Geophysical Journal International

Geophys. J. Int. (2016) **205,** 915–937
Advance Access publication 2016 February 10
GJI Marine geosciences and applied geophysics

doi: 10.1093/gji/ggw061

Frequency domain full waveform elastic inversion of marine seismic data from the Alba field using a Bayesian trans-dimensional algorithm

Anandaroop Ray,[1] Anusha Sekar,[1] G. Michael Hoversten[2] and Uwe Albertin[1]

[1] *Chevron U.S.A. Inc., Houston, TX* 77002*, USA. E-mail: a.ray@chevron.com*
[2] *Chevron U.S.A. Inc., San Ramon, CA* 94583*, USA*

Accepted 2016 February 8. Received 2016 February 4; in original form 2015 December 11

SUMMARY

We present an algorithm to recover the Bayesian posterior model probability density function of subsurface elastic parameters, as required by the full pressure field recorded at an ocean bottom cable due to an impulsive seismic source. Both the data noise and source wavelet are estimated by our algorithm, resulting in robust estimates of subsurface velocity and density. In contrast to purely gradient based approaches, our method avoids model regularization entirely and produces an ensemble of models that can be visualized and queried to provide meaningful information about the sensitivity of the data to the subsurface, and the level of resolution of model parameters. Our algorithm is trans-dimensional and performs model selection, sampling over a wide range of model parametrizations. We follow a frequency domain approach and derive the corresponding likelihood in the frequency domain. We present first a synthetic example of a reservoir at 2 km depth with minimal acoustic impedance contrast, which is difficult to study with conventional seismic amplitude versus offset changes. Finally, we apply our methodology to survey data collected over the Alba field in the North Sea, an area which is known to show very little lateral heterogeneity but nevertheless presents challenges for conventional post migration seismic amplitude versus offset analysis.

Key words: Fourier analysis; Inverse theory; Probability distributions; Wave propagation.

1 INTRODUCTION

The seismic full waveform inversion (FWI) method (e.g. see Virieux & Operto 2009, for an overview) is a geophysical characterization tool which promises to become a mainstay of the exploration and production industry. The objective is simple—instead of performing the traditional process of migration on raw shot data to provide an image of the subsurface reflectivity (Wapenaar 1996), find a model(s) of the subsurface which when input to a suitable wave propagation engine, provides a 'reasonable fit' to the observed raw seismic gathers. The propagated seismic field should match the raw gathers both in amplitude and phase, inclusive of reflections, refractions, mode conversions and all multiples (Tarantola 1984). However, in the 30 years since its inception, this idea has not completely come to fruition and FWI remains an active research problem (Aleardi & Mazzotti 2014; Alkhalifah 2014; Biondi & Almonin 2014; Vigh *et al.* 2014; Kamath *et al.* 2015, to name only a few articles). There are various reasons for this, but first and foremost, a numerical forward simulation of the full seismic field is exceedingly costly, but with the advent of parallel computing techniques (e.g. Komatitsch *et al.* 2010; Johnsen *et al.* 2015) this problem appears less daunting with every passing month. Second, the inverse problem is highly nonlinear, which manifests itself through the presence of local misfit minima in the elastic parameter model space. Here we shall attempt to overcome the non-uniqueness inherent in FWI and despite the specific assumptions used in this paper of a layered earth, provide a generally applicable method for doing so with any kind of earth model geometry.

1.1 The importance of model regularization

The problem of dealing with model non-uniqueness and characterizing uncertainty in the inverted models is one that has no easy solution (Fichtner & Trampert 2011). The problem is exacerbated by the presence of (not always) random noise, which the reflection seismology community has been historically loth to deal with. Given the traditional emphasis on imaging, it follows that noise removal makes for a cleaner image. From a numerical optimization point of view, this is understandable as we would like to remove any impediment (i.e. noise) that would prevent our favourite optimization algorithm from producing a model which predicts a seismic field that is as close to the observations as possible. However, for geophysical inversion, we

are not always justified in getting as good of a data fit as we possibly can. This has been remarkably demonstrated by Parker (1980), who proved that the model with lowest misfit for the nonlinear 1-D magnetotelluric (MT) inverse problem with discretely sampled and noisy data (a situation true for all geophysical acquisition including FWI) is a series of delta functions in the earth—which is not geologically a very reasonable model. This is where regularization (or the inclusion of prior information as we will see later) plays an important role, such that the objective function being optimized, in addition to data misfit attempts to keep deviations from a preferred model low—the generally preferred approach with deterministic FWI. However, this method does not entirely preclude over-fitting the data and producing an overly complicated subsurface model, all features of which should not be believed.

1.2 Occam's Razor and the preference for simplicity

At this juncture, one may also resort to an examination of the statistics of the data noise. If we assume, for the moment, that the noise is random and Gaussian, then we should only be fitting within a given χ^2 value or noise tolerance. This is one part of the method espoused by Constable *et al.* (1987) and the Occam's inversion philosophy. Now that we have introduced data error statistics, we can then make a leap of intuition and ask, what of the model uncertainty statistics? For a linear inverse problem with Gaussian data noise, the inverted model uncertainties are provably Gaussian (Menke 1989). In this case the model estimate corresponding to the L2 norm least squares data misfit, is equal to the mean of the post-inversion Gaussian model distribution. This however brings us to the non-trivial question of 'model selection', that is, how many model parameters do we use to solve the inverse problem? Overparametrizing the model may lead to overfitting the data and underparametrizing may not fit the observations accurately enough. Further, for a nonlinear problem, even with Gaussian data noise, the model uncertainties are no longer Gaussian—in fact, from a statistical point of view, this is what leads to multimodality in the model space (local minima from the optimization point of view). Traditional gradient based inversion methods depend highly on the start model and are guided by both the preferred model and misfit gradient to approach 'optimal' solutions in the model space, of which there could be many—leading to the infamous 'cycle skips' which plague FWI. How then does one solve a nonlinear inverse problem or hope to quantify model uncertainty for such a problem? For FWI, one may use multiscale approaches and well thought out frequency selection schemes (Sirgue & Pratt 2004; Fichtner *et al.* 2009)—but as pointed out in Fichtner & Trampert (2011), the evidence of convergence to an optimal solution using these methods is empirical, though undoubtedly useful if starting from a good initial model. An alternative approach could be to make further use of the Occam's inversion philosophy, where we overparametrize the problem beyond what one can hope to reasonably resolve, and then carry out a line search to find the smoothest (i.e. simplest) model commensurate with both the data error statistics and a prior model. We can then examine features of interest (i.e. anomalies) in the smoothest allowable solution with some confidence. A necessary caveat is that the Occam result is an extremal model (smoothest) - and care should be taken in its interpretation. It will not rid us of non-uniqueness in the inverted model space—paraphrasing Sven Treitel there is no such thing as a 'slightly nonuniqe' inverse problem (Constable 2013). Further, a single inverted model that is maximally smooth, may hide from us the full resolving capability of FWI given that the full sesimic field carries in it great amounts of information on subsurface structure and elastic properties. This motivates us then, towards examining a probabilistic approach for FWI.

1.3 Bayes' theorem and its relationship to the Occam factor

Both the model selection problem and uncertainty within inverted models can be tackled within a Bayesian framework, where information is expressed via probability density functions or PDFs. Since Bayesian probability is a measure of information (Tarantola & Valette 1982; Mosegaard & Tarantola 1995; Scales & Sneider 1997) and it is the aim of FWI to provide information about the earth's subsurface elastic parameters, it is natural to pose FWI in a Bayesian framework. Mathematically speaking, Bayesian information is contained in PDFs represented by $p(\cdot)$. Using Bayes' theorem, we write

$$p(\mathbf{m}|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m}) \cdot p(\mathbf{m})}{p(\mathbf{d})}. \tag{1}$$

For Bayesian geophysical inversion, the observed data vector $\mathbf{d}$ is a constant. All PDFs with a model dependence are functions of the random variable $\mathbf{m}$. The term $p(\mathbf{d}|\mathbf{m})$ can then be interpreted as the model likelihood, the functional form of which depends on the statistics of the noise distribution, and the value of which depends on the model $\mathbf{m}$ being sampled and its fit to observed data. For Gaussian noise, the model likelihood is given as

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left(- \frac{[\mathbf{d} - \mathbf{f}(\mathbf{m})]^T \mathbf{C_d}^{-1} [\mathbf{d} - \mathbf{f}(\mathbf{m})]}{2} \right). \tag{2}$$

Here $f(\mathbf{m})$ corresponds to the modelled data and $\mathbf{C}_d$ is the covariance matrix of the residual data errors including theory error. $[\mathbf{d} - f(\mathbf{m})]^T \mathbf{C_d}^{-1} [\mathbf{d} - f(\mathbf{m})]$ is the χ^2 misfit for the evaluated model $\mathbf{m}$. The prior model PDF $p(\mathbf{m})$ represents our state of knowledge *independent* of the survey data. The evidence term $p(\mathbf{d})$ corresponds to a constant PDF normalizing factor equal to the integral over all models of the numerator on the right-hand side of (1). The aim of Bayesian inversion is to obtain the PDF $p(\mathbf{m}|\mathbf{d})$, or the posterior model PDF, which conditions our prior notions of the subsurface to the observed seismic data. Loosely speaking, the posterior model PDF contains the model ensemble that has been sculpted out of our set of prior models, after considering the misfit due to each model.

In order to understand the second level of inference (MacKay 2003), that is, finding uncertainty on the *number* of parameters being sampled, let us rewrite eq. (1) and derive the Bayesian *Occam Factor*. First, we split the model $\mathbf{m}$ into two parts, one part containing a random variable k representing the k parameters (layers in 1-D, cells for a geobody parametrization, etc.) into which we divide the earth model, and the other part $\mathbf{m_k}$ is a random vector that contains the actual elastic parameters of interest such as the velocities, densities and the positions of these cells. We can then write the joint probability of observed data and models and use the chain rule of probabilities as follows

$$p(\mathbf{m}, \mathbf{d}) = p(\mathbf{m_k}, k, \mathbf{d}), \tag{3}$$

$$p(\mathbf{m_k}|k, \mathbf{d}) p(k|\mathbf{d}) p(\mathbf{d}) = p(\mathbf{d}|\mathbf{m_k}, k) p(\mathbf{m_k}|k) p(k), \tag{4}$$

$$p(k|\mathbf{d}) = \frac{p(\mathbf{d}|\mathbf{m_k}, k) p(\mathbf{m_k}|k) p(k)}{p(\mathbf{d}) p(\mathbf{m_k}|k, \mathbf{d})}. \tag{5}$$

At this juncture, if we recognize that the evidence term $p(\mathbf{d})$ is a normalizing constant, assume that the prior $p(k)$ on the number of cells that our models are permitted to have is uniform and therefore a constant in eq. (5), we can then write

$$p(k|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}_\mathrm{k}, k) \cdot \left[\frac{p(\mathbf{m}_\mathrm{k}|k)}{p(\mathbf{m}_\mathrm{k}|k, \mathbf{d})} \right]. \tag{6}$$

Eq. (6) is fundamental to understanding Bayesian preference for simpler explanations. The left-hand side is the probability that our models can consist of k parameters, given the observations of seismic data $\mathbf{d}$. The first term on the right-hand side is the likelihood of the models $\mathbf{m}_\mathrm{k}$ with k elastic parameters. This likelihood depends on the modelled data misfit as shown in eq. (2). The bracketed term on the right of eq. (6) is the ratio of prior model probability of a k-parameter model (independent of data observations, based on experience), to the posterior k-parameter model probability (based on fit to the observations, after drawing from within the set of prior assumptions about values of elastic parameters). To make the argument even simpler, let us assume that we are only interested in P-wave velocity in the earth, and divide the earth for a particular model such that it has k layers. We assume *a priori* that the velocities in different layers are independent and uniformly distributed in a range ΔV. However, after collecting our seismic data and drawing velocities uniformly from these k layers to fit the observations, we find that for a layer i, we occupy only a fraction f_i of the prior range of velocities ΔV in every layer for our k-layered models. We find this to be true after investigating (i.e. drawing) numerous models. We can then rewrite eq. (6) as

$$p(k|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}_\mathrm{k}, k) \cdot \left[f_1 . f_2 ... f_k \right]. \tag{7}$$

Thus if we were to compare two models with different k, but similar misfit and hence similar likelihood in eq. (7), the model with the larger k will likely have lower posterior probability $p(k|\mathbf{d})$ owing to the multiplication of a larger number of fractional quantities on the right-hand side. Thus the Occam Factor, from eq. (7) can be thought of as the ratio of posterior accessible volume to prior accessible volume, to quote MacKay (2003). It typically penalizes more complex parametrizations without explicitly requiring model regularization, as was first brought to the attention of the geophysics community in a seminal paper by Malinverno (2002) on Bayesian parsimony in geophysical inversion. Of course, the moment we select a basis set (e.g. piecewise constant functions) to represent our model, we are limiting ourselves to features representable by that basis set (Valentine & Trampert 2015), as has been pointed out by Hawkins & Sambridge (2015), but this does not invalidate our methodology on considering models with a varying number of parameters once a basis set is chosen.

At this point, we would like to point out that it may be possible to use Bayesian parsimony to perform model selection in a wider sense. For example, questions on the form of the data error statistics, the choice of a model basis set, the inclusion of anisotropy, which frequencies to use or even whether to use a 1-D, 2-D or 3-D model—fall within the purview of model selection. However, within this work, we restrict ourselves to model selection over a varying number of elastic layers (or cells in higher dimensions).

1.4 Trans-dimensional Bayesian inversion

It is impossible to obtain the posterior model PDF $p(\mathbf{m}|\mathbf{d})$ in eq. (1) by exhaustively sampling a large model space, hence we resort to using Markov chain Monte Carlo (McMC) sampling algorithms (e.g. Gilks *et al.* 1996) to somewhat circumvent the 'curse of dimensionality' and generate model samples proportional to $p(\mathbf{m}|\mathbf{d})$. We know that fixing a particular model parametrization for the inversion is known to produce posterior PDFs, *only* for the given parametrization as shown in Ray (2014, see fig. 1.13 for a striking example) and Dettmer *et al.* (2010). This is where the 'trans-dimensional' or 'reversible jump' (Geyer & Møller 1994; Green 1995) Markov chain Monte Carlo (RJ-McMC) differs from traditional McMC methods, in sampling from a posterior PDF where the number of unknowns (i.e. the parametrization) is also treated as part of the inverse problem. In other words, the parametrization is also inferred from the observed data. A review of applications which use trans-dimensional McMC can be found in Sisson (2005). Sambridge *et al.* (2006) further discuss this method in the context of evidence based model selection (Bernardo & Smith 1994; Denison 2002). An introduction to geophysical trans-dimensional Bayesian inversion (referred to as 'trans-D' from now on) can be found in Sambridge *et al.* (2013). Malinverno & Leaney (2000, 2005) use trans-D for vertical seismic profile (VSP) inversions. Bodin & Sambridge (2009) use trans-D McMC for solving the seismic surface wave tomography problem. Fontaine *et al.* (2015), Bodin *et al.* (2012, 2013) and Piana Agostinetti & Malinverno (2010) apply trans-D to the seismic receiver function inversion problem. Recent applications of this method to solve geophysical electromagnetic problems can be found in Gehrmann *et al.* (2015a), Ray *et al.* (2013a,b, 2014), Ray & Key (2012), Minsley (2011) and Brodie & Sambridge (2012). In particular, the geo-acoustics community has been extensively using trans-D Bayesian methods to invert for near surface sediment velocities at high frequencies above 300 Hz. (Dettmer & Dosso 2012; Dettmer *et al.* 2010; Dosso *et al.* 2014). We have discussed the trans-D method in some detail in Appendix B. Further, we have wrapped our trans-D McMC sampling within a parallel tempering framework (Dosso *et al.* 2012; Ray *et al.* 2013a; Sambridge 2013), to facilitate sampling of a peaky misfit space (i.e. a multimodal likelihood landscape), as elaborated upon in the Appendix B6.1

As we have noted earlier, in the exploration seismology context, trans-D has been applied to VSP data inversion by Malinverno & Leaney (2005). However, we are not aware of any applications of the trans-D method to the exploration seismology nonlinear elastic FWI problem, and consider our work to be the first. A further highlight of our work is that we do not require the source seismic wavelet to be known prior to inversion. The source signal can often be problematic to derive, especially in shallow water. For a conventional FWI method which requires a fixed source estimate, the error will propagate into the inverted models and the results can be unsatisfactory. After validating our methodology using a synthetic example, we invert ocean bottom cable (OBC) hydrophone data over the Alba field in the UK North Sea area (Sears *et al.* 2010).

