GEOPHYSICS, VOL. 81, NO. 2 (MARCH-APRIL 2016); P. ID1–ID24, 21 FIGS., 4 TABLES.
10.1190/GEO2015-0127.1

Joint stochastic constraint of a large data set from a salt dome

Alan W. Roberts[1], Richard W. Hobbs[2], Michael Goldstein[3], Max Moorkamp[4], Marion Jegen[5], and Bjørn Heincke[6]

ABSTRACT

Understanding the uncertainty associated with large joint geophysical surveys, such as 3D seismic, gravity, and magnetotelluric (MT) studies, is a challenge, conceptually and practically. By demonstrating the use of emulators, we have adopted a Monte Carlo forward screening scheme to globally test a prior model space for plausibility. This methodology means that the incorporation of all types of uncertainty is made conceptually straightforward, by designing an appropriate prior model space, upon which the results are dependent, from which to draw candidate models. We have tested the approach on a salt dome target, over which three data sets had been obtained; wide-angle seismic refraction, MT and gravity data. We have considered the data sets together using an empirically measured uncertain physical relationship connecting the three different model parameters: seismic velocity, density, and resistivity, and we have indicated the value of a joint approach, rather than considering individual parameter models. The results were probability density functions over the model parameters, together with a halite probability map. The emulators give a considerable speed advantage over running the full simulator codes, and we consider their use to have great potential in the development of geophysical statistical constraint methods.

INTRODUCTION

To map a region of earth, it is commonplace to use one or more kinds of data sets to constrain structural models parameterized by one or more proxy parameters, such as seismic velocity, density, or resistivity. An interpreter will then use their geologic insight combined with these models to make judgments about the region. This may be with a view to, for example, determining where appropriate drilling locations might lie to maximize the possibility of hydrocarbon extraction. There are many approaches used to constrain the proxy models, ranging from deterministic inverse approaches to Markov chain Monte Carlo (MCMC) search schemes (Press, 1970; Sambridge and Mosegaard, 2002; Shapiro and Ritzwoller, 2002; Hobro et al., 2003; Gallardo and Meju, 2004; Roy et al., 2005; Heincke et al., 2006; Meier et al., 2007; Moorkamp et al., 2011). Deterministic inverse schemes are optimal when the uncertainties in the data and physical system are small, and the aim is to find the optimum model as fast as possible. This approach works by repeated model update so as to minimize the difference between the observed data and the simulator's output. However, in many scenarios, there are considerable uncertainties associated with the data and physics concerned. In this case, statistical schemes may be adopted. In these methods, the aim is normally to discern the entire plausible model space for the system concerned. The character of such statistical schemes varies from the entirely forward-based screening method (Press, 1970), to the more targeted sampling strategy of the MCMC approach (Hastings, 1970; Sambridge and Mosegaard, 2002). MCMC schemes seek to sample enough of the model space to give a robust uncertainty estimate; however, often, the number of forward simulations in both of these methods required to sufficiently sample the space for large systems often makes these methods computationally impracticable. Thus, often in part due to the

Manuscript received by the Editor 27 February 2015; revised manuscript received 21 August 2015; published online 19 February 2016.
[1]Formerly Durham University, Department of Earth Sciences, Durham, UK, and Durham University, Department of Mathematics, Durham, UK; presently Geospatial Research Limited, Durham, UK. E-mail: a.w.roberts@durham.ac.uk.
[2]Durham University, Department of Earth Sciences, Durham, UK. E-mail: r.w.hobbs@durham.ac.uk.
[3]Durham University, Department of Mathematics, Durham, UK. E-mail: michael.goldstein@durham.ac.uk.
[4]Formerly GEOMAR, Kiel, Germany; presently University of Leicester, Department of Geology, Leicester, UK. E-mail: mm489@leicester.ac.uk.
[5]GEOMAR, Kiel, Germany. E-mail: mjegen@geomar.de.
[6]Formerly GEOMAR, Kiel, Germany; presently GEUS, Department of Petrology and Economic Geology, Copenhagen, Denmark. E-mail: bhm@geus.dk.

lack of a feasible method of assessing the uncertainty associated with a system and also conceptual difficulties in incorporating a particular kind of uncertainty into the constraint process, the uncertainty assessments that are fed into the decision-making processes can be ill informed.

The approach adopted here is a forward-modeling-based screening strategy. To make the process computationally feasible, we build "emulators" for each of the forward modeling ("simulator") codes, trained over the prior model space. We use the term *emulator* to mean a statistical model of the output generated by a complex forward modeling code, or *simulator*. The aim of building an emulator is to have a means of generating a fast uncertainty calibrated estimate of the simulator output (which maybe be time expensive to compute). Through providing uncertainty-calibrated rapid estimates of the full simulator output, these emulators overcome the computational barrier of making vast numbers of complex simulator runs. By iteratively rejecting an implausible model space and updating the emulators, the plausible model space is discerned. In this study, we apply the method to constrain a region of earth characterized by a salt dome using three kinds of data: seismic refraction, magnetotelluric (MT), and gravity data sets for a 1D seven-layered parameterization, and we construct a rock-type probability map based on the fractional salt versus sediment model acceptance for the region. As discussed in Osypov et al. (2011) and elsewhere, proper assessment of risk in hydrocarbon exploration requires not only the analysis of a proxy-parameter model, but also a full analysis of the structural uncertainty. The ability to construct a probability map in this manner has the potential to be of considerable value in this regard.

Joint inversion

Deterministic inversion methods (Tarantola, 2005), in which the aim is to iteratively update a model so as to reduce some objective function, are commonly used when the data come from a single technique. However, using such schemes in a joint framework in which the relationship between the physical parameters (e.g., seismic velocity, resistivity, and density) is empirical and uncertain, poses philosophical challenges regarding the coupling strategy, for example, the weighting attached to maintain structural coherency across the various methods (Gallardo and Meju, 2004). Similarly, there are also conceptual intricacies associated with properly and quantitatively including most kinds of uncertainty associated with the problem, for example, uncertainty in the data measurements and model discrepancy (due to the fact that a model is not a complete representation of nature). Recently however, a few authors such as Roy et al. (2005), Heincke et al. (2006), and Moorkamp et al. (2011, 2013) make considerable progress in developing structural coupling-based joint inversion methodologies through crossgradient and other coupling schemes. Bodin et al. (2012) also develop hierarchical Bayes approaches for joint inversion.

Statistical schemes

Statistical schemes designed to assess uncertainty, such as simulated annealing, genetic algorithms, and MCMC approaches, can be used when the number of model parameters is small. Sambridge and Mosegaard (2002) give a useful review of the varied methods that can be used and their historical development.

However, as is commented in Sambridge and Mosegaard (2002), if the number of parameters is large, then these methods become unfeasible because the number of complex, and possibly expensive, forward model simulations becomes impracticably large given the computation time required. In a few scientific fields, such as climatology, volcanic hazard prediction, ocean modeling, and cosmology (Logemann et al., 2004; Rougier, 2008; Bayarri et al., 2009; Vernon and Goldstein, 2009), in which forward simulators are also highly time expensive to run, *emulators* are often used. An emulator is a statistical representation of the forward modeling simulator, which gives a very rapid prediction of the simulator output, with a calibrated uncertainty.

Building an emulator is similar to building a neural network. Neural networks are successfully used to solve inverse problems in geophysics, for example, Meier et al. (2007), who develop a neural network system to invert S-wave data. Others have also developed methods of using quick approximations to a full forward code in inversion schemes, for example, James and Ritzwoller (1999), who use truncated perturbation expansions to approximate Rayleigh-wave eigenfrequencies and eigenfunctions, and Shapiro and Ritzwoller (2002), who take a similar methodology in an MCMC scheme to construct a global shear-velocity model of the crust and upper mantle. In each of these cases, the aim is to minimize some objective function or maximize a likelihood function.

Here, we adopt a statistical approach that is fundamentally different in that it is based entirely on forward modeling, as opposed to using any kind of objective/likelihood function or inverse step. We simply seek to discern which areas of model space are plausible and which are implausible, given the observed data. This approach has been proposed in the past (e.g., Press, 1970), and it is used in a variety of settings such as the history matching of hydrocarbon reservoir production data (Murtha, 1994; Li et al., 2012). However, in the context of structural constraint, it is largely sidelined in favor of more search-efficient schemes such as those described above. We implement this forward approach by the use of emulators to make it more computationally efficient. Roberts et al. (2012) describe our methodology for a synthetic scenario; however, here we describe a number of modifications to achieve greater stability and efficiency. We develop and apply the approach to observed 3D joint seismic, MT, and gravity data sets obtained from a salt dome region, and we ultimately determine a model probability map for the profile. The method is akin to the response surface methodologies beginning to be used in the field of reservoir simulation (Zubarev, 2009). However in this case, we seek to fully model the uncertainty in the simulator-prediction system, and hence we aim to construct response clouds, rather than surfaces. The method is shown diagrammatically in Figure 1.

The strategy here, to exclude model space, rather than build up the plausible space searching from some starting model, represents a fundamentally and philosophically different "top-down" approach, to the traditional inversion, and it relies entirely on forward computation. Because we globally sample the prior plausible model space, seeking to exclude implausible model space, rather than searching a part of the model space for plausible models, the uncertainty measures which are obtained are maxima, rather than minima, given the prior model space and choices of tuning parameters made in the analysis.

Building statistical system models, or *emulators*, successively in a multicycle fashion, we progressively refine the plausible space. Because a proper consideration of uncertainty in an inverse scheme can be conceptually difficult, often, when any consideration of uncertainty is made, it is commonly specified to be Gaussian in

character at times not because it is indeed Gaussian, but simply because of mathematical convenience. The fact that, in our method, the screening process relies entirely on forward modeling means that it is conceptually straightforward to include uncertainty pertaining to any part of the system by building the appropriate distribution over the prior model space. These uncertainties may take the form of data uncertainty, physical uncertainty, model discrepancy, or others, which in an inversion scheme, including physical uncertainties and model discrepancy, among others, can be conceptually difficult.

Emulation

To overcome computational limitations in our forward screening Monte Carlo scheme, we build and use emulators (Kennedy and O'Hagan, 2001; Vernon et al., 2009). Like the case of a neural network, an emulator is designed using training models and data sets, and it seeks to predict the output data arising for a given model parameter set. However, an emulator differs from a neural network in that it seeks to not only predict the output of a system from an input, but also to do so with a fully calibrated uncertainty. An emulator treats the parametric and nonparametric parts of the system holistically, giving a full stochastic representation of the system. Because of this focus on uncertainty calibration, emulators can be used to rapidly screen model space for implausibility (Goldstein and Wooff, 2007). This would not be the case with an uncalibrated simulator-prediction system because there is no measure or criterion to discern whether a comparative data set is sufficiently close to the observed data set to be deemed plausible. Although the method described here is very much a forward modeling philosophy, in that we are simply seeking to trial sets of model parameters for plausibility, one could consider that the fitting of parametric functions to build the emulator model of the forward simulator constitutes a partly inverse component. However, because the forward simulator itself, rather than the model parameters is being "inverted" for, the emulator screening method is fundamentally different to a traditional inversion scheme.

Although there are occasional instances of emulators being developed for earth systems (Logemann et al., 2004), they have not been widely applied in the geosciences. Here, we review and demonstrate the use of an emulator (Roberts et al., 2010, 2012) to constrain the structure of a salt diapir using 1D profiles through a 3D joint data set. Figure 1 summarizes the strategy adopted in this study. The data consist of 3D seismic data, full tensor gravity (FTG) data, and MT data from a salt dome. Examples from the three data sets are shown in Figure 2.

THEORY AND PRELIMINARIES

In performing an experiment to test a model, the scientist has a set of output data points, a set of model parameters, and a function (*simulator* or forward modeling code) f, which defines the relationship between the model parameters $\mathbf{\theta}$ and the "perfect" data $\mathbf{\psi}$ (equation 1) as follows:

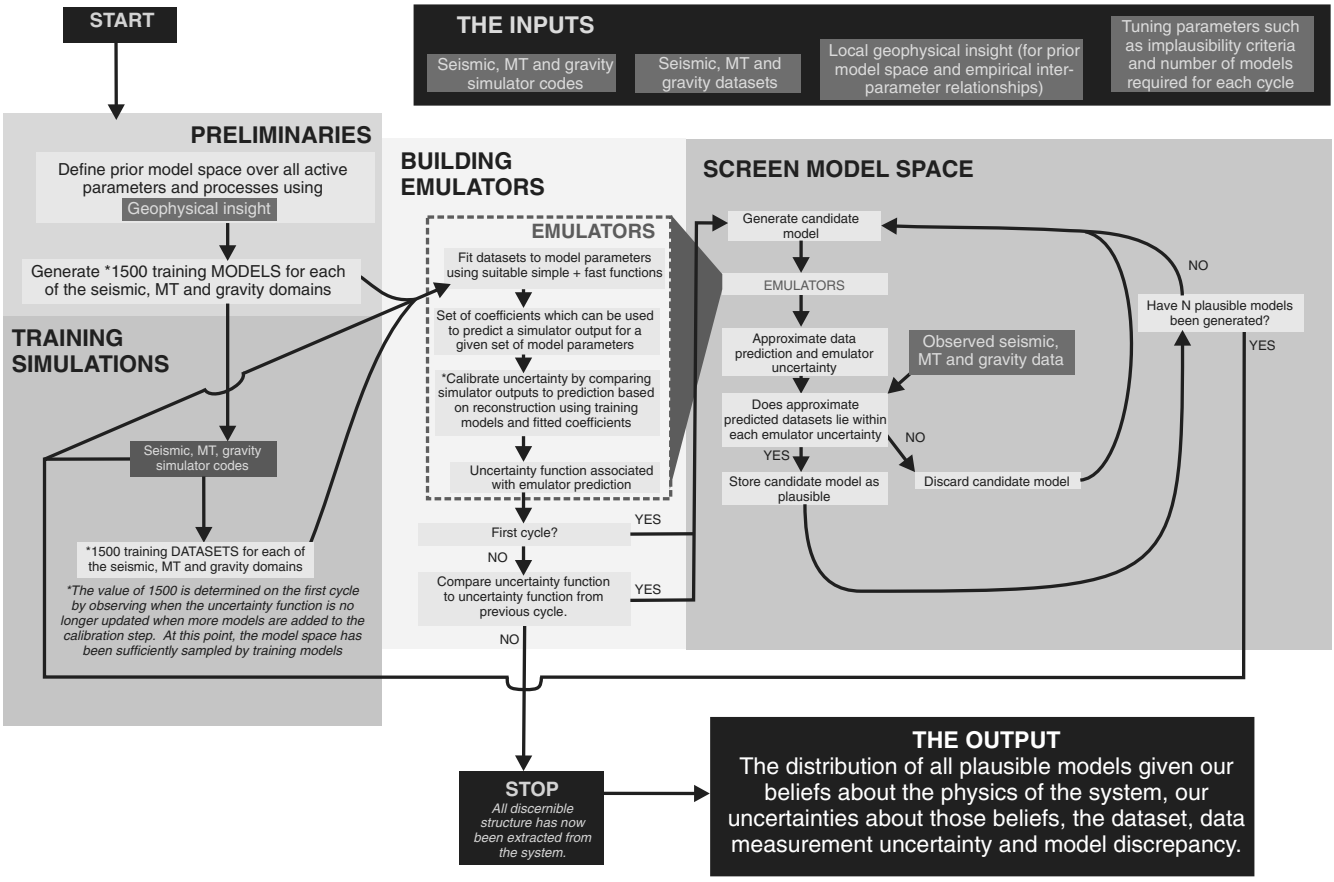

Figure 1. The emulator screening methodology.

$$\boldsymbol{\psi} = f(\boldsymbol{\theta}). \quad (1)$$

In the case of a heavily parameterized system, with many nodes in the parameter models, and many output data, this function, or *simulator* f can take a long time to evaluate. For a typical inversion problem involving seismic data, many thousands of evaluations of f maybe required, and the problem very quickly becomes impractical to solve if the time spent evaluating f is significant.

Vernon et al. (2009) and Kennedy and O'Hagan (2001) use emulators to address this kind of problem in which large numbers of complex simulator evaluations are required. An emulator seeks to represent the simulator function f as a combination of a computationally cheap deterministic function (e.g., a polynomial) h and a Gaussian process g (Rasmussen and Williams, 2010):

$$\boldsymbol{\psi} = h(\boldsymbol{\theta}) + g(\boldsymbol{\theta}). \quad (2)$$

The aim is not to completely replace the full simulator, but to develop a system such that one can very quickly glean enough information from the relationship between the model parameters and output data to make meaningful judgments about whether regions of model space can be excluded from the analysis on the basis that they would result in output data not compatible with the observed data.

Because h and g are fast to evaluate, a considerable time savings (of orders of magnitude) can typically be achieved by this approach. In the study detailed here, we adopt a multistage approach (Vernon et al., 2009) of seeking to describe the global behavior and then, as the implausible model space is excluded, to describe increasingly localized behavior as we develop more predictively accurate emulators.

