Geophysical Journal International

Geophys. J. Int. (2021) **226,** 548–563
Advance Access publication 2021 March 25
GJI General Geophysical Methods

doi: 10.1093/gji/ggab110

Two-dimensional Bayesian inversion of magnetotelluric data using trans-dimensional Gaussian processes

Daniel Blatter [1,*] Anandaroop Ray [2] and Kerry Key [1]

[1] *Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY,* 10964*, USA. E-mail: daniel.blatter@columbia.edu*
[2] *Geoscience Australia, Symonston, Australian Capital Territory,* 2609*, Australia*

Accepted 2021 March 22. Received 2021 January 8; in original form 2020 September 8

SUMMARY

Bayesian inversion of electromagnetic data produces crucial uncertainty information on inferred subsurface resistivity. Due to their high computational cost, however, Bayesian inverse methods have largely been restricted to computationally expedient 1-D resistivity models. In this study, we successfully demonstrate, for the first time, a fully 2-D, trans-dimensional Bayesian inversion of magnetotelluric (MT) data. We render this problem tractable from a computational standpoint by using a stochastic interpolation algorithm known as a Gaussian process (GP) to achieve a parsimonious parametrization of the model vis-a-vis the dense parameter grids used in numerical forward modelling codes. The GP links a trans-dimensional, parallel tempered Markov chain Monte Carlo sampler, which explores the parsimonious model space, to MARE2DEM, an adaptive finite element forward solver. MARE2DEM computes the model response using a dense parameter mesh with resistivity assigned via the GP model. We demonstrate the new trans-dimensional GP sampler by inverting both synthetic and field MT data for 2-D models of electrical resistivity, with the field data example converging within 10 d on 148 cores, a non-negligible but tractable computational cost. For a field data inversion, our algorithm achieves a parameter reduction of over 32× compared to the fixed parameter grid used for the MARE2DEM regularized inversion. Resistivity probability distributions computed from the ensemble of models produced by the inversion yield credible intervals and interquartile plots that quantitatively show the non-linear 2-D uncertainty in model structure. This uncertainty could then be propagated to other physical properties that impact resistivity including bulk composition, porosity and pore-fluid content.

Key words: Electrical properties; Magnetotellurics; Non-linear electromagnetics; Inverse theory; Probability distributions.

1 INTRODUCTION

Geophysical inversion aims to extract information about an earth model from field observations, where the earth model is a quantitative, discretized description of the spatial (and possibly temporal) distribution of physical properties of interest. For example, this might be a 1-D model with layers that contain a constant scalar property (e.g. isotropic electrical resistivity), or a 2-D or 3-D discretized model with a number of scalar or vector properties defined on a grid sufficiently fine to capture the desired degree of model complexity. The field observations are often made at the surface, yet the physical properties of interest (e.g. electrical resistivity, seismic velocity) are at depth. Deterministic inversion methods attempt to produce a single model estimate by combining data fit with *a priori* assumptions about the model in the form of model regularization, which also serve to stabilize the non-unique and usually ill-posed inverse problem (e.g. Constable *et al.* 1987; Calvetti & Somersalo 2018; Fournier & Oldenburg 2019; Newman & Alumbaugh 2000). Bayesian sampling-based inverse methods instead produce an ensemble of physical models, each of which fit the data and from which statistical information about model parameters can be inferred (Mosegaard & Tarantola 1995). As a result, these methods provide crucial information about inverted model parameter uncertainty, which is necessary to place statistically informed constraints on related physical properties of interest or their structure (de Pasquale & Linde 2017; Blatter *et al.* 2019). In addition, Bayesian methods are better able to address strong non-linearity since they do not rely upon linearization (Dettmer & Dosso 2013).

[*] Now at Institute of Geophysics and Planetary Physics, Scripps Institution of Oceanography, University of California, San Diego, La Jolla, CA, 92037, USA.

This additional information comes with an added computational cost relative to deterministic inversion methods. Deterministic linearized inversion methods typically proceed by iteratively taking small steps down the local gradient of the objective function, typically requiring tens to hundreds of forward calculations to find a good fitting model at a minimum of the objective function. Conversely, Bayesian sampling methods can require 10^5–10^6 forward computations to converge to the posterior distribution (Bodin & Sambridge 2009; Agostinetti & Bodin 2018). This cost rises as the spatial dimensions of the model increase from 1-D to 2-D to 3-D for two main reasons. First, calculating the model response (forward modelling) usually becomes significantly more expensive computationally as the model dimensionality increases. Second, the number of model parameters that must be sufficiently sampled increases exponentially since 2-D and 3-D models can represent more complicated structures and typically have many more free parameters than 1-D models. In order to extend Bayesian inversion to 2-D and 3-D models, the cost of forward modelling must be reduced so that the entire inverse problem can be solved in a practical amount of computational time; here we define practical to mean solvable on a time scale of up to a few weeks compute time. This can be achieved via high performance computing (HPC) and parallel computational architectures that permit efficient parallel programming constructs.

Even when using HPC methods to speed up forward modelling, to make Bayesian inversion tractable in 2-D and 3-D the number of parameters the Bayesian algorithm samples over (the dimension of the model space) must remain small enough to allow for sampling to converge within practical computing times. Neal (2011) provides a succinct example of the difficulties faced in sampling high dimensional distributions, especially if each dimension is independent of the others. The forward model response must be computed numerically, requiring a discrete representation of the model. Usually, this is accomplished through a mesh or grid of scalar- or vector-valued cells. This mesh must be fine enough to capture the model complexity resolvable by the data, often necessitating thousands of independent model parameters. Not all of these model parameters are truly independent, however. The model structure that can be resolved by diffusive methods such as low-frequency electromagnetics (EM), for instance, is typically lower-wavenumber than the density of the mesh required to accurately represent it. For instance, accurately representing a simple, sinusoidal variation in subsurface properties might require many thousands of discrete grid cells (Commer & Newman 2009). This is evident even when sharp, piecewise-constant parametrizations are used such as in Minsley (2011), Ray & Key (2012), Brodie & Sambridge (2012) and Ley-Cooper (2016).

That the model structure resolvable by the data is lower-wavenumber than the discrete mesh grid necessary to describe it implies that it should be possible to discover a parsimonious representation of the physical model that can still capture the necessary degree of model complexity. This low-dimension model space can be searched efficiently by the Bayesian sampler. Further, a transform from the parsimonious to the dense model parametrization can be used to connect the Bayesian sampler to the forward solver to allow the data misfit of each parsimonious model to be evaluated.

Trans-dimensional Markov chain Monte Carlo (MCMC) algorithms (Green 1995; Malinverno 2002; Bodin & Sambridge 2009; Ray *et al.* 2016), in which the number of model parameters is selected by the data during the inversion and is allowed to freely vary, are ideally suited to this task. Because of the parsimony inherent in Bayes' rule (Malinverno & Leaney 2000; Malinverno 2002; MacKay 2003; Schoniger *et al.* 2015), trans-dimensional Bayesian sampling algorithms prefer models with fewer parameters. This means that they are naturally able to discover a parsimonious parametrization that fits the observations. Thus, the user need not know beforehand the degree of simplification vis-a-vis the dense parameter grid that is appropriate for the data.

Multiple trans-dimensional approaches to reducing model space dimensionality have been proposed. Layers work well in 1-D (Malinverno & Leaney 2000; Minsley 2011; Dettmer *et al.* 2015; Blatter *et al.* 2018), while Voronoi cells have been successfully implemented in 2-D (Bodin & Sambridge 2009; Dettmer *et al.* 2014; Ray *et al.* 2014; Burdick *et al.* 2018; Galetti & Curtis 2018), although their sharp edges are not necessarily ideal for inverting some types of data, including diffusive EM techniques such as the magnetotelluric (MT) method (Hawkins *et al.* 2019). In 3-D, however, the computational geometry of Voronoi cells becomes complex and expensive (Agostinetti *et al.* 2015), though this is an active area of research (Zhang *et al.* 2018). Hawkins & Sambridge (2015) developed a tree structure for trans-dimensional inversion using wavelets as basis functions. This approach is flexible, dimension agnostic, and efficient. However, it suffers from the need to specify prior distributions on wavelet or other basis function coefficients. Because these are not physical properties, it is difficult to select appropriate prior distributions for them.

To achieve a parsimonious model representation that allows specifying priors on physical properties, we implement the trans-dimensional Gaussian process (TDGP) algorithm of Ray & Myer (2019) that utilizes a stochastic interpolation algorithm known as a Gaussian process (GP, e.g. Williams & Rasmussen 1996). We utilize HPC, including parallel tempering (Sambridge 2013; Dettmer & Dosso 2012), to reduce the time required to sufficiently sample the model space.

We apply our algorithm to both synthetic and field MT data. The TDGP sampler successfully produces an ensemble of models that fit the data. In our example application with synthetic data, the posterior model ensemble recovers the features of the true model, while in the field data case the ensemble mean includes features that are strikingly similar to those in a regularized, gradient-based inverse solution. In addition, however, our algorithm produces posterior model parameter uncertainties that do not rely on linearization. In these examples, the TDGP algorithm achieves a parameter space reduction of more than 4× for the lightly parametrized synthetic problem, and more than 32× for the more densely parametrized field data problem.

The computational cost is dominated by the cost of the forward problem and is non-negligible, yet well within the limitations of modern computer power. At 0.3 s per forward computation, the synthetic inversion took 3.5 d for convergence to the posterior distribution when run on an HPC system using 480 processing cores. The field data inversion, meanwhile, took just under 10 d at 0.85 s per forward computation when run on 168 cores. The synthetic data contained more frequencies and hence was able to benefit from a greater number of cores. Though 2-D and 3-D probabilistic inversions of MT data have been carried out by Chen *et al.* (2012) and Rosas-Carbajal *et al.* (2013), our inversions represent, to our knowledge, the first trans-dimensional 2-D Bayesian inversions of MT data.

2 METHODS

2.1 MT method

The MT method uses natural variations in the Earth's magnetic field to probe subsurface resistivity structure (Cagniard 1953; Tikhonov 1950). The frequencies at which the natural field varies, in conjunction with the subsurface resistivity, determine the depth to which MT data are sensitive. Higher frequencies are sensitive to shallower portions of the model, while lower frequencies (longer periods) sense deeper structure. In addition, the more resistive the subsurface is, the deeper MT soundings can penetrate, while a more conductive subsurface limits depth sensitivity. See Chave & Jones (2012) for a recent MT review that spans theory, data processing, modelling and interpretation techniques.

The MT method is governed by a diffusive differential equation, which means that MT data are an integrated quantity. While MT data can be *sensitive* to sharp structural gradients, the structures MT data are able to resolve will be smoothed, integrated versions of the subsurface geology. For MT examples relevant to this study, see Key *et al.* (2006) and Naif *et al.* (2013), who image a salt body in the Gulf of Mexico and a melt channel at the base of the lithosphere offshore Nicaragua, respectively.

2.2 Bayesian inversion

The posterior probability distribution that Bayesian methods seek to sample from is defined by Bayes' Rule

$$p(\boldsymbol{m}|\boldsymbol{d}) \propto p(\boldsymbol{d}|\boldsymbol{m})p(\boldsymbol{m}), \tag{1}$$

where $\boldsymbol{m}$ is a vector of parameters that constitute our model of the Earth, $\boldsymbol{d}$ is the vector of observed data, $p(\boldsymbol{d}|\boldsymbol{m})$ is the likelihood, $p(\boldsymbol{m})$ is the prior and $p(\boldsymbol{m}|\boldsymbol{d})$ is the posterior. In order for eq. (1) to represent the posterior, it must be normalized by the probability of the data. In any relative inference formulation, including trans-D, however, this quantity does not need to be calculated explicitly. The prior distribution represents *a priori* information, assumptions, and beliefs about the model, independent of field data. The likelihood is the probability that the difference between the observed data, $\boldsymbol{d}$, and the forward modelled data, $\boldsymbol{f}(\boldsymbol{m})$, is due purely to random error. Therefore, models that fit both the data and our prior assumptions on the model have a higher posterior probability.