2 THEORY

2.1 Frequency domain formulation and additive noise model

To describe a seismic shot gather in the water column, we use the inhomogeneous acoustic wave equation, subject to the boundary conditions of a free surface above the source and receivers, with an elastic seafloor underneath. The solution to this problem will account for all parts of the observations, including the direct wave, source and receiver ghosts, water layer reverberations and all mode

conversions, reflections and refractions including multiples from within the earth. We can write the time domain partial differential equation subject to these boundary conditions as

$$\left(\nabla^2 - \frac{1}{c_0^2}\frac{\partial^2}{\partial t^2}\right)p(t,\mathbf{r}) = s(t)\delta(\mathbf{r}), \tag{8}$$

where $p(t,\mathbf{r})$ is the pressure field recorded at the receivers, $s(t)$ is the source time function associated with an impulsive source at the origin of co-ordinates, with the water velocity given as c_0. Fourier transforming every receiver trace from time to frequency using the following definition,

$$P(\omega,\mathbf{r}) = \int_{-\infty}^{\infty} p(t,\mathbf{r})\mathrm{e}^{i\omega t}\mathrm{d}t, \tag{9}$$

we arrive at the inhomogeneous Helmholtz equation for the signal at angular frequency ω across all receivers:

$$\left(\nabla^2 + \frac{\omega}{c_0}^2\right)P(\omega,\mathbf{r}) = S(\omega)\cdot\delta(\mathbf{r}). \tag{10}$$

If we know the Green's $G(\omega,\mathbf{r})$ function for eq. (10) with a unit source, then our solution at every receiver location $\mathbf{r}$ is simply

$$P(\omega,\mathbf{r}) = S(\omega)G(\omega,\mathbf{r}), \tag{11}$$

where the solution obtained in the time domain after inverse Fourier transforming (11) is:

$$p(t,\mathbf{r}) = \frac{1}{2\pi}\int_{-\infty}^{\infty} P(\omega,\mathbf{r})\mathrm{e}^{-i\omega t}\mathrm{d}\omega \tag{12}$$

However, if we consider additive random noise $n(t,\mathbf{r})$ in the time domain, our observed data can be represented as

$$d(t,\mathbf{r}) = p(t,\mathbf{r}) + n(t,\mathbf{r}), \tag{13}$$

using the definition in eq. (9) and Fourier transforming eq. (13), we obtain using eq.(11)

$$d(\omega,\mathbf{r}) = G(\omega,\mathbf{r})S(\omega) + N(\omega,\mathbf{r}). \tag{14}$$

Note from eq. (14) that if $n(t,\mathbf{r})$ is independent (i.e. samples well separated in time are only weakly dependent), then by the definition (9), $N(\omega,\mathbf{r})$ which is the sum of scaled independent random variables must be Gaussian by central limiting (e.g. Mecklenbrauker & Gerstoft 2000). This is elaborated upon in section IV of Brillinger (1974), especially for the case of discrete data and a large number of time samples. This justifies our use of a Gaussian likelihood for solving FWI in the frequency domain.

2.2 Estimating the source wavelet

Let $\mathbf{d}_l \in \mathbb{C}^{n_r}$ be the complex pressure field observed at n_r receivers at frequency ω_l and $\mathbf{G}_l\,(\mathbf{m}) \in \mathbb{C}^{n_r}$ be the Green's function at those receivers at the same frequency due to a model $\mathbf{m}$. Then from eq. (14) we can see that the complex source wavelet estimate at frequency ω_l, can be obtained either using a maximum likelihood (ML) estimate (Mecklenbrauker & Gerstoft 2000; Dosso & Wilmut 2012; Dettmer *et al.* 2015; Gehrmann *et al.* 2015b), or as the solution to a linear overdetermined least-squares problem (e.g. Virieux & Operto 2009). The estimate of the source is given as

$$\hat{S}_l(\mathbf{m}) = \frac{\mathbf{G}_l^{\dagger}(\mathbf{m})\mathbf{d}_l}{\mathbf{G}_l^{\dagger}(\mathbf{m})\mathbf{G}_l(\mathbf{m})}, \tag{15}$$

where † denotes the Hermitian transpose. Note that no matrix inverse computation is required to arrive at this estimate.

2.3 Data noise estimation through maximum likelihood methods

Within a Bayesian formulation, it is essential to know the noise statistics well in order to be able to formulate a likelihood. However, within the exploration seismology community, the focus has been on noise removal as opposed to estimation. Often times, this manifests in complicated processing of the raw shot gathers, after which inversion of the resultant processed data becomes intractable as we may not know enough about this processing flow to apply the same steps to our forward modelled data. Of course, given the operational constraints of obtaining seismic data, it may not be possible to obtain repeated measurements of a shot so as to obtain variance estimates on the recorded traces. One could argue that stacking adjacent traces will reduce noise variance, but we need to know how large of a spatial aperture to stack within so as to not introduce modelling error, and that again requires significant manual intervention. An approach which has been in use by the geo-acoustics community for many years, has been to use ML methods to derive objective functions with an estimate of the spectral noise as well as the source wavelet, and then use these objective functions within likelihood functions for inversion or Bayesian sampling (Mecklenbrauker & Gerstoft 2000; Dosso & Wilmut 2012). Since we have already justified the use of Gaussian likelihoods in the frequency domain, and dealt with estimates of the source spectrum, let us now turn our attention to the last remaining piece of the likelihood estimate, the frequency domain covariance. We begin by noting that we can write a Fourier transform as defined in eq. (9) as a matrix $\mathbf{F}$ for n_f frequencies which operates on a $n_t \times 1$ time series column vector to its right in the following manner

$$\mathbf{F} = \begin{bmatrix} \mathrm{e}^{i\omega_1 t_1} & \cdots & \mathrm{e}^{i\omega_1 t_{n_t}} \\ \vdots & \ddots & \vdots \\ \mathrm{e}^{i\omega_{n_f} t_1} & \cdots & \mathrm{e}^{i\omega_{n_f} t_{n_t}} \end{bmatrix}, \tag{16}$$

The Fourier transform of a random $n_t \times 1$ time series $\mathbf{x}$ is a linear transformation and can be written as $\tilde{\mathbf{x}}$, where

$$\tilde{\mathbf{x}} = \mathbf{F}\mathbf{x}. \tag{17}$$

Let $\mathbf{x}$ be zero mean, with time domain covariance matrix $\mathbf{C}_\tau$. If the mean and covariance of a random vector $\mathbf{y}$ are given as $\mathbf{y}_0$ and $\mathbf{C}_y$ respectively, it can easily be shown that under a linear transformation $\tilde{\mathbf{y}} = \mathbf{B}\mathbf{y}$ that the mean and covariance of $\tilde{\mathbf{y}}$ are given by $\mathbf{B}\mathbf{y_0}$ and $\mathbf{B}\mathbf{C_y}\mathbf{B^t}$ respectively (e.g. Timm 2002). Thus for the frequency domain random vector $\tilde{\mathbf{x}}$, the covariance matrix $\mathbf{C}_\omega$ can be written as

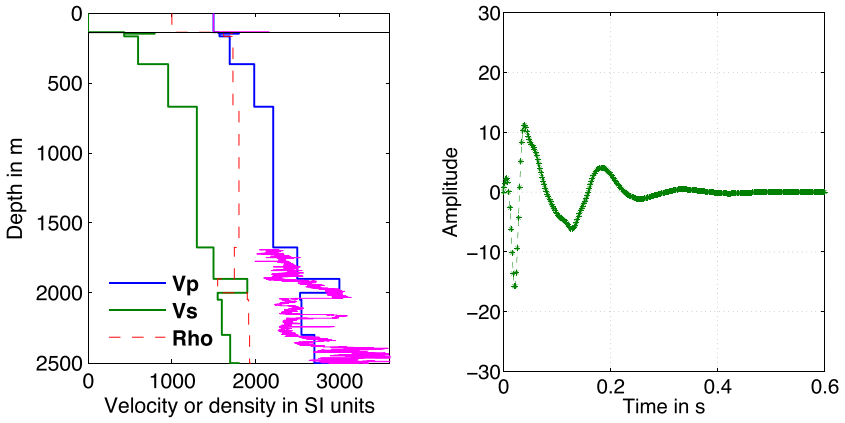
$$\mathbf{C}_\omega = \mathbf{F}\mathbf{C}_\tau{}^{\mathbf{t}}\mathbf{F}^{\dagger}, \tag{18}$$

where † represents a Hermitian transpose. Under the assumption of independent data error of variance σ^2 at every receiver, using the orthonormality of $\mathbf{F}$ this simplifies to

$$\mathbf{C}_\omega = \sigma^2\mathbf{F}\mathbf{F}^{\dagger}, \tag{19}$$

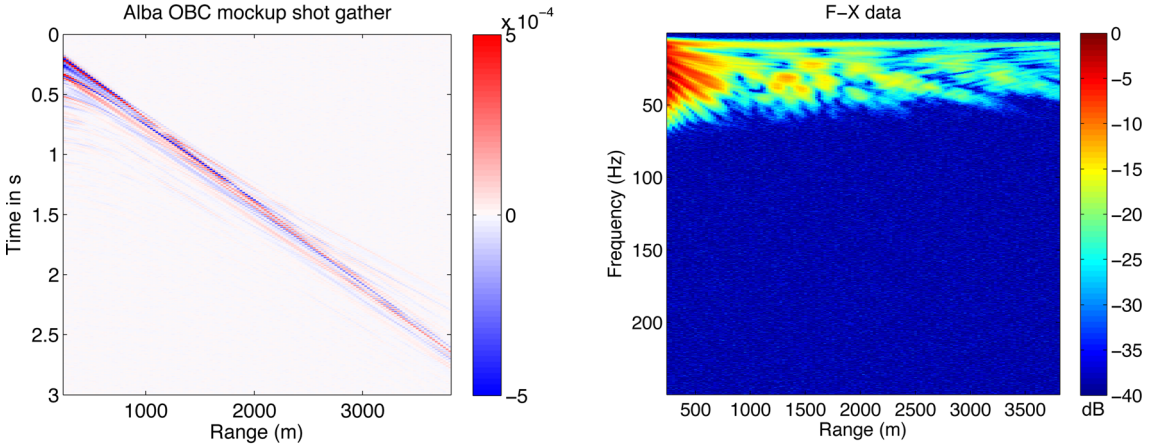
$$\Rightarrow \mathbf{C}_\omega = \frac{\sigma^2}{n_t}\mathbf{I}, \tag{20}$$

where $\mathbf{I}$ is the identity matrix. Thus we see that white data noise in the time domain leads to white data noise in the frequency domain. However, seismic data are often low passed and receivers have responses which are not flat in the frequency domain. We can represent the operation of low passing as another linear transformation



(a) (b)

Figure 1. (a) Synthetic model based on Alba well data (magenta) and background geology from Sears (2007). The high-velocity, low-density reservoir is near 2 km depth. (b) Wavelet produced from far field signature during the actual OBC survey. This wavelet was used to compute the synthetic T-X gather in Fig. 2(a).



(a) (b)

Figure 2. (a) Noisy time domain shot gather for the synthetic example. (b) The resulting F-X amplitude spectrum with every trace in (a) Fourier transformed. Amplitude colour scale is shown in dB relative to the maximum signal amplitude near 20 Hz at the first offset of 250 m.

in the frequency domain. Using a diagonal matrix $\mathbf{H}$ with purely real entries $(h_1, \ldots, h_{n_f})$, we low pass $\tilde{\mathbf{x}}$ to produce $\tilde{\mathbf{x}}_l$ as follows

$$\tilde{\mathbf{x}}_l = \mathbf{H}\tilde{\mathbf{x}}, \tag{21}$$

with covariance $\mathbf{C}'_\omega$ given by

$$\mathbf{C}'_\omega = \mathbf{H}\mathbf{C}_\omega{}^{\mathbf{t}}\mathbf{H}^{\mathbf{t}}, \tag{22}$$

$$\mathbf{C}'_\omega = \frac{\sigma^2}{n_t}\mathbf{H}\mathbf{H}, \tag{23}$$

$$\mathbf{C}'_\omega = \frac{\sigma^2}{n_t}\mathrm{diag}\left(h_1^2, \ldots, h_{n_f}^2\right) \tag{24}$$

Thus for the frequency l we have an unknown variance $\frac{\sigma_l^2}{n_t}$ and the following diagonal (i.e. uncorrelated) frequency domain covariance,

$$\mathbf{C}'_\omega = \mathrm{diag}\left(\frac{\sigma_1^2}{n_t}, \ldots, \frac{\sigma_l^2}{n_t}, \ldots, \frac{\sigma_{n_f}^2}{n_t}\right). \tag{25}$$

For random variables with uncorrelated real and imaginary parts, i.e. with circular symmetry in the Argand plane, we use a slightly modified likelihood $\mathcal{L}(\mathbf{m})$ instead of eq. (2),

$$\mathcal{L}(\mathbf{m}) = p(\mathbf{d}|\mathbf{m}) = \prod_{l=1}^{n_f}\left(\pi\frac{\sigma_l^2}{n_t}\right)^{-n_r}\exp\left(-\frac{[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]}{\sigma_l^2/n_t}\right), \tag{26}$$

where $\mathbf{f_l}(\mathbf{m})$ is the pressure field due to model $\mathbf{m}$ at frequency ω_l predicted at n_r receivers using the estimate $\hat{S}_l$ from eq. (15) in the forward eq. (11), that is,

$$\mathbf{f_l}(\mathbf{m}) = \hat{S}_l(\mathbf{m}) \cdot \mathbf{G}_l(\mathbf{m}). \tag{27}$$

To obtain the ML estimate of the noise, let us denote the residual at frequency l as

$$\mathbf{r}_l(\mathbf{m}) = \mathbf{d_l} - \mathbf{f_l}(\mathbf{m}). \tag{28}$$

As shown in the Appendix A, if we maximize eq. (26) with respect to σ_l and substitute this value back in eq. (26), we obtain

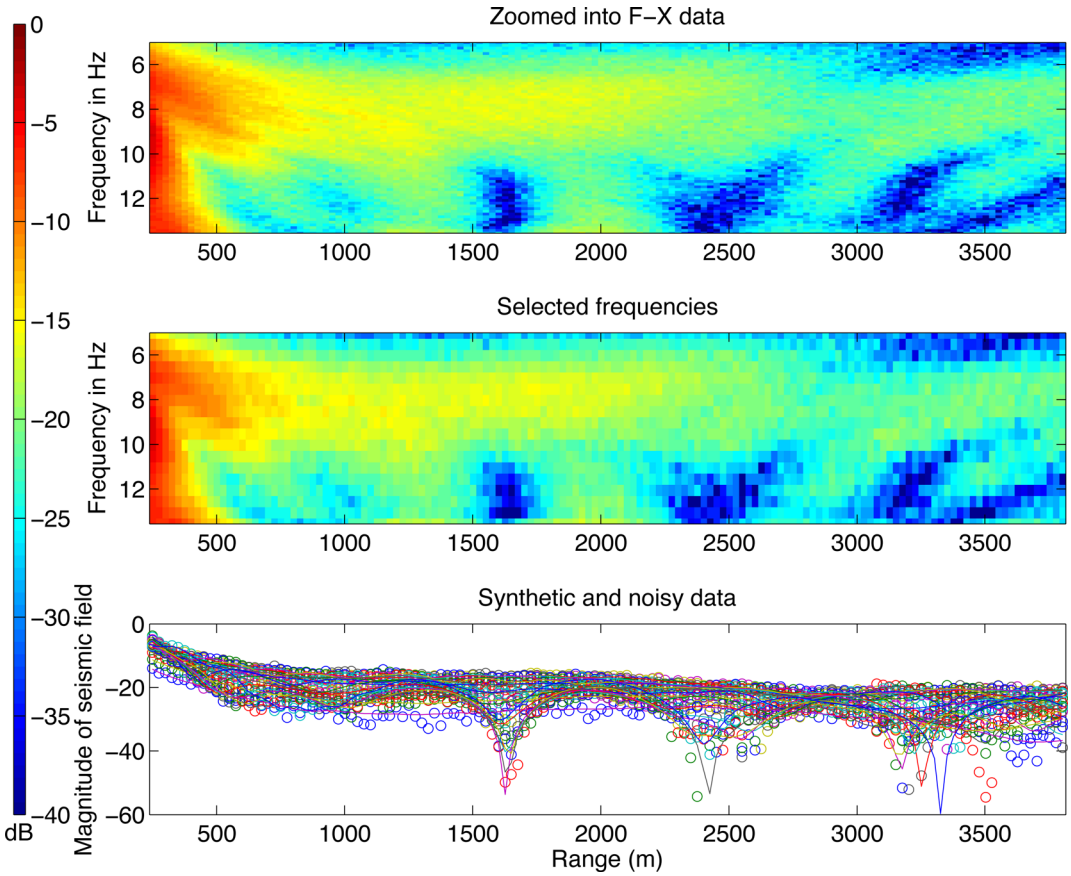


Figure 3. Selection of noisy frequency-offset data for synthetic inversion: Top: zooming into Fig. 2(b). Middle: decimated version of the top row, with frequencies sampled every 0.5 Hz. Bottom: synthetic model responses (solid lines) and noisy data (circles) are shown at each frequency. All amplitudes in dB relative to the maximum amplitude near 20 Hz and 250 m offset.

the following ML expression for the misfit square or negative log-likelihood,

$$-\log \mathcal{L}(\mathbf{m}) = n_r \sum_{l=1}^{n_f} \log(\mathbf{r}_l^{\dagger}(\mathbf{m})\mathbf{r}_l(\mathbf{m})) \tag{29}$$

Thus, using eqs (27)–(29) we can obtain an ML expression or 'processor' as it is known in the geoacoustics community (Dosso & Wilmut 2012), which implicitly samples both the data noise and the source spectrum as a function of the model being sampled. We would like to emphasize that these equations can also be used in a gradient based deterministic inversion, since a negative log-likelihood is equivalent to the squared data misfit. A similar derivation can also be found in the time domain in Sambridge (2013) for sesimic receiver function inversion. Though we could have used a hierarchical Bayesian scheme to sample the noise and source spectrum along with the earth model, the ML approach adopted here reduces the dimensionality of the search space (Sambridge 2013; Dettmer *et al.* 2015). We would like readers to note that we have not made any assumptions specific to the forward modelling geometry here, that is, our outline so far has been completely generic and applicable not only to layered earth but also wave propagation through complicated geological models.

3 APPLICATION OF THE ALGORITHM

We apply the algorithm first to a synthetic example that roughly mimics the geology of the Alba field (Newton & Flanagan 1993), including the low-density and high-velocity reservoir layer at 2 km depth of nearly 100 m thickness. We then apply the method to real marine OBC data obtained from the Alba field.

