2021

Quantifying model structural uncertainty using airborne electromagnetic data

Burke J. Minsley

Nathan Leon Foks

Paul A. Bedrosian

Geophysical Journal International

Geophys. J. Int. (2021) **224,** 590–607
Advance Access publication 2020 August 25
GJI General Geophysical Methods

doi: 10.1093/gji/ggaa393

Quantifying model structural uncertainty using airborne electromagnetic data

Burke J. Minsley[1], Nathan Leon Foks[2] and Paul A. Bedrosian[1]

[1] *U.S. Geological Survey, Geology, Geophysics, and Geochemistry Science Center, MS*973*—Denver Federal Center, Denver, CO* 80225*, USA. E-mail: bminsley@usgs.gov*
[2] *Apogee Engineering LLC. Contracted to U.S. Geological Survey, Science Analytics and Synthesis, Advanced Research Computing, Building* 810*—Denver Federal Center, Denver, CO* 80225, USA

Accepted 2020 August 15. Received 2020 August 11; in original form 2020 January 13

SUMMARY

The ability to quantify structural uncertainty in geological models that incorporate geophysical data is affected by two primary sources of uncertainty: geophysical parameter uncertainty and uncertainty in the relationship between geophysical parameters and geological properties of interest. Here, we introduce an open-source, trans-dimensional Bayesian Markov chain Monte Carlo (McMC) algorithm GeoBIPy—Geophysical Bayesian Inference in Python—for robust uncertainty analysis of time-domain or frequency-domain airborne electromagnetic (AEM) data. The McMC algorithm provides a robust assessment of geophysical parameter uncertainty using a trans-dimensional approach that lets the AEM data inform the level of model complexity necessary by allowing the number of model layers itself to be an unknown parameter. Additional components of the Bayesian algorithm allow the user to solve for parameters such as data errors or corrections to the measured instrument height above ground. Probability distributions for a user-specified number of lithologic classes are developed through posterior clustering of McMC-derived resistivity models. Estimates of geological model structural uncertainty are thus obtained through the joint probability of geophysical parameter uncertainty and the uncertainty in the definition of each class. Examples of the implementation of this algorithm are presented for both time-domain and frequency-domain AEM data acquired in Nebraska, USA.

Key words: Non-linear electromagnetics; Inverse theory; Probability distributions; Statistical methods.

INTRODUCTION

Models are useful tools for summarizing existing knowledge and for making predictions about future responses given alternate forcing or management scenarios (Bredehoeft 2005). A particular challenge in the development and use of models, however, is characterizing their uncertainty given the limitations of the data that are used to construct them (Neuman 2003; Højberg & Refsgaard 2005; Troldborg *et al.* 2007; Refsgaard *et al.* 2012; Scheidt *et al.* 2018). The very definition of a model itself may be uncertain given the underlying assumptions used in its construction. For example, most geophysical or hydrological models are created by determining a set of spatially varying physical parameters through the process of inversion or calibration to measured data, and standard techniques exist for evaluating the uncertainty in these parameter estimates. A number of studies have highlighted the importance of conceptual uncertainty, or the uncertainty in the geological model structure that may be the most important component of model uncertainty (Refsgaard *et al.* 2012). We present an approach for characterizing model structural uncertainty that accounts for both uncertainties in the underlying geophysical data used to inform model parameters and in the physical property relationship that links geophysical and geological parameters. Results presented here are intended not to produce 'best' estimates of model structure, but rather to falsify models inconsistent with observations and prior information in order to communicate a range of plausible solutions (Tarantola 2006).

Geophysical parameter uncertainty and non-uniqueness can be caused by numerous factors, including observation errors, survey design, measurement physics, numerical modelling and inversion parameter assumptions. Each of these factors contributes in various ways to the uncertainty of an estimated model parameter value, which can be assessed using any one of many well established approaches (Aster *et al.* 2005; Tarantola 2005). A second and more difficult source of uncertainty to quantify is in defining how geophysical properties are related to the geological properties of interest. For example, the typical range of electrical resistivity values associated with various geologic formations

can be estimated broadly from generic compilations (e.g. Palacky 1987), or from site-specific laboratory or borehole observations that have inherent variability in physical property relationships (Fitterman & Deszcz-Pan 1998; Bosch *et al.* 2002; Foged *et al.* 2014; Marker *et al.* 2015; Knight *et al.* 2018).

Airborne electromagnetic (AEM) data are increasingly used to inform geological and hydrological models because of their ability to cover relatively large areas—often spanning hundreds-to-thousands of square kilometres—with high-resolution information about subsurface properties. AEM data are typically inverted to recover estimates of the 3-D distribution of electrical resistivity values, which when combined with borehole logs and other observations inform interpretations of subsurface properties in support of hydrologic studies (Fitterman & Deszcz-Pan 1998; Siemon *et al.* 2009; Viezzoli *et al.* 2010; Jørgensen *et al.* 2012), geologic mapping (Best *et al.* 2006; Jørgensen *et al.* 2013; Høyer *et al.* 2015) and cryosphere investigations (Minsley *et al.* 2012a; Mikucki *et al.* 2015; Walter Anthony *et al.* 2018). The ability to accurately recover the distribution of electrical resistivity values, and their relation to geological or hydrological properties of interest, is therefore of key importance for making quantitative interpretations based on AEM data.

Numerous studies discuss quantitative approaches for interpretation of both airborne time-domain electromagnetic (TDEM) and frequency-domain electromagnetic (FDEM) data, including geophysical inversion for electrical resistivity models (Farquharson *et al.* 2003; Brodie & Sambridge 2006; Cox & Zhdanov 2008; Viezzoli *et al.* 2008; Christensen *et al.* 2009, 2010; Auken *et al.* 2015), comparison of different AEM instruments (Bedrosian *et al.* 2015), analysis of AEM spatial footprints (Beamish 2003; Reid *et al.* 2006) and depth of investigation (Spies 1989; Christiansen & Auken 2012). While the majority of algorithms for the analysis of AEM data use deterministic gradient-based methods, Bayesian Markov chain Monte Carlo (McMC) algorithms have also been developed for the rigorous quantification of parameter uncertainty associated with AEM data (Minsley 2011; Brodie 2017; Blatter *et al.* 2018; Hawkins *et al.* 2018).

The development of geological or hydrological models based on AEM data has traditionally been done in a manual way, with the interpreter using borehole or other auxiliary information along with transitions in resistivity values to identify geological boundaries (Abraham *et al.* 2012; Jørgensen *et al.* 2012). More recently, semi-automated approaches have been developed that make the interpretation process faster and less subject to the decisions of an individual modeller (Gulbrandsen *et al.* 2017). Campanyà *et al.* (2019) use a statistical approach to map borehole log data classifications onto a spatially extensive electrical resistivity model determined from magnetotelluric data. One drawback to these approaches, however, is that the result is typically a single interpretive model that does not fully reflect the uncertainties and ambiguities that inevitably lead to multiple geological interpretations that are supported by the same geophysical data. Several studies have proposed the use of geostatistical simulation tools as an objective approach for generating multiple geological realizations conditioned to geophysical models and borehole observations (He *et al.* 2014; Koch *et al.* 2014; Marker *et al.* 2017). For example, Marker *et al.* (2017) link electrical resistivity models derived from several types of ground-based and airborne geophysical data with borehole logs to produce a clay-fraction model (Foged *et al.* 2014) that, together with clustering and sequential indicator simulation (SIS) algorithms, are used to produce multiple ensemble realizations of hydrostratigraphic structure that are used in the development and calibration of groundwater models. This approach has also been extended to training image-based methods that use multi-point statistics (Vilhelmsen *et al.* 2019) or spatial generative adversarial neural network (Laloy *et al.* 2018) algorithms to generate complex realizations of model structure that can be used to forecast uncertainty in hydrologic parameters such as well catchment area.

In this paper, we combine both the uncertain relationship between geophysical properties such as electrical resistivity and geological or hydrostratigraphic parameters, as well as the uncertainty in geophysical parameters given observed data and measurement errors following the approach of Christensen *et al.* (2017). We discuss in detail the development of a new open-source trans-dimensional Bayesian McMC Python package GeoBIPy (Geophysical Bayesian Inference in Python) for geophysical parameter uncertainty quantification (Foks & Minsley 2020) that can be used with time-domain or frequency-domain AEM data that builds on the approach described by Minsley (2011).

METHODS

A two-step approach is used to quantify uncertainty in using AEM data to inform geological model structure. First, comprehensive estimates of geophysical parameter uncertainty are made using a trans-dimensional Bayesian McMC algorithm that is based on the methods described by Minsley (2011) and can be applied to either time-domain or frequency-domain AEM data. Next, a relationship that links resistivity and lithologic classes is identified. Here, this relationship is developed empirically using a semi-automated approach; however, one could also incorporate borehole or petrophysical data to develop the relationship between geological and geophysical properties (Christensen *et al.* 2017). The product of these two sources of uncertainty, in geophysical parameters and in the geophysical-geological relationship, is used to define overall uncertainty in model structure.

Quantifying geophysical parameter uncertainty

The basic McMC approach for the estimation of geophysical parameter uncertainty builds on the algorithm described by Minsley (2011) using a similar model parametrization and workflow. Important updates to this algorithm presented here include (1) capability to work with either frequency-domain or time-domain AEM data sets, (2) a hierarchical approach to estimating data errors as additional unknown parameters (Bodin *et al.* 2012b) and (3) the use of stochastic Newton methods (Martin *et al.* 2012) for more efficient model proposals in the McMC

workflow. We also present an approach for visualization and communication of geophysical uncertainties given the large amount of data typically acquired in an AEM survey, combined with the large number of models evaluated at every location in the McMC analysis.

A typical AEM survey can have on the order of 10^4–10^7 data locations. Given that the McMC algorithm is run for 10^5 steps per location, approximately 10^9–10^{12} models are sampled for a single survey. Given the commonly used 1-D layered-earth modelling assumption, every AEM survey location is treated as an independent data set, allowing for straightforward and scalable parallelization where each data point can be distributed to a single computational core. At each survey location, a 1-D depth-dependent model is estimated, where the model is defined by k piecewise constant resistivity values and k-1 interfaces where the top of the first interface is fixed at the ground surface and the bottom resistivity layer is a half-space.

As with Minsley (2011), we implement a naturally parsimonious trans-dimensional McMC algorithm that allows the number of model layers to be one of the unknown parameters (Malinverno 2002; Sambridge *et al.* 2013), where solutions with fewer model layers are implicitly favoured. By allowing the data to determine the number of necessary model layers, we remove a significant potential source of model error related to the conceptual model design which typically requires the number of model layers to be pre-defined and fixed across the entire survey area. A 1-D layered-earth model, $\mathbf{m}_e$, is defined by the combined set of $2k$ parameters,

$$\mathbf{m}_e = [k,\ \ln(\mathbf{z}), \ln(\boldsymbol{\sigma})], \tag{1}$$

where $\mathbf{z}$ is a vector of $k-1$ interface depths and $\boldsymbol{\sigma}$ is a vector of k layer conductivity values. While the algorithm is implemented in terms of conductivity, results are presented in terms of its inverse, resistivity (ρ), to be consistent with much of the AEM literature. In addition to the layered-earth model parameters, a hierarchical approach is used to estimate any of three optional parameters, $\mathbf{m}_{opt}$: (1) corrections to the AEM transmitter height above ground, h^{tx} (Minsley 2011); (2) system calibration correction factors for frequency-domain data, $\mathbf{m}_{cal}$ (Minsley *et al.* 2012b; Minsley *et al.* 2014); and (3) data error parameters, $\boldsymbol{\varepsilon}$ (Bodin *et al.* 2012b), where

$$\mathbf{m}_{opt} = \left[h^{tx}, \mathbf{m}_{cal}, \boldsymbol{\varepsilon}\right]. \tag{2}$$

In this study, we focus only on this final optional parameter, an estimate of data errors, which is described below.

The posterior probability distribution of the parameters in $\mathbf{m} = [\mathbf{m}_e, \mathbf{m}_{opt}]$ is described by Bayes' rule,

$$p(\mathbf{m}|\mathbf{d}) = c \cdot p(\mathbf{m}) \cdot p(\mathbf{d}|\mathbf{m}), \tag{3}$$

where c is a normalizing constant, $p(\mathbf{m})$ defines the prior probability distribution for each of the parameters in $\mathbf{m}$ and $p(\mathbf{d}|\mathbf{m})$ is the data likelihood function, which is a measure of data fit. Simulated forward responses and model sensitivities (used in the McMC proposal step) are calculated using existing 1-D layered-earth approximation algorithms for frequency-domain (Minsley 2011) or time-domain (Brodie 2017) data.

A detailed description of the algorithm can be found in Minsley (2011), which is summarized in the following routine and also described in greater detail below:

(1) Define a homogeneous starting model with conductivity equal to the best-fitting half-space value.
(2) Propose a new model in two steps:

(i) Propose new model geometry by changing the number or locations of layer boundaries (k, $\mathbf{z}$).
(ii) Propose new model parameter values ($\boldsymbol{\sigma}$, $\mathbf{m}_{opt}$).

