Geophysical Journal International

Geophys. J. Int. (2011) **185,** 663–675 doi: 10.1111/j.1365-246X.2011.04996.x

Non-linearity in Bayesian 1-D magnetotelluric inversion

Rongwen Guo,[1,2] Stan E. Dosso,[2] Jianxin Liu,[1] Jan Dettmer[2] and Xiaozhong Tong[1]

[1] *School of Geosciences and Info-Physics, Central South University, Changsha* 410083*, China*

[2] *School of Earth and Ocean Sciences, University of Victoria, Victoria BC, Canada. E-mail: rwguo@uvic.ca*

Accepted 2011 February 21. Received 2010 December 29; in original form 2010 November 23

SUMMARY

This paper applies a Bayesian approach to examine non-linearity for the 1-D magnetotelluric (MT) inverse problem. In a Bayesian formulation the posterior probability density (PPD), which combines data and prior information, is interpreted in terms of parameter estimates and uncertainties, which requires optimizing and integrating the PPD. Much work on 1-D MT inversion has been based on (approximate) linearized solutions, but more recently fully non-linear (numerical) approaches have been applied. This paper directly compares results of linearized and non-linear uncertainty estimation for 1-D MT inversion; to do so, advanced methods for both approaches are applied. In the non-linear formulation used here, numerical optimization is carried out using an adaptive-hybrid algorithm. Numerical integration applies Metropolis-Hastings sampling, rotated to a principal-component parameter space for efficient sampling of correlated parameters, and employing non-unity sampling temperatures to ensure global sampling. Since appropriate model parametrizations are generally not known *a priori*, both under- and overparametrized approaches are considered. For underparametrization, the Bayesian information criterion is applied to determine the number of layers consistent with the resolving power of the data. For overparametrization, prior information is included which favours simple structure in a manner similar to regularized inversion. The data variance and/or trade-off parameter regulating data and prior information are treated in several ways, including applying fixed optimal estimates (an empirical Bayesian approach) or including them as hyperparameters in the sampling (hierarchical Bayesian). The latter approach has the benefit of accounting for the uncertainty in the hyperparameters in estimating model parameter uncertainties. Non-linear and linearized inversion results are compared for synthetic test cases and for the measured COPROD1 MT data by considering marginal probability distributions and marginal profiles. In some cases, important differences are indicated, including poorer sensitivity to thin and/or low-conductivity layers for linearized inversion, and multimodal PPDs which cannot be addressed within a linearized approach.

Key words: Inverse theory; Probability distributions; Magnetotelluric.

1 INTRODUCTION

The 1-D magnetotelluric (MT) inverse problem of estimating layered earth conductivity models using surface measurements of electromagnetic fields has received much attention (e.g. Tikhonov 1950; Parker 1970; Weidelt 1972; Rokityansky 1982; Whittall & Oldenburg 1992). Many approaches to MT inversion have been based on parameter estimation by linearized inversion (e.g. Becher & Sharpe 1969; Oldenburg 1979; Constable *et al.* 1987; Smith & Booker 1988), with less attention applied to non-linear analysis and to uncertainty estimation. Bayesian inference theory provides a natural formulation for such investigations.

Bayesian inversion (Tarantola 1987; Tarits *et al.* 1994) seeks to quantify properties of the posterior probability density (PPD) which combines data information, in the form of a likelihood function, and prior information. PPD properties of interest typically include optimal parameter estimates (such as the maximum *a posteriori*, MAP, estimate), parameter uncertainties (covariance matrix and marginal probability distributions) and parameter interrelationships (correlations and joint marginal distributions). Computing these properties requires optimizing and integrating the PPD over the multidimensional parameter space. Analytic solutions exist for linear inverse problems with Gaussian-distributed data errors and Gaussian or uniform priors, where the PPD is itself a multivariate Gaussian distribution. These solutions can be extended to weakly non-linear problems using local linearization; however, the applicability of this approximation generally cannot be evaluated without comparison to fully non-linear solutions. Comparing linearized and non-linear

inversion and uncertainty estimation for 1-D MT is the primary goal of this work.

Increased low-cost computational power has made possible non-linear (numerical) PPD integration based on stochastic sampling techniques; however, care is required to ensure that the sampling procedure is unbiased and complete. Monte Carlo integration (Press *et al.* 1992), based on drawing random samples from the domain of integration, provides unbiased integral estimates but can be very slow. Multiple-MAP sampling, based on models collected via one or more runs of an optimization algorithm (Sen & Stoffa 1995), is much faster but produces biased results (considered in Section 2.4). Markov-chain Monte Carlo (MCMC) methods, such as Metropolis Hastings sampling (MHS) and Gibbs' sampling, provide unbiased integral estimates, with models drawn (asymptotically) proportional to the PPD (Gilks *et al.* 1996; Sambridge & Mosegaard 2002). Sampling efficiency and completeness are key issues for MCMC methods, with the goal of achieving a well-mixed Markov chain that effectively samples the parameter space, avoiding both small, ineffectual perturbations and high perturbation rejection rates. In particular, interparameter correlations, which result in high-probability regions oriented obliquely to the parameter axes, can inhibit mixing in some algorithms. Further, for non-linear problems with potentially multimodal PPDs, care must be taken to ensure global sampling which can encompass disjoint regions of high probability.

Quantitative parameter uncertainty estimation requires knowledge of the data uncertainties and of an appropriate model parametrization. In particular, a precise value of the variance is needed for misfit evaluation, but may not be available in practical cases. Within a Bayesian formulation, this can be addressed by estimating the variance via optimization and using this value in sampling (an empirical Bayesian approach), or by treating the variance as a hyperparameter included in the sampling (hierarchical Bayesian). The latter has the advantage that effects of the uncertainty in the hyperparameter are included in the model uncertainty estimates (Malinverno & Briggs 2004). In terms of model parametrization, adopting too few parameters (e.g. layers) can lead to underfitting the data, biasing parameter estimates and underestimating uncertainties, while adopting too many parameters yields underdetermined parameters with excessive variance. Common approaches to parametrization include underparametrization and overparametrization. Underparametrization is based on determining the minimum number of parameters consistent with the resolving power of the data (e.g. Weaver & Agarwal 1993). Overparametrization adopts a large number of parameters of dimensions below the resolution limit of the data and explicitly controls the amount of model structure (variability) via regularization or prior information (e.g. Constable *et al.* 1987; Smith & Booker 1988; Dosso & Oldenburg 1989). This approach typically involves a trade-off parameter that controls the balance between fitting the data and smoothing the model, or, equivalently, between model variance and resolution. This trade-off parameter can be treated with either empirical or hierarchical Bayesian approaches (the latter does not appear to have been done previously for MT inversion). A recent and promising alternative to under- or overparametrization is trans-dimensional inversion, which treats the number of parameters as a hyperparameter in a hierarchical (numerical) Bayesian approach, sampling over parametrizations of different dimensions proportional to their probability (Malinverno 2002; Sambridge *et al.* 2006). This approach does not yet appear to have been applied to MT, and is not applied here as it would preclude comparison to (analytic) linearized inversion.

Tarits *et al.* (1994) applied Monte Carlo integration in Bayesian 1-D MT inversion to obtain marginal distributions for layer thicknesses and conductivities (discretized over bounded prior intervals), considering parametrizations based on a small number of layers chosen subjectively. The marginal distributions obtained were unimodal but sometimes significantly non-Gaussian. Grandis *et al.* (1999) considered inversion for a stack of layers of fixed thicknesses with prior information that favoured smooth models (i.e. an overparametrized approach), employing Gibbs' sampling to sample the PPD. The effect of different (fixed) arbitrary choices of the smoothing factor (trade-off parameter) was investigated, with data variance treated as unknown. Cerv *et al.* (2007) considered three approaches to estimating uncertainty distributions for 1-D MT inversion including multiple-MAP estimation, Gibbs' sampling and neighbourhood-algorithm resampling. They noted that, while efficient, multiple-MAP and neighbourhood sampling produced biased uncertainty estimates. Gibbs' sampling provided unbiased estimates given adequate sampling; however, Cerv *et al.* (2007) noted that their algorithm was inefficient for correlated parameters which could preclude convergence.

This paper develops linearized and non-linear MT inversions based on both under- and overparametrized approaches, and examines the effects of non-linearity on parameter uncertainty distributions. For underparametrization, the Bayesian information criterion (BIC) is applied to determine the number of layers in an objective manner. For overparametrization, prior information is included which favours simple structure. In both approaches the data variance and/or prior weighting factor are treated unknowns in the inversion. An MHS algorithm is applied which samples in principal-component parameter space to overcome inefficiencies due to correlated parameters (Nolte & Frazer 1994), with high-temperature sampling (Brooks & Frazer 2005) applied to ensure wide sampling of potentially multimodal parameter spaces. Simulated and measured MT data are considered.

Although 1-D MT is presently used less than 2-D and 3-D approaches, a large body of existing work has applied this method. Further, 1-D MT is still used to investigate simple structure such as basin depths (Adam *et al.* 2007), mid-ocean ridges (Oskooi *et al.* 2005) and geothermal fields (Cumming & Mackie 2010). The method is also applied in joint inversion with other types of geophysical data (Son & Sun 2005; Jegen *et al.* 2009; Cumming & Mackie 2010), as well as in quality control and to construct starting models for higher dimensional inversions (Kumar *et al.* 2010). Finally, 1-D MT has often been used as an example problem for advances or reviews of geophysical inverse theory (e.g. Constable *et al.* 1987; Whittall & Oldenburg 1992; Tarits *et al.* 1994; Pedersen 2004; Cerv *et al.* 2007; Niwas *et al.* 2007). Much of the above work is based on linearized inversion; hence, it is worthwhile understanding the applicability of this approximation by comparison to fully non-linear methods. The non-linear approaches applied here are general and applicable to other geophysical inverse problems and, in principle, to 2-D or 3-D MT inversion. However, higher dimensional non-linear inversions are computationally demanding, and beyond the scope of this work.