3.1 Forward modelling by wavenumber integration

We model the full 3-D pressure field in the water due to an elastic subsurface, with all ghosts, internal multiples, water reverberations, mode conversions, reflections and refractions included. However, since we intend this paper as a proof of concept, we restrict ourselves to modelling laterally homogenous media using wavenumber integration methods (Ursin 1983; Schmidt & Tango 1986; Kennett 1986, 2009). To generate an elastic full field shot gather, plane wave reflectivity is modelled in the frequency-wavenumber (F-k) domain, Hankel transformed (over radial wavenumber) to the frequency-offset (F-X) domain, and then inverse Fourier transformed (over frequency) to the time-offset (T-X) domain. We must briefly mention here, that this semi-analytical solution for layered media produces a *3-D response*, the same an accurate 3-D finite difference or finite element code will produce, were it to be tasked with modelling the T-X response due to the same layer-cake geology. For our inversion purposes, observed T-X data are Fourier transformed to

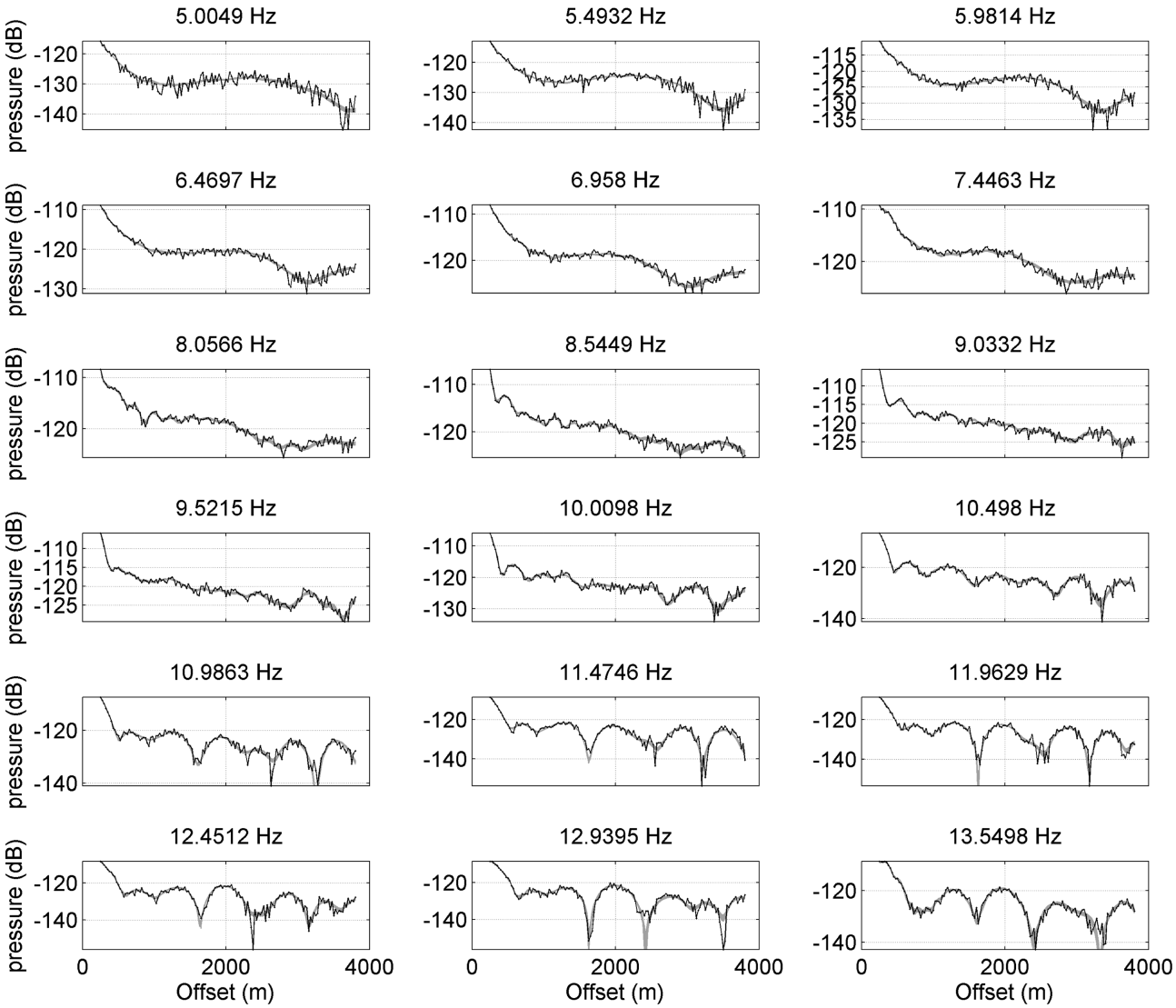


Figure 4. Pressure field amplitude with offset at the 18 inverted frequencies for the noisy synthetic experiment. The observations are shown in black. 500 responses due to models randomly selected from $p(\mathbf{m}|\mathbf{d})$ are shown in grey. Both calculated and observed fields are shown in dB relative to 1 unit of observed pressure. The lack of a constant difference between observations and predictions shows that the ML source spectral estimation scheme is working well.

the F-X domain, and only selected frequencies are modelled in the F-X domain using the algorithm described above. Given the low dip geology and lateral homogeneity exhibited by the Alba field, we are justified in making this assumption.

3.2 Synthetic data generation

For the model shown in Fig. 1(a), using the source wavelet shown in Fig. 1(b) we generated a shot gather in the time-offset (T-X) domain. 143 receivers were modelled with a receiver spacing of 25 m, with the first receiver at 250 m and the last at 3.8 km. The receivers were placed just above the seafloor at 138 m depth. The free surface was set to 6 m above the source. Uncorrelated Gaussian noise, with a standard deviation equal to −65 dB (with respect to the maximum amplitude of the first recorded offset) was added to the data, with the resultant shot gather shown in Fig. 2(a). Interfering events associated with the reservoir depths are found 1.6–1.7 s onwards at the closest offsets, not easily visible with the added noise. We then transformed this T-X gather to the frequency-offset (F-X) domain using the definition (9), by Fourier transforming the signal at each receiver. The noisy amplitude spectrum is shown in Fig. 2(b). Note how most of the signal energy is concentrated at frequencies within 50 Hz, given that the source wavelet did not contain significant energy at frequencies above 50Hz. The attraction of frequency domain modelling is that with just a few frequencies, one can capture the essential features of the seismic wavefield—such as the notches seen in the amplitude spectrum Fig. 2(b)—which are representative of the earth model. This can lead to significant computational savings compared to time domain modelling, where theoretically, many closely spaced frequencies must be considered to produce seismic traces at the required sampling rate, for the required total modelling time interval.

For the synthetic model considered here, we selected for inversion F-X data at 18 frequencies between 5 and 13.5 Hz separated by 0.5 Hz (Fig. 3). Since the generated shot gather was 8.19 s long, we could have had a frequency spacing of $1/8.19 = 0.12$ Hz. We opted not to invert such closely spaced frequencies as the wavefield features at these frequencies are not very different, and possibly will not provide extra information about the earth model sought after.

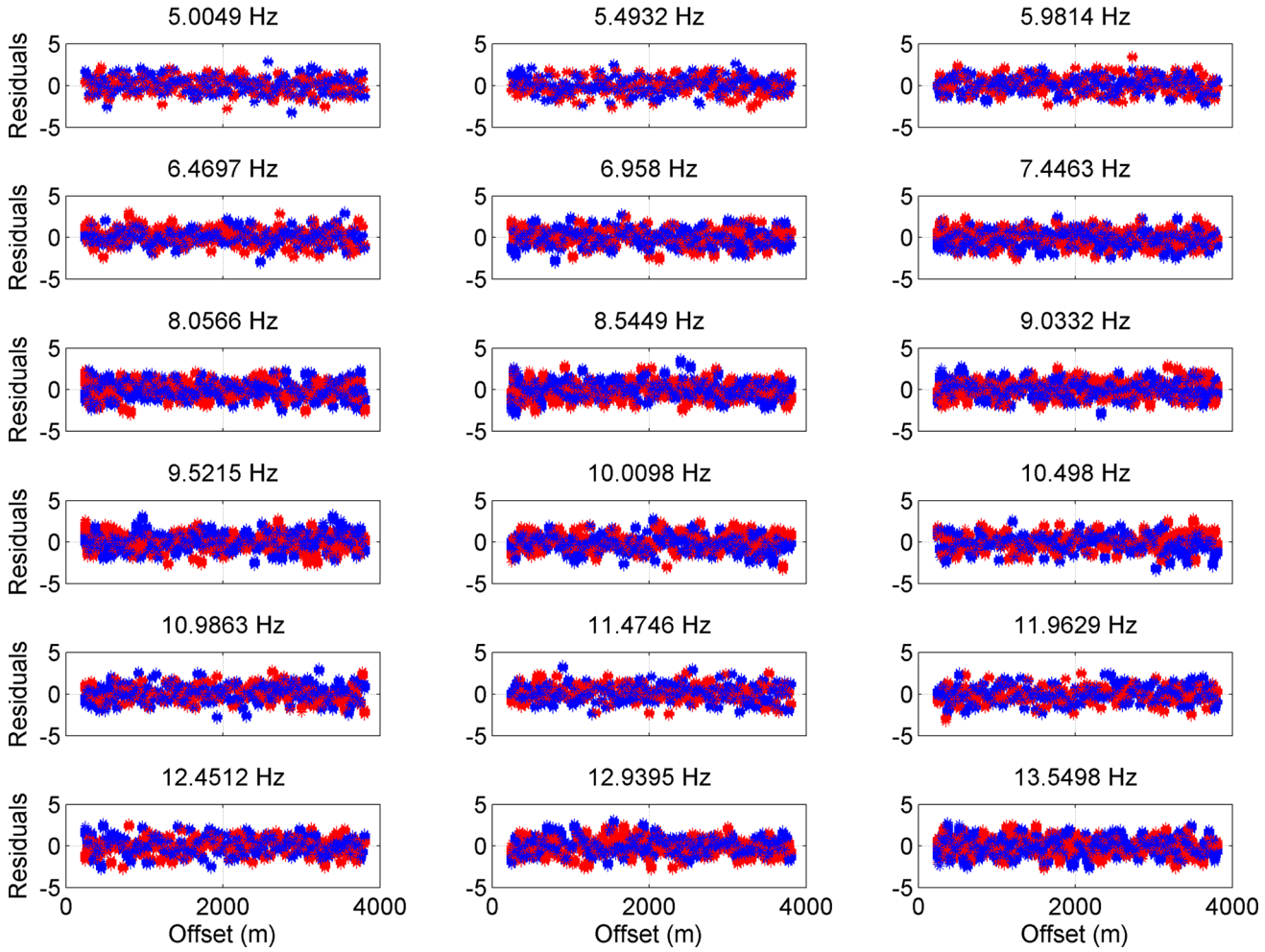


Figure 5. Inversion residuals with offset at 18 frequencies for the synthetic example. 500 real (red) and imaginary (blue) residuals due to models randomly selected from $p(\mathbf{m}|\mathbf{d})$ are shown. The complex data residuals have been normalized by the implicitly sampled ML data error.

This can be seen by comparing the top and middle rows in Fig. 3. Further discussion on the selection of frequencies is done at the end of Section 3.3.

3.3 Synthetic inversion results

Using the trans-D method, we sampled the posterior model ensemble $p(\mathbf{m}|\mathbf{d})$ given in eq. (1) with a uniform prior distribution (detailed later in this section) as well as the likelihood specified in eq. (29). Before going on to examine the results, we believe that an inversion is only as good as the data fit it produces. Since our Markov chain inversion consists of many models in the posterior model ensemble, we randomly selected 500 models and examined their data fit (Fig. 4) as well as the inversion data residuals (Fig. 5) at all 18 frequencies. In each panel of Fig. 4, the black line corresponds to the noisy pressure field observations with offset at a fixed frequency. The band of grey lines in the background are the responses from 500 models within $p(\mathbf{m}|\mathbf{d})$. The amplitude scale for each panel is simply $20\log_{10}$(field), both for the observations and predictions. The close match implies that the scaling scheme (27) is working well. In Fig. 5, an examination of the 500 corresponding real (red) and imaginary (blue) data residuals (normalized by the implicitly sampled data error given by eq. A5 in the Appendix A) at each frequency for every offset shows that the residuals are not correlated with offset and are zero mean Gaussian distributed (Fig. 6) as expected. These diagnostics give us confidence in our inversion scheme, and we can move on to an examination of the posterior models.

Since $p(\mathbf{m}|\mathbf{d})$ is mutidimensional, we cannot view it without marginalizing over the model parameters of interest. Our model parametrization consists of piecewise constant values of compressional velocity V_p, Poisson's ratio ν, density ρ, compressional and shear attenuation Q_p and Q_s. Different models sampled can have different numbers of layers. The above parameters in each layer were uniformly sampled from within fixed prior ranges at every depth. The sampled Poisson's ratio was then used to compute shear velocity V_s given the value of V_p. Prior values were uniformly assigned with depth so as to give the model physics a chance to sculpt out the background elastic trends, instead of our imposing them *a priori*. The marginalized $p(\mathbf{m}|\mathbf{d})$, with the accompanying prior bounds can be seen in Fig. 7. For the first five panels from the left, darker shading corresponds to higher probability density. In these panels, the true model is given by the magenta line, and the two thin blue lines at every depth represent the 90 per cent Bayesian credible interval. The red dashed line represents the median value of the elastic parameter at every depth. The last panel to the right shows the probability that the data require an interface to be present in the model, that is, require a change in elastic parameters with depth.

From these plots, we can see that the true model including the reservoir top is fairly well recovered with minimal uncertainty, at

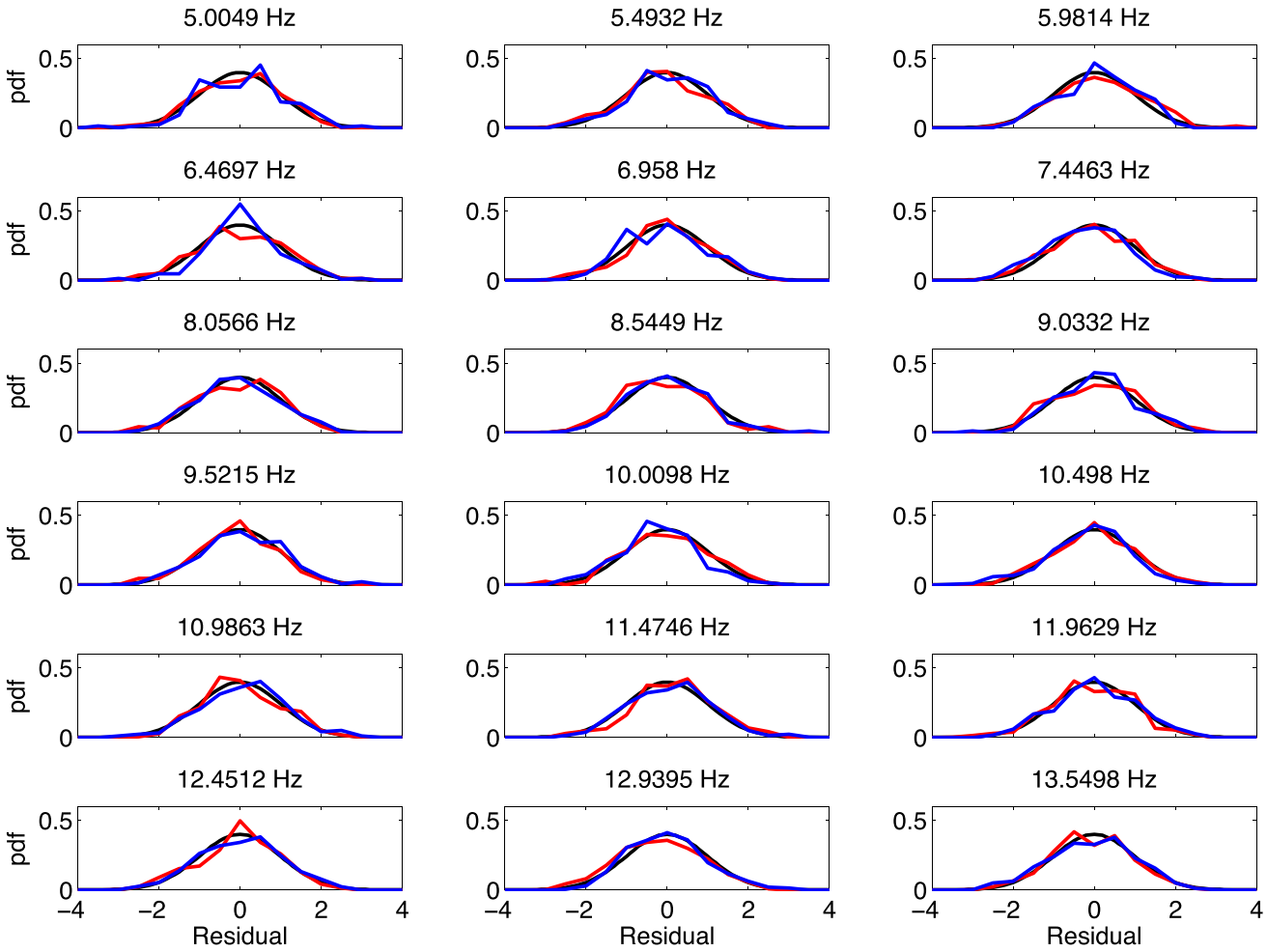


Figure 6. PDFs of normalized inversion residuals for the synthetic data obtained from 500 models. For each frequency, an analytic standard Gaussian PDF is shown in black, and the real and imaginary residuals are shown in red and blue, respectively. The complex data residuals have been normalized by the implicitly sampled ML data error. The assumption of Gaussian data error has been met as expected.

least in V_p and V_s, down to 2 km depth. ρ is less well resolved, and as expected from previous studies, Q_p and Q_s are not well resolved.