(3) Calculate data likelihood for the proposed new model.
(4) Use the Metropolis and Hastings (Metropolis *et al.* 1953; Hastings 1970) criteria to accept or reject proposed models.
(5) Repeat from (2) for 10^5 steps after burn-in. The initial burn-in phase ends when an acceptable level of data fit is achieved.

Model priors and natural parsimony

The naturally parsimonious behaviour of the algorithm is described by the prior terms for k, $\boldsymbol{\sigma}$ and $\mathbf{z}$ that compose the joint model prior $p(\mathbf{m})$ in eq. (3),

$$p(\mathbf{m}) = p(k) \cdot p(\boldsymbol{\sigma}, \mathbf{z}|k). \tag{4}$$

Although the user can choose any form of the prior distribution for k, we typically choose a uniform distribution on the interval [1, 30] such that each number of layers is equally probable. In addition, the probability of adding a layer in step 2(i) described above is given the same probability (1/6) as deleting a layer so there is no explicit bias towards larger or fewer-layer models (Malinverno 2002; Minsley 2011). However, the prior for $\boldsymbol{\sigma}$ is conditional on k—for example, the prior for conductivity using a normal distribution with covariance $\boldsymbol{C}_{\sigma_0}$ about the reference half-space model $\boldsymbol{\sigma}_0$ is

$$p(\boldsymbol{\sigma}|k) = \frac{1}{\sqrt{2\pi^k \left|\boldsymbol{C}_{\sigma_0}\right|}} e^{-\frac{1}{2}(\boldsymbol{\sigma}-\sigma_0)^T C_{\sigma_0}^{-1}(\boldsymbol{\sigma}-\sigma_0)}. \tag{5}$$

A model with many layers will therefore have a relatively low prior probability compared with a few-layer model, even if their conductivity structure is identical, owing to the fact that the more complex model with greater number of layers occupies a larger prior volume (MacKay

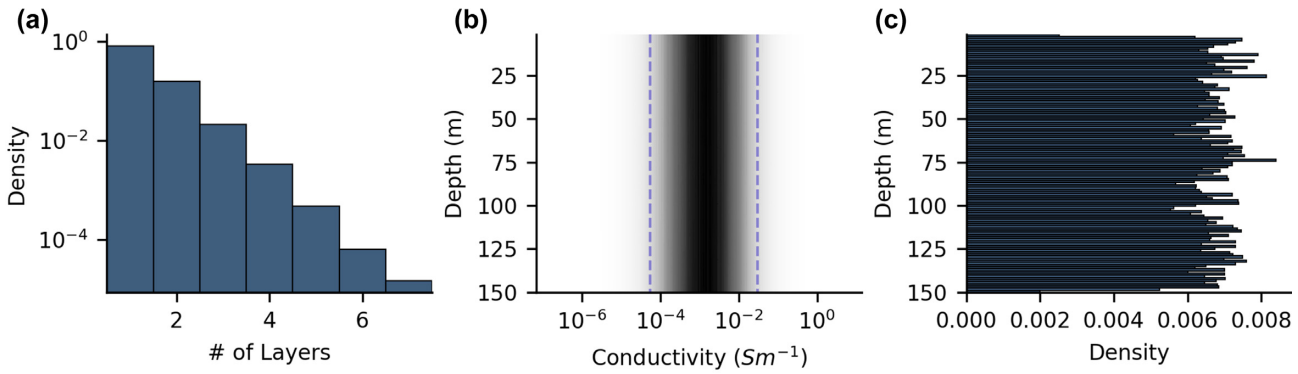

Figure 1. Samples drawn from the normal joint conditional prior distribution $\boldsymbol{p}(\mathbf{m})$ in eq. (4). (a) The sampled distribution for number of model layers favours fewer-layer models even though the explicit prior on k is uniform due to the additional parsimony included in the joint conditional prior on conductivity values or their gradients. (b) Conductivity values sample a normal distribution with mean value 0.001 S m^{-1} in this example. (c) Layer interface depths show uniform sampling.

2003). This result of simple models being implicitly favoured over more complex ones has been described as the 'Occam Factor' in greater detail in several other studies (Malinverno 2002; MacKay 2003; Ray *et al.* 2017).

This implicit favouring is illustrated in Fig. 1, where we run the algorithm for 1 000 000 steps while ignoring the data likelihood term (i.e. all values of $p(\mathbf{d}|\mathbf{m}) = 1$) in order to generate samples exclusively from the joint conditional prior distribution in eq. (4). The combined prior for all parameters, $\mathbf{m}$, strongly favours few-layer models even though $p(k)$ is uniform (Fig. 1a). An interesting consequence of this behaviour is that the weighting towards natural parsimony will depend somewhat on the choice of prior distribution. For example, if a uniform conductivity prior is used, then $p\ (\sigma|k) = (\sigma_{\max} - \sigma_{\min})^{-k}$ which still favours few-layer models, but with different relative weighting compared with the case of a normal conductivity prior distribution.

Estimating data errors

Data errors are specified using an error model that combines both relative (Σ^{re}) and additive (Σ^{ae}) components that make up the total measurement error, Σ:

$$\Sigma_i = \left[\left(\Sigma_i^{\mathrm{re}}\right)^2 + \left(\Sigma_i^{\mathrm{ae}}\right)^2\right]^{1/2}. \tag{6}$$

The index, *i,* refers to different data components; these are individual transmitter–receiver coil pairs for frequency-domain data or time gates for time-domain data. The total errors for all data components in eq. (6) are controlled by two underlying parameters, $\boldsymbol{\varepsilon} = [\varepsilon^{\mathrm{re}}, \varepsilon^{\mathrm{ae}}]$. The relative errors are defined using the same relationship for both frequency- and time-domain data,

$$\Sigma_i^{\mathrm{re}} = \varepsilon^{\mathrm{re}}\ |d_i|\,, \tag{7}$$

where d_i is the data amplitude in parts per million (ppm) for frequency-domain data or V A^{-1} m^{-4} for time-domain data and $\varepsilon^{\mathrm{re}}$ is the relative error specified as a fraction of the data amplitude. For frequency-domain data, $\varepsilon^{\mathrm{ae}}$ [ppm] defines a minimum level of error, which is important for avoiding unrealistically small total errors for low-amplitude data points. In principle, one could solve for a separate relative and additive error for each frequency; however, here we use a single pair of parameters in $\boldsymbol{\varepsilon}$ that applies to all frequencies. For time-domain data, the additive error at each time gate is specified as

$$\Sigma_i^{\mathrm{ae}} = \varepsilon^{\mathrm{ae}}\ \mathrm{e}^{-\frac{1}{2}(t_i - t_0)}, \tag{8}$$

where $\varepsilon^{\mathrm{ae}}$ is the signal amplitude [V A^{-1} m^{-4}] at the reference decay time t_0 [s]. In this study, we use $t_0 = 0.001$ s. For systems with multiple receiver coils, that is, single-moment platforms with both horizontal and vertical receivers or dual-moment systems such as SkyTEM, each component of $\boldsymbol{\varepsilon}$ has a pair of parameters.

Total errors for each data component, Σ_i, are incorporated in the data likelihood function,

$$\mathrm{p}\,(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^N\ |C_{\mathrm{d}}|}}\mathrm{e}^{-\frac{1}{2}(\mathbf{d}-F(\mathbf{m}))^T\mathbf{C}_{\mathrm{d}}^{-1}(\mathbf{d}-F(\mathbf{m}))}, \tag{9}$$

where $\mathbf{C}_{\mathrm{d}}$ is the data covariance matrix with diagonal elements

$$C_{\mathrm{d}}^{ii} = \Sigma_i^2, \tag{10}$$

N is the number of measurements in the data vector, $\mathbf{d}$, and $F(\mathbf{m})$ is the simulated data for any given set of model parameters.

Stochastic Newton model proposals

The model proposal step for conductivity parameter values, used to transition the Markov chain from model $\boldsymbol{\sigma}$ to model $\boldsymbol{\sigma}$', follows the stochastic Newton method (Martin *et al.* 2012). The proposal density for transitioning conductivity values given new model structure (k', **z**') from the original model **m** is given by

$$q\left(\boldsymbol{\sigma}'|\mathrm{k}', \mathbf{z}', \mathbf{m}\right) = \sqrt{\frac{|\mathbf{H}|}{(2\pi)^k}} \mathrm{e}^{-\frac{1}{2}\left(\boldsymbol{\sigma}'-\tilde{\sigma}-\mathbf{H}^{-1}\mathbf{g}\right)^T \mathbf{H}\left(\boldsymbol{\sigma}'-\tilde{\sigma}-\mathbf{H}^{-1}\mathbf{g}\right)}, \tag{11}$$

where **H** and **g** are approximate local Hessian and gradient information computed about the current k-layer model, **m**, according to

$$\mathbf{H} = \mathbf{J}^T \mathbf{C}_{\mathrm{d}}^{-1}\mathbf{J} + \mathbf{C}_{\sigma_0}^{-1}, \text{ and} \tag{12}$$

$$\mathbf{g} = \mathbf{J}^T \mathbf{C}_{\mathrm{d}}^{-1}\left(\mathbf{d}\left(\mathbf{m}\right) - \mathbf{d}_{\mathrm{obs}}\right) + \mathbf{C}_{\sigma_0}^{-1}\left(\boldsymbol{\sigma} - \boldsymbol{\sigma}_{\mathrm{prior}}\right). \tag{13}$$

Here, **J** is the sensitivity matrix, $\mathbf{C}_{\mathrm{d}}$ is the data covariance matrix defined in eq. (10) and $\mathbf{C}_{\sigma_0}$ is the prior covariance for conductivity. The proposal distribution in eq. (11) is centred on a conductivity model ($\tilde{\sigma}$) that is built from the current k-layer conductivity values ($\boldsymbol{\sigma}$) mapped onto the proposed model structure (k', **z**'). Although calculation of the sensitivity matrix is computationally expensive, it provides an objective and efficient way to guide model space exploration by tuning model proposals to local estimates of the target posterior probability distribution. For example, use of the Hessian as the inverse covariance matrix for the proposal distribution in eq. (11) ensures that proposals for well-constrained model parameters (such as shallow layers with high sensitivity to the data) are relatively small compared with deep layers that are poorly constrained and are consistent with a wider range of model values. In addition, the off-diagonal information in the Hessian ensures that model proposals account for correlated parameters.

Changes in model dimension for birth or death steps in the trans-dimensional algorithm require special consideration in the model proposal. In the case of a birth step from $k = 2$ to $k' = 3$, for example, the mean of the proposal distribution ($\tilde{\sigma}$) is built from the current two-layer conductivity values ($\boldsymbol{\sigma}$) mapped onto the proposed three-layer model structure (k', **z**'). In a birth step, this involves duplicating the current conductivity value into the newly created layer; in a death-step, the existing conductivity values for the layers surrounding the deleted interfaced are averaged into a single conductivity value (Malinverno 2002). The proposal density described in eq. (11) is therefore a Normal distribution, with its mean centred on a new model, and is computed by solving a Gauss–Newton step from the current model; its covariance is equal to a local estimate of the posterior model covariance. Similar considerations are made for the reverse model proposal distribution, $q(\boldsymbol{\sigma}|k, \mathbf{z}, \mathbf{m}')$, that is used in computing the McMC acceptance probability α,

$$\alpha = \frac{p\left(\mathbf{m}'\right)}{p\left(\mathbf{m}\right)} \frac{p\left(\mathbf{d}|\mathbf{m}'\right)}{p\left(\mathbf{d}|\mathbf{m}\right)} \frac{q\left(\boldsymbol{\sigma}|k, \mathbf{z}, \mathbf{m}'\right)}{q\left(\boldsymbol{\sigma}'|k', \mathbf{z}', \mathbf{m}\right)}. \tag{14}$$

Code implementation

The Bayesian inference used throughout this study relies on the GeoBIPy software (Foks & Minsley 2020). GeoBIPy is an open-source software library that uses object-oriented design patterns and high-performance computing techniques to invert large-scale geophysical data sets using distributed-memory architectures. Computational efficiency is maintained through the use of existing numerical Python packages such as Numpy (Van der Walt *et al.* 2011) and Scipy (Oliphant 2007). The frequency-domain forward modeller is written within GeoBIPy using Python with Numba support, while the external time-domain forward modeller is written in C++ (Brodie 2017). At its core, GeoBIPy relies on the 'StatArray' class, a custom class that extends the Numpy package, and allows attachment of statistical attributes to a variable such as prior distributions, proposal distributions and histograms with fast indexing to represent a variable's posterior distribution. We then use the StatArray class to build the other components that represent the geophysical inverse problem such as the 1-D layered-earth model, time- or frequency-domain data and the McMC approach of the inversion algorithm. Given the layered-earth modelling assumption, we limit our subsurface model discretization to 1-D. By not imposing lateral smoothing constraints, we can independently invert each data point in a data set providing us with a massively parallel problem. GeoBIPy uses the mpi4py package (Dalcín *et al.* 2005) wrapped around an MPI implementation to accomplish this. Typically, one of the greatest impediments to computational efficiency of a parallel algorithm is the time required to read and write data to disk. Since GeoBIPy generates relatively large amounts of analysable information (Terabytes), compared to the typical size of a data file (Megabytes), we require a fast data storage format that allows access from multiple CPU cores simultaneously, in parallel. To this end, we leverage the h5py Python package (Collette 2013) that wraps around the Hierarchical Data Format (HDF; The HDF Group 1997), which allows McMC results to be written in parallel. The HDF file is stored as binary (as opposed to human readable ascii format), and can optionally apply compression, thus minimizing file size on disk. In addition to the software benefits, performance can be bolstered by utilizing parallel file system hardware that accompanies most modern high-performance computing clusters.

