi) \quad (72)$$

A simple check of the theory is to confirm the new regressions have the same slope as the clean rock data.

8.3 Conversion to Delivery formats

For use in *Delivery*, augmented regression relationships of the form

$$v_p = A_p + B_p d + C_p \text{LFIV} + D_p \phi_{\text{flt}} + \epsilon'_p \quad (73)$$

$$\phi = A_\phi + B_\phi v_p + C_\phi \phi_{\text{flt}} + \epsilon'_\phi \quad (74)$$

will be used, using the existing notation in *Delivery* documents. The shear relation is unchanged. The pressure dependence will be effected by taking the

¹⁶Regressions for the data with floating grains can be performed using EM (Expectation-Maximisation) techniques, a possible future development

loading depth $d \rightarrow Z_{\text{eff}}$ (e.g. $(1 - \exp(-P/P_0))$) computed from e.g. basin modelling (use the `depth_rock_curves` entry in the XML, to distinguish from model depth). The LFIV will be dropped by setting $C_p = 0$. Approximate conversion to this *Delivery*-style form of the coupled regression can be obtained with

$$A_p = a_p + b_p(1 + g)a_\phi \quad (75)$$

$$B_p = b_p b_\phi (1 + g) \quad (76)$$

$$D_p = b_p \left(1 - \frac{1 + g}{1 - f_c}\right) \quad (77)$$

$$A_\phi = -\frac{a_p}{b_p(1 + g)} \quad (78)$$

$$B_\phi = \frac{1}{b_p(1 + g)} \quad (79)$$

$$C_\phi = -\frac{1}{1 + g} \quad (80)$$

$$\sigma'_{\epsilon_p} = \sqrt{\sigma_{\epsilon_p}^2 + (b_p(1 + g))^2 \sigma_{\epsilon_\phi}^2} \quad (81)$$

$$\sigma'_{\epsilon_\phi} = \frac{\sigma_{\epsilon_p}}{b_p(1 + g)} \quad (82)$$

Numerical Example

8.4 Rock trends

We use the loading depth term $Z_{\text{eff}} = (1 - \exp(-P/800\text{PSI}))$ for the loading regressions, which is input to *Delivery* as the parameter `depth_rock_curves`. At a particular location of interest, $Z_{\text{eff}} = 0.976$ at $P = 3000$ PSI, roughly around the depth 11000ft.

Modelling loading behaviour in a certain petroleum province yields the following estimates for shales and sands.

8.4.1 Shales

The loading behaviour for non-reservoir rocks (shales) has also to be converted to use the Z_{eff} variable. For simplicity, we took typical velocities from the existing depth-based trends, then set $B_p = 0$ to yield

$$v_p = 10500 \pm 470\text{ft/s} \quad (83)$$

$$\rho_m = 0.278v_p^{0.234} \pm 0.027\text{gm/cc} \quad (84)$$

$$v_s = -2333 + 0.678v_p \pm 280\text{ft/s} \quad (85)$$

8.4.2 Sands

The fits here are:

$$\phi = 1.10 - 0.88(1 - \exp(-P/800\text{PSI})) - 1.54\phi_{\text{ft}} \pm 0.0024 \quad (86)$$

$$v_p = 18717 - 32100\phi - 31060\phi_{\text{ft}} \pm 750\text{ft/s} \quad (87)$$

$$v_s = -2774 + 0.77v_p \pm 240\text{ft/s} \quad (88)$$

The coefficients we require are read off as $a_p = 18717$, $b_p = -31060$, $a_\phi = 1.1$, $b_\phi = -0.88$, $f_C = 0.33$ and $(1 + g) = 1.033$. Coefficients for use in *Delivery* are computed from equations (75)–(82) as $A_p = -16576$, $B_p = 28235$, $D_p = 16830$, $A_\phi = 0.583$, $B_\phi = -3.11 \times 10^{-5}$, $C_\phi = -0.968$. The shear coefficients are $A_s = -2774$, $B_s = 0.77$.

From the estimated noise RMS values $\sigma_{\epsilon_\phi} = 750$ and $\sigma_{\epsilon_p} = 0.0024$ we compute $\sigma'_{\epsilon_p} \approx 700\text{ft/s}$ and $\sigma'_{\epsilon_\phi} \approx 0.02$ as working guesses.

8.5 Model system

Consider a 2-layer shale–seal:sand system, where the sand layer is initially modelled as clean ($NG \sim N(1, 0)$). The well logs used to construct the prior have some clean rocks (no floating grains) and rocks with floating content of around 5%. To model the inferability of the float fraction, we construct synthetic seismic truth–case stacks for near and far stack angles of a few degrees and about 30° , using a truth case model with 5% float, and all other parameters at the most likely values from the trends. Fig 9 illustrates the system, with truth-case plus posterior near and far synthetics from the posterior of case iii) we describe shortly.

For inversion, the prior on floating fraction is taken as $N(0, 0.05^2)$, and we attempt to compute the posterior floating fraction from 3 cases: i) NG fixed at 1, near–stack only, ii) ($NG \sim N(1, 0.2^2)$), near–stack only, and iii) ($NG \sim N(1, 0.2^2)$), near and far stacks.

For case i), figs. 10–12 shows 3 possible forms of the prior – varied for illustrative purposes – on which we superpose the original log data, which illustrates how the floating grain effect smears out the regional prior. The actual prior used in the inversion cases coincides with the inset (c).

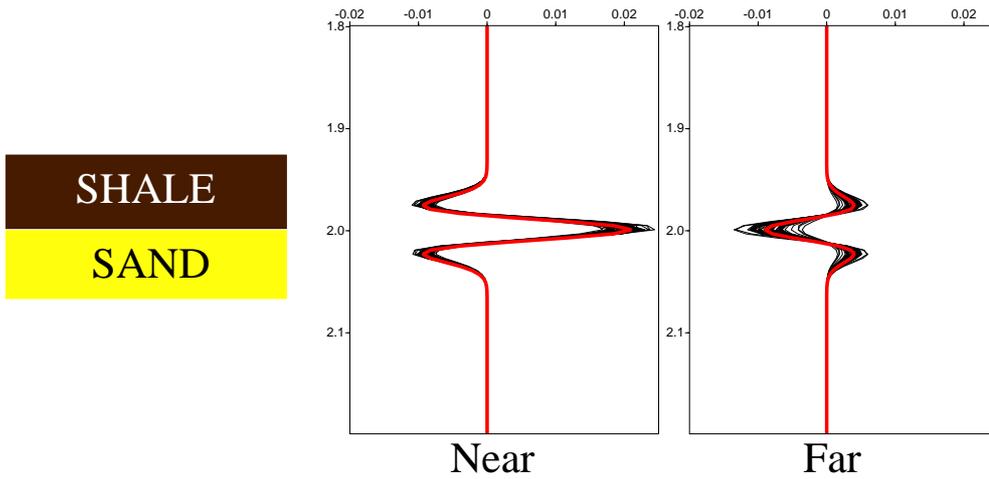


Fig. 9. Two-layer model system with truth case (thick red) seismic traces and synthetics from the posterior (black) for the inversion case iii) described in the text. The absolute noise level is set at 0.0025 for both stacks.

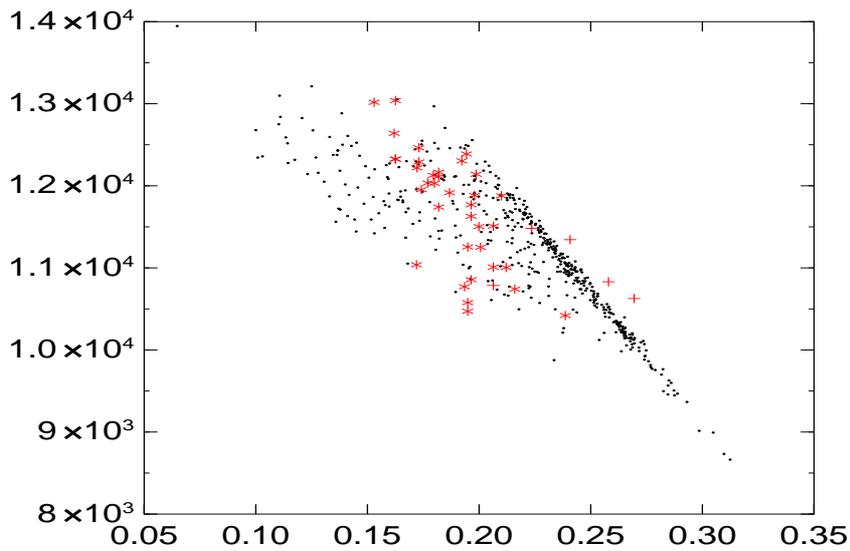


Fig. 10. This and figs. 11 and 12: three 'pedagogic' priors for velocity versus porosity in pure sand layer. Clean well data points (+) and float-polluted well data points (*) plotted on all three graphs: dots (.) are draws from model prior. (a) Prior constructed with artificially narrow $\sigma_\phi = 0.002$, showing how parameters arise from a clear 50:50 mixture of 'clean-rock regression' points and an elliptical smear from the effects of the floating grains. The clean rock trend is obviously far too narrow to embrace the clean well (+) measurements, but the 'clean trend' is clearly visible and centred.

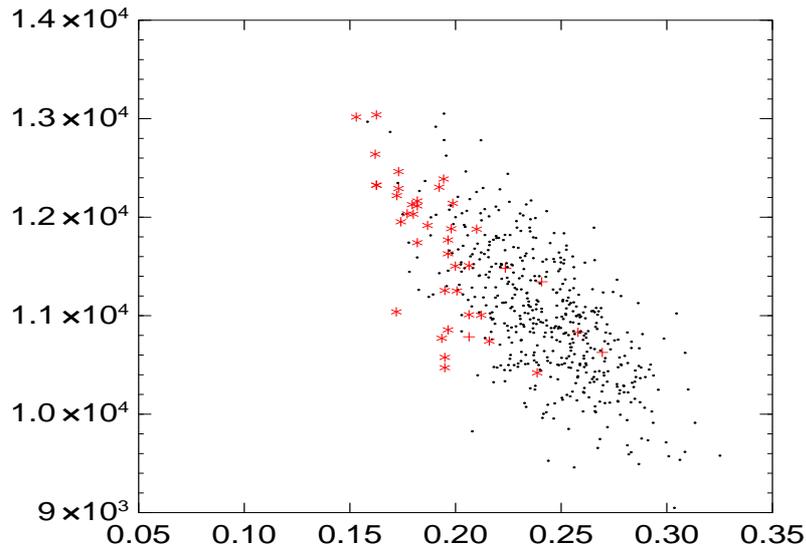


Fig. 11. Illustrative prior (b): prior drawn from clean rocks only (float-fraction $\sim N(0, 0)$), with realistic $\sigma_\phi = 0.02$. The tails of the distribution do not contain the floating grain data comfortably. Clean well data points (+) and float-polluted well data points (*), dots (.) are draws from model prior.