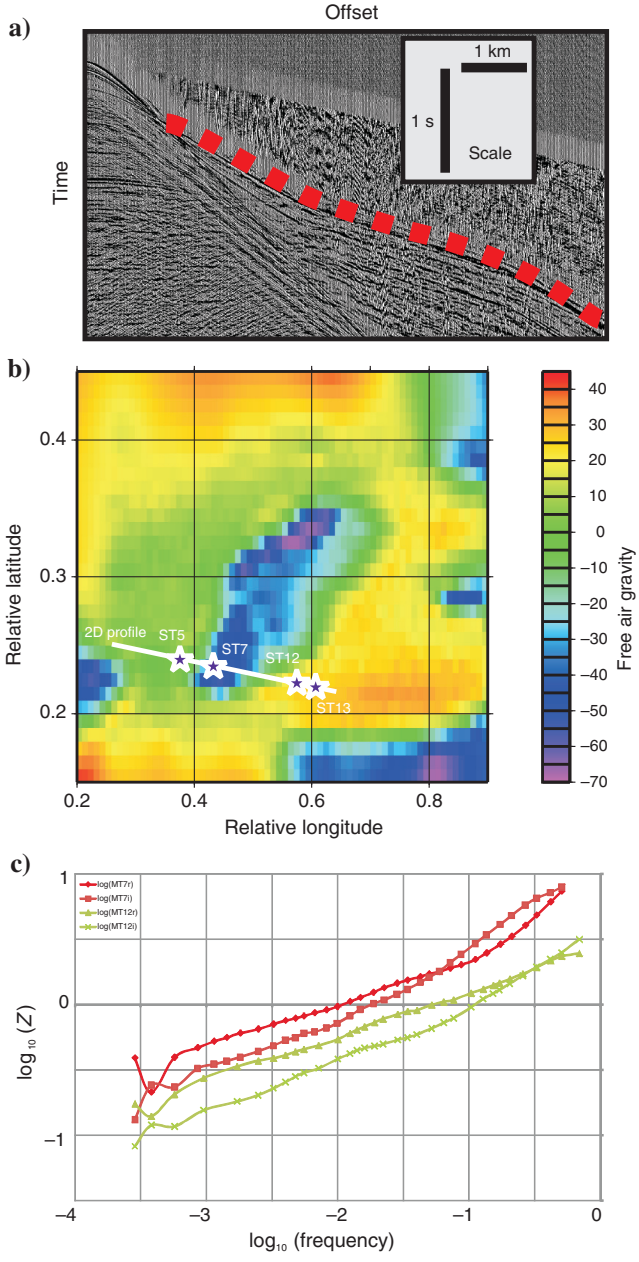

Figure 2. Data examples: (a) seismic, (b) gravity, and (c) MT. The red dots on the seismic gather show the first arrival wide-angle turning waves, which are being modeled in this study. On the gravity map in panel (b), the locations of ST5, ST7, ST12, and ST13 (which are frequently referred to in this study) are marked with purple stars, and the track of the 2D line for which profiles are shown in later figures. The MT data plot shows Berdichevsky invariant (Berdichevsky and Dmitriev, 2002) Re(Z) and Im(z) for stations ST7 (red) and ST12 (green).

DATA, MODEL SPACE, AND THE INVERSE PROBLEM

Data

A joint data set for this study was kindly supplied by Statoil. It is a joint 3D seismic, FTG and MT data set recorded over a region known to Statoil as being characterized by a salt diapiric body (Figure 2). To simplify the problem, we enforce a local 1D solution. The MT data were transformed into the directionally independent Berdichevsky invariant form (Berdichevsky and Dmitriev, 2002), seismic data were picked for 1868 shot gathers and transformed into the common midpoint (CMP) domain, and the closest CMP profiles to each MT station were identified and used as 1D seismic data for the purposes of the study. The FTG data were transformed to scalar data, and the closest measurement to each MT station was identified. Results are thus generated for the series of 1D seismic, MT, and gravity data sets collocated at the site of each MT station. In this paper, each site is labeled "STxx", where "xx" can take the value 1–14, for example, in Figure 2.

Gravity datum

In addition, it was also necessary to establish a datum for the gravity data measurements so as to make meaningful comparison between each station. This is because if the models are allowed to be of arbitrary total thickness, then the gravity reading could be considered as simply as a free parameter and afford no constraint. In practice, this is an expression of the Airy hypothesis of isostasy (Airy, 1855). This calibration requires the tying of the measured gravity point at one station to the simulator output at that station, with an assumption about the structure at that station, against which results at the other stations can be considered as being relative to. This assumption might be that the model is of a given total

thickness, or some other consideration about the model. We choose to use the assumption that at a given point along the line, there is no salt present. On the grounds that prior studies indicate that it lies over a region of sediment, we chose to use station ST12 (Figure 2) for the purpose of gravity calibration. The calibration was carried out by running the code with gravity screening disabled, only generating "sediment models" and then generating density models using the relationship in equation 7 from the plausible velocity models. Gravity measurements were then generated from these models using the gravity simulator, and the most likely gravity measurement were compared with the measured gravity value at station ST12. The difference between these two values was then used as a "correction" value for comparing screened gravity values to measured gravity values at all the other stations. In practice, this means that the gravity results in this study and the "salt content" are being measured relative to station ST12 screened with the assumption that there is no salt there (the prior probability of salt in each layer is set to zero, and thus no salt models are generated). However, for a meaningful comparison (and a test of the assumption that there is no salt at ST12), the screening process is then repeated for ST12 with the possibility of salt models included with a probability of 0.5 in each layer, as was the case for the other stations. A comparison of the results prior to the gravity calibration (without gravity constraint) and postcalibration (the main results presented in this paper) would provide an interesting study; however, for brevity, these preliminary outputs are not discussed here.

Methods and model space

Our goal is to describe the model in terms of proxy quantities (P-wave velocity, resistivity, and density) and also to obtain a rock probability map for the profile along which the MT stations are located. Although the priors in this study are somewhat illustrative, in a scenario in which the priors are well constrained and tested, such a probability map may be used as a more direct input to evaluate geologic risk, rather than simply providing proxy-parameter models that the scientist must then interpret. We discern the distribution of jointly plausible models with respect to each of the seismic, gravity, and MT data sets, given all of the uncertainties we wish to specify, by generating candidate joint models drawn from a prior model space. In this way, we effectively screen the model space using the interparameter relationship to discriminate between salt and sediment rocks.

As is noted in Roberts et al. (2012), the simulators, particularly the seismic simulator (Heincke et al., 2006), are more sophisticated than required for the problem at hand; however, to facilitate future development and allow integration and direct comparison with other work (Heincke et al., 2006; Moorkamp et al., 2011, 2013), we use these simulators.

Prior model space

The first consideration is the initial model space within which we consider the plausible models for the system to lie (Figure 1). Our focus here is on the emulation methodology as a means to screen and constrain model space, rather than on generating robust Bayesian posterior distributions for the particular region used for the case study. As such, here we have placed only a cursory emphasis on the determination and specification of the prior model space. The final result should, therefore, not be considered as a true Bayesian constraint from which a genuine geologic inference can be made about this region. For such a result, proper consideration of appropriate priors should be made, and proper sensitivity calibration through sampling those priors. Accordingly, the analysis presented here is made on the assumption that there is indeed halite present in the region of earth under consideration and on the basis that that halite, and indeed the surrounding material, has properties reflective of those seen globally and in conjunction with the borehole data set described below.

Similarly, at several points in the emulator building and screening process, the tuning parameters are set. Again, here these are chosen somewhat qualitatively and arbitrarily. In reality, the choices made for these parameters also constitute part of the model space, and so for the results of the screening process to be geologically meaningful, expert judgment should be used in the choice of these parameters, with a prior distribution that can be fully sampled. The final results of the analysis presented here should thus be treated as being illustrative of the method and should be subject to all of the explicit and implict assumptions made, rather than being authoritative as to the earth structure in question.

Our prior joint model space is constrained primarily by three influences: (1) the interparameter relationship linking the seismic velocity, density, and resistivity parameters, (2) the range of geophysically plausible values which each of these parameters may adopt, and (3) the prior probability of salt existing in each layer.

Physical parameter relationship.—For the purpose described here, a rock is characterized by its combination of physical properties (in our case, resistivity, seismic velocity, and density), which are encapsulated by the empirical physical parameter relationships that connect them. In this joint setting, we therefore propose, for a given layer, combinations of model parameters across each of the domains that are connected by either a sediment relationship or a "salt relationship. By doing this, and then assessing the fraction of models deemed plausible generated using each relationship for a given depth, we can then make a statement about the rejection ratio for models generated using each relationship regarding the probability that salt or sediment exists at different locations and constructing a salt likelihood map for the profile.

The interparameter relationship for sediments, although it is empirical and uncertain, may be relatively easily formulated by fitting a curve through well-log data from the survey area (Figure 3). However, the presence of salt complicates the situation somewhat, in that for sediment there is a monotonic increase between seismic velocity, density, and resistivity. However, salt has a very characteristic seismic velocity of 4500 m/s, a density of around 2100 kg/m^3 (Birch, 1966), and very high resistivity (>500 Ωm; Jegen-Kulcsar et al., 2009). Therefore, we define two relationships for our situation, as shown in equations 5–9. In this case, we have chosen the uncertainty to be a function added to a central value. It would also be straightforward to specify the uncertainty in other ways, for example, as uncertainty in the values of the relationship coefficients. In these relationships, r, ρ, and v refer to the resistivity, density, and seismic velocity values, respectively. The value $N(a, b)$ refers to a sample from a normal distribution of mean a and standard deviation b. The borehole data from which the sediment density/resistivity/velocity relationship was obtained (kindly provided by Statoil) is shown in Figure 3. The borehole is located adjacent to station ST5 (Figure 2). Consideration of Figure 3 suggests that given that there are a considerable number of points lying outside the bounds shown for the salt and sediment relationships, there may be a case

for including a third relationship category of "other," to accommodate these currently extremal points more effectively. However, for simplicity, here we decided to continue with the two-relationship scheme. In this case, we have used data from this single borehole for the entire line. A more rigorous study would seek if possible to consider borehole data from various points along the line to account for regional variation.

Sediment parameter relationships are given in the following equations:

$$\log_{10}(r) = -8.72487 + 0.0127274v - 6.4247 \times 10^{-6}v^2 + 1.45839 \times 10^{-9}v^3, \quad (3)$$

$$-1.47131 \times 10^{-13}v^4 + 5.32708 \times 10^{-18}v^5 + N(0, \sigma_r(v)), \quad (4)$$

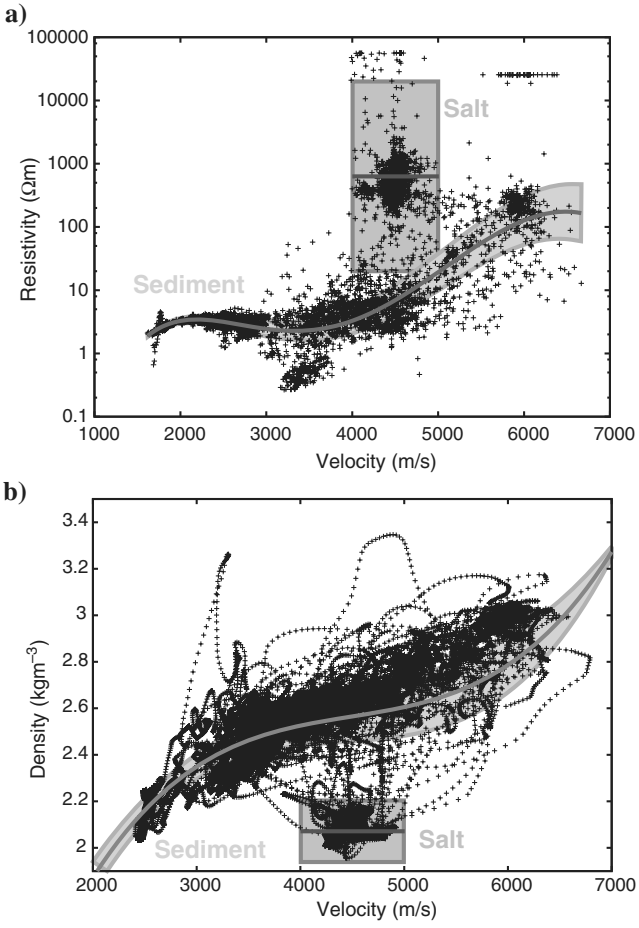

Figure 3. (a) Resistivity versus velocity and (b) density versus velocity relationships derived from well-log data. The borehole is located adjacent to station ST5 (Figure 2). Data points were characterized by location on the plot as being from salt or sediment, and regions defined from which appropriate combinations of velocity and resistivity parameters could be drawn. The salt region was defined as a rectangular box, whereas the sediment relationship was defined by fitting a polynomial curve (equation 5). The fitted relationship and associated uncertainty for resistivity are shown in equations 5–7 and the equivalent relations for density are given in equations 8–9. The bounds shown here are for the 99% confidence bound (3σ). Data are kindly provided by Statoil.

$$\sigma_r(v) = -2.931 \times 10^{-2} + 1.989 \times 10^{-5}v + 1.058 \times 10^{-9}v^2, \quad (5)$$

$$\rho = -785.68 + 2.09851v - 4.51887 \times 10^{-4}v^2 + 3.356 \times 10^{-8}v^3 + N(0, \sigma_d(v)), \quad (6)$$

and

$$\sigma_\rho(v) = 1.42693 \times 10^2 - 1.11564 \times 10^{-1}v + 3.0898 \times 10^{-5}v^2 - 2.52979 \times 10^{-9}v^3. \quad (7)$$

Salt parameter relationships are given in the following equations:

$$\log_{10}(r) = 2.8 + N(0, 0.5) \quad (8)$$

and

$$\rho = 2073 + N(0,45). \quad (9)$$

Parameter ranges.—Another important bound on the model space is our belief about the prior plausible model parameter ranges. It is implicit that the model parameterization should be chosen so as to be capable of describing the full range of prior plausible models using as few parameters as possible and also that it should be as unique as possible. Here, we consider 1D joint common structure models (Jegen-Kulcsar et al., 2009). Thus, we choose a parameterization of velocities, densities, and resistivities for a series of layers of common variable thickness. A fuller treatment would involve quantitatively trialing the ability to represent appropriate geologic formations and data sets using a range of numbers of model layers that is in itself part of the model space definition. In this case, we considered that after informal qualitative testing, models parameterized by seven layers seemed sufficient in providing the ability for the system to discern structure, particularly in the shallow region and the salt body, while not overparameterizing the system given the resolution of the observed data sets. The prior model parameter ranges for each of these layers are shown in Table 1.

Prior salt probability.—A more subtle constraint on the prior model space is the prior probability of salt existing in each layer. For this study, we specify this to be 0.5 for each layer. In each screening cycle, for each layer, salt or sediment models are generated in the ratio appropriate to the fraction of models (the likelihood of salt present) deemed plausible for that layer from the previous cycle.

Building an emulator

Having specified the prior model space from which we intend to draw candidate models, we now construct an emulator for each of the seismic, gravity, and MT cases. We describe the process in detail for the seismic case and adopt a similar approach for the MT and gravity cases. The framework for each of these, including the full set of governing equations, is given in Appendix A. The model space used for training each emulator was simply defined by the range of parameter values considered plausible in each of the velocity, resistivity, density, and thickness cases (Table 1). In other words, each

emulator was built independently for each modeling domain (seismic, gravity, and MT), using uniform distributions over the parameters in columns 1–4 of Table 1, and without any statement about the origin of the models being used.

Seismic emulator

In constructing a seismic emulator, we use a method similar to that of Roberts et al. (2010, 2012), but here we develop and improve the results by fitting weighting coefficients to Laguerre polynomial functions rather than fitting simple polynomial functions. We choose to use Laguerre polynomials because of their mutual orthogonality, which increases the efficiency in fitting the functions concerned. Other classes of orthogonal polynomials could have been chosen here; however, Laguerre polynomials were a convenient choice. The exponential weighting associated with Laguerre polynomials means that the fitting process here may be more sensitive to lower parameter values. In a more thorough treatment, this could be a focus for investigation; however, no significant issues were encountered here, and they were deemed fit for purpose. Depending on the setting, other functions may be more suitable to choose for the bases; if the aim was to fit to periodic data, a natural choice of basis functions would have been a Fourier series, for example. For the first cycle, we consider the velocity model space shown in Table 1, parameterized by 14 parameters; $(v_m, s_m)_{m=1}^{7}$ where v_i and s_i are the velocities and thicknesses ascribed to each of the seven layers, as shown in Table 1. The model space is designed such that there is finer stratification in the shallow region. This reflects the fact that as a result of having traveltime data out to approximately 10 km of offset, we expect greater seismic sensitivity in the upper 3 km or so. Our aim in building the emulator is to predict, to a calibrated uncertainty, the seismic forward code output for models drawn from this space. We generate a 1500×14 Latin hypercube (McKay et al., 1979; Stein, 1987) and use this to create a set of 1500 models over the 14-parameter space, which fill the space evenly. Each of these 1500 14-parameter models is then passed in turn to the forward seismic simulator, producing 1500 t versus x plots, each consisting of 100 (x, t) pairs. The simulator computes traveltimes using a finite element method (Podvin and Lecomte, 1991; Heincke et al., 2006). Laguerre polynomial functions are then fitted, using a least-squares algorithm, to each of these data sets (equation 10) to compute a vector of polynomial coefficients $\alpha_{x,i}$ to represent each of the $i = 1 - 1500$ data sets. It was found that Laguerre polynomials of order 3, parameterized by four $\alpha_{x,i}$ coefficients to weight the polynomials, are sufficient to recover the form of the data and keep the least-squares algorithm stable. Our code is designed such that if a singularity occurs in the fitting of the coefficients (i.e., overfitting of the data is occurring), then the number of coefficients is automatically decreased until a stable fit is achieved. In early versions of the code, simple polynomials were used as basis functions instead of Laguerre polynomials and overfitting of the data points was commonplace; however, using Laguerre polynomials, with the property of orthogonality over the space concerned, has meant that such overfitting using the number of coefficients specified here has been eliminated. Thus, we reduce each plot of 100 data points to a set of four coefficients. In using these polynomial coefficients to represent the (x, t) data, there is a misfit function that we denote as $g_x(x)$, as follows:

$$t = \left(\sum_{i=0}^{p_x} \alpha_{i,x} x^i \mathrm{e}^{-x} L_i(x) \right) + g_x(x). \tag{10}$$

Table 1. Prior parameter bounds for each layer. Ranges are shown for emulator training, and the ranges used to sample models from each of the sediment and salt cases. Velocity values are given in units of m/s, resistivity values are given in units of Ωm, density values are given in units of kg/m, and the layer thickness values are given in units of m.