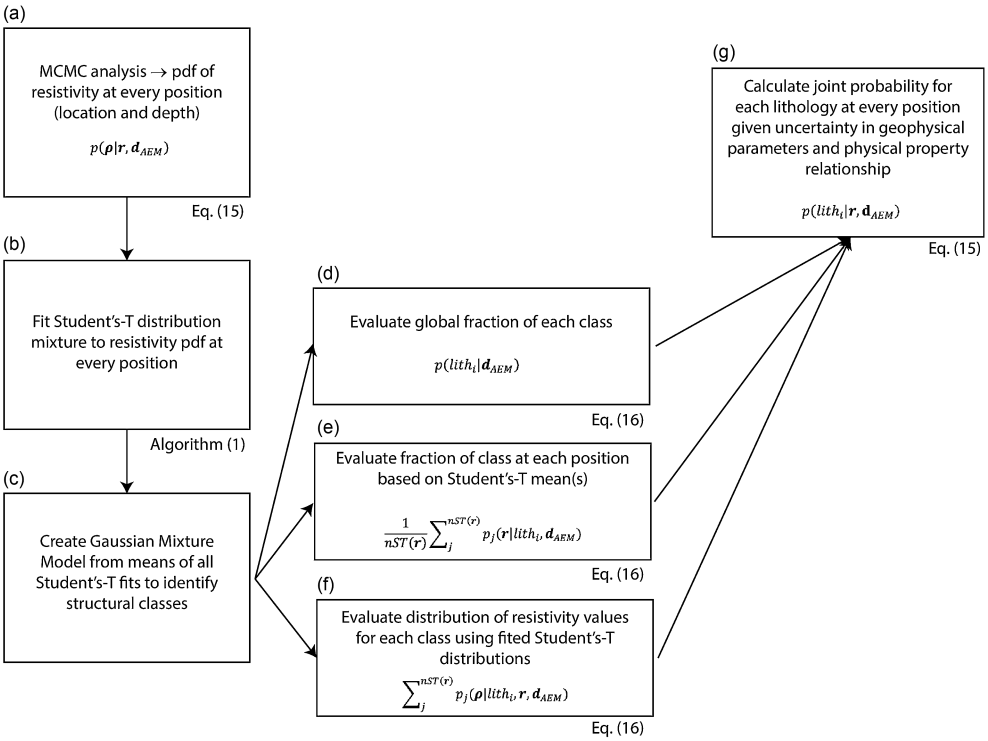

Figure 2. Workflow for posterior clustering approach to develop a resistivity–lithology relationship.

Developing a resistivity—lithology relationship

To determine the probability of the *i*th lithology, or facies class, given AEM data, we marginalize over resistivity (ρ),

$$p\left(\text{lith}_i|\boldsymbol{r}, \mathbf{d}_{\text{AEM}}\right) = \int_{\rho} p\left(\text{lith}_i, \boldsymbol{\rho}|\boldsymbol{r}, \boldsymbol{d}_{\text{AEM}}\right) \mathrm{d}\rho = \int_{\rho} p\left(\text{lith}_i|\boldsymbol{r}, \boldsymbol{\rho}, \boldsymbol{d}_{\text{AEM}}\right) \cdot p\left(\boldsymbol{\rho}|\boldsymbol{r}, \boldsymbol{d}_{\text{AEM}}\right) \mathrm{d}\rho. \tag{15}$$

Here, $p(\boldsymbol{\rho}|\boldsymbol{r}, \boldsymbol{d}_{\text{AEM}})$ is the posterior probability distribution for resistivity at any particular coordinate, $\boldsymbol{r} = [\boldsymbol{x}, \boldsymbol{y}, \boldsymbol{z}]$, from the geophysical McMC analysis, and

$$p\left(\text{lith}_i\,|\boldsymbol{r}, \boldsymbol{\rho}, \boldsymbol{d}_{\text{AEM}}\right) = \frac{\frac{1}{\text{nST}(\boldsymbol{r})} p\left(\text{lith}_i|\boldsymbol{d}_{\text{AEM}}\right) \sum_j^{\text{nST}(\boldsymbol{r})} p_j(\boldsymbol{r}|\text{lith}_i, \boldsymbol{d}_{\text{AEM}}) \cdot p_j(\boldsymbol{\rho}|\text{lith}_i, \boldsymbol{r},\ \boldsymbol{d}_{\text{AEM}})}{\sum_i \frac{1}{\text{nST}(\boldsymbol{r})} p\left(\text{lith}_i|\boldsymbol{d}_{\text{AEM}}\right) \sum_j^{\text{nST}(\boldsymbol{r})} p_j(\boldsymbol{r}|\text{lith}_i, \boldsymbol{d}_{\text{AEM}}) \cdot p_j(\boldsymbol{\rho}|\text{lith}_i, \boldsymbol{r},\ \boldsymbol{d}_{\text{AEM}})} \tag{16}$$

describes the location-dependent resistivity relationships for each lithology and is developed from posterior fitting and clustering analyses performed on the McMC results detailed below. Together, the terms in eq. (15) account for the uncertainty in model structure due to both geophysical parameter and physical property relationship uncertainties and are included in the workflow summarized in Fig. 2. Although the terms in eq. (16) are defined from an AEM data-driven clustering approach in this study, a similar relationship or alternative clustering scheme could also be used to incorporate alternative forms of geological prior information from co-located borehole data (Christensen *et al.* 2017) or other existing physical property relationships. In that case, the relationships between lithology, depth and resistivity would also be conditional on borehole data and/or other prior geological data.

A two-step approach is used to generate McMC-derived estimates of the resistivity–lithology relationships described in the first term on the right-hand side of eq. (15). First, a Student's-T mixture model is fit to the probability density function (pdf) of the resistivity posterior at each discrete location, **r**, using both the frequency- and time-domain data sets separately, where the number of components in the mixture model at each position is given by nST(r). Second, the Student's-T mixture model results are used to evaluate the probability of each lithology for every model location using eq. (16).

The first step (Fig. 2b) is summarized in Algorithm 1, which describes the process of generating a mixture model from a pdf. In Algorithm 1, a model is defined as a weight (w), mean ($\bar{\rho}$), variance (σ^2) and degrees of freedom (υ) for each Student's-T component in the mixture. The sum of the Student's-T mixture is fit to the pdf extracted from each location and depth from the McMC analysis. Scipy's 'curve_fit' algorithm, which uses a rectangular trust region dogleg approach for nonlinear bound-constrained optimization problems (Voglis & Lagaris 2004), is used to fit the sum of the Student's-T mixture to the pdf. Initially, only a single Student's-T component is fit to the pdf, and additional components are iteratively added to the mixture until an acceptable fit is achieved. Two input parameters define convergence; ϵdefines the tolerance on the absolute fit, while μdefines the tolerance on the change in absolute fit as mixture components are added and reduces overfitting of noisy pdfs. Each time a Student's-T mixture is fit to a pdf, an initial guess and lower and upper bounds are specified for

each parameter in the model. Weights (w) are initialized to one and have a lower bound of zero. For the first iteration, the initial mean ($\bar{\rho}$) is equal to the centre of the histogram bin that corresponds to the maximum of the pdf, and each subsequent mean is equal to the centre of the bin that has the highest residual after subtracting the fit of the previous model. The bounds for each mean are defined by their bin edges. By imposing these bounds on the mean, each component is assumed to have zero skewness. The variance is initialized to $\sigma^2_{\max}/2$, where $\sigma^2_{\max}$ is the upper bound on variance and calculated using a stochastic estimate of the histogram's variance. The upper bound on variance is a necessary requirement that prevents components with unfeasibly large variances that span many orders of magnitude of log resistivity. The degrees of freedom are initialized to a large value (e.g. 10 000) and have a lower bound of 2 such that the variance of each Student's-T component is well defined.

Algorithm 1. Mixture model fitting to a pdf.

Input:
$h1D \leftarrow 1D\ histogram$
$\epsilon \leftarrow \%\ absolute\ tolerance$
$\mu \leftarrow relative\ tolerance$
Result:
Mixture Model
$f(\boldsymbol{x}) \leftarrow sum\ of\ Student's - T\ distribtions$
$\boldsymbol{y} \leftarrow estimate\ pdf\ from\ histogram\ counts$
$\rho_{guess} \leftarrow bin\ centre\ with\ the\ highest\ count$
$\sigma^2_{max} \leftarrow estimate\ of\ variance\ from\ histogram$
$\boldsymbol{guess} \leftarrow [w,\ \rho_{guess},\ \frac{\sigma^2_{max}}{2},\ \nu]$
$\boldsymbol{bounds} \leftarrow set\ bounds for\ \boldsymbol{guess}$
$\boldsymbol{model} \leftarrow fit\ f(\boldsymbol{guess})\ to\ \boldsymbol{y}\ given\ \boldsymbol{bounds}$
$\boldsymbol{y^{pre}} = f(\boldsymbol{model})$
$fit_k = \frac{||\boldsymbol{y} - \boldsymbol{y^{pre}}||_\infty}{||\boldsymbol{y}||_\infty}$
while $(fit_k > \epsilon) \wedge (|fit_k - fit_{k-1}| > \mu)$ **do**
$\rho_{new} \leftarrow bin\ centre\ of\ \max(\boldsymbol{y} - \boldsymbol{y^{pre}})$
$\boldsymbol{guess} \leftarrow append\ with\ [w, \rho_{new},\ \frac{\sigma^2_{max}}{2},\ \nu]$
$\boldsymbol{bounds} \leftarrow set\ bounds\ for\ \boldsymbol{guess}$
$\boldsymbol{model} \leftarrow fit\ f(\boldsymbol{guess})\ to\ \boldsymbol{y}\ given\ \boldsymbol{bounds}$
$\boldsymbol{y^{pre}} = f(\boldsymbol{model})$
$fit_k = \frac{||\boldsymbol{y} - \boldsymbol{y^{pre}}||_\infty}{||\boldsymbol{y}||_\infty}$
End
Return model

The second step in generating lithology–resistivity relationships using eq. (16) involves the evaluation of fitted Student's-T mixture models, both locally and globally throughout the subsurface model (Fig. 2c). To achieve this, the mean of every Student's-T mixture model is input as data to a Gaussian mixture model (GMM). Gaussian mixture modelling uses expectation maximization to maximize the likelihood of each data point to each component in the Gaussian mixture, where the number of components, or classes, is defined by the user. The weights, means and variances of each component are iteratively updated until this maximization is achieved. Output from the mixture model are distributions of mean resistivity values for each class. There are three parts to this step (eq. 16, Figs 2d–f) that use the GMM and Student's-T mixture models: (1) The global fraction for each class is calculated from the relative contribution of each component in the GMM. The global fraction for each class is used as part of the prior in eq. (16) in order to weight class distributions at individual locations according to their global support volume. (2) At each position, **r**, the probability of each class is calculated by evaluating the GMM at the mean resistivity value for each fitted Student's-T distribution. For example, if the mean of the Student's-T falls near the mean of a GMM component, it will have a relatively high probability of being associated with that component. (3) At each position, **r**, the fitted Student's-T distributions are considered to represent empirical position-dependent distributions of resistivity values for each component in the GMM. The combined result of these three parts is an estimate of the spatially varying resistivity–lithology relationship given in eq. (16). Finally, this term is combined with the McMC-derived uncertainty in resistivity at the same position to estimate the joint probability of lithology at that position, marginalized over resistivity (eq. 15, Fig. 2g).