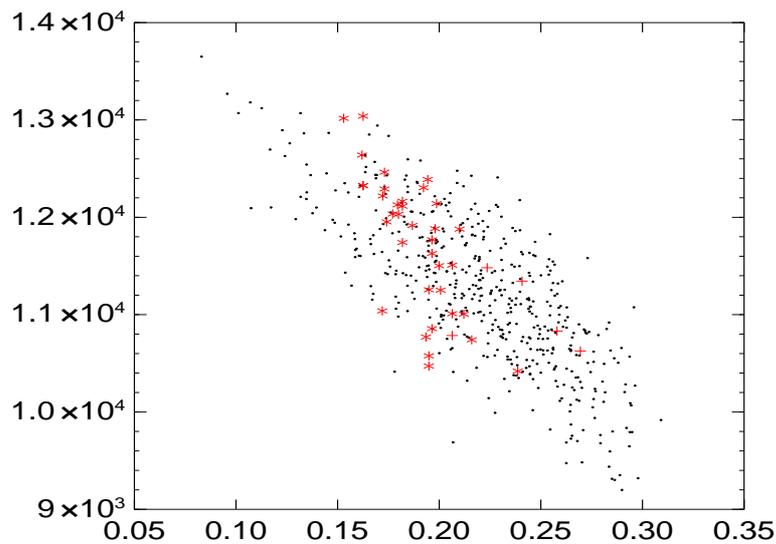


Fig. 12. Illustrative prior (c): actual prior with float-fraction distributed as $\sim N(0, 0.05^2)$. Only a few measurements appear to lie at the periphery of the distribution, but the mixture character is not as clear to the eye as in fig. 10. Clean well data points (+), float-polluted well data points (*), dots (.) are draws from model prior.

8.6 Inversion Analysis of Posteriors

8.6.1 Fixed NG, single stack

Figure 13 shows salient scatterplots of properties of the sand layer before and after inversion. The peak-signal to noise ratio is set at about 6:1.

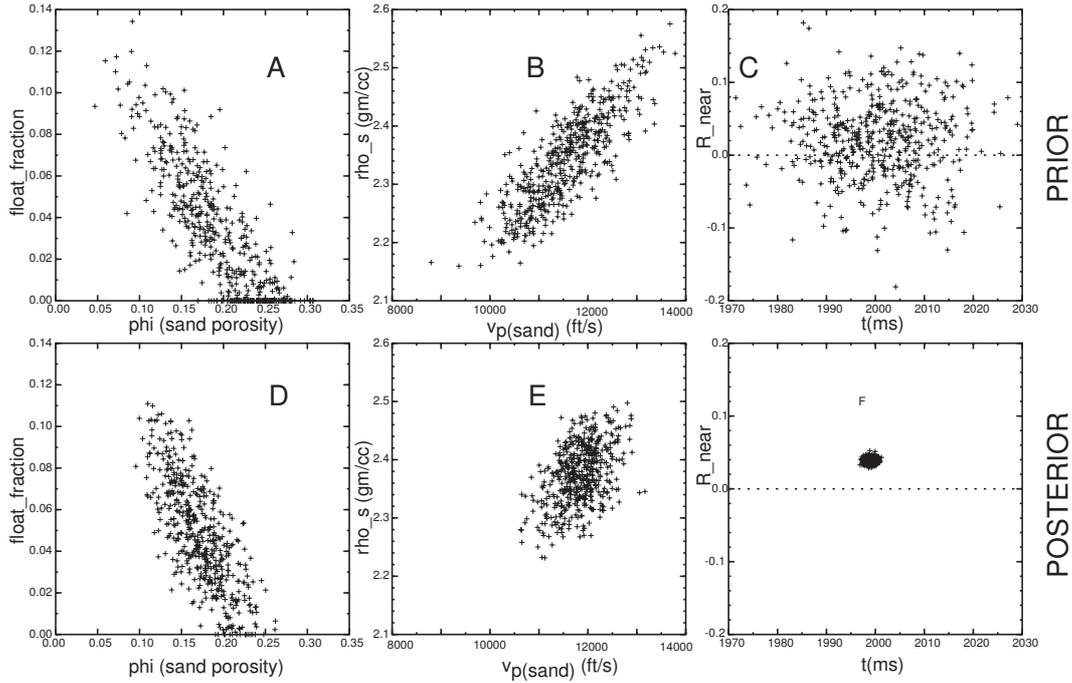


Fig. 13. Scatterplots from prior distribution (top row) and posterior (bottom row) for 2-layer model with sand below shale. (A,D) float_fraction vs sand porosity. A perceptible narrowing around the true answer of float_fraction=5% is visible. Fewer clean sands are produced in the posterior. (B,E) The velocity (v_p) density scatterplot narrows more obviously. (C,F) The layer-time (t) versus effective reflection coefficient R_{eff} is clearly pinned down sharply. As usual, these parameters are most heavily constrained.

8.6.2 Free NG, single stack

Figure 14 shows salient scatterplots of properties of the sand layer after inversion, where the model has additional net-to-gross freedom free in the prior $\text{NG} \sim N(1, 0.2^2)$.

8.6.3 Fixed NG, near+far stack

Inversion using shear data may help narrow down floating grain porosities better, as the shear carries additional information. The far stack for this test case is set at about 30 degrees (c.f. a few degrees for the near) and the reflected amplitude is much weaker (AVO effects). The noise level was set at the same value as for the far stack.

Figure 15 shows comparable scatterplots to 14.

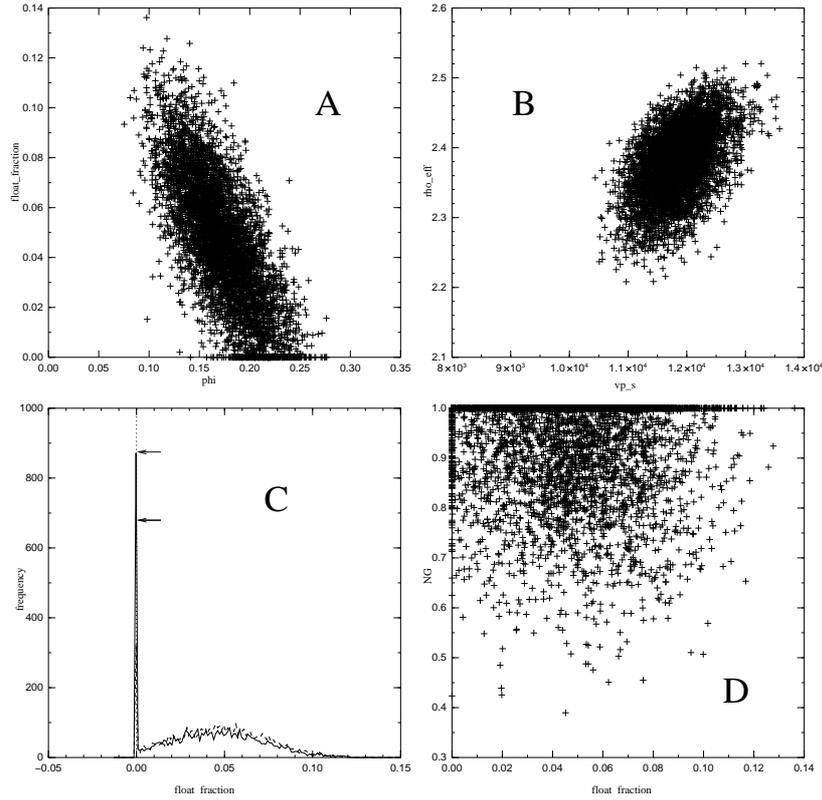


Fig. 14. Posteriors from model with loosened NG distribution in prior. (A) float–fraction vs sand porosity. (B) The velocity (v_p) vs density. (C) Histogram of float–fraction from the posteriors of the models with fixed $NG = 1$ and free $NG \sim N(1, 0.2)$. No obvious effect on the univariate float–fraction distribution is caused by allowing the NG to become free: the peaks of the two delta functions are shown with arrows. (D) Scatterplot of NG vs float–fraction. Again, no obvious strong correlation appears here, with the density strongest near the truth case values (0.05, 1.0).

8.7 Delivery Implementation details

Recall that Delivery works with two versions of the model vector \mathbf{m} . The vector \mathbf{m} has a fully Gaussian prior, with no truncations or restrictions on values. The physical model vector \mathbf{m}' , which is used in the forward model and its associated likelihoods (seismic, isopachs) is obtained by applying time orderings and truncations (e.g. of NG or saturations) to \mathbf{m} , i.e. $\mathbf{m}' = f(\mathbf{m})$, for some suitable f . The truncation effectively induces a prior which, for simple properties like NG, is a mixture of a truncated Gaussian distribution and delta functions at endpoints.

With the augmented models defined by equations (73) and (74), the linearity means the prior is still Gaussian, but the truncation of ϕ_{ft} in \mathbf{m}' must be handled with care. The extra coefficients D_p, C_ϕ have the effect of placing the prior on inclined ellipsoids in e.g. the $\{v_p, \phi_{\text{ft}}\}$ plane, so pure truncation on ϕ_{ft}

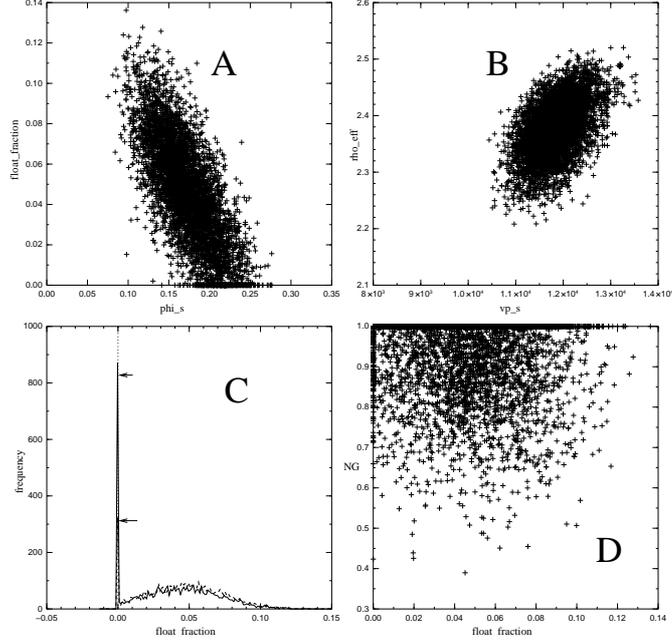


Fig. 15. Scatterplots as per figure 14. The Posterior is more concentrated on the solution $NG = 1$ (D), and similarly has less weight on the clean-rock solution $\text{float-fraction} = 0$ (C and D) (i.e. note the reduced relative height of the spike in (C)).

has the effect of smearing the tail of the distribution onto the plane $\phi_{\text{ft}} = 0$ in a direction off the principal axes. This is clearly a poor way to handle the prior. Fig 16 shows a scatter plot of points produced from a prior constructed in this naive way, with the obvious artifacts. A more reasonable way to handle the truncation is with the mappings (*only for* $\phi_{\text{ft}} < 0$):

$$v'_p = v_p - D_p \phi_{\text{ft}} \quad (89)$$

$$\phi' = \phi + B_\phi (v'_p - v_p) - C_\phi \phi_{\text{ft}} \quad (90)$$

$$v'_s = v_s + B_s (v'_p - v_p) \quad (91)$$

$$\phi'_{\text{ft}} = 0, \quad (92)$$

which forces the remapping to occur along directions parallel to the principal axes.