Layer	Parameter	Training (min)	Training (max)	Sediment (min)	Sediment (max)	Salt (min)	Salt (max)
1	Velocity	1600	5500	1600	5000	4000	5000
1	Resistivity	0.5	5000	0.5	10	100	5000
1	Density	1800	3600	1800	3600	2000	2200
1	Thickness	50	1600	50	1600	50	1600
2	Velocity	2000	5500	1600	5000	4000	5000
2	Resistivity	2.0	5000	2.0	20	100	5000
2	Density	1800	3600	1800	3600	2000	2200
2	Thickness	50	2700	50	2700	50	2700
3	Velocity	2000	6500	1600	5000	4000	5000
3	Resistivity	5.0	5000	5.0	70	100	5000
3	Density	1800	3600	1800	3600	2000	2200
3	Thickness	200	2900	200	2900	200	2900
4	Velocity	2000	6500	1600	5000	4000	5000
4	Resistivity	5.0	5000	5.0	70	100	5000
4	Density	1800	3600	1800	3600	2000	2200
4	Thickness	1200	2900	1200	2900	1200	2900
5	Velocity	2000	6500	1600	5000	4000	5000
5	Resistivity	5.0	5000	5.0	70	100	5000
5	Density	1800	3600	1800	3600	2000	2200
5	Thickness	1500	2500	1500	2500	1500	2500
6	Velocity	2000	6500	1600	5000	4000	5000
6	Resistivity	5.0	5000	5.0	70	100	5000
6	Density	1800	3600	1800	3600	2000	2200
6	Thickness	1500	2500	1500	2500	1500	2500
7	Velocity	2000	6500	1600	5000	4000	5000
7	Resistivity	5.0	5000	5.0	70	100	5000
7	Density	1800	3600	1800	3600	2000	2200
7	Thickness	1500	2500	1500	2500	1500	2500

We then fit $\boldsymbol{\alpha}_{x,i=1}^{4}$ to the model parameters (in this case, the velocity and layer thickness parameters $[v_m, s_m]_{m=1}^{7}$), again using a least-squares method to fit the weighting coefficients for Laguerre polynomials. This is similarly accomplished using Laguerre polynomials up to third order in each of the layer parameters (equations 11 and 12). The result is a set of 228 $\beta_{x,ijk}$ coefficients (four for each of the 14 model parameters, plus a zeroth-order term, for each of the four $\alpha_{x,i}$ coefficients ($= (4 \times 14 + 1) \times 4$)). Again, there is a misfit function associated with this fitting step (equation 13). Examples of the recovery of the $\boldsymbol{\alpha}_x$ coefficients using the $\boldsymbol{\beta}_x$ coefficients are shown in Figure 4. Using these α_x coefficients, we can then construct the traveltime curves for a given set of model parameters. Examples comparing the traveltime curves obtained using the recovered α_x coefficients with the simulated traveltime curves are shown in Figure 5:

$$\boldsymbol{\theta}_x = [v_1 \;\; v_2 \;\; v_3 \;\; v_4 \;\; v_5 \;\; v_6 \;\; v_7 \;\; s_1 \;\; s_2 \;\; s_3 \;\; s_4 \;\; s_5 \;\; s_6 \;\; s_7]^T \tag{11}$$

and

$$\alpha_{i,x} = \left(\sum_{k=1}^{w_x} \sum_{j=0}^{q_x} \beta_{ijk} \theta_{k,x}^{j} \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) \right) + g_{i,x}(\boldsymbol{\theta}_x). \tag{12}$$

In predicting the parametric components of the system, we have two sources of misfit in the process of building the emulator: $g_x(x)$ and $g_{x,i}(\boldsymbol{\theta}_x)$, as in equations 10 and 12, respectively. In equations 12–15, we group the terms so as to separate the parametric and nonparametric parts of the system and obtain the global misfit function $G(x, \boldsymbol{\theta}_x)$, which is a function of offset x and the model parameters $\boldsymbol{\theta}_x$. A more careful treatment of the system would involve considering this dependence. However, on the grounds of simplicity of calibration given the proof-of-concept nature of this study, we chose to compute a misfit function averaged over all model parameters. Thus, we consider the misfit function $G_x(x)$, as shown in equations 16 and 17:

$$t = \sum_{i=0}^{p_x} \sum_{k=1}^{w_x} \sum_{j=0}^{q_x} \beta_{ijk} \theta_{k,x}^{j} \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i \mathrm{e}^{-x} L_i(x) + g_{i,x}(\boldsymbol{\theta}_x) x^i \mathrm{e}^{-x} L_i(x) + g_x(x), \tag{13}$$

$$= \sum_{i=0}^{p_x} \sum_{k=1}^{w_x} \sum_{j=0}^{q_x} \beta_{ijk} \theta_{k,x}^{j} \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i \mathrm{e}^{-x} L_i(x) + \left[\sum_{i=0}^{p_x} (g_{x,i}(\boldsymbol{\theta}_x) x^i \mathrm{e}^{-x} L_i(x)) + g_x(x) \right], \tag{14}$$

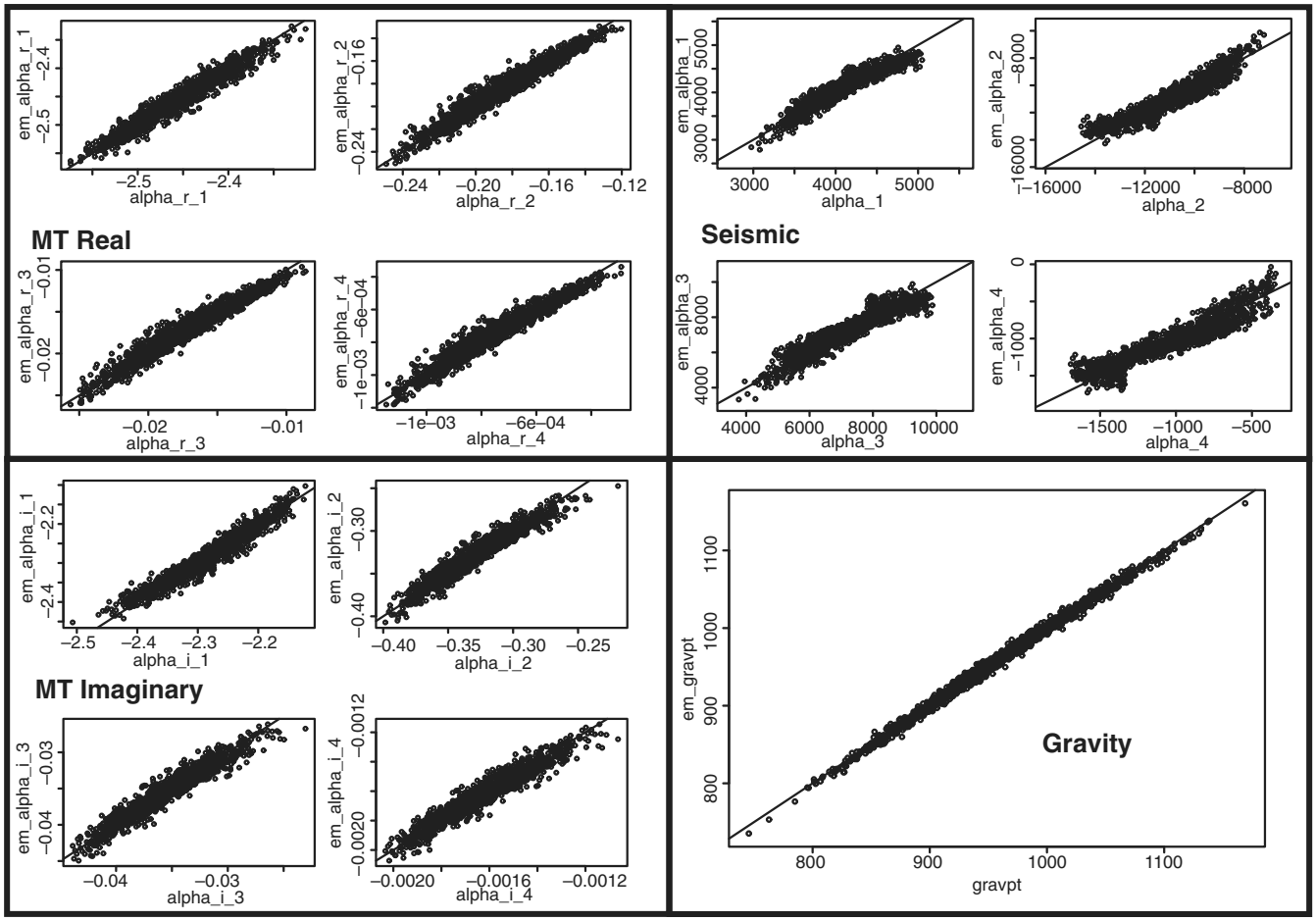

Figure 4. Example reconstruction of α coefficients from β coefficients for ST12. In the case of the gravity emulator, note that we are simply representing a single point, rather than a function represented by α coefficients, and so we plot the emulator-reconstructed points against the "actual" points generated by the gravity simulator. Note the strong correlation between emulated and simulated outputs in each case.

$$= \sum_{i=0}^{p_x} \sum_{k=1}^{w_x} \sum_{j=0}^{q_x} \beta_{ijk} \theta_{k,x}^{j} e^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i e^{-x} L_i(x) + G(x, \theta_x), \tag{15}$$

$$\approx \sum_{i=0}^{p_x} \sum_{k=1}^{w_x} \sum_{j=0}^{q_x} \beta_{ijk} \theta_{k,x}^{j} e^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i e^{-x} L_i(x) + G_x(x), \tag{16}$$

and

$$G_x(x) = \sqrt{\frac{\sum_{n=1}^{n_{\max}} (t_{\mathrm{em},n}(x_n) - t_{\mathrm{sim},n}(x_n))^2}{n_{\max}}}. \tag{17}$$

To properly effect the screening of model space using a predictive simulator proxy, it is necessary to calibrate the uncertainty on the predictor. We calibrate the emulator uncertainty in using $\boldsymbol{\beta}_x$ for prediction by computing $G_x(x)$, as in equation A-8. This is done by generating the $\boldsymbol{\beta}_x$-coefficient estimated output function and the full simulator output for each training model parameter set (Figure 5) and computing the root mean square (rms) of the residuals with respect to the traveltime functions used to train the emulator as a function of x. Examples of this misfit function are shown in Figures 6a–9a.

A key question is, "How many training models are required to correctly estimate $G_x(x)$ and thus sufficiently sample the model space?" This question is vitally important for two reasons: first, because a model's plausibility or implausibility can only be reliably determined if the emulator uncertainty with respect to the estimation of the simulator output is correct and second because the aim at each screening cycle is to exclude model space not deemed plausible; it is crucial to properly sample the whole of the remaining space to prevent the model space from being wrongly removed. If the model coverage is not sufficient, then the emulator will underestimate the predictive uncertainty. A more rigorous study would involve either a more detailed assessment of the space to be sampled or the inclusion in the sampling method of a finite probability of sampling outside the currently constrained space. Using the criterion of two samples/parameter, we would wish to use $2^{14} \approx 16000$ training models (for example, as in Sambridge and Mosegaard, 2002). For our purposes, we chose a semiqualitative and fairly rudimentary approach of considering that if the coverage is sufficient, then the addition of further model parameter sets to the training process will not significantly alter the uncertainty estimate. We therefore calibrated the number of models needed by testing cases of generating the emulator using 150, 1500, and 15,000 training models and assessing the impact on the emulator uncertainty of adding more models to the training process. For emulators trained over our prior model space (Table 1), it was found that using 150 models was insufficient (Figure 5), but that the uncertainty function estimates using 1500 and 15,000 training models give similar uncertainty functions. Over this space, therefore, 1500 models is deemed a sufficient number with which to train the emulator.

The set of $\boldsymbol{\beta}_x$ coefficients and this uncertainty function $G_x(x)$ *together* constitute the emulator, or statistical model. We use this uncertainty function to determine whether emulated output data of a proposed model lie sufficiently close to the observed data sets such that the model can be deemed plausible or not. However, $G_x(x)$ is calculated as the rms of the simulator-predictor residual, and as such it is possible (and indeed, it is certainly the case in some instances) that the actual data-representation error for a given set of model parameters may be significantly larger than this. Hence, it may be the case that potentially plausible models are rejected by the emulator screening simply because the emulator prediction for that set of model parameters was located in the tail of the uncertainty function. The emulator screening reliability is, therefore, tested by using this screening technique on 100 target data sets, produced by the simulator from 100 synthetic models. A scaling factor γ_x for the uncertainty function is then calculated by calibrating against these 100 target data sets, such that there is at least a 97% probability that the emulator screening process will include the "true" model in its selection of plausible models if the true model is included in the candidate model space. The figure of 97% is in many senses arbitrary; however, we considered it suitable for the purpose at hand. The condition for plausibility is shown in equation 18, where

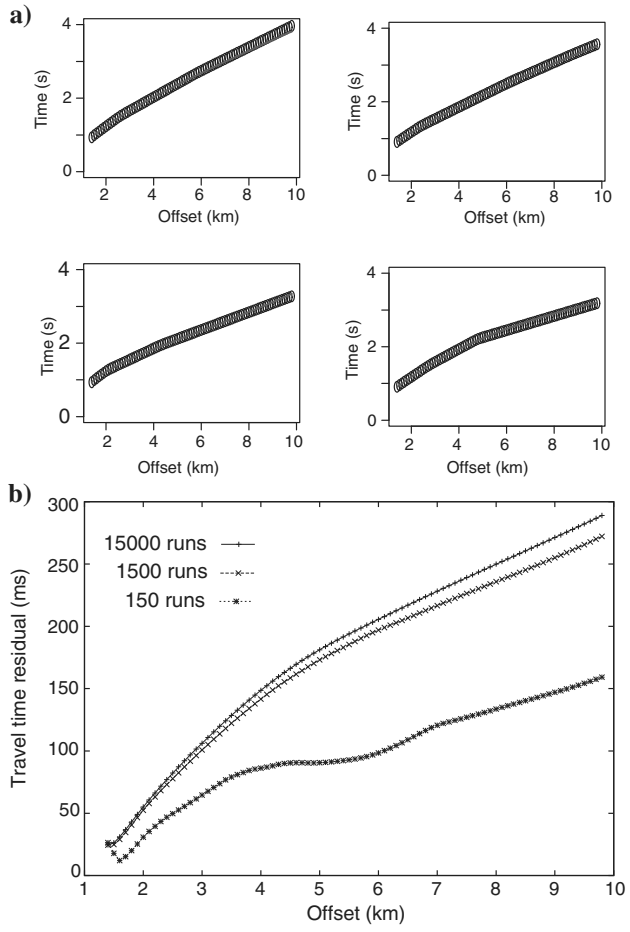

Figure 5. (a) Four example traveltime training outputs and emulator-reconstructed outputs. Black ovals show the traveltimes generated by the full simulator code, and the gray lines show the traveltime curves predicted by applying the predictive β coefficients to the same sets of model parameters. (b) Comparison of seismic emulator uncertainty function for ST13 after eight cycles using 15,000, 1500, and 150 training models.

$t_{\mathrm{em}}(x_n)$ and $t_{\mathrm{targ}}(x_n)$ are the emulated and full simulator traveltimes at an offset x_n, respectively. The weights $\kappa_{x,n}$ are user-defined weights for each traveltime point. For example, we attach greater importance to achieving a close fit to the short-offset traveltimes, compared with the long-offset measurements on the basis that the velocity gradient is typically higher in the shallow structure. Table 4 shows the values of κ used in this study. Here, we have chosen to give all points a weighting of either 1 or 0, and varied the density of points along the offset profile with value 1 to control the weight given to varying parts of the traveltime curves. If preferred, the user could easily use fractional weights:

$$\sum_{n=1}^{n_{\max}} \kappa_{x,n} \frac{\max[|(t_{\mathrm{em}}(x_i) - t_{\mathrm{targ}}(x_i))| - \gamma_x G_x(x_i), 0]}{G_x(x_n) \sum_{p=1}^{n_{\max}} \kappa_{x,p}} < n_{\max}. \tag{18}$$

Spike emulator

To locate discontinuities in the gradient of the seismic traveltime curves and thus constrain abrupt changes in velocity at layer boundaries, a "spike" emulator was built. There are a number of other approaches (Grady and Polimeni, 2010), which could have been taken to identify the boundary positions, such as the basic energy

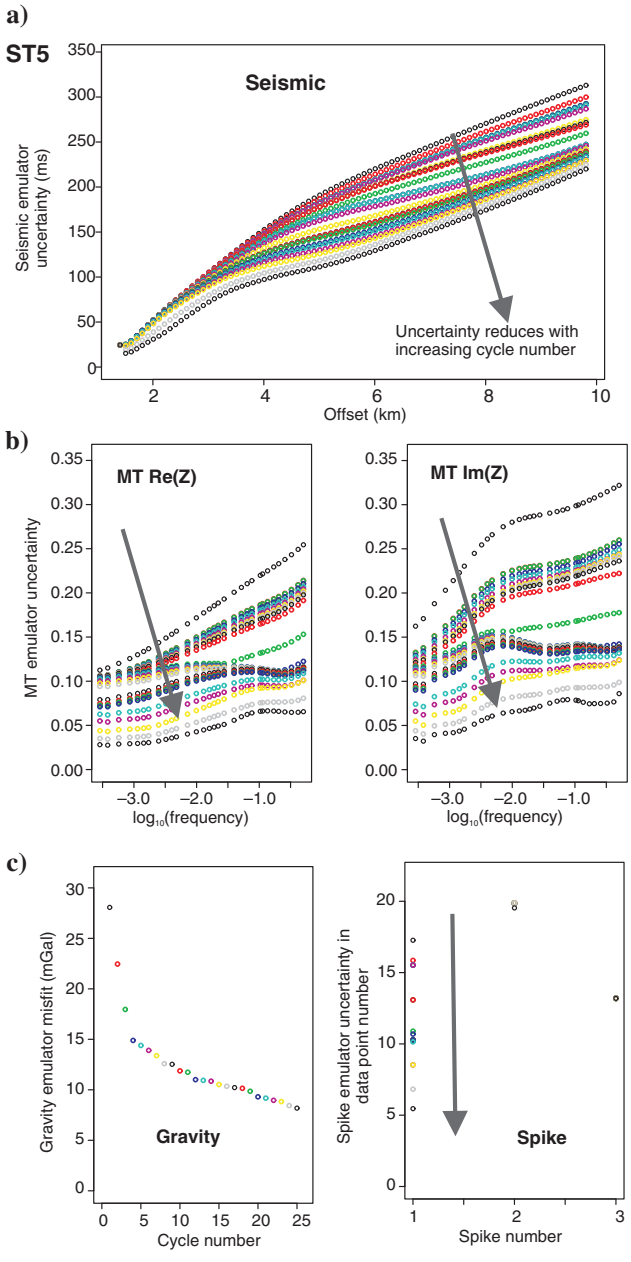

Figure 6. (a) Seismic, (b) MT, and (c) gravity uncertainty functions for ST5. Arrows show how the predictive uncertainty of the emulators reduce with the increasing screening cycle as the model space is refined.