2 INVERSE THEORY AND ALGORITHMS

2.1 Bayesian formulation

This section briefly describes the Bayesian inversion formulation used here; more detailed treatments of Bayesian theory can be found

elsewhere (e.g. Tarantola 1987; Tarits *et al.* 1994). Bayes' rule relating a set of M parameters $\mathbf{m}$, representing a model parametrization θ, and N data $\mathbf{d}$ can be written

$$P(\mathbf{m}|\mathbf{d},\theta) = \frac{P(\mathbf{d}|\mathbf{m},\theta)\,P(\mathbf{m},\theta)}{P(\mathbf{d},\theta)}. \quad (1)$$

In eq. (1), $P(\mathbf{m}|\mathbf{d},\theta)$ denotes the conditional probability density function (PDF) of $\mathbf{m}$ given $\mathbf{d}$ and θ, $P(\mathbf{d}|\mathbf{m},\theta)$ is the conditional PDF of $\mathbf{d}$, and $P(\mathbf{m},\theta)$ and $P(\mathbf{d},\theta)$ represent the model and data prior PDFs, respectively. Given observed data and a fixed parametrization, $P(\mathbf{d}|\mathbf{m})$ is interpreted as the likelihood function, $L(\mathbf{m}|\mathbf{d})$, and eq. (1) becomes

$$P(\mathbf{m}|\mathbf{d}) \propto L(\mathbf{m}|\mathbf{d})\,P(\mathbf{m}), \quad (2)$$

where $P(\mathbf{m}|\mathbf{d})$ is the PPD. The likelihood can generally be written $L(\mathbf{m}|\mathbf{d}) \propto \exp[-E(\mathbf{m},\mathbf{d})]$, where E is the data misfit function, and the PPD expressed

$$P(\mathbf{m}|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m},\mathbf{d})]}{\int \exp[-\phi(\mathbf{m}')]\,\mathrm{d}\mathbf{m}'}, \quad (3)$$

where the integration spans the parameter space and ϕ represents the generalized misfit, combining data and prior, as

$$\phi(\mathbf{m},\mathbf{d}) = E(\mathbf{m},\mathbf{d}) - \ln P(\mathbf{m}). \quad (4)$$

In a Bayesian formulation, the PPD is considered the complete solution to the inverse problem. However, understanding this multidimensional distribution typically requires calculating its properties in terms of parameter estimates, uncertainties and interrelationships, such as the MAP model, mean model, model covariance matrix and marginal distributions given, respectively, by

$$\hat{\mathbf{m}} = \mathrm{Arg}_{\max}\,P(\mathbf{m}|\mathbf{d}) = \mathrm{Arg}_{\min}\,\phi(\mathbf{m},\mathbf{d}), \quad (5)$$

$$\bar{\mathbf{m}} = \int \mathbf{m}'\,P(\mathbf{m}'|\mathbf{d})\,\mathrm{d}\mathbf{m}', \quad (6)$$

$$\mathbf{C}_m = \int (\mathbf{m}'-\bar{\mathbf{m}})(\mathbf{m}'-\bar{\mathbf{m}})^T\,P(\mathbf{m}'|\mathbf{d})\,\mathrm{d}\mathbf{m}', \quad (7)$$

$$P(m_i|\mathbf{d}) = \int \delta(m_i'-m_i)\,P(\mathbf{m}'|\mathbf{d})\,\mathrm{d}\mathbf{m}', \quad (8)$$

where δ is the Dirac delta function and higher order marginals are computed similarly to eq. (8). The above properties can be calculated analytically for linear inverse problems with Gaussian-distributed data errors and Gaussian or uniform priors, in which case the PPD is itself a Gaussian distribution. For non-linear inversion, linearization and linear inversion can be applied; the applicability of this approximation depends on the degree of non-linearity. Alternatively, numerical solutions can be applied which suffer no linearization errors and are fully general, at the cost of increased computational effort.

2.2 Underparametrized approach: model selection

In many practical applications, an appropriate model parametrization is not known *a priori*. In such cases, two approaches are commonly applied: (i) underparametrization, based on determining the minimum number of model parameters consistent with the resolving power of the data; and (ii) overparametrization, based on parametrizing at a small scale (below the data resolution limit) and explicitly controlling the amount of model structure (considered in Section 2.3).

In an underparametrized approach, model selection is ideally based on Bayesian evidence, defined as the likelihood of a particular parametrization given the observed data,

$$P(\mathbf{d}|\theta) = \int P(\mathbf{d}|\mathbf{m},\theta)P(\mathbf{m}|\theta)\,\mathrm{d}\mathbf{m}. \quad (9)$$

Maximizing the evidence naturally produces simple structure (Kass & Raftery 1995). However, evaluating evidence requires integrating the likelihood with respect to the prior, which is particularly challenging for non-linear problems. In this paper, an asymptotic point estimate of evidence, the BIC (Schwarz 1978; Kass & Raftery 1995; Dettmer *et al.* 2009) is applied for model selection.

$$-2\ln P(\mathbf{d}|\theta) \approx \mathrm{BIC} = 2E(\hat{\mathbf{m}},\mathbf{d},\theta) + M\ln N. \quad (10)$$

The BIC can be computed for a series of models with different numbers of parameters, with the parametrization that minimizes the BIC selected. The first term on the right-hand side in eq. (10) favours low data misfit; however, the second term mitigates against including unnecessary parameters. Compared to other information measures, such as the F-test (Weaver & Agarwal 1993) and the Akaike information criterion (Akaike 1973), the BIC penalizes extra parameters more strongly and avoids a bias towards unjustifiably complex models.

Applying the BIC requires computing optimal models for a variety of parametrizations. Because this involves models with relatively few parameters (compared to overparametrization), non-linear inversion is applicable. This paper uses adaptive simplex simulated annealing (ASSA, Dosso *et al.* 2001), to compute MAP estimates. ASSA is a hybrid optimization that adaptively combines elements of the downhill simplex (DHS) algorithm (Press *et al.* 1992) and the global search method of very fast simulated annealing (Ingber 1989). Compared to traditional global searches (e.g. simulated annealing and genetic algorithms), ASSA overcomes inefficiency at sensing gradients, particularly for problems involving correlated parameters, and retains a working memory of the best models encountered in the search. DHS steps are geometric operations (reflection, extension and contraction) applied to a simplex of $M+1$ models in an M-D space, providing gradient-based steps without the computational effort of evaluating partial derivatives or solving systems of equations. Random perturbations are applied to the DHS results, with perturbations for each parameter consisting of a Cauchy-distributed random variable scaled by the running average of the size of recently accepted perturbations. Perturbations are accepted with probability given by the Metropolis criterion

$$\alpha \le \exp(-\Delta\phi/T), \quad (11)$$

where α is a random number on $U(0,1)$, $\Delta\phi$ is the change in misfit due to the perturbation and T (temperature) is a control parameter which is gradually decreased. If a high perturbation rejection rate is encountered during cooling, a DHS multiple contraction is applied. This procedure is repeated until the difference between the highest and lowest mismatches in the simplex relative to their mean is less than a specified tolerance (10^{-5} is used here).

2.3 Overparametrized approach: linearized parameter estimation

Overparametrized inversion adopts many parameters and controls model structure via prior information or regularization. The large

number of parameters can make parameter estimation by numerical optimization inefficient, and a linearized approach is described here. The linearized solution can then be used as a starting point for non-linear uncertainty estimation, as described in the following section. The approach applied here treats both the data variance and the weight of the structural prior as unknowns determined by the data. This requires including variance and prior hyperparameters in the inversion, which can be solved for analytically in a linearized approach (Mitsuhata & Uchida 2002) as outlined in this section, or sampled in non-linear inversion (Section 2.4).

Consider a set of observed data contaminated by Gaussian errors with covariance matrix $\mathbf{C}_d = s\ \mathbf{C}'_d$, where $\mathbf{C}'_d$ represents a preliminary estimate of the covariance and s is a correction or scale factor (Mitsuhata *et al.* 2001). For a fixed parametrization the likelihood function is

$$L(\mathbf{m}|\mathbf{d}, s)=\frac{1}{(2\pi s)^{N/2}|\mathbf{C}'_d|^{1/2}}\exp\left[-\frac{1}{2s}(\mathbf{d}-\mathbf{Am})^T\mathbf{C}_d'^{-1}(\mathbf{d}-\mathbf{Am})\right]. \tag{12}$$

Further, consider prior information regarding model structure specified as (Mitsuhata & Uchida 2002)

$$P(\mathbf{m}|\mu, s) = \frac{1}{(2\pi s/\mu)^{M'/2}}\exp\left[-\frac{\mu}{2s}\mathbf{m}^T\mathbf{D}^T\mathbf{D}\,\mathbf{m}\right], \tag{13}$$

where μ is a scaling parameter for the prior information and $\mathbf{D}$ represents a matrix derivative operator of rank M'. In this paper $\mathbf{D}$ is taken to be a first derivative matrix, that is, $\mathbf{D} = \text{bidiag}\,[1, -1]$, and $M' = M - 1$. Substituting eqs (12) and (13) into (2) yields

$$P(\mathbf{m}|\mathbf{d}, \mu, s) \propto \exp\left[-\frac{\psi(\mathbf{m}, \mu)}{2s}+\frac{M-1}{2}\ln\mu-\frac{N+M-1}{2}\ln s\right], \tag{14}$$

where

$$\psi(\mathbf{m}, \mu) = (\mathbf{d}-\mathbf{Am})^T\mathbf{C}_d'^{-1}(\mathbf{d}-\mathbf{Am}) + \mu\mathbf{m}^T\mathbf{D}^T\mathbf{D}\,\mathbf{m}. \tag{15}$$

Setting $\partial\ \psi(\mathbf{m})/\partial\mathbf{m} = 0$ leads to the MAP solution

$$\hat{\mathbf{m}} = \left[\mathbf{A}^T\mathbf{C}_d'^{-1}\mathbf{A} + \mu\,\mathbf{D}^T\mathbf{D}\right]^{-1}\mathbf{A}^T\mathbf{C}_d'^{-1}\mathbf{d} \tag{16}$$

and the posterior model covariance matrix

$$\mathbf{C}_m = s\left[\mathbf{A}^T\mathbf{C}_d'^{-1}\mathbf{A} + \mu\,\mathbf{D}^T\mathbf{D}\right]^{-1}. \tag{17}$$

