

Crustal constraint through complete model space screening for diverse geophysical datasets facilitated by emulation

A.W. Roberts^{1,2}, R.W. Hobbs¹, M. Goldstein²,
M. Moorkamp³, M. Jegen³, B. Heincke³

¹*Dept. of Earth Sciences, Durham University, Durham, DH1 3LE, UK;*

²*Dept. of Mathematics, Durham University, Durham, DH1 3LE, UK;*

³*IFM-Geomar, Wischoffstrasse, Kiel, Germany*

Abstract

Deep crustal constraint is often carried out using deterministic inverse methods, sometimes using seismic refraction, gravity and electromagnetic datasets in a complementary or “joint” scheme. With increasingly powerful parallel computer systems it is now possible to apply joint inversion schemes to derive an optimum model from diverse input data. These methods are highly effective where the uncertainty in the system is small. However, given the complex nature of these schemes it is often difficult to discern the uniqueness of the output model given the noise in the data, and the application of necessary regularization and weighting in the inversion process means that the extent of user prejudice pertaining to the final result may be unclear. We can rigorously address the subject of uncertainty using standard statistical tools but these methods also become less feasible if the prior model space is large or the forward simulations are computationally expensive. We present a simple Monte Carlo scheme to screen model space in a fully joint fashion, in which we replace the forward simulation with a fast and uncertainty-calibrated math-

emational function, or emulator. This emulator is used as a proxy to run the very large number of models necessary to fully explore the plausible model space. We develop the method using a simple synthetic dataset then demonstrate its use on a joint data set comprising first-arrival seismic refraction, MT and scalar gravity data over a diapiric salt body. This study demonstrates both the value of a forward Monte Carlo approach (as distinct from a search-based or conventional inverse approach) in incorporating all kinds of uncertainty in the modelling process, exploring the entire model space, and shows the potential value of applying emulator technology throughout geophysics. Though the target here is relatively shallow, the methodology can be readily extended to address the whole crust.

Keywords: Bayesian, statistical methods, emulation, joint inversion, salt diapir, crustal imaging

1 **1. Introduction**

2 *1.1. Methodological background*

3 A widely used approach for determining deep crustal structure is to use a
4 deterministic non-linear inverse method (Zelt and Barton, 1998; Hole et al.,
5 2006; Roberts et al., 2009). A forward simulator code is used to compute syn-
6 thetic data and by seeking to minimise an objective function, which normally
7 includes residuals with respect to an observed dataset and some regulariza-
8 tion and smoothing terms, an update to the model is computed. This is
9 repeated iteratively until an acceptable value of data misfit, normally mea-
10 sured by the χ^2 parameter, is obtained. Where several kinds of data have
11 been recorded at the same location (e.g. seismic refraction, seismic reflection,

12 gravity, electromagnetic) these may all be used together to constrain a com-
13 mon model across various physical parameters; seismic velocity, density, re-
14 sistivity, where the parameters are coupled by some relationship (Moorkamp
15 et al., 2011). The deterministic inverse approach works well when there is
16 clear justification for the use of a particular regularization and smoothing
17 regime and a good prior understanding about the region of model space of
18 where the optimum model is to be found (i.e. the user can be confident that
19 the process will find the global minimum in the objective function, rather
20 than merely a local minimum). In a joint setting the method also works
21 well where there is no uncertainty regarding the parameter coupling. How-
22 ever, often such clarity and certainty is not possible, but in order to obtain
23 a result, overly subjective assertions are made about the degree of regular-
24 ization, smoothing, coupling, and data uncertainty. In such cases, there is
25 the strong possibility that unquantified user bias may influence the result.
26 So while linearised inversion methods are best suited to obtaining a single
27 optimum result, they do not facilitate a rigorous treatment of the uncertainty
28 associated with a system. In general it is the case that rather than a single
29 optimum model, because of various kinds of uncertainties associated with the
30 system, many structures could give rise to the observed data. Understanding
31 not just an optimum model, but the whole plausible model set, along with
32 a clear understanding of the prior beliefs we are imposing on the constraint
33 process, is important for making inference about the deep crust.

34 In recent decades, with computational development, and the importance
35 of more fully treating the uncertainty associated with geophysical results, sta-
36 tistical determination schemes are increasingly being used to constrain earth

37 structure. The most widely known and used methodology is that of the
38 Markov Chain Monte Carlo (MCMC) method, where a point in model space
39 is chosen and parameters updated according to some sampling scheme, such
40 as that of Metropolis-Hastings-Gibbs (Hastings, 1970; Smith and Roberts,
41 1993), in order to maximise the *likelihood* function. However, although re-
42 cent computational advances have made it possible to handle larger problems,
43 it is still the case that these methods are still really only practicable when the
44 number of model parameters and the size of the model space is *comparatively*
45 small (Sambridge and Mosegaard, 2002). This is because if the number of
46 parameters is large (commonly $> 10^5$), the number of (complex and there-
47 fore time-expensive) forward simulations required to sample the model space
48 and properly build up the posterior probability density functions becomes
49 infeasible, although parallel computing methods are beginning to mitigate
50 this to an extent.

51 Prior to the development of MCMC methods, authors such as Press
52 (1970) developed conceptually straightforward Monte Carlo methods based
53 on the simple sampling of model space. However, these methods were quickly
54 sidelined in favor of the more targeted sampling strategies of MCMC-related
55 methods, on account of computational efficiency. In this study, we re-adopt
56 the conceptually straightforward approach, of seeking to sample the entire
57 model space, but instead of using the full forward simulator codes, we use
58 *emulators* to rapidly sample the model space and screen it for plausibility.
59 This differs significantly from the search-based MCMC or deterministic in-
60 version methods in that rather than trying to “build up” the plausible model
61 space or search for an optimum model, here the aim is to start with the

62 entire prior model space and to exclude implausible regions. The result is
63 therefore guaranteed to include *all* potentially plausible model space, given
64 all uncertainties specified as being relevant to the problem at hand.

65 *1.2. What is an emulator?*

66 An emulator is a fast statistical representation of a forward modelling
67 code. By training the emulator with a number of runs of the full simula-
68 tor code, the emulator seeks to predict the output of the full simulator to
69 a calibrated uncertainty, using a number of simple (in our case polynomial)
70 functions. Emulators are widely used e.g. climate modelling (Rougier et al.,
71 2009), ocean modelling (Logemann et al. , 2004), and cosmological appli-
72 cations (Vernon and Goldstein, 2009). In many cases, their use is crucial
73 to the ability to effectively model the system concerned since the systems
74 are so complex that to run a full simulation of a system such as the earth’s
75 atmosphere, ocean, or the universe, would be infeasible on the grounds of
76 insufficient computational capacity, even with recent technological advances.
77 These emulator-based methods have provided considerable insight into the
78 systems concerned through making the modelling process tractable. In this
79 study we use the emulator to screen model space in order to discern all regions
80 containing models representing earth structures which could have given rise
81 to the observed data, given the specified uncertainties in data measurements,
82 physical relationships and any others which the user may wish to specify.

83 In a number of ways an emulator is similar to a neural network, in that
84 through a process of learning the relationship between model parameters and
85 the data outputs, it seeks to give a rapid prediction of the output of a com-
86 plex code for a given set of input model parameters. Neural networks have

87 been used in order to solve inverse problems in geophysics; for example Meier
88 et al. (2007) use a neural network to invert shear wave data. However, an
89 emulator differs from a conventional neural network in that the emulator is
90 fully uncertainty-calibrated. Not only does the emulator give a rapid esti-
91 mate of the complex forward code output, but it will also give a calibrated
92 estimate of the uncertainty associated with that estimate. This uncertainty
93 calibration makes it possible to use the emulator to test and accept/reject
94 model space for plausibility. Simply having a prediction of the forward code
95 output, without an uncertainty estimate would not allow robust screening of
96 model space, simply because there would be no measure as to the reliability
97 of the output estimates being tested. However, with a calibrated emulator,
98 this problem is resolved. As with a conventional neural network, an emulator
99 will typically run several orders of magnitude faster than the full simulator
100 code, and so may be used to test and accept or reject large areas of model
101 space very quickly. Other authors have also developed methods to quickly
102 approximate the output of a full forward code in order to accelerate inverse
103 methods; James and Ritzwoller (1999) use truncated perturbation expan-
104 sions to approximate Rayleigh-Wave Eigenfrequencies and Eigenfunctions;
105 and Shapiro and Ritzwoller (2002), who take a similar methodology to use in
106 a MCMC scheme to construct a global shear-velocity model of the crust and
107 upper mantle. In each of these cases the aim is to minimise some objective
108 function or maximise a likelihood function.

109 Having built an emulator, we generate sets of model parameters using
110 a space-filling sampling design (latin hypercube), and test the emulator es-
111 timates of the forward code output for the candidate model parameter-sets

112 against an observed dataset, in order to reject implausible regions of model
113 space and thus constrain the region of plausible model space. After pop-
114 ulating the plausible space, a new emulator can be built over this smaller
115 region, with smaller predictive uncertainty. This new emulator can then be
116 used to further reject implausible model space. We repeat this cyclically
117 until no further emulator uncertainty reduction achieved, at which point all
118 structure has been discerned in the system. An advantage of this kind of
119 approach, which relies entirely on forward modelling, is that it is conceptu-
120 ally straightforward to include any kind of uncertainty or prior belief about
121 the model space. This may include data uncertainty, inter-parameter rela-
122 tionship uncertainty, model discrepancy (uncertainty due to the fact that no
123 model perfectly represents nature). The user simply generates models from
124 a prior space with the required properties.

125 We present a synthetic example using a 4-layer 1D model space over
126 seismic velocity, density, and resistivity (16 parameters in total, including
127 layer thickness). The resistivity, density and velocity parameters are linked
128 by an uncertain physical relationship (based on Gardner et al. (1974); Jegen-
129 Kulcsar et al. (2009)). Using a lab PC over a period of ~ 72 hours, after
130 an initial investment of 1000 model runs using the forward *simulator* code,
131 we use 11 emulation cycles to screen $\frac{1}{4}$ billion models over our large starting
132 model space, and reduce it by a factor of 10^{-19} . We then apply the method
133 successfully to an industrial 3D dataset at selected locations.

134 While the test case here is rather simple, and indeed it would be straight-
135 forward to solve using a deterministic inversion method, or indeed sampling
136 the model space using the full simulator, our aim here is to show the value

137 of a top-down and joint approach, where the entire possible model space is
138 considered, and to present emulation as a potentially useful tool both for fa-
139 cilitating this strategy and more widely in the field of geophysical constraint.
140 Based on our experience and that of Meier et al. (2007) and others, we posit
141 that emulation techniques have the potential to be used widely in the field
142 of geophysics and more generally throughout the Earth Sciences.

143 **2. Methods**

144 The problem being presented is that of joint constraint of a synthetic 1D
145 earth structure. We begin by briefly introducing the three different modelling
146 techniques/domains being used for the study; seismic refraction, magneto-
147 telluric (MT) and gravity and the function of the respective forward sim-
148 ulators, before considering the prior model space and coupling relationship
149 which we will sample in order to constrain the structure from the synthetic
150 dataset.

151 *2.1. Seismic refraction technique*

152 The seismic modelling technique takes advantage of the fact that differing
153 rock types possess differing velocities at which sound waves travel through
154 them. By measuring the time which sound waves take to travel along different
155 paths in the earth, one can gain information about this seismic velocity field,
156 and thus make interpretative judgements about the geological structure in
157 that region.

158 A full treatment of the seismic method is given in Kennett (2001). In
159 this study, we are considering energy which turns in the earth in a manner
160 described by Snell's law (Equation 1 and Figure 1), due to the presence of a

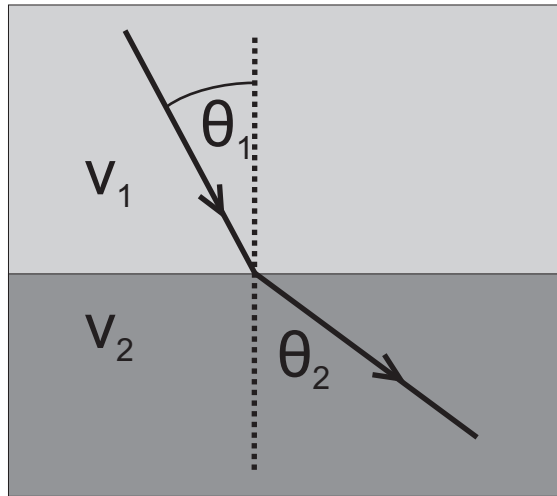


Figure 1: When sound waves encounter an impedance discontinuity, their path is altered. Where the density is unchanged, the deflection is controlled by the velocity, as shown, according to Snell's law (Equation 1). In the limiting case of a velocity gradient (infinitesimally thin layers of increasing velocity), the energy turns in the earth as shown in Figure 2.