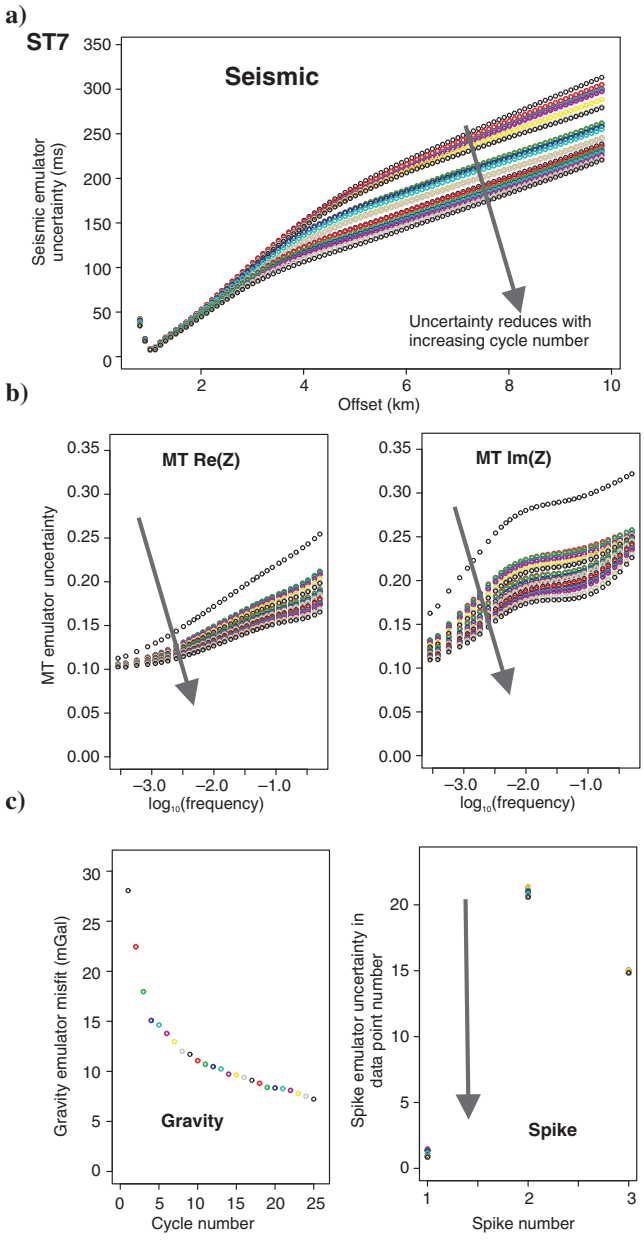

Figure 7. (a) Seismic, (b) MT, and (c) gravity uncertainty functions for ST7. Arrows show how the predictive uncertainty of the emulators reduce with increasing screening cycle as the model space is refined.

model and the total variation model. Each has strengths and weaknesses, particularly regarding how noise is regarded in association with high-frequency data. A key part of the philosophy of our method is that it should be as conceptually straightforward as possible and data driven wherever appropriate, and so we chose to implement the simple gradient detection method, described here. In building the seismic emulator, the chosen form of data reduction of using polynomial curves to represent the t versus x curves, while being suitable for describing the smooth trends (Figure 5), does not capture discontinuities in the traveltime gradient function $\mathrm{d}t/\mathrm{d}x$. Roberts et al. (2012) describe a strategy to consider these gradient discontinuities, whereby the dependence of the offset position of these gradient discontinuities is considered as a function of the seismic model parameters. We adopt the same strategy here, seeking to model such features in the data, and thus capture structural information, with a view to optimizing the positions of the model layer boundaries, thus best representing the substructure.

As in Roberts et al., (2012), instead of considering $\mathrm{d}t/\mathrm{d}x$ to probe this information, we calculate the squared second derivative of the t versus x function $\psi = (\mathrm{d}^2t/\mathrm{d}x^2)^2$ (Figure 10). In principle, given that we are using seven-layer models, to optimize the layer boundary positions, we could search for the six largest spikes. However, the presence of six discernible spikes in many of the observed seismic CMP gathers is unlikely (see Figure 2 for example), and this may yield the positions of noise spikes (the positions of which would likely be uncorrelated to any structural information). To avoid potential computational problems as a result of misattributing structurally sourced gradient discontinuities to noise, we choose to only estimate the offset positions x of the three largest spikes in this $\psi = (\mathrm{d}^2t/\mathrm{d}x^2)^2$ function. We preferentially use $(\mathrm{d}^2t/\mathrm{d}x^2)^2$ as opposed to $\mathrm{d}^2t/\mathrm{d}x^2$ to ensure that ψ is positive, simplifying the process of picking the extrema, in addition to exaggerating the relative magnitudes of the spikes in question. A key assumption of this method is that the largest spikes do represent layer boundaries, rather than noise. For cases in which there is a high degree of noise, it may be necessary to either consider other methods for the detection of structural boundaries or reduce the number of model layers and the expected output resolution.

For each seismic emulator training data set, we therefore compute (numerically) $\psi = (\mathrm{d}^2t/\mathrm{d}x^2)^2$ and then search for the offset x

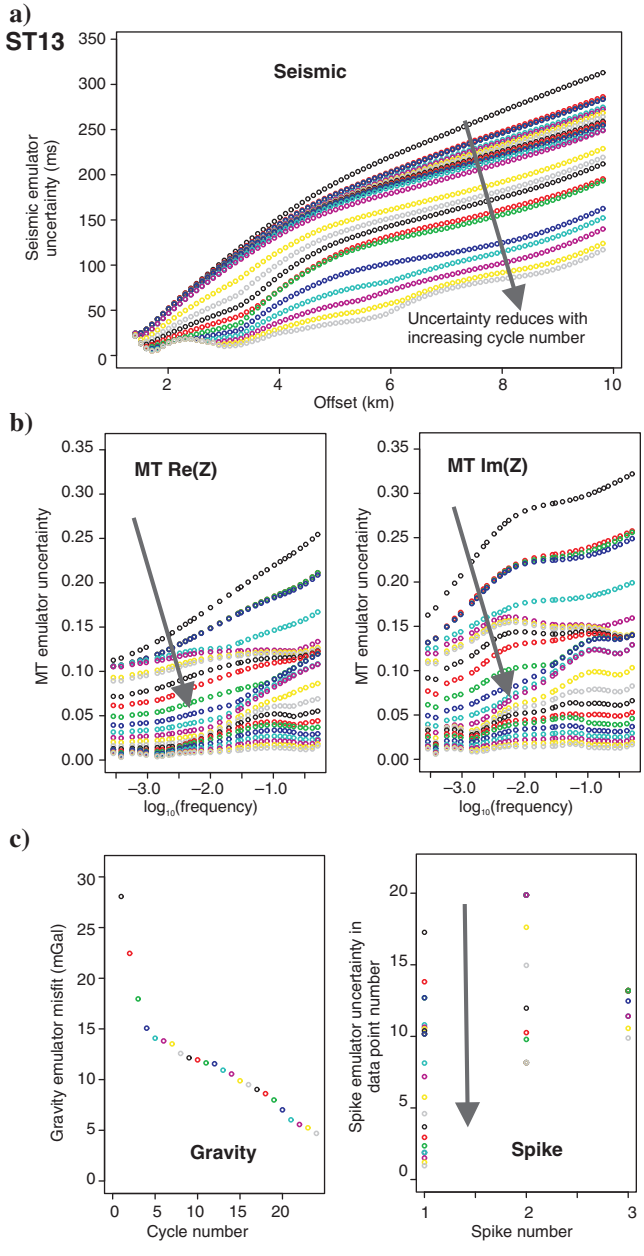

Figure 8. (a) Seismic, (b) MT, and (c) gravity uncertainty functions for ST13. Arrows show how the predictive uncertainty of the emulators reduce with increasing screening cycle as the model space is refined.

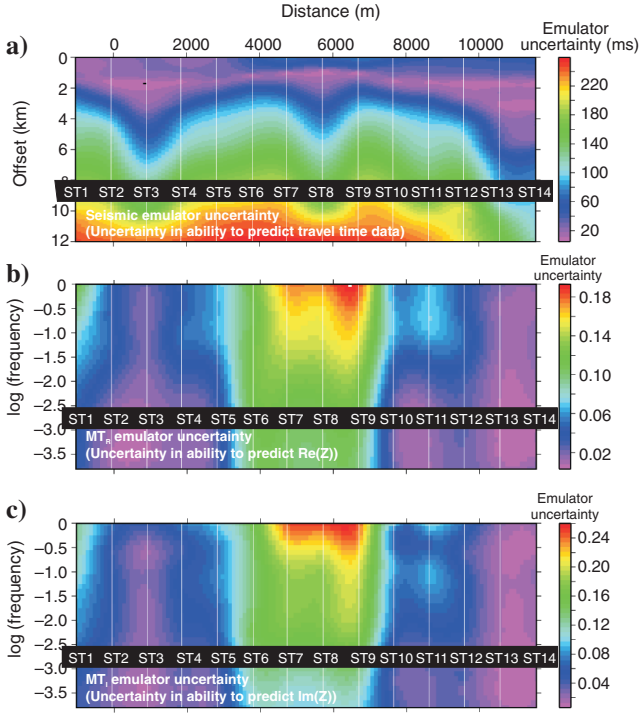

Figure 9. (a) Seismic and (b-c) MT emulator uncertainty functions ($G_x(x)$ and $G_\omega(\omega)$) for each station. These data maps represent the predictive uncertainty of the emulator in predicting the simulator output. White lines show the positions of the stations between which the function is interpolated.

positions of the three largest spikes. These are then fitted to the model parameters $(v_m, s_m)_{m=1}^7$ according to the formulation in equations A-9–A-12 and the predictive uncertainty $G_{\psi,i}$, computed by comparing the "actual" positions of the spikes to the estimates given by $\boldsymbol{\beta}_\psi$.

MT and gravity emulators

We built emulators for each of the MT and gravity data sets in an analogs way; to predict complex impedance and gravitational field strength as a function of the resistivity and density models, respectively (Roberts et al., 2012). For the MT case, unlike with the seismic data, the $(\omega, \mathrm{Re}(Z))$ and $(\omega, \mathrm{Im}(Z))$ functions are smooth functions, and so we do not construct an analog for the seismic spike emulator for the MT data set. In the case of the gravity emulator, because there is simply a single gravity measurement at each location, rather than a function such as (x, t) or $(\omega, \mathrm{Re}(Z))$ or $(\omega, \mathrm{Im}(Z))$, the initial data reduction step is not necessary, and so we simply fit the output simulator gravity value to polynomials in the model parameters, in a similar fashion to the method used to construct the spike emulator.

Screening phase

Having trained an emulator for each of the seismic, spike, gravity, and MT cases using the method described above over models generated from the parameter ranges in Table 1, we generate candidate models to test for implausibility across the three data sets: seismic, gravity, and MT. Our goal is to generate candidate joint models to test for implausibility and thus discern the commonly plausible set of models (as illustrated in Figure 11). To maximize the constraint afforded by the process, rather than performing a single screening phase, we repeat the screening process in a cyclic scheme, each time using the remaining plausible model space from the previous screening cycle to build a new emulator, which due to being trained over a smaller space, will have a smaller $G_x(x)$ and thus be more predictively accurate (see Roberts et al. [2010] and Figures 6a–9).

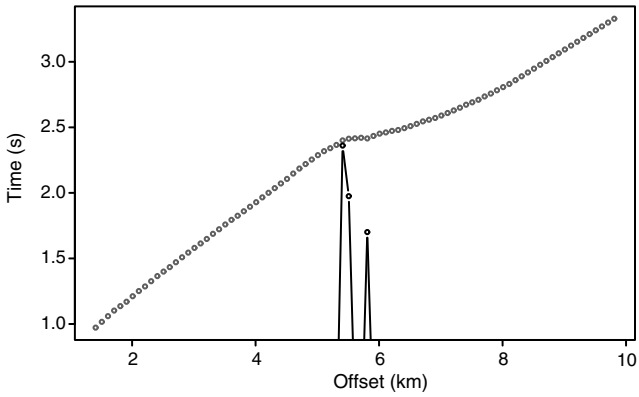

Figure 10. First arrival traveltime picks from station ST12 (gray), with (scaled) $\psi = (\mathrm{d}^2t/\mathrm{d}x^2)^2$ overlaid (black). The aim of the spike emulator is, to within a known uncertainty, predict the positions of the maxima of the (x, ψ) function. In this example, $\psi_{\max}$ are seen at $x \approx 5500$, $x \approx 5900$.

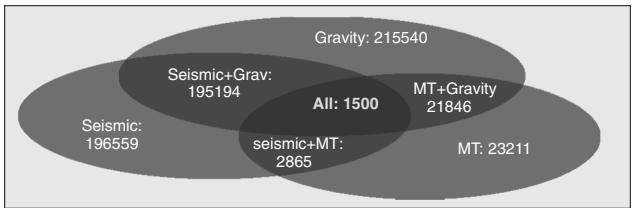

Figure 11. Commonly plausible model statistics using the seismic, gravity, and MT emulators to generate a population of 1500 plausible models for one screening cycle at station ST12. In each case, the numbers show the number of models deemed plausible by each screening method: seismic (including spike), gravity, and MT.

For each of the seismic, spike, gravity, and MT cases, we use the respective emulator to rapidly test sets of model parameters to see if the emulated output from each set of model parameters lies within a given range $\{\gamma_q G_q\}_{q=x,\psi,\omega,\rho}$ of the observed wide-angle traveltime data, gravity measurement, and $\mathrm{Re}(Z)$ and $\mathrm{Im}(Z)$ data observed closest to each of the stations ST1-14.

The generation of joint models may be accomplished in a variety of ways. We choose to generate such 28-parameter models $(v_m, r_m, \rho_m, s_m)_{m=1}^7$ by choosing a set of (v_m, s_m). Resistivity and density values (r_m and ρ_m) are then generated using the appropriate (salt or nonsalt) relationship from v_m according to equations 5–9. In generating joint candidate models; therefore, our first question for each layer of the model we are generating is whether we wish that layer to be characterized by a salt relationship or a sediment relationship (Figure 3). At the start of our analysis, we therefore specify a probability for each layer of the model, p_m with which to generate candidate models using the salt relationship or the sediment relationship. Here, we set this probability to 0.5 for each layer. In the first screening cycle, $(v_m, s_m)_{m=1}^7$ are generated on the fly using a Sobol algorithm (Bratley and Fox, 1988), and then the emulator outputs are tested for plausibility against the observed data set at the station in question. In using common layer thicknesses for the models, we are imposing the additional constraint of structural coherency across the models. We then generate (r_m, ρ_m) from v_m according to the salt/sediment probability vector p_m. Note that in doing so, the prior probability distributions for resistivity and density specified in Table 1 become largely implicit in the screening process; however, in the initial emulator training step, models for each domain are drawn independently of the coupling relationships in equations 5–9.

On generating each candidate model, we use each of the four emulators to generate an estimated data output in each case. We define the condition for implausibility for an individual method as follows: A weighted mean of the emulator-predicted data residuals with respect to the observed data is less than $\gamma_x G_x(x)$, $\gamma_\psi G_\psi$, $\gamma_\omega G_\omega(\omega)$, *or* $\gamma_\rho G_\rho$ as appropriate. A joint model is considered *commonly* plausible, and thus suitable for use in the subsequent cycle, if it is found to be not implausible with respect to all three methods: seismic (including spike), MT, and gravity, based on the plausibility conditions shown in equations A-31–A-35 (Figure 11). By generating and testing model parameter sets in this way, we reject the implausible model space and we build up a population of plausible models. This is then repeated in a cyclic fashion; when 1500 plausible models have been found, they are used to build a new emulator, which is used alongside the previous emulators to screen further models from the reduced model space. A more careful treatment would calibrate the number of training runs for each cycle (by determining when the addition of further training runs has a negligible impact on the uncertainty estimate).