If μ is unknown but the variance scaling s is known, a common approach is to choose μ such that the χ^2 data misfit

$$\chi^2 = s^{-1}(\mathbf{d}-\mathbf{A}\hat{\mathbf{m}})^T\mathbf{C}_d'^{-1}(\mathbf{d}-\mathbf{A}\hat{\mathbf{m}}), \tag{18}$$

obtains its expected value of N (Constable *et al.* 1987). However, prior knowledge of the data variance is not always available. In such cases, Akaike's Bayesian information criterion (ABIC) can be minimized to estimate both μ and s (Mitsuhata & Uchida 2002). Based on entropy maximization, the ABIC is defined

$$\text{ABIC} = -2\ln P(\mathbf{d}|\mu, s) + 2\cdot(\text{number of hyperparameters}), \tag{19}$$

where the number of hyperparameters here is two (i.e. μ and s), and $P(\mathbf{d}\ |\mu,\ s)$ is the evidence for μ and s, for example, eq. (9). For linear problems, Mitsuhata *et al.* (2001) solve the evidence integral to give

$$\text{ABIC} = (N-1)\ln(2\pi s) - 2\ln|\mathbf{C}_d'^{-1/2}| - (M-1)\ln\mu + \ln|\mathbf{A}^T\mathbf{C}_d'^{-1}\mathbf{A} + \mu\mathbf{D}^T\mathbf{D}| + \psi(\hat{\mathbf{m}}, \mu)/s + 4. \tag{20}$$

Setting $\partial\text{ABIC}/\partial s = 0$, the optimum value for the variance scaling is

$$s = \frac{\psi(\hat{\mathbf{m}}, \mu)}{N-1}. \tag{21}$$

Substituting eqs (21) and (16) into (20) gives the ABIC as a function of μ only, which can be minimized numerically using line-search methods [e.g. the golden section algorithm (Press *et al.* 1992)]. The optimal μ value can then used to estimate $\hat{\mathbf{m}}$, s and $\mathbf{C}_m$ via eqs (16), (21) and (17), respectively.

Non-linear problems can be solved (approximately) via linearization and iteration. Expanding a non-linear problem $\mathbf{d} = \mathbf{d}(\mathbf{m})$ in a Taylor series to first order about a starting model $\mathbf{m}_0$ leads to $\mathbf{d}' = \mathbf{Am}$, where $\mathbf{A}$ is the Jacobian matrix, $A_{ij} = \partial\ d_i(\mathbf{m}_0)/\partial\ m_j$, and $\mathbf{d}' = \mathbf{d} - \mathbf{d}(\mathbf{m}_0) + \mathbf{Am}_0$. The estimate for s is evaluated using eq. (21) but with ψ given by

$$\psi(\mathbf{m}, \mu) = (\mathbf{d}-\mathbf{d}(\mathbf{m}))^T\mathbf{C}_d'^{-1}(\mathbf{d}-\mathbf{d}(\mathbf{m})) + \mu\,\mathbf{m}^T\mathbf{D}^T\mathbf{D}\mathbf{m}. \tag{22}$$

The linearized MAP and covariance matrix are given by eqs (16) and (17) with $\mathbf{d}$ replaced by $\mathbf{d}'$, and the linearized PPD is an M-D Gaussian distribution

$$P(\mathbf{m}|\mathbf{d}) = \frac{1}{(2\pi)^{M/2}|\mathbf{C}_m|^{1/2}}\exp\left[-(\mathbf{m}-\hat{\mathbf{m}})^T\mathbf{C}_m^{-1}(\mathbf{m}-\hat{\mathbf{m}})/2\right]. \tag{23}$$

2.4 Non-linear inversion: numerical integration

This section considers numerical evaluation of integral properties of the PPD, eqs (6)–(9), of the form

$$I = \int f(\mathbf{m})P(\mathbf{m}|\mathbf{d})\,\mathrm{d}\mathbf{m}. \tag{24}$$

In uniform Monte Carlo integration (Press *et al.* 1992), Q models are drawn uniformly at random from the domain of integration. The probability of each sample is assigned

$$P(\mathbf{m}_i|\mathbf{d}) = \frac{\exp[-\phi(\mathbf{m}_i)]}{\sum_{j=1}^{Q}\exp[-\phi(\mathbf{m}_j)]}, \quad i = 1, Q, \tag{25}$$

and the integral is estimated

$$I \approx \sum_{i=1}^{Q} f(\mathbf{m}_i)P(\mathbf{m}_i|\mathbf{d}). \tag{26}$$

However, it is inefficient to draw models uniformly if the integrand is concentrated in localized regions of the parameter space. Monte Carlo importance sampling is based on preferentially drawing samples from regions that contribute significantly to the integral. Let $g(\mathbf{m})$ represent a sample-generating distribution [with $\int g(\mathbf{m})\mathrm{d}\mathbf{m} = 1$] from which Q models are drawn; eq. (24) can then be written

$$I = \int \frac{f(\mathbf{m})P(\mathbf{m}|\mathbf{d})}{g(\mathbf{m})}g(\mathbf{m})\,d\mathbf{m} \approx \frac{1}{Q}\sum_{i=1}^{Q}\frac{f(\mathbf{m}_i)P(\mathbf{m}_i|\mathbf{d})}{g(\mathbf{m}_i)}. \tag{27}$$

It is worth noting that multiple MAP integration methods (Sen & Stoffa 1995), based on collecting models using optimizations with unknown sample-generating functions (e.g. simulated annealing and genetic algorithms) do not correct for $g(\mathbf{m})$ in the denominator of eq. (27). Hence, even if the parameter space is sampled adequately, the integral estimates are (potentially strongly) biased.

The MCMC method of Metropolis-Hastings sampling (MHS, Metropolis *et al.* 1953; Hastings 1970) provides an effective sample

generator based on repeatedly perturbing model parameters and accepting perturbations according to the Metropolis criterion, eq. (11), at a fixed temperature $T = 1$. Markov-chain analysis verifies that an unbiased sampling of the PPD is achieved in the limit of a large number of perturbations (Gilks *et al.* 1996). Using MHS as the importance sampler, that is, $g(\mathbf{m}) = P(\mathbf{m}|\mathbf{d})$ in eq. (27), the integral estimate becomes

$$I \approx \frac{1}{Q}\sum_{i=1}^{Q} f(\mathbf{m}_i). \tag{28}$$

Convergence can be monitored by comparing integral estimates from two or more independent MCMC chains collected in parallel: when the difference is sufficiently small, the procedure is considered to have converged and the integral estimated from the union of samples. Since marginal distributions are of primary interest in this paper, the convergence criterion applied here is that the maximum difference between cumulative marginal distributions for all parameters of two chains is less than 0.05 (an almost imperceptible difference if plotted).

In applying the Metropolis criterion, model perturbations can be drawn from any symmetric proposal distribution; the choice of proposal does not affect the integral estimates at convergence but can strongly affect sampling efficiency. In particular, the optimal proposal distribution is given by the PPD itself, but this is obviously unavailable. However, a local linear PPD approximation can be estimated as an M-D Gaussian distribution, eq. (23), where the linearized model covariance matrix is given by eq. (17) for overparametrized problems and by

$$\mathbf{C}_m = s\left[\mathbf{A}^T\mathbf{C}_d'^{-1}\mathbf{A}\right]^{-1} \tag{29}$$

for underparametrized problems, where $\mathbf{A}$ is the Jacobian matrix. To overcome inefficient sampling caused by the correlated parameters, perturbations are applied in a principal-component parameter space where the axes align with the dominant correlation directions (Nolte & Frazer 1994). The orthogonal transformation between physical parameters $\mathbf{m}$ and rotated parameters $\tilde{\mathbf{m}}$ is given by

$$\tilde{\mathbf{m}} = \mathbf{U}^T\mathbf{m}, \quad \mathbf{m} = \mathbf{U}\tilde{\mathbf{m}}, \tag{30}$$

where $\mathbf{U}$ is a column-eigenvector matrix, $\mathbf{C}_m = \mathbf{U}\Lambda\mathbf{U}^T$, and $\Lambda = \text{diag}\,[\lambda_1, \ldots, \lambda_M]$ is the eigenvalue matrix with λ_i representing the variance projected along eigenvector $\mathbf{u}_i$. The linearized PPD also provides effective perturbation magnitudes. For example, the ith rotated parameter can be perturbed using a Gaussian distribution with variance λ_i. However, to achieve some larger perturbations for non-linear problems, a Cauchy distribution, scaled according to (rotated) variances, is used here. An initial (local) estimate for the model covariance matrix is obtained from the linearized solution evaluated at the MAP model. Burn-in sampling is then performed to compute a non-linear estimate of $\mathbf{C}_m$, eq. (7), which better represents the overall covariance structure of the parameter space and is then used in the sampling for integral estimates (burn-in samples are not retained). The above procedure can improve MHS efficiency by orders of magnitude for correlated parameters (Dosso & Wilmut 2008), which was found to be a limiting factor in earlier applications to 1-D MT inversion (Cerv *et al.* 2007).