Example data sets

We use a portion of the AEM data presented by Bedrosian *et al.* (2015) to illustrate the methods developed here. Data sets include both frequency-domain (RESOLVE) and time-domain (AeroTEM-IV) AEM surveys acquired over the Morrill Test Block (MTB) in western Nebraska in 2008. The MTB covers an area of approximately 40 km^2 (Fig. 3) and comprises ten 22-km-long flight lines oriented north-to-

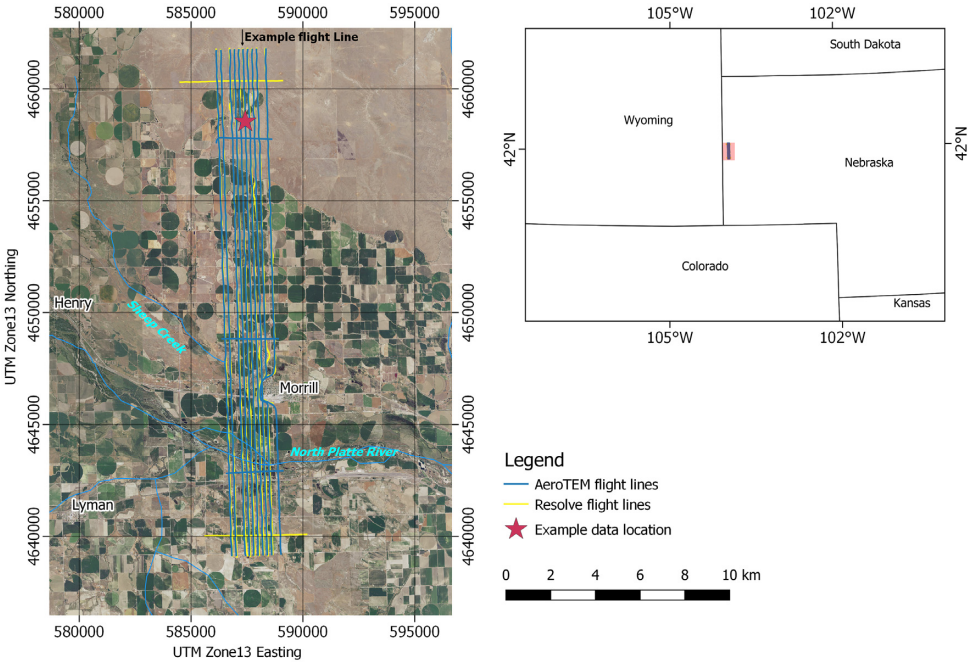

Figure 3. Survey area map. Coincident flight paths for the Resolve and AeroTEM systems are indicated along with nearby boreholes. The red star indicates the location of the frequency- and time-domain inversion results in Figs 4, 5 and 7, while the arrow indicates the flight lines used in Figs 6, 8 and 11.

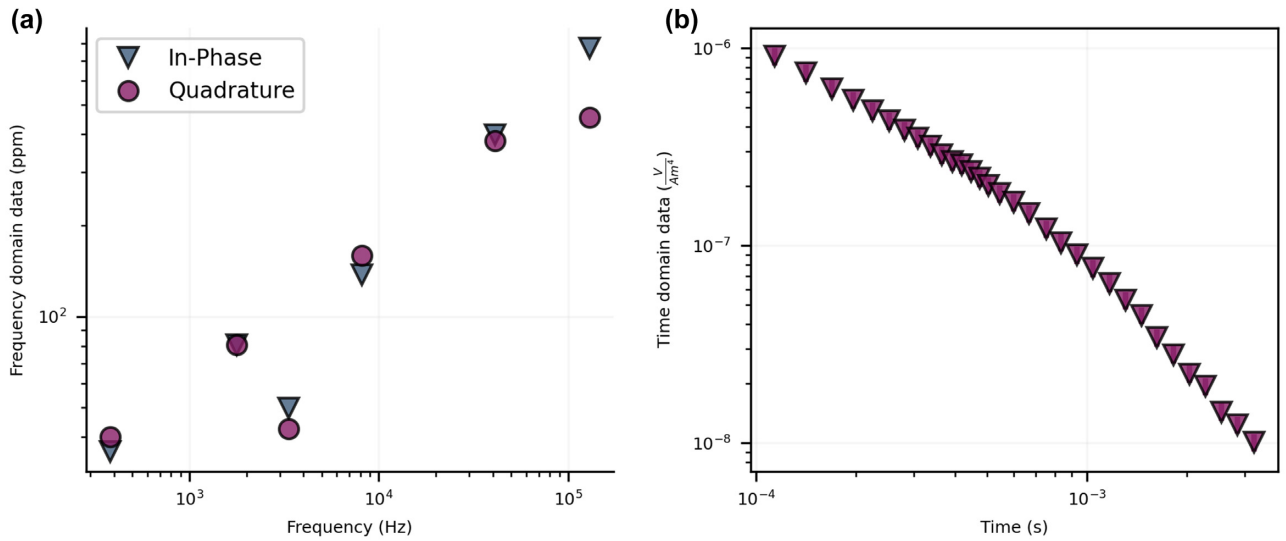

Figure 4. Single sounding curve examples for different AEM data types. (a) Resolve frequency-domain data shown as in-phase and quadrature responses for six different measurement frequencies and (b) AeroTEM time-domain decay curve. Data and subsequent results are from the starred location indicated in Fig. 3.

south with 200-m line spacing. Quaternary sands and gravels make up a shallow aquifer, with broad channels of the present-day and ancestral North Platte River system incised into the Tertiary Brule Formation (Verstraeten *et al.* 2001; Cannia *et al.* 2006).

The RESOLVE frequency-domain instrument records in-phase and quadrature data on six separate coil pairs with frequencies that range from 0.4 to 140 kHz. Five of the coil pairs are oriented as vertical magnetic dipoles with 7.9 m separation between transmitter and receiver, and the sixth coil pair uses a horizontal magnetic dipole with 9.3 m separation. The AeroTEM-IV is a rigid frame central-loop time-domain instrument with 115 m^2 transmitter area and 237 000 A m^2 peak moment. Off-time measurements are recorded from 58 to 3200 μs (with zero time defined at the end of the transmitter turn-off). Additional system details and data processing steps are described by Bedrosian *et al.* (2015). The data analysed here include 67 470 frequency-domain and 93 218 time-domain locations, with an example of a single sounding of each datatype in Fig. 4. The only difference with the input data used here compared with Bedrosian *et al.* (2015) is that errors are estimated as part of the McMC algorithm and are not assumed to be known *a priori.*

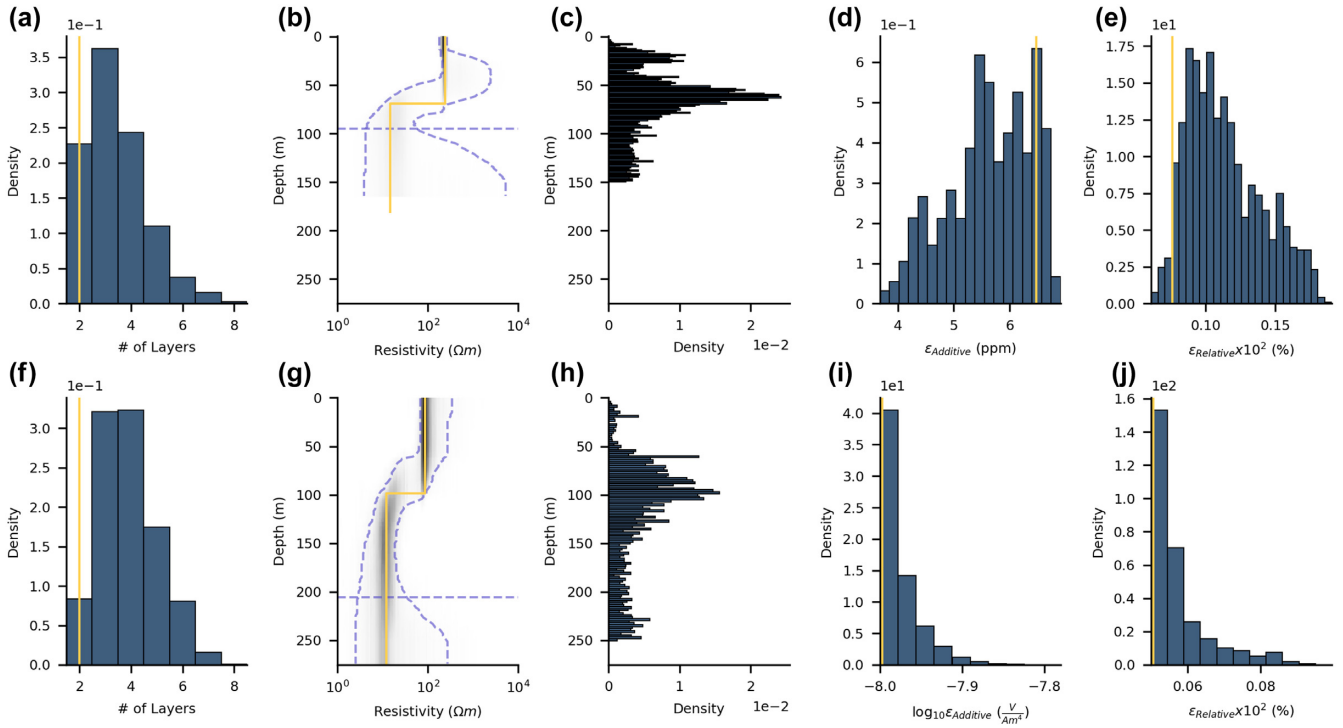

Figure 5. 1-D posterior statistics for (a–e) FDEM and (f–j) TDEM data. Distribution of the number of layers used in the ensemble of McMC models for the (a) FDEM and (f) TDEM data sets at a single location indicated in Fig. 3. Solid yellow lines indicate the number of layers for the model with the greatest posterior probability. The posterior probability distribution of resistivity models as a function of depth is shown for (b) FDEM and (g) TDEM. Dashed blue lines indicate the 90 per cent credible region, and the most probable model is shown in yellow. The estimated depth of investigation is indicated by the horizontal dashed line. The posterior probability distribution of layer interface depths for the entire ensemble of models for (c) FDEM and (h) TDEM indicates the likely locations of model interfaces. Output distributions for (d and i) additive and (e and j) relative errors show estimates of optional hierarchical model parameters.

RESULTS

For each data location, the McMC algorithm samples 100 000 models post burn-in according to the methods described in this paper. Burn-in is reached when the normalized chi-squared measure of data misfit reaches a target value of 1.0. Model parameters include the number of layers (k), conductivity ($\boldsymbol{\sigma}$) and interface depths ($\mathbf{z}$) for each layer, as well as relative and additive errors (ε^{re}, ε^{ae}). For any measurement location, posterior statistics can be compiled for any of the estimated model parameters such as the marginal distribution for the number of model layers (Figs 5a and f), the distribution of sampled resistivity models as a function of depth (Figs 5b and g), or the distribution of interface locations (Figs 5c and h). The depth of investigation, or an estimate of the depth to which model parameters are no longer sensitive to the data, is calculated from the ensemble of McMC models by computing the depth at which the range in the 90 per cent credible interval reaches 67 per cent of the range at the deepest depth (Figs 5b and g). The posterior distribution of additive (Figs 5d and i) and relative (Figs 5e and j) errors indicates the McMC-estimated errors for the FDEM and TDEM soundings. Sampled TDEM errors (Figs 5i and j) tend to be biased towards the lower end of the prior range for these parameters (10^{-8} and 5 per cent for additive and relative errors, respectively). Although the McMC algorithm could sample smaller error values with acceptable data fit, the lower bounds used here are intended to avoid overfitting of unrealistic model structure caused by smoothing filters applied during pre-processing of the TDEM data (Bedrosian *et al.* 2015).

Comparison of our recovered resistivity cross-section with a deterministic inversion along the flight path marked in Fig. 3 (Resolve line number 30 060/AeroTEM line number 1060) is shown in Fig. 6. The deterministic inversion (Figs 6a and b) used a spatially constrained inversion (Auken *et al.* 2005; Viezzoli *et al.* 2008) and is originally presented in Bedrosian *et al.* (2015). Summary cross-sections of McMC results are shown as the mean resistivity model calculated from the posterior ensemble of resistivity models at each individual depth and measurement location (Figs 6c and d) and the maximum *a posteriori* model for each location (Figs 6e and f) identified as the model with the greatest product of prior probability and likelihood. The conservative depth of investigation metric described by Christiansen & Auken (2012) is used to limit the depth to which models are displayed for the deterministic results (Figs 6a and b). Transparency shading is applied to the McMC models at depth according to the width of the 90 per cent credible interval of the posterior conductivity distribution at each measurement location and depth (Figs 5b and e). Where the credible interval is large, model confidence is low (Figs 6g and h), typically due to low data sensitivity at depth or in highly resistive layers. The use of a Bayesian formulation helps to highlight some of the differences in uncertainty between the frequency and time domain inversions.