This mapping minimises the difference $(\mathbf{m}' - \mathbf{m})^T C_P^{-1} (\mathbf{m}' - \mathbf{m})$, subject to the positivity constraint, which seems a reasonable formulation.

Note that the actual Gassman fluid substitution calculation that occurs later in the forward model uses only the pure fluids (oil, gas etc), as the Gassman-like effect of the floating grain presence is implicitly accounted for by the floating-grain terms in the modified regressions.

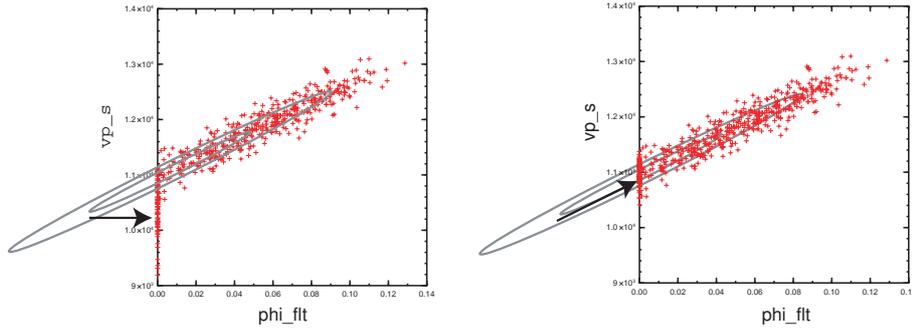


Fig. 16. Left: pure truncation of ϕ_{fit} resulting in smearing of prior along $\phi_{\text{fit}} = 0$ plane. The right figure illustrates the remappings of equations (89)–(92) which seem more reasonable.

8.8 Inputs/Outputs

The prior distribution of the floating grain volume fraction is specified on a layer–basis for the reservoir rock, like, say, the saturation of a hydrocarbon type. It’s default is $N(0,0)$, so the model reduces to the standard model by default. Outputs will be the *modified porosity* $(1 - S_L)\phi_s$ (actual fluid porosity) and the additional field f_L . The SU header word **duse** will encode output file format variations for compatibility with older models (it is rarely used, and defaults to zero).

Appendix 9: Vertical model–stacking functionality

An additional feature added to *Delivery* since the published paper is the capacity to *vertically stack* models, which is a clandestine way to study models where transverse correlations between parameters at different spatial locations can be important, or at least more tightly constrained than the allowable (regression-error) variations in the regional rock curves.

The idea here is that the coupled inversion at two or more separate spatial locations – where there exists significant transverse correlation of layer properties – can be effected by stacking several models on top of one another at a single spatial location, and including coupling terms for the likelihood which connect properties in arbitrarily different layers. The idea is illustrated in figure 17.

The seismic data from the several locations is “stacked” in a similar way. Consider a “two–stack” case for example: a three–layer model with time–window of interest $t \in [1, 1.5]\text{s}$, can be stacked into the total interval $t \in [1, 2.5]$ with location 1 in $[1, 1.5]\text{s}$ and location 2 in $[2, 2.5]\text{s}$. A suitable buffering time interval (in this case $[1.5, 2]\text{s}$) is used between the two packages. Dummy shale packages can also be inserted if desired.

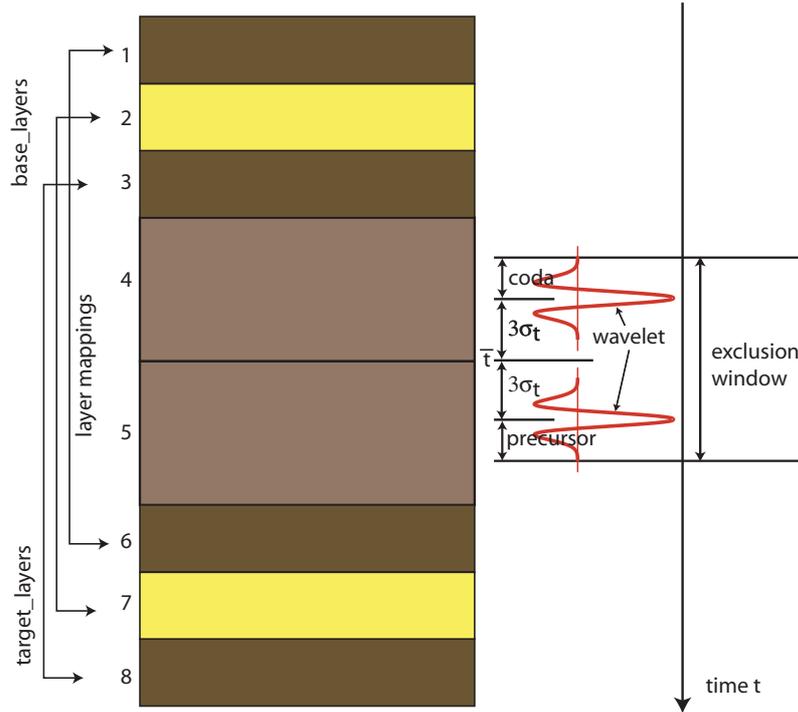


Fig. 17. Illustration of construction of a stacked 8–layer model from two 3–layer models with buffering shale layers. Mappings between comparable layers are shown on the left. To exclude seismic–mismatch coupling effects from the region where the two packages join, a seismic–likelihood exclusion–window is constructed, where the XML specifies the parameters (`start_layer=5`, `end_layer=4`). The constructed exclusion zone starts from a “safe” distance above the *top* of the `start_layer` (i.e. including prior time uncertainties and the effect of the wavelet coda) down to a “safe” time below the *bottom* of the named `end_layer` (using the wavelet precursor and 3 standard deviations of the prior time pick).

The user must specify a set of one or more layer mappings, each of which specifies a *base_layer* and one or more *target_layers* via an xml entry like `<mapping>1:5,6</mapping>` which has `base_layer=1` and `target_layers = 5,6`. These then enable the coupling of properties between the base layer and the target layers. This is a quite general form – in simple two–location models there is only one *target_layer* for each *base_layer*.

Functionality to couple the properties $p \in \{v_p, v_s, \phi_s, \rho_m, \text{NG}, \text{thickness}\}$ exists. For each of these properties, the user provides a between–layers “difference” variance σ_p (*in the same units as the trend curves*), and can (optionally) sub-select the set of *applicable* *base_layers* over which it is desirable to couple the property. If the variance σ is small or comparable to the typical errors in the regional curves, the coupled inversion can be expected to give quite different answers to inversions carried out independently at each spatial location.

For the properties $p \in \{v_p, v_s, \phi_s, \rho_m, \text{NG}\}$, the difference constraint can be absorbed rigorously in the prior. The extra term in the Gaussian prior exponent

(c.f. equation (28), with the fluid-choice subscript k suppressed for simplicity) corresponding to this construction is of form

$$-\log(P(\mathbf{m})) = \frac{1}{2} \sum_{\text{properties } p} \sum_{\substack{\text{applicable} \\ \text{base.layers } i_p}} \sum_{\text{target.layers } j_{i_p}} \frac{(\mathbf{m}_{I(p,i_p)} - \mathbf{m}_{I(p,j_{i_p})})^2}{\sigma_p^2}. \quad (93)$$

where $I(p, j)$ denotes the index of property p , layer j , in \mathbf{m} . The RHS in (93) is a quadratic form which we may write as $\mathbf{m}^T C_{VS}^{-1} \mathbf{m} / 2$, where C_{VS} is a matrix assembled from a sum of block matrix type entries of the general type $\begin{pmatrix} 1 & -1 \\ -1 & 1 \end{pmatrix} / \sigma_p^2$, inserted in the appropriate rows and columns.

The net effect of this term is that the existing prior specified by the rock-physics curves and time-picks can be simply updated using the usual Bayesian formulas to the new Gaussian prior with parameters

$$\begin{aligned} C'_m &= (C_{VS}^{-1} + C_m^{-1})^{-1} \\ \bar{\mathbf{m}}' &= (C_{VS}^{-1} + C_m^{-1})^{-1} C_m^{-1} \bar{\mathbf{m}} \end{aligned} \quad (94)$$

Thus, the prior is the same multi-Gaussian expression as equation (28), but with the modified mean and covariance as per (94). In particular, specification of “vertical stacking” terms for $p \in \{v_p, v_s, \phi_s, \rho_m, \text{NG}\}$, or their modification, does not alter the overall prior probability of the model.

Note that the *thickness* constraint does not contribute a pure quadratic term that can be absorbed rigorously into the Gaussian prior. Though we would like to treat all vertical-stacking requirements as part of the prior (and this is also often true for isopachs), this requires that the normalisation constants are explicitly available, and these are not analytically available for terms that do not appear directly as Gaussian. Hence the vertical-stacking constraint for thickness is imposed via a likelihood term, very similarly to the isopach constraint. This means that changes in the vertical-stacking thickness term *will* alter the underlying probability of a model, just as the isopach and seismic data does. For completeness, we write the likelihood from the thickness difference as

$$L_{\text{VSth}}(\mathbf{m}) = \frac{1}{(2\pi)^{d_{\text{th}}/2} |C_{\text{VSth}}|^{1/2}} \exp\left(-\frac{1}{2} \sum_{\substack{\text{applicable} \\ \text{base.layers } i_p}} \sum_{\text{target.layers } j_i} \frac{(\Delta Z_i(\mathbf{m}) - \Delta Z_{j_i}(\mathbf{m}))^2}{\sigma_{Z_i}^2}\right). \quad (95)$$

where $C_{V\text{sth}} = \text{diag}\{\sigma_{z_i}^2\}$ contains d_{th} entries, one for each term of the double sum in the exponent.

A minor complication with this trick of vertical stacking is that the reflected seismic amplitudes at the interfaces between each “package” of layers will couple the properties between the top and bottom of adjoining packages, which is usually undesirable. In this context, the amplitudes at these times are of no interest, and should be excluded from the inversion. The user has the option of excluding these using one or more xml entries of typical form

```
<seismic_likelihood_exclusion_range>
<start_layer>5</start_layer>
<end_layer>4</end_layer>
</seismic_likelihood_exclusion_range>
```

Each such entry defines a time-window where seismic-mismatch terms are not accumulated in the likelihood equation (25). The window is defined by the interval

$$\begin{aligned} t &> \bar{t}_{\text{start_layer}} - 3\sigma_{t_{\text{start_layer}}} - |\text{wavelet-coda}| \\ t &< \bar{t}_{\text{end_layer}+1} + 3\sigma_{t_{\text{end_layer}}} - |\text{wavelet-precursor}| \end{aligned} \quad (96)$$

where the mean and standard deviations \bar{t}_i and σ_{t_i} are those of the prior picks of layer times, c.f. section 2.2.1. This definition enables the exclusion-window(s) to track the model over space. Figure 17 illustrates how this works, for a 8-layer model.

