Geophysical Journal International

Geophys. J. Int. (2019) **218,** 1822–1837
Advance Access publication 2019 June 03
GJI Marine Geosciences and Applied Geophysics

doi: 10.1093/gji/ggz253

Bayesian joint inversion of controlled source electromagnetic and magnetotelluric data to image freshwater aquifer offshore New Jersey

Daniel Blatter[1], Kerry Key[1], Anandaroop Ray[2], Chloe Gustafson[1] and Rob Evans[3]

[1]*Lamont-Doherty Earth Observatory, Columbia University, Palisades, NY, USA. E-mail: daniel.blatter@columbia.edu*
[2]*Geoscience Australia, Canberra, Australia*
[3]*Department of Geology and Geophysics, Woods Hole Oceanographic Institution, Woods Hole, MA* 02543, USA

Accepted 2019 May 30. Received 2019 May 24; in original form 2018 December 7

SUMMARY

Joint inversion of multiple electromagnetic data sets, such as controlled source electromagnetic and magnetotelluric data, has the potential to significantly reduce uncertainty in the inverted electrical resistivity when the two data sets contain complementary information about the subsurface. However, evaluating quantitatively the model uncertainty reduction is made difficult by the fact that conventional inversion methods—using gradients and model regularization—typically produce just one model, with no associated estimate of model parameter uncertainty. Bayesian inverse methods can provide quantitative estimates of inverted model parameter uncertainty by generating an ensemble of models, sampled proportional to data fit. The resulting posterior distribution represents a combination of a priori assumptions about the model parameters and information contained in field data. Bayesian inversion is therefore able to quantify the impact of jointly inverting multiple data sets by using the statistical information contained in the posterior distribution. We illustrate, for synthetic data generated from a simple 1-D model, the shape of parameter space compatible with controlled source electromagnetic and magnetotelluric data, separately and jointly. We also demonstrate that when data sets contain complementary information about the model, the region of parameter space compatible with the joint data set is less than or equal to the intersection of the regions compatible with the individual data sets. We adapt a trans-dimensional Markov chain Monte Carlo algorithm for jointly inverting multiple electromagnetic data sets for 1-D earth models and apply it to surface-towed controlled source electromagnetic and magnetotelluric data collected offshore New Jersey, USA, to evaluate the extent of a low salinity aquifer within the continental shelf. Our inversion results identify a region of high resistivity of varying depth and thickness in the upper 500 m of the continental shelf, corroborating results from a previous study that used regularized, gradient-based inversion methods. We evaluate the joint model parameter uncertainty in comparison to the uncertainty obtained from the individual data sets and demonstrate quantitatively that joint inversion offers reduced uncertainty. In addition, we show how the Bayesian model ensemble can subsequently be used to derive uncertainty estimates of pore water salinity within the low salinity aquifer.

Key words: Controlled source electromagnetics (CSEM); Joint inversion; Magnetotellurics; Statistical methods; Marine electromagnetics; Probability distributions.

1 INTRODUCTION

Electromagnetic (EM) methods use passive or active source EM fields to probe subsurface conductivity structure. Depending on the frequency content of the source fields, the depth of investigation can range from hundreds of kilometres to just a few meters. Commonly employed EM methods include the Magnetotelluric (MT) method, which measures the Earth's EM response to natural variations in the geomagnetic field, and the controlled source electromagnetic method (CSEM), which actively generates EM fields and measures their attenuation and phase shift at receivers offset from the transmitter.

The measured EM fields must be inverted to recover an estimate of subsurface conductivity. The standard deterministic inversion technique for EM geophysical data relies upon an objective function that is typically the noise-weighted L2 norm of the data misfit between measured and forward modelled data, and a regularization penalty term against model roughness and/or difference from

a reference model. Gradients in the objective function are then used to attempt to discover a model that minimizes this function (e.g. Constable *et al.* 1987; Newman & Alumbaugh 2000). Using this approach, the inversion algorithm searches until it produces a model that fits the data to within a user specified tolerance—ideally to within the data uncertainty. Other nearby models can also be found by adjusting the data fit tolerance or by adjusting the relative weighting of the model roughness norm, which results in finding one or more additional models that are usually in a neighbourhood of the original best-fitting model.

What these methods leave unaddressed is the uncertainty in the estimated model parameters. Most geophysical inverse problems are non-unique—meaning that an infinite number of models may adequately explain the measured data. In fact, regularization must be introduced in the formulation of the objective function precisely to constrain this non-uniqueness and allow the inverse algorithm to converge to a solution. Bayesian sampling-based inverse methods are a class of algorithms that provide an estimate of model parameter uncertainty by generating an ensemble of models—each of which fits the data, and from which statistical properties of the model parameters can be inferred (Mosegaard & Tarantola 1995). Additionally, they do not require linearization, nor do they require explicit regularization—but these benefits come at the price of significant, additional computational cost. For an illuminating discussion of the equivalence between Bayesian and deterministic methods, see Calvetti & Somersalo (2018). Recent examples of Bayesian inverse methods applied to CSEM data are provided in, *inter alia*, Ray & Key (2012) and Gehrmann *et al.* (2015), and MT examples are given in Mandolesi *et al.* (2018) and Xiang *et al.* (2018).

The non-uniqueness of the inverse problem often results in significant model parameter uncertainty. This uncertainty can arise from multiple sources, including simplifications in the governing physics equations, computational inaccuracies in the forward calculations, and measurement uncertainty. One approach for reducing this uncertainty is to collect multiple kinds of data and invert them jointly. Joint inversion aims to reduce the contribution of measurement error and governing physics to inverted model parameter uncertainty. Because models under joint inversion are required to fit all data sets simultaneously, the range of parameter values compatible with the joint data set should be smaller than the range compatible with any of the individual data sets, if they contain complementary information about the model (Fig. 1).

Jointly inverting EM data sets with standard inverse methods—using gradients to find a minimizer of an objective function—often yields significant, qualitative improvements in the inverted model parameters (e.g. Abubakar *et al.* 2011). Because these methods do not provide true, non-linear estimates of model parameter uncertainty, however, this qualitative improvement is difficult to quantify. Bayesian sampling-based inverse methods, on the other hand, can easily be adapted to joint inversion, allowing the improvement in model parameter uncertainty to be evaluated quantitatively.

Joint inversion methods using gradients to minimize an objective function have been applied to geophysical problems since at least the 1970s (Vozoff & Jupp 1975). The range of data types that have been jointly inverted is extensive, and includes: DC resistivity and MT data (Sasaki 1989); DC resistivity and CSEM (Gómez Trevino & Edwards 1983); marine seismic and CSEM data (Hoversten *et al.* 2006); DC resistivity and seismic tomography (Gallardo & Meju 2004); seismic traveltime and gravity (Lelièvre *et al.* 2012); strong motion, teleseismic, geodetic, and tsunami data (Yokota *et al.* 2011); electrical resistance and ground-penetrating radar data (Bouchedda *et al.* 2012); glacial isostatic adjustment and mantle convection

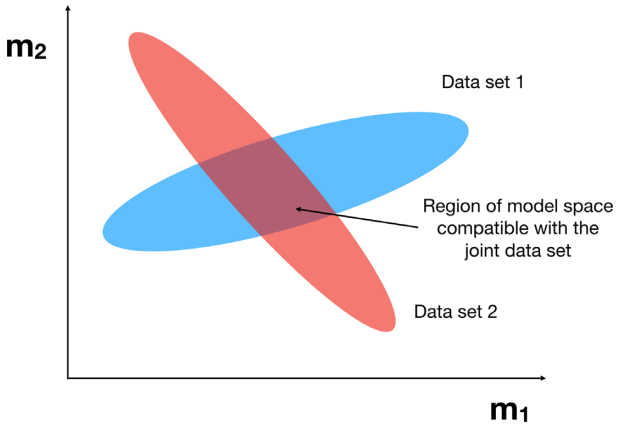

Figure 1. Simple cartoon illustrating how joint inversion might reduce the size of parameter space compatible with the measured data. The region of this 2-D parameter space that is compatible with a hypothetical data set 1 is shown in blue, while the region compatible with data set 2 is shown in orange. These data sets contain complementary information about the model (their compatible regions do not overlap completely), so the region of parameter space compatible with the joint data set should be smaller than either of them (the shaded, overlapping region).

(Mitrovica & Forte 2004); and, most recently, marine MT and CSEM data collected at a slow spreading mid-ocean ridge (Johansen *et al.* 2019). Most joint inversion studies can be classified according to whether all data sets inform the same underlying model parameters (e.g. receiver function and surface wave traveltime data inverted for shear velocity) or whether multiple physical quantities are being inverted for (e.g. CSEM and seismic tomography making inference on electrical resistivity and *P*-wave velocity, respectively). In the former case, the data fitting term can be expanded to include multiple data sets in a relatively straightforward manner, and each data set contributes information to the model directly. In the latter case, the two separate models must be made to 'communicate' in some manner, often through the use of cross gradients or a statistical or analytic relationship between the different physical quantities. More recently, Agostinetti & Bodin (2018) develop a method to permit the two models to share structure only where allowed by the data. In this study, both data sets inform subsurface electrical resistivity, so we will focus our attention on the first approach.

Bayesian sampling-based inverse methods are readily adaptable to a joint inversion framework, which will be discussed in detail in the following section. Jardani *et al.* (2010) inverted synthetic seismic and seismo-electric data for reservoir properties in a 1-D layered model; Rosas-Carbajal *et al.* (2013) inverted synthetic radio frequency MT and electrical resistivity tomography (ERT) data for 2-D electrical resistivity models; Rabben *et al.* (2008) estimated subsurface elastic parameters from synthetic PP and PS reflection coefficients; Bodin *et al.* (2012) recovered estimates of 1-D shear wave velocity profiles from measured surface wave dispersion (SWD) and receiver function (RF) data, while Agostinetti & Bodin (2018) invert electrical resistivity and shear wave velocity. Of the foregoing, all but Bodin *et al.* (2012) and Agostinetti & Bodin (2018) use a fixed-dimensional MCMC sampler. We follow Bodin *et al.* (2012) and use a trans-dimensional MCMC sampler, where the number of model parameters is itself an unknown. Trans-dimensional MCMC was introduced to the geophysics literature by Malinverno (2002) and Sambridge *et al.* (2006).

In this work, we sample both the individual and joint Bayesian posterior probability density functions from MT and CSEM data

using a trans-dimensional Markov Chain Monte Carlo (MCMC) method based on the Metropolis–Hastings algorithm (Metropolis *et al.* 1953; Hastings 1970; Geyer & Moller 1994; Green 1995). The MT and CSEM data were collected during a marine survey offshore New Jersey (Fig. 2), whose objective was to map a freshwater aquifer in the shallow continental shelf. It is well-known that MT and CSEM data contain complementary information about subsurface conductivity. MT is more sensitive to conductive regions while CSEM is more sensitive to resistive features (Constable & Weiss 2006). We use a simple 1-D model to demonstrate this by illustrating the shapes of the regions of parameter space compatible with each data set individually, as well as the region compatible with the joint data set. This synthetic test demonstrates that the space compatible with the joint data set is within the intersection of the two individual regions. Applying our algorithm to the New Jersey field data, we identify a region of high resistivity consistent with a relatively freshwater aquifer and quantify—using the posterior distribution of probability density and the 90 per cent credible interval width—the model uncertainty reduction obtained by jointly inverting the data over inverting either of the data sets individually. Additionally, in conjunction with porosity data from nearby well logs, we also produce quantitative estimates of pore fluid salinity within the aquifer, with associated uncertainty.

2 MODEL PARAMETRIZATION AND FORWARD CALCULATIONS

We choose a 1-D parametrization of the Earth such that our model consists of k_s subsurface layer interfaces and k_s + 1 subsurface layers, each with an associated electrical resistivity. The subsurface model, then, consists of k_s interface depths $\mathbf{z_s} = [z_1, z_2, ...z_{k_s}]$ and k_s + 1 layer resistivities $\boldsymbol{\rho}_s = [\log(\rho_1), \log(\rho_2), ...\log(\rho_{k_s}), \log(\rho_{k_s+1})]$. The last layer is assumed to be a semi-infinite halfspace.