The mean and maximum *a posteriori* models (Figs 6c–f) are a useful summary of results, but do not convey the comprehensive details on parameter uncertainty obtained through the McMC analysis presented here. An assessment of the uncertainty in resistivity values within any

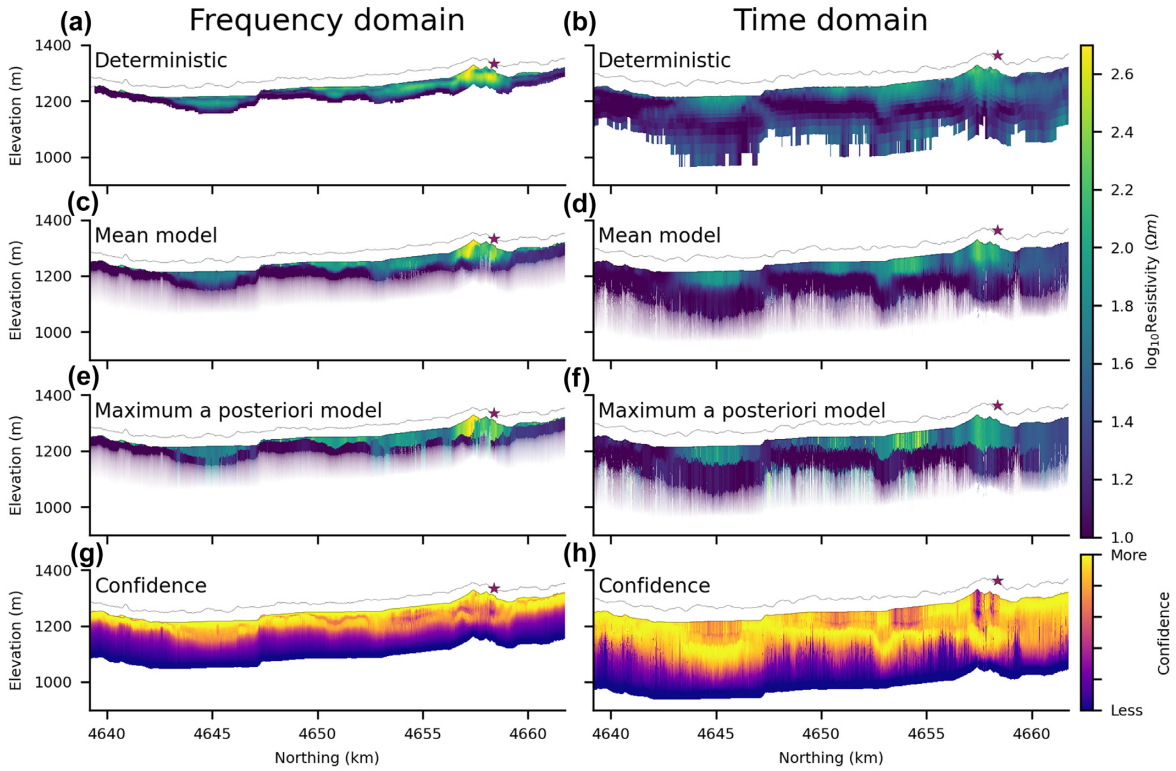

Figure 6. Inversion results for the flight line indicated in Fig. 3. Deterministic inversion results for the (a) FDEM and (b) TDEM systems are reproduced from Bedrosian *et al.* (2015). McMC results include the (c and d) mean and (e and f) maximum *a posteriori* models. The confidence, derived from the width of the 90 per cent credible interval, is shown in (g) and (h). The star indicates the location of results shown in Figs 4, 5 and 7. Thin grey line denotes flight elevation for each AEM system.

depth interval can further be made by extracting values from Fig. 5(b) or Fig. 5(g) within the desired depth window. Fig. 7 shows examples taken from four different depths, with each panel comparing the range of plausible resistivity values for the FDEM and TDEM systems. As expected, the FDEM instrument shows less uncertainty in resistivity in the shallowest depth (Figs 7a and b), whereas the TDEM system is able to resolve features at greater depths where the FDEM system has low sensitivity (Figs 7c and d). Although the TDEM instrument cannot distinguish between low- and high-resistivity layers at 100 m depth (Fig. 7c), it identifies sensitivity to both layers whereas the FDEM instrument does not. While a joint inversion of both data sets may improve overall resolution and compatibility of posterior results (Bodin *et al.* 2012a; JafarGandomi & Binley 2013; Blatter *et al.* 2019; Visser & Markov 2019), we have not focused on that approach here since co-located AEM data sets are extremely uncommon. Instead, we focus on a comparison of posterior uncertainty estimates from each data set independently.

In cross-section view, a grey-scale image of the probability of interface depths at each location (Figs 5c and h) indicates the geometry of model interfaces along with information about their uncertainty (Fig. 8). Shaded regions indicate locations with high probability of model interfaces, where the intensity and width of the shaded region indicates uncertainty in the interface location.

The McMC outputs are used to empirically infer the probability of different lithological units identified by the AEM data based on the posterior clustering approach described in Fig. 2 and summarized in Fig. 9. Both data types are combined here in order to maintain the same cluster groups for comparison in subsequent analyses; however, this comes at some expense of greater uncertainty within each class because of the differences in resolution for each method (Figs 5 and 7). In most cases only a single AEM data set is available for one study area, and this approach would be carried out for a single data type.

Resistivity posteriors are extracted from the McMC results at three different locations, **r** (black curves; Figs 9a–c). The dashed white line shows the Student's-T mixture components that were fit to the posterior using Algorithm 1. In each case, Algorithm 1 determined the number of components required to fit the posterior given tolerance values of $\epsilon = 0.05$ and $\mu = 0.05$. We applied Algorithm 1 to the McMC results at all data locations and discrete depths throughout both the frequency- and time-domain data sets ($N = 20\,960\,002$). This step results in a Student's-T mixture model at each position with nST(r) components, where each component has a weight, mean, variance and degrees of freedom.

For the purpose of demonstration, a five-component GMM was chosen to represent the variability within and between the identifiable lithologic units in the MTB survey. Here, classes 1–2 map to spatially varying patterns within the deeper Brule Formation, while classes 3–5 capture spatial trends within the Quaternary aquifer. In practice, selection of the most appropriate number of layers may come from testing different values informed by knowledge of site-specific geology. The intent here is not to identify an optimal number of classes, but rather

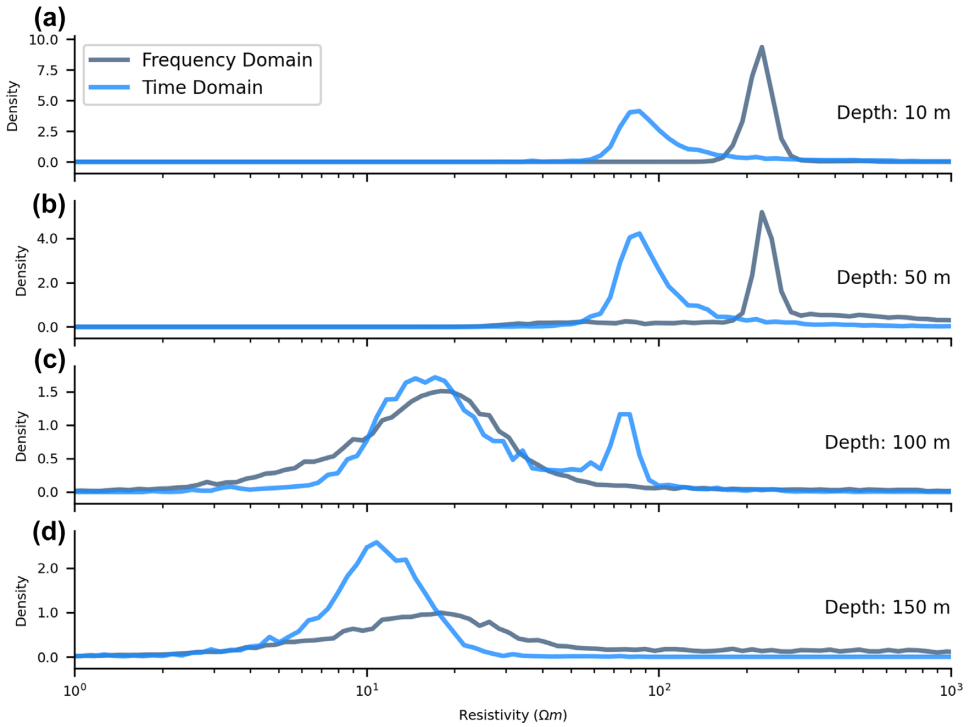

Figure 7. 1-D marginal probability distribution of resistivity at four discrete depths. The posterior probability distribution of resistivity values extracted from the FDEM (Fig. 5b) and TDEM (Fig. 5g) results indicates the ability of the different instruments to resolve resistivity values at different depths in this geologic environment.

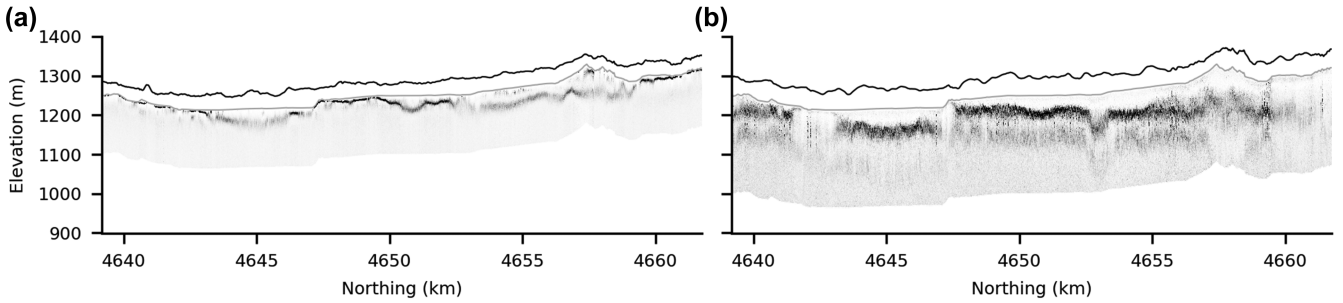

Figure 8. Interface probability cross-sections. Shading indicates the probability of an interface at different locations along the flight profile for (a) FDEM and (b) TDEM data. Sharp, dark-shaded areas indicate well-defined peaks in the posterior interface distributions shown in Figs 5(c) and (h), whereas diffuse shading indicates lower probability of an interface occurring at a particular location. Solid black and grey lines indicate AEM system and land surface elevation, respectively.

to present a workflow that accurately quantifies model structural uncertainty given the subjective choice of how many classes is appropriate. Fig. 9(d) presents the five global distributions that result from fitting the GMM to the histogram representation of all Student's-T mixture means, with the relative fraction of each component, $p(\text{lith}_i|\boldsymbol{d}_{\text{AEM}})$, annotated in Fig. 2(d). At each position, the other components in eq. (16) are evaluated (Figs 2e and f). Fig. 9(e) shows the combined numerator of eq. (16) for the two-component Student's-T mixture model in Fig. 9(c).

Using eqs (15) and (16) to combine the uncertainty in resistivity values (Figs 5b and g, and 7) with the uncertainty in the physical property relationship (Fig. 9), an estimate of uncertainty in predicting lithology at any location and depth can be generated (Fig. 2g). An example of the probability for each of the five lithologies (which sum to 1) as a function of depth is shown for three locations along the frequency-domain flight line (Figs 10d–f) for the corresponding resistivity–depth probability distributions in Figs 10(a)–(c) and resistivity–lithology relationships developed at each position. Cross-sections for the entire flight profile of the marginal probability for each lithology class are shown in Fig. 11 using both FDEM (Figs 11a–e) and TDEM (Figs 11g–k) data sets, along with the pointwise highest marginal facies probability at each location (Figs 11f and l). Similar results for the entire frequency-domain data set are illustrated as map-view depth slices (Fig. 12).