Note that for cycle n, where $n > 1$, the condition for a candidate model to be considered plausible is not simply that it is not considered implausible by the test described using the emulators constructed for cycle n using the plausible models from cycle $n-1$, but that it is also not considered implausible for each of the preceding cycles. Thus, for a given candidate model at cycle n, it must pass the screening test for $4n$ sets of emulated data outputs (one for each of the n cycles, over each of the four methods [seismic, spike, gravity, and MT]). However, as soon as a model is deemed implausible by a single screening, then that candidate model can be discarded and a new candidate model can be generated. To maximize screening efficiency, we choose to compute the emulated output for a given model for whichever of the emulators (seismic, spike, gravity, or MT) is fastest to run, including computation of the plausibility condition (equations A-31–A-35). In this case, testing against the gravity measurement was fastest because it is just a single point.

After the first cycle, to maximize efficiency, rather than using the Sobol sampling strategy, we use a method to sample from the joint model parameter distribution. Sampling from the joint distribution is not a trivial task, and there are a number of ways of accomplishing this, for example, the method of Osypov et al. (2011). One of the key considerations is that although we have 1500 models, which sufficiently sample the model space, we do not wish to sample in a bootstrap manner from this distribution of point values, but from the *continuous* distribution described by these points. If we were to repeatedly sample simply from the distribution of point values from the previous cycle, then after a few cycles, the resulting distribution will tend toward that of a number of discrete spikes. To avoid this issue, we use a scheme of sampling the combinations of velocity and thickness parameters ($v_{m,i}$ and $s_{m,i}$) from the previous cycle, each perturbed by a value sampled from a uniform distribution with a width of 1% of the marginal plausible parameter range from the previous cycle. In this manner, we generate new sets of $(v_m, s_m)_{m=1}^{7}$ values which are close to those deemed plausible in the previous cycle, according to the formulation in equation A-36. The perturbation of 1% was chosen after testing a range of values. The greater the value that is chosen, the greater is the continuity of the overall distribution, at the expense of smearing the information available. We make the perturbation using a uniform distribution rather than a normal distribution to avoid issues relating to leakage, particularly when sampling from close to the bounds of the previous cycle parameter distributions on the grounds that $U(a,b)$ is bounded between a and b, whereas $N(\bar{x},\sigma)$ is unbounded. Other distributions, such as a β-distribution, could be used here, and they may be considered to be better choices; however, in this case, we chose to use a uniform distribution for conceptual and computational simplicity. The choice of perturbation method affects how the distributions are sampled, and thus they are something upon which the final results are dependent, and so they should be given proper thought.

The total number of screening cycles used can be determined by one of several methods: that the size of the emulator uncertainty functions $\{G_q\}_{q=x,\psi,\omega,\rho}$ fall below some threshold value, or that they cease to reduce further (at which point, all discernible parametric information, given our emulator parameterization, has been extracted from the system), or at some arbitrary fixed number. In our case, we choose to use a fixed number of 25 cycles. We consider that setting an arbitrary cutoff in this fashion is not the most rigorous method, and for a robustly interpretable result, this should be given greater consideration; however, we deemed it suitable for the purpose at hand of demonstrating the screening methodology. The result from a screening cycle is the joint distribution of models not deemed implausible. Figure 12 shows example marginal distributions of model parameters for the plausible models after 25 screening cycles for station ST5. These do not represent the full information available from the joint distribution; however, they are useful in understanding how the parameter space is being constrained. We consider it to be of much greater interpretative value, however, to consider the acceptance ratio plots of Figures 13–16 (after Flecha et al., 2013), which show the prior model space and the acceptance ratio for model parameters in parameter-depth space.

After each screening cycle, the population of plausible models is analyzed to ascertain the proportion of models that are characterized by salt or sediment for each layer (this can be considered an extra model parameter). This proportion is then used to update the probability vector $p_{m=1}^{7}$ during the subsequent screening cycle that a particular layer in each candidate model will be generated using the salt or sediment relationship (equations 5–9). The parameter bounds are similarly updated based on the newly refined model space.

By averaging, or computing particular parameter quantiles over the depth range, and interpolating between the station, parameter maps such as Figures 17–19 can be generated. These plots are

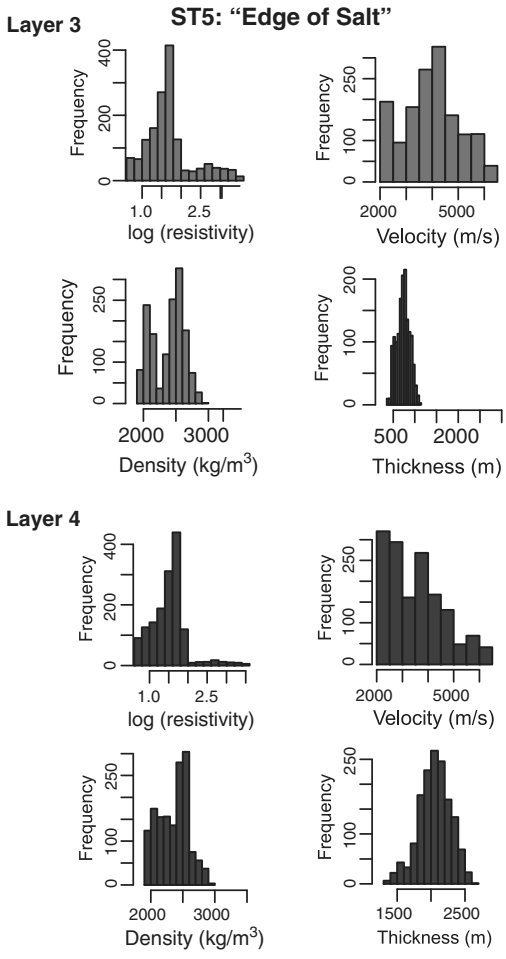

Figure 12. Marginal velocity, density, resistivity, and thickness parameter histograms for ST5, located over the transition zone between predominantly salt (layer 3) and sediment rock (layer 4).

useful for gaining a broad overview of the geophysical features along the profile; however, particularly in cases where there is strong multimodality in the plausible model parameter distributions (as is seen in several cases in Figures 13–16), they can be strongly misleading in that in such cases the average model may lie close to a minimum in acceptance ratio, rather than representing a maximum likelihood estimate. This is considered further in the "Discussion" section.

As well as the ability to consider the distribution of parameter likelihood, this method also allows the construction of a rock probability map, as in Figure 20. This is constructed by calculating the fraction of salt and sediment physical parameter relationships that were deemed plausible at depth nodes down the profile.

Model validation

Well-log data

Figure 16 shows log data from a well located close to station ST5, the location of which is shown in Figure 2. Note that, as can be seen from Figures 17–19, the well track is not vertical at the location of ST5 (in fact, it is much closer to ST6 at depth), and so this overlay is only semiapplicable, and it should not be used for detailed comparison purposes. However, it can be seen from Figures 16–19 that the screening process has done a reasonable job in seeing the structural variation observed in the borehole.

In the case of resistivity, the well-log plot is close to the upper end of the 10%–90% band through most of the depth range and the median quantile is far from the well-log plot. Note that these quantiles are not used for any screening purpose, but they are shown for illustrative and visualization purposes. The presence of the "ground truth" well-log plot toward the edge of the plausible model space further emphasizes that simply adopting some maximum likelihood, mean, or median model can be misleading, and that simply adopting a median/mean/modal model as being representative of the model space would not be appropriate in this case.

The density log overlay of density profile shown in Figures 16 and 17–19 is less useful because the log only starts near the top of the salt body. However, within the joint setting of the screening process, the method correctly identifies the top of the salt body structure. Note that the density log is best described by the lower density quantile shown in Figure 19c, reflecting the likely presence of salt.

Deterministic inversion

Coworkers on the Joint Inversion with Bayesian Analysis (JIBA) project carried out a 3D deterministic inversion of the same joint data set using the method of Moorkamp et al. (2011, 2013). Their result for the velocity profile is shown in Figure 21. For the MT and gravity results associated with the deterministic inversion, the

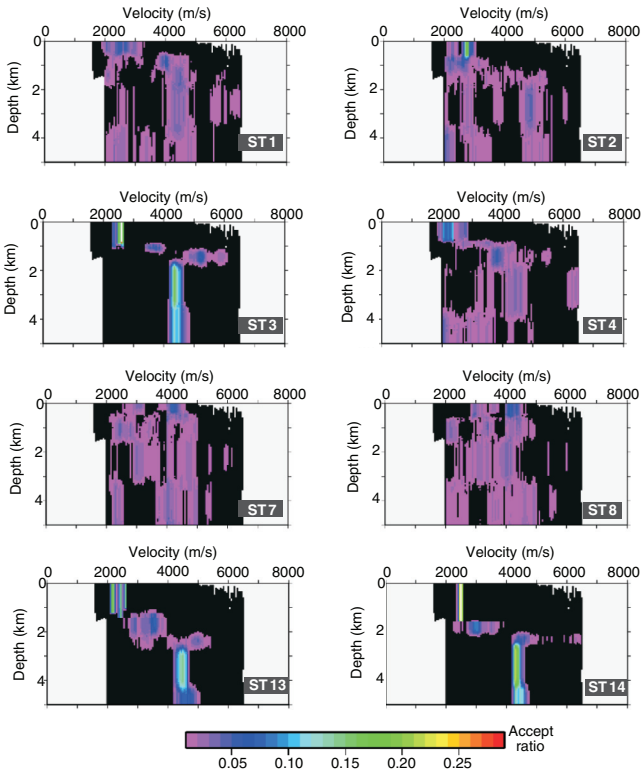

Figure 13. Acceptance ratio of velocity models at a range of stations. The black region shows the area sampled by the prior model space (generated using all the models sampled in the first screening cycle). The multimodality of the velocity distribution at depth in some cases (e.g., ST7–ST8) shows how simply taking a central average value to represent the parameter information would be inappropriate and would hinder the interpretation process.

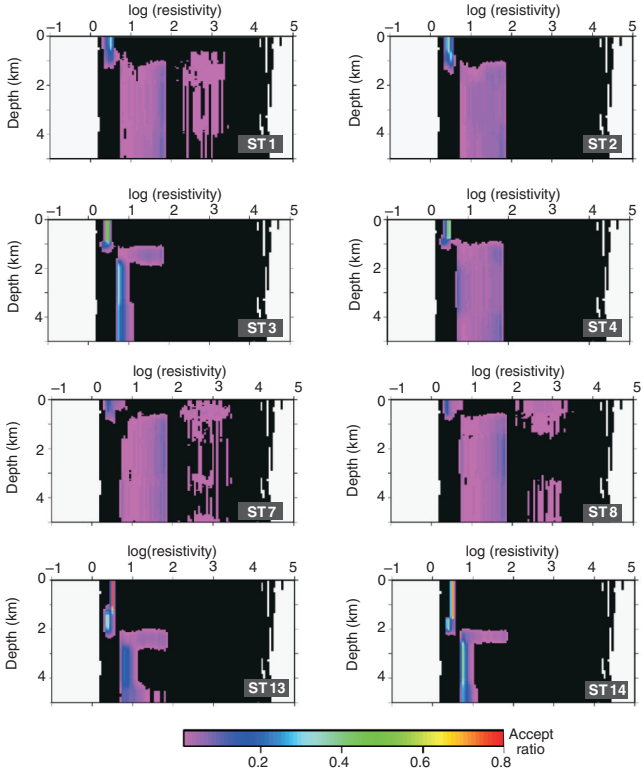

Figure 14. Acceptance ratio of resistivity models at a range of stations. The black region shows the area sampled by the prior model space (generated using all the models sampled in the first screening cycle). The multimodality of the resistivity distribution at depth in some cases (e.g., ST7–8) shows how simply taking a central average value to represent the parameter information would be inappropriate and would hinder the interpretation process.

reader is referred to Moorkamp et al. (2011). Comparing this with Figures 17a and 20, it is seen that the salt body is collocated with that discerned through the model screening method.

DISCUSSION

Emulator versus simulator run time

A key aim of the emulator screening method is that the use of an emulator should afford a considerable time savings over the full simulator code for the computation of outputs for a given set of model parameters. To quantify this, an emulator construction and screening cycle was timed for one of the stations (ST14) for the seismic simulator code (the most complex to fit). To run the 1500 training models took 353 s with the full simulator. In the subsequent screening cycle, 11,243 models were then screened using the seismic emulator in 22 s to obtain 1500 plausible models for the next cycle. The number of models computed per second by the simulator and emulator are thus those given in Table 2.

In this particular case, therefore, the emulator can screen models around 100 times more rapidly than using the full simulator. This illustrates the value of using an emulator in this setting, and indeed although the probem at hand in this case is relatively simple, the more complex the simulator code is, the greater will be the time saving that a well-designed emulator can produce.

Recovery of α coefficients

In general, the recovery of the coefficients (Figure 4) for the seismic emulator is reasonable and there is a clear correlation between the "real" α_x coefficients obtained from fitting curves to the training data sets, and the emulator-reconstructed $\alpha_{x,\mathrm{em}}$, obtained by using the predictive β coefficients with the same model parameter sets. The gravity data points are also well reconstructed. There is a slightly higher scatter on the MT coefficient recovery plots, meaning that the MT emulator appears to be slightly less effective at predicting the form of the output data for a given input model; this does not mean that the MT screening is less reliable. This is because the uncertainty in prediction is absorbed by a larger uncertainty function. Hence, although the rate at which the emulator can exclude plausible model space is lower, the reliability of the screening process itself is not affected. The function of the emulator is not to just rapidly predict the simulator output for a given model, but to do so within a calibrated/known uncertainty. This highlights how conceptually different the emulation Monte Carlo approach is to many of the current schemes, which seek to simply model the system as accurately as possible and find the best model. Here, instead, we seek to iteratively exclude implausible model space, until further exclusion is not possible, and the ability to do this relies not only on the ability to predict the data for a given model, but also on knowing the

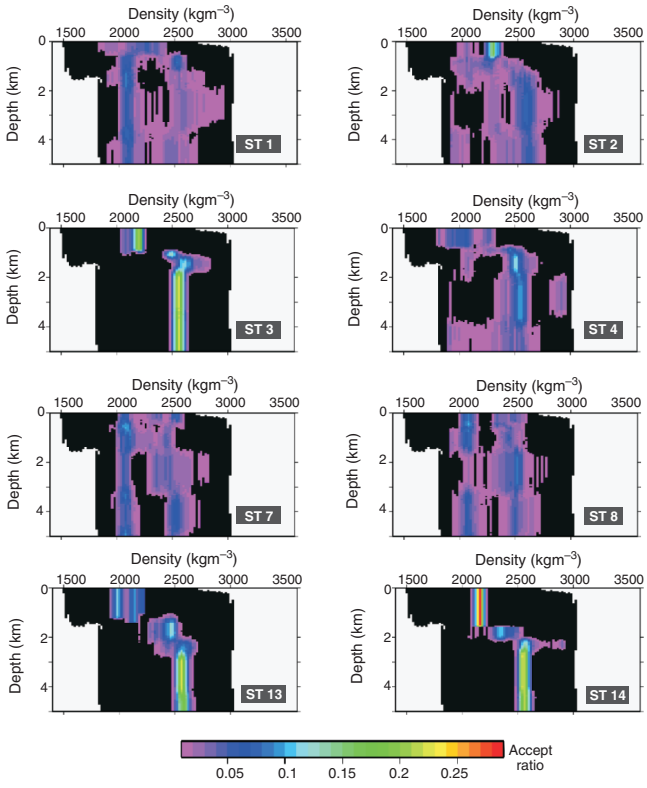

Figure 15. Acceptance ratio of density models at a range of stations. The black region shows the area sampled by the prior model space (generated using all the models sampled in the first screening cycle). The multimodality of the density distribution at depth in some cases (e.g., ST7–8) shows how simply taking a central average value to represent the parameter information would be inappropriate, and would hinder the interpretation process.

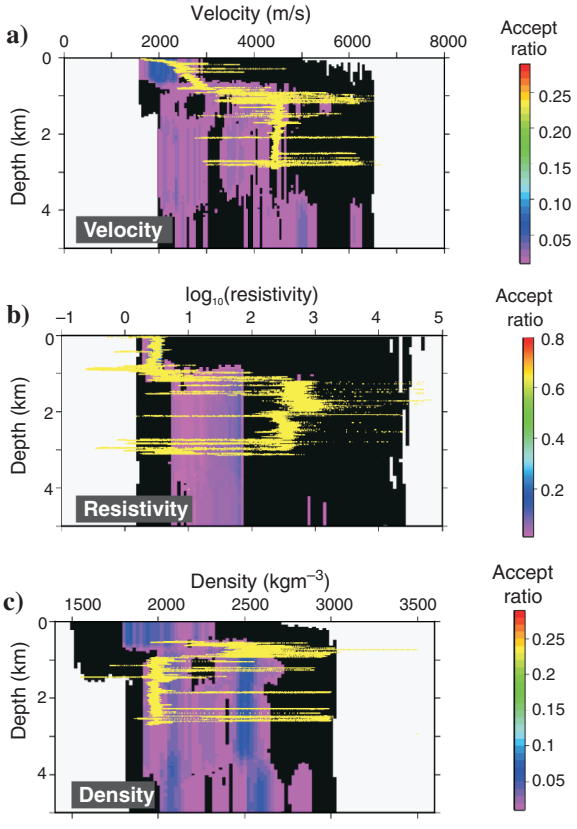

Figure 16. Acceptance ratio of (a) velocity, (b) resistivity, and (c) density models at station ST5, overlaid with the wellbore measurement (yellow line). The black region shows the area sampled by the prior model space (generated using all the models sampled in the first screening cycle). Note that the overlay is not entirely consistent with the location (compare the borehole track in Figures 17–19), and so the comparison should be considered as approximate.

uncertainty of that prediction. The reliability of the emulator in doing this is in fact thus determined as described earlier, by the scaling factor $\{\gamma_q\}_{q=x,\psi,\omega,\rho}$, which is applied to the uncertainty function $\{G_q(q)\}_{q=x,\psi,\omega,\rho}$, which is calibrated when the emulator is built.