A typical XML chunk to perform two-location stacking for the 8-layer model shown in this figure would be

```
<stacked_models_info>
  <layer_mappings>
    <mapping>1:5</mapping>
    <mapping>2:6</mapping>
    <mapping>3:7</mapping>
  </layer_mappings>
  <property_difference_std_deviations>
    <vp_m applicable_layers="1,3">350</vp_m>
    <rho_m applicable_layers="1,3">0.02</rho_m>
    <vp_s applicable_layers="2">250</vp_s>
    <phi_s applicable_layers="2">0.1</phi_s>
    <NG applicable_layers="2">0.1</NG>
    <thickness applicable_layers="2">15</thickness>
  </property_difference_std_deviations>
```

```

    <seismic_likelihood_exclusion_range>
      <start_layer>5</start_layer>
      <end_layer>4</end_layer>
    </seismic_likelihood_exclusion_range>
  </stacked_models_info>

```

For convenience, the runtime flag `-IVS` switches off all the inter-layer coupling induced by the terms (93),(95), which is convenient for quick comparisons.

Appendix 10: Internal fluid–contact modelling

A common scenario in inverting is one in which the vertical sequence of possible fluids is known from particular fluid contacts observed in wells or inferred from pressure data. Often the relative uncertainty in the contact depth(s) is not large relative to the layer–depth uncertainties in the inversion, and reflections from the fluid contact(s) can be very strong, so using this additional contact data is very useful.

The default operating mode of the inversion uses the supplied fluid probabilities to construct an enumerated list of fluid combinations for the entire model, where each combination assumes that layers are completely occupied by the nominated fluid type. In situations where a particular fluid contact intersects many different layers across the inversion domain, this style of modelling can be messy to construct, especially if layers are introduced as a clandestine way of capturing the fluid contacts.

As an alternative, a simple fluid–contacts inversion mode is supported, where the fluid configuration of any layer is determined by specified fluid contacts *in depth*, and the computed layer–depth using the model velocities, times, and specified reference depth layer. In this mode, reflections arising from within-layer fluid contrasts are incorporated into the forward model.

In this “contact–mode” inversion, fluid combinations are not enumerated as before, rather, the fluid sequence is deterministically computed from a set of 1 or more “fluid units”, as depicted in figure 18.

Fluid units satisfy the following rules.

- A fluid unit is a set of contiguous layers, and within a fluid unit there may be 1, 2 or 3 of the following contacts {go, gl, gb, ol, ob, lb} (oil (o), gas (g), brine (b), low–saturation gas (l)), under usual full–density ordering.
- The set of fluid units must account for all the layers in the model
- Storage efficiencies require the *total* number of contacts from all fluid units

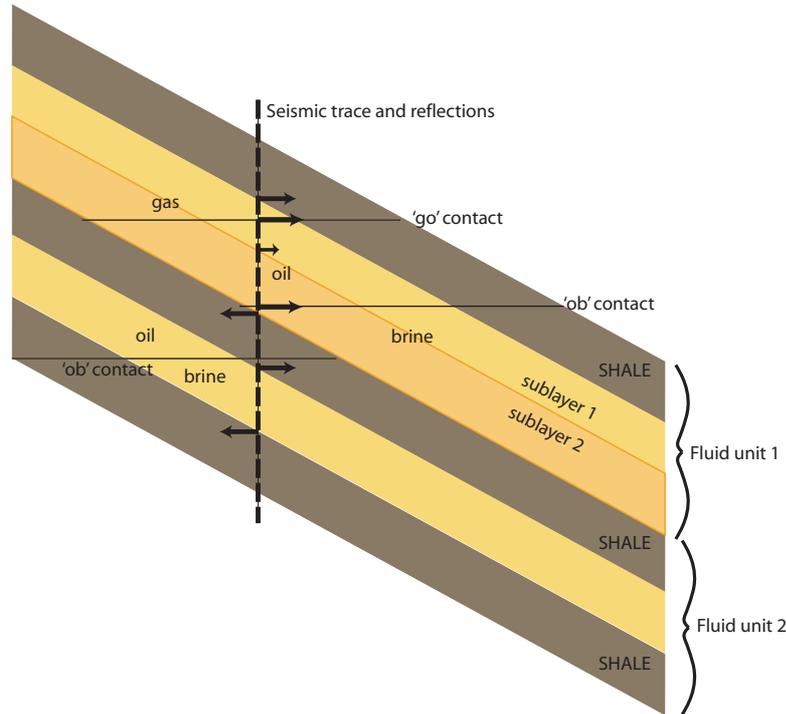


Fig. 18. Illustration of concept of fluid units and depth-specified contacts.

not exceed the total number of layers in the model (in practice a very unlikely restriction).

- Two adjacent fluid units must have an impermeable layer at the bottom of the upper fluid unit or top of the lower fluid unit. This allows unusual fluid sequences between fluid units, analogous to the 'partial-density-ordering' which models layer sequences not in hydraulic communication
- Fluid units are itemised in the XML in descending depth order
- Contacts in each fluid unit specify the upper and lower fluid type, and are modelled by Gaussian distributions $N(Z_c, \sigma_{Z_c}^2)$ whose mean and standard deviation have default values in the XML, but which may be spatially variable via use of the model-prior traces SU file. Contacts contribute a diagonal elements $\sigma_{Z_c}^2$ to C_k in equation (28) with associated mean Z_c inserted in the mean vector $\bar{\mathbf{m}}_k$.
- Any layer which can in-principle host a fluid phase inferred from the fluid units and contacts must specify the fluid in the fluid nameable in each `<reservoir_endmember>` subsection. E.g., if fluid unit 1 identifies a 'bo' contact, all permeable layers spanned by fluid unit 1 must specify the oil name and saturation (the probability field is ignored). The oil name is expected to match throughout the fluid unit.
- The fluid unit XML attribute *fluid_properties_synchronised* can force the stochastic properties of all hydrocarbon properties within a fluid unit to be mapped from a common underlying variable if set to true. This may be useful, if, e.g. similar saturations are expected for, say, gas, in all reservoir layers in the fluid unit. If synchronisation is activated, the XML+model_traces

specification of the *uppermost active reservoir layer*¹⁷ controls the prior distribution of the fluid properties, even if this layer is above a contact excluding the fluid.

It is important to note that *the supplied fluid probabilities are ignored* when fluid occupations are determined using contacts in this mode.

10.1 Upscaling rules

The code populates a variety of layer-effective fields in the output realisations-style SU files, so upscaling rules must be used when there are internal contacts. We use the following:

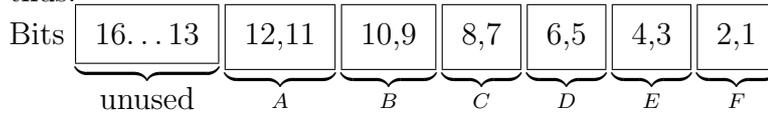
- velocities **vp_eff** and **vs_eff** are time-thickness weighted averages of sublayer velocities, so subsequent multiplication by the overall layer time-thickness gives the true layer thickness, i.e. equal to the sum of the thicknesses of the sublayers delimited by contact depths and the layer top/bottom depth.
- density **rho_eff** and saturation **S_h** are thickness weighted averages (the latter lumping all non-brine fluids), which should give reasonable results for forward calculation of rock mass and hydrocarbon in-place. For mixed hydrocarbon phases, the fields **S_oil**, **S_lsg**, and **S_gas** are probably more useful. These last three are phase-specific thickness weighted averages.

10.2 New outputs in the realisations SU file

- Reflection coefficients **R_near_fc**, **R_far_fc**, which are near and far stack reflection coefficients arising *only* from fluid contact reflections occurring within layers.
- Times **t_fc** of these reflection coefficients, in the same order. No association of these times with a particular layer, via the layer-slot in the SU file from which they are retrieved, can be made. The layer “owning” the reflections must be determined by comparison with the usual “t” entries. Times are computed and written even for zero-reflections occurring within purely impermeable layers.
- Depths **d_fc** of these reflection coefficients, in the same order.
- **fluid_type** is interpreted as an 2 byte integer (cast from the float value in the SU trace), with legacy values 0=brine, 1=oil, 2=lsg, 3=gas. These values require only 2 bits of information. For layers split by contact(s), the spare

¹⁷ A reservoir layer is defined by a layer with $\sigma_{NG} > 0$ or $\overline{NG} > 0$. Active layers are defined in Appendix 5.1. Note this can vary spatially, so the safest (and least ambiguously documented) option is setting the priors identically in all reservoir layers.

bits of the integer are encoded to communicate the sub-layer fluid sequence thus:



- A = number of internal reflections (= number of sublayers-1)
- B = fluid_type, sublayer 4, if A>2
- C = fluid_type, sublayer 3, if A>1
- D = fluid_type, sublayer 2, if A>0
- E = fluid_type, sublayer 1, if A>0
- F = fluid_type of thickest sublayer ("effective" fluid type) (97)

This encoding reduces to the legacy 0,1,2,3 values for layers not split by contacts.

- **S_oil**, **S_lsg**, and **S_gas**. These "effective" saturation components conserve correct hydrocarbon-phase volume if multiplied by the layer thickness, in cases where internal contacts exist. For example, if a oil-brine contact bisects a layer (in depth), **S_oil** is discounted by 1/2. For contact-free layers, their values are sensible and equal the legacy **S_h** field, for the appropriate phase.
- The *Analysr* program can now generate **net-oil**, **net-lsg** and **net-gas**, which are correct special cases of **net-hydrocarbon**, using the **S_oil**, **S_lsg**, and **S_gas** fields described above

Reconstruction of some fraction of the internal detail of the layers is then possible using these new or modified fields.

10.3 Example Inversion

Consider the simple 3 layer test problem of figure 19 (see the *Tests/InverterTests/Simple3LayerContactsTestProblem* directory in the distribution). The reservoir conditions here are typical of north west shelf Australia, with 30% porosity sands at about 2km TVD. Agnostically, the reference depth layer is at the top of the model, with a depth standard deviation of 20m.