We assume the data noise to be Gaussian and define the likelihood as follows

$$p(\boldsymbol{d}|\boldsymbol{m}) = \frac{1}{\sqrt{|2\pi \boldsymbol{C}_d|}} \times exp\left(-\frac{1}{2}[\boldsymbol{f}(\boldsymbol{m}) - \boldsymbol{d}]^T \boldsymbol{C}_d{}^{-1}[\boldsymbol{f}(\boldsymbol{m}) - \boldsymbol{d}]\right), \tag{2}$$

where $\boldsymbol{C}_d$ is the data error covariance matrix. The prior distribution, which depends on the number of model parameters, k, and on the details of how the model is parametrized, will be discussed in the next section following Ray & Myer (2019).

2.3 TDGP Markov chain Monte Carlo

The MCMC method is an iterative algorithm that draws samples from the posterior distribution by means of a guided random walk through the model space. At each step, a proposal model $\boldsymbol{m}'$ is generated from a proposal distribution $q(\boldsymbol{m}'|\boldsymbol{m})$, and is then accepted or rejected by means of an acceptance probability $\alpha(\boldsymbol{m}'|\boldsymbol{m})$. α depends on the posterior probability of $\boldsymbol{m}'$ and $\boldsymbol{m}$, as well as q and the number of model parameters in $\boldsymbol{m}'$ and $\boldsymbol{m}$, which can change at each iteration. The acceptance probability guides the Markov chain towards the posterior distribution. The proposal is selected by the user, and needs to be carefully chosen as it can affect the rate of convergence (the number of iterations before the algorithm has sufficiently sampled the model space). In general, a proposal that closely resembles the posterior can accelerate convergence (Mosegaard 2019).

We implement here a trans-dimensional variant of Metropolis–Hastings–Green MCMC (Metropolis *et al.* 1953; Hastings 1970; Green 1995), in which the number of model parameters is itself a parameter that the data select for. A detailed description of the algorithm is found in Ray & Myer (2019). Our algorithm implements a 'birth-death' scheme (Geyer & Moller 1994) that allows the number of model parameters to increase, decrease, or remain unchanged at each step of the MCMC algorithm. The trans-dimensional nature of the algorithm is reflected in the acceptance probability

$$\alpha(\boldsymbol{m}'|\boldsymbol{m}) = \min\left[1, \quad \left(\frac{p(\boldsymbol{d}|\boldsymbol{m}')}{p(\boldsymbol{d}|\boldsymbol{m})}\right)^{1/T} \frac{p(\boldsymbol{m}')}{p(\boldsymbol{m})} \frac{q(\boldsymbol{m}|\boldsymbol{m}')}{q(\boldsymbol{m}'|\boldsymbol{m})} \times |J|\right], \tag{3}$$

where T is an annealing temperature as part of our implementation of parallel tempering (see Ray *et al.* 2013; Blatter *et al.* 2018, for details). Only samples from $T = 1$ chains are used as part of the model ensemble. The matrix J is the Jacobian of the jump from $\mathbf{m}$ to $\mathbf{m}'$. For the types of trans-dimensional jumps used in this paper, $|J| = 1$ (Agostinetti & Malinverno 2010).

In order to converge to the posterior probability distribution, trans-dimensional MCMC must adequately sample the parameter space. This means that the time to convergence can grow rapidly with the number of model parameters. There is a natural parsimony built into Bayes' rule, meaning that for models that fit the data equally well, Bayesian inversion prefers those with fewer parameters. In this vein, trans-D Bayesian algorithms also attempt to solve the problem of model selection, ensuring there is sufficient model complexity to fit the data. As a result, if the model space is not appropriately parametrized, Bayesian parsimony may not be enough to prevent inversions for models with two or more spatial dimensions from becoming prohibitively expensive computationally.

Here we utilize a TDGP MCMC sampler to achieve a parsimonious model parametrization (Ray & Myer 2019). A GP is the interpolation of a Gaussian random field using a spatial autocorrelation function, and allows inference to be made on the probable value of an arbitrary continuous function at locations where the value of the function is unknown. These inferences are made on the basis of knowledge of the function at a finite number of locations, in conjunction with a kernel function that specifies how the ensemble of function values correlate spatially. GPs view functions as realizations from spatially-correlated, multivariate Gaussian distributions, where the number of variables is in principle infinite. This is mathematically equivalent to the concept of Kriging, which has been applied in mining and reservoir modelling (Krige 1952; Pyrcz & Deutsch 2014, for an overview), as well as least squares collocation in geodesy (Moritz 1980). A Bayesian treatment of this kind of interpolation can be found in Tarantola & Valette (1982). In practice, since we implement GPs on computers, we know the value of the function at a finite number of points $\boldsymbol{x}$, which we here refer to (for reasons that will become clear later) as 'interpolation nodes', and we seek the probable value of the function at another finite set of points $\boldsymbol{x}^*$, which we will call 'gridpoints'.

The spatial correlation is key to defining a GP, and is specified by a kernel function, which can take any number of forms. Here we

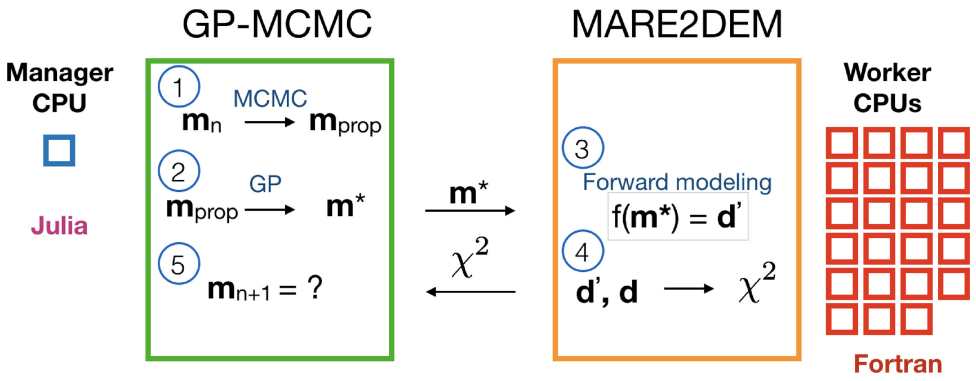

Figure 1. Description of the high performance computing implementation and workflow of the trans-D MCMC algorithm. Loading the data and the dense finite element parameter mesh is handled in MARE2DEM. At iteration n, Birth-death MCMC (step 1), the parsimonious to dense GP transformation (step 2), and the acceptance/rejection of the proposed next model (step 5) are performed on the parsimonious model, $\boldsymbol{m}$, which provides the GP mean through eq. (9), on the manager CPU of the compute node, in Julia. Forward modelling in MARE2DEM (step 3) and misfit calculation (step 4) are performed on the dense model ($\boldsymbol{m}^*$) on the worker CPUs, in Fortran. Communication between Julia and Fortran is handled through a direct Julia-to-Fortran interface subroutine. The numbers 1–5 represent the chronological order of operations at each iteration, as described in the text.

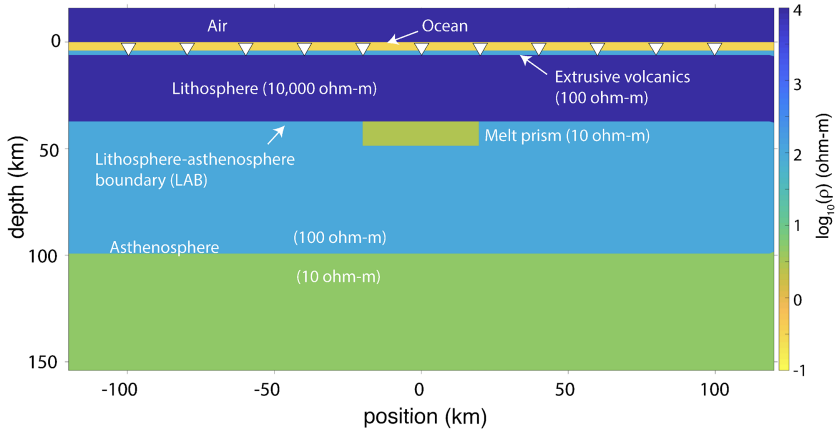

Figure 2. Resistivity model from which synthetic data were generated. The model consists of shallow, extrusive volcanics overlying a resistive lithosphere and conductive asthenosphere. A conductive anomaly at the base of the lithosphere represents a partial melt prism. Synthetic data were generated at 11 MT sites (white triangles) over a period range of 10–10 000 s. The model grid $\boldsymbol{x}^*$ consists of 818 independent resistivity cells.

use the following:

$$\boldsymbol{K}(\boldsymbol{y}, \boldsymbol{y}') = \exp\left(-\frac{1}{2}[\boldsymbol{y} - \boldsymbol{y}']^T \boldsymbol{C}_\lambda{}^{-1}[\boldsymbol{y} - \boldsymbol{y}']\right), \quad (4)$$

where $\boldsymbol{y}$ and $\boldsymbol{y}'$ are points in $\mathbb{R}^{n_d}$, where n_d is the number of spatial dimensions. In this context, $\boldsymbol{y}$ and $\boldsymbol{y}'$ are general and either can represent $\boldsymbol{x}$ or $\boldsymbol{x}^*$. This kernel defines an exponential spatial correlation and is appropriate for inverting data with a smooth resolution kernel, such as low-frequency MT data. For sharper transitions the Matern 3/2 or 5/2 kernels may also be used, but are not investigated here (Williams & Rasmussen 1996).

The matrix $\boldsymbol{C}_\lambda$ is diagonal with n_d entries that specify the spatial length scale over which the function of interest is correlated. These correlation length scales must be selected by the user and represent some of the key adjustable parameters in our TDGP approach. Their value should be at least as small as the size of the smallest features resolvable by the data. Correlation length scales smaller than this will not improve model resolution but may result in additional time to convergence, since more interpolation nodes may be required to fit the data. Correlation length scales larger than this will fail to adequately capture the degree of model complexity resolvable by the data. Since it is not always possible to know precisely what the maximum resolution of the data is, some care should be taken in the choice of correlation length scales.

A related issue is the need for an explicit, discrete parameter grid. While the GP representation of the model is in principle continuous (it defines the model at an arbitrary, infinite number of locations), the need to solve the 2-D or 3-D forward problem numerically on computers, (using for example, the finite element or finite difference methods), requires a discrete model representation. This usually takes the form of a grid or mesh of cells, each with a constant physical property value or vector of values, with each cell independent of the others. This mesh must be designed with the model complexity resolvable by the data in mind. A parameter grid that is too coarse to adequately represent the model structures that the data can resolve will be unable to achieve an acceptable fit to the data. For most realistic synthetic models and for field data, this usually requires a dense mesh with thousands or even millions of cells. The design of an adequate parameter mesh is problem-specific and can be guided by gradient-based inversion, repeatedly refining the mesh until the data fit is adequate.