Beyond 2 km depth, the reservoir bottom is not well resolved. However, the presence of reservoir can be inferred, given that we see the PDFs of the elastic parameters change in concordance with the true model. We see a spike in the PDFs of V_p and V_s and a sharp decrease in the PDF of ρ. Note that besides the prior ranges of elastic parameters, no accurate starting model is required for a Bayesian McMC inversion.

The number of posterior interfaces required in $p(\mathbf{m}|\mathbf{d})$ to explain the data are shown in Fig. 8(a). In truth there are 12 layers with 11 interfaces, and this plot shows the importance of using a trans-D formulation in estimating the background and anomaly velocities correctly without resorting to regularization or imposition of arbitrary vertical correlation lengths from the seafloor down to the basement. Fig. 8(b) shows 500 estimates of the amplitude spectrum of the ML wavelet sampled using eq. (15) and the spectrum of the wavelet actually used to compute the T-X response. We see that the recovered wavelet spectrum is also an excellent match to the true wavelet spectrum, further validating our methodology in using an ML estimate.

We must mention here, that our choice of frequencies is open to question—we tried the same frequency range, with 1 Hz spacing. The results were unsatisfactory, with the reservoir velocity spike barely discernible in $p(\mathbf{m}|\mathbf{d})$. We have experimented with skip ahead in frequency schemes as described in Sirgue & Pratt (2004), but met with limited success in their use. We did not invert frequencies higher than 13.5 Hz, owing to the computationally demanding requirement of using a greater number of wavenumbers to model the pressure field with increasing frequency. We recommend carrying out forward modelling and sensitivity analysis to different subsurface elastic parameters in order to select the frequencies for inversion. The synthetic study just discussed was motivated by our goal of inverting field data from the Alba OBC survey, to be discussed next.

3.4 Results from the Alba field

The Alba field in the UK North Sea exhibits very low dip geology and remarkably horizontal layering. With a water depth of 138 m, it lies at an approximate depth of 2 km within the Alba formation towards the base of a sequence ranging in age from late Palaeocene to Miocene, dominated by hemipelagic mudstones and occasional sandstones. Details about the geology can be found in Newton & Flanagan (1993). We invert a shot gather from over the 'Area 12' outlier oil bearing area, close to the 12Z well. Full waveform OBC data from this field have been inverted by Sears *et al.* (2010) and

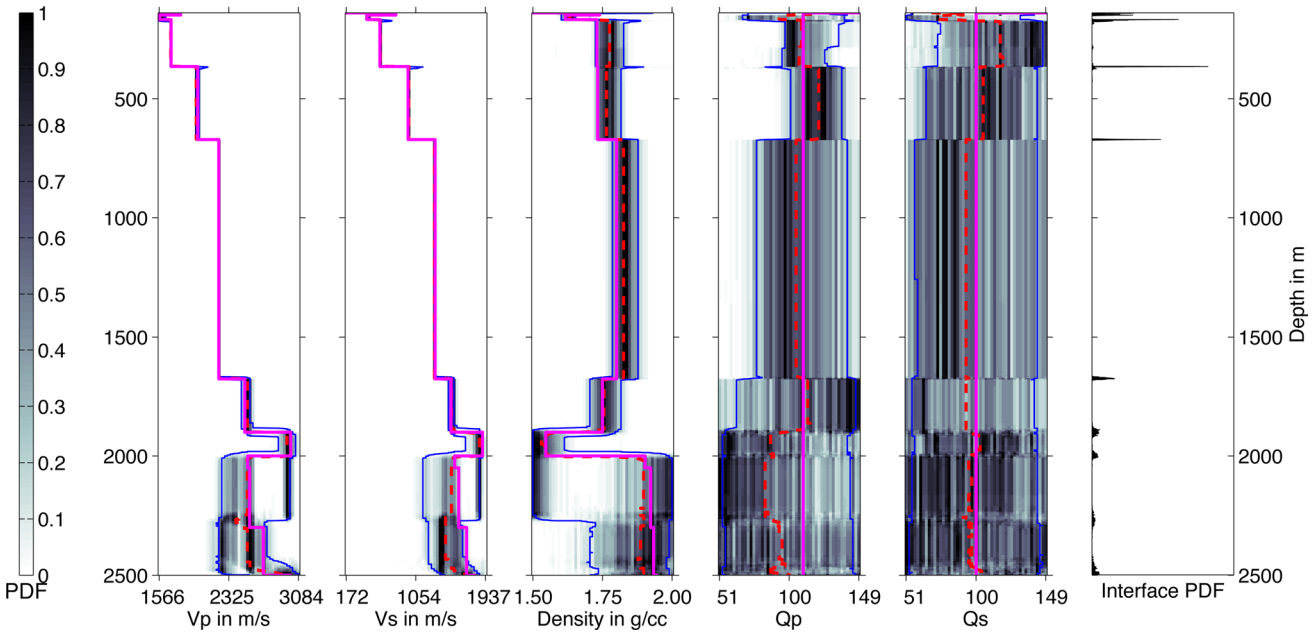
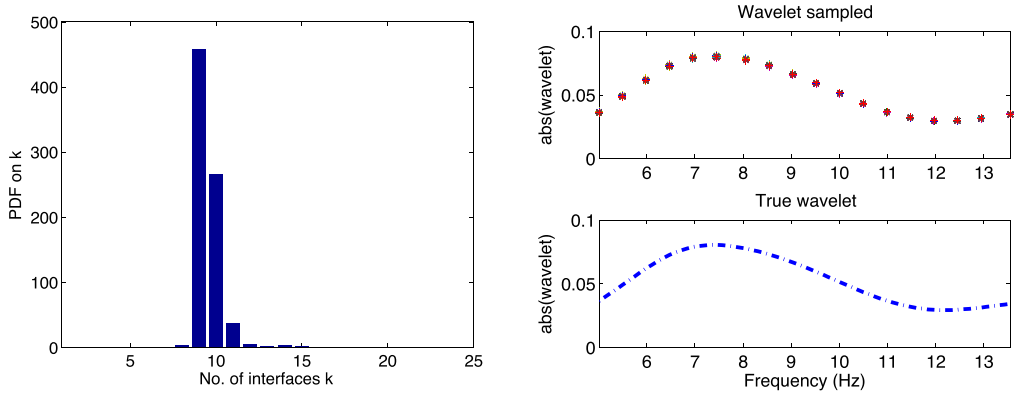


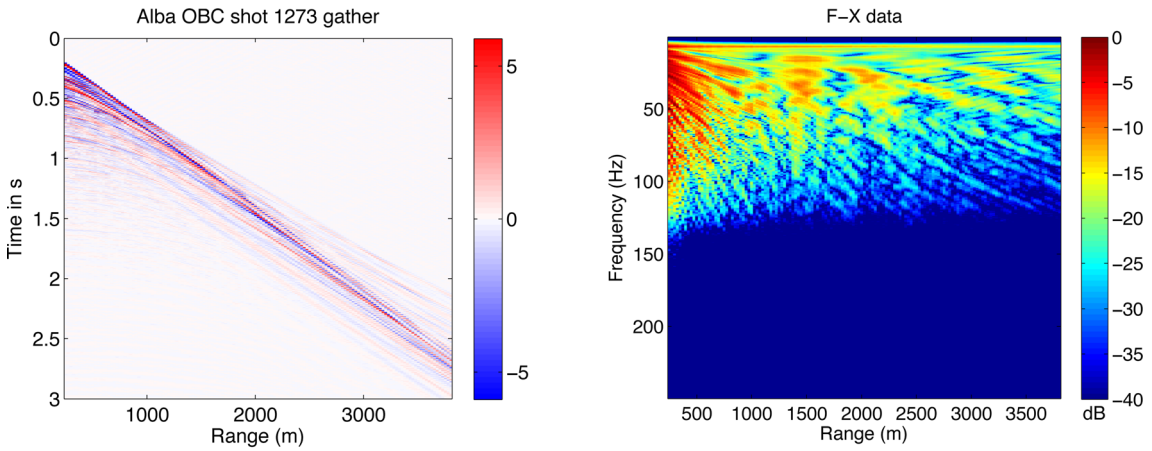
Figure 7. Marginal posterior PDFs with depth for the synthetic inversion. Darker shading corresponds to higher probability density in the first five panels from the left. The true model is given by the magenta line, and the two thin blue lines at every depth represent the 90 per cent Bayesian credible interval. The red dashed line represents the median value of the elastic parameter at every depth. The last panel to the right shows the probability that the data require an interface to be present in the model, that is, require a change in elastic parameters with depth. Changes *en masse* in the parameter PDFs of V_p, V_s and ρ enable us to delineate the reservoir with associated uncertainty.



(a) (b)

Figure 8. (a) PDF on the number of interfaces required to fit the observations. The true number is 11, and this plot displays the uncertainty in the parametrization and the rationale for using trans-D. (b) The implicitly sampled ML wavelet (top) spectrum from 500 models, and the true wavelet used to calculate the synthetics (bottom). The inversion was not privy to knowledge of the true source spectrum. This again shows that our ML source spectrum estimation scheme is working well.

Sears (2007). Their work inverts multicomponent data from the whole cable within a deterministic framework with careful pre-processing (which is significantly less work than the conventional seismic processing and migration workflow), and a three-stage inversion approach requiring inversion of short and then intermediate wavelengths and then again short wavelength features. They also require a good starting velocity model as do most deterministic FWI methods. We restrict ourselves to inverting only the pressure field (with subsurface mode conversions) from one raw shot gather with virtually no pre-processing of the raw, unmigrated data as we received it. As with the synthetic example, we transform raw T-X data (Fig. 9a) to the F-X domain (Fig. 9b), and then perform inversion at the same set of 18 frequencies from 5 to 13.5 Hz (Fig. 10) within a similar set of prior elastic parameters as for the synthetic case. We also start our inversion with a randomly selected two layered model within the prior bounds. For this shot, the inverted pressure field was recorded at hydrophones along the cable at a receiver spacing of 25 m, with the first offset at 246 m and the last at 3.8 km. It must be noted at the outset from Fig. 10 that the shot gather as we received it, shows 'ripples' with offset. This is the hallmark of some kind of wavenumber based noise removal filtering—this artefact is present in the T-X domain data and cannot be eliminated. Unfortunately the data are at this point 17 yr old (see Sears *et al.* 2010, for detailed acquisition parameters) and we could not go back any further to more 'raw' data. We may expect some degree of correlation in the inversion residuals right at the outset, as the forward model physics will only be able to reproduce the propagation part of the observed signal, not the subsequent (unknown) processing. Thus, unlike the



(a) (b)

Figure 9. (a) Alba OBC shot 1273. We invert offsets from 246 m to 3.8 km, at a hydrophone spacing of 25 m. (b) Like for the synthetic example, we Fourier transform every trace in panel (a) to the frequency domain. The resulting F-X amplitude spectrum is shown in dB relative to maximum amplitude near 20 Hz at 246 m offset.

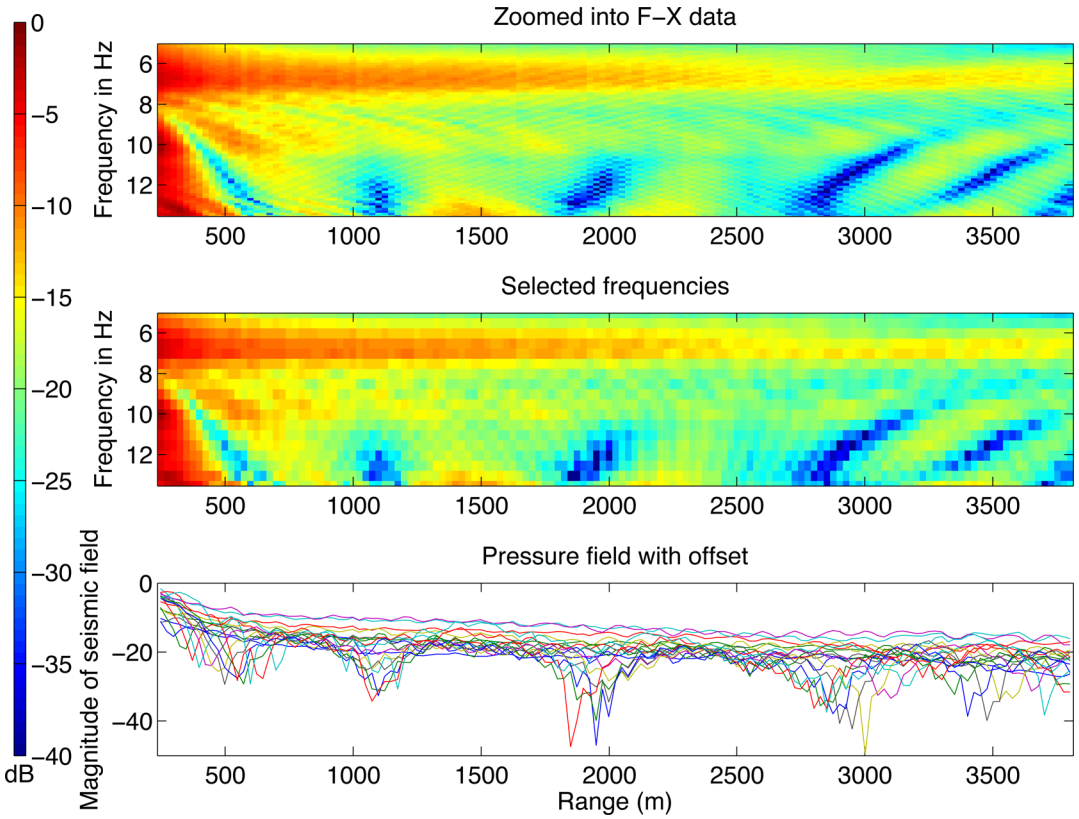


Figure 10. Selection of noisy frequency-offset data for Alba OBC hydrophone data inversion: Top: zooming into Fig. 9(b). Middle: decimated version of the top row, with frequencies sampled every 0.5 Hz. Bottom: the same as the middle figure, pressure field variation with offset at each of the frequencies for inversion. All amplitudes are in dB relative to the maximum amplitude near 20 Hz and 246 m offset. Note the presence of ripples in the observations with offset at all frequencies—either a filtering or acquisition artefact. We expect this ripple to manifest in correlated inversion residuals as it cannot be modelled by the physics of the wave equation.

synthetic noisy data inversion, we would expect our data residuals post inversion to be correlated, but like in the synthetic inversion, they should be unbiased and zero mean Gaussian.

As with the synthetic data, we first examine 500 randomly selected models from the posterior model ensemble $p(\mathbf{m}|\mathbf{d})$ for this inversion. Their data fits and residuals are shown in Figs 11 and 12. From the data fits of the pressure field amplitude, it is evident that our inversion scheme has worked well. The high frequencies have been fit remarkably, reproducing satisfactorily all the notches in the wavefield at frequencies higher than 9 Hz. For the lower frequencies, the effects of the input data ripple are evident, though the long wavelength part is again well accounted for. The 'dinosaur back' like features in the observed wavefield cannot be modelled, and lead to spatial correlation in the residuals at all frequencies

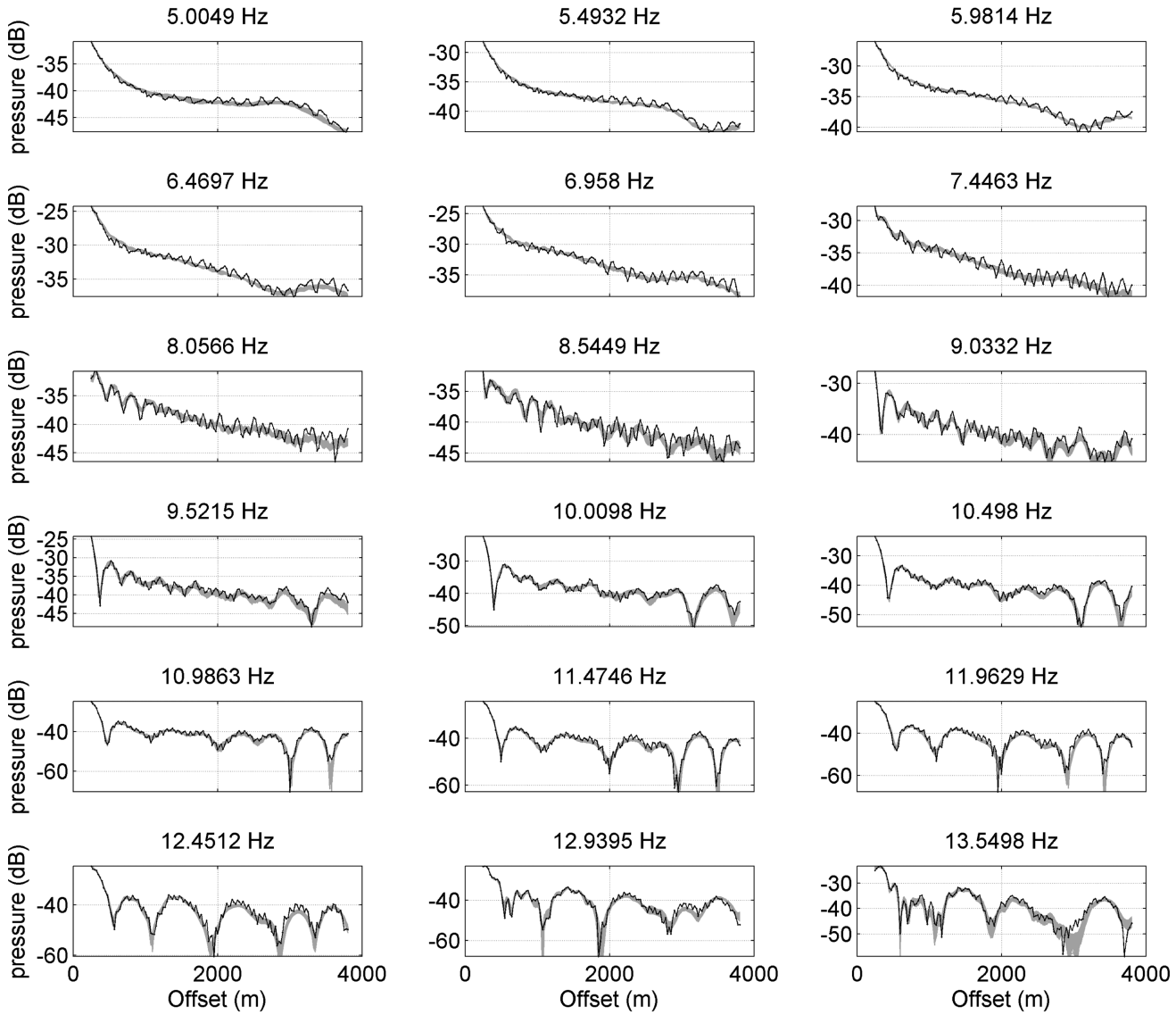


Figure 11. Pressure field amplitude with offset at the 18 inverted frequencies for the Alba OBC hydrophone data. The observations are shown in black. 500 responses due to models randomly selected from $p(\mathbf{m}|\mathbf{d})$ are shown in grey. Both calculated and observed fields are shown in dB relative to 1 unit of observed pressure. The data match is quite healthy, ignoring the expected ripple that cannot be modelled. Long wavelength trends at low frequencies are well reproduced, as are the notches at high frequencies.