The resistivity of the seawater has a strong impact on the modelled CSEM data. Because it is a function of temperature, the seawater resistivity can vary significantly from ocean surface to seafloor in the shallow waters above the North American continental shelf. To avoid biasing the inversion, we model the water column as a series of layers rather than assume a water column resistivity profile. We model the water column in the same way we model the subsurface, as a vector of layer interfaces $\mathbf{z_w} = [z_1, z_2, ...z_{k_w}]$ and a vector of resistivities $\boldsymbol{\rho}_w = [\log(\rho_1), \log(\rho_2), ...\log(\rho_{k_w}), \log(\rho_{k_w+1})]$, except that the water column model terminates at the seafloor, rather than as a half-space. The combined model is formed by adding the water column model above the subsurface model, which begins at the seafloor and extends downward. The combined model, therefore, consists of $k_s + k_w + 1$ interfaces (the extra '1' counts the seafloor) and $k_s + k_w + 2$ layer resistivities.

The EM field components produced by a finite length, horizontal electric dipole are calculated using the open source code Dipole 1-D, described in Key (2009). The effect of the 336 m long transmitter wire was simulated by numerically integrating point dipoles distributed along the wire's length using an efficient Gauss quadrature approach. Forward modelling for MT data were done using the standard impedance recursion approach (Ward & Hohmann 1987).

3 JOINT INVERSION FRAMEWORK

The non-uniqueness of the inverse problem often results in significant model parameter uncertainty. This is particularly true of many EM methods in geophysics that are governed by a diffusion-type equation, where precise resolution of fine-scale structure is difficult. In particular, inversion of EM geophysical data often suffers from trade-offs between model parameters. The MT method experiences a trade-off between the thickness and conductivity of a conductive layer. The CSEM method has an opposite, resistivity-thickness trade-off. As a result, there is often a wide range of model parameter values that can fit a given field data set of one type equally well. Physical or other constraints on the values model parameters are allowed to take can limit the extent of this range, but joint inversion can go further.

3.1 Bayesian joint inversion framework

Bayesian probability describes how information provided by measured data modifies our prior assumptions to produce posterior probabilities on the model parameters. This process is described in Bayes' rule

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

where the variables to the left of | are conditional on those to the right. In other words, $p(a|b)$ is the probability of a given that b is known. $\mathbf{m}$ is a vector of parameters that constitute our model of the Earth (in our case, k_s, k_w, $\mathbf{z_s}$, $\mathbf{z_w}$, $\boldsymbol{\rho}_s$, and $\boldsymbol{\rho}_w$), while $\mathbf{d}$ is the vector of observed data.

The $p(\mathbf{m})$ term on the right-hand side of (1) is known as the prior and is a probability density function of the model parameters that represents all the information known and assumptions made about the model—independent of the measured data. In this study we assume uniform prior distributions on all model parameters (number of interfaces, interface depths and layer resistivities)—though the prior is only uniform when considering a fixed value of k. By using a minimally informative prior, we allow the information about the model parameters contained in the data to inform the posterior distribution of probability.

The $p(\mathbf{d}|\mathbf{m})$ term on the right-hand side is called the model likelihood, and is a measure of data fit. The likelihood is the probability that the modelled data differ from the measured data purely due to random measurement error. We use a simple chi-squared (χ^2) measure of misfit

$$\chi^2 = (\mathbf{d} - \mathbf{f}(\mathbf{m}))\mathbf{C}_d^{-1}(\mathbf{d} - \mathbf{f}(\mathbf{m})), \tag{2}$$

where $\mathbf{C_d}$ is the matrix of data covariances and $\mathbf{f}(\mathbf{m})$ is the forward modelling function applied to a model $\mathbf{m}$. This choice of misfit corresponds to a Gaussian likelihood function, justified by the use of stacked data:

$$p(\mathbf{d}|\mathbf{m}) \propto \exp\left(\frac{-\chi^2}{2}\right). \tag{3}$$

Finally, the product of the prior and the likelihood yields the posterior—a probability density function of the model parameters, written $p(\mathbf{m}|\mathbf{d})$, that reflects all the information about the model contributed by the measured data and our prior information and assumptions. Obtaining a good approximation to the posterior is the objective of sampling-based Bayesian inverse methods. For nonlinear inverse problems, this is done by sampling from the posterior, whose analytical form is unknown.

Algorithms have been developed to draw samples from unknown distributions. One of the more successful—due to its simplicity and robustness—is MCMC, which we utilize in this study to invert CSEM and MT data. Our method is a trans-dimensional variant of the Metropolis–Hastings–Green MCMC sampler (Metropolis *et al.*

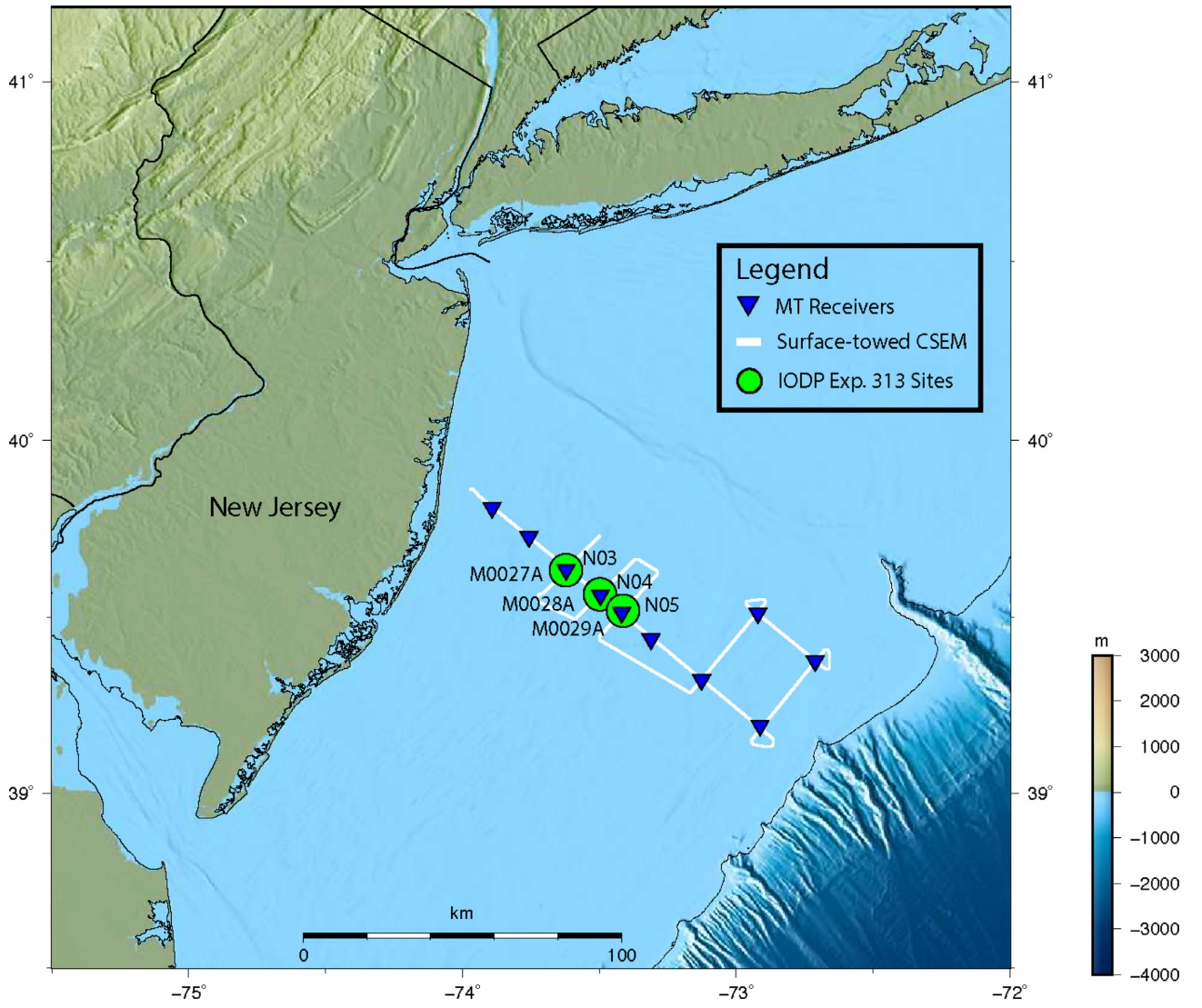

Figure 2. Map of a marine EM survey offshore New Jersey (upper-left-hand corner). The survey included 10 seafloor MT stations (blue squares) and over 200 line km of CSEM data, recorded on four receivers towed on the surface behind the ship. The main tow-line, extending from near the NJ shoreline out to near the edge of the continental shelf, was chosen to cross the locations of three recent IODP drill sites (Expedition 313—green circles), for which porosity and salinity data are available (Mountain *et al.* 2010).

1953; Hastings 1970; Green 1995), meaning that the number of model parameters is itself a parameter that we invert for, and its selection is guided by the data. Here we adopt the 'birth-death' scheme (Geyer & Moller 1994) whereby, at each step, the number of interfaces can increase by one, decrease by one, or remain the same. For more robust and accelerated convergence, we implement parallel tempering (PT), whereby multiple Markov chains are run in parallel, each with its own 'temperature' T. The chains at $T = 1$ are unmodified and will make up the model ensemble; the chains at temperatures $T > 1$ explore a 'tempered' model space and, by swapping models with colder chains after each MCMC step, allow for more robust and accelerated convergence to the high probability regions of model space. For a detailed description of our algorithm see Blatter *et al.* (2018), Ray *et al.* (2013a) and Ray & Key (2012); for a helpful discussion of MCMC, see Gilks *et al.* (1995); and for a thorough discussion of PT with numerical examples, see Sambridge (2013).

The algorithm used in this work is the same as in Blatter *et al.* (2018), except that here there is a water layer, which is sampled separately from the subsurface. At each step of the algorithm, a uniform random number is generated. If it is greater than β (a pre-determined constant which has a value between 0 and 1), the subsurface model only is perturbed; otherwise, the water layer only is modified. Because the water layer in the New Jersey survey is shallow (20–100 m) relative to the depths of interest in the subsurface (the uppermost kilometer of the crust), we chose $\beta = 0.2$ to prioritize sampling the subsurface.

Although MT and CSEM data contain fundamentally different information about subsurface electrical resistivity, only the relative uncertainty associated with a measurement is relevant in a joint Bayesian inversion framework, not the 'type' of data it represents. This is true regardless of the relative information content of the data. This makes incorporating multiple data sets in a joint Bayesian inversion a straightforward process, where $\mathbf{d}$ and $\mathbf{C}_d$ are simply extended to include the new data, with their associated uncertainties. When measurement errors between two data sets are uncorrelated, eqs (2) and (3) can be used to show that the joint likelihood is simply