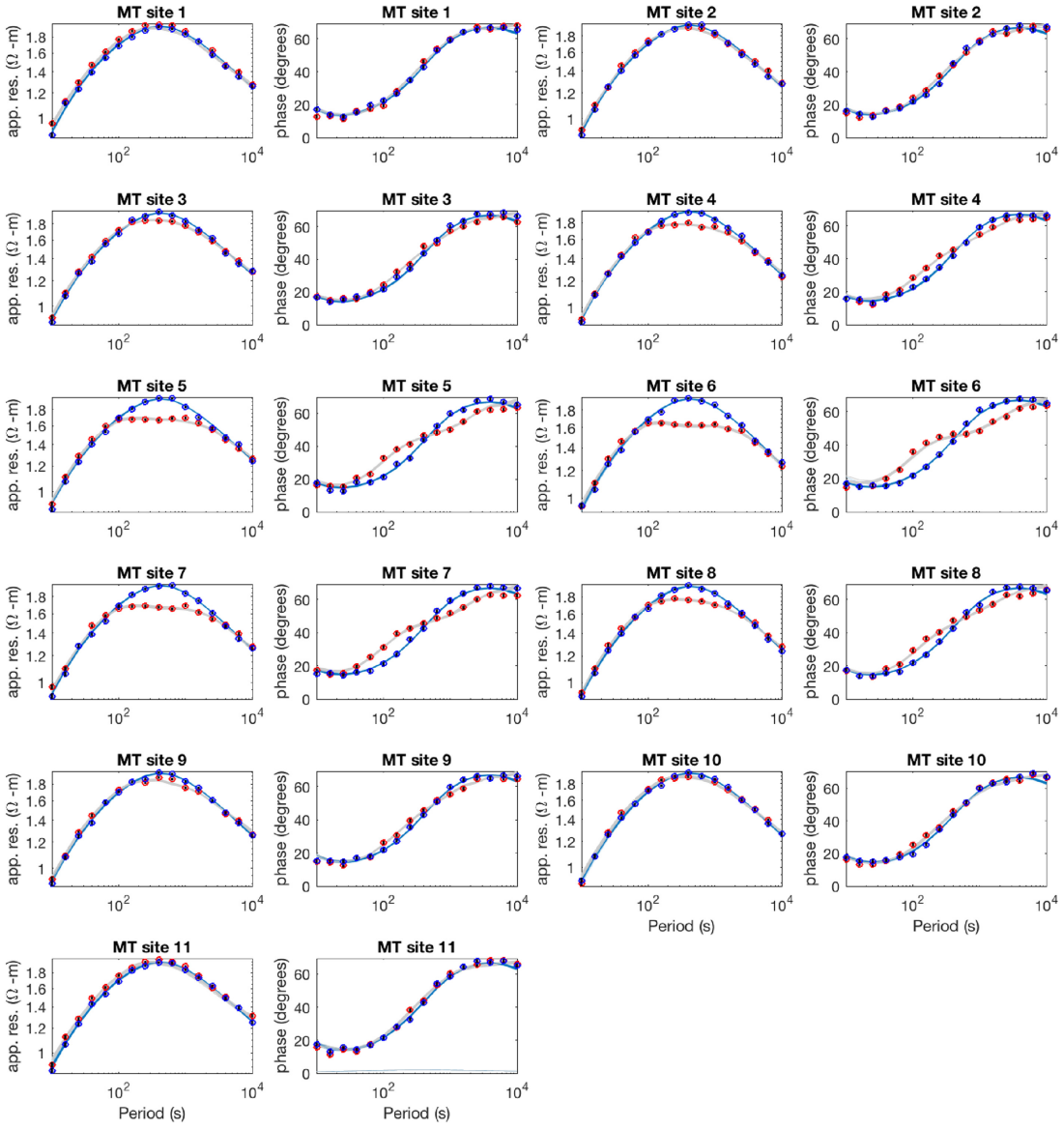

Figure 3. Synthetic TE (red circles) and TM (blue circles) mode apparent resistivity and phase for 11 MT sites (see white triangles in Fig. 2). In addition, the forward responses for 50 randomly selected models from the posterior ensemble are plotted in grey (TE mode) and blue (TM mode). The distribution of the RMS fit to the data across the ensemble had a mode of 1.25.

Once this mesh is defined, we let the locations of the mesh cells be the gridpoints, $\boldsymbol{x}^*$, enabling the GP to connect the parsimonious model parametrization defined at the interpolation nodes, $\boldsymbol{x}$, to the denser, computational model. A detailed description follows:

Given a kernel function $\boldsymbol{K}$, a GP can be specified as

$$\begin{bmatrix} \boldsymbol{g}(\boldsymbol{x}) \\ \boldsymbol{g}(\boldsymbol{x}^*) \end{bmatrix} = N\left(\boldsymbol{0}, \begin{bmatrix} \boldsymbol{K_m} & \boldsymbol{K_*} \\ \boldsymbol{K_*}^T & \boldsymbol{K_{**}} \end{bmatrix}\right), \tag{5}$$

where the posterior covariance matrix is made up of sub-matrices defined as follows:

$$\boldsymbol{K_m} = \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}) + {\boldsymbol{\sigma_m}}^2 \tag{6}$$

$$\boldsymbol{K_*} = \boldsymbol{K}(\boldsymbol{x}, \boldsymbol{x}^*) \tag{7}$$

$$\boldsymbol{K_{**}} = \boldsymbol{K}(\boldsymbol{x}^*, \boldsymbol{x}^*). \tag{8}$$

Here, $\boldsymbol{x} \in \mathbb{R}^{n_{interp}}$ and $\boldsymbol{x}^* \in \mathbb{R}^{n_{grid}}$, where n_{interp} is the number of interpolation nodes and n_{grid} is the number of gridpoints. The sub-matrices $\boldsymbol{K_m}$, $\boldsymbol{K_*}$ and $\boldsymbol{K_{**}}$ have sizes $(n_{\text{interp}} \times n_{\text{interp}})$, $(n_{\text{interp}} \times n_{\text{grid}})$, and $(n_{\text{grid}} \times n_{\text{grid}})$, respectively. For example, suppose we have a 2-D earth model with 1000 model parameters in the dense computational model, and 50 interpolation nodes in the parsimonious parametrization. Then $\boldsymbol{x}$ has length 50, $\boldsymbol{x}^*$ has length 1000, and the full posterior covariance matrix $\boldsymbol{K}$ has size 1050×1050. ${\boldsymbol{\sigma_m}}^2$ is a diagonal covariance matrix (or 'nugget', for details see Pyrcz & Deutsch 2014) that specifies how closely the GP interpolation must adhere to the function value at the interpolation nodes, $\boldsymbol{x}$.

In essence, via eqs (5)–(8) a GP defines the values of the function at the gridpoints $\boldsymbol{g}(\boldsymbol{x}^*)$ as conditional Gaussian realizations—conditional on the value of the function at the interpolation nodes $\boldsymbol{g}(\boldsymbol{x})$. The relationship between them is determined by the chosen kernel function (eq. (4)), such that nearby points will be similar in value while distant points will be independent of one another.

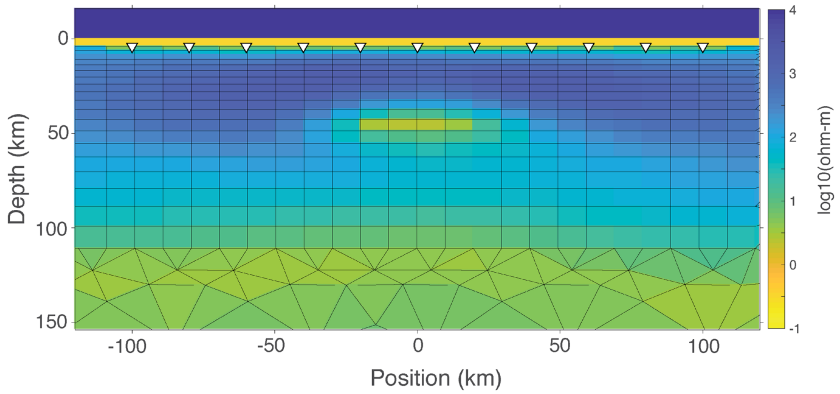

Figure 4. Smooth resistivity model obtained from regularized inversion of the synthetic MT data using MARE2DEM. The smooth model indicates the scale of features resolvable by the data and can assist in selecting a resistivity mesh (outlined in black) sufficiently dense to capture the model complexity, as well as appropriate choices for the GP correlation length scales. Smooth versions of the main features of the melt prism model are present in the regularized estimate, including the extrusive volcanics, lithosphere, melt prism, and asthenosphere. The LAB is not sharply resolved.

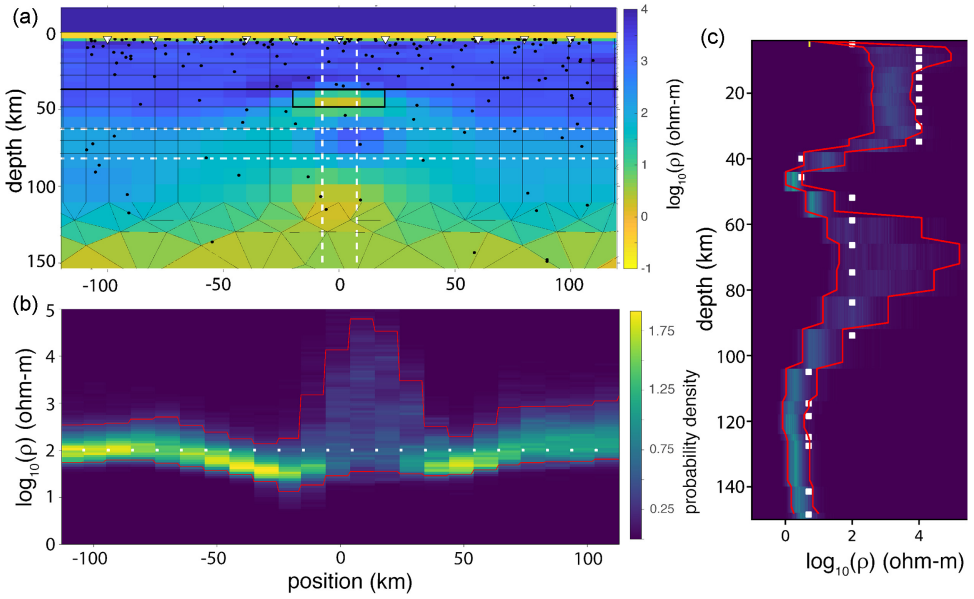

Figure 5. Synthetic inversion results. (a) The mean of the model ensemble captures the conductive melt prism (black box). The LAB (thin black line) is less well resolved. The resistivity mesh (outlined in black) and the interpolation node locations (black dots) for a model selected at random from the model ensemble are shown. (b) The marginal posterior distribution between 60 and 80 km depth [the region between the horizontal dashed lines in (a)] indicates that the uncertainty in the distribution beneath the conductive melt prism is greater than on either side of it. The true model values are indicated by white squares. (c) Marginal distribution through the melt prism [the region between the vertical dashed lines in (a)] shows a tightening of the uncertainty where the model is conductive, relative to where it is resistive. The red lines represent the 5th and 95th percentiles, and indicate the 90 per cent credible interval.

Because $\boldsymbol{g}(\boldsymbol{x}^*)$ is viewed as a conditional random variable, many different realizations of the function are possible. Here we use the mean of the distribution of realizations, analytically provided by the equation

$$\boldsymbol{\mu}^* = \boldsymbol{K}_*{}^T \boldsymbol{K}_{\boldsymbol{m}}{}^{-1} \boldsymbol{g}(\boldsymbol{x}), \tag{9}$$

where $\boldsymbol{\mu}^* = E[\boldsymbol{g}(\boldsymbol{x}^*|\boldsymbol{x})]$ is the expected value of $\boldsymbol{g}$ at the gridpoints $\boldsymbol{x}^*$, given the value of $\boldsymbol{g}$ at the interpolation nodes $\boldsymbol{x}$. If we now take the function locations to be subsurface coordinates and $\boldsymbol{g}$ to be some subsurface property of interest—in our case, electrical resistivity—we can use eq. (9) to connect a parsimonious model parametrization described at $\boldsymbol{x}$ by $\boldsymbol{g}(\boldsymbol{x})$ to the dense computational model described at the mesh cells $\boldsymbol{x}^*$ by $\boldsymbol{\mu}^*$. Essentially, the GP acts as both a parsimonious description of the subsurface, via the interpolation nodes $(\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))$, and as the transform between it and the dense computational model $\boldsymbol{\mu}^*$, via eq. (9) and the kernel function in eq. (4). From here on, we use the identities $\boldsymbol{m} = (\boldsymbol{x}, \boldsymbol{g}(\boldsymbol{x}))$ and $\boldsymbol{m}^* = (\boldsymbol{x}^*, \boldsymbol{\mu}^*)$ to refer to the parsimonious and dense subsurface resistivity models, respectively (steps 1–5 in Fig. 1).