as seen in Fig. 12. However, we expected this at the outset, and encouragingly, our normalized inversion residuals are unbiased and zero mean Gaussian (Fig. 13). Methods of sampling correlated noise during the inversion or using a sampled correlation matrix do exist (Dosso *et al.* 2006; Steininger *et al.* 2013; Dettmer *et al.* 2015; Gehrmann *et al.* 2015a,b), but given our experience with them (Ray *et al.* 2013b) and the challenges associated with them (Bodin *et al.* 2012; Dettmer *et al.* 2015), we consider these methods to be outside the scope of this paper. The posterior ensemble of models $p(\mathbf{m}|\mathbf{d})$ have been marginalized with depth and shown in Fig. 14. As with the synthetic example, darker shading in the first five panels from the left correspond to higher PDF values. We see that for both the compressional and shear velocities, the 90 per cent credible intervals (between the blue lines) are bounded quite tightly till about 1400 m depth. Thus it appears that the background velocity trends have been quite well estimated in the shallower sections. They are in good agreement with the work of Sears (2007). As expected, the PDFs of velocity get broader with depth, indicating a loss of resolution with depth. The seismic quality factors are not well resolved, as expected from the synthetic examples. The pronounced velocity changes between 1400 m and 1500 m depth correspond to the inferred depth of the Upper Oligocene horizon (Lonergan & Cartwright 1999, blue dashed horizontal line in the interface PDF plot). The reservoir sections in Area 12 correspond to depths between 1.8 to 2.5 km, as reported by (Sears *et al.* 2010). To our obvious delight, we do see the PDFs of V_p increase and those of ρ decrease, exhibiting similar behaviour as in our synthetic examples (Fig. 7), at an approximate depth of just over 2 km, indicative of reservoir. The results of (Sears *et al.* 2010) also show a low velocity interval and then a high velocity spike in V_p at about 2.5 km s^{-1}, indicative of reservoir in Area 12, at a depth below 2 km (magenta dashed horizontal line in the interface PDF plot). Further, the PDFs of V_s show a small spike near 2 km depth at approximately 1.5 km s^{-1} in the reservoir section, also reported by Sears *et al.* (2010).

The number of interfaces required to fit the data are shown in Fig. 15(a). We allowed for a max of 45 interfaces (46 layers), and

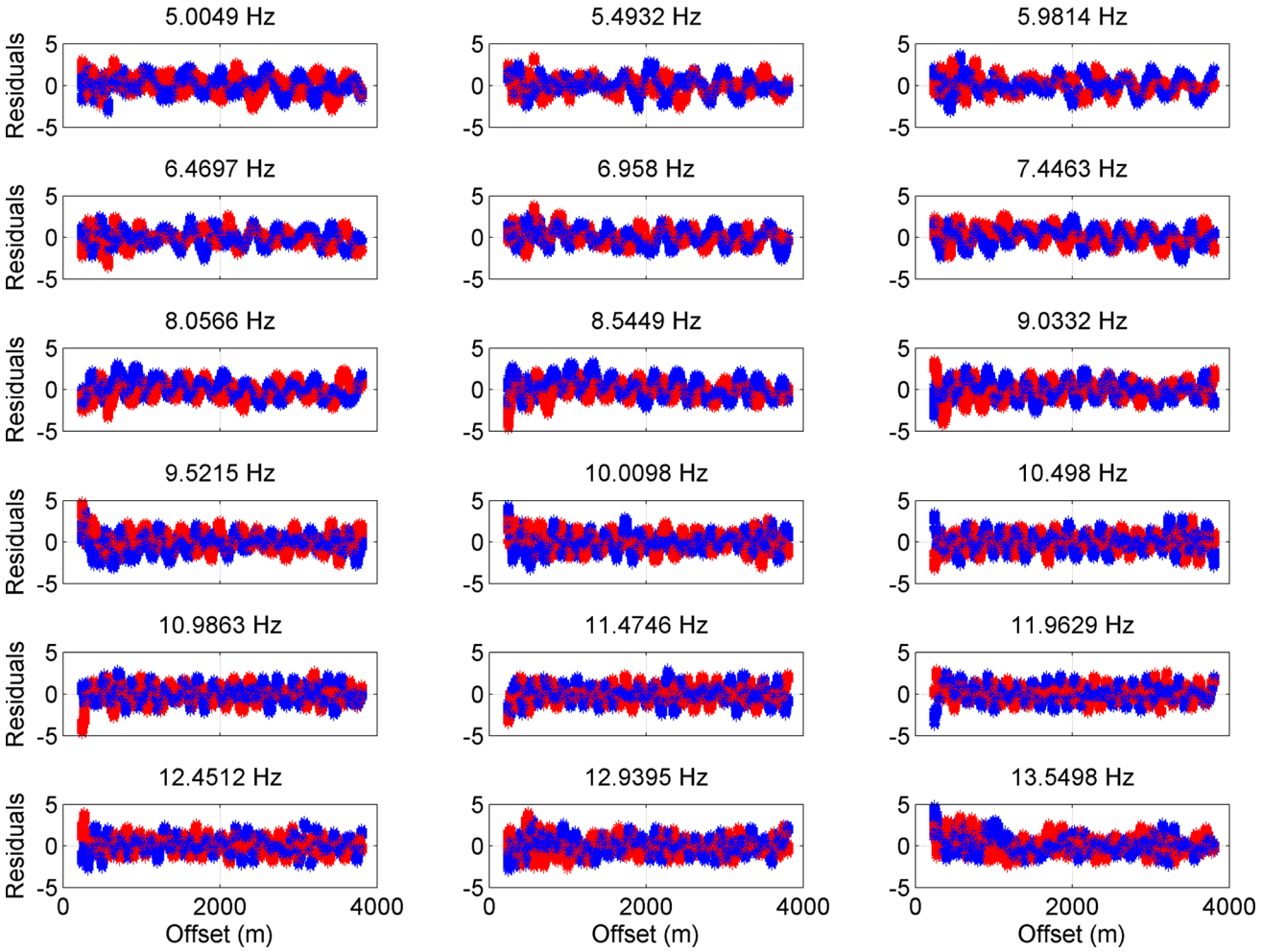


Figure 12. Inversion residuals with offset at 18 frequencies for the Alba OBC hydrophone data. 500 real (red) and imaginary (blue) residuals due to models randomly selected from $p(\mathbf{m}|\mathbf{d})$ are shown. The complex data residuals have been normalized by the implicitly sampled ML data error. The residuals are correlated as expected from Fig. 10, but importantly, they are unbiased about zero.

find that the mode is at 39 interfaces, with a spread around this value indicating the non-uniqueness in the number of layers. Fig. 15(b) displays a spectral comparison between a deghosted wavelet used for standard post-migration amplitude versus offset work obtained from a far field signature (Sears 2007), and the wavelet we sampled implicitly using eq. (15) from 500 models. The results are similar, with our sampled wavelet showing more 'detail' between 7 Hz and 10 Hz. If we had used the provided wavelet, it is likely that this would have biased our results in some manner.

3.4.1 Time domain data fit for Alba shot gather

Since we collect and are familiar with data in the time domain, we felt it would be useful to transfer our frequency domain results to the time domain. From the point of view of physics, both domains offer equivalent descriptions of the same wave phenomena. However, we only inverted a discrete set of frequencies and found the ML wavelet for this set of frequencies between 5 Hz and 13.5 Hz. Therefore, the forward modelled data for any inverted model in $p(\mathbf{m}|\mathbf{d})$ do not contain the full spectrum of responses for inverse Fourier transformation to the time domain. Under a certain set of assumptions however, we can forward calculate a shot gather in time, from any of the models in $p(\mathbf{m}|\mathbf{d})$.

We assume that the inverted models at our discrete set of frequencies f, will provide 'reasonable' forward responses for *all* frequencies in the band [min (f), max (f)]. After computing the ML wavelet at frequencies *not* inverted, we can forward calculate the pressure field at these frequencies using eqs (15) and (27). We calculated these responses at a frequency interval of 0.12 Hz, corresponding to a record length of 8.19 s, and constructed the frequency response from 0.12 Hz till 40 Hz. We multiplied these responses within [0.12, 40] Hz with a zero-phase bandpass filter corresponding to the range [5, 13.5] Hz since we assume that our frequency responses are valid only in this range. We then inverse Fourier transformed these responses at every hydrophone location to produce a time domain shot gather. Finally for comparison's sake, we applied the same bandpass to the input shot gather, and display both the input gather and the modelled time domain gather (per inverted model) using the same colour scale, in the same frequency band, side by side. The results are shown in Fig. 16 for two models, together with the mean calculated data from 60 inverted models. Interfering events associated with the reservoir depths are found 1.6 s –1.7 s onwards at the closest offsets.

These 'butterfly' plots show a remarkable fit in the time domain, especially the headwaves, diving waves and early near offset reflections, indicating that our inversion scheme together with the elastic model physics work well in reproducing reflections, wide angle

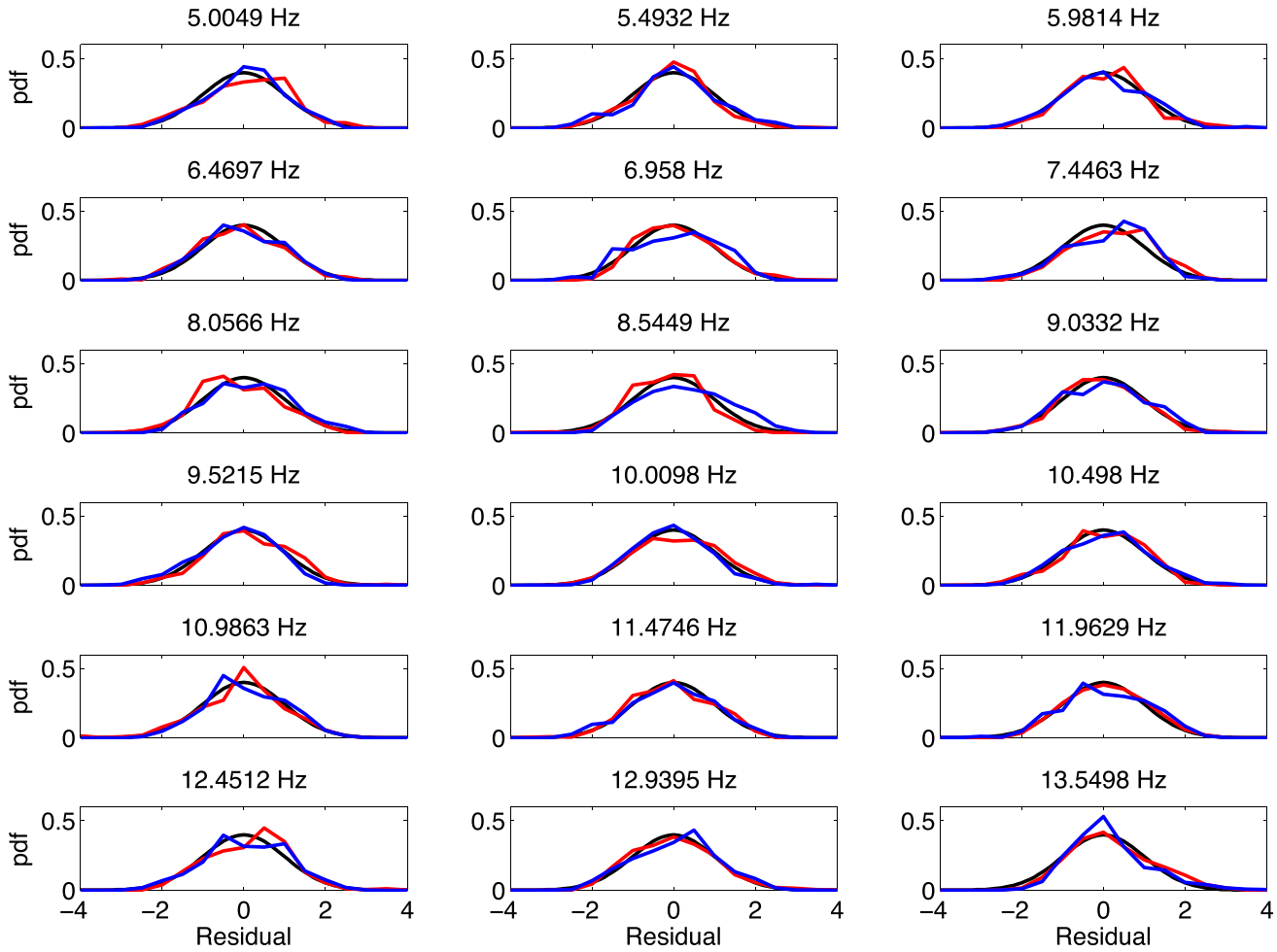


Figure 13. PDFs of normalized inversion residuals for the Alba OBC hydrophone data obtained from 500 models. For each frequency, an analytic standard Gaussian PDF is shown in black, and the real and imaginary residuals are shown in red and blue, respectively. The complex data residuals have been normalized by the implicitly sampled ML data error. To a large degree, we are not far removed from the assumption of Gaussian data error.

refractions, multiples as well as mode conversions. This also confirms that we do not need finer resolution in frequency for our inversion purposes - though we would probably have produced an even better time domain match if we had.

4 CONCLUSIONS

We have successfully implemented a flexibly parametrized trans-D Bayesian inversion using reversible jump McMC for the frequency domain elastic FWI problem. We have also demonstrated the equivalence of time domain and frequency domain approaches by showing data fits in both domains. The flexibility in model parametrization and the absence of arbitrary correlation or regularization are essential for a data-driven inversion of anomalous elastic parameters and the associated background, especially when starting with poor prior knowledge of the subsurface. The noise model we describe and the ML methodology for estimating source and noise parameters (Section 2) is quite general and applicable to wave propagators for 2-D or 3-D earth models. Synthetic studies were carried out to understand the model space and uncertainty associated with a deep target in a shallow water environment. As established in Ray *et al.* (2014), following the spatial changes in posterior PDFs of model parameters (e.g. towards high velocities or low densities) is far more valuable in delineating anomalies within the earth as opposed to following the changes in either a single model (as in deterministic inversion) or a statistic of the posterior PDF (such as the mode). Our inversion of shallow water OBC data from the Alba survey was successfully able to recover both the background V_p and V_s, with a good indication of the depth of the reservoir as well as the Upper Oligocene horizon in the Area 12 part of the field. Comparison with previous work (Sears 2007; Sears *et al.* 2010) showed good agreement between the Bayesian and deterministic results. It must be noted that unlike (Sears *et al.* 2010) we did not require an accurate start model, nor careful data pre-conditioning (Virieux & Operto 2009) to arrive at our results, which were in good agreement with nearby well log data.