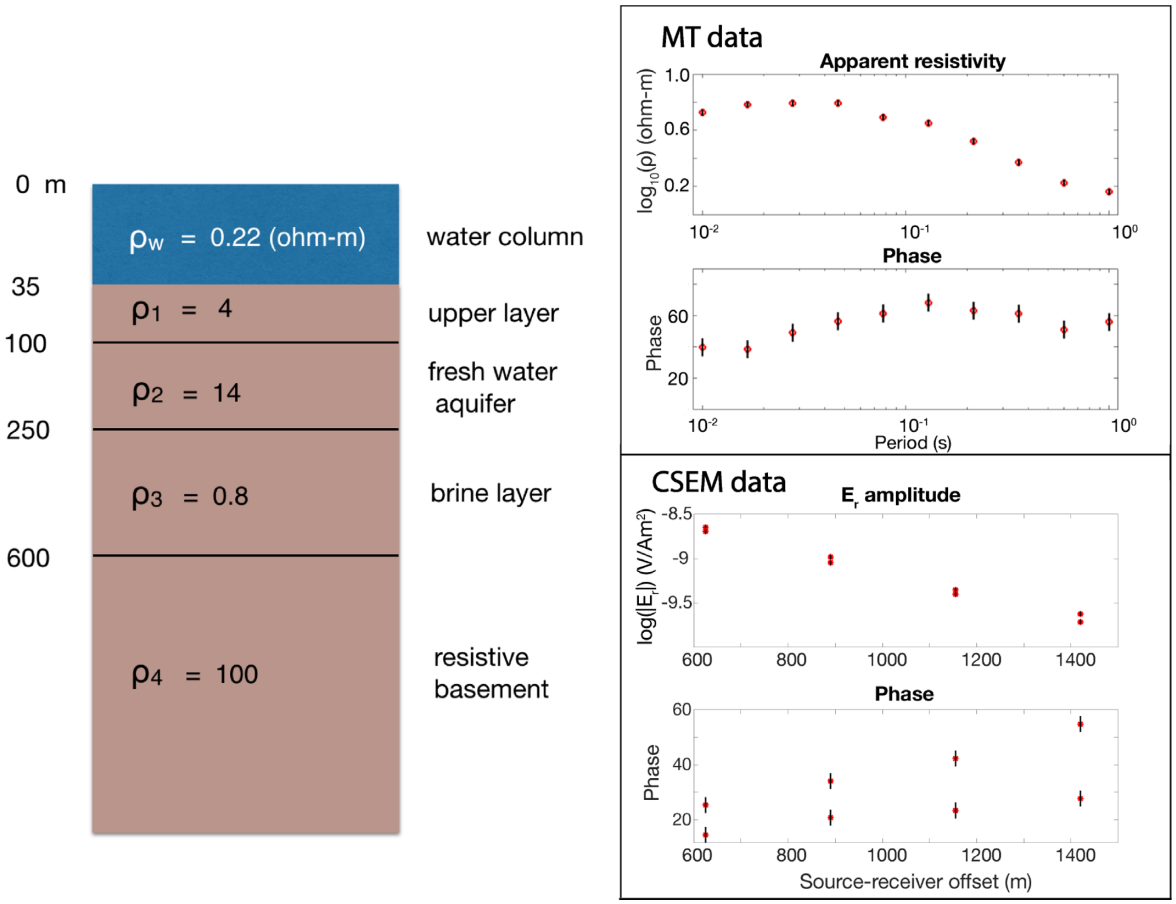

Figure 3. True model (left) from which synthetic MT (upper right) and CSEM (lower right) data were generated. The CSEM frequencies inverted for were 0.75 and 1.75 Hz. The synthetic data include 5 per cent Gaussian noise. The true model was chosen to simulate the nearly 1-D geology in the region where the New Jersey marine survey was conducted. The synthetic data were inverted using a modified, fixed-dimension version of the MCMC sampler. The water column depth and thickness, as well as the number and depths of the subsurface interfaces were fixed during the inversion, so that only the four subsurface layer resistivities ($\rho_1 - \rho_4$) were inverted for.

proportional to the product of the individual likelihoods:

$$p(\mathbf{d}|\mathbf{m}) = p(\mathbf{d_1}|\mathbf{m})\, p(\mathbf{d_2}|\mathbf{m}), \tag{4}$$

where the joint data vector $\mathbf{d} = [\mathbf{d_1}; \mathbf{d_2}]$.

Because of the likelihood function's insistence that only the data error determine a datum's weight in the inversion process, it is imperative that the data covariance matrix, C_d, be accurately estimated. Indeed, both the degree of posterior model complexity (Agostinetti & Malinverno 2010) and the posterior model parameter variance (Guo *et al.* 2014) are to a large extent determined by the data errors. Estimating C_d involves determining the data variances and the degree and nature of data error correlation.

The data variances (the diagonal elements of C_d) were determined through standard MT (Egbert 1997) and CSEM (Myer *et al.* 2010) data processing methods. The CSEM data inverted in this study are the log-amplitude of the in-line electric field (in the frequency domain) and its phase; the MT data inverted are the log of the apparent resistivity (also frequency domain) and its phase. While the errors in such data may not be normally distributed, the departure from Gaussianity is negligibly small if the data errors are small relative to the magnitude of the data themselves (Wheelock *et al.* 2015, suggest 10 per cent relative error as a reasonable threshold). Here we assume the data errors are normally distributed.

Although Bayes' rule admits no differentiation of data based on 'type', there is a concern that the measurement errors estimated from data processing workflows do not always accurately reflect the total data error. For example, inaccuracy due to forward modelling or assuming 1-D or 2-D physics also contribute to the overall discrepancy between forward modelled and measured data. To prevent overly optimistic error estimates from producing spurious or overly complex model structure, we used a relative error floor of 1 per cent for CSEM data and 5 per cent for MT data.

It is possible, in a Bayesian context, to go one step further and posit that, even after careful data processing and error estimation, some of the information about the data errors remains unknown. Usually, this is done by introducing a scale factor that multiplies C_d and which can be estimated through maximum likelihood (Mecklenbrauker & Gerstoft 2000; Dosso & Wilmut 2006; Sambridge 2013; Ray *et al.* 2016) or through Hierarchical Bayes methods (Bodin *et al.* 2012; Malinverno & Briggs 2004; Agostinetti *et al.* 2015). In the case of joint inversion, there is usually a different scale factor applied to each data set separately (Agostinetti & Bodin 2018). Under the Hierarchical Bayes approach, these scale factors are hyperparameters inverted for, allowing the data and hyperparameters to select the optimal contribution for each data set. Hierarchical Bayes is similar in many respects to the weighting terms introduced

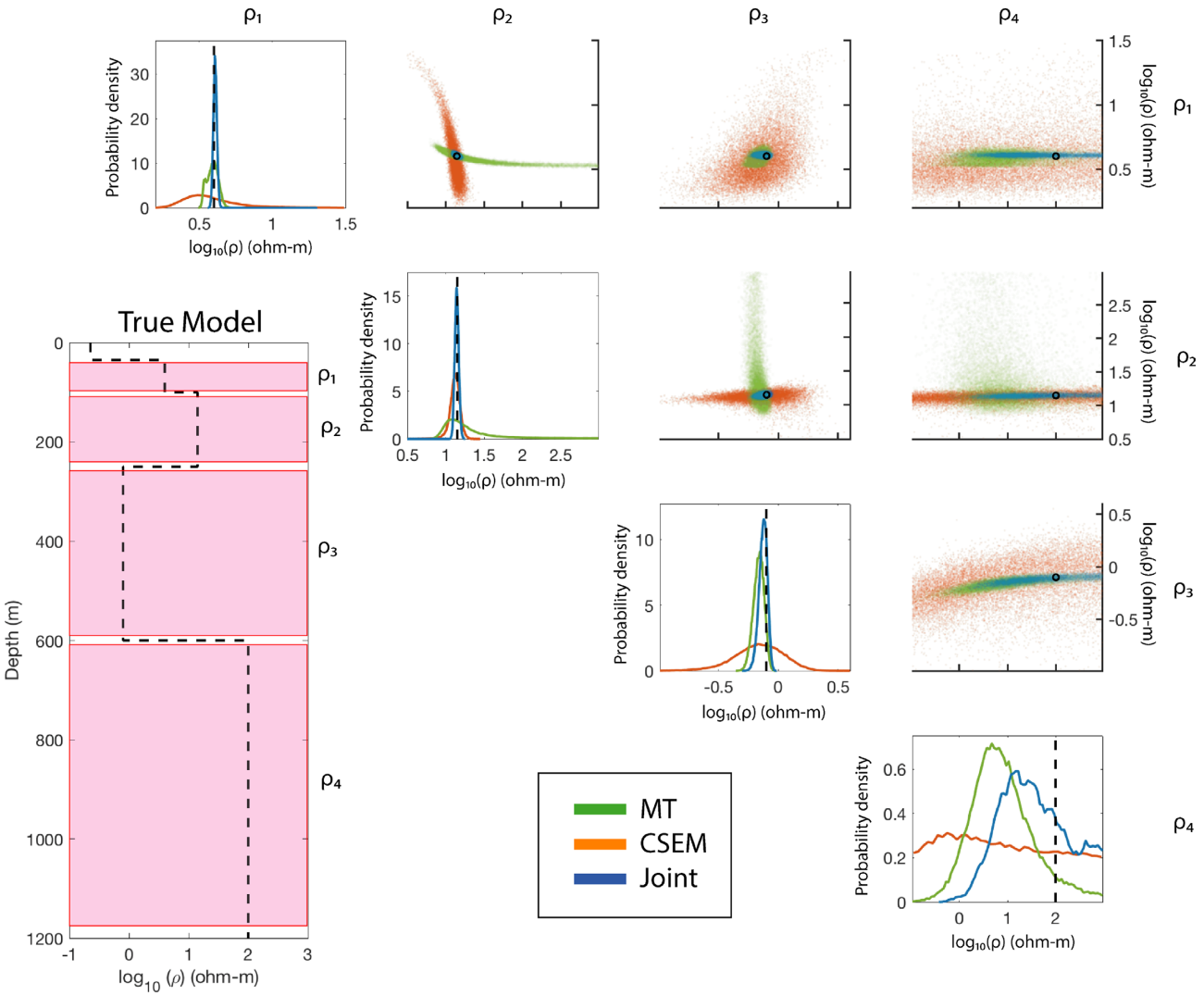

Figure 4. Scatter plots of pairs of model parameters from 70 000 randomly selected models from the posterior model ensembles of individually inverted MT and CSEM data (green and orange points, respectively) as well as a joint inversion of both data sets (blue points). The true value is indicated in each off-diagonal plot by the black circle. The plots along the diagonal are the model parameter probability distributions, with the true value indicated by the black dashed line. In each column, the *x*-axis of each plot corresponds to the x-axis of the probability-distribution plot in that column, while in each row the *y*-axes of the off-diagonal plots are the same as the *y*-axis of the final plot in that row.

in standard, linearized joint inversion to weight the relative contribution of each data set to the objective function—the main difference being that in the latter case the choice of what values to assign each weight is largely arbitrary (Commer & Newman 2009; Key 2016).

We chose not to use Hierarchical Bayes to estimate a scale factor weighting the MT against the CSEM data as there is evidence, in the case that the data errors have been robustly estimated, that the posterior distribution estimated via Hierarchical Bayes converges to that obtained using standard trans-dimensional MCMC (Agostinetti & Malinverno 2018). In such cases, if the data errors are normally distributed, C_d is diagonal, and the data variances have been robustly estimated, then the maximum likelihood estimate for a scale factor is equal to the rms of the residuals, and should be close to unity (see the Appendix for a derivation). We subsequently show that the distribution of RMS misfit across all models in the joint ensemble reaches its maximum value in the neighbourhood of one, for both data sets individually and for the joint data set. Sampling for data error scale factors using Hierarchical Bayes would therefore not significantly affect the joint posterior probability distribution in our case.

One final issue in estimating C_d involves the problem of correlated noise (the off-diagonal elements of C_d). Although it is standard practice to assume independent, normally distributed noise, it is likely that the measurement error is correlated to some degree. To account for this, one could assume a Gaussian, exponential, or other parametric correlation between the data noise at different offsets or frequencies (Bodin *et al.* 2012; Xiang *et al.* 2018). Another approach is to sample the posterior residuals to estimate a non-parametric correlation (Ray *et al.* 2013b). While we acknowledge the importance of accurately estimating the data error statistics, the choice of which scheme to pursue to account for data error correlation is not obvious. Indeed, overly complex or unjustified modelling of error correlations can lead to spurious inverted model structure (Xiang *et al.* 2018). Given the lack of a definitive alternative, we follow the standard choice in the literature and assume C_d is diagonal.