In the context of trans-dimensional MCMC, we do not use the GP by itself to infer the value of subsurface properties. As a result, there is no need to generate realizations from a large covariance matrix, which has traditionally been a computational bottleneck in the use of high dimensional GPs (See Lindgren *et al.* 2011; Emzir *et al.* 2020, for possible workarounds). Instead, the MCMC sampling is performed on $\boldsymbol{m}$, the parsimonious model specified by the GP nuclei (i.e. their positions and resistivities), while the data misfit of each of these models is calculated using their dense representation, $\boldsymbol{m}^*$. Thus our approach strikes a balance between a parsimonious model representation $\boldsymbol{m}$ that is suitable for efficient MCMC sampling with

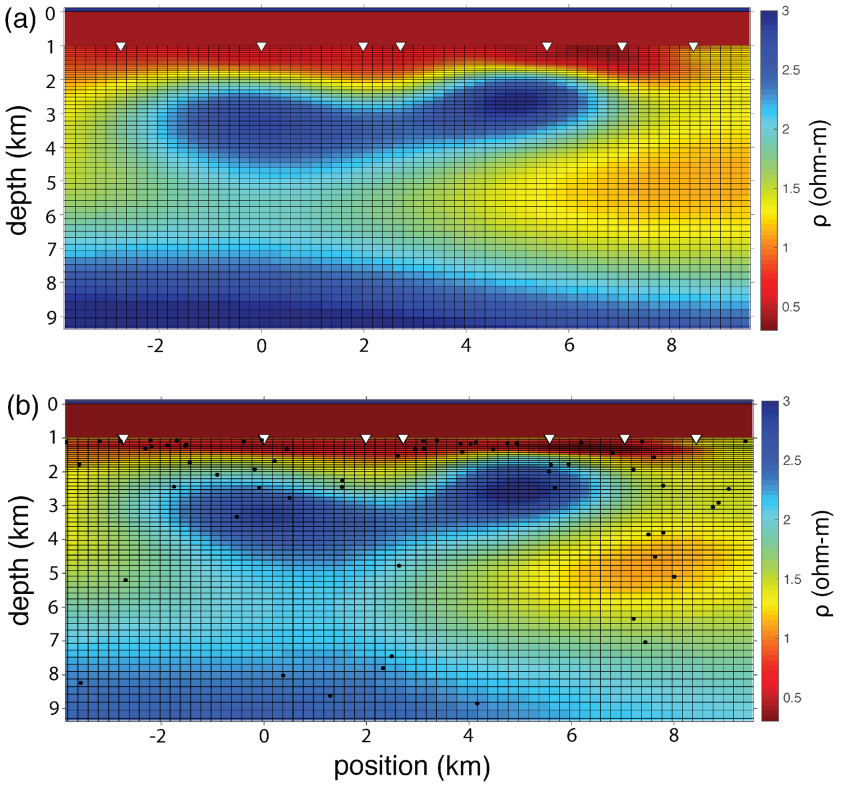

Figure 6. Gemini MT data inversion results. (a) Deterministic inversion of the Gemini data, obtained using MARE2DEM. (b) Mean of the TDGP model ensemble. The dense resistivity mesh is outlined in black, while the interpolation node locations for a model randomly chosen from the ensemble are shown in (b) as black circles. The MT sites are represented by white triangles, in order from station 1 on the left to station 7 on the right. The uncertainty on the base of the salt (blue object in the center) is one of the key questions left unresolved by this inversion. The similarities between (a) and (b) are striking given the highly varied nature of the models in the ensemble (Fig. 10). Note the use of a linear colour scale given the small range of inverted resistivity.

the necessary spatial structure required to fit the data in the dense model $\boldsymbol{m}^*$.

The prior distribution, $p(\boldsymbol{m})$, can now be properly defined, following Ray & Myer (2019), in terms of the number of interpolation nodes, k, their locations, $\boldsymbol{x}_k$, and their resistivities, $\boldsymbol{m}_k$, as

$$p(\boldsymbol{m}) = p(\boldsymbol{m}_k, \boldsymbol{x}_k, k). \tag{10}$$

To be more explicit, $\boldsymbol{m}_k$ represents the base-10 logarithm of electrical resistivity, ρ, at the k interpolation nodes, $\boldsymbol{x}_k$. Using the chain rule of conditional probabilities, we can rewrite this as

$$p(\boldsymbol{m}_k, \boldsymbol{x}_k, k) = p(\boldsymbol{m}_k|\boldsymbol{x}_k, k)p(\boldsymbol{x}_k|k)p(k). \tag{11}$$

If we now assume that the resistivities are uniformly distributed over a log-resistivity range $\Delta\rho$ and that the interpolation nodes are uniformly distributed over a length, area, or volume (determined by the value of n_d) defined by $\prod_{i=1}^{n_d} \Delta x_i$, we can write the prior as

$$p(\boldsymbol{m}_k, \boldsymbol{x}_k, k) = \frac{1}{\Delta\rho^k} \frac{k!}{(\prod_{i=1}^{n_d} \Delta x_i)^k} p(k). \tag{12}$$

There are a number of choices for the prior distribution on the number of interpolation nodes, $p(k)$. Here we choose it to be uniform

$$p(k) = \frac{1}{k_{\max} - k_{\min} + 1}, \tag{13}$$

where $k_{\max}$ and $k_{\min}$ are the maximum and minimum allowed number of interpolation nodes, respectively. We select $k_{\min} = 2$ and $k_{\max}$ such that the vast majority of models found by the inversion have fewer than $k_{\max}$ interpolation nodes.

2.4 HPC and code implementation

We implement the TDGP sampler in Julia, a computationally efficient, modern software language designed for scientific and HPC (Bezanson *et al.* 2017). The MT forward modelling in our code leverages the MARE2DEM (Key 2016) code, a parallel, goal-oriented adaptive finite-element solver written in modern Fortran that uses the message-passing interface (MPI) standard for efficient parallel forward calculations. MARE2DEM uses a dual grid approach and thus is well suited for our TDGP method. The dense resistivity grid calculated from the parsimonious GP parametrization is input to MARE2DEM as regions of piecewise constant conductivity and then MARE2DEM automatically generates a finite element mesh that conforms to the dense grid. This mesh is automatically adaptively refined until MARE2DEM obtains accurate EM responses given the input resistivity structure and data parameters. Further, since MARE2DEM uses an unstructured triangular mesh, it can readily handle complicated model structures such as topography or seismically imaged geologic surfaces. In our implementation, the input data and model grid files use the same format as used for regularized inversion with MARE2DEM; thus a TDGP inversion can be run after regularized inversion with only a small additional amount of user effort. To communicate between Julia and Fortran in a HPC environment, we utilize a Julia-to-Fortran interface subroutine that directly passes variables in memory. See Fig. 1 for a description of the workflow.

The GP-based transform from parsimonious ($\boldsymbol{m}$) to dense ($\boldsymbol{m}^*$) parametrization is computationally efficient, and for parsimonious models with a few thousand parameters can be accomplished on a

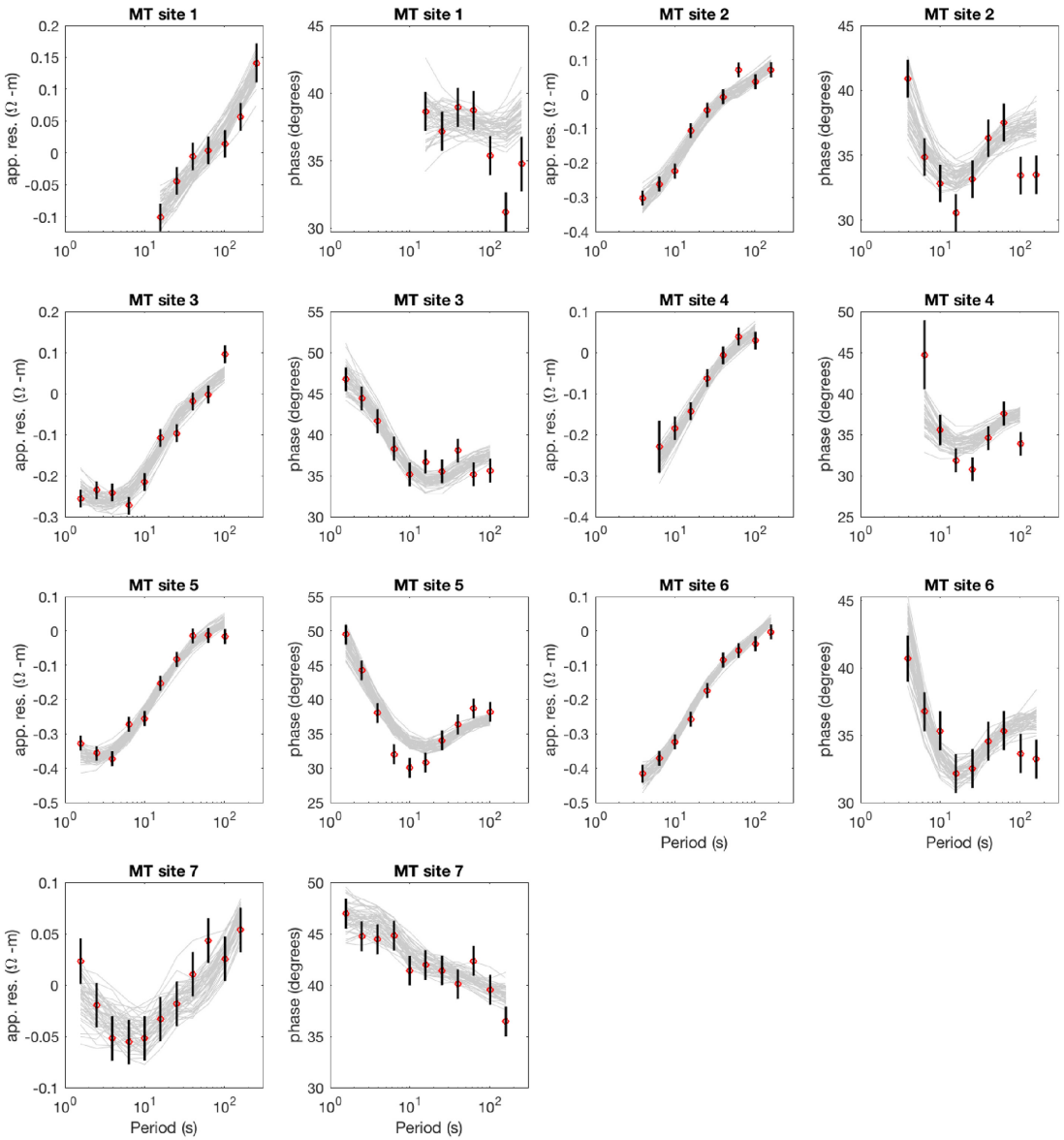

Figure 7. Gemini TE mode apparent resistivity and phase for 7 MT sites (see white triangles in Fig. 6). In addition, the forward responses for 50 randomly selected models from the ensemble are plotted in grey. The distribution of the RMS fit to the data across the ensemble was 1.1–1.5, with a peak at 1.28.

single CPU in under 0.1 s via Cholesky factorization. As described in Ray & Myer (2019), only parts of the kernel matrices $\boldsymbol{K}_m$ and $\boldsymbol{K}_*$ have to be updated at each MCMC step, leading to significant computational savings in the computation and storage of the kernel function in eq. (4). This means that the computational heavy lifting for 2-D MT occurs during forward modelling. As a result, all the MCMC and GP steps can be efficiently performed on a single CPU, which we refer to as the 'manager'. This includes producing the proposal model, $\boldsymbol{m}_{prop}$, from the current model via birth-death MCMC, as well as computing $\boldsymbol{m}^*$ using eq. (9).

Once $\boldsymbol{m}^*$ has been calculated on the manager in Julia, these model parameters are passed via the direct Julia-to-Fortran interface to worker CPUs, who compute the forward response and data misfit in parallel using MARE2DEM. This information is then passed back to the manager to determine whether $\boldsymbol{m}_{prop}$ or the current model, $\boldsymbol{m}_n$, is accepted as the next model in the Markov chain.