The 2 contacts are set up with an XML section thus

```
<fluid_units_and_contacts>
  <fluid_unit fluid_properties_synchronised="false">
    <top_layer>1</top_layer>
    <bottom_layer>3</bottom_layer>
    <contact>
      <type>G0</type>
      <depth_varies_areally="false">2050</depth>
    </contact>
  </fluid_unit>
</fluid_units_and_contacts>
```

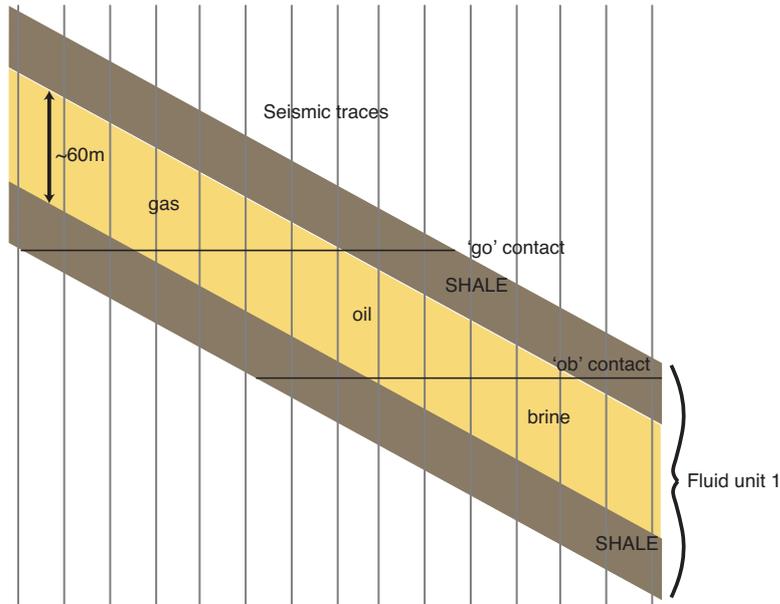


Fig. 19. Simple 3 layer, single fluid unit, 2 contact problem: gas over oil over brine

```

    <sigma_depth varies_areally="false">0</sigma_depth>
  </contact>
  <contact>
    <type>OB</type>
    <depth varies_areally="false">2120</depth>
    <sigma_depth varies_areally="false">0</sigma_depth>
  </contact>
</fluid_unit>
</fluid_units_and_contacts>

```

Fig. 20 shows synthetics for this model, where the Gassmann fluid effect is very strong: indeed, fluid-contrasts are stronger than shale-sand contrasts.

Fig. 21 shows typical realisations from the inversion of this model (which includes an isopach constraint of $60 \pm 10\text{m}$) on the reservoir. The brighter reflections, including fluid contrasts, yield sharper layer and internal contact times. In the brine leg, the dimmer amplitudes produce greater uncertainties. The peak-signal to RMS noise ratio is about 14 in the gas, diminishing to about 2-3 in the brine leg. A fascinating aspect of the use of contacts is that the depth uncertainty of layers is seriously reduced when a visible contact occurs in the trace, as per the lower graph in fig.21. Here, the bright peaks sharply define the reflectivities (thus velocities) and reflection times, thus the layer thickness can be better nailed, and all that is needed is a known event in depth upon which to hang the model. Notice the top depth of layer 1 is appreciably better pinned by this mechanism, even though the reflection occurs in the layer below.

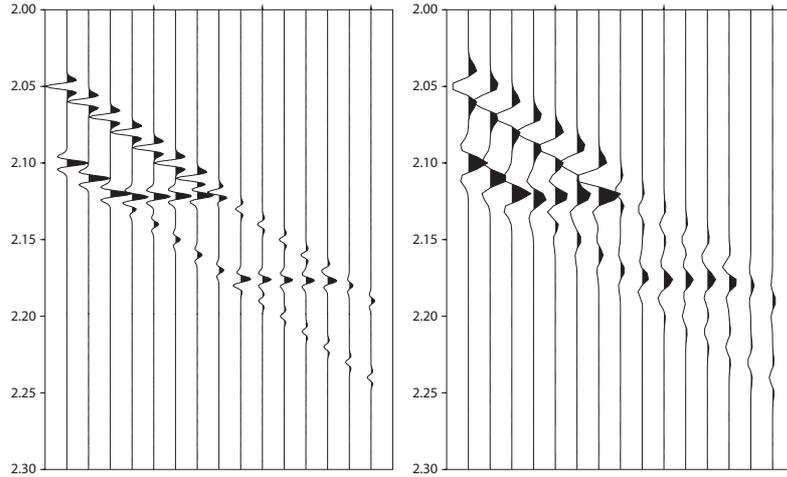


Fig. 20. Synthetic seismic data for the model of fig. 19 at 200Hz (unrealistic) and 80Hz (realistic) bandwidth.

10.4 Likely Issues

The simple fluid contact model described here uses the usual optimisation machinery which enumerates multiple modes by accruing local-optimisation solutions started from an assortment of dispersed starting points. Multimodality associated with loop skipping will still occur, and can be expected at traces where internal fluid contacts are close to layer boundaries. Tunnelling-like behaviour of the MCMC sampler is not unlikely if the modes are very different in curvature and the mode-jumping strategies of section 4.3 are not well-adapted. Some of this character is already evident in the brine leg of the example (fig. 21). Careful examination of sample outputs for good mixing is encouraged.

Another not-yet thoroughly tested issue is that the posterior surface will have jump discontinuities associated with contacts pinching out, analogous to the discontinuities arising from layer pinchouts. The latter are dealt with via quadratic-programming (QP) techniques tacked on to the Gauss-Newton schemes, but this strategy is much harder for internal fluid-contact discontinuities, as the fluid contact is not specified in time. Possible linearised QP strategies are under consideration.

Appendix 11: Model selection facilities

It is often desirable to compare the relative statistical significance of several candidate models $k = 1 \dots N$, where higher level basic assumptions in the model may vary. The existing apparatus implicitly performs model comparison

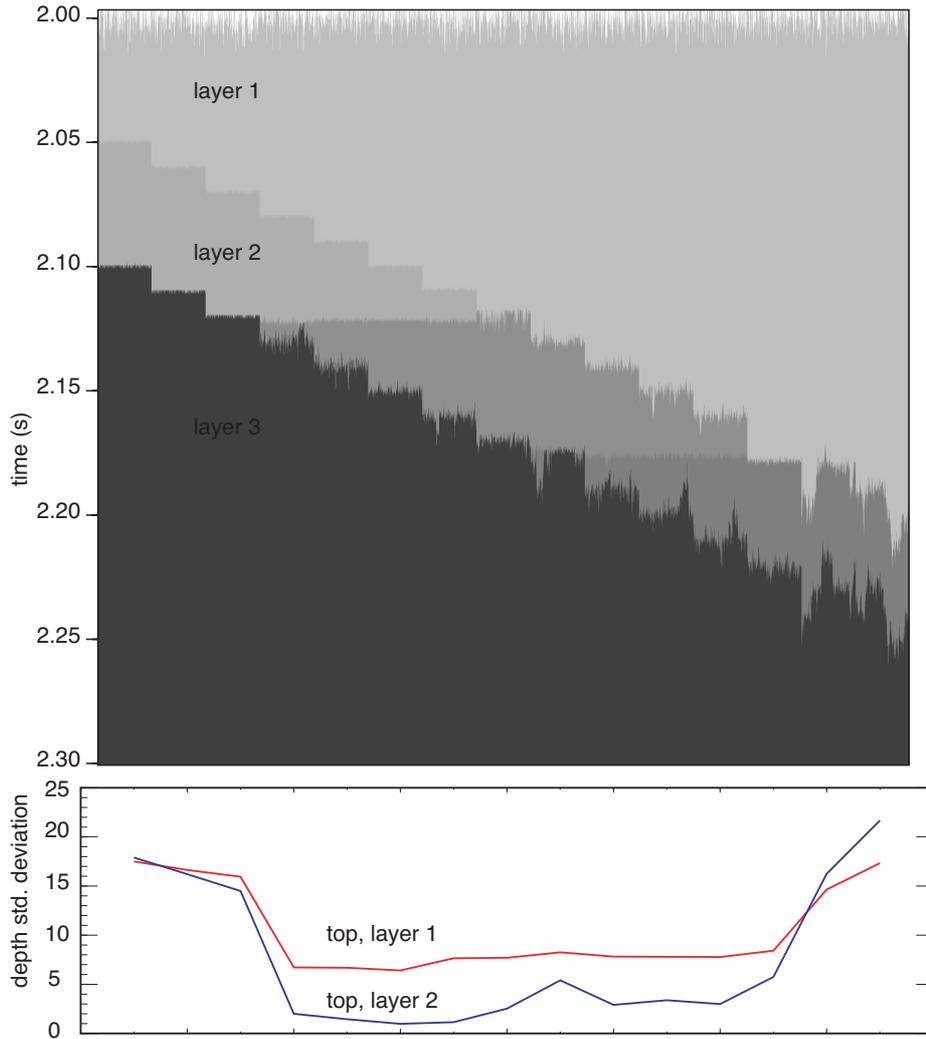


Fig. 21. Typical inversion realisations showing internal contacts well resolved in general. The lower graph is depth uncertainty, which reduces markedly when a contact is visible.

in a single run, where the changes involve different fluid combinations k , but higher level model changes will need to be assessed by comparing outputs from separate runs of the code. Henceforth, we use k to label a general model choice.

The fundamental entity in Bayesian model selection strategies (Gilks et al., 1996), (Denison et al., 2002) is the *marginal model likelihood*, obtained by the marginalisation integral

$$\Pi(k|S) = \int \Pi(\mathbf{m}_k|S) d\mathbf{m}_k, \quad (98)$$

using the notation of equation (27), and interpreting S liberally to mean all the “data”: seismic, isopachs etc. The priors $P(\mathbf{m}_k)$ and likelihoods used to construct the posterior $\Pi(\mathbf{m}_k|S)$ must be normalised correctly.

Between any two models k and j , the marginal likelihoods can be used to construct the Bayes Factor

$$B_{kj} = \frac{\Pi(k|S)}{\Pi(j|S)}$$

which is usefully interpreted as the posterior odds ratio of model k to model j . More generally, among $k = 1 \dots N$ models with prior probability p_k , the posterior probability of model k is

$$P(k|S) = \frac{\Pi(k|S)p_k}{\sum_{j=1}^N \Pi(j|S)p_j}.$$

The chief challenge is performing the integral in (98). A first useful approximation is the Laplace approximation (Raftery, 1996), based on a quadratic expansion about the MAP point $\tilde{\mathbf{m}}$ obtained at the termination of the Newton optimisation scheme (omitting the label k where context permits):

$$\Pi(\mathbf{m}_k|S)p_k = e^{-\chi^2(\tilde{\mathbf{m}})} \exp(-(\mathbf{m} - \tilde{\mathbf{m}})^T \tilde{C}_m^{-1}(\mathbf{m} - \tilde{\mathbf{m}})/2). \quad (99)$$