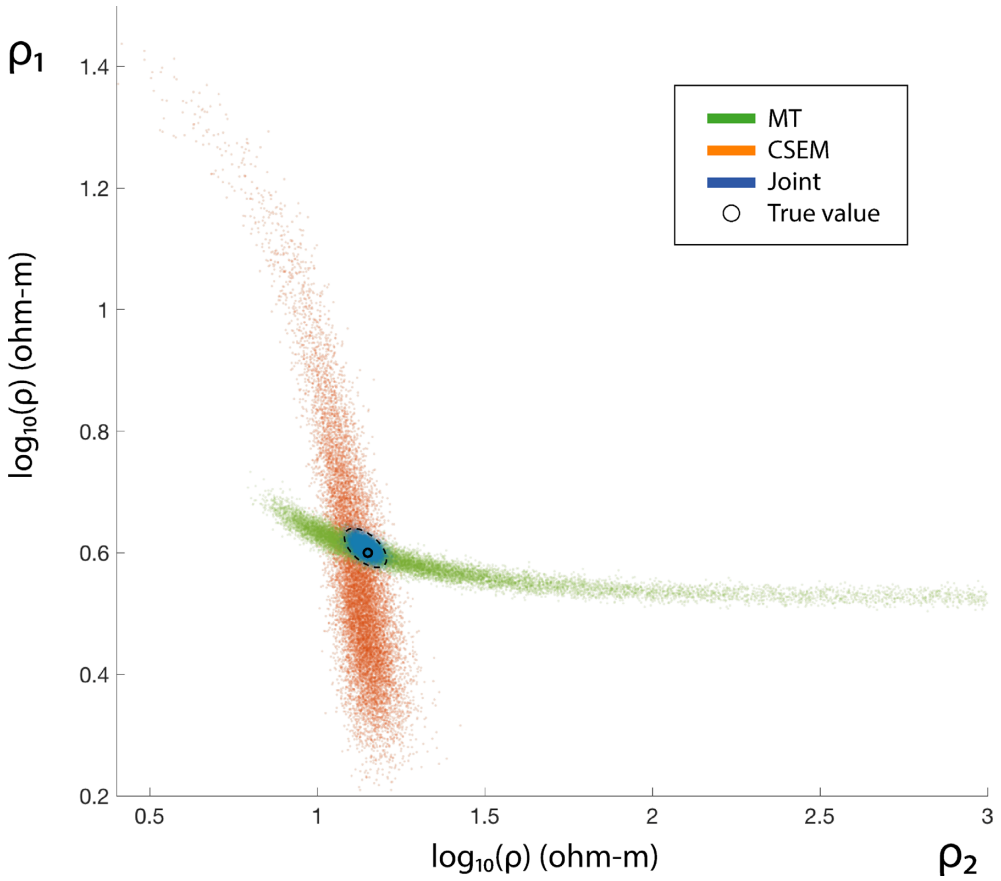

Figure 5. Larger version of a single plot from Fig. 4. The orange and green dots indicate possible parameter estimates that gradient-based inversion of only the MT or CSEM data, respectively, might produce, depending on the starting model, choice of regularization and other parameters. A joint inversion of both data sets, however, would be restricted to searching within the region highlighted by the black dashed oval, ensuring a model estimate much closer to the true value (small black circle).

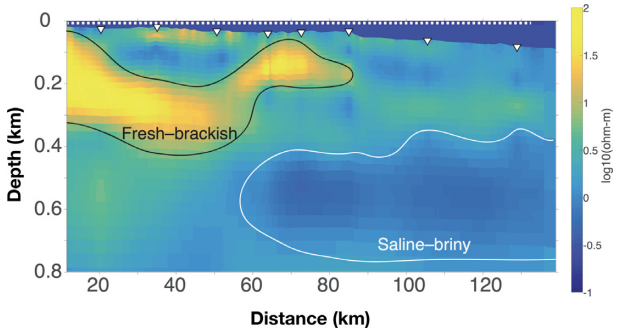

Figure 6. 2-D joint inversion of the main New Jersey tow-line (white dashed line) obtained using the freely available inversion code MARE2D. See Gustafson et al (accepted). White triangles are MT sites as indicated in Fig. 2. Warm colours indicate resistive structures, while cooler colours indicate more conductive features. The shallow resistor (indicating the presence of the fresh-brackish water aquifer) is clearly visible in the upper 400 m, out to a distance of 90 km along the tow-line. The highly conductive region (brine-saturated sediments) begins approximately 60 km offshore, at a depth of about 500 m.

4 SYNTHETIC TEST

To demonstrate how joint inversion reduces the size of parameter space compatible with the data, we constructed a simple toy model (see Fig. 3) consisting of a shallow water layer above three subsurface layers and a half-space. This model was designed to emulate the expected resistivity profile of the upper kilometer of the continental shelf, based on 2-D regularized inversions of the New Jersey data by Gustafson et al (accepted). We generated synthetic MT and surface-towed CSEM data from this model by adding 5 per cent Gaussian noise to the model responses. We assumed an isotropic Earth, an assumption discussed in Gustafson et al. The frequencies and other parameters used were taken from the field data set to keep the synthetic data realistic. We then inverted this synthetic data using a version of our MCMC code modified to keep the number and depth of the layer interfaces, as well as the thickness and resistivity of the water column, constant—that is, fixed at their true values. In other words, only the resistivities of the subsurface layers were allowed to vary. The MT and CSEM data were inverted separately and jointly.

To visualize the size and shape of the trade-offs inherent in the (non-linear) inversion of these synthetic data, we selected models at random from the posterior model ensemble from each inversion (individual and joint) and made scatter plots of each pair of model parameters—(ρ_1, ρ_2), (ρ_1, ρ_3), etc. Fig. 4 shows the scatter plots and marginal model parameter probability distributions arranged like a covariance matrix. The true model is plotted in the lower left of Fig. 4, with each layer labeled and highlighted. A larger version of the (ρ_1, ρ_2) plot is shown in Fig. 5.

A few conclusions can be drawn from Fig. 4. First, the synthetic MT and CSEM data do indeed contain complementary information

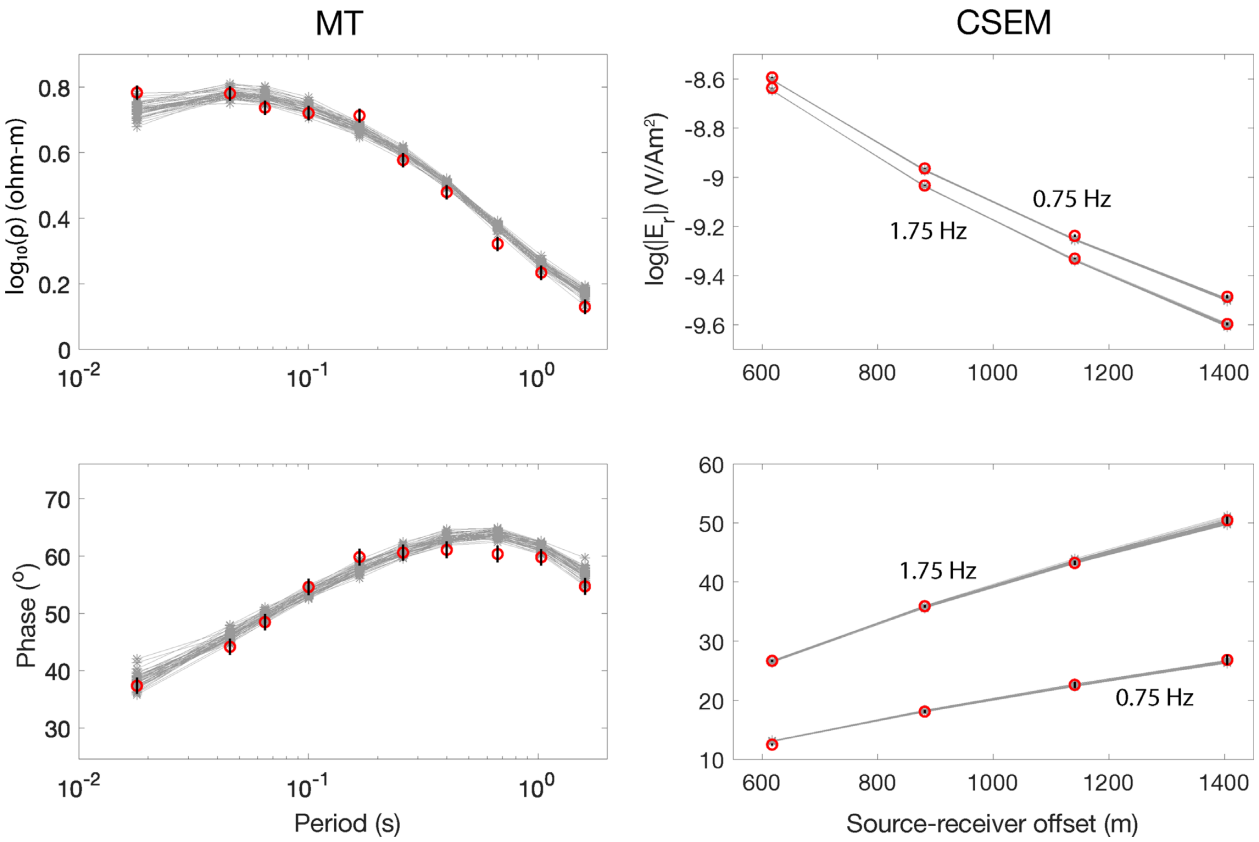

Figure 7. Field data collected at location N05 (Fig. 2). MT apparent resistivity (top) and phase (bottom) are plotted on the left, while the CSEM inline electric field amplitude (top) and phase (bottom) are plotted on the right. The measured data are plotted in red (error bars in black), while the forward responses of 50 models randomly selected from the joint model ensemble are plotted in grey.

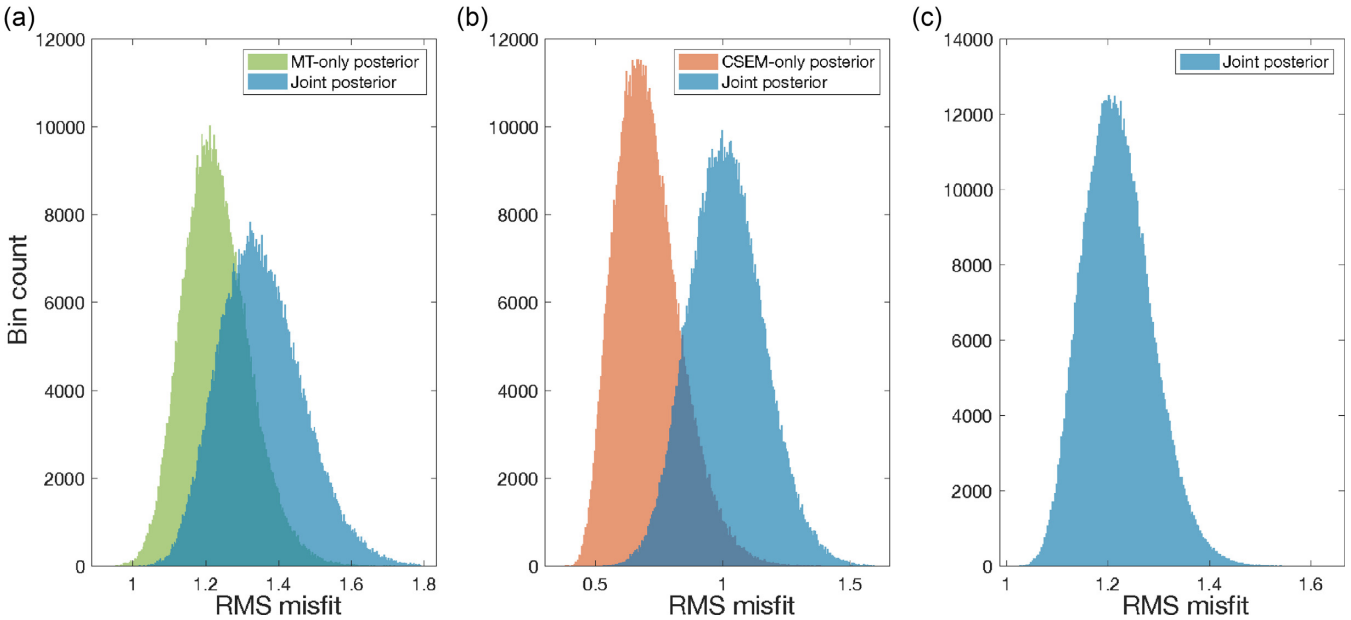

Figure 8. RMS misfit for the data at location N05 across all models in the MT-only (green), CSEM-only (orange), and joint (blue) ensembles. (a) shows the distribution of fit to the MT data; (b) shows the distribution of fit to the CSEM data and (c) shows the distribution of fit to the joint data. The data fit to each individual data set deteriorates somewhat when the models in the ensemble are required to satisfy both data sets simultaneously. Compare with Table 1.

about the subsurface model parameters, as indicated by the differing shapes and locations of the green and orange point clouds. The extent of their intersection is, in all cases, less than their individual extents. Secondly, it appears that the region of model space compatible with the joint data set is indeed roughly equivalent to the intersection of the regions of model space compatible with the individual data sets, and in some cases appears to be even smaller. Furthermore, examining the model parameter variances further indicates that the range of acceptable values for each model parameter has indeed been reduced by joint inversion—in some cases, by a great deal.

While Fig. 4 is arranged in the style of a covariance matrix, it is important to recognize that it represents more information than a covariance matrix can provide. The latter boils down the wealth of information displayed in each plot of Fig. 4 to a single number. While this approximation may be adequate locally (e.g. in the neighbourhood of a particular dot in Fig. 5), it clearly does not hold far from that model, as indicated by the complex shapes traced out in Fig. 4.