The numerical integration described above is based on MHS at $T = 1$. However, for strongly non-linear problems involving isolated high-probability regions (multiple modes) in the parameter space, this may not provide sufficient global sampling. Wider sampling can be achieved by applying MHS at $T > 1$ and correcting (weighting) the samples for unbiased integral estimates. Following Brooks & Frazer (2005), eq. (24) is written

$$I = \int f(\mathbf{m})\frac{\exp[-\phi(\mathbf{m})]}{Z_1}\,\mathrm{d}\mathbf{m} = \frac{Z_T}{Z_1}\int f(\mathbf{m})\exp[-\phi(\mathbf{m}) \times (1-1/T)]\frac{\exp[-\phi(\mathbf{m})/T]}{Z_T}\,\mathrm{d}\mathbf{m}, \tag{31}$$

where

$$Z_1 = \int \exp[-\phi(\mathbf{m})]\,\mathrm{d}\mathbf{m}, \qquad Z_T = \int \exp[-\phi(\mathbf{m})/T]\,\mathrm{d}\mathbf{m}. \tag{32}$$

Drawing Q models from $\exp[-\phi(\mathbf{m})/T]/Z_T$ via MHS at temperature T, eq. (31) can be approximated

$$I \approx \sum_{i=1}^{Q} f(\mathbf{m}_i)\exp[-\phi(\mathbf{m}_i)(1-1/T)] \,/\, \sum_{i=1}^{Q}\exp[-\phi(\mathbf{m}_i)(1-1/T)]. \tag{33}$$

For $T = 1$, eq. (33) simplifies to standard MHS, eq. (28), while for $T \to \infty$, eq. (33) becomes uniform Monte Carlo sampling, eq. (26), which represents the widest (but slowest) sampling. In this paper, inversions at a sequence of temperatures $T = 1, 2, 3, \ldots$ are considered with the convergence criterion described [note that the denominator in eq. (33) simply represents normalization and can be neglected when computing marginals]. A change in results due to an increase in T indicates the MCMC chains did not sufficiently wander the parameter space at the lower temperature; no change is consistent with satisfactory sampling. For uni-modal PPDs, the convergence rate typically decreases with increasing temperature due to increased sampling of low-probability regions. However, for multimodal PPDs, the convergence rate can improve substantially due to improved transitions between modes. In other inversion applications (Dosso & Wilmut 2008), high-temperature sampling allowed successful sampling of multimodal PPDs which was not possible at $T = 1$. For inversions in this paper, high-temperature sampling was significantly slower than sampling at unit temperature; however, all runs converged on a single-processor desktop computer in a few hours.

Finally, cases where the variance and/or prior scale factors (s and μ) are unknown can be treated either by applying fixed estimates computed via optimization (an empirical Bayesian approach), or by including them as hyperparameters in the MHS algorithm (hierarchical Bayesian). The latter has the benefit of accounting for the uncertainty in s and/or μ in estimating model parameter uncertainties.

3 LINEARIZED AND NON-LINEAR INVERSION EXAMPLES

3.1 Three-layer test case

This section compares linearized and non-linear inversion for a simple three-layer test case in which the parametrization and data variance are known (unknown parametrization and variance are considered in the following sections). The three-layer model consists of a 0.6-km-thick layer of conductivity 0.001 S m^{-1} embedded in a uniform half-space of 0.01 S m^{-1}. Synthetic data are generated at 11 logarithmically spaced periods from 0.0025 to 25 s, with 2 percent Gaussian errors added. The unknown parameters are the conductivities of the three layers (σ_1, σ_2, σ_3) and the thicknesses of the upper two layers (h_1, h_2), with the third layer consisting of

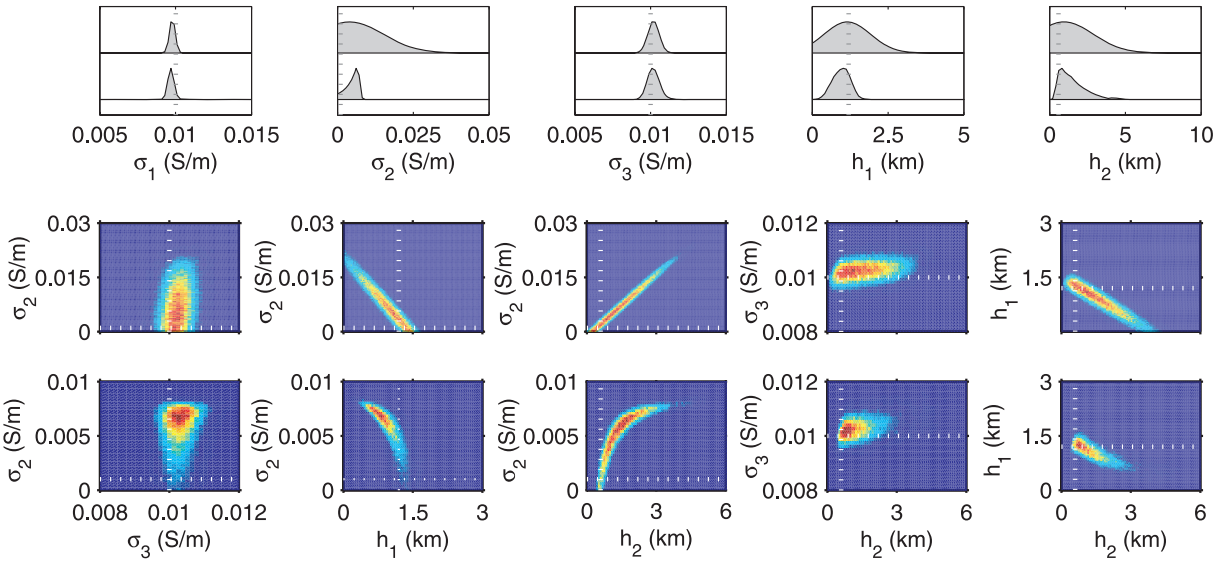

Figure 1. Marginal distributions for the three-layer case. The first row shows 1-D marginals from linearization (upper distributions in each panel) and non-linear inversion (lower distributions). Rows 2 and 3 show selected 2-D marginals from linearization and non-linear inversion, respectively. Dotted lines indicate true parameter values.

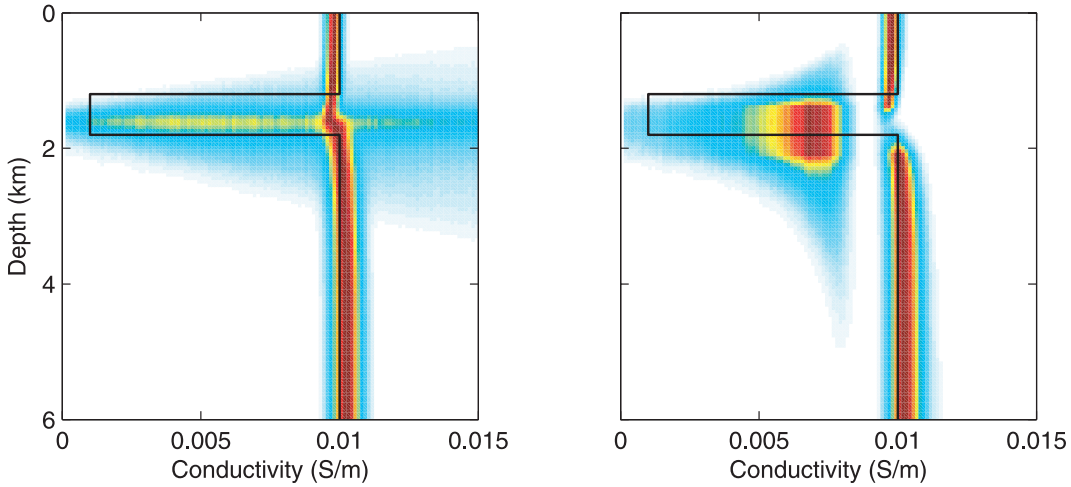

Figure 2. Marginal probability profiles for the three-layer test case from linearization (left-hand side) and non-linear inversion (right-hand side). Solid lines indicate true parameter values.

a half-space. Prior distributions for all parameters are taken to be bounded uniform distributions over $[0, 1]$ S m^{-1} for conductivities and $[0, 10]$ km for thicknesses.

Marginal probability distributions computed via linearized and non-linear inversion are compared in Fig. 1. To obtain linearized marginal distributions which include the bounded uniform priors, a large sample of models is drawn at random from the product of the prior and the multivariate Gaussian PPD defined by eqs (23) and (29). The upper row of Fig. 1 shows that linearized and non-linear 1-D marginal distributions are similar for σ_1 and σ_3. However, for σ_2, h_1 and h_2, the marginals differ significantly, with linearization overestimating the parameter uncertainties relative to the (correct) non-linear uncertainties. Selected joint marginal distributions for linearized and non-linear inversion are compared in rows 2 and 3 of Fig. 1, and illustrate interparameter correlations which are integrated out in 1-D marginals. For instance, within the linearized approximation, the joint marginal distributions show a strong positive correlation between h_2 and σ_2, and strong negative correlations between h_1 and σ_2, and between h_1 and h_2. Note that the negative correlation between h_1 and σ_2 together with the prior lower limit of 0 S m^{-1} for σ_2 results in the non-Gaussian shape of the 1-D linearized marginal distribution for h_1 in Fig. 1. The non-linear analysis shows similar correlations to the linearized analysis; however, significant non-linear (i.e. non-Gaussian) behaviour is indicated for the strongly correlated parameters with correlations that vary over the parameter space producing 'banana-shaped' 2-D marginals. Finally, the inversion results are displayed as marginal probability profiles in Fig. 2. This figure shows that while the low-conductivity layer is not well resolved in general, the layer is resolved much better by non-linear inversion than by linearized inversion.

Table 1. The five-layer model.

Layer	1	2	3	4	5
σ (S m^{-1})	0.004	0.04	0.01	0.1	0.04
h (km)	0.6	1.4	4.0	4.0	∞

3.2 Five-layer test case

A more general and practical test case is based on a five-layer model (Dosso & Oldenburg 1989; Whittall & Oldenburg 1992) with the model parametrization and variance scaling unknown (the true model is given in Table 1 and the true $s = 1$). Synthetic data were generated at 25 logarithmically spaced periods from 0.0025 to 250 s with 2 percent noise added. Underparametrized inversion is considered first. Non-linear optimizations were carried out for

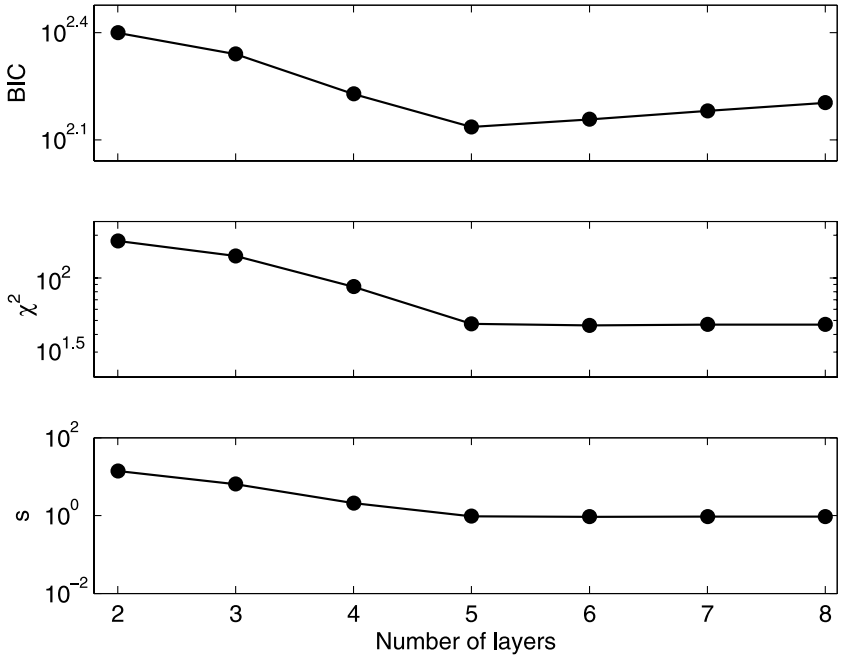

Figure 3. BIC, data misfit and variance factor as a function of the number of layers for the five-layer case.