Here $\chi^2(\tilde{\mathbf{m}})$ is the minimum of the objective function used in the optimiser, evaluated at the termination point $\tilde{\mathbf{m}}$. For completeness, we define this as

$$\chi^2(\mathbf{m}) = \chi_{\text{full-seismic}}^2(\mathbf{m}) + \chi_{\text{iso}}^2(\mathbf{m}) + \chi_{\text{VSth}}^2(\mathbf{m}) + \chi_{\text{prior}}^2(\mathbf{m}) + \chi_{\text{fluid-prior}}^2 \quad (100)$$

where

$$\begin{aligned} \chi_{\text{full-seismic}}^2(\mathbf{m}) &= \frac{1}{2} \sum_{\substack{\text{stacks } i \\ \text{error-sampling points } j}} \left\{ \frac{f_{\text{error}}(\mathbf{S}_{\text{syn},j}(\mathbf{m}) - \mathbf{S}_{ij})^2}{\sigma_{s,i}^2} + [\log 2\pi + \log(\sigma_{s,i}^2/f_{\text{error}})] \right\} \\ \chi_{\text{iso}}^2(\mathbf{m}) &= \frac{1}{2} \sum_{\text{isopachs } j} \frac{(v_{j,p,\text{eff}}(t_j - t_{j-1}) - \Delta Z_j)^2}{\sigma_{\Delta Z_j}^2} + [\log 2\pi + \log(\sigma_{\Delta Z_j}^2)] \\ \chi_{\text{VSth}}^2(\mathbf{m}) &= \frac{1}{2} \sum_{\substack{\text{applicable} \\ \text{base.layers } i_p, \\ \text{target.layers } j_i}} \left\{ \frac{(\Delta Z_i(\mathbf{m}) - \Delta Z_{j_i}(\mathbf{m}))^2}{\sigma_{Z_i}^2} + [\log 2\pi + \log(\sigma_{Z_i}^2)] \right\} \end{aligned}$$

$$\begin{aligned}\chi_{\text{prior}}^2(\mathbf{m}) &= \frac{1}{2}(\mathbf{m} - \bar{\mathbf{m}}')^T C'_m{}^{-1}(\mathbf{m} - \bar{\mathbf{m}}') + \left[\frac{d_m}{2} \log 2\pi + \frac{1}{2} \log(|C'_m|)\right] \\ \chi_{\text{fluid-prior}}^2 &= -\log(p_k)\end{aligned}\tag{101}$$

The likelihood and priors are exactly as per the original equations (25),(26),(28) (with (94), and (95) if vertical stacking is used), but with the normalisation constants (terms in $[\]$ brackets) explicitly delared. The $\log 2\pi$ terms matter since models of different dimensionality may be compared. All non-thickness terms from vertical stacking, and/or fluid-contact terms (if contacts are part of the model), are explicitly absorbed in the multi-Gaussian prior. If there is no vertical stacking, we drop the primes on C'_m , $\bar{\mathbf{m}}'$, and omit χ_{VSth}^2 .

The integration for the marginal model likelihood in equation (98), using this approximation, is then straightforward, and yields

$$\Pi(k|S)p_k = (2\pi)^{d_m/2} |\tilde{C}_m|^{1/2} e^{-\chi^2(\tilde{\mathbf{m}})}\tag{102}$$

The quantity we believe is of most practical use to users is

$$\begin{aligned}-\log(\text{MML}) &\equiv -\log \Pi(k|S) \\ &\approx \chi_{\text{full-seismic}}^2(\tilde{\mathbf{m}}) + \chi_{\text{iso}}^2(\tilde{\mathbf{m}}) + \chi_{\text{VSth}}^2(\tilde{\mathbf{m}}) + \chi_{\text{prior}}^2(\tilde{\mathbf{m}}) \\ &\quad - \frac{d_m}{2} \log 2\pi - \frac{1}{2} \log(|\tilde{C}_m|)\end{aligned}\tag{103}$$

since this is independent of the prior model probability p_k , and scores model significance on a useful logarithmic scale. This quantity should be able to be meaningfully compared between different runs of the code, with different modelling assumptions. Ratios of exponentials of $-\log(\text{MML})$ can then be interpreted directly as Bayes factors.

As a simple example, suppose we run model 1, with oil-rich sandstone, using the options `--MAP -v 3`. This will produce a table of *Normalised local modes*, containing an entry for the $-\log(\text{MML})$ term. The *smallest* $-\log(\text{MML})$ entry represents the most likely mode (assuming multiple modes from e.g. fluid combinations or loop-skips), and will coincide usually with the largest mode *weight*. Then we run with an alternative model, say model 2, with soft shale instead of “oil-rich sandstone”. Again, we read off the smallest $-\log(\text{MML})$ entry. The posterior odds ratio of the soft-shale model is then

$$\text{odds ratio} = \exp(\log(\text{MML})_{\text{soft-shale}} / \log(\text{MML})_{\text{oil-rich sandstone}}).$$

In principle, the marginal model likelihood is robust to many changes in the prior. A quick check of its sensibility can be obtained by “switching off” the data terms in a model comparison experiment (seismic and isopachs via `-IS`

-II , and the thickness constraints in vertical stacking): no posterior odds ratio preferences ought to occur in such experiments. Comparisons of models with different data windows (e.g. varying time gates) is in principle legal, but is usually fraught with danger and is not recommended. One may expect very complex models (relative to the Nyquist rate which yields the seismic data count) to be increasingly penalised by this measure, in typical BIC-like behaviour.

Simple comparisons, like the substitution of rock types, or variations in the chosen net-to-gross models, should in general be meaningful. In situations where severe pinchouts occur, the accuracy of the Laplace approximation may be problematic, so careful investigations are recommended. For example, pseudo parabolic cross sections through the mode optima can be dumped to ascii files *PosteriorXsection.** by invoking `--Xsections 2`, which is useful as a quick check of the adequacy of the quadratic approximation.

Appendix 12: MAP point estimates and fluid model probabilities in noisy problems

Bayesian inversion codes are often required to supply “point estimates”, or “most-likely” type values for the parameters they try to infer. When the problem has a model-selection aspect, it is usually tempting to try and provide the most-likely parameters in the model that appears to have the largest posterior probability. This can be quite different to the parameters for the model that best fits the data at a particular point.

Model probabilities are usually found by trying to marginalise/integrate over all the continuous parameters in the problem, yielding a model “evidence”, or marginal model likelihood (MML) (see Appendix 11). This is the underlying machinery in trying to infer fluid types in *Delivery*. The Bayes factor in comparing two models is then usually written as a likelihood-ratio term, with an extra compensation factor reflecting the probability contribution that comes from the width of the posterior, usually called the “Occam factor” (see e.g. Ch 28 of Mackay (2002), and Denison et al. (2002)). In Laplace-like approximations to the MML this Occam factor is identified with the determinant of the mode Hessian (e.g. eqn (102)), so, for example, a model-comparison Bayes-factor for the posterior probabilities of models 1 and 2 might look like

$$B_{1,2} = \underbrace{\frac{(2\pi)^{d_{m,1}/2} |\tilde{C}_{m,1}|^{1/2}}{(2\pi)^{d_{m,2}/2} |\tilde{C}_{m,2}|^{1/2}}}_{\text{Occam factor}} \cdot \underbrace{\frac{e^{-\chi_1^2(\tilde{\mathbf{m}}_1)}}{e^{-\chi_2^2(\tilde{\mathbf{m}}_2)}}}_{\text{Likelihood ratio}}. \quad (104)$$

This “Occam factor” boosts the model probability for models that have diffuse posterior distributions, relative to models that may have better maximum-likelihood points, but narrow posteriors. Thus it is not necessarily true that the model with the smallest χ^2 fit is the most probable model, which may be confusing to some users.

This behaviour is logically consistent, but slightly counter-intuitive, and highlights the problems that can be caused by placing undue significance on particular point estimates. The following examples illustrate what is going on.

12.1 An illustrative toy problem

Consider the problem of Bayesian estimation for fitting a box-car signal to “truth-case” data given by the box-car function of fig.22. The true signal is $B(t)$, which we fit to a two-parameter model $bB(t-\tau)$, so $\{b, \tau\} = \{\text{amplitude,}$

timing–offset}. Naively, we expect the best parameters to be $b = 1$, $\tau = 0$, but some interesting behaviour arises when the noise gets large.

We will assume Gaussian noise, even though the signal in this case is noise-free. We will also use flat Bayesian priors on the parameters b, τ , but we need a finite range on τ , so $-1 < \tau < 1$ seems reasonable.

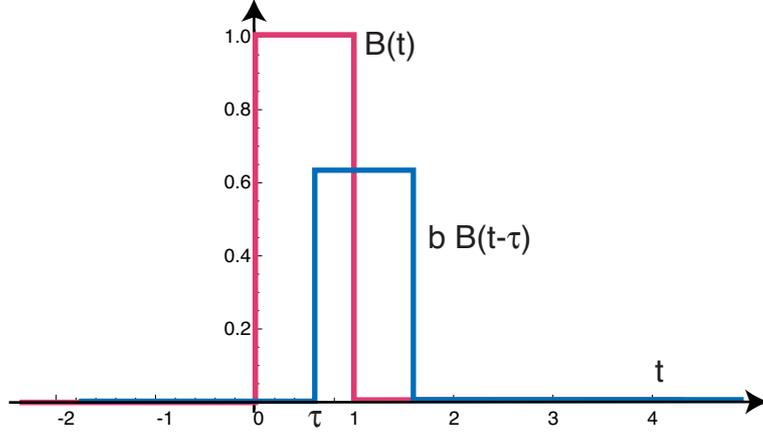


Fig. 22. Boxcar function $B(t)$ and synthetic model $bB(t - \tau)$.

If the noise level is σ , we can define the posterior by

$$\Pi(b, \tau) \sim \exp\left(-\frac{N}{2\sigma^2} \int_{-\infty}^{\infty} (B(t) - bB(t - \tau))^2 dt\right) \quad (105)$$

corresponding to Gaussian noise and an equivalent χ^2 misfit error term

$$\chi^2 = \frac{N}{2\sigma^2} \int_{-\infty}^{\infty} (B(t) - bB(t - \tau))^2 dt.$$

Here, N is something like the “number of equivalent data per unit time”, and, in a practical problem, the integral would be approximated by a sum.

This problem is interesting, because the parameter τ is related to layer–time detection, and b to fluid/saturation detection in *Delivery*. For a situation with several candidate fluid types, b would have a few discrete values, and we might be trying to determine the most likely value of b and τ .

The boxcar problem is trivially simple. We get

$$\chi^2 = \frac{N}{2\sigma^2} \begin{cases} 2b|\tau| + (1 - b)^2 & |\tau| < 1 \\ 1 + b^2 & \text{otherwise} \end{cases}$$

So the joint maximum-likelihood point (minimum of χ^2) is $\{b = 1, \tau = 0\}$, rather obviously. Using this solution for χ^2 , the joint-posterior probability surface of eqn (105) looks like fig. 23.