Our computations required 100 CPU nodes within a high performance computing cluster to sample the trans-D posterior model PDFs. As this is a first effort, we did not try to optimize our resource utilization. The synthetic example required 1 day to converge to a stationary posterior model sample, whereas the real data inversion required 3 days to achieve stationarity. More details on convergence can be found in the Appendix B6. Though not intuitive, it is well known that McMC methods require longer sampling times for data with very high SNR (or artificially high SNR after processing as in the case of Alba), as discussed by Brooks & Neil Frazer (2005). We recommend an approach of being as close to the raw gathers as

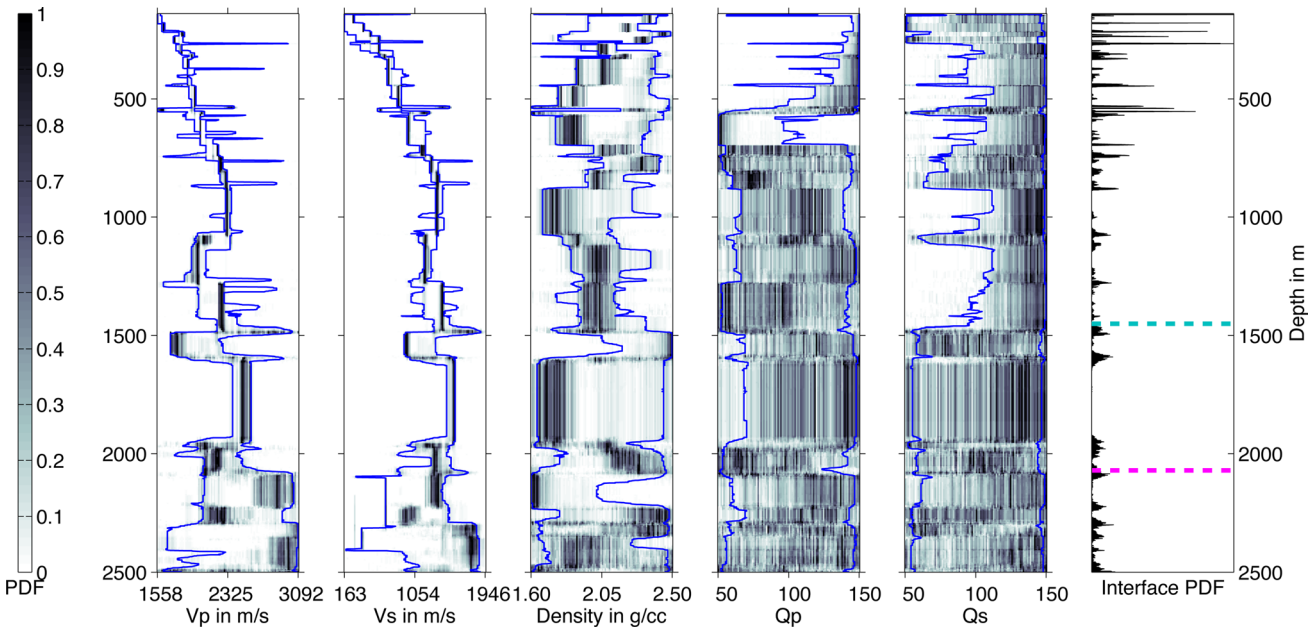
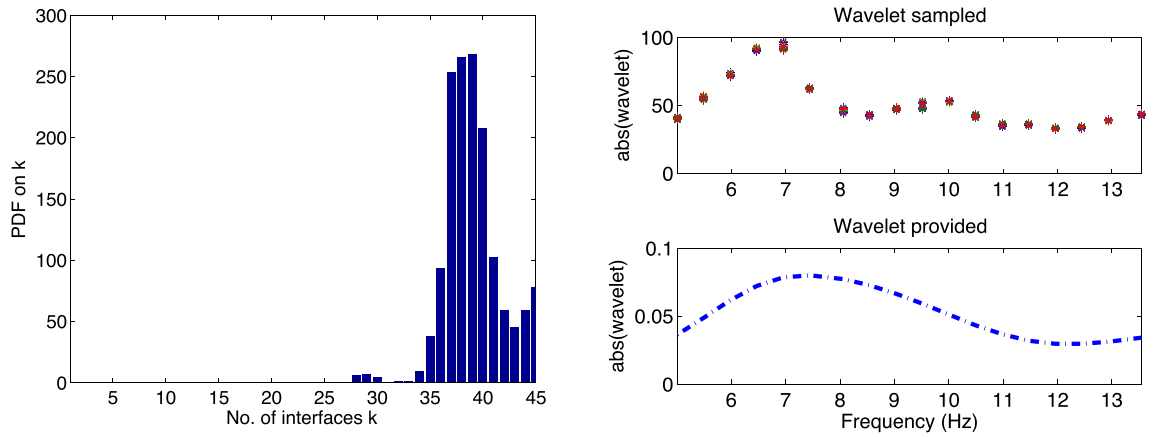


Figure 14. Marginal posterior PDFs with depth for the Alba OBC data inversion. Darker shading corresponds to higher probability density in the first five panels from the left. The two thin blue lines at every depth represent the 90 per cent Bayesian credible interval. The last panel to the right shows the probability that the data require an interface to be present in the model, that is, require a change in elastic parameters with depth. Changes *en masse* in the parameter PDFs of V_p, V_s and ρ enable us to delineate a 'reservoir like interval' just below 2000 m depth (magenta dashed line in rightmost column, as also inverted by Sears *et al.* (2010)), with associated uncertainty and a thickness of about 100 m. Also well estimated is a sharp velocity change corresponding to the Upper Oligocene horizon (dashed green line in rightmost column) at a depth of ∼1.5 km.



(a) (b)

Figure 15. (a) PDF on the number of interfaces required to fit the Alba OBC observations. This plot displays the uncertainty in the parametrization and the rationale for using trans-D. (b) The implicitly sampled ML wavelet (top) spectrum from 500 models, and the source deghosted wavelet provided from a far-field signature (bottom). If we had used the provided wavelet, it is likely that this would have biased our results in some manner, further showing the usefulness of the ML estimated source spectrum.

is possible, removing obvious outlier data which would not satisfy an L2 norm and applying minimal processing to the data, so as to remain true to our likelihood function (29). As discussed recently by Dettmer *et al.* (2015) for seismic receiver function inversion, it is better to invert data directly than to process it for making inferences about the earth, as processing can lead to a true model likelihood that cannot accurately be represented by our theory. Our forward code was not particularly optimized for parallel computation, and with rapid advances in GPU computing (e.g. Komatitsch *et al.* 2010; Johnsen *et al.* 2015) as well as machine learning methodologies, together with the application of adjoint based methods (Plessix 2006) to provide local gradients (as discussed in the Appendix B6) we are already at the point where it is possible to postulate solving geophysical Bayesian inversion problems with 2-D or 3-D propagators and/or earth models (e.g. Bodin *et al.* 2009; Cordua *et al.* 2012; Dettmer & Dosso 2013; Ray *et al.* 2014; Rosas-Carbajal *et al.* 2015).

ACKNOWLEDGEMENTS

We would like to thank Chevron Energy Technology Company for providing permission to publish this work. Alistair Harding (UC San Diego) and Kurt Nihei (Lawrence Berkeley National Labs) are

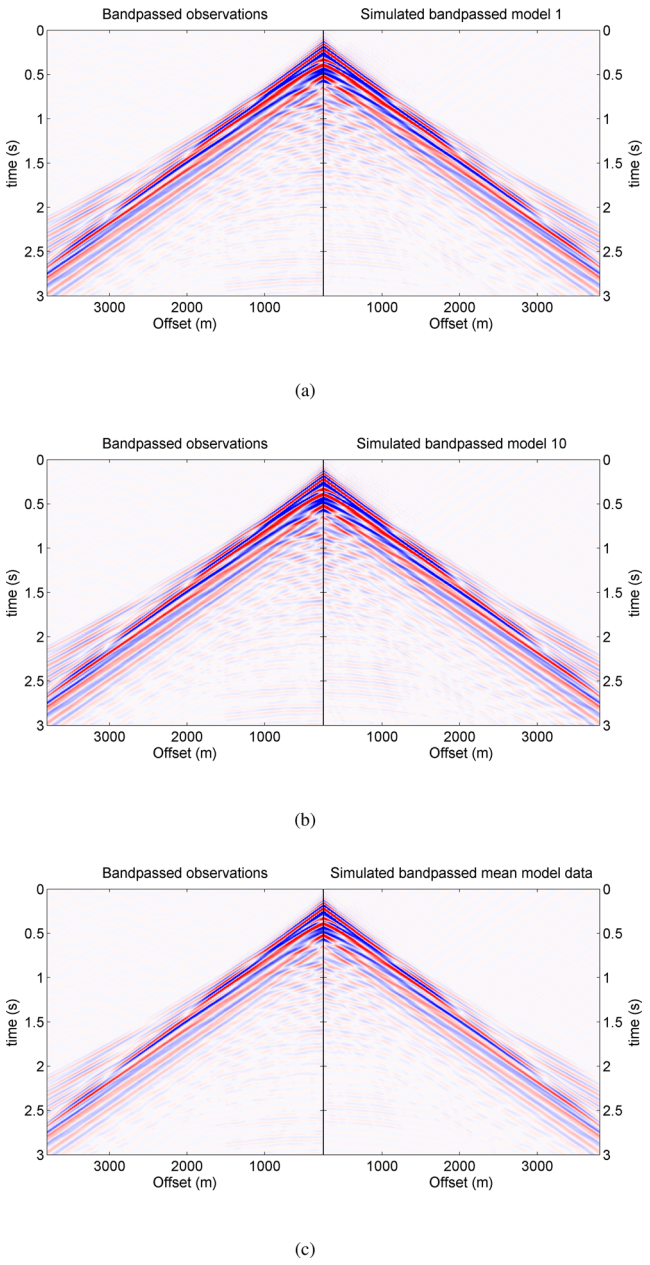


Figure 16. (a,b) Bandpassed time domain comparison of observed shot gather 1273 from the Alba field OBC data (left) and similarly bandpassed response (right) from one model randomly selected from $p(\mathbf{m}|\mathbf{d})$. (c) Bandpassed time domain comparison of observed shot gather (left) and the mean of similarly bandpassed responses (right) from 60 models randomly selected from $p(\mathbf{m}|\mathbf{d})$.

thanked for deeply insightful discussions into seismic wave propagation. John Washbourne and Joe Stefani at Chevron ETC are thanked for providing access to their elastic full waveform modeller. Vanessa Brown and Jake Deeds at Chevron ETC are thanked for providing access to the Alba OBC data. Chevron Upstream Europe are acknowledged for advancing research by allowing their data to be used for publication purposes. Malcolm Sambridge and Jan Dettmer at Australian National University provided valuable advice on implementing an efficient parallel tempering sampler. Finally, we thank Malcolm Sambridge and Alberto Malinverno for very constructive reviews which served to make this manuscript significantly more complete in its exposition of the trans-D method.

REFERENCES

Albert, I., Espié, E., de Valk, H. & Denis, J.-B., 2011. A Bayesian evidence synthesis for estimating campylobacteriosis prevalence, *Risk Anal.,* **31**(7), 1141–1155.

Aleardi, M. & Mazzotti, A., 2014. 1D Elastic FWI and Uncertainty Estimation by Means of a Hybrid Genetic Algorithm-Gibbs Sampler Approach, in *EAGE Conference & Exhibition,* Amsterdam, doi:10.3997/2214-4609.20140858.

Alkhalifah, T., 2014. Monofrequency waveform acquisition and inversion: a new paradigm, in *SEG Technical Program Expanded Abstracts,* pp. 1002–1006.

Bernardo, J.M. & Smith, A.F.M., 1994. *Bayesian Theory,* Wiley.

Biondi, B. & Almonin, A., 2014. Simultaneous inversion of full data bandwidth by tomographic full waveform inversion (TFWI), *Geophysics,* **79**(3), WA129–WA140.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178**(3), 1411–1436.

Bodin, T., Sambridge, M. & Gallagher, K., 2009. A self-parametrizing partition model approach to tomographic inverse problems, *Inverse Probl.,* **25**(5), 055009, doi:10.1088/0266-5611/25/5/055009.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117**(B2), doi:10.1029/2011JB008560.

Bodin, T., Yuan, H. & Romanowicz, B., 2013. Inversion of receiver functions without deconvolution–application to the Indian craton, *Geophys. J. Int.,* **196,** 1025–1033.

Brillinger, D., 1974. Fourier analysis of stationary processes, *Proc. IEEE,* **62**(12), 1628–1643.

Brodie, R.C. & Sambridge, M., 2012. Transdimensional Monte Carlo Inversion of AEM Data, in *22nd International Geophysical Conference and Exhibition, Brisbane, Australia,* no. 1.

Brooks, B.A. & Neil Frazer, L., 2005. Importance reweighting reduces dependence on temperature in Gibbs samplers: an application to the coseismic geodetic inverse problem, *Geophys. J. Int.,* **161**(1), 12–20.

Chib, S. & Greenberg, E., 1995. Understanding the Metropolis–Hastings algorithm, *Am. Stat.,* pp. 327–335.

Constable, S., 2013. A Young Person's Guide to Regularized Inversion, in *ASEG Workshop,* Melbourne.

Constable, S.C., Parker, R.L. & Constable, C.G., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics,* **52**(3), 289–300.

Cordua, K.S., Hansen, T.M. & Mosegaard, K., 2012. Monte Carlo full-waveform inversion of crosshole GPR data using multiple-point geostatistical a priori information, *Geophysics,* **77**(2), H19–H31.

Denison, D.G.T., 2002. *Bayesian Methods for Nonlinear Classification and Regression,* vol. 386, John Wiley & Sons.

Dettmer, J. & Dosso, S.E., 2012. Trans-dimensional matched-field geoacoustic inversion with hierarchical error models and interacting Markov chains, *J. acoust. Soc. Am.,* **132**(4), 2239–2250.

Dettmer, J. & Dosso, S.E., 2013. Probabilistic two-dimensional water-column and seabed inversion with self-adapting parameterizations, *J. acoust. Soc. Am.,* **133**(5), 2612–2623.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128**(6), 3393–3405.

Dettmer, J., Dosso, S.E., Bodin, T., Stipčević, J. & Cummins, P.R., 2015. Direct-seismogram inversion for receiver-side structure with uncertain source–time functions, *Geophys. J. Int.,* **203,** 1373–1387.

Dosso, S.E. & Wilmut, M.J., 2012. Maximum-likelihood and other processors for incoherent and coherent matched-field localization, *J. acoust. Soc. Am.,* **132,** 2273–2285.

Dosso, S.E., Nielsen, P.L. & Wilmut, M.J., 2006. Data error covariance in matched-field geoacoustic inversion, *J. acoust. Soc. Am.,* **119**(1), 208–219.

Dosso, S.E., Holland, C.W. & Sambridge, M., 2012. Parallel tempering for strongly nonlinear geoacoustic inversion., *J. acoust. Soc. Am.,* **132**(5), 3030–3040.

Dosso, S.E., Dettmer, J., Steininger, G. & Holland, C.W., 2014. Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation, *Inverse Probl.,* **30**(11), 114018, doi:10.1121/1.4899489.

Earl, D.J. & Deem, M.W., 2005. Parallel tempering: theory, applications, and new perspectives., *Phys. Chem. Chem. Phys.,* **7**(23), 3910–3916.

Fichtner, A. & Trampert, J., 2011. Resolution analysis in full waveform inversion, *Geophys. J. Int.,* **187**(3), 1604–1624.

Fichtner, A., Kennett, B.L.N., Igel, H. & Bunge, H.-P., 2009. Full seismic waveform tomography for upper-mantle structure in the Australasian region using adjoint methods, *Geophys. J. Int.,* **179**(3), 1703–1725.

Fontaine, F.R., Barruol, G., Tkalčić, H., Wölbern, I., Rümpker, G., Bodin, T. & Haugmard, M., 2015. Crustal and uppermost mantle structure variation beneath La Réunion hotspot track, *Geophys. J. Int.,* **203**(1), 107–126.

Galetti, E., Curtis, A., Meles, G.A. & Baptie, B., 2015. Uncertainty loops in travel-time tomography from nonlinear wave physics, *Phys. Rev. Lett.,* **114**(14), 1–5.

Gehrmann, R.A.S., Dettmer, J., Schwalenberg, K., Engels, M., Dosso, S.E. & Özmaral, A., 2015a. Trans-dimensional Bayesian inversion of controlled-source electromagnetic data in the German North Sea, *Geophys. Prospect.,* **63,** 1314–1333.

Gehrmann, R.A.S., Schwalenberg, K., Riedel, M., Spence, G.D., Spiess, V. & Dosso, S.E., 2015b. Bayesian inversion of marine controlled source electromagnetic data offshore Vancouver Island, Canada, *Geophys. J. Int.,* **204**(1), 21–38.

Gelman, A. & Rubin, D.B., 1992. Inferences from iterative simulations using multiple sequences, *Stat. Sci.,* **7**(4), 457–511.

Geyer, C.J. & Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Stat.,* **21,** 359–373.

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood, in *Proceedings of the 23rd Symposium on the Interface,* pp. 156–163, ed. Keramidas, E.M., Interface Foundation, Fairfax Station.

Gilks, W.R., Richardson, S. & Spiegelhalter, D.J., 1996. Introducing Markov chain Monte Carlo, in *Markov Chain Monte Carlo in Practice,* eds Gilks, W.R., Richardson, S. & Spiegelhalter, D.J., pp. 1–19, Chapman and Hall.

Green, P.J., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika,* **82**(4), 711–732.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *Geophys. J. Int.,* **203,** 972–1000.

Johnsen, T., Loddoch, A. & Liu, W., 2015. *System and Method of Implementing Finite Difference Time Domain Models with Multiple Accelerated Processing Components (APCs)*.

Kamath, N., Tsvankin, I. & Esteban, D., 2015. Elastic full-waveform inversion for a synthetic VTI model from Valhall field, in *SEG Technical Program Expanded Abstracts*, pp. 457–461.

Kennett, B., 2009. *Seismic Wave Propagation in Stratified Media,* ANU E Press.

Kennett, B.L.N., 1986. Seismic wave propagation in stratified media, *Geophys. J. R. astr. Soc.,* **86**(1), 219–220.

Komatitsch, D., Erlebacher, G., Göddeke, D. & Michéa, D., 2010. High-order finite-element seismic wave propagation modeling with MPI on a large GPU cluster, *J. Comput. Phys.,* **229**(20), 7692–7714.

Liang, F., Liu, C. & Carroll, R., 2011. *Advanced Markov Chain Monte Carlo Methods: Learning From Past Samples,* vol. 714, Wiley-Blackwell.

Lonergan, L. & Cartwright, J.A., 1999. Polygonal faults and their influence on deep-water sandstone reservoir geometries, Alba Field, United Kingdom Central North Sea, *AAPG Bull.,* **83**(3), 410–432.