Spike emulator methodology

In modeling the positions of discontinuities in the seismic traveltime gradient function, one approach could have been to fit a polynomial function to the derivative dt/dx. However, we chose to adopt the described approach of fitting the offset positions of the spikes in the $\psi = (d^2t/dx^2)^2$ function. This latter approach was favored because, given we are using a polynomial to represent the t versus x function, if we try to fit a polynomial to the derivative of this function, dt/dx, the result of the least-squares fit is likely to be the derivative of the function given by our α-coefficient polynomial representation, which we could calculate analytically, and so we would not gain further useful information. In addition, the parts of the gradient function containing the most useful information are the steepest-turning regions, "which are the most difficult parts to fit using smooth functions. Another advantage of the spike-fitting approach over trying to predict the gradient function itself is that the maximum number of data points we are aiming to fit for an n-layered model is n–1. Whether we choose to fit all n–1 points in this way or as described in the "Methodology" section, in our case, only three data points (the x-positions of the three largest spikes in the $\psi = (d^2t/dx^2)^2$ function), the emulator screening process is considerably more efficient than in the case of fitting the derivative function dt/dx to the entire set of traveltime offsets.

Emulator uncertainty reduction and model space rejection rate

We have seismic data with traveltime offsets to 10 km, so we expect (by rule of thumb) a seismic resolution down to approximately a 3 km depth. In designing the seven-layer space, we there-

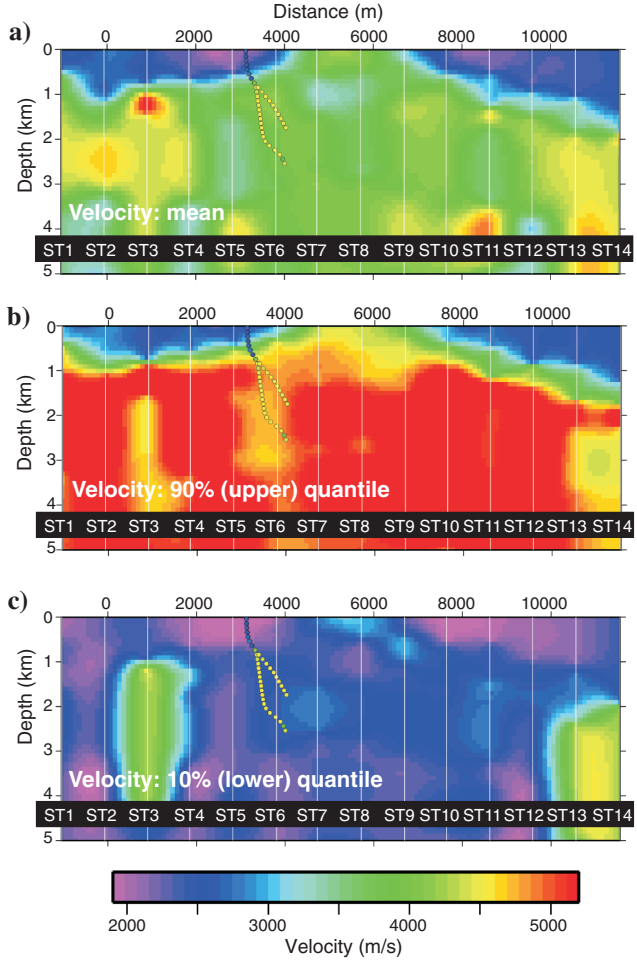

Figure 17. (a) Mean, (b) upper, and (c) lower 90% quantile velocity models. Generated by calculating the distribution of velocity parameters at depth nodes for each station and interpolating between stations. The borehole track is overlaid with colors indicating the log velocity. Note that the log velocity is very similar to the upper velocity quantile in panel (b), consistent with the presence of salt in that region.

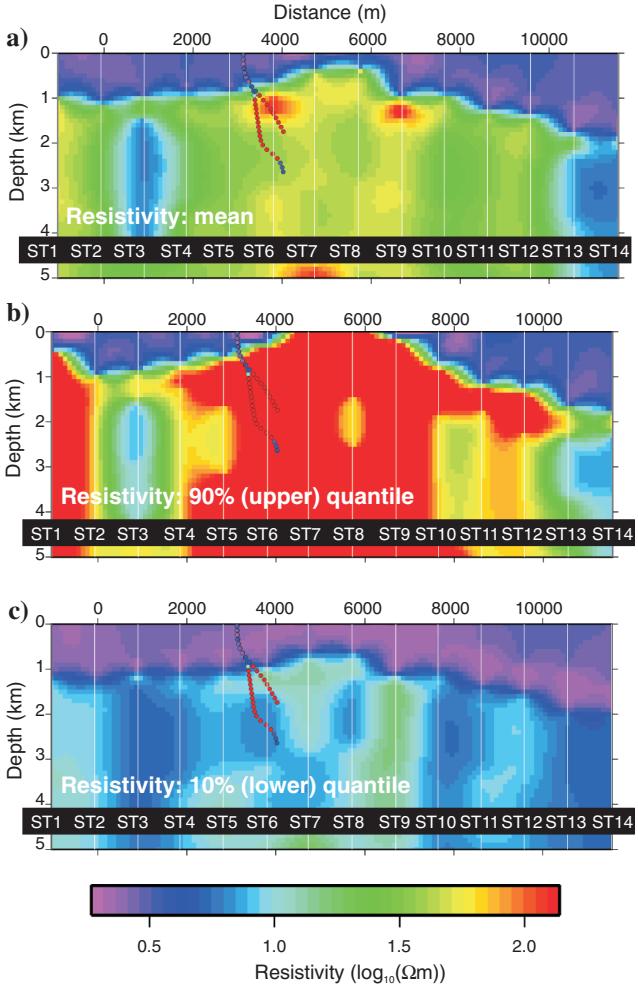

Figure 18. (a) Mean, (b) upper, and (c) lower 90% quantile resistivity models. Generated by calculating the distribution of resistivity parameters at depth nodes for each station and interpolating between stations. The borehole track is overlaid with colors indicating the log resistivity. Note that the log resistivity is very similar to the upper resistivity quantile in panel (b), consistent with the presence of salt in that region.

fore concentrated thin layers toward the upper 1 km of the models (due to the nature of the method, we expect the seismic data to afford the greatest constraint at shallow depths in comparison to the MT and gravity methods). This can be seen in Table 1 by looking at the permitted layer thickness ranges. In the histograms of Figure 12, the prior parameter ranges are represented by the horizontal axis widths of each histogram. These correspond to the ranges shown in Table 1. The fact that the distribution widths of the final plausible parameter ranges are smaller than these ranges in most cases demonstrates how the system is constraining the plausible model space. This is seen even more strikingly in the acceptance ratio plots (Figures 13–16), where the black background shows the prior model space. This model space reduction is also reflected in the uncertainty functions in Figures 6–8, which reduce at each new cycle. The seismic and MT uncertainty functions at the end of the final screening cycle for each station location are shown in the form of predictive data uncertainty maps in Figure 9.

Figures 6–8 show how the predictive uncertainty of the emulator in representing the simulator output reduces with each cycle as the model space is refined. It is clear that the rate of this reduction (also visible in Figure 9) varies across the profile: The rate of uncertainty reduction at station ST7 is much lower than at station ST13, for example. This variation is directly related to the changing plausible model parameter space, and the nature of the interparameter relationships in equations 5–9 and Figure 3; a high uncertainty in these will reduce the rate at which model space is rejected as implausible. In this case, there is a large uncertainty on the resistivity of salt (Figure 3) in comparison with the uncertainty for sediment. Also in the case of salt, there is no correlation between the resistivity, density, and velocity values in that the value of resistivity does not further constrain the value of velocity or density. This can be seen in Figure 3, in which the relationship bounds for salt are simply rectangular boxes, as opposed to the nonzero gradient on the relationships for sediment.

As a result of these two considerations, in regions where there is a strong salt presence, the system will be comparatively slow in rejecting the model space because the three methods (seismic, gravity, and MT) are not strongly mutually cooperative in reducing model space. In the case of sediment, there is lower uncertainty in the relationship and a much stronger correlation between the parameter values. As a result, where there is sediment, the three methods assist one another much more effectively in reducing the size of the model space. The cooperation between the methods can be seen in Figure 8, in which each of the methods is seen to be clearly more effective in reducing the emulator predictive uncertainty at varying screening cycles. This illustrates the value of a joint approach over a single parameter analysis in that each method contributes information that can be used to constrain the parameter values for the other methods. The results of Moorkamp et al. (2011, 2013) demonstrate this point well. In regions where there is salt, the emulator uncertainty reduction rate is much lower (Figure 8), the result of which is that in such regions, sediment models will be rejected less easily. The effect of this is that where salt is present, the probability of salt presence (Figure 20) will be underestimated, or rather, it will be biased toward 0.5, which is the prior specified salt probability across the model. The result is that, due to the nature of each of the physical parameter coupling relationships, the system will more easily discern the presence of sediment than the presence of salt. This is reflected in Figure 20, where the system seems more confident of the suitability of the sediment relationship being appropriate in regions of sediment. As was commented earlier regarding Figure 3, there is a case for perhaps including a third rock-type relationship in the analysis given there are a number of observations lying outside the confidence ranges of the fitted relationships.

Nonmarginal information and sampling strategies

The histograms of Figure 12, however, do not show the full extent by which the parameter space has been shrunk. This is because they simply show the distribution of marginal model parameters for the plausible model space at each cycle. To maximize the efficiency of the sampling scheme, rather than simply sampling parameters from the univariate marginal distributions of each of the layers, as is the case here, the scientist could sample from the full joint distribution of parameters across all layers. This would mean that, for example, emergent correlations between, for example, the velocity of layer 1 and that of layer 2, could be used. Such a scheme is discussed for a synthetic case in Roberts et al. (2010, 2012).

Although the principle of sampling from a prior model space and testing models for plausibility is conceptually straightforward, the manner in which this sampling is carried out (in particular the prior distribution of parameters) requires some careful thought in each case. This is because the shape of this distribution (whether it be normal, uniform, or some other class) is, itself a positive prior statement of belief about the system. It is thus of considerable importance that the sampling strategy and prior parameter distributions are given careful thought before embarking on this kind of method.

Sampling issues — One fundamental weakness of the method as currently used is the assumption that the plausible model space for a given emulation cycle has been sufficiently sampled when 1500 successful models have been found. A proper treatment of the problem would include the consideration that at any given emulation cycle, there is the possibility that plausible areas of model space have not been sampled. A significant improvement to this proof-of-concept methodology would thus include, at each emulation cycle, a finite probability of sampling models outside the currently constrained plausible model space (a "jumping" distribution), as is commonly implemented in Metropolis-Hastings-based sampling schemes (Metropolis et al., 1953; Hastings, 1970).

Given the nature of the screening-cycling method in which the first n successful models are chosen for screening in the next cycle, it should be borne in mind that a weakness of this screening scheme is that over time, for bimodal distribution sampling, such as is the case here where salt and sediment populations are being sampled, there will be bias toward asymmetry when the local maxima are not equal.

In addition to the prior model space and coupling parameters (Table 1 and equations 5–9), the method presented here also includes a number of tuning parameters as inputs (Figure 1), such as the number of models used for each screening cycle (1500) or the perturbation of 1% in sampling model parameters from the previous cycle. As we have presented it here, these tuning parameters have been arbitrarily selected, rather than fully sampled. Because the results are affected by the values chosen for these tuning parameters, although they are probabilistic and the probability map shown in Figure 20 is not a truly Bayesian result in that it does not represent the product of the prior salt probability and a likelihood function. A valuable further development of this methodology would be to consider the sampling of these parameters more rigorously.

Model representation

One of the key advantages of this kind of approach is that, rather than a single average model, and perhaps an estimate of the uncertainty on it, here, the result is the plausible model space. Although the information contained in this result is considerably richer than in the case of an average model and an uncertainty estimate, representing this information presents a challenge. Figures 13–15 (after those of Flecha et al., 2013) show the acceptance ratio for model parameters with depth at a number of stations. It is quickly seen that, while in some cases (e.g., station ST13), there is a clear, well-defined unimodal parameter distribution over most of the depth range, in other cases (particularly where it is thought that there is a considerable salt presence, e.g., stations ST7–8), the distribution of plausible parameters is multimodal. This reflects the fact that, given the observed data and the specified parameter relationships, there may be either sediment or salt present at given depths. This multimodality shows how it can often be inappropriate to represent the resulting geophysical parameter constraint as some uncertainty around a central average value. In recent years, however, a few authors, such as Zhdanov et al. (2012), have made progress in developing methods for joint inversion schemes for multimodal parameter spaces.

This is further highlighted by Figures 16–19. From Figure 16a and 16c, we see that over much more of the depth range, there are two distinct populations of parameters accepted by the screening process; in the case of velocity, the higher valued velocities are from salt models, and the lower valued velocities are from sediment models. In the case of density, the reverse is the case (compare with Figure 3). In this case, it is clear that representing the result by a central average value, where the acceptance ratio is zero, or negligible, would be highly misleading. In this case, it is both of the extrema of the accepted parameters that more appropriately represent the result. Figures 17 and 19 demonstrate this clearly, in that the well-log velocity and density re much better represented by the 90% and 10% quantiles, respectively, as opposed to the means in each case.

Note that in the case of resistivity (Figures 16b and 18), there is a more unimodal output over the depth range. It can also be seen that the system consistently accepts resistivity values lower than does the recorded well log. This is thought to be due to the fact that the MT

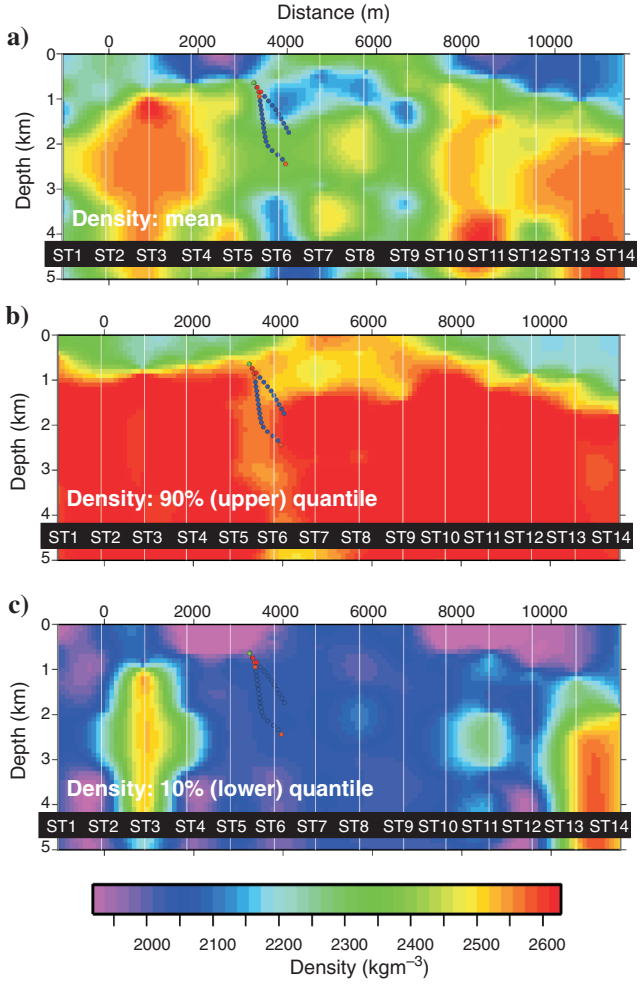

Figure 19. (a) Mean, (b) upper, and (c) lower 90% quantile density models. Generated by calculating the distribution of density parameters at depth nodes for each station and interpolating between stations. The borehole track is overlaid with colors indicating the log density. Note that the log density is very similar to the lower density quantile in panel (c), consistent with the presence of salt in that region.

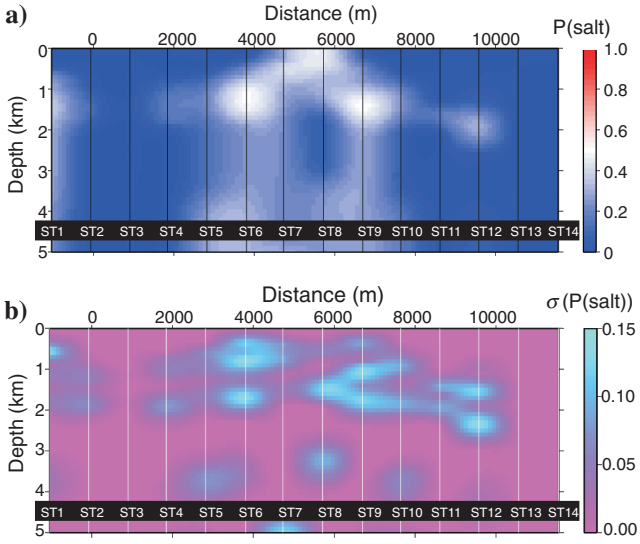

Figure 20. (a) Salt probability map: generated by averaging the model count for "salt" models deemed plausible at depth nodes for each station, and interpolating between stations. (b) Standard deviation of the probability estimate.