161 velocity gradient, as in Figure 2. For our purposes, we are only considering
 162 energy which is NOT reflecting off some boundary/step change in impedance,
 163 ie. we are only considering energy which is travelling along paths akin to that
 164 of the blue ray in Figure 2. Note that in the scenario being described here,
 165 where we are using constant velocity model parameters, in order to create
 166 turning waves with the forward simulator, we add a small velocity gradient
 167 of 0.5%.

$$v_1 \sin(\theta_1) = v_2 \sin(\theta_2) \quad (1)$$

168 The forward simulator, which takes the input model and generates a set
 169 of output travel time data, is written by Björn Heincke (Heincke et al., 2006)

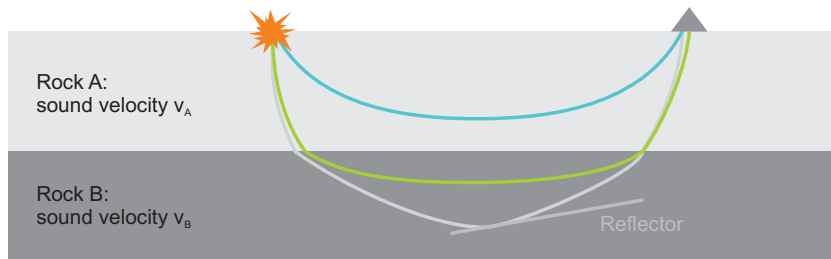


Figure 2: Rays showing the propagation path of sound waves through the earth. Sound waves turn within a velocity gradient (eg. blue and green rays), and they can also reflect from impedance contrasts (grey ray). Here we are considering only seismic energy turning as shown by the blue and green rays, and are not considering reflected energy. Note that in general sound waves have a frequency spectrum with finite width. Describing the process in terms of rays is only strictly valid when considering an infinite frequency spectrum.

170 and is based on the Finite Element (FE) method described by Podvin and
 171 Lecomte (1991). One of the major advantages of using a FE method such as
 172 this is that, unlike with a ray-tracing approach, travel times are calculated
 173 for ALL receiver locations, rather than simply where the infinite frequency
 174 ray-tracing approach finds a ray-path. For our purposes, this means that
 175 problems associated with a potentially differing number of data points with
 176 each run of the simulator are avoided. For the 1D purpose at hand, this
 177 simulator may be considered rather more complex than necessary, and if the
 178 aim of this paper were to simply find an efficient solution to the 1D problem, a
 179 less complex and computationally intensive code would easily suffice. Indeed,
 180 as is commented elsewhere in this paper, an emulator is not strictly necessary
 181 to solve this class of problem. However, although we are using a simple earth
 182 model, our intention is to show as far as possible an emulator being built
 183 for a complex 3D simulator which may be used on a large dataset. The

184 work was also carried out contemporaneously with that of Moorkamp et al.
185 (2011) with a view to exploring complementarity between the deterministic
186 and stochastic methods. This complementarity is not the main focus of this
187 paper, however in the context of the collaborative nature of the work, we
188 decided to use a common set of simulator codes.

189 The FE engine requires a gridded volume, and we chose to set this up with
190 dimensions appropriate for a reasonably high resolution large 2D refraction
191 seismic survey. The simulator was thus configured to use a 50,000 x 1,000 x 30,000 m
192 volume with a 100 m grid spacing. Travel times were generated/‘recorded’
193 every 500 m along the main axis of the model in a line colinear with the shot
194 position (0 m, 500 m, 10 m). Note that the receiver and shot positions are
195 slightly below the surface of the model, since placing these positions along
196 model boundaries/nodes can cause simulator instability. Because in our ex-
197 ample we use 1D layer-parameterized models, the model-input function of the
198 simulator was modified so that it could read in a series of 4 velocity values
199 and 4 layer thicknesses and use these values to generate the required for the
200 simulator. With this parameterization 1,000 runs of the simulator typically
201 took about 4.5 minutes. The output data for each run of the simulator are
202 thus a set of 100 (offset, time) points.

203 The output data from the simulator for a given model input are a series of
204 travel times recorded at a number of offsets (in our case 100) from the receiver
205 position. Note that this may seem a high density of travelttime points given
206 the problem under consideration. However, in the context of demonstrating
207 the use of the data reduction technique described in Section 2.7.1, and in
208 order to present a data density closer to that seen in 3D datasets, we have

209 chosen to use a high traveltime density.

210 2.2. MT technique

211 The Magneto-Telluric (MT) method aims to probe the resistivity struc-
212 ture of the Earth by measuring the terrestrial electric and magnetic fields
213 at the Earth's surface. Cagniard (1953) give a comprehensive description
214 of the method. The key output parameter from an MT experiment is a
215 measurement of the Earth's complex impedance $Z = \frac{E_x}{H_y}$ over a range of
216 electromagnetic frequencies. The raw output is normally in the form of
217 $R(\omega) = \text{Re}(Z(\omega))$ and $I(\omega) = \text{Im}(Z(\omega))$, however this is normally plot-
218 ted in terms of the transformed functions *apparent resistivity* r and *phase* ϕ ,
219 as defined in Equation 3.

$$\log_{10}(r_{app}(\omega)) = \log_{10}\{(R(\omega)^2 + I(\omega)^2)/(2\pi\mu_0 10^\omega)\} \quad (2)$$

$$\phi(\omega) = \frac{180}{\pi} \arctan(I(\omega)/R(\omega)) \quad (3)$$

220 Because the form of the $R(\omega)$ and $I(\omega)$ plots are much simpler in form
221 than the apparent resistivity/phase plots, we choose to consider R and I , as
222 shown in Figure 4.

223 Our simulator, written by Avdeev et al. (2002), takes as inputs a list of
224 resistivity r and layer thickness s values, along with a list of frequencies,
225 and outputs a list of R and I values, evaluated at each of the values of
226 ω . In our case, we have four layers (so, as with the seismic case, 8 model
227 parameters), and we choose to evaluate the output functions $R(\omega)$ and $I(\omega)$
228 at 20 frequencies ω_{1-20} . ω is in some sense a proxy for depth in the Earth,
229 in that the MT signal at higher frequency gives information about the upper

230 structure and the signal at low frequency gives information about the deeper
 231 structure. We therefore choose a fairly wide frequency range, from $\log_{10}(\omega) =$
 232 -6 to $\log_{10}(\omega) = 0.5$. Because of this large frequency range, we emulate
 233 $R(\log(\omega))$ and $I(\log(\omega))$ as a function of the model parameters r_{1-4} and
 234 s_{1-4} .

235 2.3. Gravity technique

236 This technique uses the fact that the fine-scale gravitational field is sen-
 237 sitive to density variations in the Earth. The gravitational field at a given
 238 point, a distance r from a point source of mass m is given by Equation 4.

$$g = -\nabla\phi = -\frac{Gm}{r^2} \quad (4)$$

239 As with other inverse square law phenomena, we can apply Gauss' the-
 240 orem (Equation 5) to obtain, in the 1D case, for our four-layer model, with
 241 $\vec{\theta} = [\rho_{1-4}, s_{1-4}]^T$, Equation 6.

$$\iint \vec{g} \cdot d\vec{A} = -4\pi G \iiint \rho dV \quad (5)$$

$$g = \sum_{i=1}^4 \left(\frac{\rho_i s_i}{\left(\sum_{j=1}^i s_j\right)^2} \right) \quad (6)$$

242 Our gravity simulator is extracted from the Full Tensor Gravity (FTG)
 243 inversion code by Moorkamp et al. (2011). In this 1D scalar setting, the
 244 simulator runs quickly. The output dataset, for a given $\vec{\theta}$ in the 1D case,
 245 thus consists of a single point measurement.

Layer 1: v_1 : 2800-4400 ms^{-1} (2837) s_1 : 500-4000 m (1200)	r_1 : 0.75-150 Ωm (1.00) ρ_1 : 2225-2525 kgm^{-3} (2259)
Layer 2: v_2 : 2800-4400 ms^{-1} (3760) s_2 : 500-4000 m (2200)	r_2 : 0.75-150 Ωm (20.00) ρ_2 : 2225-2525 kgm^{-3} (2424)
Layer 3: v_3 : 2800-4400 ms^{-1} (3057) s_3 : 500-4000 m (2500)	r_3 : 0.75-150 Ωm (2.00) ρ_3 : 2225-2525 kgm^{-3} (2301)
Layer 4: v_4 : 2800-4400 ms^{-1} (4186) s_4 : 500-4000 m (2000)	r_4 : 0.75-150 Ωm (100.00) ρ_4 : 2225-2525 kgm^{-3} (2490)

Figure 3: Schematic diagram of the joint model, with parameter ranges and the values used to produce the target dataset given in parentheses.

246 2.4. Model Space

247 Before building an emulator to screen model space, the bounds of the
 248 model space over which the emulator is to be used, must first be specified.
 249 In defining the prior model space, it should be borne in mind that in order to
 250 build a reliable emulator, the emulator should be trained over a slightly larger
 251 space than is required for screening purposes. This is in order to ensure that
 252 the edges of the prior plausible parameter space are sufficiently sampled and
 253 that information from these regions is included in the emulator construction.
 254 The prior plausible parameter space used here is shown in Figure 3.

255 2.5. The datasets

256 The synthetic scenario we are considering is that we have seismic re-
 257 fraction traveltimes data, complex impedance data from a MT survey, and
 258 a gravity measurement. The seismic and MT datasets, generated using the
 259 model parameter set shown in Figure 3, are shown in Figure 4.

260 In the case of the seismic data, the (x,t) curve is seen to consist of four
261 “segments”, reflecting the fact that the synthetic model is parameterized by
262 four constant velocity layers. The discontinuities in gradient contain struc-
263 tural information on account of the fact that deeper-propagating seismic
264 energy emerges at larger offsets. In a layered system, therefore, the posi-
265 tions of these gradient discontinuities can be exploited to constrain the layer
266 thicknesses. We exploit this in Section 2.7.2.

267 The MT dataset consists of two curves showing $\text{Re}(Z(\omega))$ and $\text{Im}(Z(\omega))$
268 varying with the driving frequency of the MT instrument. The electromag-
269 netic skin depth is inversely proportional to $\sqrt{\omega}$, and so lower frequency
270 signals propagate further into the Earth.

271 The gravity datum consists of a single synthetic measurement computed
272 to be 78.73896 mgal. Note that the absolute value of this synthetic measure-
273 ment should not be interpreted physically because the datum has not been
274 specified. It has simply been computed using the densities and layer thick-
275 nesses shown in Figure 3. In a real earth scenario, the measurement includes
276 considerable contribution from the deeper earth, and so for the purposes of
277 earth inference it is the value relative to a datum which is significant. In our
278 case we will be generating candidate density models and computing gravity
279 measurements to test against this synthetic value without any contribution
280 from a half space below, so we are in effect using a zero reference datum. In
281 the real data example presented later, we consider the gravity measurement
282 over the salt relative to that over sediment (Section 3).

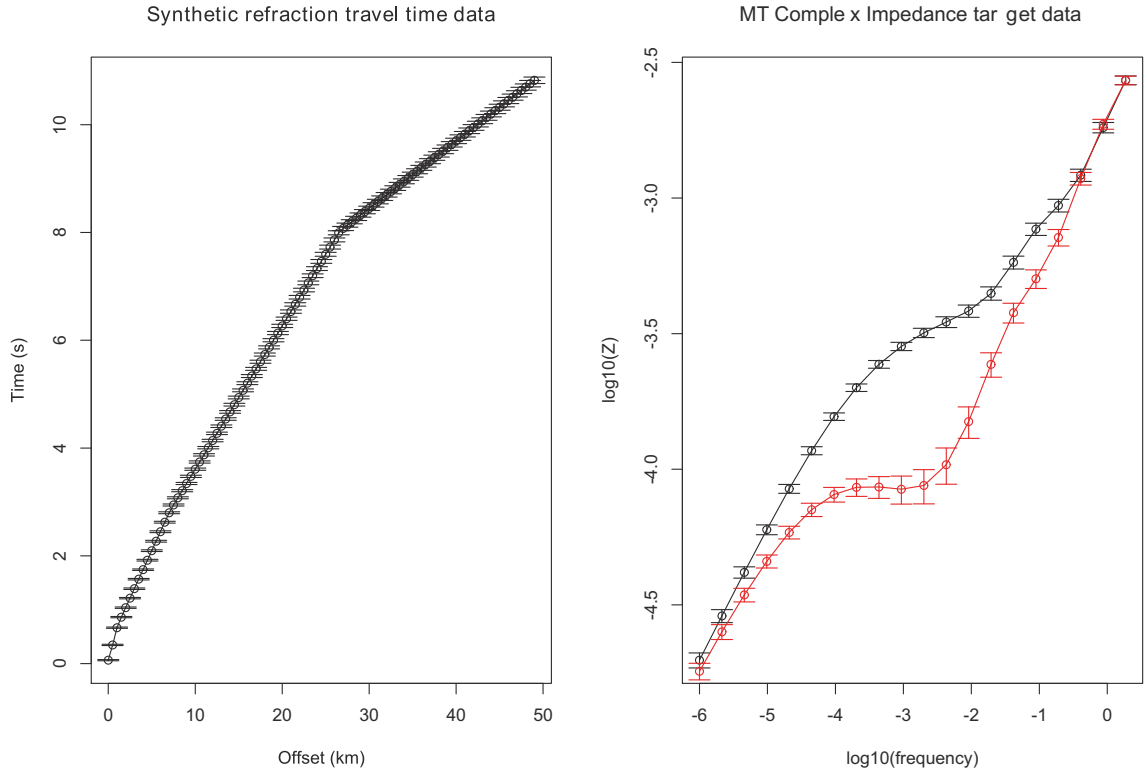


Figure 4: Synthetic datasets used for the study. Left: Refraction travel time data. Note that the travel time curve consists of four distinct straight segments. This is due to the model used to generate it consists of four constant velocity layers. Right: MT Complex Impedance data. The black line/points show $\text{Re}(Z(\omega))$ and the red line/points show $\text{Im}(Z(\omega))$. The data were generated by running the respective forward simulator codes with the parenthesized model parameters shown in Figure 3.