MacKay, D.J.C., 2003. *Information Theory, Inference and Learning Algorithms,* Cambridge Univ. Press.

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.,* **151**(3), 675–688.

Malinverno, A. & Leaney, S., 2000. A Monte Carlo method to quantify uncertainty in the inversion of zero-offset VSP data, in *SEG Annual Meeting,* no. 3, pp. 2393–2396.

Malinverno, A. & Leaney, W.S., 2005. Monte-Carlo Bayesian look-ahead inversion of walkaway vertical seismic profiles, *Geophys. Prospect.,* **53**(5), 689–703.

Mecklenbrauker, C.F. & Gerstoft, P., 2000. Objective functions for ocean acoustic inversion derived by likelihood methods, *J. Comput. Acoust.,* **8**(2), 259–270.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory,* vol. 45, Academic Press.

Minsley, B.J., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data, *Geophys. J. Int.,* **187**(1), 252–272.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Newton, S.K. & Flanagan, K.P., 1993. The Alba Field: evolution of the depositional model, in *Proceedings of the 4th Conference on Petroleum Geology of Northwest Europe*, Geological Society, London, pp. 161–171.

Parker, R.L., 1980. The inverse problem of electromagnetic induction : existence and construction of solutions based on incomplete data, *J. geophys. Res.,* **85,** 4421–4428.

Piana Agostinetti, N. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.,* **181,** 858–872.

Plessix, R.E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophys. J. Int.,* **167**(2), 495–503.

Ray, A., 2014. Trans-dimensional Bayesian Inversion of Controlled Source Electromagnetic Data, *Ph.D. thesis*, University of California, San Diego.

Ray, A. & Key, K., 2012. Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm, *Geophys. J. Int.,* **191,** 1135–1151.

Ray, A., Alumbaugh, D.L., Hoversten, G.M. & Key, K., 2013a. Robust and accelerated Bayesian inversion of marine controlled-source electromagnetic data using parallel tempering, *Geophysics,* **78**(6), E271–E280.

Ray, A., Key, K. & Bodin, T., 2013b. Hierarchical Bayesian inversion of marine CSEM data over the Scarborough gas field—A lesson in correlated noise, in *SEG Technical Program Expanded Abstracts,* no. 1, pp. 723–727, Houston.

Ray, A., Key, K., Bodin, T., Myer, D. & Constable, S., 2014. Bayesian inversion of marine CSEM data from the Scarborough gas field using a transdimensional 2-D parametrization, *Geophys. J. Int.,* **199,** 1847–1860.

Rosas-Carbajal, M., Linde, N., Peacock, J., Zyserman, F., Kalscheuer, T. & Thiel, S., 2015. Probabilistic 3-D time-lapse inversion of magnetotelluric data: application to an enhanced geothermal system, *Geophys. J. Int.,* **203**(3), 1946–1960.

Sambridge, M., 2013. A Parallel Tempering algorithm for probabilistic sampling and multimodal optimization, *Geophys. J. Int.,* **196**(1), 357–374.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Trans-dimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167**(2), 528–542.

Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A,* **371**(1984), 20110547, doi:10.1098/rsta.2011.0547.

Scales, J.A. & Sneider, R., 1997. To Bayes or not to Bayes?, *Geophysics,* **62**(4), 1045–1046.

Schmidt, H. & Tango, G., 1986. Efficient global matrix approach to the computation of synthetic seismograms, *Geophys. J. Int.,* **84**(2), 331–359.

Sears, T.J., 2007. Elastic full waveform inversion of multi-component ocean-bottom cable seismic data, *PhD thesis,* University of Cambridge.

Sears, T.J., Barton, P.J. & Singh, S.C., 2010. Elastic full waveform inversion of multicomponent ocean-bottom cable seismic data: application to Alba Field, UK North Sea, *Geophysics,* **75**(6), R109–R119.

Sen, M.K. & Stoffa, P.L., 2013. *Global Optimization Methods in Geophysical Inversion,* Cambridge Univ. Press.

Sirgue, L. & Pratt, R.G., 2004. Efficient waveform inversion and imaging: A strategy for selecting temporal frequencies, *Geophysics,* **69**(1), 231–248.

Sisson, S.A., 2005. Transdimensional Markov Chains, *J. Am. Stat. Assoc.,* **100**(471), 1077–1089.

Steininger, G., Dettmer, J., Dosso, S.E. & Holland, C.W., 2013. Trans-dimensional joint inversion of seabed scattering and reflection data, *J. acoust. Soc. Am.,* **133**(3), 1347–1357.

Swendsen, R.H. & Wang, J.S., 1987. Nonuniversal Critical Dynamics in Monte Carlo Simulations, *Phys. Rev. Lett.,* **58**(2), 86–88.

Tarantola, A., 1984. Inversion of seismic reflection data in the acoustic approximation, *Geophysics,* **49**(8), 1259–1266.

Tarantola, A. & Valette, B., 1982. Inverse problems = quest for information, *J. Geophys.,* **50,** 159–170.

Timm, N.H., 2002. *Applied Multivariate Analysis,* Springer.

Ursin, B., 1983. Review of elastic and electromagnetic wave propagation in horizontally layered media, *Geophysics,* **48**(8), 1063–1081.

Valentine, A.P. & Trampert, J., 2015. The impact of approximations and arbitrary choices on geophysical images, *Geophys. J. Int.,* **204**(1), 59–73.

Vigh, D., Jiao, K., Watts, D. & Sun, D., 2014. Elastic full waveform inversion application using multicompoment measurements of seismic data collection, *Geophysics,* **79**(2), R63–R77.

Virieux, J. & Operto, S., 2009. An overview of full-waveform inversion in exploration geophysics, *Geophysics,* **74**(6), WCC1–WCC26.

Wapenaar, C.P.A., 1996. Inversion versus migration: a new perspective to an old discussion, *Geophysics,* **61**(3), 804–814.

APPENDIX A: MAXIMUM LIKELIHOOD

In order to derive the ML objective function (29) of Section 2 we begin with the model likelihood $\mathcal{L}(\mathbf{m})$ for a circularly symmetric Gaussian complex random variable:

$$\mathcal{L}(\mathbf{m}) = p(\mathbf{d}|\mathbf{m}) = \prod_{l=1}^{n_f}\left(\pi\frac{\sigma_l^2}{n_t}\right)^{-n_r}\exp\left(-\frac{[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]}{\sigma_l^2/n_t}\right), \quad \text{(A1)}$$

$$-\log\mathcal{L} = \sum_{l=1}^{n_f}\left\{\log\left(\pi^{n_r}\left(\frac{\sigma_l^2}{n_t}\right)^{n_r}\right) + \frac{n_t}{\sigma_l^2}[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]\right\}, \quad \text{(A2)}$$

$$= \sum_{l=1}^{n_f}\left\{n_r\log\pi + n_r\log\left(\frac{\sigma_l^2}{n_t}\right) + \frac{n_t}{\sigma_l^2}[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]\right\}. \quad \text{(A3)}$$

For minimum of the negative log likelihood or square misfit, that is, maximum of the log likelihood,

$$\frac{\partial(-\log\mathcal{L})}{\partial\sigma_l} = n_r\frac{n_t}{\sigma_l^2}\frac{2\sigma_l}{n_t} - \frac{2}{\sigma_l^3}n_t[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})] = 0, \quad \text{(A4)}$$

$$\Rightarrow \frac{\sigma_l^2}{n_t} = \frac{1}{n_r}[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})], \quad \text{(A5)}$$

Substituting eq. (A5) into eq. (A3),

$$-\log\mathcal{L} = \sum_{l=1}^{n_f}\left[n_r\log\pi + n_r\log\left\{[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]\right\} + n_r\right], \quad \text{(A6)}$$

$$= \sum_{l=1}^{n_f} n_r\log\left\{[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]\right\}, \quad \text{(A7)}$$

ignoring the constant terms not dependent on the data residual, we have

$$-\log\mathcal{L}(\mathbf{m}) = n_r\sum_{l=1}^{n_f}[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]^\dagger[\mathbf{d_l}-\mathbf{f_l}(\mathbf{m})]. \quad \text{(A8)}$$

APPENDIX B: MATHEMATICAL DETAILS FOR THE TRANS-DIMENSIONAL ALGORITHM

B1 Trans-dimensional Bayesian inversion

In the simplest terms, the objective of trans-dimensional Bayesian inversion is to sample the model PDF given by eq. (1). This is achieved with the reversible jump algorithm by drawing candidate earth models from a proposal PDF. These models are then examined to see if they fall within geophysically sensible uniform bounds. If a proposed model falls outside the prior PDF, it is rejected and the Markov chain retains the previous model as the next model. If a proposed model is within the prior bounds, an acceptance probability is calculated using a ratio of the proposal probability, the prior probability and the likelihood of the candidate model with respect to the previous model. The proposed model is either accepted with the calculated probability and it becomes the next model in the chain, or it is rejected and the previous model is retained as the next in the chain. Complete details of this process are given in the following section for the interested reader. As the algorithm proceeds, hundreds of thousands of models are sampled, with a data-driven addition or deletion of layer interfaces ('birth/death' in RJ-McMC parlance), such that a chain of models, most of which fit the data well within the noise, are retained at the end. To ensure thorough sampling of this multidimensional parameter space, a parallel tempering algorithm (Swendsen & Wang 1987; Geyer 1991) has been used. Details of our implementation can be found in Ray *et al.* (2013a) with a particularly illuminating discussion given in Sambridge (2013).

B2 The prior probability

The prior PDF contains information on our knowledge about the subsurface independent of the survey data. This can be based on data from well logs, seismic surveys, etc. In the trans-dimensional formulation, we split the prior into 2 parts. One part contains information about the number of interfaces k in the model, $p(k)$. The other part $p(\mathbf{m_k}|k)$ in our particular case, contains information about the physical parameters associated with a model $\mathbf{m_k}$ of dimension k, such as where these interfaces are in depth z, and what the $k+1$ elastic layer parameters of these cells are V_p, v_s, ρ, Q_p, Q_s. Please note that since we parametrize with interfaces, for n interfaces we will have $n+1$ layers including a terminating half-space. In practice, we sample Poisson's ratio ν uniformly and calculate V_s for the given value of V_p and Poisson's ratio. Using the chain rule of probabilities, we write

$$p(\mathbf{m}) = p(\mathbf{m_k}, k) = p(\mathbf{m_k}|k)\times p(k). \quad \text{(B1)}$$

where

$$\mathbf{m_k} = [\mathbf{z}, \mathbf{V_p}, \boldsymbol{\nu}, \boldsymbol{\rho}, \mathbf{Q_p}, \mathbf{Q_s}], \quad \text{(B2)}$$

$$= [\mathbf{z}, \mathbf{m}_e], \quad \text{(B3)}$$

where

$$\mathbf{z} = [z_1, \ldots, z_k], \tag{B4}$$

$$\mathbf{V_p} = [V_{p_1}, \ldots, V_{p_{k+1}}], \tag{B5}$$

$$\boldsymbol{\nu} = [\nu_1, \ldots, \nu_{k+1}], \tag{B6}$$

$$\boldsymbol{\rho} = [\rho_1, \ldots, \rho_{k+1}], \tag{B7}$$

$$\mathbf{Q_p} = [Q_{p_1}, \ldots, Q_{p_{k+1}}], \tag{B8}$$

$$\mathbf{Q_s} = [Q_{s_1}, \ldots, Q_{s_{k+1}}]. \tag{B9}$$

We use a uniform prior on k, given by

$$p(k) = \begin{cases} \frac{1}{k_{\max} - k_{\min} + 1} & \text{if } k_{\min} \leq k \leq k_{\max} \\ 0 & \text{else} \end{cases}. \tag{B10}$$

We assume no *a priori* knowledge between the locations of interfaces in the model and the elastic parameters $\mathbf{m}_e$. These random variables are independent and therefore their PDFs can be separated in the following product form,

$$p(\mathbf{m_k}|k) = p(\mathbf{z}|k)p(\mathbf{m}_e|k). \tag{B11}$$

Interfaces can be located anywhere in the depth range $[z_{\min}, z_{\max}]$. A given interface can be at any of k points within this depth. This defines $k+1$ subintervals, the sum of which is equal to $\Delta z = z_{\max} - z_{\min}$. The probability distribution of the segment lengths for the analogous case of a rope of unit length cut randomly at k points, is the Dirichlet distribution (e.g. Albert *et al.* 2011; Steininger *et al.* 2013). For the case in which all segment lengths are equally probable, the appropriate form of $p(\mathbf{z}|k)$ transformed to a length interval Δz is derived in detail by Dosso *et al.* (2014) and is given as:

$$p(\mathbf{z}|k) = \frac{k!}{\Delta z^k}. \tag{B12}$$

This can intuitively be understood as there being $k!$ possible arrangements for interfaces uniformly distributed within a length Δz. Assuming that all $k+1$ elastic parameters within a given model are uniformly distributed, independent of each other, we can write

$$p(\mathbf{m}_e|k) = \begin{cases} \left[\frac{1}{\Delta m_e}\right]^{k+1} & \text{if } m_{e_{\min}} \leq m_e \leq m_{e_{\max}} \\ 0 & \text{else} \end{cases}, \tag{B13}$$

where m_e is any of the elastic parameters within a layer, and

$$\Delta m_e = \Delta V_p \Delta \nu \Delta \rho \Delta Q_p \Delta Q_s, \tag{B14}$$

corresponding to uniform property ranges of distribution within each layer. To obtain the explicit expression for the prior model probability, we write $\Delta k = k_{\max} - k_{\min} + 1$ and substitute eqs (B10)–(B13) into (B1) to get

$$p(\mathbf{m}) = \begin{cases} \frac{k!}{\Delta z^k} \cdot \frac{1}{\Delta m_e^{k+1}} \cdot \frac{1}{\Delta k} & \text{if } z \in [z_{\min}, z_{\max}], \\ & m_e \in [m_{e_{\min}}, m_{e_{\max}}], \\ & \forall k \in [k_{\min}, k_{\max}] \\ 0 & \text{else} \end{cases}. \tag{B15}$$

We mention here that we can parametrize V_p such that we sample the range of V_p in the logarithm of the range. This can facilitate sampling of a large range of V_p from the seafloor down to basement depth. We found that this did not have a large effect on our posterior PDF.

B3 Metropolis–Hastings algorithms and the acceptance probability

What guides an McMC sampler like the Metropolis–Hastings (MH) algorithm (Hastings 1970) to convergence upon the posterior model PDF is the acceptance probability α (e.g. Liang *et al.* 2011). At every step of the Markov chain, a candidate model is sampled by perturbing the current model from a known PDF (the proposal PDF q) and the acceptance α is calculated. A random number r is then sampled uniformly from the interval [0,1]. If $r < \alpha$ the proposed perturbation is accepted, else the old model is retained. The rationale behind this algorithm can be explained by examining in more detail the expression for α (Bodin & Sambridge 2009), where

$$\alpha(\mathbf{m}'|\mathbf{m}) = \min\left[1, \frac{p(\mathbf{m}')}{p(\mathbf{m})} \times \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \times \frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} \times |\mathbf{J}|\right]. \tag{B16}$$

Here $\mathbf{m}'$ is the new proposed model and $\mathbf{m}$ is the old model (throughout this paper, primes will denote new model values). Specifically, $\frac{p(\mathbf{m}')}{p(\mathbf{m})}$ is the prior ratio, $\frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})}$ is the likelihood ratio and $\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})}$ is the proposal ratio. The Jacobian term $|\mathbf{J}|$ is not to be confused with the model Jacobian needed for gradient based inversions (e.g. Constable *et al.* 1987), but is a matrix that incorporates changes in model dimension when moving from $\mathbf{m}$ to $\mathbf{m}'$. In a classic MH algorithm with a fixed number of dimensions, the prior ratio (for uniform priors), proposal ratio (for symmetric proposals), and Jacobian term are all 1 (Dettmer *et al.* 2010). Hence the algorithm always moves toward areas of higher posterior probability if the data misfit improves (likelihood ratio >1). However, it can also move to areas of lower posterior probability with a probability α if the misfit does not improve (likelihood ratio <1).

To be able to compare likelihoods between models with different numbers of parameters (i.e. with different dimensions), the Jacobian in the acceptance term in eq. (B16) needs to be evaluated. There are various implementations of RJ-McMC, and in all the examples cited so far, a 'birth-death' scheme has been used. As shown in Bodin & Sambridge (2009) and Dettmer *et al.* (2010) for the 'birth-death' RJ-McMC scheme, this Jacobian term is unity. We have adopted the 'birth-death' algorithm in this paper and shall not concern ourselves with this Jacobian term any further.

B4 Outline of our algorithm

We start the algorithm with a very simple model, with $k = k_{\min}$, typically a layer in contact with a half-space. We then allow the algorithm to iteratively add interfaces ('birth') or remove them ('death'), perturbing the layer properties, as the data may demand via the acceptance probability α in (B16). In brief, this is how we proceed:

B4.1 Initialization

Start the algorithm with $k = k_{\min}$ and sample all parameters for each layer uniformly from within our prior bounds.

B4.2 Choose one of four moves

(1) Update: Perturb a randomly chosen layer property about its current value using a Gaussian proposal $q(\mathbf{m}'|\mathbf{m})$ with a standard deviation Σ_{m_e}, where

$$q(\mathbf{m}'|\mathbf{m}) = \frac{1}{\sqrt{2\pi}\,\Sigma_{m_e}} \exp\left[-\frac{1}{2\Sigma_{m_e}^2}(m'_e - m_e)^2 \right]. \tag{B17}$$

Note that this update move does not involve a change in the number of interfaces.