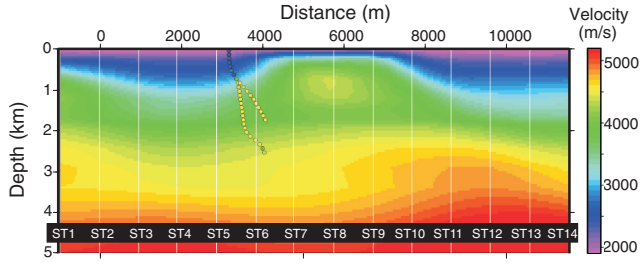

Figure 21. Velocity model result from the joint inversion carried out by Moorkamp et al. (2011, 2013). The borehole track is overlaid with colors indicating the log velocity.

recording instruments are sensitive to conductivity rather than resistivity, so that layers of high conductivity/low resistivity are seen with greater resolution than low-conductivity/high-resistivity layers.

Beyond parameter models: A rock probability map

A major benefit of this model sampling approach is that, from the start, the question of discerning between two types of rock (described by the parameter relationships in equations 5–9) has been integral. Rather than simply seeking to test model parameters, we have tested joint models consistent with either one parameter relationship or another (in this case, sediment or salt). This allows the acceptance ratio for each relationship to be considered and a map of this, as well as of the acceptance spread, to be constructed (Figure 20). This result should be understood in context, in that it gives the model acceptance fraction given all of the prior model space specifications and tuning parameters used in this study. As has been commented in the methodological sections, here, the model space and tuning parameters have been selected somewhat illustratively, with the aim being to focus on the screening method. The result presented here should therefore not be understood as a robust estimation for the region in question. For the result to be geologically meaningful and robust, suitable expert consideration should be given to the prior model space and tuning parameters.

From Figure 20, we immediately see how the system has much more definitely discerned the presence of sediment (indicated by blue colors and acceptance ratios < 0.5) than the presence of salt (red colors and acceptance ratios >0.5). We consider this to be due to the nature of the physical parameter relationships (Figures 3), and it is something worth future investigation. We see that the central region (ST6–9) has a much lower likelihood of sediment presence in the shallow subsurface in comparison to the rest of the profile. It is very striking, on comparing Figure 20 with Figures 13–19, that the region in which there is apparently more salt is not necessarily characterized by a well-defined parameter model (velocity, resistivity, or density). If the uncertainty shown in Figures 13–19 were presented as the result of an inversion, it would likely be considered that the data were at fault, or that (rightly) it was not possible to constrain the region from the observed data. However, on considering Figure 3, we see that it is this very uncertainty in the parameter map, which reflects the fact that salt is more likely to be present. However, as has been noted, the peak acceptance ratio is approximately 0.5, which is the same as the prior probability specification, suggesting that the system has also struggled to add discernment.

Large 3D models and data sets

The study presented here may be considered relatively simplistic given the size of typical 3D models and data sets, and it is a pertinent question as to how this methodology may be scaled up to a fully 3D treatment of the problem at hand. We have shown how using emulators instead of the full simulators can increase greatly the efficiency of screening a large model space. However, the practical usefulness of this method depends on the time required to build the emulators. For large model spaces with large numbers of parameters, even running the number of simulations required to build an emulator may be impractical. In

Table 2. Model computation rate for the seismic simulator versus emulator for one screening cycle.

	Data set realizations/second
Simulator	4.25
Emulator	511.04

Table 3. Mathematical symbols used in this paper.

Symbol	Description
n	Data point reference number (for seismic, gravity, or MT)
$n_{\max}$	Total number of data points (for seismic, gravity, or MT)
x, x_n	Offset at which traveltime t is observed
t, t_n	Traveltime (at data point reference n)
ω, ω_n	Frequency points at which MT impedance Z is observed
Z, Z_n	MT impedance measurements
R, I	Real and imaginary parts of Z
ψ, ψ_n	Squared second-derivative of the traveltime curve
ϕ	Gravity measurements
θ	Model parameters (v, r, ρ, s)
v, v_m	P-wave velocity of layer m
ρ, ρ_m	Density of layer m
r, r_m	Resistivity of layer m
s, s_m	Thickness of layer m
p, p_m	Probability that layer m is "salt"
α, α_i, $\alpha_{q,i}$	Coefficients used to fit curve to data points
β, β_{ijk}, $\beta_{ijk,q}$	Coefficients used to fit $\alpha_{i,q}$ to model parameters θ
f	Function representing the simulator code (seismic, MT, or gravity)
h	Function representing the parametric part of the emulator (seismic, MT, or gravity)
g_q	Theoretical Gaussian residual function for case q
G_q	Computed approximation to g_q
L_i	Laguerre polynomial of order i
$\kappa_{q,n}$	Weighting for data point n in computing plausibility condition
Superscripts p and q and w	Numbers of α and β coefficients and layers
Subscripts $q = x$, ωr, $\omega\imath$, ρ, ψ	Denote seismic, MT (real or imaginary), gravity, or seismic spike domains
Subscripts em, sim, obs, targ	Denote emulated, simulated, observed, and target values

addition, in using an emulator-based method such as this, the scientist is ultimately trading simulator accuracy with the ability to screen a large area of model space. Ultimately, for the problem and purpose at hand, it comes down to the questions of (1) how one can most efficiently extract sufficient information from the system as to which areas of the model space are plausible and which are implausible, and (2) at what stage in the model-screening process it is optimal to use the information contained in the system. A key avenue for investigation in this regard may lie in considering the way in which 3D earth models are parameterized. In this study, for example, we reduced the number of traveltime data points by representing the traveltime functions as polynomials parameterized by a series of α coefficients, on the basis that in general, due to a high sampling density, there is a strong correlation between adjacent traveltime measurements, and particularly in the early stages of model-space screening, the resolution of or information content in the traveltime data is much larger than is necessary to exclude a very large amount of model space. In a similar manner, at the early stages of a typical model-screening process, the information content in a high-resolution 3D model is undoubtedly much larger than is necessary to exclude a very large region of model space. The natural way to take account of this may be to consider adaptive parameterization of the model space, for example, as considered by Trinks et al. (2005) and others, beginning with a coarse parameterization and including more information from the data set and introducing more model parameters as the plausible model space is refined. Along with considering parallelization of the emulator-building and screening processes, we consider this to be fundamental to the future development and practical application of this method.

Emulator automation and tuning considerations

Although the construction and use of emulators give considerable computer runtime savings, in this study, the emulator construction process has itself required considerable investment of user time and thought. In many ways, the designing of an emulator is never quite a one-size-fits-all scenario, for example, the class of functions chosen to represent the data functions (in this case, Laguerre polynomials), the number of emulation cycles to cut off the analysis, the criterion for a model to be considered plausible, or implausible, the number of simulator runs to train an emulator within any given setting, and many others. For a real-world scenario, formal sampling of these tuning parameters would need to be carried out, including the incorporation of insight from an expert geophysicist.

This study has sought to demonstrate the potential for emulator-type technology as an effective tool to facilitate the rapid screening of model space. Although the choice of emulator design can never really be divorced from a consideration of the particular scenario at hand, through further development, including a more robust treatment of the prior model space, and the various tuning parameters, we envisage that it may be possible to develop a semiautomated system for particular types of geophysical settings, parameterizable by, for example, a variety of function classes with which to fit the data to the model parameters, coupling relationships, and implausibility criteria. However, at the moment, this kind of screening approach still seems to be in its infancy in the world of geophysics, and so considerable further work is needed before this could be realized and the method becomes commercially feasible.

Why use this kind of approach

In cases in which system uncertainty is large, and when there are several kinds of joint data sets, as in this example, the ability to discern the full plausible model space greatly adds to the understanding of the system concerned. The advantage of a forward Monte Carlo approach such as that taken here is first that the whole prior plausible model space is considered and second that joint models generated using different physical relationships can be tested. Based on the fraction of models generated using each relationship that are accepted and rejected, conclusions can be made regarding the probability that a particular set of rock properties are present across the profile. This is demonstrated by the salt probability map shown in Figure 20. It is noticeable that although each of the velocity, resistivity, and density models shown in Figures 17–19 may be liable to be individually misinterpreted (and indeed here, each of these is an average model, and so they are not in themselves best fits to the data sets) the probability map shows us that although a wide range of velocity, resistivity, and density models may individually fit the data sets, if one considers the question of which set of physical relationships are preferred across the profile, the result in Figure 20 shows where the presence of salt is most probable.

We consider the ability to not just ask the question of what the optimum velocity, density, or resistivity models are, and rely on intuition to then make judgments, but to ask the direct question "What is the probability, given the data and prior understanding, that salt exists across the profile?" to be extremely powerful. This study demonstrates that it represents a more robust, satisfactory, and useful way of informing decisions than considering a central average parameter model, about which there is some specified uncertainty, which is the currently accepted pseudo-standard in many settings.

Presenting the two results together, the probabilistic analysis described in this paper, and the result from the deterministic inversion (Figure 21), provides a powerful tool for the geologic interpreter. From the deterministic result, we have an optimum map for the rock properties, and from probabilistic modeling, we can make informed judgments about the kind of rock that is present and the uncertainty associated with drawing particular inferences about the rock types present. The implementation of the probabilistic approach presented here has considerable potential for development, technically in terms of incorporating 2D and 3D information, visualisation in terms of viewing joint multimodal information, and in terms of the design of the emulators used to interpolate the model space.

Table 4. The n weights for data points used in computation of plausibility conditions for seismic and MT screening.

Seismic traveltime number (n)	$\kappa_{x,n}$	MT point number (n)	$\kappa_{\omega\{ri\},n}$
−10, 12, 14, 16, 18, 20, 22, 24, 26, 28, 30, 34, 38, 42, 46, 50, 54, 58, 62, 66, 70, 76, 82, 88	1	all (1–20)	1
All others in the range 1–99 not listed above	0		

CONCLUSIONS

In this study, we have demonstrated the use of emulator technology in the field of geophysical constraint. Through an initial investment of training simulator runs, we have built emulators for seismic, gravity, and MT systems, which give a rapid uncertainty-calibrated estimate of the forward code outputs. These emulators give a speed increase of several orders of magnitude over the full forward code and, as a result of the uncertainty calibration, can be used to screen large areas of model space for plausibility.

The increased constraint using several methods in a joint regime over using a single method has been seen in that each method (seismic, MT, and gravity) provides complementary information for the exclusion of an implausible model space.

We have seen that in cases in which a multimodal distribution of plausible model parameters is observed, it is inappropriate, misleading, and incorrect to present the results as some central average value about which there is an uncertainty, as is often considered the normal accepted practice in the field of geophysical imaging.

Screening the whole of the model space, rather than a small part, in a forward scheme such as this allows not simply for the determination of an optimum model. Because the method enables the trialing of alternative candidate rock-physics relationships (which, in many situations, characterize the rock itself, rather than one property of that rock), the question of "With what probability can we say that a given kind of structure exists?" can be directly answered, as shown in Figure 20. Presented perhaps in conjunction with the result of a deterministic parametric joint inversion, this represents a very powerful tool for the purpose of informing geoeconomic decisions, particularly in relation to risk.

ACKNOWLEDGMENTS

The authors would like to thank ITF and the sponsors of the JIBA Consortium, through which this work was funded, for their financial provision and advice. JIBA sponsors are as follows: Statoil, Chevron, ExxonMobil, Nexen, Wintershall, RWE, and Shell. We would also like to thank Statoil in particular for providing a data set on which to test the method. Most of the coding for this study was carried out using R, an open-source statistical coding environment, along with several additional modules (R Development Core Team, 2008; Carnell, 2009; Dutang, 2009). We are also grateful to two anonymous reviewers for their comments on a previous version of this paper, and their suggestion to use Laguerre polynomials as basis functions for fitting the data sets. Inspiration for the acceptance ratio maps of Figures 13–16 came from S. Pearse's Ph.D. thesis.

APPENDIX A

SYMBOLOGY

The mathematical symbols used throughout this paper are defined in Table 3.

Seismic emulator

Here, t refers to seismic traveltime (first-arrival wide-angle refraction), x refers to the source-receiver offset, and v_m and s_m refer to the velocity and thickness of layer m, respectively. The p_x refers to the order of polynomial used in fitting curves to each training data set using the coefficients α_x. Equation A-1 shows how we rerepresent the traveltime curves as a set of polynomial coefficients $\alpha_{i,x}$ and a misfit function $g_x(x)$. Equations A-2 and A-3 show how the data coefficients $\alpha_{i,x}$ are then represented as a polynomial in the model parameters $\boldsymbol{\theta}_x$, parameterized by a further set of polynomial coefficients $\boldsymbol{\beta}_x$, and a further misfit function $g_{i,x}(\boldsymbol{\theta}_x)$:

$$t = \left(\sum_{i=0}^{p_x} \alpha_{i,x} x^i \mathrm{e}^{-x} L_i(x)\right) + g_x(x), \tag{A-1}$$

$$\boldsymbol{\theta}_x = [\,v_1 \;\; v_2 \;\; v_3 \;\; v_4 \;\; v_5 \;\; v_6 \;\; v_7 \;\; s_1 \;\; s_2 \;\; s_3 \;\; s_4 \;\; s_5 \;\; s_6 \;\; s_7\,]^T, \tag{A-2}$$

and

$$\alpha_{i,x} = \left(\sum_{k=1}^{w_x}\sum_{j=0}^{q_x} \beta_{ijk}\theta_{k,x}^j \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x})\right) + g_{i,x}(\boldsymbol{\theta}_x). \tag{A-3}$$

Equations A-4–A-8 show the reorganization of the formulation described in equations A-1–A-3 to compute the global emulator misfit function $G(x, \boldsymbol{\theta}_x)$. In practice, we calculate a variant of this function $G_x(x)$ averaged over all model parameters. The q_x and w_x are the order of polynomial used to write the data coefficients α as a function of the model parameters $\boldsymbol{\theta}$ and the number of model parameters $\boldsymbol{\theta}$, respectively.

$$t = \sum_{i=0}^{p_x}\sum_{k=1}^{w_x}\sum_{j=0}^{q_x} \beta_{ijk}\theta_{k,x}^j \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i \mathrm{e}^{-x} L_i(x) + g_{i,x}(\theta_x) x^i \mathrm{e}^{-x} L_i(x) + g_x(x), \tag{A-4}$$

$$= \sum_{i=0}^{p_x}\sum_{k=1}^{w_x}\sum_{j=0}^{q_x} \beta_{ijk}\theta_{k,x}^j \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i \mathrm{e}^{-x} L_i(x) + \left[\sum_{i=0}^{p_x} (g_{x,i}(\theta_x) x^i \mathrm{e}^{-x} L_i(x)) + g_x(x)\right], \tag{A-5}$$

$$= \sum_{i=0}^{p_x}\sum_{k=1}^{w_x}\sum_{j=0}^{q_x} \beta_{ijk}\theta_{k,x}^j \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i \mathrm{e}^{-x} L_i(x) + G(x, \theta_x), \tag{A-6}$$

$$\approx \sum_{i=0}^{p_x}\sum_{k=1}^{w_x}\sum_{j=0}^{q_x} \beta_{ijk}\theta_{k,x}^j \mathrm{e}^{-\theta_{k,x}} L_i(\theta_{k,x}) x^i \mathrm{e}^{-x} L_i(x) + G_x(x), \tag{A-7}$$

and

$$G_x(x) = \sqrt{\frac{\sum_{n=1}^{n_{\max}} (t_{\mathrm{em},n}(x_n) - t_{\mathrm{sim},n}(x_n))^2}{n_{\max}}}, \tag{A-8}$$

where (t_n, x_n) denote the individual (traveltime, offset) observation points.

Equations A-13–A-30 show the analogs formulation for the MT and gravity emulators. In each case, as with the seismic case, the emulators consist of a set of parametric coefficients $\beta_{ijk,\omega}$ or $\beta_{jk,\rho}$ and an uncertainty function $G_\omega(\omega)$ or G_ρ.