283 *2.6. Relationship uncertainty*

284 At the heart of a joint constraint method is the coupling between the
285 different kinds of model parameters. With the screening approach presented
286 here, any kind of coupling can easily be implemented. Here we build three
287 emulators; one for each of the seismic, gravity, and MT forward simulators.
288 We choose to build each emulator independently of each of the other mod-
289 elling domains so we do not need to invoke inter-parameter coupling in order
290 to build each emulator. Then, having built an emulator for each modelling
291 domain, we invoke the inter-parameter relationship with a specified uncer-
292 tainty to generate joint candidate models for screening, and the emulators
293 are used to discern which of these joint models are commonly plausible in
294 all three domains. It is not strictly necessary to specify such a physical rela-
295 tionship or coupling scheme. However, in such a situation, the “joint” nature
296 of the problem reduces to three independent modelling scenarios. Here we
297 seek to jointly constrain density, resistivity and seismic velocity parameters
298 for each of four layers (Figure 5).

299 In practice, the relationships linking these parameters are normally de-
300 rived empirically, by fitting a relationship to pre-existing borehole log data,
301 for example. This means that the relationship between the parameters is
302 uncertain. In linearised inversion schemes, using an uncertain physical rela-
303 tionship can create significant conceptual and technical challenges, however
304 because this method relies entirely on forward modelling, we can naturally
305 include this uncertain relationship by using it to generate our distribution of
306 candidate models at the screening stage.

307 The coupling relationship shown in Figure 5 applies in a sub-basalt set-

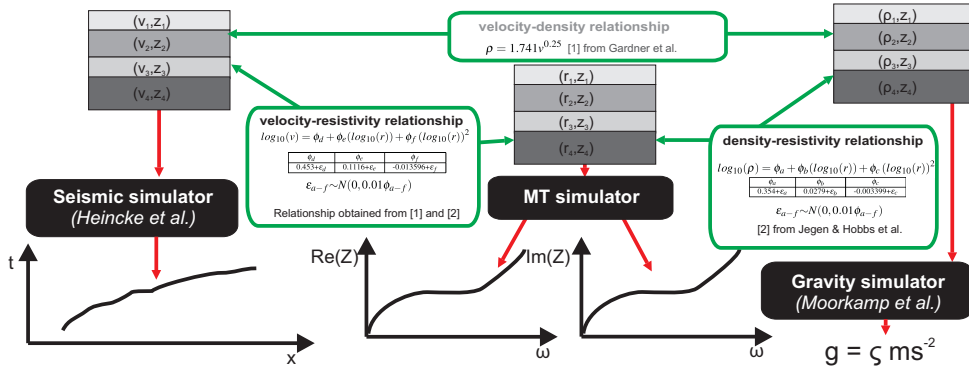


Figure 5: Joint modelling setting.

308 ting, such as that characteristic of the North East Atlantic in the vicinity of
 309 the Faroe Islands. We chose this scenario as a test case because of the particu-
 310 lar value a joint approach can add to constraint of this kind of structure. The
 311 challenges associated with sub-basalt seismic imaging are well documented
 312 (Roberts et al., 2009, for example). Recent technological and methodological
 313 advances (Lunnon et al., 2003; Ziolkowski et al., 2001) have given rise to some
 314 improvements to intra- and sub-basalt seismic images. However, the typi-
 315 cally highly heterogeneous nature of basalt, which gives rise to significant
 316 scattering of the seismic wavefield, means that the fundamental challenge
 317 remains. Jegen-Kulcsar et al. (2009), among others, have demonstrated the
 318 value of joint inversion methods, by virtue of the complementary information
 319 provided by different kinds of data.

320 2.7. Building the emulators

321 Jegen-Kulcsar et al. (2009) showed the value of using MT and seismic
 322 datasets together to constrain a sub-basalt problem. Because of the normally

323 non-unique nature of seismic models in a sub-basalt context, a Bayesian
324 approach such as this is particularly suited to be applied in this regime. We
325 design each of our emulators in three stages. We first employ a data-reduction
326 method. This is because, for example, in the seismic dataset, which consists
327 of (offset, travelttime) data, the data points, particularly at large offsets are
328 highly correlated. Having reduced the dataset, we then use a least-squares
329 fitting routine in order to fit the output data from a series of training runs of
330 the full simulator code to the model parameters in order to build a predictor.
331 Simply having a predictor, however, is not very useful in practice, however,
332 unless one has an estimate of the uncertainty of the predicted output. We
333 therefore, thirdly, calibrate the uncertainty of the predictor by calibrating it
334 against the output from the full simulator code for the training runs. The
335 result (the emulator) is a framework with which the output of the full forward
336 code can be predicted rapidly, with a calibrated uncertainty estimate. This
337 can then be used to screen candidate model parameter sets for plausibility.

338 *2.7.1. Building a seismic emulator*

339 To construct an emulator, a batch of training runs using the full forward
340 simulator code is required. The seismic simulator code used to train the
341 emulator was that used by Heincke et al. (2006). We generate 1,000 sets
342 of the 8 velocity model parameters (4 velocities and 4 layer thicknesses for
343 each model) using a space-filling latin hypercube design over the model space
344 shown in Figure 3. Each of these 1,000 models is then passed through the
345 seismic *simulator* code, which generates a set of 100 (offset, travelttime) points
346 for each set of model parameters. Because of the typically highly correlated
347 nature of the (x, t) points, we then chose to re-represent each of these travel

348 time data curves as a set of $p_{seis} = 8$ polynomial coefficients $\alpha_{1-p_{seis},seis}$. The
349 value of p_{seis} was determined by trial and error to obtain the optimum order
350 of polynomial with which to fit the (x, t) curves. These coefficients, $\alpha_{i,seis}$,
351 are generated by least-squares fitting each curve to the functional form shown
352 in Equation 7. $g_{seis}(x)$ is an uncertainty function, which we calibrate later.

$$\log(t^2) = \left(\sum_{i=0}^{p_{seis}} \alpha_{i,seis} (\log(x^2 + 1))^i \right) + g_{seis}(x) \quad (7)$$

$$\theta_{seis}^{\rightarrow} = \left[v_1 \ v_2 \ v_3 \ v_4 \ s_1 \ s_2 \ s_3 \ s_4 \right]^T \quad (8)$$

$$\alpha_{i,seis} = \left(\sum_{k=1}^{w_{seis}} \sum_{j=0}^{q_{seis}} \beta_{ijk} \theta_{k,seis}^j \right) + g_{i,seis}(\theta_{seis}^{\rightarrow}) \quad (9)$$

353 Having re-represented the training datasets as sets of $\alpha_{1-8,seis}$, we seek
354 to predict these coefficients for a given set of model parameters. We do this
355 by fitting the α_i to functions of the model parameters $\theta_{seis,1-8} = (v_{1-4}, s_{1-4})$,
356 using coefficients β_{ijk} , as shown in Equations 8-9. The functions $g_{i,seis}(\theta_{seis}^{\rightarrow})$
357 are uncertainty functions associated with the prediction of each $\alpha_{i,seis}$ coeffi-
358 cient, w_{seis} is the number of model parameters being considered (8), and q_{seis}
359 is the number of β_{ijk} coefficients used to fit each α_i coefficient (4). The value
360 of q_{seis} was determined by trial and error, by examining the optimum order
361 of polynomial for reconstructing the $\alpha_{i,seis}$. Here we have chosen q_{seis} to be
362 the same for each α_i , however this need not be the case in that it would be
363 straightforward to use a different number of β coefficients for each i .

364 In the coefficients β_{ijk} , we now have a means to predict the output of the
365 simulator code for a given set of input model parameters θ_{seis} . Examples of
366 the ability to predict or reconstruct the data using these β_{ijk} coefficients are
367 shown in Figure 6.

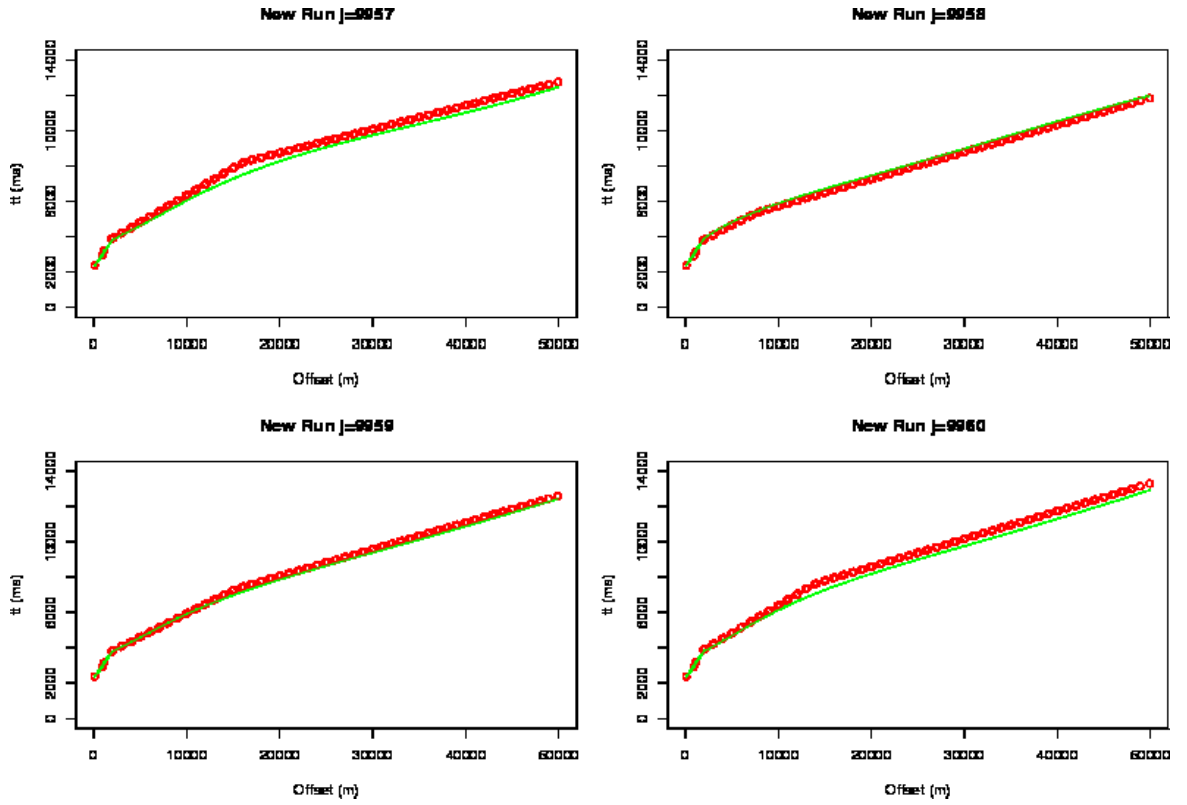


Figure 6: Sample emulator output traveltime data plotted with the real travel times. Red dots mark the real data points and the green lines represent the result of predicting the travel times using the emulator with the relevant model parameters. Note that there is a generally good reconstruction and much of the misfit which is present occurs where there is a discontinuity in the travel time vs offset gradient function (in this case due to the fact that the models consist of four constant velocity layers).

368 However, simply having a prediction of the output of the forward code is
369 of little value, unless one can also specify the uncertainty of the prediction.
370 Our next step is therefore to calibrate the uncertainty of the predictor out-
371 put. We do this by using our predictor to generate outputs for the training
372 models and comparing these to the outputs obtained from the full simula-
373 tor code through construction of a residual function, as shown in Figure 7.
374 Note that the residual function we are constructing here is not simply ei-
375 ther $g_{seis}(x)$ or $g_{i,seis}(\vec{\theta}_{seis})$, as in Equations 7 and 9. Rather, it is given by
376 $G_x(x)$ (Equation 14 where n is the number of emulator training runs), which
377 we use to approximate $G(x, \theta_{seis})$ (Equations 12-13). Note that although in
378 this example the traveltime plots themselves consist of four distinct segments
379 (Figure 6) due to the use of constant velocity four layer models, the residual
380 plots do not manifest these segments since over the set of 1000 models used
381 to construct the residuals, different layer thicknesses are used, drawn from
382 across the model space shown in Figure 3.

$$\log(t^2) = \sum_{i=0}^{p_{seis}} \left(\left(\sum_{k=1}^{w_{seis}} \sum_{j=0}^{q_{seis}} \beta_{ijk} \theta_{k,seis}^j \right) + g_{i,seis}(\vec{\theta}_{seis}) (\log(x^2 + 1))^i \right) + g_{seis}(x) \quad (10)$$

$$= \sum_{i=0}^{p_{seis}} \sum_{k=1}^{w_{seis}} \sum_{j=0}^{q_{seis}} \beta_{ijk} \theta_{k,seis}^j + \left[\sum_{i=0}^{p_{seis}} \left(g_{i,seis}(\vec{\theta}_{seis}) (\log(x^2 + 1))^i \right) + g_{seis}(x) \right] \quad (11)$$

$$= \sum_{i=0}^{p_{seis}} \sum_{k=1}^{w_{seis}} \sum_{j=0}^{q_{seis}} \beta_{ijk} \theta_{k,seis}^j + G(x, \vec{\theta}_{seis}) \quad (12)$$

$$\approx \sum_{i=0}^{p_{seis}} \sum_{k=1}^{w_{seis}} \sum_{j=0}^{q_{seis}} \beta_{ijk} \theta_{k,seis}^j + G_x(x) \quad (13)$$

$$G_x(x) = \sqrt{\frac{\sum_{n=1}^{n_{max}} (t_{em,n}(x) - t_{sim,n}(x))^2}{n_{max}}} \quad (14)$$

383 2.7.2. The seismic second derivative

384 It can be seen from Figure 6 that the largest discrepancy between the
 385 emulated output t vs x function occurs where there is a discontinuity in $\frac{dt}{dx}$.
 386 In our scenario the positions of these gradient discontinuities contain useful
 387 information, since the offset can be thought of as a proxy for the depth in
 388 the model. As a result, the offset positions of these gradient discontinuities
 389 contain information about the depths of the boundaries in our structure.