Finally, we implement parallel tempering, which allows multiple Markov chains running in parallel to swap temperatures at the end of each iteration. This facilitates more efficient sampling of the model space and accelerates convergence of the Markov chains to the posterior distribution. In addition, parallel tempering is easily implemented in an HPC environment, enabling convergence to be sped up by harnessing more CPUs. Further details of our parallel tempering and MCMC implementations can be found in Ray *et al.* (2013) and Blatter *et al.* (2018).

3 RESULTS

3.1 Synthetic inversion

We test our algorithm on a synthetic data set generated from a resistivity model designed to simulate a prism of partial melt at the lithosphere–

asthenosphere boundary (LAB), inspired by the asthenospheric melt channel beneath the Cocos Plate seen in Naif *et al.* (2013).

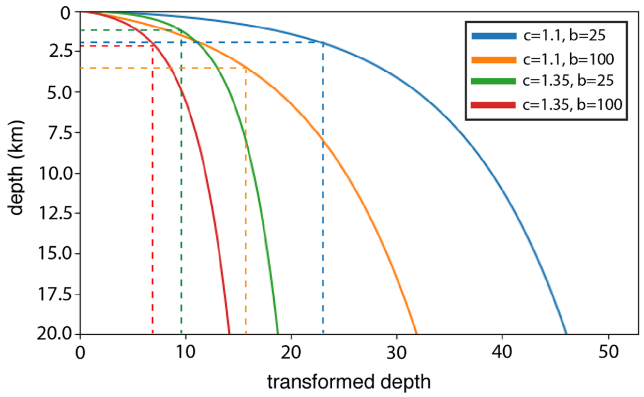

Figure 8. Transformation from linear depth (km) to sampled depth, $\hat{z}$, given in eq. (15). The choice of b and c determines the fraction of a given interval $[0, \hat{z}_{\max}]$ in the transformed domain that maps to shallow depths in the untransformed domain. The dashed lines indicate for each combination of b and c the point $(\frac{\hat{z}_{\max}}{2}, h(\frac{\hat{z}_{\max}}{2}))$, the linear depth that corresponds to the midpoint of $[0, \hat{z}_{\max}]$.

Fig. 2 shows the true model, which includes shallow, extrusive volcanics; resistive lithosphere; conductive asthenosphere; and deeper mantle. At the base of the lithosphere is a 10 km thick, 40 km wide, conductive melt prism. Synthetic apparent resistivity and phase data were generated using MARE2DEM at 11 MT sites (white triangles) spaced 20 km apart, over a period range from 10 to 10 000 s to which uncorrelated Gaussian noise (4 per cent of the absolute value of each datum) were added. This period range was chosen since it represents the working bandwidth of broadband (*cf.* long-period) marine MT receivers in the deep ocean (Constable 2013), which are able to measure the shorter period signals sensitive to asthenosphere conductive structure. Fig. 3 shows the synthetic data for all 11 MT sites, which exhibit a characteristic lowering of the TE mode apparent resistivity and larger TE phases at 100–1000 s period for stations located over the conductive melt prism.

We first conduct a gradient-based inversion of the synthetic data using MARE2DEM (Fig. 4). MARE2DEM utilizes the Occam method that searches for the smoothest model that fits the data (Constable *et al.* 1987). The smooth inversion model indicates the features of the model that the synthetic data are likely able to resolve. Plotted on top of the smooth inversion model in Fig. 4 are the outlines of the discrete resistivity mesh. Because MT data are most sensitive to shallow structure near the MT receivers, the resistivity mesh must be finer near the receivers than at depth. This means that for data broadband sets that have sensitivity spanning from shallow to deep depths the TDGP algorithm will need to spend equal portions of the time sampling the small, near surface intervals and large, deep intervals. We therefore sample the model space in the base-10 logarithm of depth rather than linear depth. This has the added benefit of allowing a correlation length scale that varies with depth. We choose λ_z to be 0.1 log units in depth to roughly match the depth scale of features in the smooth inversion, and 15 km in the horizontal direction to match the scale of features resolvable given the 20 km station spacing.

The ideal choice of these parameters is not immediately obvious. We selected them by first doing a 'pixel-matching' test—taking the true resistivity at each resistivity cell (plus noise) as the 'data' and minimizing the square of the distance between the TDGP model and true model resistivity. Solving this low-cost problem by inverting these 'data' using different choices of correlation length-scale is an efficient way of finding values of these parameters that ensure that the GP can adequately represent the features of the true model. For this synthetic model, this scale length selection process took less than 10 min when running on a single processor. In the case of inverting field data, where the true model is not known, a deterministic inversion model (e.g. Fig. 4) can be used instead since this model should contain the scale of features resolvable by the field data, though this amounts to a choice of prior length scale.

We invert the synthetic data on a cluster using 480 processors and 20 parallel tempering chains, with three chains at the coldest, unmodified annealing temperature $T = 1$. See Ray *et al.* (2013) and Blatter *et al.* (2018) for background on temperature settings. The average time per MCMC step was 0.3 s, owing to the relative simplicity of the resistivity mesh, which contained 818 resistivity cells. We ran the algorithm for one million MCMC steps totaling 3.5 d of compute time. We eliminated the first 300 000 models of each $T = 1$ chain as burn-in (i.e. the time it takes to find the first models with an acceptably low misfit), with the rest comprising the model ensemble. The TDGP algorithm required 150–200 interpolation nodes to fit the data compared with 818 model parameters in the dense model, representing a model-space reduction of over 4×. Model responses from 50 models randomly selected from the model ensemble are shown in Fig. 3. The data misfit across all models in the ensemble is χ^2-distributed with a peak implying a RMS misfit of 1.25.

The mean of the ensemble is shown in Fig. 5(a). While the mean of the ensemble is not always an accurate representation of the subsurface, in this case it contains the main features of the model. The melt prism is well recovered, albeit represented as a smooth anomaly rather than one with sharp edges. This is expected, given our choice of an exponential spatial correlation kernel function to represent the model. The interpolation node locations, $\boldsymbol{x}$, for a model chosen at random from the ensemble are shown in Fig. 5(a) as well. The bulk of them are clustered in the shallow portion of the model, as expected given our choice to sample log-depth, as well as the greater data sensitivity there. The mean of the ensemble is, in many respects, similar to the regularized inversion result (Fig. 4). In neither model is the LAB sharply imaged.

Bayesian inversion produces far more than just the mean, however. Marginal distributions obtained by taking horizontal and vertical slices through the model are shown in Figs 5(b) and (c), respectively. The brighter colours indicate regions of higher probability density, while the red lines indicate the 5th and 95th percentiles and delineate the 90 per cent credible interval. The vertical profile through the conductive anomaly quantitatively shows the well-known fact that MT data are better able to constrain the resistivity of conductors than resistors, as shown by the tightening of the 90 per cent credible interval over the anomaly and again at depth. The model uncertainty is significantly higher in the more resistive portions of the model. Interestingly, there is high uncertainty directly beneath the melt prism, visible in the mean as the resistive body beneath the anomaly, as well as in the broadening of the 90 per cent credible interval in this region in both the vertical and horizontal profiles. This is due to the loss of signal strength due to the overlying conductor, which effectively shields sensitivity to the region beneath it. The deterministic inversion provides an estimate of the resistivity in this region, but without this corresponding uncertainty information, whereas the Bayesian inversion shows this region could be moderately resistive (around 100 Ωm) or highly resistive (around 10 000 Ωm).

Generally, the true model lies within the 90 per cent credible interval defined by the model ensemble (see 5b). One exception is in the shallowest portion of the model (the conductive sediments,

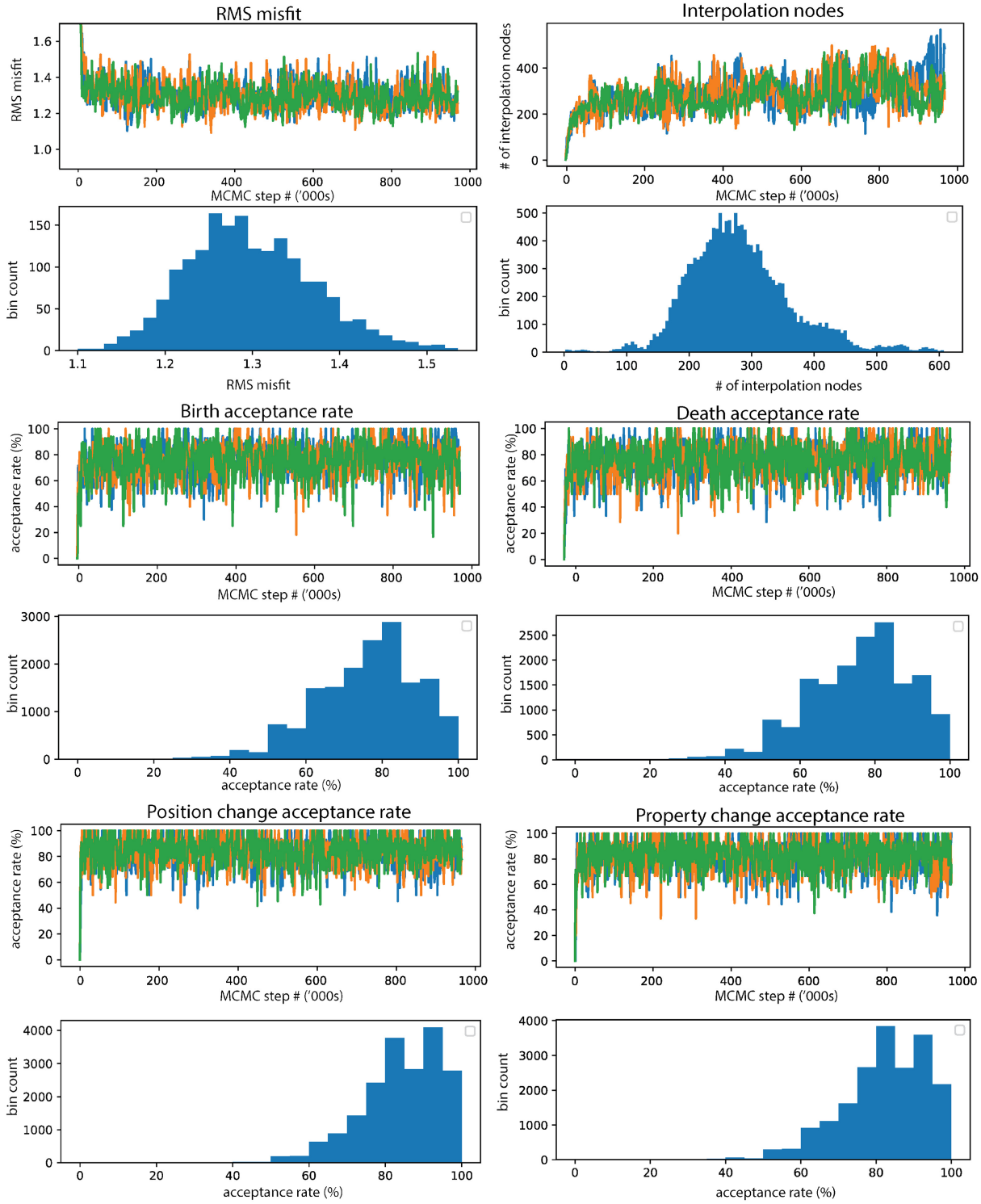

Figure 9. Convergence properties of the trans-D GP inversion of the Gemini data set. Convergence properties (RMS misfit, number of interpolation nodes, and acceptance rates for the birth, death, resistivity update and position update moves) as a function of MCMC iteration number for the three Markov chains at temperature $T = 1$ are plotted in green, orange and blue. Histograms of these properties across all three chains are shown beneath. Models in the ensemble fit the data in the range RMS 1.1–1.5. The bulk of the models contained 125–475 interpolation nodes, compared with 8424 finite elements used by MARE2DEM.

as seen in the shallowest part of Fig. 5c). This is likely because λ_z, the correlation-length scale in depth, was insufficiently small in the shallow part of the model. Finding a value of λ_z that is adequate for all portions of the model given the depth-dependent sensitivity of the data to subsurface structure is challenging. Sampling the base-10 logarithm of depth proved largely adequate, given that the synthetic model spanned a depth range of 150 km. The ideal choice depends on the geology, the data sensitivity, and the dimensional scale of the problem. In the next section we adopt a different approach to selecting λ_z, given a shallower region of interest.