5 FIELD DATA INVERSIONS

We now illustrate these points by applying the full, trans-dimensional MCMC sampler to a field data set. The New Jersey EM data set (see Fig. 2) is part of a larger freshwater marine EM data set that was collected to understand the spatial extent of low salinities observed in a series of nearby wells. Surface-towed CSEM data was collected along a 130-km-long profile across the continental shelf, beginning near the New Jersey shoreline and ending at the shelf edge. In addition, eight seafloor MT receivers were deployed roughly 10–20 km apart to constrain the broad scale structure and better resolve the more conductive sediments. High quality MT data were collected from 10^{-4} to 10^{2} Hz—unusually high for marine MT data. The skin depth, $z_s \approx (500\,\mathrm{m})\sqrt{\rho/f}$, is the depth at which an EM signal of frequency f will have attenuated to $1/e$ of its original amplitude within a homogeneous medium of resistivity ρ. This relation indicates that the thicker the conductive sea water, the more attenuation the MT signal will suffer before reaching the subsurface. For instance, at 1 Hz in 0.2 ohm-m seawater, the skin depth is $\approx$ 220 m. The relatively shallow waters of the continental shelf (18–100 m) allowed higher-than-normal frequencies to penetrate the water column. Here we invert only the data from 10^{-1} to 10^{2} Hz, as we are primarily interested in the upper kilometer of the subsurface. The surface-towed CSEM system featured a 336-m-long electric dipole antenna broadcasting a 0.25 Hz fundamental frequency waveform. Four broad-band receivers measuring the inline horizontal electric field were towed behind the antenna at offsets of 600, 870, 1120 and 1380 m. Further details on the survey design, data collection, processing, and inversion can be found in Gustafson et al (accepted).

Regularized, deterministic 2-D joint inversions of the data (Fig. 6) obtained using the freely available code MARE2D (Key 2012) reveal a laterally continuous, 90 km long, 10–110 ohm-m resistive layer within the upper 400 m (Gustafson et al, accepted). At around 60 km offshore, this resistive zone becomes shallower, occupying depths of 50–200 m. It continues for roughly 30 km more before petering out. A conductive zone ($<$1 ohm-m) was identified that begins near the centre of the survey line and extends out to the shelf edge at depths beneath the shallower resistive layer (450–750 m). Although there are 2-D features in the gradient-based inversion results of the New Jersey survey data, these variations are on the order of tens to hundreds of meters vertically over a length scale of some 120 km, justifying the use of a 1-D layered earth model.

Table 1. Average rms misfit across all models in the CSEM-only, MT-only and joint posterior model ensembles for data collected at location N05, arranged by data type.

Data type	CSEM-only	MT-only	Joint
MT	—	1.23	1.35
CSEM	0.71	—	1.02
Total	0.71	1.23	1.22

We used the full, trans-dimensional MCMC sampler to invert the MT data collected offshore New Jersey, as well as the surface-towed CSEM data from the 130-km-long main line. All 8 MT stations (stations N01-N08) were inverted individually for 1-D models. The surface-towed CSEM data from the main tow line were divided into groups of 15 adjacent soundings (no overlap between groups) and averaged. These average soundings were then inverted individually for 1-D models. In addition, each of the averaged CSEM soundings was inverted jointly with the data from the nearest MT station. For each inversion, either joint or individual, we allowed the MCMC sampler to draw one million samples from the posterior distribution. We then removed the first 300 000 samples from each chain (known as the 'burn in') and combined the remaining models from all of the PT chains at T=1 to form the posterior model ensemble. We ran each inversion with eight PT chains, three of them at T=1 (for a total combined posterior ensemble of $3 \times 700\,000 = 2.1$ million models). The trans-dimensional MCMC algorithm was allowed to place interfaces down to a depth of 3000 m, but here we only show the upper 1200 m since our data sensitivity and target of interest are in this region.

All the MCMC inversions were computed in serial (no parallelization—either over the different PT chains or during the forward calculations—was implemented). The MT-only inversions each took approximately 2 hr to complete, while the CSEM-only and joint inversions each took approximately 1.5 d. The computation time could be greatly reduced by implementing parallel computing strategies across PT chains and in the forward calculations (e.g. across frequencies).

The MT and CSEM data from location N05 (see Fig. 2) are plotted in Fig. 7. Also plotted in Fig. 7 are the forward responses of 50 randomly selected models from the joint inversion model ensemble. For each data set, and for each model ensemble (CSEM-only, MT-only and joint), we computed the average RMS misfit to the data (Table 1). Histograms of RMS misfit to the MT and CSEM data for each model in their respective model ensembles are shown in Fig. 8.

As might be expected, the joint ensemble is less able to fit the individual MT and CSEM data sets relative to the MT-only and CSEM-only model ensembles (as indicated by the orange histograms in Fig. 8). Just because a model is able to fit one data set very well, however, does not mean that its estimates of subsurface model parameters are necessarily the most accurate. Indeed, when complementary information from a second data set is added, it becomes clear that the model parameter estimates derived from the first data set were based on an incomplete picture of the subsurface. To demonstrate this, we randomly selected 50 models from the CSEM-only and MT-only model ensembles for the data at location N05. We then computed the MT forward responses of the models derived from the CSEM data, and the CSEM forward responses of the models derived from the MT data, and plotted them against the measured MT and CSEM data (Fig. 9). Comparing Fig. 9 to Fig. 7 indicates the

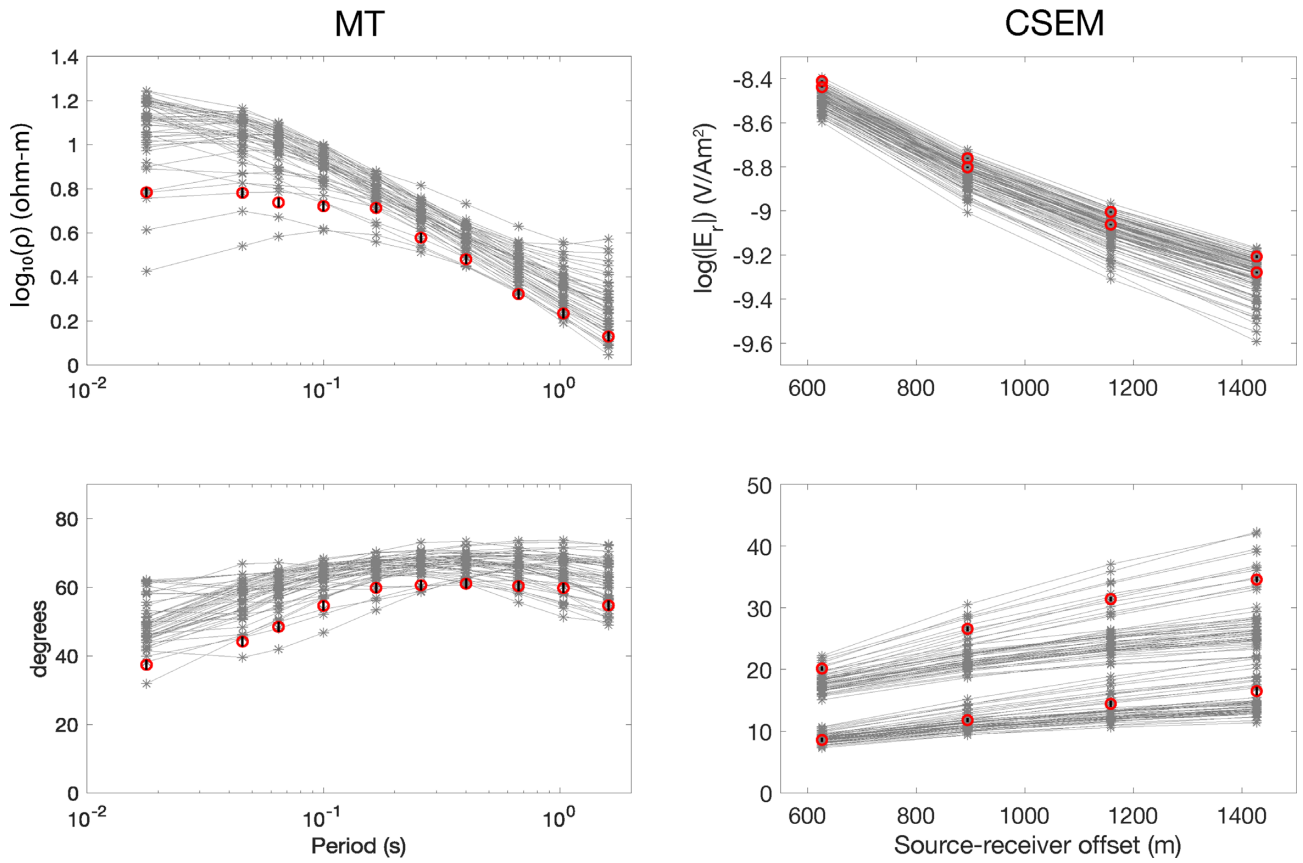

Figure 9. MT and CSEM data from location N05 (compare with Fig. 7). The measured data are plotted in red (error bars in black). The MT forward responses of 50 models randomly selected from the CSEM-only posterior model ensemble are plotted on the left, while the CSEM forward responses of 50 models randomly selected from the MT-only posterior model ensemble are plotted on the right. The poor fit of these forward responses indicates that many of these models will not be compatible with the joint data set.

degree to which joint inversion constrains the region of parameter space compatible with the data. In particular, it is interesting to note that the CSEM-only models tend to overpredict the apparent resistivity while the MT-only models tend to underpredict the electric field amplitude (suggesting that these models are underestimating the subsurface resistivity). This does not, however, prevent them from fitting their respective data sets better than the models in the joint model ensemble. In fact, the CSEM-only models tend to overfit the CSEM data, while the MT-only models tend to under-fit the MT data.

Ultimately, the goal of geophysical inversion is to obtain estimates of model parameters—in this case, the electrical resistivity of the upper kilometer of the continental shelf offshore New Jersey. Gradient-based methods provide a single estimate of each model parameter without assessing the trade-offs inherent in inversion and illustrated in Figs 4 and 5. The trans-dimensional MCMC sampler, by contrast, produces a posterior model ensemble, from which an estimate of the posterior distribution of probability density can be estimated.

Fig. 10 shows the marginal posterior probability density for resistivity as a function of depth for the CSEM-only, MT-only and joint model ensembles at location N05. At each depth, the marginal distribution of probability density across resistivity is indicated by colour: warmer colours indicate higher probability density and cooler colours indicate lower density. The width of the warm colour region at a given depth is indicated by the left and right red lines, which indicate the 5th and 95th percentiles of the distribution, respectively, and bound the 90 per cent credible interval. This means that 90 per cent of the models in the ensemble have electrical resistivity values that fall between the red lines, making the 90 per cent credible interval a reasonable proxy for how well the data can constrain the model.

The geological interpretation of Fig. 10(c) is as follows: conductive seawater (≈40 m deep) followed by a thin (≈40 m thick), conductive layer of shallow, seawater-saturated sediments; a resistive layer (≈175 m thick) of sediments filled with relatively fresh water; another layer of relatively conductive, seawater-saturated sediment (≈175 m thick); and a thick (≈400 m) layer of highly conductive, briney sediments. Beneath the thick conductive layer, the subsurface appears to become increasingly resistive with depth as the sensitivity of the data to the subsurface resistivity decreases. This interpretation, based on the posterior, is corroborated by 2-D inversions of the data by Gustafson et al (accepted); see also the white line in Fig. 10(c).