390 Rather than seek to emulate the gradient function to probe this informa-
 391 tion, we calculate the second derivative of the t vs x function (Figure 8), and
 392 given we are using 4-layer models, we aim to estimate the offset positions x
 393 of the three largest spikes in this $\psi = \left(\frac{d^2t}{dx^2}\right)^2$ function. Note that we use
 394 $\left(\frac{d^2t}{dx^2}\right)^2$ rather than $\frac{d^2t}{dx^2}$ in order that the ψ is positive definite, simplifying

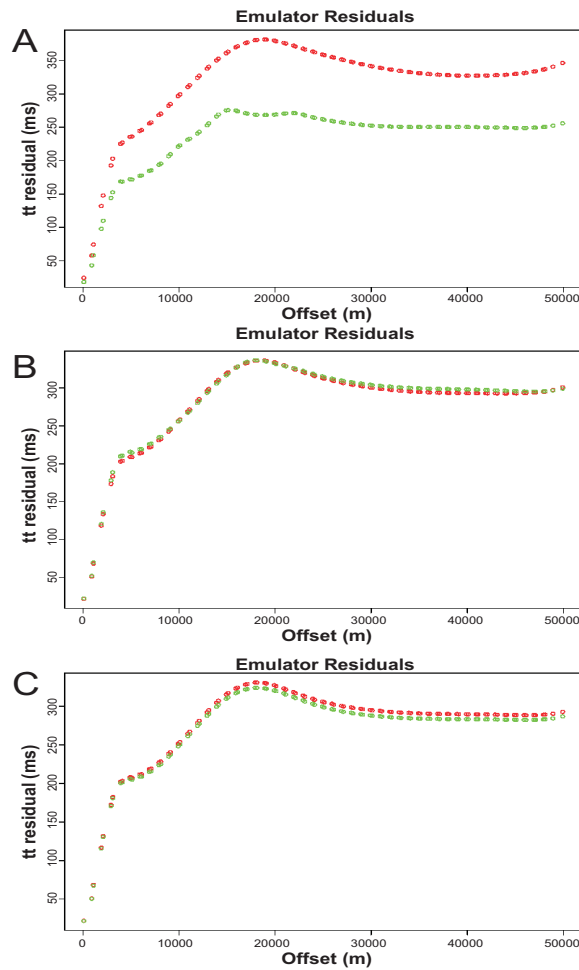


Figure 7: Sample seismic emulator output traveltide data residuals using (A) 100, (B) 1,000 and (C) 10,000 simulator runs to generate the emulator. Green dots are the residuals for the models used to create the emulator and red dots the residuals from running the emulator for 10,000 further models not used for training. Notice two effects: 1) The emulator traveltide residual for new models (red dots) decreases with increased training runs. 2) On using 100 runs, the emulator traveltide residual computed using the models used to construct the emulator is smaller than that for running the 10,000 further models, whereas on using 1,000 training runs, the emulator traveltide residual function for those models well represents that obtained on running further models, implying that $\sim 1,000$ training models are sufficient to cover the model space.

395 the process of picking the extrema.

396 We choose to do this because, given we are using a polynomial to rep-
397 resent the t vs x function, if we try to fit a polynomial to the derivative of
398 this function, $\frac{dt}{dx}$, the result of the Least Squares fit is likely to simply be
399 the derivative of the function given by our α -coefficient polynomial repre-
400 sentation, which we could calculate analytically, and so we would not gain
401 further useful information. Also, the parts of the gradient function contain-
402 ing the most useful information are the steepest-turning regions, which are
403 the most difficult parts to fit using smooth functions. Another advantage
404 of the ‘spike’-fitting approach over trying to predict the gradient function
405 itself is that the number of data points we are aiming to fit for an n -layered
406 model is $n-1$, so in our case we are trying to fit only three datapoints (the
407 x -positions of the three largest spikes in the $\psi = \left(\frac{d^2t}{dx^2}\right)^2$ function). Having
408 only three datapoints to fit thus makes the emulator construction process
409 considerably more efficient.

410 2.8. MT and Gravity emulators

411 A similar method was used to build emulators for the MT and gravity
412 modelling scenarios. The forward simulator codes used for training these
413 emulators were based on those by Avdeev et al. (2002) and Moorkamp
414 et al. (2011), respectively. In the case of MT, the forward simulator gen-
415 erates a complex impedance as a function of frequency $Z(\omega) = \text{Re}(Z(\omega)) +$
416 $i.\text{Im}(Z(\omega))$, in an analogous way to the seismic simulator, which generates
417 travel times as a function of offset $t(x)$. We therefore use a similar strategy as
418 for the seismic case in order to predict the MT simulator output for a given
419 set of model parameters. Firstly, we reduce the dataset by fitting a set of

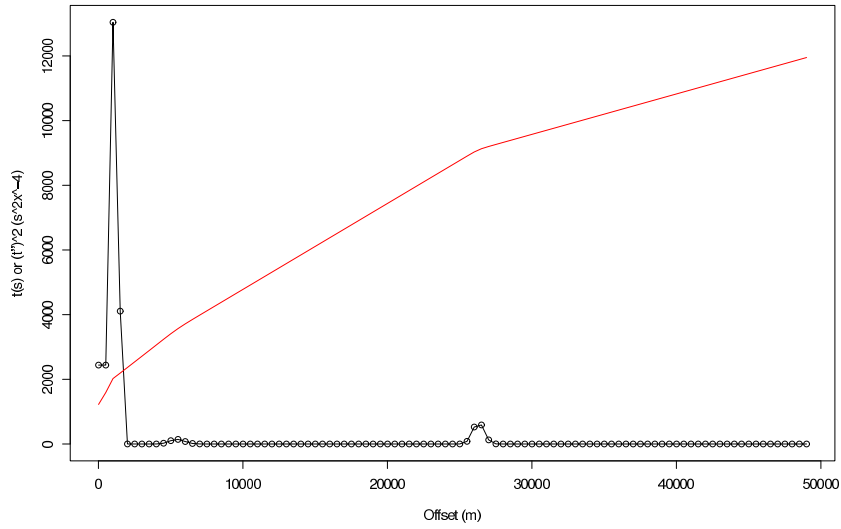


Figure 8: Synthetic t vs x plot (red), with (scaled) $\psi = \left(\frac{d^2t}{dx^2}\right)^2$ overlaid (black). The aim is, to within a known uncertainty, predict the positions of the three largest maxima of the (x, ψ) function. In this example, ψ_{max} are seen at $x \approx 1500$, $x \approx 5500$, $x \approx 27000$.

420 polynomial coefficients α_{MTR} and α_{MTI} to the simulator outputs $\text{Re}(Z(\omega))$
 421 and $\text{Im}(Z(\omega))$ respectively. We then fit these α_{MTR} and α_{MTI} to the model
 422 parameters $\vec{\theta}_{MT}$. The formulation is shown in Equations 15-28.

Emulating $R(\omega) = \text{Re}(Z(\omega))$:

$$\log(R(\omega)) = \left(\sum_{i=0}^{p_{MT}} \alpha_{iR} (\log(\omega))^i \right) + g_{p_{MT}R}(\omega) \quad (15)$$

$$\theta_{MT} = \left(r_1 \ r_2 \ r_3 \ r_4 \ s_1 \ s_2 \ s_3 \ s_4 \right)^T \quad (16)$$

$$\alpha_{iR} = \left(\sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkR} \theta_{ijk,MT}^j \right) + g_{iR}(\theta_{MT}^{\rightarrow}) \quad (17)$$

$$\log(R(\omega)) = \left(\sum_{i=0}^{p_{MT}} \left(\left(\sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkR} \theta_{ijk,MT}^j \right) + g_{iR}(\theta_{MT}^{\rightarrow}) \right) (\log(\omega))^i \right) + g_{p_{MT}R}(\omega) \quad (18)$$

$$= \sum_{i=0}^{p_{MT}} \sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkR} \theta_{ijk,MT}^j (\log(\omega))^i + G(\omega, \theta_{MT}^{\rightarrow}) \quad (19)$$

$$\approx \sum_{i=0}^{p_{MT}} \sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkR} \theta_{ijk,MT}^j (\log(\omega))^i + G_{\omega,R}(\omega) \quad (20)$$

$$G_{\omega,R}(\omega) = \sqrt{\frac{\sum_{n=1}^{n_{max}} (R_{em,n}(\omega) - R_{sim,n}(\omega))^2}{n_{max}}} \quad (21)$$

and $I(\omega) = \text{Im}(Z(\omega))$:

$$\log(I(\omega)) = \left(\sum_{i=0}^p \alpha_{iI} (\log(\omega))^i \right) + g_{pI}(\omega) \quad (22)$$

$$\theta_{MT} = \left(r_1 \ r_2 \ r_3 \ r_4 \ s_1 \ s_2 \ s_3 \ s_4 \right)^T \quad (23)$$

$$\alpha_{iI} = \left(\sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkI} \theta_{ijk,MT}^j \right) + g_{iI}(\theta_{MT}) \quad (24)$$

$$\log(I(\omega)) = \left(\sum_{i=0}^{p_{MT}} \left(\left(\sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkI} \theta_{ijk,MT}^j \right) + g_{iI}(\theta_{MT}) \right) (\log(\omega))^i \right) + g_{p_{MT}I}(\omega) \quad (25)$$

$$= \sum_{i=0}^{p_{MT}} \sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkI} \theta_{ijk,MT}^j (\log(\omega))^i + G(\omega, \vec{\theta}_{MT}) \quad (26)$$

$$\approx \sum_{i=0}^{p_{MT}} \sum_{k=1}^{w_{MT}} \sum_{j=0}^{q_{MT}} \beta_{ijkI} \theta_{ijk,MT}^j (\log(\omega))^i + G_{\omega,I}(\omega) \quad (27)$$

$$G_{\omega,I}(\omega) = \sqrt{\frac{\sum_{n=1}^{n_{max}} (I_{em,n}(\omega) - I_{sim,n}(\omega))^2}{n_{max}}} \quad (28)$$

423 After testing we chose $p_{MT} = 9$ and $q_{MT} = 3$.

424 Examples of the uncertainty function calibrated over 1000 models after a
425 single emulation cycle are shown in Figure 9.

426 In the case of gravity, because we are in a 1D setting, there is only a
427 single gravity point. There is therefore no need to perform the first stage,
428 of reducing the dataset, since there is only one point. In the case of gravity,
429 we therefore simply fit the simulator gravity outputs ϕ directly to the model
430 parameters $\vec{\theta}_{grav}$ and calibrate the predictive uncertainty, as in Equations
431 30-31.

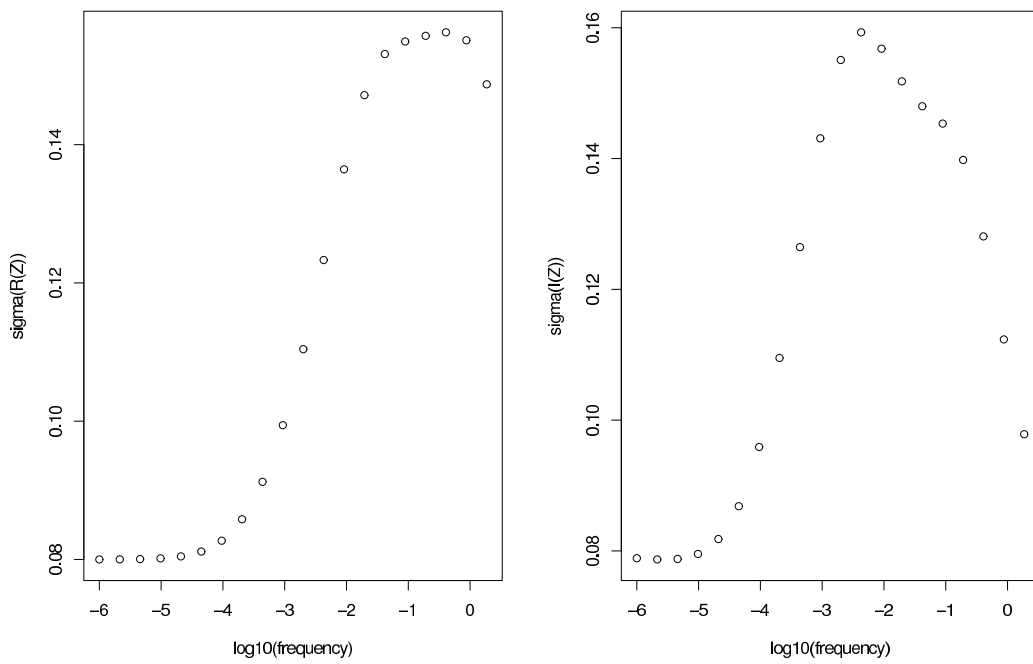


Figure 9: Standard deviation of MT emulator residual at each frequency point for $\text{Re}(Z)$ (left) and $\text{Im}(Z)$ (right).

$$\phi = \left(\sum_{k=1}^{w_{grav}} \sum_{j=0}^{q_{grav}} \beta_{jk,grav} \theta_{k,grav}^j \right) + g_{grav}(\theta_{grav}^{\rightarrow}) \quad (29)$$

$$g_{grav}(\theta_{grav}^{\rightarrow}) \approx \sqrt{\frac{\sum_{i=1}^n (g_{i,em} - g_{i,sim})^2}{n}} \quad (30)$$

$$(31)$$

432 2.9. Using the emulators

433 Having built emulators for each of the seismic, gravity and MT simulators,
434 we then use the four of them in order to screen model space for implausibility.