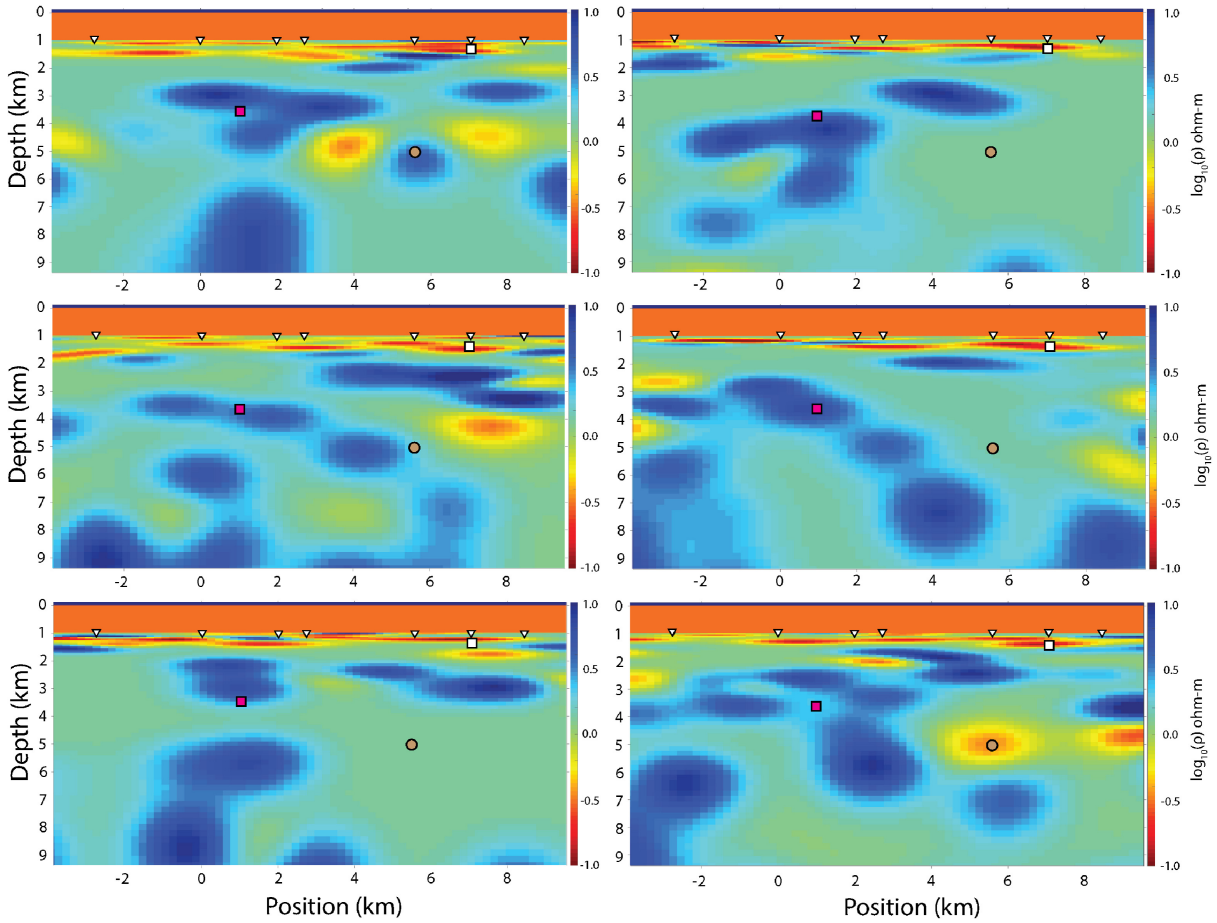

Figure 10. Several resistivity models chosen at random from the model ensemble, evaluated at the gridpoints. Despite the wide degree of variability, the mean of the ensemble coherently images the salt body. The degree of variability provides crucial knowledge of uncertainty in the subsurface resistivity. For example, note how in the vicinity of the magenta square there is always a resistor; similarly, in the vicinity of the white square there is always a conductor, whereas in some other locations the model exhibits more variability among the samples shown here (e.g. the tan circle).

3.2 Field data inversion

To test our algorithm on field data, we invert a selection of seven MT sites from the Gemini data set. This data was collected in the Gulf of Mexico to image a salt body surrounded by conductive sediments and has been analysed in several past studies (Constable *et al.* 1998, 2015; Hoversten *et al.* 2000; Key *et al.* 2006). A new gradient-based inversion of the 1–250 s period data using MARE2DEM is shown in Fig. 6(a), while the mean of a TDGP inversion is shown in Fig. 6(b). Because MT data have low sensitivity to resistive bodies in conductive sediments, the resolved salt body appears as only a small increase in resistivity over the background, despite the actual salt likely having resistivity greater than 1000 Ωm. This is enough to deflect MT currents and saturate the MT responses (e.g. Hoversten *et al.* 2000). Thus here we show resistivity on a linear scale. The relatively resistive (>3 Ωm) upper salt body is clearly imaged, surrounded by conductive (<1 Ωm) sediments. At depth, a larger resistor is present, possibly including a thin layer of salt that may or may not connect to the upper salt body. While the regularized inversion images the top and sides of the salt, the base of the salt is not clear.

We invert the TE mode apparent resistivity and phase since these have the strongest sensitivity to the resistive salt body (Key *et al.* 2006). The data errors were estimated using standard MT data processing workflows, with an error floor imposed of 2.5 per cent for apparent resistivity and phase. We assume our data errors to be uncorrelated. The data are shown in Fig. 7.

3.2.1 *Choosing correlation length scales*

Choosing λ_z, the correlation length-scale in depth, proved challenging owing to the fact that the near-surface is characterized by small-scale structure while at depth the features of the model resolvable by the data are larger-scale. Simply sampling the base-10 logarithm of depth proved in this case insufficient to adequately represent the variable length-scales of the model structure resolvable by the MT data. When sampling the base-10 logarithm of depth, the TDGP chains were unable to achieve a satisfactory level of misfit. To achieve a λ_z that appropriately increases with depth, we used the transform

$$\hat{z} = \log_c \left(1 - \frac{z(1-c)}{b}\right), \tag{14}$$

where z is linear depth, in meters, $\hat{z}$ is transformed depth sampled by the TDGP algorithm, and b and c are constants to be selected. This transform can be intuitively understood by examining its inverse

$$z = h(\hat{z}) = b\,\frac{1-c^{\hat{z}}}{1-c} \tag{15}$$

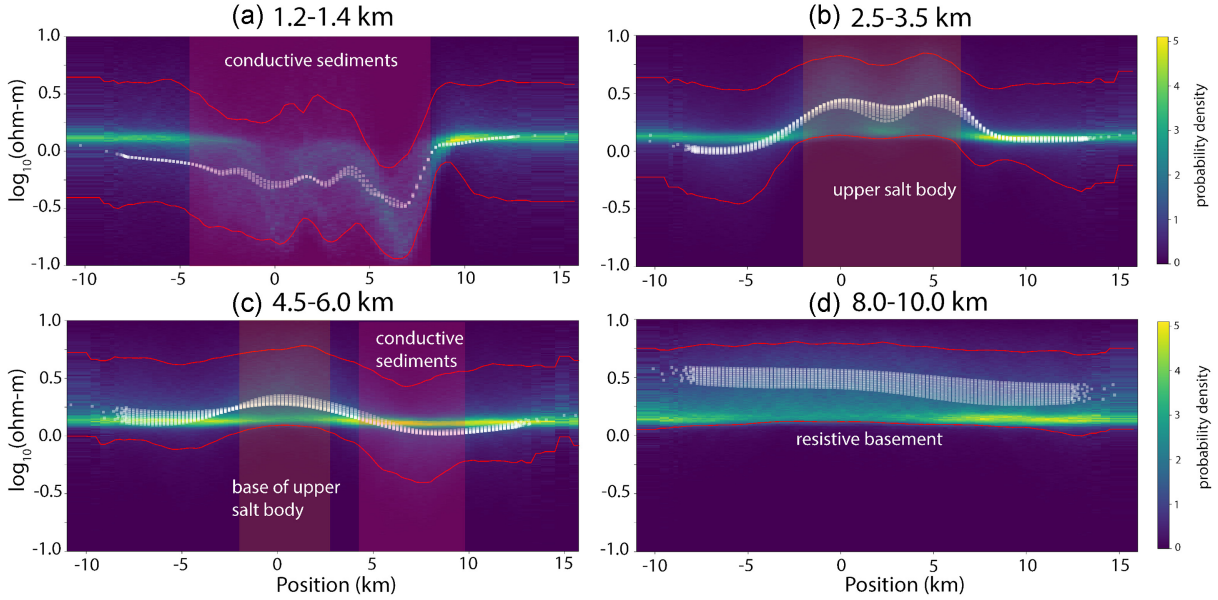

Figure 11. Marginal distributions for subsurface resistivity taken at various depth slices through the model: (a) 1.2–1.4 km; (b) 2.5–3.5 km; (c) 4.5–6.0 km and (d) 8.0–10.0 km. Warmer colours indicate higher probability density while the red lines denote the 90 per cent credible interval. The white squares represent gradient-based resistivity estimates at the gridpoints within the depth slice. In general, model regions consistent with the salt body (brown shaded regions) exhibit an increase in the 5th percentile to above 1 Ωm, while regions consistent with sediments (pink shaded regions) have a 5th percentile that is significantly lower.

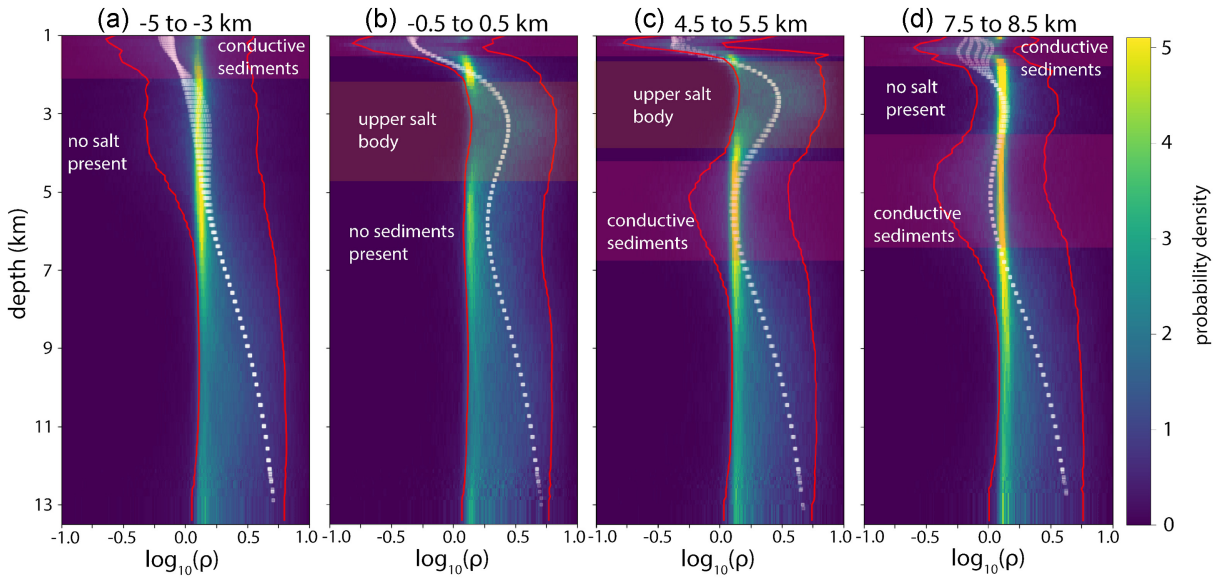

Figure 12. Marginal distributions for subsurface resistivity taken at various vertical slices through the model: (a) –5 to –3 km; (b) –0.5 to 0.5 km; (c) 4.5 to 5.5 km and (d) 7.5 to 8.5 km. Warmer colours indicate higher probability density while the red lines denote the 90 per cent credible interval. The white squares represent gradient-based resistivity estimates at the gridpoints within the depth slice. In general, model regions consistent with the salt body (brown shaded regions) exhibit an increase in the 5th percentile to above 1 Ωm, while regions consistent with sediments (pink shaded regions) have a 5th percentile that is significantly lower.

which is the sum of a geometric series. It is clear from eq. (15) that z increases geometrically with $\hat{z}$. As a consequence, it can be shown (see the Appendix) that for a fixed correlation length in the transformed variable, $\lambda_{\hat{z}}$, the correlation length in linear depth, λ_z, increases linearly with depth

$$\lambda_z = a_0 + a_1 z, \tag{16}$$

where $a_0 = \frac{b(1-c^{\lambda_{\hat{z}}})}{1-c}$ and $a_1 = c^{\lambda_{\hat{z}}} - 1$. It follows that sampling $\hat{z}$ with a constant $\lambda_{\hat{z}}$ is equivalent to sampling z with a λ_z that grows linearly with depth in the model, matching the natural loss of model resolution with depth inherent in the data.