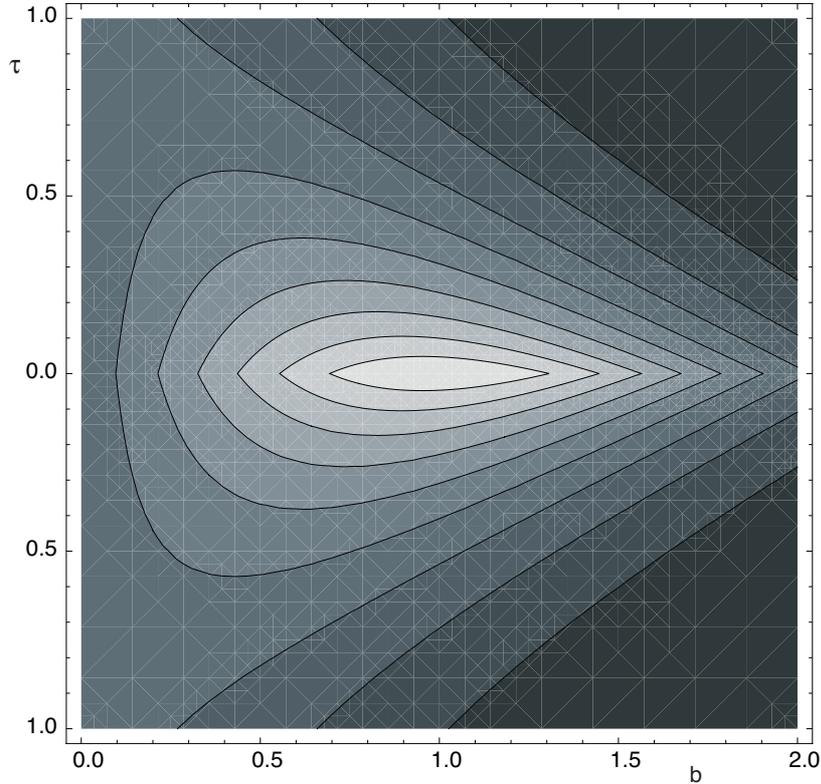


Fig. 23. Joint posterior of boxcar detection problem.

Since we have in mind a problem where b might take on a few discrete values, we might like to find the relative probability of those values by integrating out all the other parameters in the problem. Thus, an interesting quantity to compute is the marginal distribution of b ,

$$\Pi(b) = \int_{-1}^1 \Pi(b, \tau) d\tau \sim \frac{e^{-N(1-b)^2/2\sigma^2} (1 - e^{-Nb/\sigma^2})}{Nb}.$$

For $N = 1.1$, curves of $\Pi(b)$ at various noise levels are shown in fig. 24. One can see from this figure that for low noise levels, the “most-likely” value from the marginal for b is close to 1. However, for high noise levels, it shrinks away from 1 towards smaller values (but remembering that the uncertainty is rather high). This comes from the extra probability mass from larger τ values that are reasonable if the event amplitude is smaller, as per the asymmetry of the joint density in fig. 23. Thus we can see that if the support of b was confined to a few discrete values, the “most-likely” choices of b would fall on weaker signals if the noise level is high.

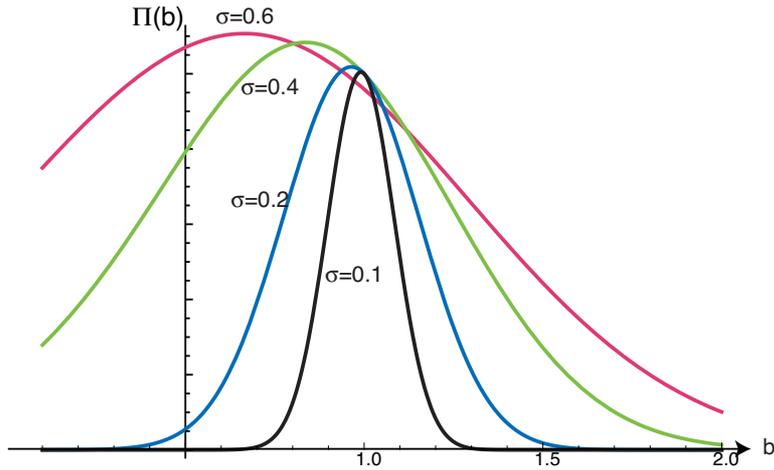


Fig. 24. Marginal distributions of b for various noise levels.

As a concrete example, we might consider two distinct cases $b = 1$ and $b = 0.25$, where, for noise $\sigma = 0.5$ and $N = 1.15$, the difference in χ^2 at the joint MAP point $\chi^2 = \frac{N(1-b)^2}{2\sigma^2}$ is

$$\Delta\chi^2 = \frac{N[(1-1)^2 - (1-0.25)^2]}{2\sigma^2} \approx 1.3$$

Based on this alone one might imagine a $\approx 3 : 1$ model-selection preference for the case $b = 1$. However, the marginal distribution for b puts the odds ratio close to 1.3. Two slices through the joint posterior for $b = 1$ and $b = 0.25$ are shown in fig. 25, showing how the extra probability for the “poorer fit” model comes from the wide posterior.

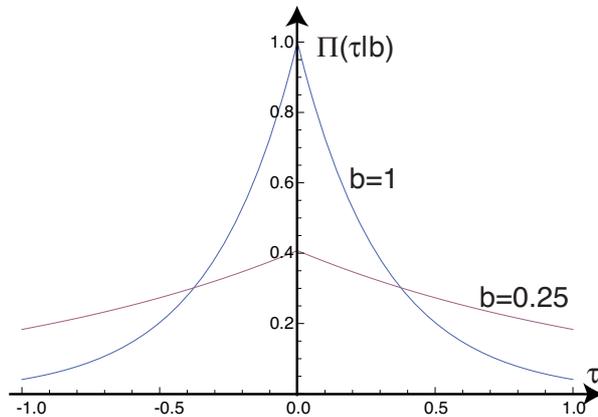


Fig. 25. Conditional distributions $\Pi(\tau|b)$ for $b = 1$, $b = 0.25$ (slices through Fig 23). Both cases have comparable overall probability, even though the maximum likelihood point for the case $b = 1$ seems dominant.

The phenomenon is generic. For example, one can do the calculation for the “stretched box-car” model with a box-car synthetic plateau between $\tau_1 <$

$t < \tau_2$, so free parameters are $\{b, \tau_1, \tau_2\}$. With everything as before, but just including an extra marginalisation over τ_2 as well as τ_1 , the marginal for b is much the same as before: see fig. 26.

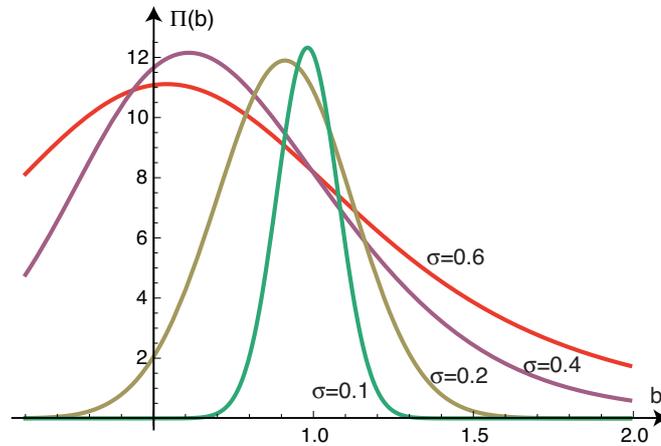


Fig. 26. Marginal distributions of b for various noise levels, for stretched box-car problem (τ_1, τ_2)

A few points are helpful to clarify the counter-intuitive nature of this behaviour. 1) The MML machinery is the most logical approach, granted the assumptions of the noise model. The box-car functions used above are clearly “error-free”, so our instinctive reaction as to what the right criterion should be is formed by a model with very small noise levels – which we automatically intuit when inspecting the data. Had the signal and synthetic been drawn with noise deviates added which are consistent with the noise level in the likelihood function, the MML procedure would seem much less counter-intuitive. Fig. 27 shows a cartoon of the sorts of signals we might be trying to model with, for example, a noise level of $\sigma = 0.5$. 2) The behaviour is a fairly sensitive function of the peak/noise-RMS ratio, and for most signals where the ratio is 4:1 or better, the effect is weak. This is the usual regime we like to work in.

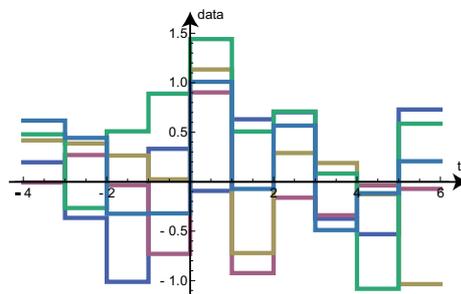


Fig. 27. Example random telegraph signal realisations with boxcar truth case masked by noise of roughly $\sigma = 0.5$

12.2 Demonstration on the Gas–Oil–Flank toy Delivery problem

A practical “walk-through” of the issue on the context of a simple *Delivery* application is outlined in the file `au/csiro/JamesGunning/Tests/InverterTests/GasOilFlank/README_MAP_CLASSIFICATION`, supplied with the *Delivery* source code ¹⁸. The demo problem is a two-layer model of a flank of a hydrocarbon deposit, with shale cap-rock, and a sand halfspace underneath. The “true” fluid changes from brine to oil to gas as we go updip. For reference, the peak seismic amplitude (gas) is about 0.15, and the oil and brine peaks are about 0.05 and 0.03 respectively. With small noise ($\sigma = 0.015$), the MAP model and MAP fluid classification is “correct”, and shown in fig. 28. With the noise at

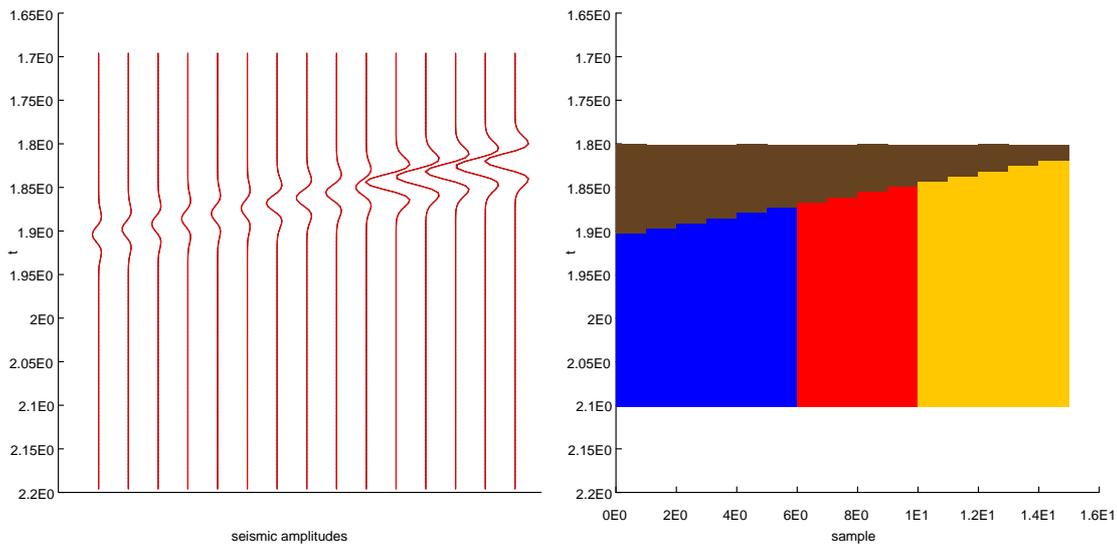


Fig. 28. Gas–oil flank problem. Left: true (red) and synthetic (black) MAP traces for 15 cdp locations, going updip. Right: fluid classification in lower reservoir under shale (brown), for small noise level (peak/noise-rms \approx 10). Fluids codes: blue=brine, red=oil, orange=gas.