435 Figure 10 shows the results of using a preliminary seismic emulator to
436 screen a set of 10,000 models drawn from the model space over which the
437 emulator was trained (Figure 3). The screening is carried out by generating a
438 candidate model, computing the emulator output predicted dataset, and then
439 comparing the predicted dataset to the target dataset. In the seismic case,
440 for example, if the target dataset lies within $\gamma_{seis}G_x$ of the emulator predicted
441 output, where γ_{seis} is a scaling factor designed to ensure a high probability
442 of the true model parameter set being selected in a synthetic test, then the
443 model parameter set is deemed to be plausible. Here we choose γ_{seis} to be 3,
444 and so the plausibility condition is as shown in Equation 32.

$$\frac{\sum_n \min [|(t_{em}(x_n) - t_{targ}(x_n))| - \gamma_{seis}G_x(x_n), 0]}{n} < 1 \quad (32)$$

445 Statistics pertaining to the construction of the emulator are shown in
446 Table 1. Note that using the emulator we can rapidly reduce the plausible
447 model space to 213/10,000 (2.13%) of the original “prior” model space, and
448 that we have done this in a time of around 5 minutes (most of which is used in

449 building the emulator itself rather than actually screening the models) rather
450 than about 45 minutes, which is how long it would have taken to run the
451 same number of models through the full simulator on the same computer.
452 In order to verify that the system was behaving in a sensible manner, i.e.
453 that it was selecting plausible models for which the full simulator output (as
454 opposed to simply the emulator output), the full simulator data outputs were
455 then computed and plotted for each of the 10,000 models. Figure 10 shows
456 the full simulator outputs for the models which the emulator deemed to be
457 plausible (green) and those it deemed implausible (blue), as well as the target
458 simulated dataset (red). Note that all the simulated data outputs for models
459 deemed plausible by the emulator lie close to the target dataset whereas the
460 simulated outputs for the models deemed implausible by the emulator lie far
461 from the target dataset. This suggests that the emulator is indeed useful for
462 screening the model space, as intended.

463 The seismic, spike, MT and gravity emulators can together be used to
464 screen model space. Candidate models are generated from the prior model
465 space shown in Figure 3, where the density, resistivity and seismic velocity
466 parameters are linked by the uncertain physical relationship as shown in
467 Figure 5. We then use each of the four emulators to screen these models to
468 discern which models are jointly plausible given our assertions regarding the
469 data uncertainty, relationship uncertainty, model discrepancy and potentially
470 other uncertainties. In this example, because we are using three methods to
471 screen models, we use a starting model pool of 100,000 models. Figure 11
472 shows the contribution each method is making to constraining the plausible
473 space. If the user wishes to bias the selection weighting towards one of

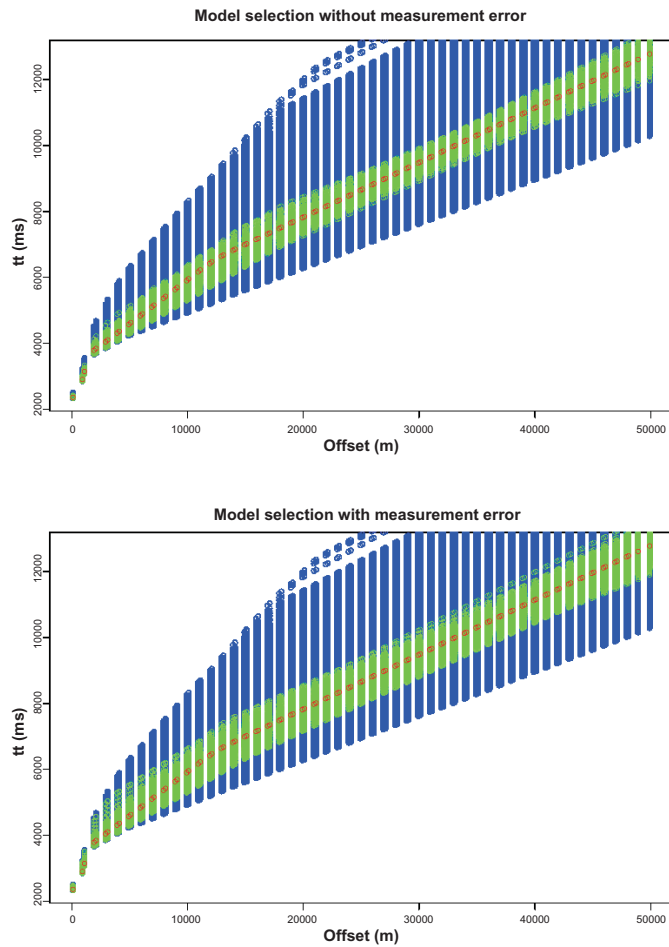


Figure 10: Rejection of implausible model space using the seismic emulator. The upper plot shows the result of assuming zero measurement error in the screening function, and the lower plot assumes a measurement error of 50 ms. The red dots show the result of running the model simulator for a chosen set of model parameters, the “target” travel time dataset. The emulator output is then compared for each of 10000 models which were not used to construct the emulator. For comparison, these 10000 models were then run through the full simulator and the outputs for those which were selected by the emulator are plotted using green dots and those which were rejected from running the emulator are plotted in blue. In the case where no measurement error is assumed, 213/10000 models were selected, and in the case where measurement error of 50ms is assumed, 498/10000 models were selected.

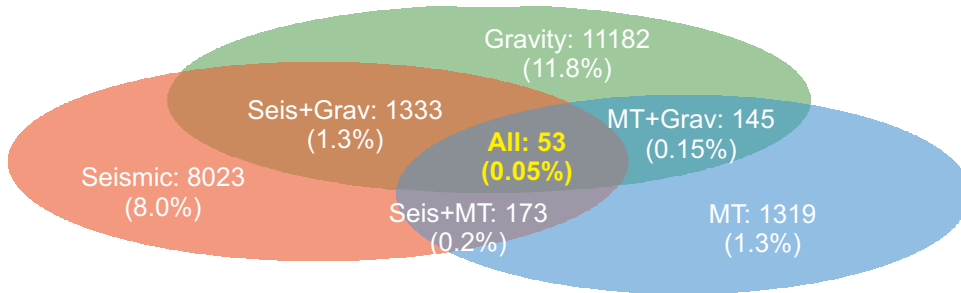


Figure 11: Model selection statistics using the seismic, gravity and MT emulators to jointly select plausible model space. Using these three emulators, it is seen that only 53/100000 (0.05%) of the model parameter space is plausible after a single cycle.

474 the methods, then the scaling factor γ_{seis} , γ_{MT} and γ_{grav} can be changed
 475 accordingly at the user’s discretion. However, to ensure reliability of the
 476 system, each γ_x should be calibrated such that there is a high probability
 477 that, should for a synthetic example, the “true” set of model parameters be
 478 presented to the system, there is a very high probability that the model will
 479 be accepted. In our scenario here, we chose $\gamma_{seis} = \gamma_{MT} = \gamma_{grav} = 3$, which
 480 meant that 97% of the time, for a large number of target models, the “true”
 481 model parameter set (that used to generate the synthetic data) was deemed
 482 plausible.

483 2.9.1. Multi-cycle screening

484 Having used the emulators to test for plausibility and sampled the prior
 485 model space, we can then use this plausible region to construct new emula-
 486 tors, which, because they are built over a smaller model space, will generally
 487 have better predictive accuracy, and thus can be used to further constrain the
 488 plausible model space for a given target dataset. At each cycle, we therefore
 489 expect the uncertainty functions associated with each emulator to reduce in

490 magnitude. We ran the scheme over 11 cycles (after which there was no
491 discernible reduction in the emulator uncertainty functions), at each cycle
492 seeking to find 1000 plausible models with which to both build a new emu-
493 lator, and reduce the size of the model space being screened. The number of
494 candidate models generated at each cycle in order to find 1000 plausible ones
495 is shown in Table 1. Because we are able to exclude a model as implausible
496 on the basis that it is deemed implausible by any one of the seismic, spike,
497 MT or gravity cases, we do not need to generate the emulator output for
498 each method for every candidate model. However, for a candidate model to
499 be deemed plausible and so be used in the subsequent cycle, it must “pass”
500 the plausibility test in each case.

501 Note also that after the first cycle, the plausible model space to be
502 searched isn’t defined simply by the marginal parameter bounds, but also
503 by the condition that a model was “passed” by the emulator screening from
504 the previous cycle. Therefore in each cycle, for a candidate model to be
505 deemed plausible, it must not simply be screened by the emulators gener-
506 ated using the plausible model space from the immediately previous cycle,
507 but by those generated by all previous cycles. So, for a model to be deemed
508 plausible on the 11th cycle, it must pass 44 screening tests (using each of the
509 seismic, spike, MT, and gravity emulators from each of the previous cycles).
510 However, if a model fails at any one of these stages, it is deemed implausible.

511 Because for a model to deemed implausible, only one “failure” is required,
512 the computational efficiency of the screening process can be maximised by,
513 for example, ordering the screening process such that the apparently most
514 stringent screening method (seismic, spike, MT or gravity) is run first. This

515 can be tested using each screening method separately. In Figure 11, the
516 MT selection is clearly the most stringent. However, in considering screening
517 efficiency, it is also the case that each emulator (seismic, spike, MT or gravity)
518 can take differing times to run (this is a function of the number of coefficients
519 requiring computation in each case), as shown in Table 1. The efficiency of
520 candidate model rejection is thus a function both of the emulator run time
521 and the stringency of that emulator. In the example presented here, for each
522 candidate model we screened using the gravity and spike emulators first,
523 since these were the faster to run emulators (due to computing single data
524 points rather than coefficients for data functions in the case of the seismic
525 and MT cases).

526 On implementing this screening strategy, the marginal parameter his-
527 tograms for the models deemed plausible are plotted in Figure 12. Taking a
528 measure of the model space simply defined by a 16-dimensional box around
529 the plausible parameter sets, the plausible model space volume has been
530 reduced by a factor of 10^{-19} (or about 0.06 on average per model parame-
531 ter). These histograms contain some information about the distribution of
532 plausible models and in some instances be used to update the user's beliefs
533 about the plausible parameter bounds. However, the marginal distributions
534 often do not convey a large amount of the total information present in the
535 distribution of plausible models. In particular, in the full joint distribution,
536 there are likely to be inter-parameter relationships which emerge, for exam-
537 ple, if the $v_2 \approx v_3$ then due to pseudo-non-uniqueness regarding the model
538 specification, there may be a strong trade-off between s_2 and s_3 . In this case
539 the histograms would show wide distributions for s_2 and s_3 but this does

540 not imply that the system is at fault in failing to constrain these layers, but
541 given the data uncertainty, and the fact that $v_2 \approx v_3$, the data do not con-
542 strain the thickness of these layers particularly tightly, whereas $s_2 + s_3$ may
543 be constrained very effectively. Such dependencies are, however, indicative
544 that it may be appropriate to reduce the number of model parameters since
545 there is functional dependence between two or more of them.

546 The results presented here were obtained by generating candidate models
547 using a marginal parameter sampling scheme; implementing a Sobol algo-
548 rithm (Bratley and Fox, 1988) over a uniform distribution to select combina-
549 tions of model parameters lying within the bounds. However, the screening
550 process efficiency can be greatly increased by sampling from the joint dis-
551 tribution from the previous cycle, in that much less time is spent sampling
552 redundant model space. We have tested some strategies for doing this with
553 some success, though there are number of questions regarding the use of
554 non-uniform prior parameter distributions at each cycle which need careful
555 consideration, since the choice of prior distribution at each cycle influences
556 the output parameter distributions: $P(X|A) = P(X).P(A|X)/P(A)$. Along
557 with exploring appropriate parameterizations for a 3D structural scenarios,
558 this is a subject of ongoing development.