The posterior distribution obtained from the CSEM data alone (Fig. 10a) indicates a much simpler structure: namely, the resistive freshwater aquifer and the conductive brine layer. Of the two, the CSEM data appears more sensitive to the presence of the resistor than the conductor, as indicated by the narrower extent of the 90 per cent credible interval over the depth range of the resistor. By contrast, Fig. 10(b)—obtained by inverting only the MT data—indicates that the MT data is preferentially sensitive to conductors. Clearly indicated in the MT-only posterior distribution are

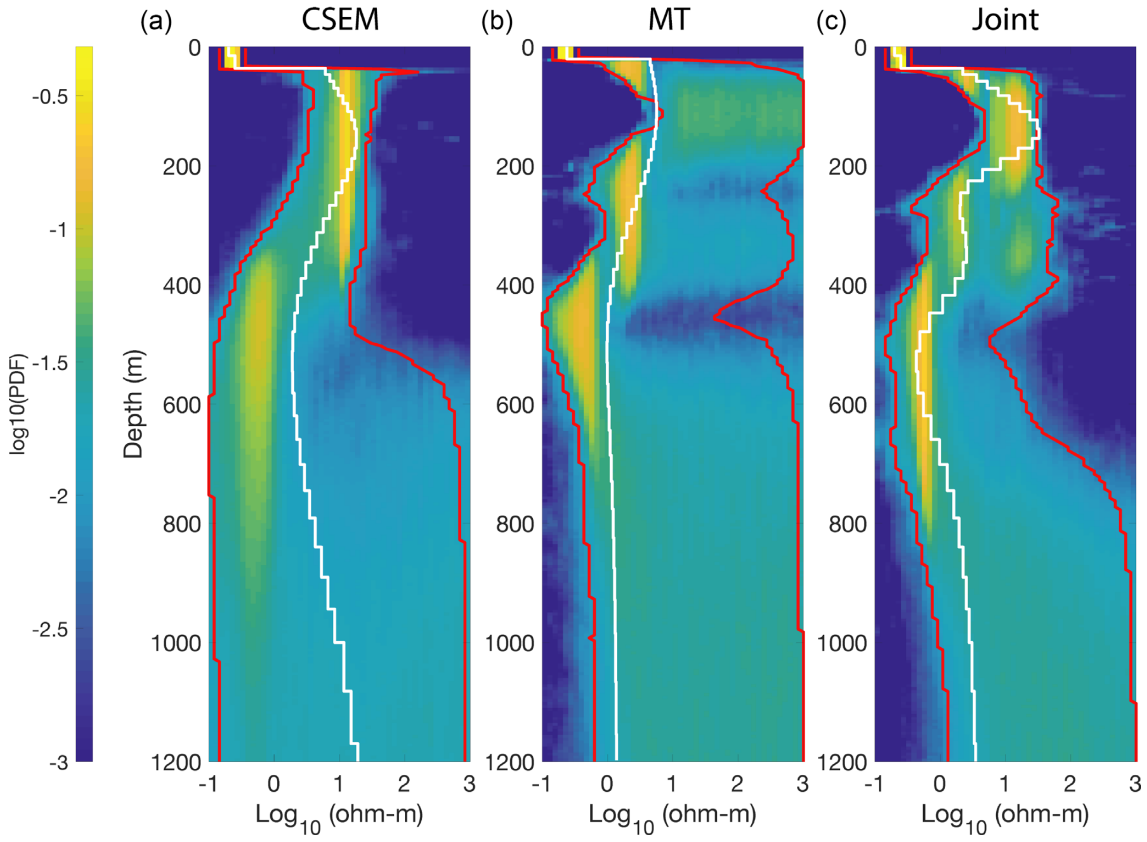

Figure 10. Posterior probability density functions obtained from CSEM-only (a), MT-only (b) and joint (c) inversions of EM data acquired at location N05 (see Fig. 2). The warmer colours indicate higher probability density, the cooler colours lower probability density. The left-most and right-most red lines represent the 5th and 95th percentiles, respectively, of the posterior PDF at each depth. The white line is a 1-D profile from the respective gradient-based 2-D inversions of Gustafson et al (accepted), plotted for reference. The CSEM data is primarily sensitive to the shallow resistor, while the MT data has greater sensitivity to the more conductive layers. The joint inversion, meanwhile, better resolves both the resistor and the conductors.

the upper and lower seawater-filled layers and the highly conductive brine layer. The resistive layer, although present, is very poorly constrained; the MT data appears to be able to place only a lower bound on the freshwater aquifer's resistivity.

That the joint posterior resolves all four subsurface layers indicates that complementary information from each data set has altered and constrained the regions of parameter space compatible with the data. To further illustrate this we computed, for each model in the three ensembles, the median resistivity over depth ranges consistent with the four layers described earlier: the uppermost conductive sediments (seafloor-60 m); the freshwater aquifer (80–200 m); the lower conductive sediments (225–375 m); and the highly conductive brine-saturated sediments (400–750 m). We labeled these median resistivities ρ_1–ρ_4 and made covariance-like scatter plots and variance-like probability distributions (Figs 11 and 12), similar to the synthetic examples shown previously.

While the trans-dimensional nature of the MCMC algorithm prevents a direct comparison of Figs 11 and 12 to Figs 4 and 5 (where the number of layers and their thicknesses was held constant), the main conclusion remains the same: the region of parameter space compatible with the joint data set is roughly equivalent to the overlap between the regions compatible with the data sets separately. In Fig. 12, for example, the resistivity of the uppermost layer trades off with that of the freshwater aquifer in the models obtained from the CSEM data only; as the freshwater aquifer becomes more resistive, the uppermost layer must become more conductive to compensate and still fit the CSEM data. In the meantime, while the MT data is relatively insensitive to the freshwater aquifer resistivity, it is much better able than the CSEM data to uniquely constrain the resistivity of the uppermost layer. This, in turn, allows the joint data set to constrain the resistivity of the freshwater aquifer to the portion of the trade-off curve that overlaps with the MT data's knowledge of the uppermost layer resistivity.

Finally, the contrast between the posterior distributions in Fig. 10 and the model estimates obtained by inverting the same data using gradient-based methods (the white lines in Fig. 10) is illustrative of the ways in which Bayesian sampling methods can improve our understanding of model parameter uncertainty. A common method of estimating the uncertainty of model parameter estimates obtained from gradient-based inverse algorithms assumes that the distribution of probability for each model parameter is normal with mean equal to the gradient-based parameter estimate. Yet at ≈ 100 m depth in Fig. 10(b), a normal distribution of probability density centred at the white line clearly fails to capture the uncertainty indicated by the Bayesian posterior distribution, no matter the variance. Likewise, a normal distribution centred at the gradient-based model estimate in Fig. 10(c) at ≈375 m depth fails to capture the bimodal nature of the Bayesian posterior. We found this second mode—at resistivities >10 ohm-m—to be robust in the sense that it is present in the vast majority of models in the joint ensemble. Seismic data show reflections at ≈370 and 400 m (see Gustafson *et al*. accepted), indicating the presence of another resistive layer. Yet this layer would be invisible to any linearized uncertainty estimate.

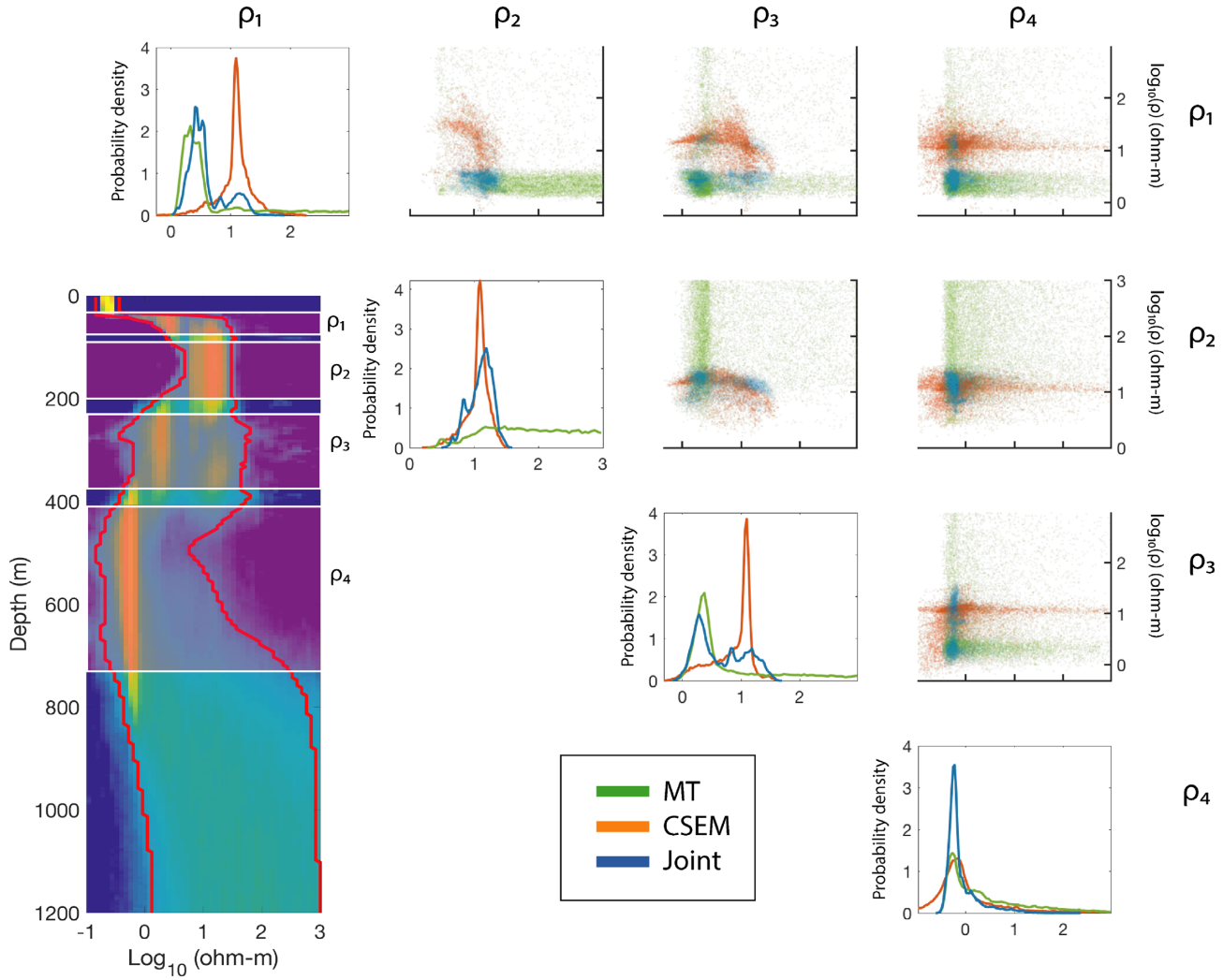

Figure 11. Off-diagonal: scatter plots of median resistivity (calculated over the depth intervals shown in the lower-left) for ≈ 70 000 models randomly-selected from the MT-only, CSEM-only, and joint posterior distributions obtained from the data at location N05. The joint posterior distribution is shown (lower-left). Each point represents a model from its respective ensemble: MT-only (green); CSEM-only (orange); joint (blue). Diagonal: distributions of probability density for each median resistivity. Compare with Fig. 4. The *x*-axes of the plots in a given column are identical; the *y*-axes of the plots in a given row are identical with the *y*-axis of the final plot in that row. As in Fig. 4, the region of parameter space compatible with the joint data set (blue) appears to be restricted to the overlap between the regions compatible with the data sets individually.

6 QUANTIFYING UNCERTAINTY IN PORE FLUID SALINITY

It is often the case that physical properties other than bulk electrical resistivity are of interest. In the case of this study, the salinity of the pore fluids in the aquifer is desired since it can be used to characterize the aquifer as a potential resource and as a key to explaining the water's provenance through palaeo-hydrological modelling (Cohen *et al.* 2010). Archie's Law (Archie 1942) relates bulk resistivity, ρ_b, and other rock parameters such as porosity, ϕ, to the resistivity of the pore fluids, ρ_f

$$\rho_f = \rho_b \phi^m, \tag{5}$$

where m is an empirically derived constant that is related to the connectedness of the pore spaces. Salinity can, in turn, be computed from ρ_f if temperature is also known. Here, we assumed a linear temperature profile as a function of depth from the seafloor down

$$T(z) = T_0 + \alpha z, \tag{6}$$

where $T_0 = 11.4\,^\circ$C and $\alpha = \frac{3}{100}\frac{^\circ\text{C}}{\text{m}}$. We used the well-known 1978 Practical Salinity Scale (PSS-78) conversion from temperature and electrical resistivity to salinity (Perkin & Lewis 1980).