$\sigma = 0.08$, the MAP classification misses oil, and puts oil where the gas is: see fig. 29. The maximum likelihood solution, (or minimum χ^2) solution, found using `--ML` instead of `--MAP` makes the visual coherence of the fits closer, as one would expect: see fig. 30. The misclassification of the MAP method is of no particular significance in view of the uncertainty of the fluid type at these noise levels: compare figs 31 and 32.

By way of comparison to the preceding box-car problem, for trace 15, which has gas, at a noise level $\sigma = 0.059$, the $\Delta\chi^2$ between the gas and oil cases is again 1.3 (in favour of gas), but the posterior Bayes factor in favour of gas is only 1.45, less than half of the likelihood ratio $e^{1.3} \approx 3.7$.

¹⁸ The file location is relative to the root of the source tree

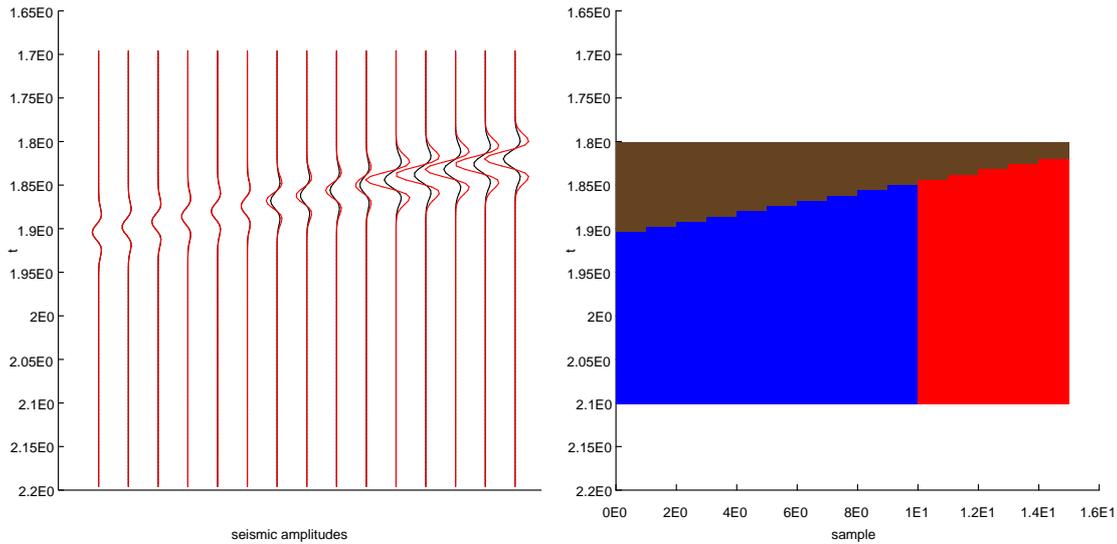


Fig. 29. Gas–oil flank problem. MAP solution and fluid classification, for large noise level $\sigma = 0.08$ (peak/noise–rms ≈ 1.9). True seismic in red, MAP synthetic in black. Fluids codes: blue=brine, red=oil, orange=gas.

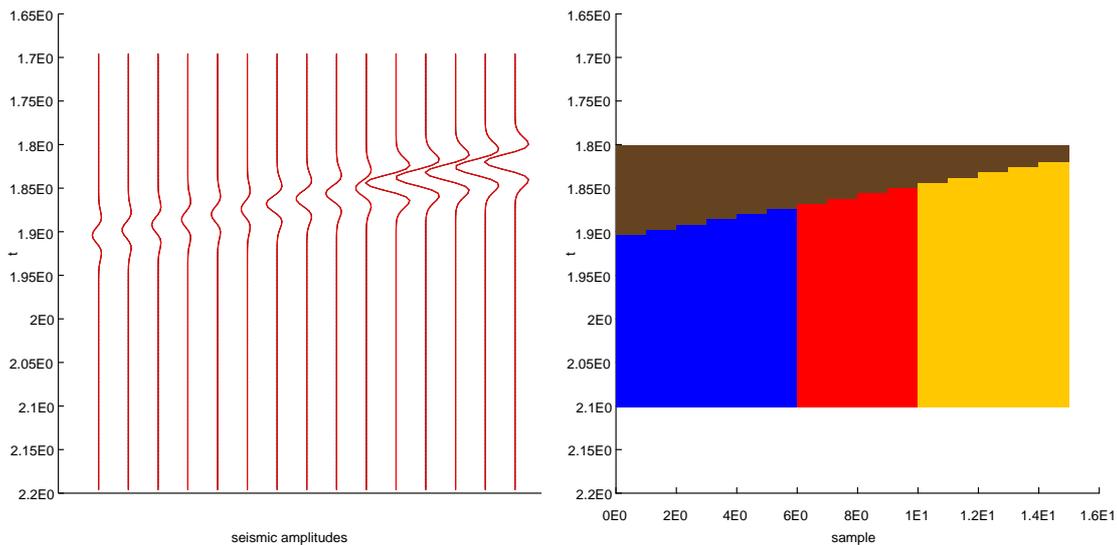


Fig. 30. Gas–oil flank problem. Maximum likelihood solution and fluid classification, for large noise level $\sigma = 0.08$.

12.3 Recommendations

There are clearly cases where, if the noise level is high enough, a model–selection based on marginalising out free parameters will favour weaker signals in spite of the χ^2 likelihood preference for models with stronger amplitudes, even when comparing models with matching dimension and flat or comparable priors. In this regime, whether one uses the MAP point or the maximum–likelihood point as the estimate of b is rather moot, as it is likely the model

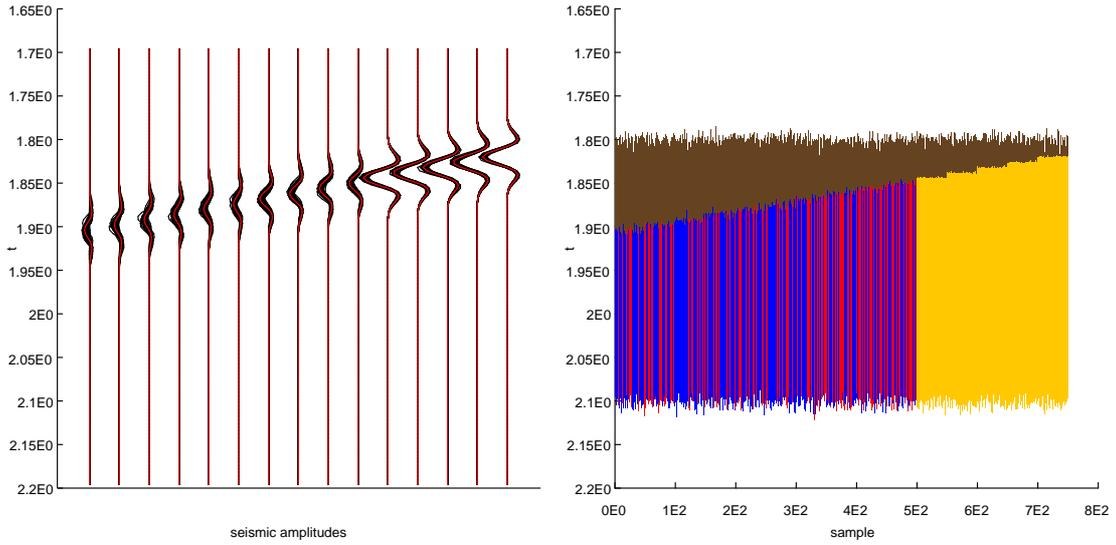


Fig. 31. Gas–oil flank problem. Posterior samples for small noise $\sigma = 0.015$. True seismic in red, MAP synthetic in black. Fluids codes: blue=brine, red=oil, orange=gas.

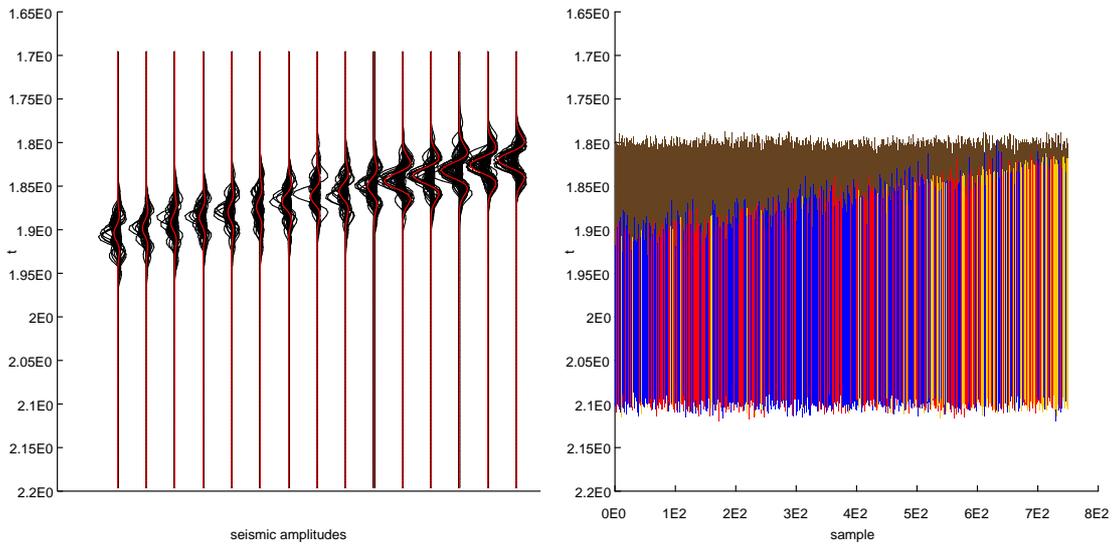


Fig. 32. Gas–oil flank problem. Posterior samples for large noise $\sigma = 0.08$. True seismic in red, MAP synthetic in black. Fluids codes: blue=brine, red=oil, orange=gas.

probabilities are broadly comparable. Further, the probability estimates provided by either the Laplace approximation or MCMC sampling will likely have errors of the same order as the difference in model probabilities, especially in high-dimensional problems. More philosophically, this highlights the weakness of relying on point estimates: in this regime, the relative model probabilities are so comparable that any forecast of interest must average over all of them in some useful way.

Some users like to see maximum-likelihood (minimum χ^2) fits, and are not

comfortable with the use of the “Occam” factor. A practical solution is to provide each of these outputs:

- the model and parameters with the maximum likelihood point, omitting marginalisation, obtained with the flag `--ML`.
- the MAP point of the most-likely model, as judged from the marginalised evidence. Use `--MAP`.
- full samples from the joint posterior over all models and parameters, so any forecast can take into account the uncertainty in both models and parameters. Typically this is some combination like `-RWS -N 1000` with perhaps `--Raftery-auto-decimation`

This way users have access to whichever quantities are of interest to them, and are free to interpret as they please.

The `--no-descent-test` flag also forces more exhaustive exploration in the optimisation phase, so this can be helpful. The transition into this conservative “dimming” behaviour in the inference is also a rather sensitive function of the noise level, so it may be worth revisiting the estimation procedures for that parameter.