559 **3. Application to a real dataset**

560 A dataset was kindly provided by Statoil, consisting of seismic, MT, and
561 gravity data over a salt body. The free air gravity data are shown in Figure
562 14. We start by constructing emulators over the joint model space shown
563 in Table 2. This model space was chosen after preliminary examination of

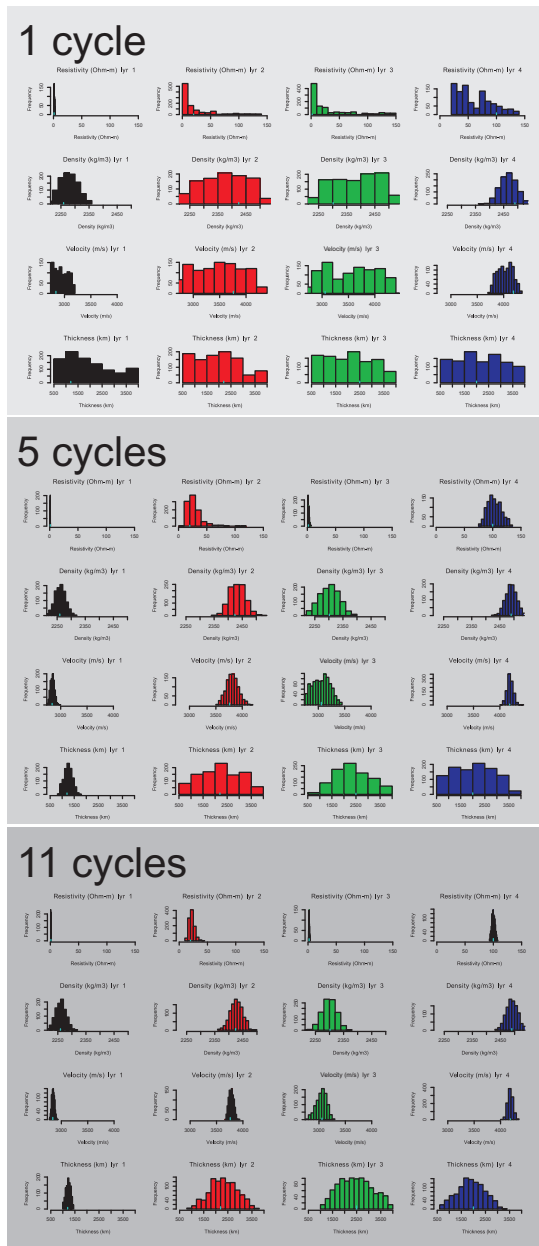


Figure 12: Histograms showing the marginal parameter distributions after 1 cycle (top), 5 cycles (middle) and 11 cycles (bottom). Notice that as further emulation cycles are carried out the plausible model space is reduced. Note that the widths of the distribution axes are the initial ranges defined by our prior model space (Figure 3).

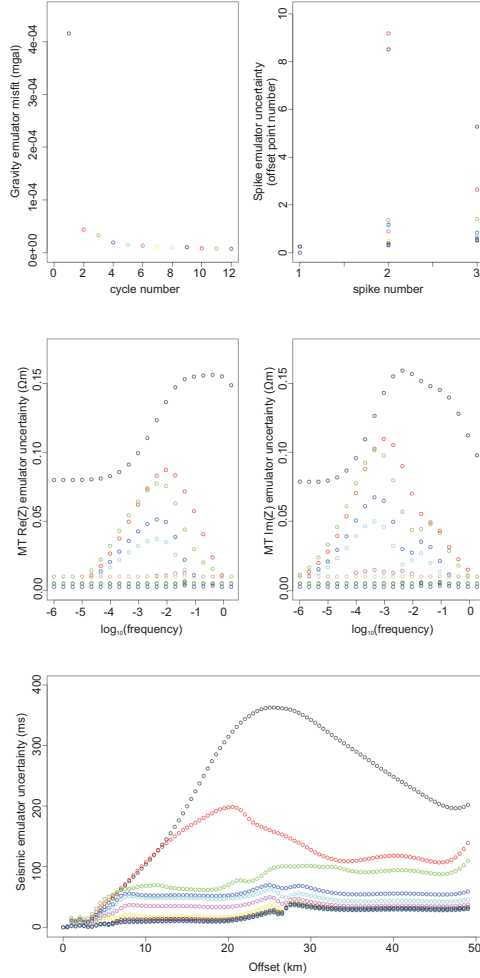


Figure 13: Misfit functions for each emulator. Bottom: seismic, Middle left: MT (real part), Middle right: MT (imaginary part), Top left: Gravity, Top right: spike. Misfit functions for cycles 1-11 are marked by the colours black, red, green, blue, cyan, magenta, yellow, grey, black, red, green and blue, respectively. Note that the uncertainty for successive emulation cycles decreases, reflecting the fact that each successive emulator for each technique is being built over a smaller model space and is thus able to capture more subtle data output variation than the emulator built in the previous cycle.

564 the data and some preliminary modelling to see which general class of mod-
565 els may be suitable. In order to test the methodology and avoid difficulties
566 associated with designing a suitable conditional sampling strategy, at each
567 emulation cycle we use a simple marginal sampling strategy, again sampling
568 using a Sobol algorithm, from uniform distributions over the range of model
569 parameters from the models deemed plausible from the previous cycle. Fig-
570 ures 15 and 16 show results from using only the *seismic* emulator to screen
571 model parameter sets from a region over salt (MT7) and a region where there
572 is no salt (MT12). The locations of MT7 and MT12 are shown in Figure 14.
573 It can be seen from Figures 15 and 16 that the system is clearly detecting
574 the salt body by virtue of the fact that layer 3 clearly has a salt-like velocity
575 at MT7 whereas at MT12 this is not the case.

576 We then included the MT and gravity emulators in the screening pro-
577 cess. To generate joint candidate models, we utilise other data provided by
578 Statoil to elicit a relationship between resistivity and seismic velocity for the
579 local region for both the salt and sedimentary regimes. The relationship for
580 the sedimentary case is shown in Equation 34. As in the synthetic exam-
581 ple, we use Gardner’s relation for the relationship between seismic velocity
582 and density. Using the information provided by Statoil, and because the
583 relationship is empirical and uncertain, we also specify an uncertainty, this
584 time normally distributed, on both the resistivity and density as a function
585 of velocity (Equation 34).

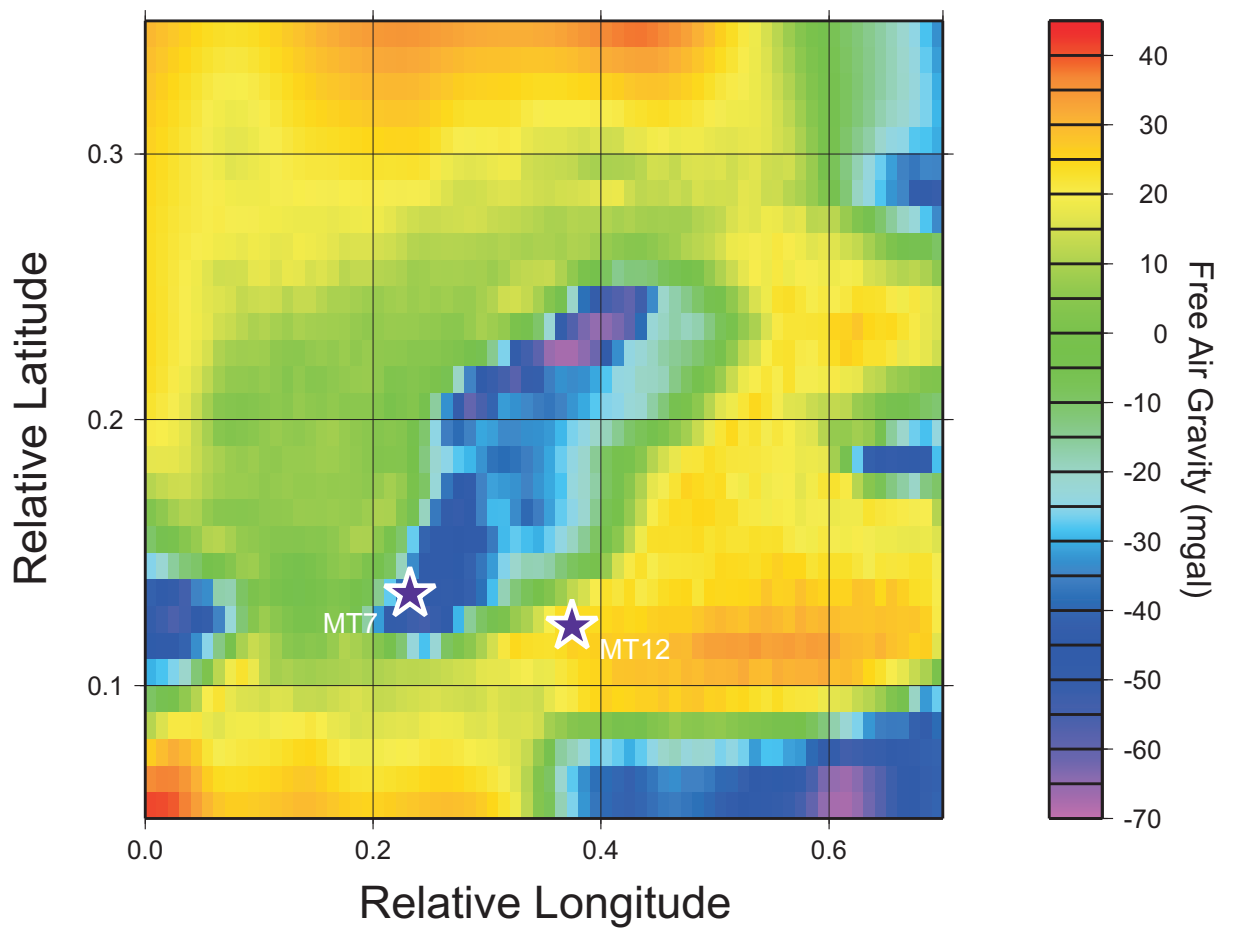


Figure 14: Gravity map in the region of MT7 and MT12. Note the presence of the salt body indicated by the low gravity anomaly under MT7.

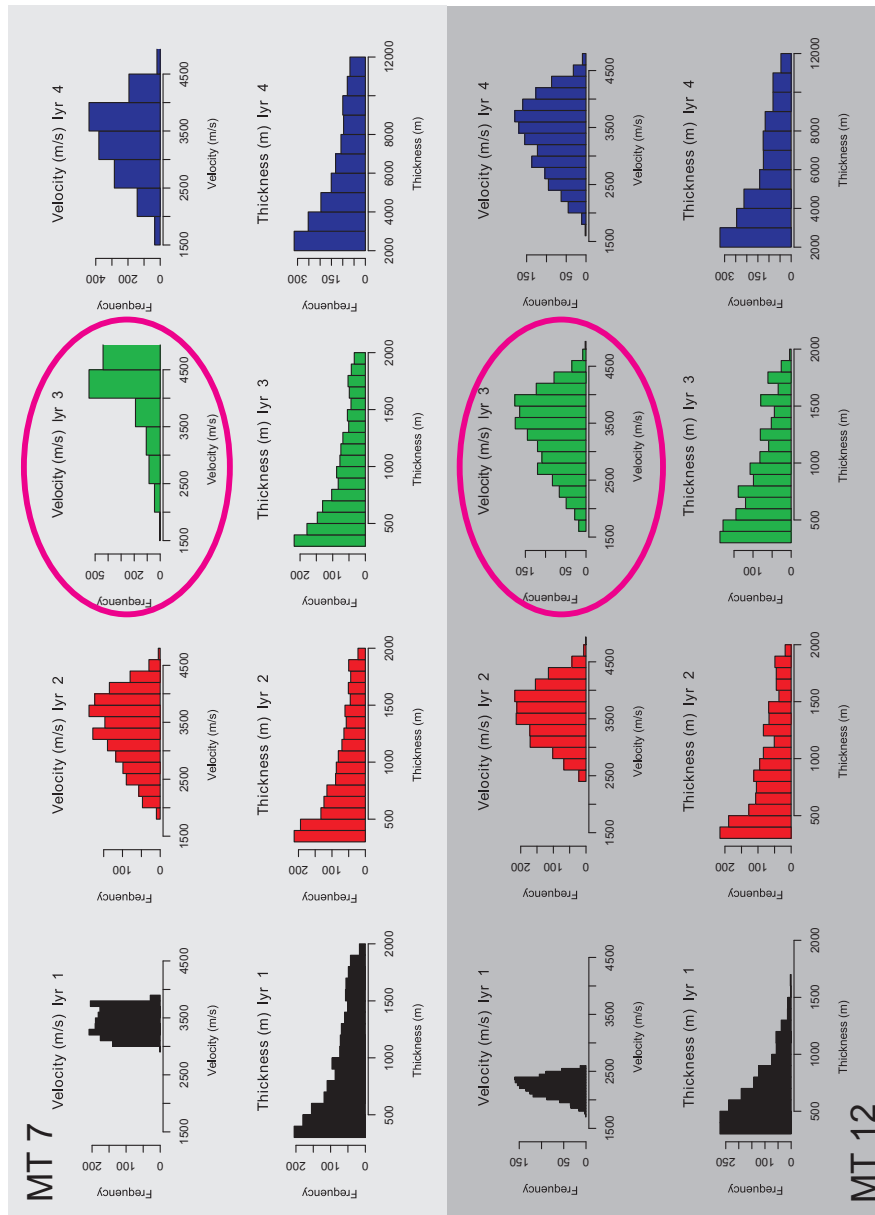


Figure 15: Histograms showing the marginal parameter distributions after 11 cycles screening only on seismic data for MT7 (left) and MT12 (right). MT7 is a profile over a salt body and MT12 is over sediment. Note that the salt structure is clearly being detected by the screening process in that a higher velocity in layer 3 (circled) is favored.