Because gradient-based inversions produce single estimates of each model parameter, they can produce only single estimates of related parameters—in this case pore fluid salinity. Bayesian inversion, by contrast, produces an ensemble of models which fit the data and which are sampled according to the posterior probability density. Consequently, an ensemble of estimates of related parameters can be produced, from which quantitative uncertainty estimates can be obtained that reflect the underlying uncertainty in the inverted model parameters. In addition, if measurements of other relevant parameters for which there is uncertainty are available, such as porosity measurements from a well log, these can be further incorporated into the uncertainty estimate.

We utilized a straightforward Monte Carlo approach to obtain uncertainty estimates of pore fluid salinity in the upper several hundred meters of the subsurface along the main tow line offshore New

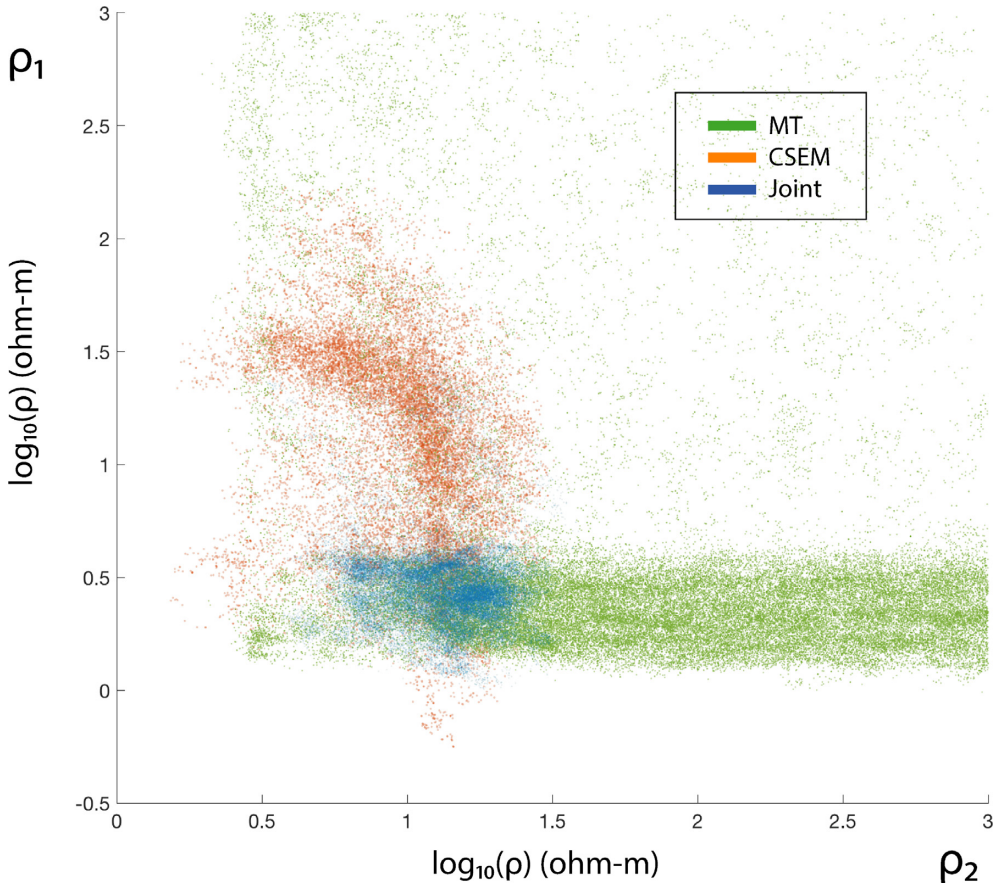

Figure 12. Larger version of a single plot from Fig. 11 (location N05). As in Fig. 5, the complementary nature of the information contained in the MT and CSEM data sets allows the joint data set to better constrain the resistivity of the both the upper layer of sediments (ρ_1) and the freshwater aquifer (ρ_2).

Jersey. First, we chose CSEM and MT soundings nearest to the wells M0027A, M0028A and M0029A, for which we have recent porosity and salinity measurements at varying depths (Expedition 313 Scientists 2010; Mountain *et al.* 2010), and inverted them to obtain a posterior model ensemble. Next, we defined a probability density distribution for the cementation factor, m: lacking field measurements, we assumed m is normally distributed with mean of 2.0 and stand deviation of 0.2. We also assume the well log porosity ϕ is normally distributed with standard deviation $\sigma_\phi = 0.05$ (see Fig. 13). Finally, for each well log, we executed the following procedure:

(i) for each porosity measurement in the well log (each at a unique depth), repeat the following steps N times:
(ii) draw a random value of ρ_b from the posterior distribution by randomly choosing a model from the ensemble and identifying its resistivity at the depth of the porosity measurement.
(iii) draw a random sample from the distribution for m.
(iv) draw a random sample from the distribution for ϕ.
(v) compute ρ_f from eq. (5), then use it to compute salinity using T from eq. (6).

We then binned the computed salinity values with depth and, for each bin, plotted a normalized histogram (see Fig. 14). These salinity estimates reflect the posterior uncertainty in the inverted bulk resistivity as well as the uncertainty in the porosity and the degree of cementation of the geological units. The freshwater aquifer is clearly visible in the measured salinity log for well M0027A (the dashed lines in Fig. 14 are measured salinity logs). This is also

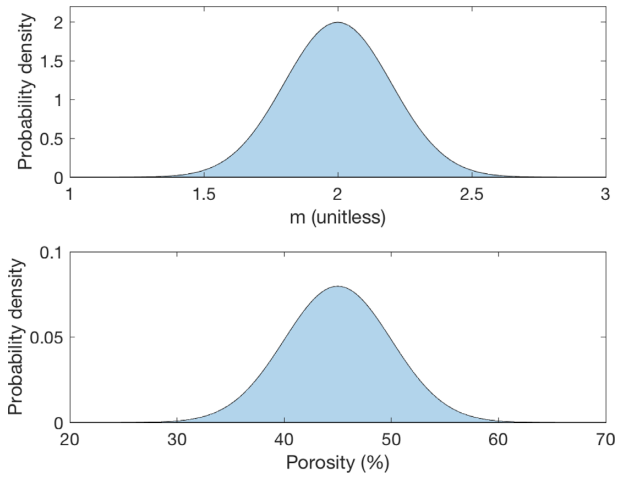

Figure 13. Probability distributions for the cementation factor, m, in Archie's Law (top panel) and the porosity of continental shelf sediments (bottom panel)—shown here for a porosity of 45 per cent. These distributions were used in the Monte Carlo process for estimating the uncertainty in the pore fluid salinity.

reflected in the uncertainty estimates near M0027A by a concentration of probability density below 20 practical salinity units—and, at some depths, well below 10 units. The salinity estimates from inversions of CSEM data only tend to predict lower salinities, probably due to the method's relative sensitivity to resistors, as well as

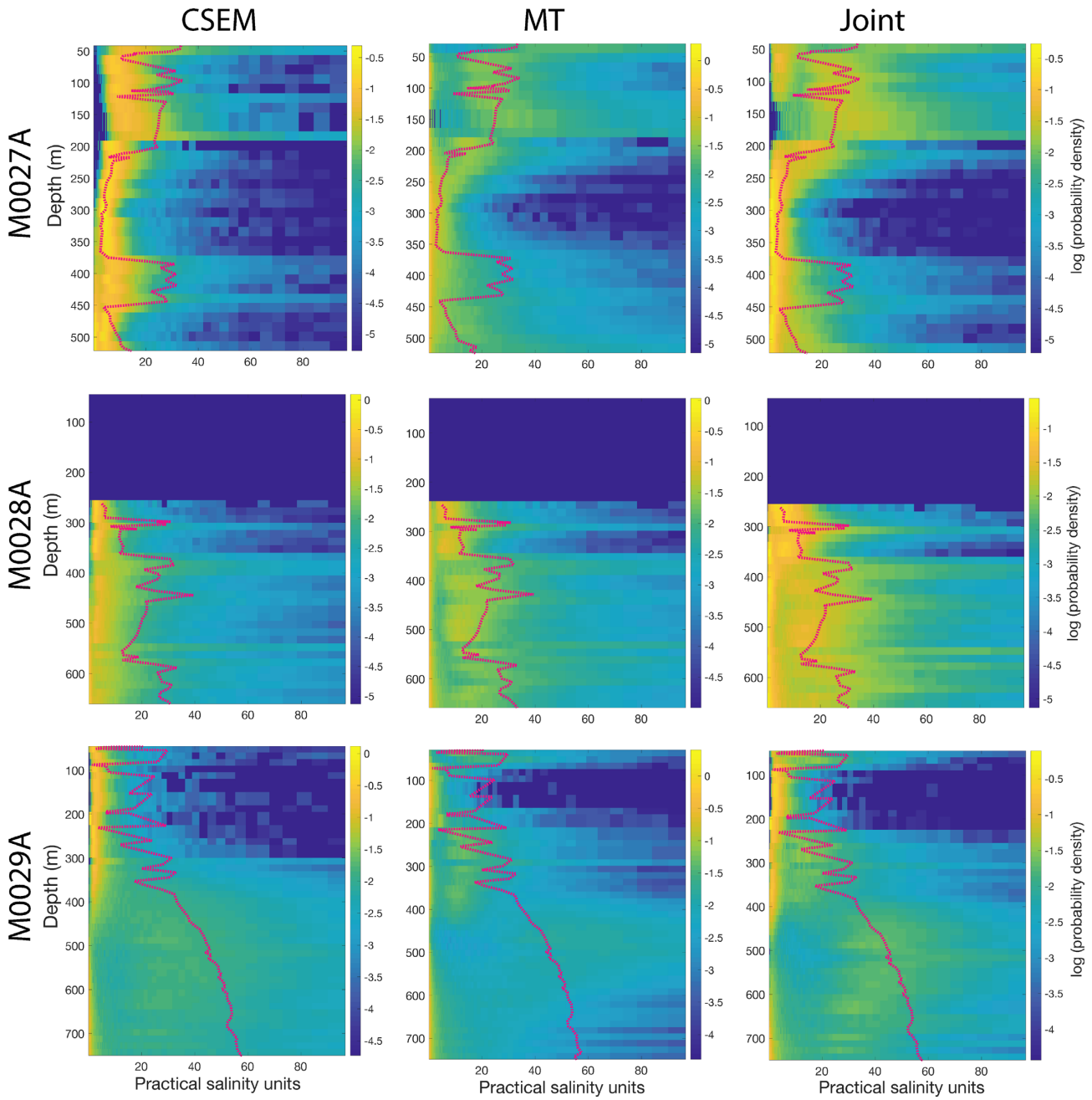

Figure 14. Distributions of probability density for pore fluid salinity as a function of depth at the locations of wells M0027A, M0028A and M0029A (see Fig. 2). Pore fluid resistivity (ρ_f) was obtained from Archie's Law (5) via a Monte Carlo method. Values of m and ϕ were drawn from their respective distributions (Fig. 13). Values of ρ_b were drawn from their respective model ensembles (obtained by inverting data nearest to the drilling locations—N03–N05). Pore fluid salinity was finally obtained using the PSS-78 standard, assuming a linear temperature gradient. The salinity values were binned in depth. In each plot, the well log salinity is plotted over the probability density (red dashed line). In each case, the joint inversion yields the most accurate estimate of salinity. This is important since it is often the case that reliable well data are not available.

the surface-towed system's relatively shallow sensitivity. The MT data-derived estimates, on the other hand, tend to predict higher salinities, and with a larger spread in probability density, likely the result of the MT method's preferential sensitivity to conductors. The joint dataset-derived salinity estimates, however, tend to provide the most accurate assessments of the uncertainty. That Bayesian sampling methods can provide a reliably accurate estimate of the uncertainty on related parameters of interest—and that joint inversion offers improved estimates of this uncertainty—is important as reliable well data are not uniformly available.

7 CONCLUSION

Bayesian sampling methods provide quantitative estimates of non-linear inverted model parameter uncertainty. As a result, they allow the uncertainty-reduction provided by joint inversion of multiple data sets to be visualized and assessed, which we demonstrated on a synthetic data set.

We provided Bayesian inference on the electrical resistivity of the upper kilometre of the North American continental shelf offshore New Jersey. We applied a trans-dimensional MCMC algorithm to

invert surface-towed CSEM and seafloor MT data for an ensemble of 1-D models, from which we obtained 1-D estimates of probability density for resistivity as a function of depth. We clearly identify the zone of relatively fresh (brackish) water indicated on well logs and in previous, gradient-based inversions of these same CSEM and MT data. In addition, we combined information from the posterior model ensembles and porosity measurements from nearby well-logs in a Monte Carlo scheme to estimate the salinity of the pore fluids in the upper kilometre of continental shelf sediments. These estimates were then compared to *in situ* measurements. This ability to provide accurate uncertainty estimates for parameters of interest to scientists outside the EM geophysics community is a strong argument in favour of applying Bayesian sampling methods to invert EM data whenever computationally feasible.