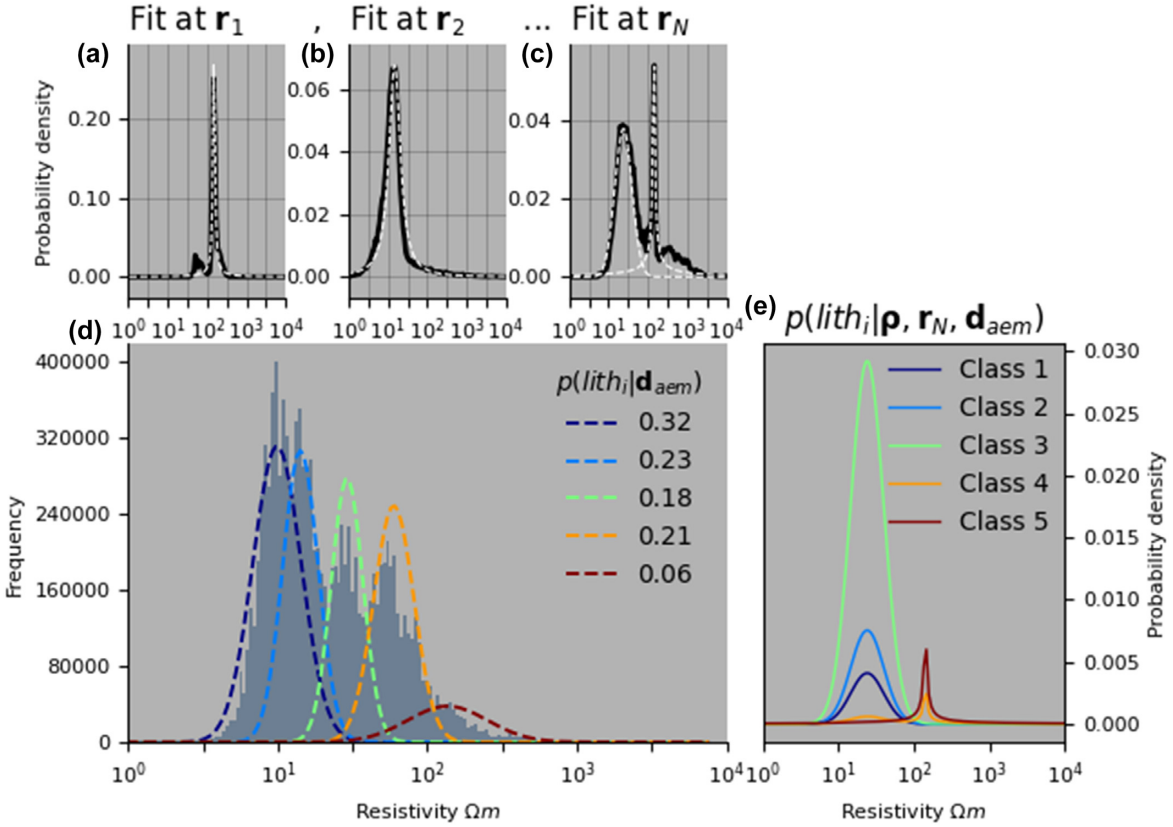

Figure 9. Building a resistivity–lithology relationship. (a–c) Examples of Student's-T mixture models fit to three different positions extracted from the McMC output. (d) Gaussian mixture model with five components (coloured curves) fit to all Student's-T means from the entire data set (grey histogram). The relative fraction of each Gaussian mixture model class is indicated in the legend. (e) Combined resistivity–lithology estimate from eq. (16) at position *N* indicated in (c).

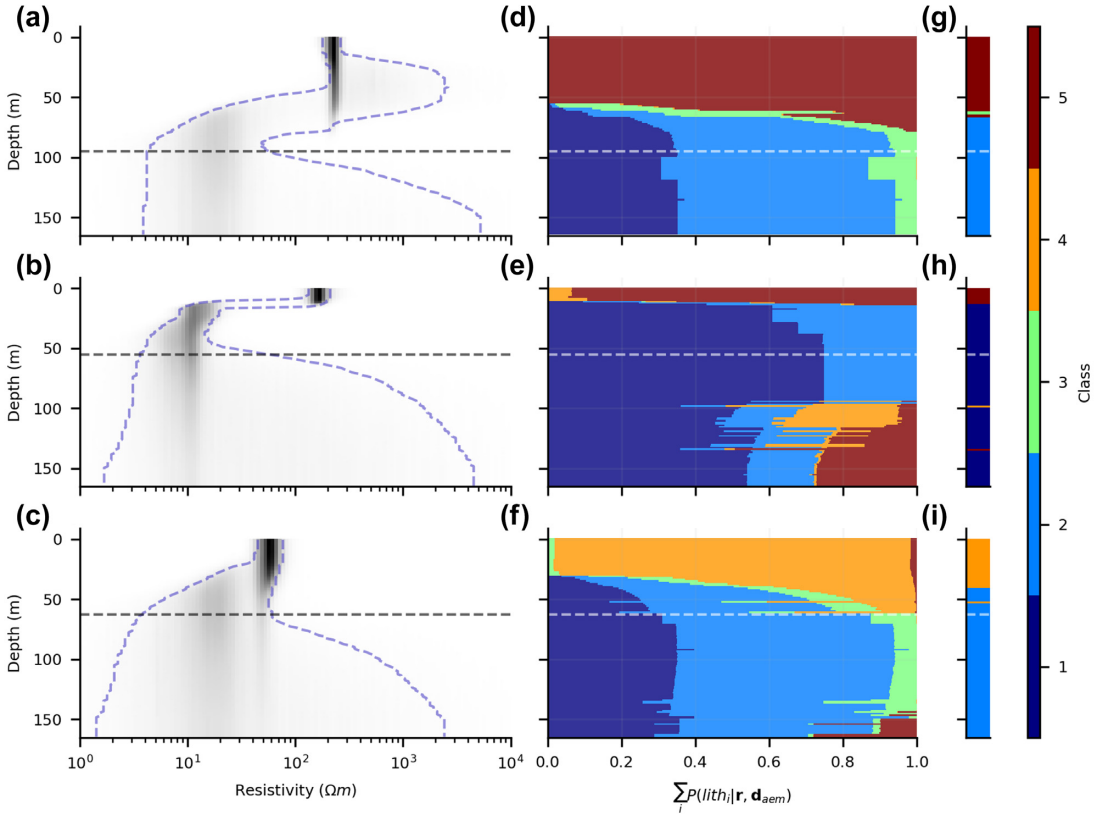

Figure 10. Marginal joint probability of $p(\text{lith} \mid \rho, z)$ for each of the five identified classes at three different locations along the flight line shown in Fig. 4. Posterior resistivity–depth distributions for the three locations are shown in (a)–(c). The probabilities for each lithology class shown in the middle column (d–f) are computed using eqs (15) and (16), with the most probable lithology class at each depth in the right-hand column (g–i).

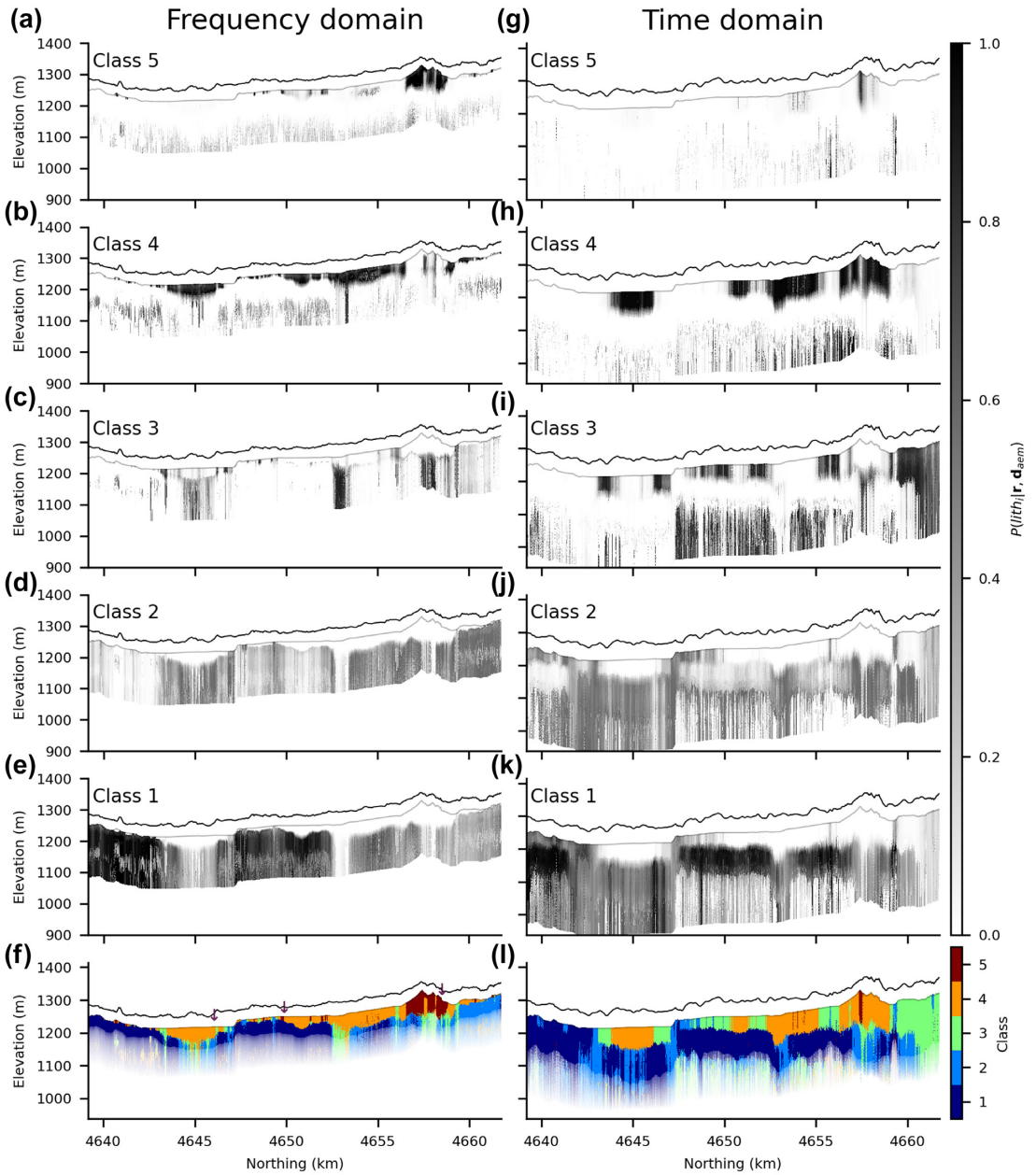

Figure 11. Probabilistic estimates of model structure. Results combine uncertainties in geophysical parameters and physical property relationships (Fig. 2). Frequency-domain results are shown in the left-hand column and time-domain results are in the right-hand column. Each row is a different facies class, starting with (a and g) class 5 through (e and k) class 1. At each pixel, the class with the pointwise highest marginal probability is illustrated in (f) and (l). Arrows in (f) indicate the locations of 1-D results in Fig. 10.

DISCUSSION

One challenge with the analysis of McMC results is the communication of uncertainty and various derived statistics calculated from the large ensemble of models sampled from the posterior distribution. For the modest-size AEM surveys presented here, 6.4×10^9 and 9.3×10^9 resistivity models were generated from the FDEM and TDEM data sets, respectively. Maps and cross-sections of mean or most probable results (Fig. 6) are useful for summarizing trends in model structure, but do not convey the wealth of information about parameter uncertainty contained in the 100 000 sampled models at each measurement location in the survey (Figs 5 and 7). Spatial information about parameter uncertainty (Figs 6g and h, and 8) can be useful, but are only two examples of the type of information that can be produced. For example, alternative questions that can be addressed from analysis of the ensemble of McMC model results can include: What is the probability that resistivity is less than X within a particular depth interval? Or, what is the depth at which resistivity drops below Y with a 50, 75, or 90 per cent confidence? These are just a few examples of the many types of analyses that can be facilitated by querying the posterior distribution of model results (e.g. Ray & Key 2012).

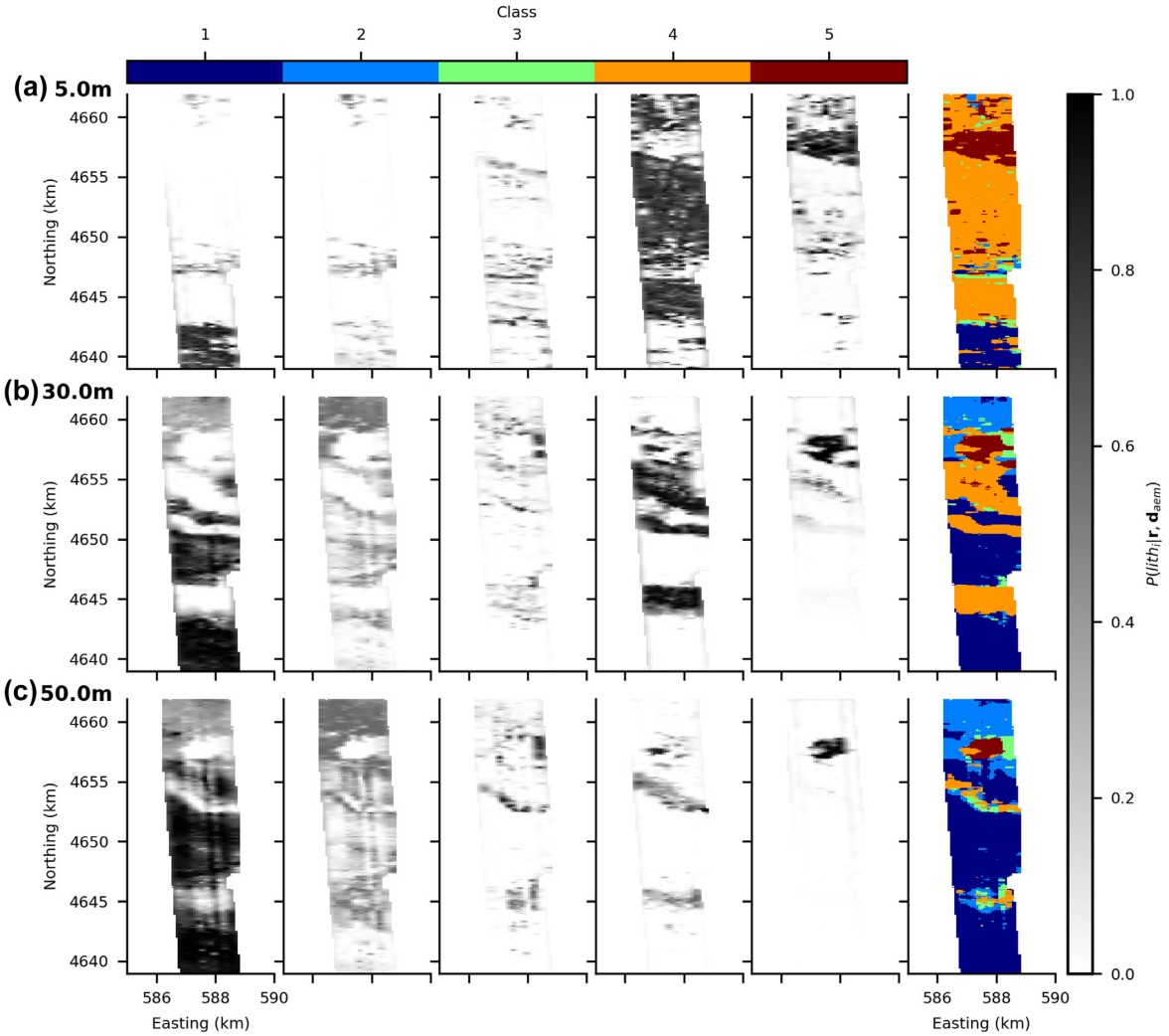

Figure 12. Map-view representation of the probabilistic estimates of model structure. Results combine uncertainties in geophysical parameters and physical property relationships for the frequency-domain data set. (a–c) Each row is a different depth while each column is a different class, starting with class 1 through class 5. The class with the pointwise highest marginal probability is illustrated in the right-most column.