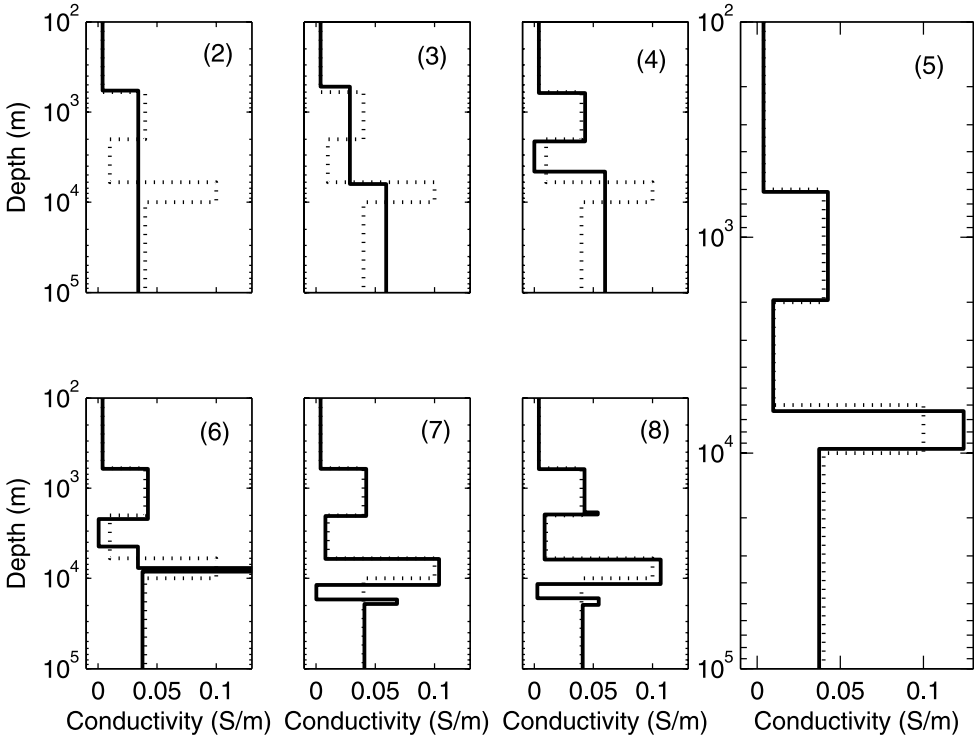

Figure 4. MAP models for different parametrizations of the five-layer case (number of layers indicated on panels). Solid lines indicate inversion solutions; dotted lines indicate the true model.

models with 2–8 layers using uniform priors with bounds of [0, 1] S m^{-1} for conductivities and [100, 10 000] m for thicknesses. The BIC, data misfit and variance factor are shown in Fig. 3 as a function of the number of layers. As the number increases, both χ^2 and s decrease, although the decreases are small beyond five layers. The BIC has a minimum at five layers, indicating this is the preferred parametrization. For five layers, $\chi^2 = 46.8$, which is close to the expected values of 50, and $s = 0.97$, close to the true value of 1. It is also interesting to consider the MAP models for different numbers of layers, as shown in Fig. 4. The two-layer model provides good estimates of the first layer thickness and the conductivities of layers 1 and 2. For the three-layer model, the first layer closely matches the true model, while layers 2 and 3 each average over two layers of the true model. The four-layer model matches the first three layers of the true model quite well, and averages over the two deepest layers. The five-layer model is an excellent approximation to the true model for all layers. When more than five layers are included, small-scale fluctuations are added at large depths.

Marginal probability distributions are shown in Fig. 5. The 1-D marginals compare results for linearized inversion, non-linear inversion with s fixed at its estimated value, and non-linear inversion with unknown s. The marginal distributions for the two non-linear inversions are similar, indicating that accounting for the uncertainty in s does not significantly affect results for the model parameters. The non-linear and linearized marginals are similar for most parameters; however, the linearized distributions for h_4 and σ_4 (defining the relatively-thin high-conductivity layer) are much broader. Fig. 5 also shows selected joint marginals from non-linear inversion with unknown s; non-linear (i.e. non-Gaussian) behaviour is indicated for several parameters. Marginal probability profiles are shown in Fig. 6 for the three inversion approaches considered in Fig. 5. Results are similar for linearized and non-linear inversion except for the thin high-conductivity layer (layer 4), which is poorly resolved via linearized inversion.

For overparametrized inversion, the model was discretized into 50 layers with (fixed) logarithmically increasing thicknesses from 0 to 10^5 m. Fig. 7 compares marginal probability profiles computed via linearized and non-linear inversion with several approaches for treating s and μ. Figs 7(a) and (b) show linearized distributions when true and estimated values of s are used, respectively. In both cases, the final μ value from the iterated linearized inversion is used. Figs 7(c)–(f) show non-linear marginal profiles. In Fig. 7(c), the values for s and μ are the same as in Fig. 7(b) (i.e. fixed estimates from linearized inversion). In Fig. 7(d), s is fixed at the linearized

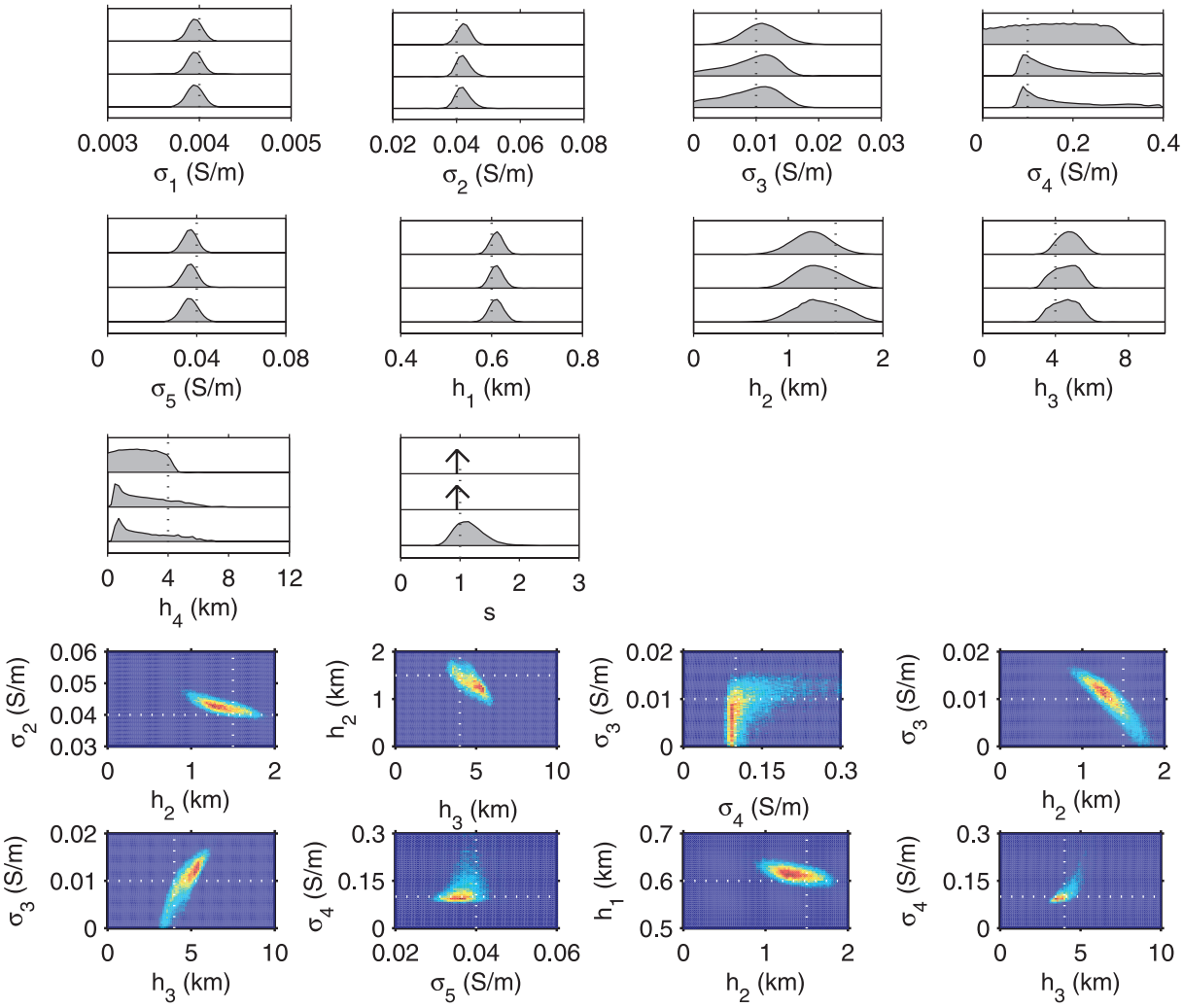

Figure 5. 1-D and 2-D marginal distributions for the five-layer case, underparametrized approach. Upper, middle and lower distributions in each panel of the 1-D marginals show results for linearized inversion with known data variance, non-linear inversion with known variance and non-linear inversion with unknown variance. 2-D marginals are computed via non-linear inversion with unknown variance. Dotted lines indicate true parameter values.