ACKNOWLEDGEMENTS

We gratefully acknowledge funding support from National Science Foundation grants 1458392 and 1459035. We thank the captain and crew of the R.V. Marcus G. Langseth for a successful cruise and the Marine EM Lab at Scripps Institution of Oceanography for providing the instrumentation. We also thank Chris Armerding, Marah Dahn, John Desanto, Jimmy Elsenbeck, Matt Folsom, Keiichi Ishizu, Jeff Pepin, Charlotte Wiman and Georgie Zelenak for participating in the cruise. We gratefully acknowledge Alberto Malinverno for the idea to use a Monte Carlo scheme to estimate the distribution of pore fluid salinity, and William Menke for many constructive conversations and suggestions.

REFERENCES

Abubakar, A., Li, M., Pan, G., Liu, J. & Habashy, T.M., 2011. Joint MT and CSEM data inversion using a multiplicative cost function approach, *Geophysics,* **76**(3), F203–F214.

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

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

Agostinetti, N.P. & Malinverno, A., 2018. Assessing uncertainties in high-resolution, multifrequency receiver-function inversion: A comparison with borehole data, *Geophysics,* **83**(3), KS11–KS22.

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

Archie, G.E., 1942. The electrical resistivity log as an aid in determining some reservoir characteristics, *Trans. AIME,* **146**(01), 54–62.

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, *Geophys. J. Int.,* **214**(3), 1919–1936.

Bodin, T., Sambridge, M., Tkalcic, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.: Solid Earth,* **117**(B2), n/a–n/a.

Bouchedda, A., Chouteau, M., Binley, A. & Giroux, B., 2012. 2-D joint structural inversion of cross-hole electrical resistance and ground penetrating radar data, *J. appl. Geophys.,* **78,** 52–67.

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

Cohen, D. *et al.*, 2010. Origin and extent of fresh paleowaters on the Atlantic Continental Shelf, USA, *Groundwater,* **48**(1), 143–158.

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

Constable, S. & Weiss, C.J., 2006. Mapping thin resistors and hydrocarbons with marine EM methods: insights from 1D modeling, *Geophysics,* **71**(2), G43–G51.

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.

Dosso, S.E. & Wilmut, M.J., 2006. Data uncertainty estimation in matched-field geoacoustic inversion, *IEEE J. Oceanic Eng.,* **31**(2), 470–479.

Egbert, G.D., 1997. Robust multiple-station magnetotelluric data processing, *Geophys. J. Int.,* **130**(2), 475–496.

Expedition 313 Scientists, 2010. Expedition 313 summary, in *Proceedings of the IODP, 313,* Integrated Ocean Drilling Program.

Gallardo, L.A. & Meju, M.A., 2004. Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints, *J. geophys. Res.: Solid Earth,* **109**(B3), 106–11.

Gehrmann, R.A.S., Dettmer, J., Schwalenberg, K., Engels, M., Dosso, S.E. & Özmaral, A., 2015. Trans-dimensional Bayesian inversion of controlled-source electromagnetic data in the German North Sea, *Geophys. Prospect.,* **63**(6), 1314–1333.

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

eds Gilks, W.R., Richardson, S. & Spiegelhalter, D., 1995. *Markov Chain Monte Carlo in Practice,* CRC Press.

Gómez Trevino, E. & Edwards, R.N., 1983. Electromagnetic soundings in the sedimentary basin of southern Ontario—a case history, *Geophysics,* **48**(3), 311–330.

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

Guo, R., Dosso, S.E., Liu, J., Liu, Z. & Tong, X., 2014. Frequency- and spatial-correlated noise on layered magnetotelluric inversion, *Geophys. J. Int.,* **199**(2), 1205–1213.

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov Chains and Their Applications, *Biometrika,* **57**(1), 97.

Hoversten, G.M., Cassassuce, F., Gasperikova, E., Newman, G.A., Chen, J., Rubin, Y., Hou, Z. & Vasco, D., 2006. Direct reservoir parameter estimation using joint inversion of marine seismic AVA and CSEM data, *Geophysics,* **71**(3), C1–C13.

Jardani, A., Revil, A., Slob, E. & Söllner, W., 2010. Stochastic joint inversion of 2D seismic and seismoelectric signals in linear poroelastic materials: a numerical investigation, *Geophysics,* **75**(1), N19–N31.

Johansen, S.E., Panzner, M., Mittet, R., Amundsen, H.E.F., Lim, A., Vik, E., Landrø, M. & Arntsen, B., 2019. Deep electrical imaging of the ultraslow-spreading Mohns Ridge, *Nature,* **567**(7748), 379–383.

Key, K., 2009. 1D inversion of multicomponent, multifrequency marine CSEM data: methodology and synthetic studies for resolving thin resistive layers, *Geophysics,* **74**(2), F9–F20.

Key, K., 2012. Marine EM inversion using unstructured grids: a 2D parallel adaptive finite element algorithm, in *SEG Technical Program Expanded Abstracts 2012,* pp. 1–5, Society of Exploration Geophysicists.

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

Lelièvre, P.G., Farquharson, C.G. & Hurich, C.A., 2012. Joint inversion of seismic traveltimes and gravity data on unstructured grids with application to mineral exploration, *Geophysics,* **77**(1), K1–K15.

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

Malinverno, A. & Briggs, V.A., 2004. Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes, *Geophysics,* **69**(4), 1005–1016.

Mandolesi, E., Ogaya, X., Campanyà, J. & Agostinetti, N.P., 2018. A reversible-jump Markov chain Monte Carlo algorithm for 1D inversion of magnetotelluric data, *Comput. Geosci.,* **113,** 94–105.

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

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.

Mitrovica, J.X. & Forte, A.M., 2004. A new inference of mantle viscosity based upon joint inversion of convection and glacial isostatic adjustment data, *Earth planet. Sci. Lett.,* **225**(1-2), 177–189.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.: Solid Earth,* **100**(B7), 12431–12447.

Mountain, G. & Proust, J.-N., Expedition 313 Science Party, 2010. The New Jersey Margin Scientific Drilling Project (IODP Expedition 313): untangling the record of global and local sea-level changes, *Scient. Drilling,* **10,** 26–34.

Myer, D., Constable, S. & Key, K., 2010. Broad-band waveforms and robust processing for marine CSEM surveys, *Geophys. J. Int.,* **184**(2), 689–698.

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

Perkin, R. & Lewis, E., 1980. The practical salinity scale 1978: fitting the data, *IEEE J. Ocean. Eng.,* **5**(1), 9–16.

Rabben, T.E., Tjelmeland, H. & Ursin, B., 2008. Non-linear Bayesian joint inversion of seismic reflection coefficients, *Geophys. J. Int.,* **173**(1), 265–280.

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

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

Ray, A., Key, K. & Bodin, T., 2013b. Hierarchical Bayesian inversion of marine CSEM data over the Scarborough gas field—a lesson in correlated noise, in *SEG Technical Program Expanded Abstracts 2013,* pp. 723–727, Society of Exploration Geophysicists.

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, *Geophys. J. 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, *Geophys. J. Int.,* **196**(3), 1508–1524.

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

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

Sasaki, Y., 1989. Two-dimensional joint inversion of magnetotelluric and dipole-dipole resistivity data, *Geophysics,* **54**(2), 254–262.

Vozoff, K. & Jupp, D.L.B., 1975. Joint inversion of geophysical data, *Geophys. J. R. Astron. Soc.,* **42**(3), 977–991.

Ward, S.H. & Hohmann, G.W., 1987. Electromagnetic theory for geophysical applications, in *Electromagnetic Methods in Applied Geophysics,* pp. 130–311, Society of Exploration Geophysicists.

Wheelock, B., Constable, S. & Key, K., 2015. The advantages of logarithmically scaled data for electromagnetic inversion, *Geophys. J. Int.,* **201**(3), 1765–1780.

Xiang, E., Guo, R., Dosso, S.E., Liu, J., Dong, H. & Ren, Z., 2018. Efficient hierarchical trans-dimensional Bayesian inversion of magnetotelluric data, *Geophys. J. Int.,* **213**(3), 1751–1767.

Yokota, Y., Koketsu, K., Fujii, Y., Satake, K., Sakai, S., Shinohara, M. & Kanazawa, T., 2011. Joint inversion of strong motion, teleseismic, geodetic, and tsunami datasets for the rupture process of the 2011 Tohoku earthquake, *Geophys. Res. Lett.,* **38**(7), doi:10.1029/2011GL050098.

APPENDIX: MAXIMUM LIKELIHOOD ESTIMATE OF DATA COVARIANCE SCALE FACTOR

The joint likelihood function is the product of the individual likelihood functions:

$$p(\mathbf{d}|\mathbf{m}) = L_1(\mathbf{m})\, L_2(\mathbf{m}). \tag{A1}$$

If we substitute in the exponential form of the likelihood for one of these, we have:

$$p(\mathbf{d}|\mathbf{m}) = L_1(\mathbf{m}) \frac{1}{\sqrt{(2\pi)^{N_2}\sigma^{2N_2}|\mathbf{C_{d_2}}|}} \exp\left(-\frac{\mathbf{r_2}^{\mathrm{T}}\mathbf{C_{d_2}^{-1}}\mathbf{r_2}}{2\sigma^2}\right), \tag{A2}$$

where the residual vector $\mathbf{r_2} = \mathbf{d_2} - \mathbf{f_2}(\mathbf{m})$ and N_2 is the number of data in the second data set. Here we have introduced a constant factor, σ, that scales all the entries of the data covariance matrix for the second data set, $\mathbf{C_{d_2}}$. To find a maximum likelihood estimate of σ, we first take the negative logarithm of the likelihood function:

$$\begin{aligned} -\log(p(\mathbf{d}|\mathbf{m})) = &-\log(L_1(\mathbf{m})) + \frac{1}{2}\log\left[(2\pi)^{N_2}|\mathbf{C_{d_2}}|\right] \\ &+N_2\log(\sigma) + \frac{\mathbf{r_2}^{T}\mathbf{C_{d_2}}^{-1}\mathbf{r_2}}{2\sigma^2}. \end{aligned} \tag{A3}$$

Next, we differentiate with respect to σ:

$$\frac{\partial}{\partial\sigma}\left(-\log(p(\mathbf{d}|\mathbf{m}))\right) = \frac{N_2}{\sigma} - \frac{\mathbf{r_2}^{T}\mathbf{C_{d_2}}^{-1}\mathbf{r_2}}{\sigma^3}. \tag{A4}$$

To find the value of the scale factor σ that maximizes the likelihood, we set this derivative equal to zero and solve for σ:

$$\hat{\sigma}_{max} = \sqrt{\frac{\mathbf{r_2}^{T}\mathbf{C_{d_2}}^{-1}\mathbf{r_2}}{N_2}}. \tag{A5}$$

In other words, the scale factor that maximizes the likelihood function—also a value that Hierarchical Bayes approach would sample—is the rms residual between the forward modelled and measured data.

Finally, we can insert $\hat{\sigma}_{max}$ from eq. (A5) into eq. (A3), giving:

$$\begin{aligned} -\log(p(\mathbf{d}|\mathbf{m})) = &-\log(L_1(\mathbf{m})) + \frac{N_2}{2}\log\left(\frac{\mathbf{r_2}^{T}\mathbf{C_{d_2}^{-1}}\mathbf{r_2}}{N_2}\right) \\ &+R(\mathbf{C}_{\mathrm{d}_2}, N_2), \end{aligned} \tag{A6}$$

where $R(\mathbf{C}_{\mathrm{d}_2}, N_2)$ contains the terms that do not depend on $\mathbf{m}$. Thus the MCMC sampling could be done using A6 as the negative log likelihood, without artificially fixing the scaling factor between the two data types or using Hierarchical Bayes. We could also carry out the same treatment for each data type's likelihood individually and arrive at a similar equation to (A6). An analogous example would be maximum likelihood data errors for every frequency in seismic full waveform inversion as shown in Ray *et al.* (2016, eq. 29).