Estimates of model confidence calculated from the width of the 90 per cent credible region (Figs 6g and h) are useful, but again do not fully convey all of the details about parameter uncertainty. For example, the McMC algorithm accurately reflects the asymmetry in parameter uncertainty, where uncertainty is greater to the resistive side of the mean or maximum a posteriori model value compared with the conductive side, due to the limited sensitivity of AEM data to resistive layers (Figs 5b and g). Parameter uncertainty calculated in a deterministic algorithm from the posterior covariance matrix will always be symmetric about the estimated parameter value. Asymmetry in uncertainty values towards the resistive end can lead to a large credible interval (low confidence) for shallow resistive layers (Figs 6g and h). However, while the resistivity of the resistive layer may not be very well constrained, we can say with relatively high confidence that the layer is in fact resistive, and that the lower-bound of resistivity is well defined (Figs 5b and g). In that case, there may be high confidence in the geological interpretation of this layer (i.e. its resistivity is greater than a certain value and thus too high to contain clay), even though its exact resistivity may be uncertain. This is one example of how the McMC approach along with derived geological interpretation is intended to falsify possible scenarios (Tarantola 2006), rather than to focus on any single particular result supported by the data.

It is well known that the choice of a prior distribution can be highly consequential in the outputs of a Bayesian algorithm (Scales & Snieder 1997; Tarantola 2006). Here, we have implemented relatively simple prior options that follow traditional deterministic algorithms where model parameters that are close to a reference half-space value and/or models that are smooth are favoured. However, the Bayesian McMC modelling framework presented here allows for more structured or geologically based prior information to be incorporated into the inference problem (Hansen & Minsley 2019).

Of the two main sources of uncertainty in quantifying model structure addressed here—uncertainty in geophysical parameters and uncertainty in the relationship between geophysical and geological properties of interest—the latter is a far more challenging and subjective

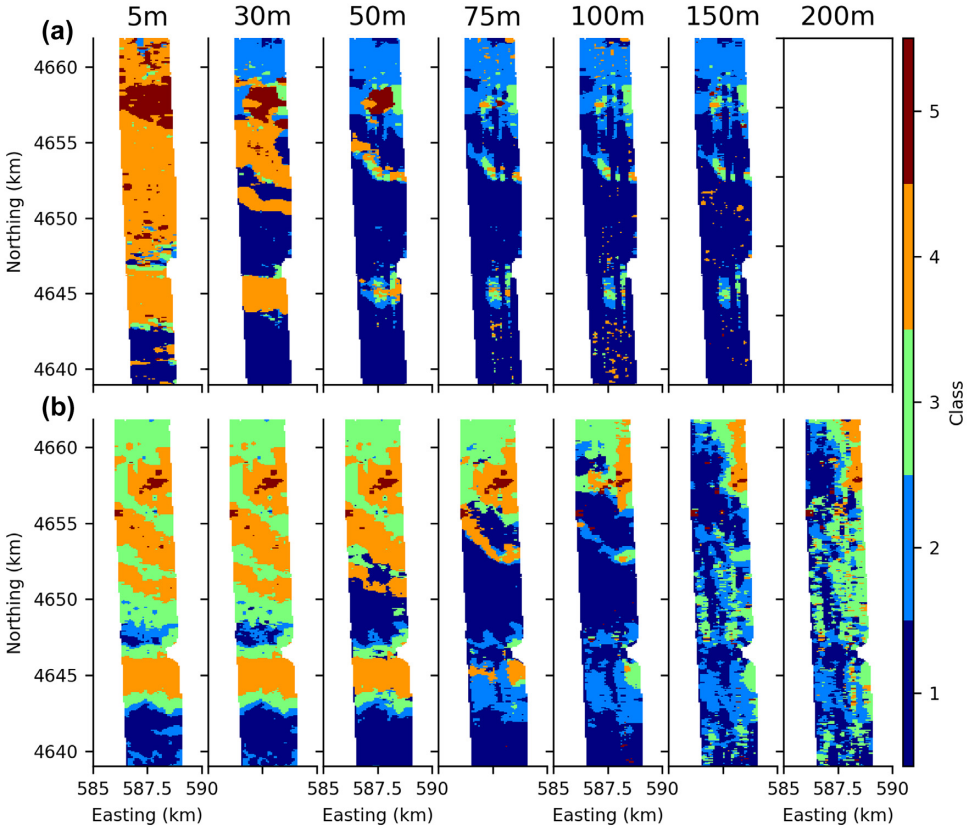

Figure 13. Comparison of the most probable class for frequency-domain and time-domain data sets. Depth-slice maps of the most probable class for seven discrete depths through the FDEM results (row a) and TDEM results (row b). The frequency domain models do not extend as deep as the time domain models, hence the missing map at 200 m in (a).

task. Quantifying geophysical parameter uncertainty using the GeoBIPy algorithm largely relies on the well-defined physics of electromagnetic induction that governs AEM instruments. Making the link between resistivity and geological properties, however, is far more challenging because of the sometimes ambiguous, uncertain, and site-specific relationships involved. Where possible, making use of co-located geophysical and borehole data can be advantageous (Christensen *et al.* 2017); however, borehole observations must be relevant to the physical property being modelled (e.g. geology, lithology and hydrostratigraphy) Our approach is to provide a framework that allows flexibility in building the relationship between resistivity and physical properties of interest, where relationships may also be allowed to vary spatially according to eq. (16).

As with any algorithm, subjective choices and assumptions must be made. In addition to the use of a 1-D resistivity model with natural parsimony and priors described earlier, subjective decisions are also necessary in mapping resistivity values to geologic classes, and these choices have implications for the resulting estimates of uncertainty in model structure. For example, the user must decide how many classes to fit; ideally, this choice is guided by knowledge of site-specific geological structure but may also be informed by empirical observations of the number of distinguishable features in the McMC outputs. In practice, several trials with different numbers of classes, or modification to incorporate additional geological prior information following eq. (16), is necessary to understand the consequences of different subjective choices on the ultimate assessment of uncertainty in model structure. In this study, we find that classes 4 and 5 clearly capture the resistive sands and gravels of the Quaternary aquifer system, with class 1 primarily representing the Tertiary Brule Formation (Figs 11 and 12). Although these three classes capture the majority of structural variability in this data set, subtle transitions possibly related to textural changes within class 1 are captured in classes 2 and 3.

Realizations of model structure, such as the maximum *a posteriori* structural model (Fig. 13), provide useful insight into the combined resolving power of mapping structure given both the physical differences in AEM instruments along with the assumptions about resistivity–lithology relationships. For example, the FDEM result contains greater detail of shallow features, whereas the TDEM system is better able to resolve deep structures. In practice, however, the probabilistic estimates of model structure (Figs 11 and 12) may be incorporated as soft data in geostatistical simulations using sequential indicator simulation or multipoint statistics to generate multiple plausible realizations of model structure consistent with all observations, assumptions, and prior information (He *et al.* 2014; Christensen *et al.* 2017; Marker *et al.* 2017; Vilhelmsen *et al.* 2019). This same approach can be extended to classification and uncertainty quantification of other physical properties as well, such as groundwater salinity (Ball *et al.* 2020).

CONCLUSIONS

AEM data are increasingly used to characterize subsurface structure over watershed-basin scales because of their unique capability to map subsurface properties over large areas. The ability to quantify uncertainty in AEM results—both in terms of geophysical parameter estimation and interpretations of model structure—are critically important for advancing the value of these data in geologic and hydrologic studies. We have addressed geophysical parameter uncertainty through the development of an open-source Python software package, GeoBIPy. GeoBIPy allows for uncertainty quantification of frequency- or time-domain AEM data using a Bayesian McMC approach that solves for 1-D resistivity structure and can run many individual soundings in parallel to analyse large airborne data sets efficiently. Future efforts will incorporate functionality in GeoBIPy for uncertainty quantification using other ground-based geophysical methods including electromagnetics, magnetotellurics and nuclear magnetic resonance.

In addition to quantifying geophysical parameter uncertainty, we also present an approach for incorporating uncertainty in the physical relationship between resistivity, depth and lithology through a posterior clustering analysis of the McMC results that could also be adapted to include other geological prior information. Uncertainty in the physical property relationship does not involve a well-defined, physics-based relationship in the same way that geophysical parameter uncertainty quantification does and therefore requires subjective decisions about how geophysical and geological properties are related. By combining estimates of geophysical parameter uncertainty with relationships between geophysical and geological properties, we provide a framework for quantifying uncertainty in model structure that can be used to inform management decisions or guide future areas of data collection in regions of high uncertainty.

ACKNOWLEDGEMENTS

Support for this work was provided by the U.S. Geological Survey (USGS) Interdisciplinary Methods and Applications in Geophysics (IMAGe) project through Mineral Resources Program funding, as well as the USGS Advanced Research Computing group. The inversion results for this study were produced using the high-performance computing resources of the USGS Advanced Research Computing group (Falgout & Gordon 2017). The authors thank Ross C. Brodie at Geoscience Australia for his support of the TDEM forward solver (Brodie 2017) used in our algorithm. Reviews provided by Ben Bloss, Anandaroop Ray and one anonymous reviewer contributed significantly to the paper. Any use of trade, firm or product names is for descriptive purposes only and does not imply endorsement by the U.S. Government.

Author contributions: BJM conceived the initial McMC algorithm and methods developed in this paper. NLF created the open-source Python implementation of the code and enhanced its functionality through parallelization and other efficiencies and developed the resistivity–lithology relationship. PAB facilitated use of the example data set illustrated here and provided input on functioning of the algorithm.

REFERENCES

Abraham, J.D., Cannia, J.C., Bedrosian, P.A., Johnson, M.R., Ball, L.B. & Sibray, S.S., 2012. Airborne electromagnetic mapping of the base of aquifer in areas of western Nebraska, U.S. Geological Survey Scientific Investigations Report, No. SIR-2011-5219.

Aster, R.C., Borchers, B. & Thurber, C.H., 2005. *Parameter Estimation and Inverse Problems,* Elsevier.

Auken, E., Christiansen, A.V., Jacobsen, B.H., Foged, N. & Sorensen, K.I., 2005. Piecewise 1D laterally constrained inversion of resistivity data, *Geophys. Prospect.,* **53,** 497–506.

Auken, E. *et al.*, 2015. An overview of a highly versatile forward and stable inverse algorithm for airborne, ground-based and borehole electromagnetic and electric data, *Explor. Geophys.,* **46,** 223–235.

Ball, L.B., Davis, T.A., Minsley, B.J., Gillespie, J.M. & Landon, M.K., 2020. Probabilistic categorical groundwater salinity mapping from airborne electromagnetic data adjacent to California's Lost Hills and Belridge oil fields, *Water Resour. Res.,* **56,** doi:10.1029/2019WR026273.

Beamish, D., 2003. Airborne EM footprints, *Geophys. Prospect.,* **51,** 49–60.

Bedrosian, P.A., Schamper, C. & Auken, E., 2015. A comparison of helicopter-borne electromagnetic systems for hydrogeologic studies, *Geophys. Prospect.,* **64,** doi:10.1111/1365-2478.12262.

Best, M.E., Levson, V.M., Ferbey, T. & McConnell, D., 2006. Airborne electromagnetic mapping for buried quaternary sands and gravels in Northeast British Columbia, Canada, *J. Environ. Eng. Geophys.,* **11,** 17–26.

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,** 1919–1936.