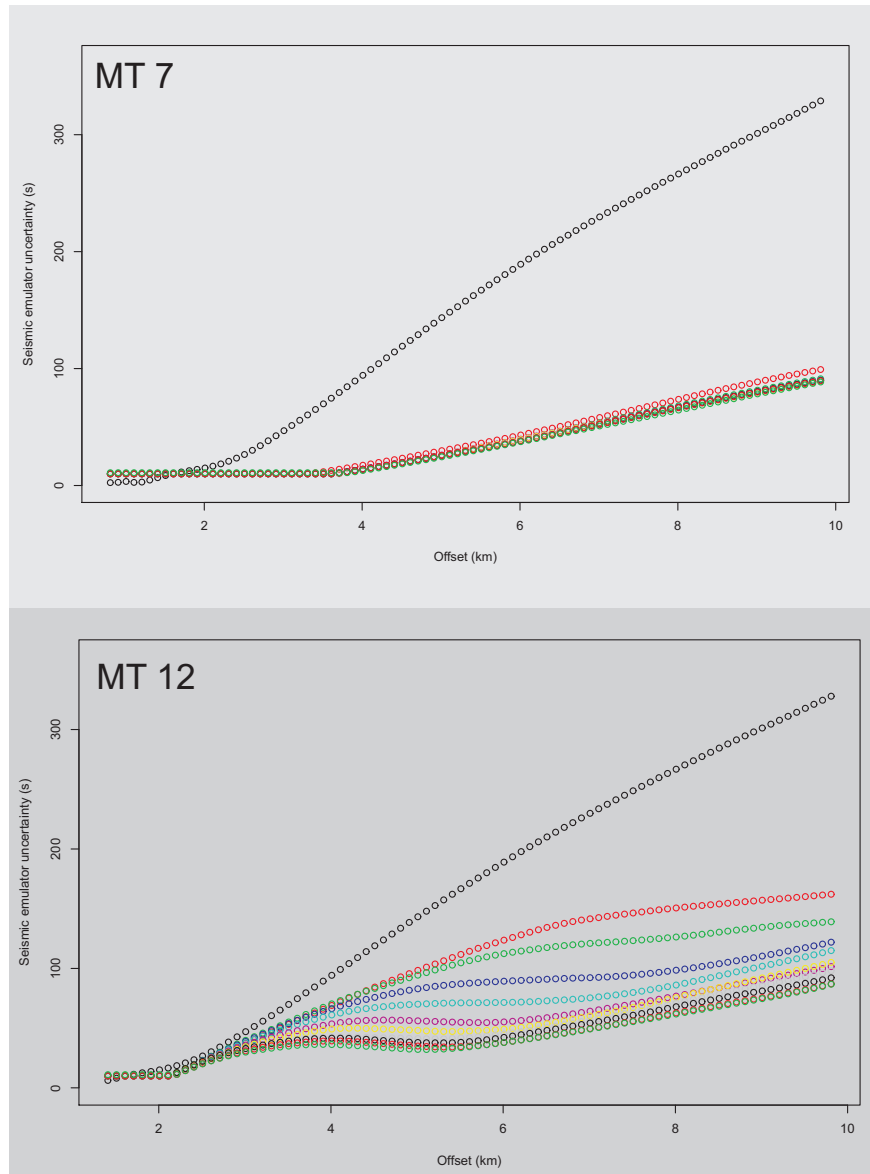


Figure 16: Misfit functions for each emulator screening only using the seismic emulator. Top: MT7 (over salt) Bottom: MT12 (over sediment).

$$\log(r) = 1.9421 + 2.4514 \times 10^{-5}v - 2.9587 \times 10^{-7}v^2 + 6.3142 \times 10^{-11}v^3 + N(0, 0.44) \quad (33)$$

$$\log(\rho) = 0.554 + 0.25 \log(v) + N(0, 0.0002) \quad (34)$$

586

587 From the same relationship data, the relationship for salt did not follow
 588 a discernible trend, but was seen to typically comprise velocities close to
 589 4500 ms⁻¹, very high resistivity (> 100Ωm) and density around 2160 ms⁻¹.
 590 After our initial analysis (Figure 16), we identified stations which appeared
 591 to lie over a salt body. In implementing the joint screening method, for these
 592 stations we specified that layer 3 should comprise salt.

593 On studying the system and observing that our method of model param-
 594 eterization was causing a number of non-unique models to be generated as
 595 a result of the distribution of layer thickness parameters, we modified the
 596 sampling scheme to use gamma distributions (with shape parameter 3) in
 597 order to sample over the thickness parameter ranges.

598 Use of real field data necessitates the normalization of the gravity data to
 599 a particular total model thickness. This was not the case with the synthetic
 600 experiment because the “true” data were generated using the same simulator
 601 which was being used to generate the gravity emulator training datasets. The
 602 real data, however, were not generated with such a simulator and so the sim-
 603 ulator output must be calibrated against the data. This was accomplished
 604 firstly by fixing the total model thickness to 12 km, and secondly by using
 605 one MT station (MT12) as a calibration station, in that we performed model
 606 screening simply using the MT and seismic datasets, then used the observed

607 relationship to give a distribution of plausible gravitational models at this
608 station. The gravity simulator outputs for these models were then gener-
609 ated and the modal (most likely) gravity value from this output distribution
610 used to calibrate an offset against the observed gravity value at this station.
611 This calibration was then applied at each of the other stations, where grav-
612 ity screening was included. Figure 17 shows the parameter histograms and
613 uncertainty functions for MT7, located over a major salt feature. These were
614 obtained using the seismic, gravity and MT emulators to screen the model
615 space. Note that the spike emulator was not used here because the data
616 did not contain sufficiently discernible gradient discontinuities for the spike
617 emulator to be built reliably. Note, on comparing Figure 17 with Figure 15,
618 we have also modified the prior velocity bounds (given by the range of the
619 histogram axes in Figure 15). This was in order to ensure that joint models
620 on the tail of the interparameter relationship distribution (Equation 34) were
621 included in the distribution of candidate models.

622 4. Discussion

623 The synthetic example described in Section 2.9, demonstrates how the
624 method is effective at screening model space for plausibility and how, in
625 a multi-cycle regime, this emulator-based approach provides a means for
626 quantifying the uncertainty associated with a modelling scenario.

627 In both the synthetic and real data examples, we see from the uncer-
628 tainty plots in Figures 13 and 16 that the emulator uncertainty functions are
629 reducing with increasing cycle. This shows that at each cycle, more struc-
630 ture is being obtained about the relationship between the model parameters

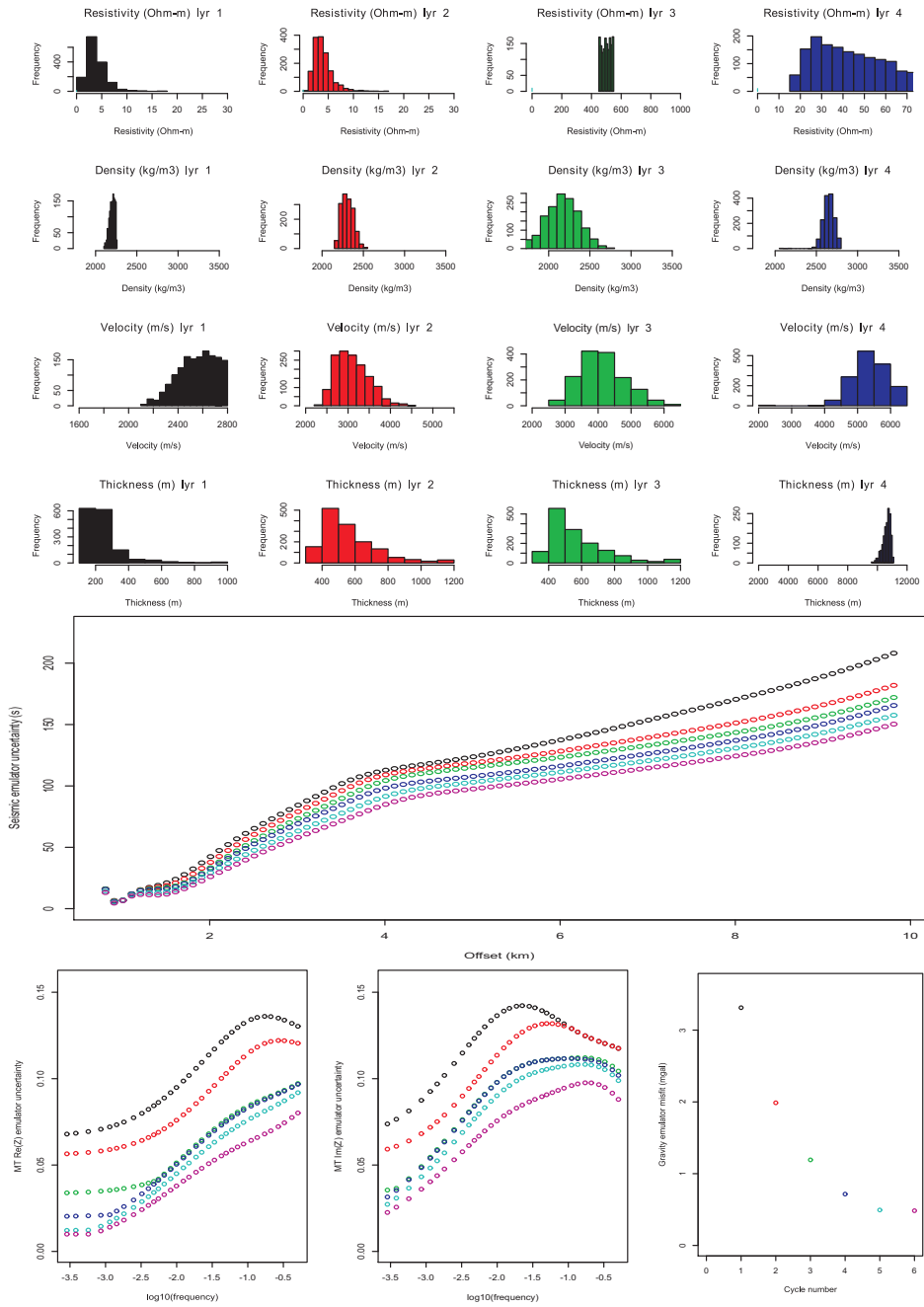


Figure 17: Histograms (top) showing plausible model parameters and emulator uncertainty functions (bottom) for a 1D profile at MT station MT7.

631 and the data. The limit of these uncertainty functions represents the total
632 uncertainty associated with the system, given user specifications about the
633 model space, the emulator parameterization, inter-parameter relationships,
634 data uncertainty, model discrepancy and others.

635 Comparing Figures 13, 16 and 17 shows that in the case of the synthetic
636 example the emulator uncertainty functions converge more efficiently than
637 in the case of the real data. Primarily this is due to the fact that in the
638 synthetic case, the class of model parameterization used to generate the syn-
639 thetic dataset (4 layers) is of the same class as the candidate models being
640 screened, whereas in the case of the real data, it is not the case that the earth
641 structure which generated the data consists of four distinct layers. Hence,
642 we also did not implement the spike emulator for screening in the case. This
643 highlights a limitation in our current implementation and how by incorpo-
644 rating more user knowledge about the target structure, more informative
645 results will be obtained through use of an intelligent and appropriate class
646 of parameterization for the candidate models.

647 As discussed briefly in section 2.9.1, we have implemented a relatively
648 simple marginal parameter sampling strategy. This means that model space
649 is excluded conservatively and that the considerable amount of joint informa-
650 tion contained in the full distribution of plausible models from the previous
651 cycle is not used. In developing the methodology further, particularly in the
652 case of more highly parameterized models, investigating conditional sampling
653 strategies will doubtless increase the efficiency of the method further.

654 Our results show the value of a top-down Monte Carlo model screen-
655 ing approach, where the constraint process begins by considering the whole

656 model space, rather than seeking to iteratively search the model space for
657 increasingly likely models. In particular, adopting such an approach removes
658 the possibility of underestimating the uncertainty associated with the sys-
659 tem through undersampling of the model space. The adoption of a simple
660 Monte Carlo screening scheme also means that all kinds of uncertainty can
661 be included in the analysis simply by generating the appropriate class of
662 candidate models, and allows for straightforward and explicit specification
663 of prior beliefs and uncertainties associated with the system, which is often
664 not the case with a search-based deterministic approach.

665 *4.1. Monte Carlo sampling or inverse solvers?*

666 Although we are strongly advocating the use of a forward screening ap-
667 proach in order to fully sample prior model space, we are not proposing that
668 this is the only valid “prescription” for geophysical constraint. The appro-
669 priate set of tools clearly depends on the system concerned, the degree of
670 importance attached to understanding the uncertainty associated with the
671 system, and the scale of the uncertainty in the problem relative to the size of
672 the potential space within which plausible models are being sought. For sys-
673 tems where there is no, or very little, uncertainty present in the system, and
674 thus where the plausible model space reduces to a point, or perhaps a col-
675 lection of separated points, then a sampling method, such as that advocated
676 here, will be very inefficient, since the probability of the system generat-
677 ing the one correct model will be very small (a ‘needle in a haystack’). In
678 this case, an inverse search-based solver will rapidly and usefully find an ac-
679 ceptable model. This result may then either be deemed sufficient in and of
680 itself, or be used to update the scientist’s understanding and beliefs about

681 the system, and allow a contraction of the relevant region of model space,
682 over which a full sampling may then be carried out, perhaps facilitated by
683 using an MCMC method.