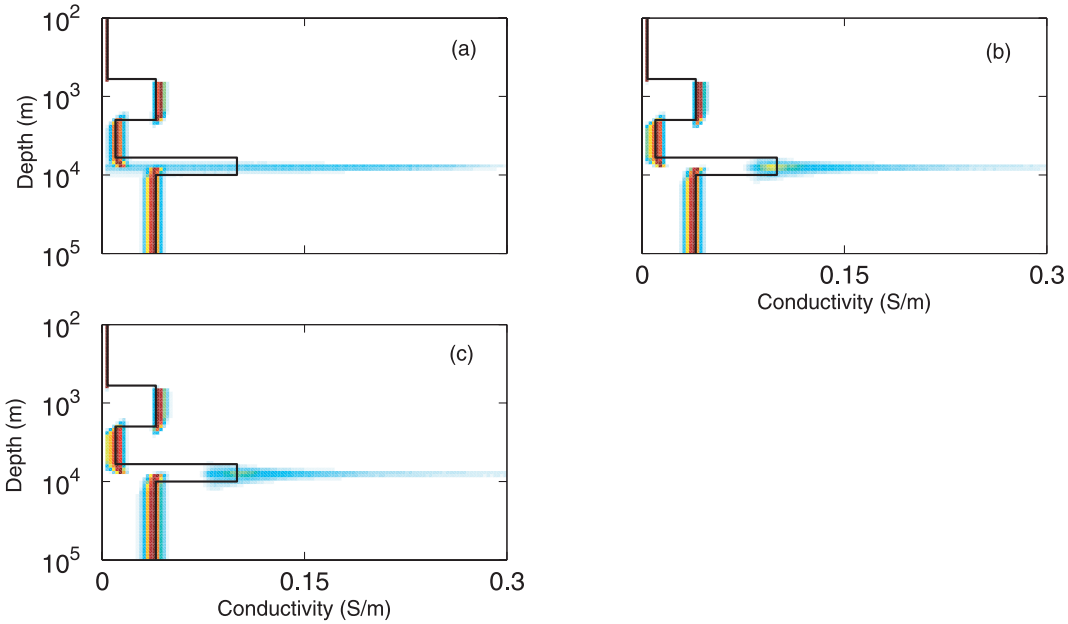

Figure 6. Marginal probability profiles for the five-layer case, underparametrized approach, for (a) linearization, (b) non-linear inversion with known variance and (c) non-linear inversion with unknown variance. Solid lines indicate the true model.

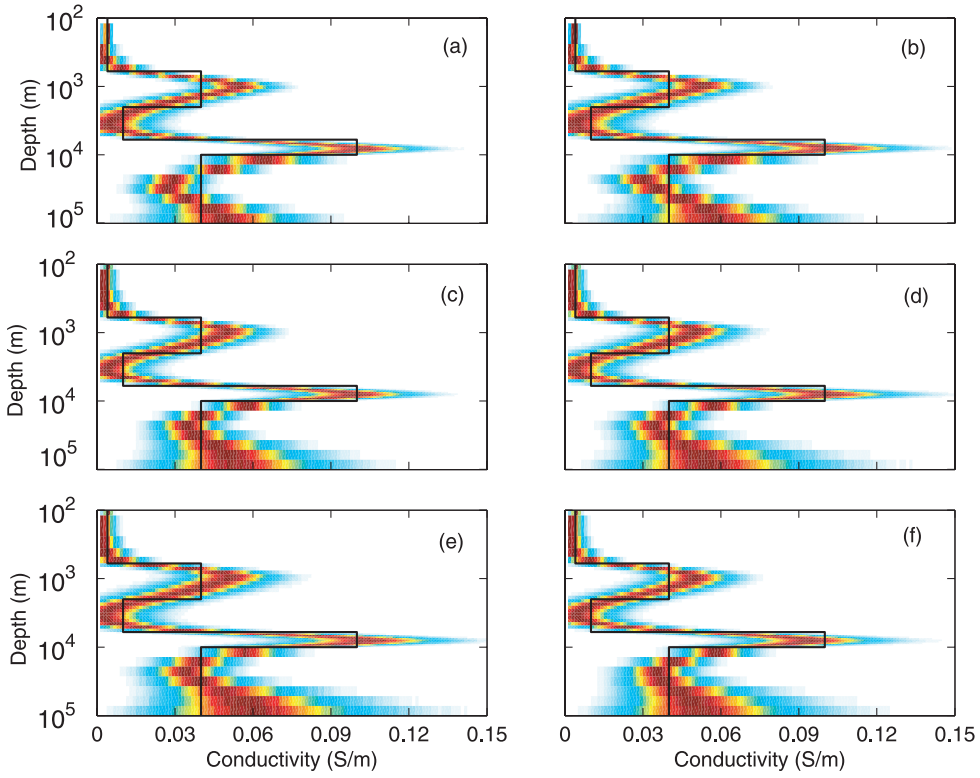

Figure 7. Marginal probability profiles for the five-layer case, overparametrized approach, for (a) linearization with true s, (b) linearization with estimated s, (c) non-linear inversion with μ and s fixed at estimated values, (d) non-linear inversion with s fixed and μ marginalized, (e) non-linear inversion with μ fixed and s marginalized and (f) non-linear inversion with μ and s marginalized. Solid lines indicate the true model.

estimate, but μ is included as a hyperparameter sampled in the inversion. In Fig. 7(e), μ is fixed and s is sampled. Fig. 7(f) shows the most general result: non-linear inversion sampling both μ and s. Overall, the marginal probability profiles in Fig. 7 are generally similar (although not identical). This indicates that non-linearity is not a major issue for this problem, and that accounting for the uncertainty in μ and s by marginalizing over these hyperparameters has only a small effect. Finally, Fig. 8 shows marginal distributions for μ and s from the non-linear inversion in Fig. 7(f). The 1-D marginal for s is reasonably well constrained close to the linearized estimate, and appears to follow a χ^2 distribution, as expected. The 1-D marginal distribution for μ extends over a fairly wide band, which can be explained, in part, by the strong correlation between μ and s, illustrated by the joint marginal in Fig. 8.

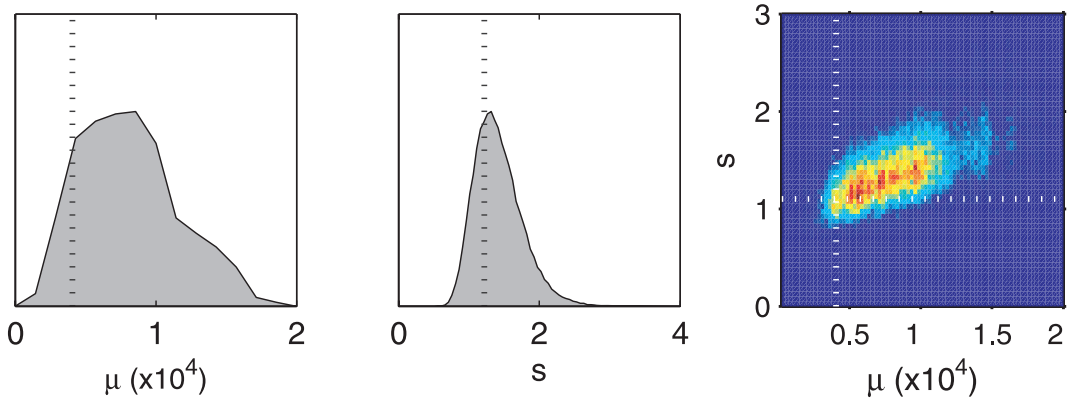

Figure 8. Marginal distribution for μ and s for the five-layer test case, non-linear inversion. Dotted lines indicate estimates from linearized inversion.

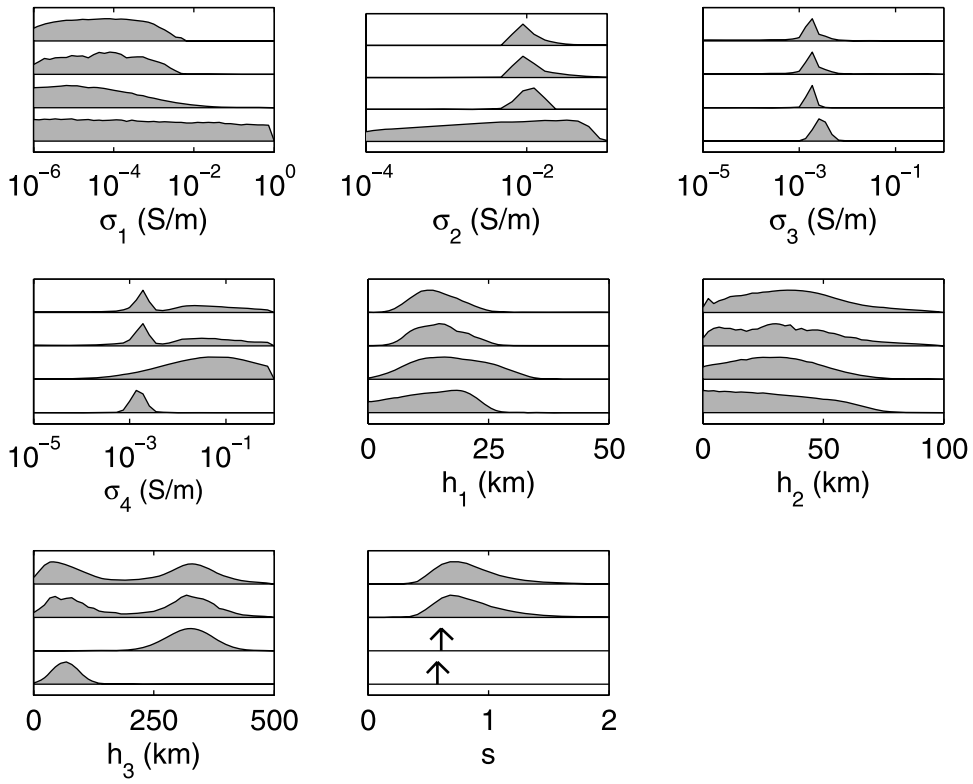

Figure 9. Marginal distributions for the COPROD1 data, underparametrized approach. From the top down in each panel, the first two distributions represent non-linear inversion with sampling temperatures $T = 1$ and 3, respectively, the third distribution represents linearized inversion with base model corresponding to the MAP model for $h_3 > 200$ km, and the fourth distribution represents linearized inversion with base model corresponding to the MAP model for $h_3 < 200$ km.