(2) Birth of a new layer: $k' = k + 1$. In the depth interval defined by $[z_{\min}, z_{\max}]$ randomly and with uniform probability we select an unoccupied point and insert an interface. Uniformly sample V_p, ν_s, ρ, Q_p, Q_s to assign properties for this layer.

(3) Death of an interface: $k' = k - 1$. An existing interface is selected at random and deleted.

(4) Move a layer interface: an existing interface is selected at random and its position is perturbed by a Gaussian proposal with standard deviation Σ_z . Note that this step does not involve a change in the number of cells either.

At each step, one of these moves is chosen with a certain probability, such that the move probabilities sum to unity. In addition, the birth and death probabilities must be set equal. We set the probabilities as follows:

$$[\text{update, birth, death, move}] \equiv \left[\frac{4}{10}, \frac{1}{4}, \frac{1}{4}, \frac{1}{10} \right].$$

At each step of the Markov chain, the proposed model is evaluated for acceptance. If it is accepted, it becomes the current model. If it is rejected, the current model is preserved and the algorithm moves on to the next step. In order to compute the acceptance, one needs to evaluate eq. (B16), for which we explicitly describe the proposal PDFs and their ratios in the next section.

B5 Proposal PDFs and acceptance probabilities

B5.1 Fixed dimension moves

For all moves that are neither birth nor death, the number of cells remain fixed. In these moves, we have elected to use Gaussian proposals to suggest the new model parameters by centring the proposals on the old parameters and drawing a random number from a normal PDF with a given standard deviation (step size). We can see from eq. (B17) these kinds of moves are symmetric, implying that the probability to go from the old state to the new state is the same as it would be in going from the new state to the old state:

$$\left[\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} \right]_{\text{fixed}} = 1. \tag{B18}$$

Since the number of dimensions remains constant, the prior ratio in eq. (B16) is 1. Hence for fixed dimension moves, we find that the acceptance probability is simply the ratio of the likelihoods:

$$\alpha_{\text{f}} = \begin{cases} \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \right] & \text{if } z \in [z_{\min}, z_{\max}], \\ & m_e \in [m_{e_{\min}}, m_{e_{\max}}], \\ 0 & \text{else} \end{cases}. \tag{B19}$$

B5.2 Birth move

For a birth move, one can select any depth in the interval $\Delta z = z_{\max} - z_{\min}$. The birthed interface's property values are drawn from uniform distributions. Since the selection of a position and the selection of property values are independent, we can write

$$q(\mathbf{m}'|\mathbf{m}) = \frac{1}{\Delta z}\frac{1}{\Delta m_e}. \tag{B20}$$

For the reverse move in a birth, keeping in mind that the current state has k interfaces, there were $k + 1$ interfaces to delete from, and the probability of removing elastic parameters in a layer in the reverse move is 1. Thus we have

$$q(\mathbf{m}|\mathbf{m}') = \frac{1}{(k+1)} \times 1. \tag{B21}$$

It follows in a birth move, from eqs (B20) and (B21), that the proposal ratio can be written as

$$\left[\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} \right]_{\text{birth}} = \frac{\Delta z \Delta m_e}{k+1}. \tag{B22}$$

Finally from eqs (B15), (B16) and (B22) we get for the birth move the following acceptance probability

$$\alpha_{\text{b}} = \begin{cases} \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \right] & \forall k \in [k_{\min}, k_{\max}] \\ 0 & \text{else} \end{cases}. \tag{B23}$$

B5.3 Death move

In a death move, one can select one of k interfaces for deletion. Further, the probability of removing properties in a layer is certain. Thus,

$$q(\mathbf{m}'|\mathbf{m}) = \frac{1}{k} \times 1. \tag{B24}$$

In the reverse move for death (i.e. birth) an interface is inserted at random uniformly in the range Δz. Further, the elastic properties are uniformly assigned from the prior ranges. Hence,

$$q(\mathbf{m}|\mathbf{m}') = \frac{1}{\Delta z}\frac{1}{\Delta m_e}. \tag{B25}$$

Thus we can see from eqs (B24) and (B25) that the proposal ratio for death can be written as

$$\left[\frac{q(\mathbf{m}|\mathbf{m}')}{q(\mathbf{m}'|\mathbf{m})} \right]_{\text{death}} = \frac{k}{\Delta z \Delta m_e}. \tag{B26}$$

Again from eqs (B15), (B16) and (B26) we get for the death moves the following acceptance probability

$$\alpha_{\text{d}} = \begin{cases} \min\left[1, \frac{p(\mathbf{d}|\mathbf{m}')}{p(\mathbf{d}|\mathbf{m})} \right] & \forall k \in [k_{\min}, k_{\max}] \\ 0 & \text{else} \end{cases}. \tag{B27}$$

Thus, we get an acceptance probability for both birth and death that is dependent only on the likelihood ratio, as must be the case if we propose models from their prior distributions (e.g. Mosegaard & Tarantola 1995; Malinverno 2002). Further, it should be noted that the derived expressions for α in eqs (B19), (B23) and (B27) do not involve a fictitious grid and are different from the expressions derived in Ray *et al.* (2014) or Bodin & Sambridge (2009). This is because we accept the findings of Dosso *et al.* (2014) and find that

we obtain better sampling efficiency in proposing property values from the prior for birth and death moves.

Though our derivation has not introduced the notion of a minimum layer thickness, it could be valuable either for reasons of numerical stability in the modelling algorithm or to increase sampling efficiency by precluding models that are thought to be unresolvable. We can easily incorporate the idea into the prior and proposal formulations. The derivations of the acceptance probabilities for the birth and death moves would change and involve order statistics as shown by Malinverno (2002), but neither of the derived acceptance probability formulae (B27) or (B23) would change. We advocate that some care should be exercised in enforcing a minimum thickness value. This is because thin layers may be resolved with low shear velocities at high frequencies, and the full nonlinear problem may have resolvability not readily apparent if we were to ignore refractions and mode conversions—a similar point has been beautifully phrased in the conclusions section of Mosegaard & Tarantola (1995). Nonlinear resolution using trans-D is also discussed by Galetti *et al.* (2015) for a tomography problem.

Finally, we mention here that in practice, the log of the acceptance probabilities are calculated and compared against the log of a uniform random number between 0+ and 1. Taking log avoids many problems of numerical stability in the evaluation of eq. (B16).

B6 Convergence to the posterior PDF

The algorithm is run for a given number of steps until it is deemed to have collected enough samples to provide a reasonable estimate of the posterior model PDF. There are a couple of caveats in this regard, as there are with any McMC sampler (Liang *et al.* 2011). If the algorithm is seeded with an initial model that is in a low posterior probability region, it may take quite a few steps till it reaches a region of high posterior probability, such that it begins to sample models, most of which fit the data within the given data error. The number of such required steps (which are subsequently discarded in the final chain) is known in McMC parlance as the 'burn-in' period, which depends on how well the proposal PDFs have been scaled (Chib & Greenberg 1995). This brings us to the step sizes (scaling) in the proposal PDFs in the form of the standard deviations Σ_{m_e} and Σ_z required in the various proposals to generate a new candidate model. The form of the proposal PDFs should 'emulate' the posterior for efficient sampling, but since the posterior PDF may be complicated (and unknown *a priori*), any kind of simple PDF, symmetric where possible, can be used. The exact form of the proposal does not affect the final solution, at least in theory.

The suitability of the step size for the problem at hand can be examined by looking at the number of samples accepted in a large interval of steps, referred to as the acceptance rate. If the acceptance rate is too low, it means that the step sizes are too large as lots of steps are falling outside the prior bounds or are being rejected as they land in low probability (high misfit) areas. If the acceptance rate is too high, then it implies that the algorithm is not exploring the model space enough and will again be slow to converge upon the posterior PDF. We discuss the details of acceptance rates attained with our real data experiment after the discussion on parallel tempering at the end of this section.

An attractive alternative could be to use adjoint based methods (Plessix 2006) to provide local gradients at minimal computational expense and construct model covariance matrices (Sen & Stoffa 2013) to step around the parameter space more efficiently (e.g. Malinverno 2002; Minsley 2011).

While the sampled posterior should not depend on the size of the steps taken, one has to factor in the optimality of the step size as otherwise convergence will be very slow. For an illuminating discussion on this matter, one can refer to Chib & Greenberg (1995). For further discussions on convergence diagnostics such as Potential Scale Reduction Factors (Gelman & Rubin 1992) and the practicality of their application, one can refer to Liang *et al.* (2011). When the number of parameters is changing, as in RJ-McMC, it is very challenging to assess convergence and this is still an active area of research (Bodin & Sambridge 2009). A heuristic method to examine speed of convergence is to monitor the 'distance' travelled from one sampling step to the next over a window of a certain number of steps. We can project model parameter values on to an underlying grid, and then look at the norm of the difference between grid values for two successive models. We can then find the average distance travelled over a given step size window.

The algorithm should be run long enough such that there is a measure of stationarity achieved in the square misfit (negative log likelihood) with iteration number, and that the posterior sample does not change appreciably on the addition of more samples.

B6.1 Parallel tempering

Finally, to ensure that the inferred posterior is not biased due to being trapped in local maxima (of the posterior probability), we recommend that the algorithm be run with numerous McMC chains at different likelihoods, ideally in parallel for computational efficiency. The different likelihoods should be 'tempered' (Swendsen & Wang 1987; Geyer 1991; Earl & Deem 2005) so as to provide a sequence of easier to sample likelihoods which enable different chains to exchange models between the different misfit spaces (i.e. negative log likelihoods) according to another MH exchange criterion:

$$\alpha = \frac{\mathcal{L}_{\rm hot}(\mathbf{m}_{\rm cold})}{\mathcal{L}_{\rm hot}(\mathbf{m}_{\rm hot})} \cdot \frac{\mathcal{L}_{\rm cold}(\mathbf{m}_{\rm hot})}{\mathcal{L}_{\rm cold}(\mathbf{m}_{\rm cold})} \tag{B28}$$

Chains running at flattened likelihoods ('hot') which are easier to sample can exchange models with chains running at the difficult to sample ('cold'), unbiased peaky likelihood (Dosso *et al.* 2012; Ray *et al.* 2013a; Dettmer *et al.* 2015). Details of this method and easy to implement pseudo-code can be found in Sambridge (2013). An added attraction of the above exchange criterion (B28) is that it can also run within an optimization framework using eq. (29)—this could prove quite useful in escaping local misfit minima and cycle skips for FWI.

In our implementation, we used 99 parallel tempering chains, at equal intervals in the log of temperature from $T = 1$ (the target unbiased sampling temperature) till $T = 8$ (the hottest temperature). Posterior inferences are made from the target temperature. For the Alba data example, the negative log likelihood (a measure of the square misfit) with iteration number for these chains can be seen at the top of Fig. B1, with each temperature being shown in a different colour. The negative log likelihood (NLL) can be negative as additive constants not dependent on the sampled model in eq. (A8) have been ignored (as also noted by Sambridge 2013). The NLL for the target chain at $T = 1$ has been shown in black. The higher temperature chains have higher NLL values as they sample models with worse misfit in an easier to sample (i.e. tempered) likelihood space. The fact that chains transition from one NLL level to another rapidly is indicative of good chain mixing. This can be seen at all NLLs after about 200 000 iterations. Details on implementing a temperature ladder which facilitates chain mixing can be found in

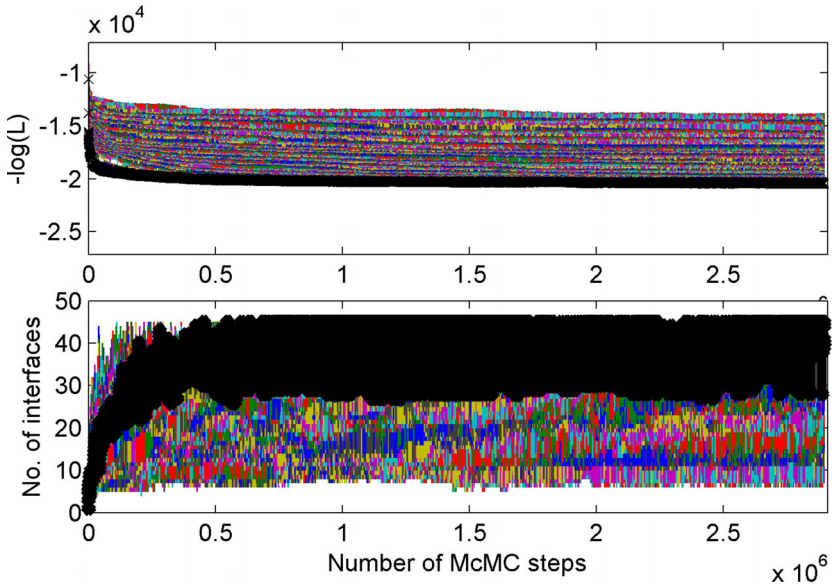


Figure B1. Top: for the Alba data inversion, negative log likelihood (representative of the squared misfit) per chain for all McMC chains, with the target chain shown in black. Model exchanges between chains in parallel tempering can be seen in the NLL transitions between chains. The high-temperature chains converge slowly, sampling a flattened likelihood. Model exchanges enable the lower temperature chains sampling peaky likelihoods to escape local misfit minima. Bottom: the number of interfaces sampled by different chains as sampling progresses, with the target chain shown in black. Lower temperature chains, sampling a peaky multimodal likelihood tend to have more interfaces.

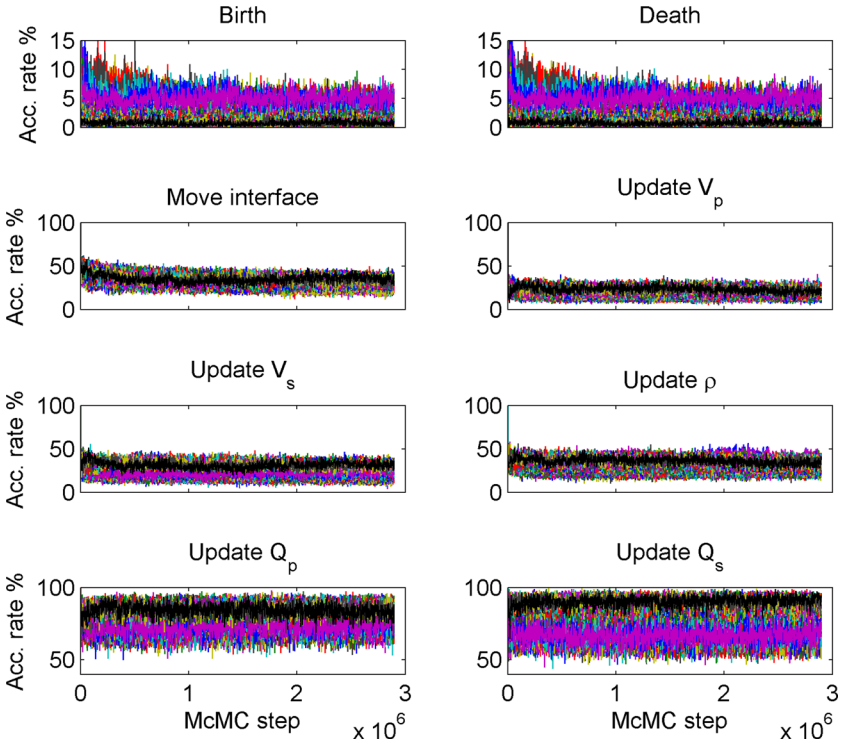


Figure B2. Alba data inversion: percentage of accepted moves for each move type per 2000 steps, with each chain shown in a different colour and the target chain in black. The hard to configure birth/death moves show a range of acceptance rates, with low acceptance rates of 1–2 per cent at the target temperature (peakiest likelihood). As the higher temperature chains achieve birth/death rates of 5–10 per cent, exchanges with higher temperature chains ensure that the target chain samples over a large range of interface numbers (as seen in the bottom row of Fig. B1). Q_p and Q_s show very high acceptance rates because of the relative insensitivity of the data to these parameters.

Sambridge (2013). Taking a conservative approach, we deemed the chains to have converged beyond 2 million samples (though we see the chains begin to flatten much earlier) as an acceptable measure of stationarity in the NLL was achieved. Our final posterior PDF was inferred from the last 800000 samples in the target chain. We applied chain thinning, using only every 100th sample to avoid correlations between successively sampled models, thus making our final posterior collection of models 8000 in number. The bottom plot in Fig. B1 is a plot of the number of interfaces sampled with iteration number, with each chain shown in a different colour. Again,

the target chain has been shown in black. In general, we find that higher temperature chains tend to have a fewer number of interfaces and be 'rougher,' since they sample flattened likelihoods at higher misfits (e.g. Bodin *et al.* 2012; Ray *et al.* 2013a).

Finally, we present in Fig. B2 the acceptance rate per move type for each of the McMC chains within our parallel tempering framework. We get acceptance rates between 20–50 per cent for all moves except birth/death and Q_p, Q_s attenuation updates. The high rates for attenuation can be understood since the data are not very sensitive to them, as the synthetic experiments have shown. Conversely, our experience across different geophysical methods seems to indicate that the more informative the problem about the subsurface (i.e. higher resolution the method), the more difficult it is to obtain a high birth/death acceptance rate. For the full wave elastic seismic problem, if the current model somewhat fits the observed seismogram, it is hard to arbitrarily place or remove an interface across which a wave can reflect, transmit and mode convert, while still fitting the observations. Thus, we get very low acceptance rates for birth/death moves, especially at the target temperature. This is where parallel tempering ensures that the target chain explores the posterior space of models with different numbers of interfaces efficiently. As can be seen in the bottom row of Fig. B1, the target chain transitions quite easily between models that have anywhere between 28 and 45 interfaces owing to the fact that lower temperature chains are easily able to exchange models with higher temperature chains that have a different number of interfaces.