Spike emulator

$$\psi = \left(\frac{\mathrm{d}^2 t}{\mathrm{d}x^2}\right)^2, \tag{A-9}$$

$$x(\psi_{\max_i}) = \left(\sum_{k=1}^{3}\sum_{j=0}^{q_\psi} \beta_{ijk,\psi}\theta_{k,x}^j\right) + g_{i,\psi}(\boldsymbol{\theta}_x), \tag{A-10}$$

$$\approx \left(\sum_{k=1}^{w_\psi}\sum_{j=0}^{q_\psi} \beta_{ijk,\psi}\theta_{k,x}^j\right) + G_{\psi,i}, \tag{A-11}$$

and

$$G_{\psi,i} = \sqrt{\frac{\sum_{n=1}^{n_{\max}} (x(\psi_{\mathrm{em},\max_i}) - x(\psi_{\mathrm{sim},\max_i}))^2}{n_{\max}}}. \tag{A-12}$$

MT emulator

$$R = \log(Z_r) = \left(\sum_{i=0}^{p_\omega} \alpha_{i,\omega_r}(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega))\right) + g_r(\omega), \tag{A-13}$$

$$I = \log(Z_\imath) = \left(\sum_{i=0}^{p_\omega} \alpha_{i,\omega_\imath}(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega))\right) + g_\imath(\omega), \tag{A-14}$$

$$\boldsymbol{\theta}_{\mathrm{MT}} = [r_1 \;\; r_2 \;\; r_3 \;\; r_4 \;\; r_5 \;\; r_6 \;\; r_7 \;\; s_1 \;\; s_2 \;\; s_3 \;\; s_4 \;\; s_5 \;\; s_6 \;\; s_7]^T, \tag{A-15}$$

$$\alpha_{i,\omega} = \left(\sum_{k=1}^{w_\omega}\sum_{j=0}^{q_\omega} \beta_{ijk,\omega}\theta_{k,\omega}^j \mathrm{e}^{-\theta_{k,\omega}} L_j(\theta_{k,\omega})\right) + g_{i,\omega}(\boldsymbol{\theta}_\omega), \tag{A-16}$$

$$R = \sum_{i=0}^{p_{\omega r}}\sum_{k=1}^{w_{\omega r}}\sum_{j=0}^{q_{\omega r}} \beta_{ijk,\omega r}\theta_{k,\omega}^j \mathrm{e}^{\theta_{k,\omega}} L_j(\theta_{k,\omega})(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega)) + \sum_{i=0}^{p_{\omega r}} (g_{i,\omega r}(\boldsymbol{\theta}_\omega)(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega))) + g_{\omega r}(\omega), \tag{A-17}$$

$$= \sum_{i=0}^{p_{\omega r}}\sum_{k=1}^{w_{\omega r}}\sum_{j=0}^{q_{\omega r}} \beta_{ijk,\omega r}\theta_{k,\omega}^j \mathrm{e}^{\theta_{k,\omega}} L_j(\theta_{k,\omega})(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega)) + \left[\sum_{i=0}^{p_{\omega r}} (g_{i,\omega r}(\boldsymbol{\theta}_\omega)(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega))) + g_{\omega r}(\omega)\right], \tag{A-18}$$

$$= \sum_{i=0}^{p_{\omega r}}\sum_{k=1}^{w_{\omega r}}\sum_{j=0}^{q_{\omega r}} \beta_{ijk,\omega r}\theta_{k,\omega}^j \mathrm{e}^{\theta_{k,\omega}} L_j(\theta_{k,\omega}) \times (\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega)) + G(\omega, \boldsymbol{\theta}_\omega), \tag{A-19}$$

$$\approx \sum_{i=0}^{p_{\omega r}}\sum_{k=1}^{w_{\omega r}}\sum_{j=0}^{q_{\omega r}} \beta_{ijk,\omega r}\theta_{k,\omega}^j \mathrm{e}^{\theta_{k,\omega}} L_j(\theta_{k,\omega}) \times (\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega)) + G_\omega(\omega), \tag{A-20}$$

and

$$G_{\omega r}(\omega) = \sqrt{\frac{\sum_{n=1}^{n_{\max}} (R_{\mathrm{em},n}(\omega_n) - R_{\mathrm{sim},n}(\omega_n))^2}{n_{\max}}}, \tag{A-21}$$

where (ω_n, R_n) denotes individual (frequency, impedance) observation points.

$$I = \sum_{i=0}^{p_{\omega\imath}}\sum_{k=1}^{w_{\omega\imath}}\sum_{j=0}^{q_{\omega\imath}} \beta_{ijk,\omega}\theta_{k,\omega}^j \mathrm{e}^{\theta_{k,\omega}} L_j(\theta_{k,\omega})(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega)) + \sum_{i=0}^{p_\omega \imath} (+g_{i,\omega\imath}(\boldsymbol{\theta}_\omega)(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega))) + g_\omega(\omega), \tag{A-22}$$

$$= \sum_{i=0}^{p_{\omega\imath}}\sum_{k=1}^{w_{\omega\imath}}\sum_{j=0}^{q_{\omega\imath}} \beta_{ijk,\omega\imath}\theta_{k,\omega}^j \mathrm{e}^{\theta_{k,\omega}} L_j(\theta_{k,\omega})(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega)) + \left[\sum_{i=0}^{p_{\omega\imath}} (g_{i,\omega\imath}(\boldsymbol{\theta}_{\omega\imath})(\log(\omega))^i \mathrm{e}^{\log(\omega)} L_i(\log(\omega))) + g_\omega(\omega)\right], \tag{A-23}$$

$$= \sum_{i=0}^{p_{\omega t}} \sum_{k=1}^{w_{\omega t}} \sum_{j=0}^{q_\omega} \beta_{ijk,\omega t} \theta_{k,\omega}^{j} e^{\theta_{k,\omega}} L_j(\theta_{k,\omega}) \times (\log(\omega))^i e^{\log(\omega)} L_i(\log(\omega)) + G(\omega, \boldsymbol{\theta}_\omega), \quad \text{(A-24)}$$

$$\approx \sum_{i=0}^{p_{\omega t}} \sum_{k=1}^{w_{\omega t}} \sum_{j=0}^{q_{\omega t}} \beta_{ijk,\omega t} \theta_{k,\omega}^{j} e^{\theta_{k,\omega}} L_j(\theta_{k,\omega}) \times (\log(\omega))^i e^{\log(\omega)} L_i(\log(\omega)) + G_\omega(\omega), \quad \text{(A-25)}$$

and

$$G_{\omega t}(\omega) = \sqrt{\frac{\sum_{n=1}^{n_{\max}} (I_{\mathrm{em},n}(\omega_n) - I_{\mathrm{sim},n}(\omega_n))^2}{n_{\max}}}, \quad \text{(A-26)}$$

where (ω_n, I_n) denotes individual (frequency, impedance) observation points.

Gravity emulator

$$\boldsymbol{\theta}_\rho = [\rho_1 \ \rho_2 \ \rho_3 \ \rho_4 \ \rho_5 \ \rho_6 \ \rho_7 \ s_1 \ s_2 \ s_3 \ s_4 \ s_5 \ s_6 \ s_7]^T, \quad \text{(A-27)}$$

$$\phi = \left(\sum_{k=1}^{w_\rho} \sum_{j=0}^{q_\rho} \beta_{jk,\rho} \theta_{k,\rho}^{j} \right) + g_\rho(\boldsymbol{\theta}_\rho), \quad \text{(A-28)}$$

$$\approx \sum_{k=1}^{w_\rho} \sum_{j=0}^{q_\rho} \beta_{jk,\rho} \theta_{k,\rho}^{j} + G_\rho, \quad \text{(A-29)}$$

and

$$G_\rho = \sqrt{\frac{\sum_{n=1}^{n_{\max}} (\phi_{\mathrm{em},n} - \phi_{\mathrm{sim},n})^2}{n_{\max}}}, \quad \text{(A-30)}$$

where ϕ_n denotes individual gravity observations.

Plausibility conditions

The seismic screening plausibility condition is as follows:

$$\sum_{n=1}^{n_{\max}} \kappa_{x,n} \frac{\max[|(t_{\mathrm{em}}(x_n) - t_{\mathrm{obs}}(x_n))| - \gamma_x G_x(x_n), 0]}{G_x(x_n) \sum_{p=1}^{n_{\max}} \kappa_{x,p}} < n_{\max}. \quad \text{(A-31)}$$

The weights $\kappa_{x,n}$ are shown in Table 4.

For the seismic spike screening, the plausibility condition that must be met for each of the i spikes is as follows:

$$|x_{\mathrm{em}}(\psi_{\max_i}) - x_{\mathrm{obs}}(\psi_{\max_i})| - \gamma_\psi G_{\psi,i} < 0. \quad \text{(A-32)}$$

For the MT screening, the plausibility condition is that both of the following conditions must be met:

$$\sum_{n=1}^{n_{\max}} \kappa_{\omega r,n} \frac{\max[|(R_{\mathrm{em}}(\omega_n) - R_{\mathrm{obs}}(\omega_n))| - \gamma_\omega G_{\omega r}(\omega_n), 0]}{G_{\omega r}(\omega_n) \sum_{p=1}^{n_{\max}} \kappa_{\omega r,p}} < n_{\max}, \quad \text{(A-33)}$$

and

$$\sum_{n=1}^{n_{\max}} \kappa_{\omega t,n} \frac{\max[|(I_{\mathrm{em}}(\omega_n) - I_{\mathrm{obs}}(\omega_n))| - \gamma_\omega G_{\omega t}(\omega_n), 0]}{G_{\omega t}(\omega_n) \sum_{p=1}^{n_{\max}} \kappa_{\omega t,p}} < n_{\max}, \quad \text{(A-34)}$$

and The weights $\kappa_{\omega\{rt\},n}$ are shown in Table 4.

The gravity screening plausibility condition is as follows:

$$|\phi_{\mathrm{em}} - \phi_{\mathrm{obs}}| - \gamma_\rho G_\rho < 0. \quad \text{(A-35)}$$

Joint sampling method

The methodology for generating a new set of velocity and thickness values θ_i' from the joint parameter distribution from the previous screening cycle (where the parameter values are given by θ_i) is shown in equation A-36. The values $\theta_{i,\max}$ and $\theta_{i,\min}$ are the maximum and minimum bounds of $\boldsymbol{\theta}$ from the previous cycle as follows:

$$\theta_i' = \theta_i + U(-0.01(\theta_{i,\max} - \theta_{i,\min}), 0.01(\theta_{i,\max} - \theta_{i,\min})). \quad \text{(A-36)}$$

REFERENCES

Airy, G. B., 1855, On the computation of the effect of the attraction of mountain masses: Philosophical Transactions of the Royal Society, **145**, 101–104, doi: 10.1098/rstl.1855.0003.

Bayarri, M. J., J. O. Berger, E. S. Calder, K. Dalbey, S. Lunagomez, A. K. Patra, B. Pitman, E. T. Spiller, and R. L. Wolpert, 2009, Using statistical and computer models to quantify volcanic hazards: Technometrics, **51**, 402–413, doi: 10.1198/TECH.2009.08018.

Berdichevsky, M. N., and V. I. Dmitriev, 2002, Magnetotellurics in the context of the theory of ill-posed problems: SEG.

Birch, F., 1966, Compressibility, elastic constants, in Handbook of physical constants: Geological Society of America Memoir 97, 97–174, doi: 10.1130/MEM97.

Bodin, T., M. Sambridge, H. Tkali, P. Arroucau, K. Gallagher, and N. Rawlinson, 2012, Transdimensional inversion of receiver functions and surface wave dispersion: Journal of Geophysical Research: Solid Earth, **117**.

Bratley, P., and B. L. Fox, 1988, Algorithm 659: Implementing Sobol's quasirandom sequence generator: ACM Transactions on Mathematical Software, **14**, 88–100, doi: 10.1145/42288.214372.

Carnell, R., 2012, lhs: Latin Hypercube Samples, R package version 0.10, http://CRAN.R-project.org/package=lhs.

Dutang, C., and P. Savicky, 2015, randtoolbox: Generating and testing random numbers, R package version 1.17.

Flecha, I., R. Carbonell, and R. W. Hobbs, 2013, Study on the limitations of travel-time inversion applied to sub-basalt imaging: Solid Earth, **4**, 543–554, doi: 10.5194/se-4-543-2013.

Gallardo, L. A., and M. A. Meju, 2004, Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints: Journal of Geophysical Research, **109**, B03311, doi: 10.1029/2003JB002716.

Goldstein, M., and D. A. Wooff, 2007, Bayes linear statistics: Theory and methods: Wiley.

Grady, L., and J. Polimeni, 2010, Discrete calculus: Springer.

Hastings, W. K., 1970, Monte Carlo sampling methods using Markov chains and their applications: Biometrika, **57**, 97–109, doi: 10.1093/biomet/57.1.97.

Heincke, B., M. Jegen, J. Chen, and R. W. Hobbs, 2006, Joint inversion of MT, gravity and seismic data applied to sub-basalt imaging: 76th Annual International Meeting, SEG, Expanded Abstracts, 784–789.

Hobro, J., S. Singh, and T. Minshull, 2003, Three-dimensional tomographic inversion of combined reflection and refraction seismic traveltime data: Geophysical Journal International, **152**, 79–93, doi: 10.1046/j.1365-246X.2003.01822.x.

James, M. B., and M. H. Ritzwoller, 1999, Feasibility of truncated perturbation expansions to approximate Rayleigh-wave eigenfrequencies and eigenfunctions in heterogeneous media: Bulletin of the Seismological Society of America, **89**, 433–441.

Jegen-Kulcsar, M., R. W. Hobbs, P. Tarits, and A. Chave, 2009, Joint inversion of marine magnetotelluric and gravity data incorporating seismic constraints: Preliminary results of sub-basalt imaging off the Faroe Shelf: Earth and Planetary Science Letters, **282**, 47–55, doi: 10.1016/j.epsl.2009.02.018.

Kennedy, M., and A. O'Hagan, 2001, Bayesian calibration of computer models (with discussion): Journal of the Royal Statistical Society, Series B, **63**, 425–464, doi: 10.1111/rssb.2001.63.issue-3.

Li, H., S. Chen, Y. Yang, and P. Tontiwachwuthikui, 2012, Estimation of relative permeability by assisted history matching using the ensemble Kalman filter method: Journal of Canadian Petroleum Technology, **51**, 205–214, doi: 10.2118/156027-PA.

Logemann, K., J. O. Backhaus, and I. H. Harms, 2004, SNAC: A statistical emulator of the north-east Atlantic circulation: Ocean Modelling, **7**, 97–110, doi: 10.1016/S1463-5003(03)00039-8.

McKay, M. D., R. J. Beckman, and W. J. Conover, 1979, A comparison of three methods for selecting values of input variables in the analysis of output from a computer code: Technometrics, **21**, 239–245.

Meier, U., A. Curtis, and J. Trampert, 2007, Global crustal thickness from neural network inversion of surface wave data: Geophysical Journal International, **169**, 706–722, doi: 10.1111/gji.2007.169.issue-2.

Metropolis, N., A. W. Rosenbluth, M. N. Rosenbluth, A. H. Teller, and E. Teller, 1953, Equations of state calculations by fast computing machines: Journal of Chemical Physics, **21**, 1087–1091, doi: 10.1063/1.1699114.

Moorkamp, M., B. Heincke, M. Jegen, A. W. Roberts, and R. W. Hobbs, 2011, A framework for 3-D joint inversion of MT, gravity and seismic refraction data: Geophysical Journal International, **184**, 477–493, doi: 10.1111/gji.2010.184.issue-1.

Moorkamp, M., A. W. Roberts, M. Jegen, and R. W. Hobbs, 2013, Verification of velocity-resistivity relationships derived from structural joint inversion with borehole data: Geophysical Research Letters, **40**, 3596–3601, doi: 10.1002/grl.50696.

Murtha, J. A., 1994, Incorporating historical data into Monte Carlo simulation: SPE Computer Applications, **6**, 11–17, doi: 10.2118/26245-PA.

Osypov, K. D., D. Nichols, M. Woodward, O. Zdraveva, F. Qiao, E. Yarman, Y. Yang, Y. Liu, and N. Ivanova, 2011, From quantifying seismic uncertainty to assessing E&P risks and the value of information: 81st Annual International Meeting, SEG, Expanded Abstracts, 3683–3688.

Podvin, P., and I. Lecomte, 1991, Finite difference computation of traveltimes in very contrasted velocity models: A massively parallel approach and its associated tools: Geophysical Journal International, **105**, 271–284, doi: 10.1111/gji.1991.105.issue-1.

Press, F., 1970, Earth models obtained by Monte Carlo inversion: Journal of Geophysical Research, **75**, 6575–6581, doi: 10.1029/JB075i032p06575.

Rasmussen, C. E., and C. K. I. Williams, 2010, Gaussian processes for machine learning: The MIT Press.

R Development Core Team, 2014, R: A language and environment for statistical computing: R Foundation for Statistical Computing, http://www.R-project.org/.

Roberts, A., R. Hobbs, M. Goldstein, M. Moorkamp, M. Jegen, and B. Heincke, 2010, Emulation: A Bayesian tool for joint inversion: 80th Annual International Meeting, SEG, Expanded Abstracts, 2805–2809.

Roberts, A. W., R. W. Hobbs, M. Goldstein, M. Moorkamp, M. Jegen, and B. Heincke, 2012, Crustal constraint through complete model space screening for divers geophysical datasets facilitated by emulation: Tectonophysics, **572–573**, 47–63, doi: 10.1016/j.tecto.2012.03.006.

Rougier, J. C., 2008, Comment on "Inferring climate system properties using a computer model," by Sansó et al: Bayesian Analysis, **3**, 45–56, doi: 10.1214/08-BA301B.

Roy, L., M. K. Sen, K. McIntosh, P. L. Stoffa, and Y. Nakamura, 2005, Joint inversion of first arrival seismic travel-time and gravity data: Journal of Geophysics and Engineering, **2**, 277–289, doi: 10.1088/1742-2132/2/3/011.

Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysical inverse problems: Reviews of Geophysics, **40**, 1–24, doi: 10.1029/2000RG000089.

Shapiro, N. M., and M. H. Ritzwoller, 2002, Monte Carlo inversion for a global shear-velocity model of the crust and upper mantle: Geophysical Journal International, **151**, 88–105, doi: 10.1046/j.1365-246X.2002.01742.x.

Stein, M., 1987, Large sample properties of simulations using Latin hypercube sampling: Technometrics, **29**, 143–151, doi: 10.1080/00401706.1987.10488205.

Tarantola, A., 2005, Inverse problem theory and methods for model parameter estimation: SIAM.

Trinks, I., S. C. Singh, C. H. Chapman, P. J. Barton, M. Bosch, and A. Cherrett, 2005, Adaptive traveltime tomography of densely sampled seismic data: Geophysical Journal International, **160**, 925–938, doi: 10.1111/gji.2005.160.issue-3.

Vernon, I., R. Bower, and M. Goldstein, 2009, Calibrating the universe: A Bayesian uncertainty analysis of a galaxy simulation: Galform Project.

Vernon, I., and M. Goldstein, 2009, Bayes linear analysis of imprecision in computer models, with application to understanding galaxy formation: Proceedings of the Sixth International Symposium on Imprecise Probability: Theories and Applications, 441–450.

Zhdanov, M. S., A. Gribenko, and G. Wilson, 2012, Generalized joint inversion of multimodal geophysical data using Gramian constraints: Geophysical Research Letters, **39**, L09301, doi: 10.1029/2012GL051233.

Zubarev, D. I., 2009, Pros and cons of applying proxy-models as a substitute for full reservoir simulations: Presented at the SPE Annual Technical Conference and Exhibition.