Examples of the transform described in eqs (14) and (15) are shown in Fig. 8 for a few choices of b and c. It is clear that the midpoint of a given interval in transformed depth $[0,\ \hat{z}_{\max}]$ maps to much less than the midpoint of the corresponding interval in linear depth $[0,\ h(\hat{z}_{\max})]$. The dashed lines in Fig. 8 indicate, for each combination of b and c, the point in linear depth that maps to the mid-point of the transformed interval, that is $(\frac{\hat{z}_{\max}}{2},\ h(\frac{\hat{z}_{\max}}{2}))$.

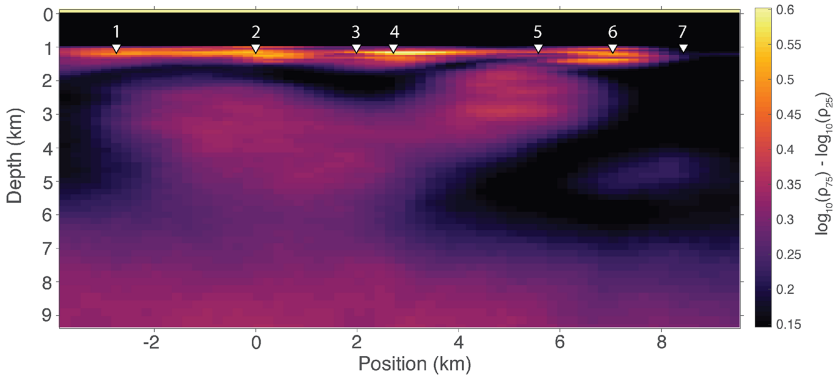

Figure 13. Interquartile range of the model ensemble obtained by TDGP inversion of the Gemini data set. Higher uncertainty generally occurs in the more resistive parts of the model (upper and lower salt bodies). Relative high uncertainty also occurs in the very shallow parts of the model where structure is shallower than the skin depths at the shortest periods for each station's data (shown as white triangles labelled 1–7). The transition between sediments to salt appears clearly demarcated.

$h(\frac{\hat{z}_{\max}}{2}) < \frac{h(\hat{z}_{\max})}{2}$ is a desirable property since we wish to ensure that small, near-surface depth intervals are given the same sampling importance as larger, deeper depth intervals.

Choosing a suitable sampling domain/transform for horizontal distance proved equally important. While the majority of the grid-points are clustered in the region of interest, to achieve an accurate forward response the model domain must extend much further, both horizontally and in depth. This means that a few gridpoints are located tens to hundreds of kilometers away from the region of interest. To avoid the TDGP algorithm placing large quantities of interpolation nodes beyond the region of interest, where data sensitivity is low, we constructed a transform such that λ_y is constant within the region of interest (between –11 and 14 km), but grows rapidly beyond that. While the subject of variable length scales is discussed in Ray & Myer (2019), warping the GP input space as we have done here has a long history (Sampson & Guttorp 1992; MacKay 1998). Here we find that an *a priori* determined scale length worked well, but Ray (in-revision) shows how the scale length can be automatically calculated using a nested GP approach.

3.2.2 *Gemini TDGP inversion results*

We used the TDGP algorithm to invert this data on 168 processors and 12 parallel chains, with three chains at $T = 1$. The average time per MCMC step was 0.85 s, slightly higher than when inverting the synthetic data due to the greater complexity of the model. To accurately represent the model structure resolvable by the MT data, the resistivity mesh needed 8424 independent resistivity cells. This grid was designed based on the complexity of the structure seen in the regularized inversion. We ran the algorithm for one million MCMC steps, for a total of just under 10 d of compute time. We eliminated the first 300 000 models of the $T = 1$ chains as burn-in, the rest comprising the model ensemble. To fit the Gemini data set, the TDGP algorithm required 125–475 interpolation nodes, with a mode of 260—a model-space reduction of approximately 32×. The inversion convergence properties for the $T = 1$ chains are shown in Fig. 9. Six models randomly selected from the ensemble are shown in Fig. 10.

The high acceptance rates for the trans-dimensional moves shown in Fig. 9 could be a sign that the TDGP sampler is taking small steps, possibly suggesting that the algorithm has yet to converge. The fact that the models in Fig. 10 look highly dissimilar from one another, however, leads us to believe that this is not the case. Rather, the high acceptance rates are likely due to the fact that the MT data are relatively uninformative, leading most changes to the model to have only a limited effect on the posterior probability. These MT data have only limited sensitivity to slightly more resistive bodies embedded in conductive sediments and thus there is a large range of resistive features that are compatible with the data set, as can be seen in Fig. 10. Additionally, there is a slight upward trend in the number of interpolation nodes over time, but since there is no corresponding change in RMS misfit, we conclude that this increase in model complexity does not indicate the presence of as-yet-unexplored high probability modes in model space.

The results of inverting the Gemini data set with our TDGP algorithm are shown in Fig. 6(b) and Figs 7–13. The upper salt body, conductive sediments and deeper resistor are clearly visible in the mean (Fig. 6b). Interestingly, while the mean of the ensemble resembles the deterministic model estimate (Fig. 6a), the individual models are highly varied and hardly look like the mean at all (Fig. 10). Yet, if the models in the ensemble were displayed on a screen one after another, well-constrained regions of the model would hardly appear to change from one model to the next, while poorly constrained regions would appear to fluctuate wildly [Tarantola's 'movie strategy', see Koren *et al.* (1991)]. In regularized inversions a model roughness penalty is often used to stabilize the algorithm such that it avoids extremal model structures that are not required to fit the data. By contrast, Fig. 10 indicates that the TDGP algorithm samples such extremal structures as well in order to describe the full range of resistivity that can fit the data, rather than giving only a single smooth model. As a result, the TDGP model ensemble provides a fuller picture of non-linear resolution, as well as critical information on the range of resistivity compatible with the MT data.

Figs 11 and 12 show marginal distributions of resistivity taken over horizontal and vertical slices through the model, respectively. The degree of model structure resolvable by the data as indicated by fluctuations in the 90 per cent credible interval (bounded by the red lines) decreases with depth as well as with horizontal distance from the region of interest as data sensitivity declines and the subsurface becomes gradually more resistive at depth. The gradient-based inversion estimate (white squares) falls within the 90 per cent credible interval at nearly all depths. Unlike the gradient-based estimate, however, the Bayesian inversion provides quantitative uncertainty at each depth. There is a high degree of heterogeneity (and model uncertainty) in the shallow sediments (Fig. 11a, upper 600 m of

Fig. 12), due largely to a lack of high frequency data to constrain the shallow subsurface. By contrast, the salt body is clearly visible as twin peaks in the resistivity credible interval at 0 and 5 km horizontal position (Fig. 11b) and as peaks in the resistivity credible interval between 2.5–4.5 km depth and 2–4 km depth (Figs 12b and c, respectively).

The marginal distributions in Figs 11 and 12 also shed light on whether the base of the salt body is connected to the deeper resistor. The distribution in Fig. 11(c) is taken from the depth range consistent with the base of the salt body and indicates an increase in the 5th percentile to >1 Ωm in the vicinity of 0 km horizontal position. By contrast, the regions of the model characterized by conductive sediments feature a 5th percentile that is considerably lower, around 0.3 Ωm or even less. The absence of a conductive drop in the 5th percentile to <1 Ωm in Fig. 12(b) is evidence that the subsurface there is likely not composed of conductive sediments, as the MT data would be sensitive to them, as in Fig. 12(c). In addition, the mean resistivity (Fig. 6b) and the marginal distribution in Fig. 12(c) image what appears to be a salt overhang at 4.5–5.5 km along the line, a structure that even the most advanced seismic imaging techniques (Zhou *et al.* 2018) struggle to image, and which require far more data and compute power.

Fig. 13 shows another measure of uncertainty, the interquartile range (IQR) of the model ensemble, as a function of depth and horizontal position. Except at very shallow depths, the IQR is generally lowest in the conductive regions of the model and highest in the more resistive structures, illustrating MT's well known sensitivity to conductive features. The IQR is largest in the near-surface, where there are many small-scale heterogeneities and the MT data lack the high-frequencies with small enough skin depths necessary to resolve them. In fact, the subtle IQR variations in the shallow section correlate with the shortest period data available at each station (Fig. 7). Stations 3, 5 and 7 have data to around 1 s period and we see that the shallow IQR values are smallest near these stations; conversely, all other stations only obtained usable data at 5 s and longer periods and have considerably larger IQR values. This clearly illustrates the importance of the MT data frequency bandwidth for reducing uncertainty in inversion models, as well as the crucial role of Bayesian inversion in quantifying model parameter uncertainty.

4 CONCLUSIONS

In this study, we successfully demonstrate fully 2-D, trans-dimensional Bayesian inversion of MT data. We invert synthetic and field MT data for 2-D models of electrical resistivity using the TDGP sampler—to the best of our knowledge, the first 2-D trans-dimensional inversions of MT data to be published. We render this problem tractable from a computational standpoint by achieving model parameter reduction through a trans-dimensional formulation that utilizes GPs. Future work could compare the degree of parameter reduction to that of other trans-dimensional parametrizations. Our algorithm is computationally efficient, converging in less than 10 d for the real data and 3.5 d for the synthetic example on a cluster while using only modest HPC resources. This is largely due to the fact that the GP allowed our TDGP algorithm to sample model spaces 32 and four times smaller than those used by the forward solver for the field and synthetic examples, respectively.

A further advantage of our TDGP sampler is that it is spatial dimension-agnostic, meaning the same basic code and theory are applicable for parametrizing 1-D, 2-D and 3-D problems. Although not shown here, we have tested the model parametrization and code for 3-D problems as well. While the parametrization is efficient, the full 3-D forward MT problem is not yet tractable for the <1 s compute times per forward call needed for practical run times. We expect that our parametrization will prove useful for other geophysical inverse problems such as seismic tomography.

The output of our algorithm—the Bayesian posterior distribution—provides quantitative uncertainties on subsurface bulk resistivity. This information is vital for constraining subsurface properties that impact bulk resistivity. Depending on the particular geologic target under investigation, the uncertainties in resistivity can be propagated into uncertainties in pore fluid resistivity, groundwater salinity, porosity, melt fraction and bulk volatile content (Blatter *et al.* 2018, 2019; Blatter 2020).

ACKNOWLEDGEMENTS

We gratefully acknowledge Alberto Malinverno and Bill Menke for many constructive conversations and suggestions, and Jan Dettmer and Andrea Licciardi for their helpful reviews. This work was supported by the Electromagnetic Methods Research Consortium at Columbia University. We also acknowledge computing resources from Columbia University's Shared Research Computing Facility project, which is supported by NIH Research Facility Improvement Grant 1G20RR030893-01, and associated funds from the New York State Empire State Development, Division of Science Technology and Innovation (NYSTAR) Contract C090171, both awarded 15 April 2010.

DATA AVAILABILITY

The data underlying this paper will be shared on request to the corresponding author.