684 Often, however, there is a considerable amount of uncertainty in the sys-
685 tem, but in order to obtain a result this may be ignored and it is often the
686 case that a single result is presented with a weakly substantiated statement
687 regarding the uncertainty on the result. In this study we have presented a
688 means by which a full sampling of the model space, which is required in order
689 to make robust statements regarding the uncertainty, may be tackled.

690 *4.2. Full simulator code or emulator?*

691 For the 1D problem shown here, it would doubtless be more efficient to
692 obtain a solution using the full simulator codes, and although we have used
693 a simple scenario to demonstrate the method, we do not suggest that the
694 emulation approach is necessarily the optimum way to obtain a solution for
695 this particular problem.

696 For the purposes of screening model space, given that an emulator is
697 trained using runs of the full simulator, it is a pertinent question as to whether
698 it would simply be more efficient to use the full simulator outputs. In many
699 ways the emulator described here can be thought of as interpolating the
700 model space between ‘known’ points sampled by the full simulator by rapidly
701 computing approximate outputs for the region between known points. For a
702 very large model space and where the simulator is computationally expensive,
703 and so the full simulator sampling is relatively sparse, it can greatly aid the
704 ability to discern the extent of the plausible region to estimate the region
705 between the sparsely known points. In cases where it is feasible to use the

706 full simulator code to sample the prior model space to a sufficient degree,
707 then in deeming the sampling density sufficient, there is by definition no
708 need to build an emulator. An emulator can therefore be considered useful
709 whenever it would be helpful to be able to quickly interpolate between points
710 of known model space. Once the plausible regions of model space have been
711 clearly identified as plausible given the uncertainties in the system, then the
712 user may wish to sample this region using the full simulator code.

713 As parallel computation methods are becoming increasingly common-
714 place, running a forward Monte Carlo screening scheme using the full sim-
715 ulators on a parallel system is clearly feasible. Indeed, the increasing use
716 of parallel systems will doubtless make larger numbers of forward simula-
717 tions feasible for larger problems too. However, given the size of the prior
718 model spaces for many modern real-world geophysical problems, in order to
719 screen the entirety of the model space, rather than adopting an MCMC or
720 deterministic bottom-up search-based strategy, it is difficult to see how this
721 could be achieved in the medium-term without a proxy-based method such
722 as emulation, and based on the results presented here, and from examples in
723 other fields, for example in Rougier et al. (2009) and Vernon and Goldstein
724 (2009), we propose that emulation may provide a valuable tool for geophys-
725 ical structural constraint.

726 **5. Conclusions**

727 In this paper we have shown that the method of emulation has the poten-
728 tial to make a large contribution in the field of geophysical modelling. From
729 the results presented, it is clear that use of an emulation method makes pos-

730 sible the handling of entire model spaces, rather than small portions thereof.
731 In the synthetic example, shown in Section 2.9.1, we have shown how, using
732 a desktop computer workstation over a period of ~ 72 hours, 245 million sets
733 of model parameters could be screened for plausibility, and that in general
734 use of an emulator affords a speed increase of several orders of magnitude in
735 terms of the rate at which models can be screened for plausibility.

736 In applying the method to a real dataset, we have also shown that an
737 emulator-based approach can be used to discern the plausible region of model
738 parameter space for a practical problem. The approach taken here is quite
739 simplistic in terms of the methodology and parameterization, however these
740 results show that with the application of larger computational resources,
741 emulators may make possible the handling very large model spaces for 3D
742 systems throughout geophysics.

743 In the field of deep crustal imaging, it is commonly the case that un-
744 certainties in data and in physics are not handled in a robust way, and
745 prior beliefs or assertions about the system are often not explicitly stated.
746 The result is that we often present an optimum model without a thorough
747 assessment of the associated uncertainty. We have presented, using a sim-
748 ple geophysical example, an approach which seeks to practicably tackle the
749 screening of the entire model space, with the aim of discerning all plausible
750 regions, rather than adopting a search-based approach, which may be prone
751 to not considering useful regions of model space, and to implicit conditioning
752 through sometimes weakly justified choice of regularization and smoothing
753 parameters.

754 Our aim here is not to present the model screening approach as the only

755 useful means for structural constraint. Indeed, as we discuss in Section 4,
756 the inverse and search-based approaches are highly useful particularly when
757 there is little uncertainty associated with the system and where there is
758 strong information about the plausibility of the prior model space. We are,
759 however, advocating that although they have not generally been employed
760 in recent decades on the grounds of computational expense, model screening
761 approaches such as that presented here, are increasingly feasible, and that
762 proxy-based emulators and other similar tools have the potential to help
763 facilitate this kind of screening method.

764 **6. Acknowledgements**

765 The authors would like to thank ITF and the sponsors of the JIBA (Joint
766 Inversion with Bayesian Analysis) consortium, through which this work was
767 funded, for their financial provision and advice. The JIBA sponsors are:
768 Statoil, Chevron, ExxonMobil, Nexen, Wintershall, RWE and Shell. We
769 would also like to thank Statoil in particular for providing a dataset on which
770 to test the method. Most of the coding for this study was carried out using R,
771 an open source statistical coding environment, along with several additional
772 modules (R Development Core Team, 2008; Dutang, 2009; Carnell, 2009).

773 **7. Bibliography**

774 **References**

775 Avdeev, D. B., Kuvshinov, A. V., Pankratov, O. V., Newman, G. A., 2002.
776 Three-dimensional induction logging problems, part I: An integral equation
777 solution and model comparisons. *Geophysics* 67, 413–426.

- 778 Bratley, P., Fox, B. L., 1988. Algorithm 659: Implementing sobol's quasi-
779 random sequence generator. *ACM Transactions on Mathematical Software*
780 14, 88–100.
- 781 Cagniard, L., 1953. Theory of the magneto–telluric method of geophysical
782 prospecting. *Geophysics* 18, 605–635.
- 783 Carnell, R., 2009. randtoolbox: Latin Hypercube Samples. R package version
784 0.5.
- 785 Dutang, C., 2009. randtoolbox: Generating and Testing Random Numbers.
786 R package version 1.07.
- 787 Gardner, G. H. F., Gardner, L. W., Gregory, A. R., 1974. Formation velocity
788 and density – the diagnostic basics for stratigraphic traps. *Geophysics* 39,
789 770–780.
- 790 Hastings, W. K., 1970. Monte Carlo sampling methods using Markov Chains
791 and their applications. *Biometrika*, 57, 97–109.
- 792 Heincke, B., Jegen, M., Chen, J., Hobbs, R. W., 2006. Joint Inversion of
793 MT, Gravity and Seismic Data applied to sub-basalt Imaging. In: *SEG*
794 *Technical Program Expanded Abstracts*. Vol. 25. Society of Exploration
795 Geophysicists, pp. 784–789.
- 796 Hole, J. A., Ryberg, T., Fuis, G. S., Bleibinhaus, F., and Sharma,
797 A. K., 2006. Structure of the San Andreas fault zone at SAFOD from
798 a seismic refraction survey. *Geophysical Research Letters*, 33, L07312,
799 doi:10.1029/2005GL025194.

- 800 James, M. B., and M. H. Ritzwoller, 1999, Feasibility of truncated perturba-
801 tion expansions to approximate rayleigh-wave eigenfrequencies and eigen-
802 functions in heterogeneous media: Bulletin of the Seismological Society of
803 America, **89**, 433–441.
- 804 Jegen-Kulcsar, M., Hobbs, R. W., Tarits, P., Chave, A., 2009. Joint in-
805 version of marine magnetotelluric and gravity data incorporating seismic
806 constraints: preliminary results of sub-basalt imaging off the Faroe Shelf.
807 Earth and Planetary Science Letters 282, 47–55.
- 808 Kennett, B. L. N., 2001. The Seismic Wavefield. Volume 1: Introduction and
809 Theoretical Development. x+370 pp. Cambridge, New York, Melbourne:
810 Cambridge University Press. ISBN 0 521 80945 2
- 811 Logemann, K., Backhaus, J. O., and Harms, I. H., 2004. A statistical emu-
812 lator of the north-east Atlantic circulation. Ocean Modelling, 7, 97–110.
- 813 Lunnø, Z. C., Christie, P. A. F., White, R. S., 2003. An evaluation of peak
814 and bubble tuning in sub-basalt seismology: modelling and results. First
815 Break 21, 51–56.
- 816 Meier, U., A. Curtis, and J. Trampert, 2007. Global crustal thickness from
817 neural network inversion of surface wave data: Geophys. J. Int., **169**, 706–
818 722.
- 819 Moorkamp, M., Jegen, M., Roberts, A., Hobbs, R., 2011. A framework for
820 3-D joint inversion of MT, gravity and seismic refraction data. Geophysical
821 Journal International 184(1): 477493

- 822 Podvin, P., Lecomte, I., 1991. Finite difference computation of traveltimes
823 in very contrasted velocity models: a massively parallel approach and its
824 associated tools. *Geophysical Journal International* 105, 271–284.
- 825 Press, F., 1970, Earth models obtained by Monte Carlo inversion: *J. Geo-*
826 *phys. Res.*, **75**, 6575–6581.
- 827 R Development Core Team, 2008. *R: A Language and Environment for Sta-*
828 *tistical Computing*. R Foundation for Statistical Computing, Vienna, Aus-
829 *tria*, ISBN 3-900051-07-0.
- 830 URL <http://www.R-project.org>
- 831 Roberts, A. W., White, R. S., Christie, P. A. F., 2009. Imaging magmatic
832 rocks on the North Atlantic rifted continental margin. *Geophysical Journal*
833 *International* 179, 1024–1038.
- 834 Rougier, J., Sexton, D., Murphy, J. M., and Stainforth, D., 2009. Analyzing
835 the Climate Sensitivity of the HadSM3 Climate Model Using Ensembles
836 from Different but Related Experiments. *J. Climate*, 22, 3540–3557.
- 837 Sambridge, M., and K. Mosegaard, 2002, Monte Carlo methods in geophysi-
838 cal inverse problems: *Reviews of Geophysics*, **40(3)**,1009, 1–24.
- 839 Shapiro, N. M., and M. H. Ritzwoller, 2002, Monte-carlo inversion for a global
840 shear-velocity model of the crust and upper mantle: *Geophys. J. Int.*, **151**,
841 88–105.
- 842 Smith, A. F. M., and Roberts, G. O., 1993, Bayesian computation via the
843 Gibbs sampler and related Markov chain Monte-Carlo methods (with dis-
844 cussion). *J. Roy. Stat. Soc. Series B* 55, 3–23.

- 845 Vernon, I., Goldstein, M., 2009. Bayes Linear Analysis of Imprecision in Com-
846 puter Models, with Application to Understanding Galaxy Formation. In:
847 Augustin, T., Coolen, F. P. A., Moral, S., Troffaes, M. C. M. (Eds.), Pro-
848 ceedings of the Sixth International Symposium on Imprecise Probability:
849 Theories and Applications. No. S061. Society for Imprecise Probability:
850 Theories and Applications, pp. 441–450.
- 851 Zelt, C. A., and Barton, P. J., 1998. Three-dimensional seismic refraction
852 tomography: a comparison of two methods applied to data from the Faeroe
853 Basin. *J. Geophys. Res.*, 103, 7187–7210.
- 854 Ziolkowski, A., Hanssen, P., Gatliff, R., Li, X., Jakubowicz, H., 2001. The use
855 of low frequencies for sub-basalt imaging. In: 71st Annual International
856 Meeting. SEG, pp. 74–77.

Cycle	Models tested	Number of runs	Seismic simulator time	Seismic emulator time
1	29,302	10^4	45 minutes	15s
2	1,011,997	2.4×10^6	7.5 days	1 hour
3	1,536,936	5.76×10^7	180 days	1 day
4	3,188,711			
5	5,815,746	Number of runs	MT simulator time	MT emulator time
6	14,041,545	10^4	12.8 minutes	1 minute
7	10,080,014	2.4×10^6	2.1 days	4 hours
8	7,010,975	5.76×10^7	50.4 days	4 days
9	145,511,481	Number of runs	Gravity simulator time	Gravity emulator time
10	34,384,460	10^4	3s	3s
11	31,993,618	2.4×10^6	12 minutes	12 minutes
		5.76×10^7	4.8 hours	4.8 hours
Total	245,532,785			

Table 1: Left: Number of models tested in each emulator cycle (by each of the spike, gravity, MT and seismic emulators) in order to generate a population of 1000 plausible models. The histograms of the final selection are shown in Figure 12. Right: Typical run times for the simulator vs emulator for each of the seismic, MT and gravity cases on a high end workstation. The typical time required to construct the emulators ranged from 30 s in the case of the gravity emulator to about 10 minutes in the case of the seismic emulator. Note that the main factor controlling the emulator times is the number of α and β coefficients being used. As a result, while the MT simulator is considerably faster than the seismic simulator, the seismic emulator is faster to run than the MT emulator. In the case of the gravity emulator, the simulator is very simple and outputs only one point, and hence the emulator affords little advantage over the full simulator in this case. Comparison times for the spike emulator are not shown since it was run “piggy-backing” on the seismic emulator.

Layer:	1	2	3	4
Velocity (ms^{-1}):	1600-2800	2000-5500	2000-6500	2000-6500
Density (kgm^{-3}):	1800-3600	1800-3600	1800-3600	1800-3600
Resistivity (Ωm):	0.5-30	0.5-30	0.5-30	0.5-30
Thickness (m):	300-1000	300-1000	300-1000	300-1000

Table 2: Prior model space used for real dataset.