3.3 COPROD1 data

This section considers the measured COPROD1 MT data (Jones & Hutton 1979a; Jones & Hutton 1979b), which have been used widely as a 1-D inversion example (e.g. Constable *et al.* 1987; Weaver & Agarwal 1993; Grandis *et al.* 1999; Cerv *et al.* 2007). The data are MT responses at 15 periods from 28.5 to 1960.7 s. The standard data errors given by Constable *et al.* (1987) are used here as preliminary estimates, with variance scaling s included in the inversion. The model is taken to be log conductivity to determine structure over several orders of magnitude. In an underparametrized approach, bounded uniform priors are applied over $[10^{-6}, 1]$ S m^{-1} for conductivities and over [1, 1000] km for layer thicknesses. The BIC indicated four layers, as the preferred parametrization (2–8 layers were considered). Marginal probability distributions for the four-layer model are shown in Fig. 9. The two top distributions in each panel show the non-linear inversion results for sampling temperatures of $T = 1$ and 3. The nearly identical results suggest that sufficient sampling of the parameter space is achieved. Of particular interest are the bimodal marginal distributions for h_3 and σ_4 obtained by non-linear inversion. The non-linear 2-D marginal distribution for h_3 and σ_4 in Fig. 10 (top right panel) indicates two distinct regions of high probability in the parameter space. Linearized uncertainty analysis is a local approximation about a given base model and cannot represent a multimodal PPD. To examine linearized solutions for each of the two high-probability regions of the parameter space, 1-D and 2-D marginal distributions are shown

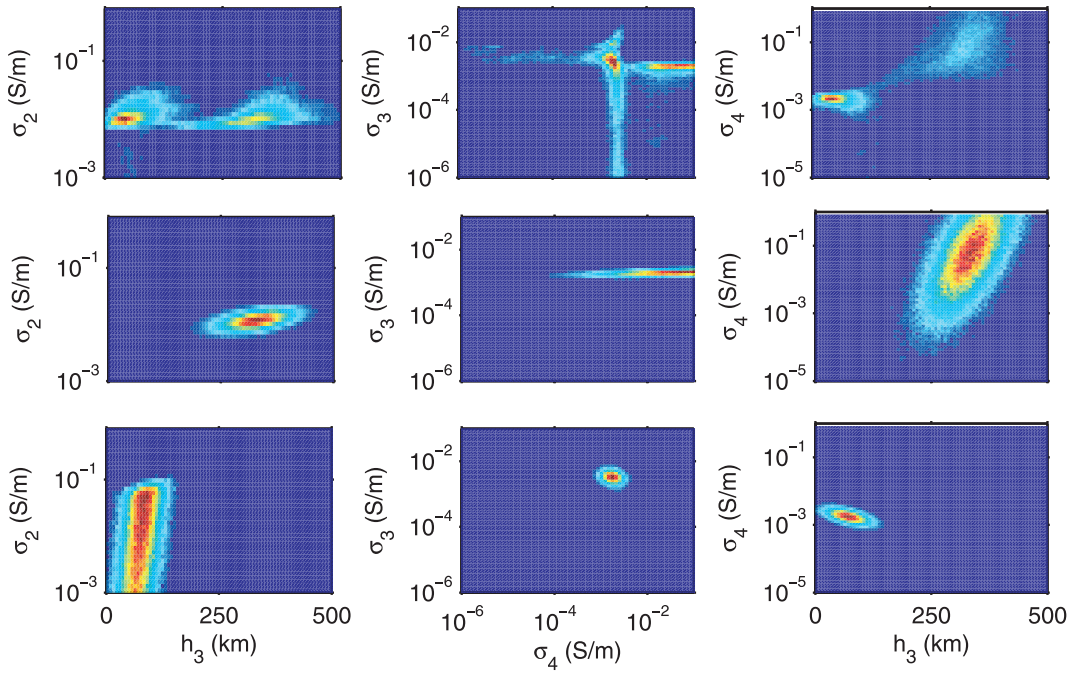

Figure 10. Selected 2-D marginal distributions for the COPROD1 data, underparametrized approach. The first row represents non-linear inversion, the second row represents linearized inversion with base model corresponding to the MAP model for $h_3 > 200$ km and the third row represents linearized inversion with base model corresponding to the MAP model for $h_3 < 200$ km.

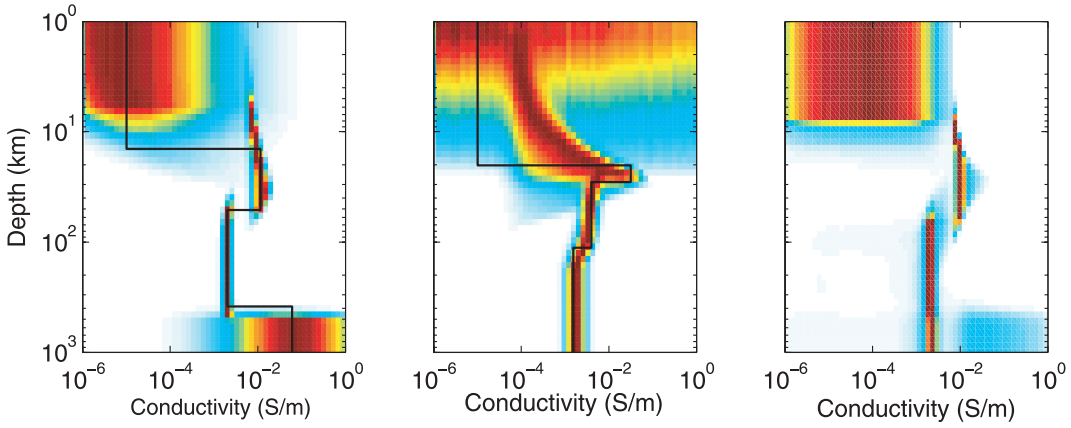

Figure 11. Marginal probability profiles for the COPROD1 data, underparametrized approach. The first panel represents linearized inversion with base model (solid line) corresponding to the MAP model for $h_3 > 200$ km, the second panel represents linearized inversion with base model corresponding to the MAP model for $h_3 < 200$ km and the third panel represents non-linear inversion.

in Figs 9 and 10 based on the minimum-misfit model in each region (regions are defined according to whether h_3 is less or greater than 200 km). Fig. 9 shows that the two linearized marginals for h_3 and for σ_4 each approximate the mode of the corresponding non-linear distribution from which the base model was drawn. Joint marginals for h_3 and σ_4, given in Fig. 10 (right-hand column), show this in 2-D.

Fig. 11 shows marginal probability profiles for the four-layer parametrization. Figs 11(a) and (b) show marginal profiles from linearized analysis about base models from the two high-probability regions, and clearly illustrate how linearized marginals represent local estimates. By comparison, the non-linear marginal profile in Fig. 11(c) provides a global uncertainty analysis, encompassing both high-probably regions at depth. It is interesting to note that previous inversions for the COPROD1 data generally resemble the base model shown in Fig. 11(b), which is drawn from the high-probability region with higher h_3 and σ_4 values. This can be understood since the basin of attraction for this model appears to be larger, given the much wider marginals for the higher σ_4 region in Figs 8 and 10. However, the base model in Fig. 11(a) has a slightly lower misfit, and hence represents the true MAP solution. Further, Cerv *et al.*'s (2007) four-layer Bayesian inversion results are unimodal, indicating the MCMC sampling may have been incomplete.

For the overparametrized approach, the model was discretized into 50 layers with logarithmically increasing thicknesses from 0 to 10^3 km. Fig. 12 compares marginal probability profiles computed via linearized and non-linear inversion. Results are similar, although the linearized uncertainty is slightly underestimated in the top few kilometres and overestimated at intermediated depths compared with the non-linear inversion. At a great depth, the non-linear inversion favours lower conductivities than the linearized inversion.

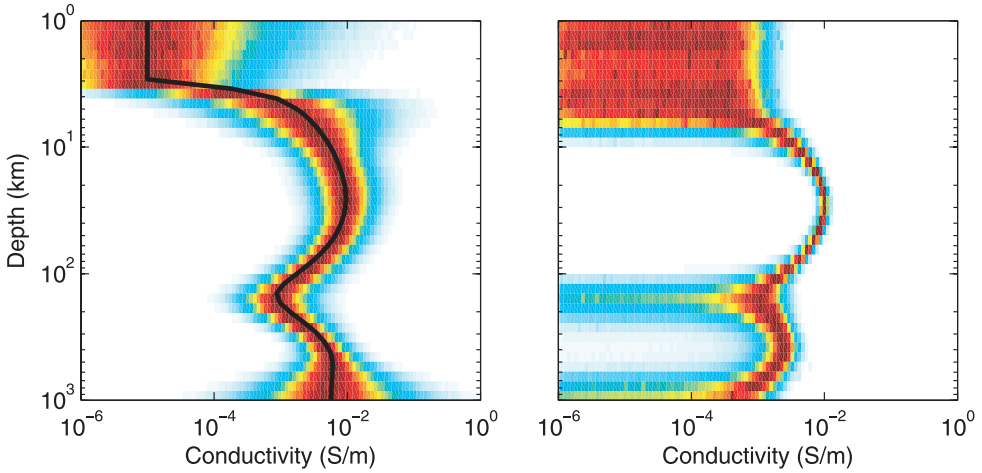

Figure 12. Marginal probability profiles for the COPROD1 data, overparametrized approach, for, from linearization (left-hand side, solid line indicates base model) and non-linear inversion (right-hand side).