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

Bodin, T., Sambridge, M., Gallagher, K. & Rawlinson, N., 2012a. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117,** doi:10.1029/2011JB008560.

Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012b. Transdimensional tomography with unknown data noise, *Geophys. J. Int.,* **189,** 1536–1556.

Bosch, M., Zamora, M. & Utama, W., 2002. Lithology discrimination from physical rock properties, *Geophysics,* **67,** 573–581.

Bredehoeft, J., 2005. The conceptualization model problem—surprise, *Hydrogeol. J.,* **13,** 37–46.

Brodie, R., 2017. *GA-AEM Source Code Repository- Geoscience Australia Airborne Electromagnetics Programs,* Geoscience Australia, Retrieved from https://github.com/GeoscienceAustralia/ga-aem.

Brodie, R. & Sambridge, M., 2006. A holistic approach to inversion of frequency-domain airborne EM data, *Geophysics,* **71,** G301–312.

Campanyà, J. *et al.*, 2019. Subsurface characterization of the Pennsylvanian Clare Basin, western Ireland, by means of joint interpretation of electromagnetic geophysical data and well-log data, *J. geophys. Res.,* **124,** 6200–6222.

Cannia, J., Woodward, D. & Cast, L., 2006. Cooperative hydrology study COHYST hydrostratigraphic units and Aquifer characterization report, *Publ. US Geol. Surv.,* **102,** 96.

Christensen, N.B., Reid, J.E. & Halkjaer, M., 2009. Fast, laterally smooth inversion of airborne time-domain electromagnetic data, *Surf. Geophys.,* **7,** 599–612.

Christensen, N.B., Fitzpatrick, A. & Munday, T., 2010. Fast approximate 1D inversion of frequency domain electromagnetic data, *Surf. Geophys.,* **8,** 1–15.

Christensen, N.K., Minsley, B.J. & Christensen, S., 2017. Generation of 3-D hydrostratigraphic zones from dense airborne electromagnetic data to assess groundwater model prediction error, *Water Resour. Res.,* **53,** 1019–1038.

Christiansen, A.V. & Auken, E., 2012. A global measure for depth of investigation, *Geophysics,* **77,** WB171–WB177.

Collette, A., 2013. *Python and HDF5: Unlocking Scientific Data,* 1st edn, O'Reilly Media.

Cox, L.H. & Zhdanov, M.S., 2008. Advanced computational methods of rapid and rigorous 3-D inversion of airborne electromagnetic data, *Commun. Comput. Phys.,* **3,** 160–179.

Dalcín, L., Paz, R. & Storti, M., 2005. MPI for Python, *J. Parallel Distrib. Comput.,* **65,** 1108–1115.

Falgout, J.T. & Gordon, J.M., 2017. USGS Yeti Supercomputer, U.S. Geological Survey, https://doi.org/10.5066/F7D798MJ.

Farquharson, C.G., Oldenburg, D.W. & Routh, P.S., 2003. Simultaneous 1D inversion of loop–loop electromagnetic data for magnetic susceptibility and electrical conductivity, *Geophysics,* **68,** 1857–1869.

Fitterman, D.V. & Deszcz-Pan, M., 1998. Helicopter EM mapping of saltwater intrusion in Everglades National Park, Florida, *Explor. Geophys.,* **29,** 240–243.

Foged, N., Marker, P.A., Christansen, A.V., Bauer-Gottwein, P., Jørgensen, F., Høyer, A.-S. & Auken, E., 2014. Large-scale 3-D modeling by integration of resistivity models and borehole data through inversion, *Hydrol. Earth Syst. Sci.,* **18,** 4349–4362.

Foks, N.L. & Minsley, B.J., 2020. *GeoBIPy—Geophysical Bayesian Inference in Python,* Retrieved from https://www.github.com/usgs/geobipy.

Gulbrandsen, M., Ball, L., Minsley, B. & Hansen, T., 2017. Automatic mapping of the base of aquifer—a case study from Morrill, Nebraska, *Interpretation,* **5,** T231–T241.

Hansen, T.M. & Minsley, B.J., 2019. Inversion of airborne EM data with an explicit choice of prior model, *Geophys. J. Int.,* **218,** 1348–1366.

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

Hawkins, R., Brodie, R.C. & Sambridge, M., 2018. Trans-dimensional Bayesian inversion of airborne electromagnetic data for 2D conductivity profiles, *Explor. Geophys.,* **49,** 134–147.

He, X., Koch, J., Sonnenborg, T.O., Jørgensen, F., Schamper, C. & Christian Refsgaard, J., 2014. Transition probability-based stochastic geological modeling using airborne geophysical data and borehole data, *Water Resour. Res.,* **50,** 3147–3169.

Højberg, A.L. & Refsgaard, J.C., 2005. Model uncertainty—parameter uncertainty versus conceptual models, *Water Sci. Technol.,* **52,** 177–186.

Høyer, A.-S., Jørgensen, F., Sandersen, P.B.E., Viezzoli, A. & Møller, I., 2015. 3D geological modelling of a complex buried-valley network delineated from borehole and AEM data, *J. Appl. Geophys.,* **122,** 94–102.

JafarGandomi, A. & Binley, A., 2013. A Bayesian trans-dimensional approach for the fusion of multiple geophysical datasets, *J. Appl. Geophys.,* **96,** 38–54.

Jørgensen, F. *et al.*, 2012. Transboundary geophysical mapping of geological elements and salinity distribution critical for the assessment of future sea water intrusion in response to sea level rise, *Hydrol. Earth Syst. Sci.,* **16,** 1845–1862.

Jørgensen, F., Møller, R.R., Nebel, L., Jensen, N.-P., Christiansen, A.V. & Sandersen, P.B.E., 2013. A method for cognitive 3D geological voxel modelling of AEM data, *Bull. Eng. Geol. Environ.,* **72,** 421–432.

Knight, R., Smith, R., Asch, T., Abraham, J., Cannia, J., Viezzoli, A. & Fogg, G., 2018. Mapping Aquifer systems with airborne electromagnetics in the central valley of California, *Groundwater,* **56,** 893–908.

Koch, J., He, X., Jensen, K.H. & Refsgaard, J.C., 2014. Challenges in conditioning a stochastic geological model of a heterogeneous glacial aquifer to a comprehensive soft data set, *Hydrol. Earth Syst. Sci.,* **18,** 2907–2923.

Laloy, E., Hérault, R., Jacques, D. & Linde, N., 2018. Training-image based geostatistical inversion using a spatial generative adversarial neural network, *Water Resour. Res.,* **54,** 381–406.

MacKay, D.J.C., 2003. *Information Theory, Inference, and Learning Algorithms,* Cambirdge Univ. Press.

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

Marker, P.A., Foged, N., He, X., Christiansen, A.V., Refsgaard, J.C., Auken, E. & Bauer-Gottwein, P., 2015. Performance evaluation of groundwater model hydrostratigraphy from airborne electromagnetic data and lithological borehole logs, *Hydrol. Earth Syst. Sci.,* **19,** 3875–3890.

Marker, P.A., Vilhelmsen, T., Foged, N., Wernberg, T., Auken, E. & Bauer-Gottwein, P., 2017. Probabilistic predictions using a groundwater model informed with airborne EM data, *Adv. Water Resour.,* **103,** 86–98.

Martin, J., Wilcox, L.C., Burstedde, C. & Ghattas, O., 2012. A stochastic Newton MCMC method for large-scale statistical inverse problems with application to seismic inversion, *SIAM J. Sci. Comput.,* **34,** A1460–A1487.

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

Mikucki, J.A. *et al.*, 2015. Deep groundwater and potential subsurface habitats beneath an Antarctic dry valley, *Nat. Commun,* **6** 6831, doi:10.1038/ncomms7831.

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

Minsley, B.J. *et al.*, 2012a. Airborne electromagnetic imaging of discontinuous permafrost, *Geophys. Res. Lett.,* **39,** L02503, doi:10.1029/2011GL050079.

Minsley, B.J., Smith, B.D., Hammack, R., Sams, J.I. & Veloski, G., 2012b. Calibration and filtering strategies for frequency domain electromagnetic data, *J. Appl. Geophys.,* **80,** 56–66.

Minsley, B.J., Kass, M.A., Hodges, G. & Smith, B.D., 2014. Multielevation calibration of frequency-domain electromagnetic data, *Geophysics,* **79,** E201–E216.

Neuman, S.P., 2003. Maximum likelihood Bayesian averaging of uncertain model predictions, *Stoch. Environ. Res. Risk Assess.,* **17,** 291–305.

Oliphant, T.E., 2007. Python for scientific computing, *Comput. Sci. Eng.,* **9,** 10–20.

Palacky, G.J., 1987. Resistivity characteristics of geologic targets, in *Electromagnetic Methods in Applied Geophysics: Voume 1, Theory,* pp. 53–129, ed. Nabighian, M.N., Society of Exploration Geophysicists.

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

Ray, A., Kaplan, S., Washbourne, J. & Albertin, U., 2017. Low frequency full waveform seismic inversion within a tree based Bayesian framework, *Geophys. J. Int.,* **212,** 522–542.

Refsgaard, J.C., Christensen, S., Sonnenborg, T.O., Seifert, D., Højberg, A.L. & Troldborg, L., 2012. Review of strategies for handling geological uncertainty in groundwater flow and transport modeling, *Adv. Water Resour.,* **36,** 36–50.

Reid, J.E., Kimber, B.A. & Worby, A.P., 2006. Airborne electromagnetic footprints in 1D earths, *Geophysics,* **71,** G63–G72.

Sambridge, M., Bodin, T., Gallagher, K. & Tkalčić, H., 2013. Transdimensional inference in the geosciences, *Phil. Trans. R. Soc. A,* **371,** 20110547, doi:10.1098/rsta.2011.0547.

Scales, J.A. & Snieder, R., 1997. To Bayes or not to Bayes? *Geophysics,* **62,** 1045–1046.

Scheidt, C., Li, L. & Caers, J.(eds), 2018. Decision making under uncertainty, in *Quantifying Uncertainty in Subsurface Systems,* pp. 29–43, American Geophysical Union (AGU).

Siemon, B., Christiansen, A.V. & Auken, E., 2009. A review of helicopter-borne electromagnetic methods for groundwater exploration, *Surf. Geophys.,* **7,** 629–646.

Spies, B.R., 1989. Depth of investigation in electromagnetic sounding methods, *Geophysics,* **54,** 872–888.

Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation,* SIAM.

Tarantola, A., 2006. Popper, Bayes and the inverse problem, *Nat. Phys.,* **2,** 492–494.

The HDF Group, 1997. 'Hierarchical Data Format, version 5'. Available at: http://www.hdfgroup.org/HDF5/, accessed date August 2020.

Troldborg, L., Refsgaard, J.C., Jensen, K.H. & Engesgaard, P., 2007. The importance of alternative conceptual models for simulation of concentrations in a multi-aquifer system, *Hydrogeol. J.,* **15,** 843–860.

Van der Walt, S., Colbert, S.C. & Varoquaux, G., 2011. The numpy array: a structure for efficient numerical computation, *Comput. Sci. Eng.,* **13,** 22–30.

Verstraeten, I.M., Steele, G.V., Cannia, J.C., Hitch, D.E., Scripter, K.G., Böhlke, J.K., Kraemer, T.F. & Stanton, J.S., 2001. Interaction of surface water and ground water in the Dutch Flats area, western Nebraska, 1995–99 (Report No. 2001–4070). Water-Resources Investigations Report, doi:10.3133/wri014070.

Viezzoli, A., Christiansen, A.V., Auken, E. & Sørensen, K., 2008. Quasi-3D modeling of airborne TEM data by spatially constrained inversion, *Geophysics,* **73,** F105–F113.

Viezzoli, A., Tosi, L., Teatini, P. & Silvestri, S., 2010. Surface water-groundwater exchange in transitional coastal environments by airborne electromagnetics: the Venice Lagoon example, *Geophys. Res. Lett.,* **37,** L01402, doi:10.1029/2009GL041572.

Vilhelmsen, T.N., Auken, E., Christiansen, A.V., Barfod, A.S., Marker, P.A. & Bauer-Gottwein, P., 2019. Combining clustering methods with MPS to estimate structural uncertainty for hydrological models, *Front. Earth Sci.,* **7,** doi:10.3389/feart.2019.00181.

Visser, G. & Markov, J., 2019. Cover thickness uncertainty mapping using Bayesian estimate fusion: leveraging domain knowledge, *Geophys. J. Int.,* **219,** 1474–1490.

Voglis, C. & Lagaris, I.E., 2004, A rectangular trust region dogleg approach for unconstrained and bound constrained nonlinear optimization, *presented at the WSEAS International Conference on Applied Mathematics,* Corfu, Greece.

Walter Anthony, K. *et al.*, 2018. 21st-century modeled permafrost carbon emissions accelerated by abrupt thaw beneath lakes, *Nat. Commun.,* **9,** 3262, doi:10.1038/s41467-018-05738-9.