REFERENCES

Agostinetti, N.P. & Bodin, T., 2018. Flexible coupling in joint inversions: a Bayesian structure decoupling algorithm, *J. geophys. Res.,* **123**(10), 8798–8826.

Agostinetti, N.P. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *J. geophys. Int.,* **181**(2), 858–872

Agostinetti, N.P., Giacomuzzi, G. & Malinverno, A., 2015. Local three-dimensional earthquake tomography by trans-dimensional Monte Carlo sampling, *J. geophys. Int.,* **201**(3), 1598–1617.

Bezanson, J., Edelman, A., Karpinski, S. & Shah, V.B., 2017. Julia: a fresh approach to numerical computing, *SIAM Rev.,* **59**(1), 65–98.

Blatter, D., 2020, Constraining fluid properties in the mantle and crust using Bayesian inversion of electromagnetic data, *PhD thesis*, Columbia University, New York City.

Blatter, D., Key, K., Ray, A., Foley, N., Tulaczyk, S. & Auken, E., 2018. Trans-dimensional Bayesian inversion of airborne transient EM data from Taylor Glacier, Antarctica, *J. geophys. Int.,* **214**(3), 1919–1936.

Blatter, D., Key, K., Ray, A. & Gustafson, C., 2019. Bayesian joint inversion of controlled source electromagnetic and magnetotelluric data to image freshwater aquifer offshore New Jersey, *J. geophys. Int.,* **218**(3), 1822–1837.

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

Brodie, R. & Sambridge, M., 2012. Transdimensional Monte Carlo inversion of AEM data, *ASEG Extend. Abstr.,* **2012**(1), 1–4.

Burdick, S., Lekic, V. & Waszek, L., 2018 . Exploring Deep Earth structure and its uncertainty with transdimensional tomography, in *Proceedings of the 20th EGU General Assembly, EGU2018,* 4–13 April 2018, Vienna, Austria, p. 11389.

Cagniard, L., 1953. Basic theory of the magneto-Telluric method of geophysical prospecting, *Geophysics,* **18**(3), 605–635.

Calvetti, D. & Somersalo, E., 2018. Inverse problems: from regularization to Bayesian inference, *Wiley Interdiscip. Rev. Comput. Stat.,* **10**(3), e1427.

Chave, A. & Jones, A.G.eds, 2012. *The Magnetotelluric Method: Theory and Practice,* Cambridge Univ. Press.

Chen, J., Hoversten, G.M., Key, K., Nordquist, G. & Cumming, W., 2012. Stochastic inversion of magnetotelluric data using a sharp boundary parameterization and application to a geothermal site, *Geophysics,* **77**(4), E265–E279.

Commer, M. & Newman, G.A., 2009. Three-dimensional controlled-source electromagnetic and magnetotelluric joint inversion, *J. geophys. Int.,* **178**(3), 1305–1316.

Constable, S., 2013. Review paper: Instrumentation for marine magnetotelluric and controlled source electromagnetic sounding, *Geophys. Prospect.,* **61**(Part I), 505–532.

Constable, S., Orange, A. & Key, K., 2015. And the geophysicist replied: 'which model do you want?", *Geophysics,* **80**(3), E197–E212.

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

Constable, S.C., Orange, A.S., Hoversten, G.M. & Morrison, H.F., 1998. Marine magnetotellurics for petroleum exploration Part I: a sea-floor equipment system, *Geophysics,* **63**(3), 816–825.

de Pasquale, G. & Linde, N., 2017. On structure-based priors in Bayesian geophysical inversion, *J. geophys. Int.,* **208**(3), 1342–1358.

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

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

Dettmer, J., Benavente, R. & Cummins, P.R., 2014. Trans-dimensional finite-fault inversion, *J. geophys. Int.,* **199,** 735–751.

Dettmer, J., Dosso, S.E., Bodin, T. & Stipčević, J., 2015. Direct-seismogram inversion for receiver-side structure with uncertain source–time functions, *J. geophys. Int.,* **203**(2), 1373–1387.

Emzir, M., Lasanen, S., Purisha, Z., Roininen, L. & Särkkä, S., 2020. Non-stationary multi-layered Gaussian priors for Bayesian inversion, *Inverse Probl.,* **37**(1), 015002.

Fournier, D. & Oldenburg, D.W., 2019. Inversion using spatially variable mixed ℓp norms, *J. geophys. Int.,* **218**(1), 268–282.

Galetti, E. & Curtis, A., 2018. Transdimensional electrical resistivity tomography, *J. geophys. Res.,* **123**(8), 6347–6377.

Geyer, C.J. & Moller, J., 1994. Simulation procudures and likelihood inference for spatial point processes, *Scand. J. Stat.,* **21**(4), 359–373.

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

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

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using trans-dimensional trees, *J. geophys. Int.,* **203**(2), 972–1000.

Hawkins, R., Bodin, T., Sambridge, M., Choblet, G. & Husson, L., 2019. Trans-dimensional surface reconstruction with different classes of parameterization, *Geochem. Geophys. Geosyst.,* **20**(1), 505–529.

Hoversten, G.M., Constable, S.C. & Morrison, H.F., 2000. Marine magnetotellurics for base-of-salt mapping: Gulf of Mexico field test at the Gemini structure, *Geophysics,* **65**(5), 1476–1488.

Key, K., 2016. MARE2DEM: a 2-D inversion code for controlled-source electromagnetic and magnetotelluric data, *J. geophys. Int.,* **207**(1), 571–588.

Key, K.W., Constable, S.C. & Weiss, C.J., 2006. Mapping 3D salt using the 2D marine magnetotelluric method: case study from Gemini prospect, Gulf of Mexico, *Geophysics,* **71**(1), B17–B27.

Koren, Z., Mosegaard, K., Landa, E., Thore, P. & Tarantola, A., 1991. Monte Carlo estimation and resolution analysis of seismic background velocities, *J. geophys. Res.,* **96**(B12), 20 289–20 299.

Krige, D.G., 1952. A statistical approach to some basic mine valuation problems on the Witwatersrand, by D.G. Krige, published in the Journal, December 1951 : interim reply by the author to the discussion, *J. S. Afr. Inst. Min. Metall.,* **52**(11), 264–266.

Ley-Cooper, A.Y., 2016. Dealing with uncertainty in AEM models (and learning to live with it), *ASEG Extend. Abstr.,* **2016**(1), 1–6.

Lindgren, F., Rue, H. & Lindström, J., 2011. An explicit link between Gaussian fields and Gaussian Markov random fields: the stochastic partial differential equation approach, *J. R. Stat. Soc., B,* **73**(4), 423–498.

MacKay, D., 1998. Introduction to Gaussian processes, *NATO ASI Ser. F Comp. Syst. Sci.,* **168,** 133–166.

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

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

Malinverno, A. & Leaney, S., 2000. A Monte Carlo method to quantify uncertainty in the inversion of zero-offset VSP data, in *Proceedings of the SEG Annual Meeting,* pp. 2393–2396, Society of Exploration Geophysicists.

Metropolis, N., Rosenbluth, A.W., Rosenbluth, M.N., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21**(6), 1087–1092.

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

Moritz, H., 1980. *Advanced Physical Geodesy,* Wichmann, Karlsruhe.

Mosegaard, K., 2019. Efficient Monte Carlo uncertainty quantification through problem-dependent proposals, in *Proceedings of the 81st EAGE Conference and Exhibition 2019 Workshop Programme,* pp. 1–1, European Association of Geoscientists & Engineers, Netherlands.

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

Naif, S., Key, K., Constable, S. & Evans, R.L., 2013. Melt-rich channel observed at the lithosphere-asthenosphere boundary, *Nature,* **495**(7441), 356–359.

Neal, R., 2011. MCMC using Hamiltonian dynamics, in *Handbook of Markov Chain Monte Carlo,* eds Brooks, S., Gelman, A., Jones, G.L. & L, M.X., Chapman & Hall/CRC.

Newman, G.A. & Alumbaugh, D.L., 2000. Three-dimensional magnetotelluric inversion using non-linear conjugate gradients, *J. geophys. Int.,* **140**(2), 410–424.

Pyrcz, M.J. & Deutsch, C.V., 2014. *Geostatistical Reservoir Modeling,* Oxford Univ. Press.

Ray, A. & Key, K., 2012. Bayesian inversion of marine CSEM data with a trans-dimensional self parametrizing algorithm, *J. geophys. Int.,* **191**(3), 1135–1151.

Ray, A. & Myer, D., 2019. Bayesian geophysical inversion with trans-dimensional Gaussian process machine learning, *J. geophys. Int.,* **217**(3), 1706–1726.

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

Ray, A., Key, K., Bodin, T., Myer, D. & Constable, S., 2014. Bayesian inversion of marine CSEM data from the Scarborough gas field using a transdimensional 2-D parametrization, *J. geophys. Int.,* **199**(3), 1847–1860.

Ray, A., Sekar, A., Hoversten, G.M. & Albertin, U., 2016. Frequency domain full waveform elastic inversion of marine seismic data from the Alba field using a Bayesian trans-dimensional algorithm, *J. geophys. Int.,* **205**(2), 915–937.

Rosas-Carbajal, M., Linde, N., Kalscheuer, T. & Vrugt, J.A., 2013. Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data, *J. geophys. Int.,* **196**(3), 1508–1524.

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

Sampson, P.D. & Guttorp, P., 1992. Nonparametric estimation of nonstationary spatial covariance structure, *J. Am. Stat. Assoc.,* **87**(417), 108–119.

Schoniger, A., Illman, W.A., Wöhling, T. & Nowak, W., 2015. Finding the right balance between groundwater model complexity and experimental effort via Bayesian model selection, *J. Hydrol.,* **531**(Part 1), 96–110.

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys.,* **20**(2), 219–232.

Tikhonov, A.N., 1950. On investigation of electrical characteristics of deep strata of Earth's crust, *Dokl. Akad. Nauk. SSSR,* **73,** 295–297.

Williams, C.K.I. & Rasmussen, C.E., 1996. Gaussian processes for regression, in *Advances in Neural Information Processing Systems,* pp. 514–520, eds Touretzky, D.S., Mozer, M.C. & Hasselmo, M.E., MIT Press.

Zhang, X., Curtis, A., Galetti, E. & de Ridder, S., 2018. 3-D Monte Carlo surface wave tomography, *J. geophys. Int.,* **215**(3), 1644–1658.

Zhou, H.-W., Hu, H., Zou, Z., Wo, Y. & Youn, O., 2018. Reverse time migration: a prospect of seismic imaging methodology, *Earth-Sci. Rev.,* **179,** 207–227.

APPENDIX: DERIVATION OF LINEARLY INCREASING CORRELATION LENGTH SCALE USING GEOMETRIC TRANSFORM

We define linear depth in terms of the transform variable $\hat{z}$,

$$z = h(\hat{z}) = b\,\frac{1-c^{\hat{z}}}{1-c} \tag{A1}$$

$$= a_2(1-c^{\hat{z}}), \tag{A2}$$

where $a_2 = \frac{b}{1-c}$. The correlation length scale in linear depth, λ_z, is defined in terms of $\hat{z}$ as

$$\lambda_z = h(\hat{z}+\lambda_{\hat{z}}) - h(\hat{z}). \tag{A3}$$

Substituting in from eq. (A2) above, we have

$$\lambda_z = a_2(1-c^{\hat{z}+\lambda_{\hat{z}}}) - a_2(1-c^{\hat{z}}) \tag{A4}$$

$$= a_2 c^{\hat{z}}(1-c^{\lambda_{\hat{z}}}). \tag{A5}$$

Using the identity

$$c^{\hat{z}} = \frac{a_2 - z}{a_2} \tag{A6}$$

from eq. (A2), we obtain

$$\lambda_z = (a_2 - z)(1-c^{\hat{z}}) \tag{A7}$$

$$= a_0 + a_1 z, \tag{A8}$$

where $a_0 = b\,\frac{1-c^{\lambda_{\hat{z}}}}{1-c}$ and $a_1 = c^{\lambda_{\hat{z}}} - 1$.