4 CONCLUSIONS

This paper applied a Bayesian formulation to compare linearized and non-linear inversion for the 1-D MT inverse problem. In particular, marginal probability distributions were computed for model parameters (conductivities and in some cases layer thicknesses), and combined to produce marginal distributions for conductivity-depth profiles. Linearized uncertainty estimates were based on the analytic approximation of a multivariate Gaussian likelihood function. Bounded uniform priors were imposed on linearized inversions by random sampling from the product of the analytic likelihood and the bounded priors. Non-linear uncertainty estimates were computed numerically using MCMC methods to sample the PPD. In particular, MHS was applied in principal-component space to efficiently sample strongly correlated parameters, an issue which has limited other recent 1-D MT inversions. To ensure global sampling, inversions were carried out at several temperatures, with sample weighting applied to compute unbiased integral estimates. In both linearized and non-linear inversions, the effect of unknown data-variance scaling was considered. Finally, both underparametrized and overparametrized approaches were applied to the inverse problem. The underparametrized approach used the BIC to determine the minimum number of layers consistent with the data, producing uncertainty distributions for discontinuous, blocky models. The overparametrized approach applied prior information on the amount of model structure resolvable by the data, consistent with fitting the data according to the ABIC. This produced uncertainty distributions for continuous, gradient-based models.

The above inversion approaches were applied to a simple three-layer case with the parametrization and variance known, to a more realistic five-layer case with parametrization and variance unknown, and to the measured COPROD1 MT data set. In all cases, similar results were obtained regardless of whether the variance and/or trade-off parameters were held fixed at the true or an estimated value, or included as hyperparameters and marginalized in the inversion. This indicates that the MT data have sufficient information content to constrain both the conductivity model and the hyperparameters. In some cases, the uncertainty distributions obtained via linearized and non-linear inversion were similar; however, important differences were also observed. In the simulated cases, parameters of thin and/or low-conductivity layers were better constrained via non-linear inversion (in the underparametrized approach). Inversion of the COPROD1 data was found to be strongly non-linear, with a bimodal PPD which precludes meaningful uncertainty estimation with linearized approaches.

The non-linearity results for 1-D MT inversion do not transfer directly to 2-D or 3-D inversion. However, this paper raises the issue of investigating linearization errors in MT inversion, regardless of dimensionality, and illustrates a general approach to the problem.

ACKNOWLEDGMENTS

We would like to thank Martyn Unsworth, Malcolm Sambridge and one anonymous reviewer for their comments improving the manuscript significantly. Thanks to Senior Production Editor Janet Marriott for the professional help with great patience during the production process. The work presented in this paper was founded by Chinese Scholarship Council, Central South University (1343-74334002003) and University of Victoria, School of Earth and Ocean Sciences. [Correction added after online publication 2011 April 8: Acknowledgments were added to the article.]

REFERENCES

Adam, A., Novak, A. & Szarka, L., 2007. Basement depths of 3-D basins, estimated from 1-D magnetotelluric inversion, *Acta Geodaetica et Geophysica Hungarica*, **42**, 59–67.

Akaike, H., 1973. Information theory and an extension of the maximum likelihood principle, in *Second International Symposium on Information Theory*, pp. 267–281, eds Petrov, B. & Csaki, B., Academiai Kiado, Budapest.

Becher, W.D. & Sharpe, C.B., 1969. A synthetic approach to magnetotelluric exploration, *Radio Sci.*, **3**, 1089–1094.

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

Cerv, V., Menvielle, M. & Pek, J., 2007. Stochastic interpretation of magnetotelluric data, comparison of methods, *Ann. Geophys.*, **50**, 122–134.

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

Cumming, W. & Mackie, R., 2010. Resistivity imaging of geothermal resources using 1-D, 2-D and 3-D MT inversion and TDEM static shift correction illustrated by a Glass Mountain case history, in *Proceedings World Geothermal Congress 2010*, Bali, Indonesia, 1–10.

Dettmer, J., Dosso, S.E. & Holland, C.W., 2009. Model selection and Bayesian inference for high resolution seabed reflection inversion, *J. acoust. Soc. Am.*, **125**, 706–716.

Dosso, S.E. & Oldenburg, D.W., 1989. Linear and non-linear appraisal using extremal models of bounded variation, *Geophys. J. Int.*, **99**, 483–495.

Dosso, S.E. & Wilmut, M.J., 2008. Uncertainty estimation in simultaneous Bayesian tracking and environmental inversion, *J. acoust. Soc. Am.*, **124**, 82–97.

Dosso, S.E., Wilmut, M.J. & Lapinski, A.L.S., 2001. An adaptive-hybrid algorithm for geoacoustic inversion, *IEEE J. Oceanic Eng.*, **26**, 324–336.

Gilks, W.R., Richardson, S. & Spiegelhalter, G.J., 1996. *Markov Chain Monte Carlo in Practice*, Chapman and Hall, London.

Grandis, H., Menvielle, M. & Roussignol, M., 1999. Bayesian inversion with Markov chains, I. The magnetotelluric one-dimensional case, *Geophys. J. Int.*, **138**, 757–768.

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

Ingber, L., 1989. Very fast simulated annealing, *Math. Comput. Model.*, **12**, 967–973.

Jegen, M.D., Hobbs, R.W., Tarits, P. & Chave, A., 2009. Joint inversion of marine magnetotelluric and gravity data incorporating seismic constraints: preliminary results of sub-basalt imaging off the Faroe Shelf, *Earth planet. Sci. Lett.*, **282**, 47–55.

Jones, A.G. & Hutton, R., 1979a. A multi-station magnetotelluric study in southern Scotland, I., Fieldwork, data analysis and results. *Geophys. J. R. astr. Soc.*, **56**, 329–349.

Jones, A.G. & Hutton, R., 1979b. A multi-station magnetotelluric study in southern Scotland, II. Monte-Carlo inversion of the data and its geophysical and tectonic implications. *Geophys. J. R. astr. Soc.*, **56**, 351–368.

Kass, R.E. & Raftery, A.E., 1995. Bayes factors, *J. Am. Stat. Assoc.*, **90**, 773–795.

Kumar, D., Hoversten, G.M., Gregg Nordquist, G. & Cumming, W., 2010. Role of 1-D MT inversion in a 3-D geothermal field, *SEG Technical Program Expanded Abstracts*, **29**, 1107–1111.

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

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes, *Geophysics*, **69**, 1005–1016.

Metropolis, N., Rosenbluth, A., Rosenbluth, M., Teller, A. & Teller, E., 1953. Equations of state calculations by fast computing machines, *J. Chem. Phys.*, **21**, 1087–1091.

Mitsuhata, Y. & Uchida, T., 2002. 2-D Inversion of frequency-domain EM data caused by a 3-D source, in *Three-Dimensional Electromagnetics*, pp. 153–172, eds Zhdanov, M.S. & Wannamaker, P.E., Elsevier, New York.

Mitsuhata, Y., Uchida, T., Murakami, Y. & Amano, H., 2001. The Fourier transform of controlled-source time-domain electromagnetic data by smooth spectrum inversion, *Geophys. J. Int.*, **144**, 123–135.

Niwas, S., Gupta, P.K. & Gaur, V.K., 2007. Straightforward inversion of MT data using a normalized impedance function, *Geophysics*, **72**, F19–F24.

Nolte, B. & Frazer, L.N., 1994. Vertical seismic profile inversion with genetic algorithms, *Geophys. J. Int.*, **117**, 162–179.

Oldenburg, D.W., 1979. One-dimensional inversion of natural-source magnetotelluric observations, *Geophysics*, **44**, 1218–1244.

Oskooi, B., Pedersen, L.B., Smirnov, M., Arnason, K., Eysteinsson, H. & Manzella, A., 2005. The deep geothermal structure of the Mid-Atlantic Ridge deduced from MT data in SW Iceland, *Phys. Earth planet. Int.*, **150**, 183–195.

Parker, R.L., 1970. The inverse problem of electrical conductivity in the mantle, *Geophys. J. R. astr. Soc.*, **22**, 38–121.

Pedersen, L.B., 2004. Determination of the regularization level of truncated singular-value decomposition inversion: the case of 1-D inversion of MT data, *Geophys. Prospect.*, **52**, 261–270.

Press, W.H., Teukolsky, S.A., Vetterling, W.T. & Flannery, B.P., 1992. *Numerical Recipes in FORTRAN*, 2nd edn, Cambridge Univ. Press, Cambridge.

Rokityansky I.I., 1982. *Geoelectromagnetic Investigation of the Earth's Crust and Mantle*, Springer, New York.

Son, W. & Sun, S., 2005. Magnetotelluric data inversion with seismic data constraint, *Acta Seimologica Sinica*, **18**, 678–685.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.*, **40**, doi:10.1029/2000RG000089.

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

Schwarz, G., 1978. Estimating dimension of a model, *Ann. Statist.*, **6**, 461–464.

Sen, M.K. & Stoffa, P.L., 1995. *Global Optimization Methods in Geophysical Inversion*, Elsevier, Amsterdam.

Smith, J.T. & Booker, J.R., 1988. Magnetotelluric inversion for minimum structure, *Geophysics*, **53**, 1565–1576.

Tarantola, A., 1987. *Inverse Problem Theory: Methods For Data Fitting and Model Parameter Estimation*, Elsevier, Amsterdam.

Tarits, P., Menvielle, J.V. & Roussignol, M., 1994. Bayesian statistics of non-linear inverse problems: example of the magnetotelluric 1-D inverse problem, *Geophys. J. Int.*, **119**, 353–368.

Tikhonov, A.N., 1950. On the determination of the electric characteristics of deep layers of the Earth's crust, *Dokl. Akad. Nauk SSSR*, **151**, 295–297.

Weaver, J.T. & Agarwal, A.K., 1993. Automatic 1-D inversion of magnetotelluric data by the method of modelling, *Geophys. J. Int.*, **112**, 115–123.

Weidelt, P., 1972. The inverse problem of geomagnetic induction, *Z. Geophys.*, **38**, 257–289.

Whittall, K.P. & Oldenburg, D.W., 1992. Inversion of magnetotelluric data for a one-dimensional conductivity, *Geophys. Monogr. Ser. 5*, ed. Fitterman, D.V., Soc. Exp. Geophysicists, Tulsa, OK.