An information theoretic Bayesian uncertainty analysis of AEM systems over Menindee Lake, Australia

Anandaroop Ray [ORCID], Yusen Ley-Cooper, Ross C. Brodie, Richard Taylor,[*] Neil Symington and Negin F. Moghaddam[*]

Geoscience Australia, Symonston, Symonston ACT 2609, Australia. E-mail: anandaroop.ray@ga.gov.au

SUMMARY

Long-range, active-source airborne electromagnetic (AEM) systems for near-surface conductivity imaging fall into two categories: helicopter (rotary-wing) borne or fixed-wing aircraft borne. A multitude of factors such as flying height, transmitter loop area and current, source waveforms, aerodynamic stability and data stacking times contribute to the geological resolvability of the subsurface. A comprehensive comparison of the relative merits of each system considering all such factors is difficult, but test flights over well-constrained subsurface geology with downhole induction logs are extremely useful for resolution studies. However, given the non-linear nature of the electromagnetic inverse problem, handling transmitter–receiver geometries in fixed-wing aircraft is especially challenging. As a consequence of this non-linearity, inspecting the closeness of downhole conductivities to deterministic inversion results is not sufficient for studying resolvability. A more comprehensive picture is provided by examining the variation in probability mass of the depth-wise Bayesian posterior conductivity distributions for each kind of AEM system within an information theoretic framework. For this purpose, probabilistic inversions of data must be carried out. Each acquiring system should fly over the same geology, survey noise levels must be measured and the same prior probabilities on conductivity must be used. With both synthetic models as well as real data from over the Menindee calibration range in New South Wales, Australia, we shed new light on the matter of AEM inverse model uncertainty. We do this using two information theoretic attributes derived from different Kullback–Leibler divergences—Bayesian information gain, and a strictly proper scoring rule, to assess posterior probabilities estimated by a novel Bayesian inversion scheme. The inversion marginalizes fixed-wing geometry attributes as generic nuisance parameters during Markov chain sampling. This is the first time-domain AEM study we know of, that compares nuisance marginalized subsurface posterior conductivities from a fixed-wing system, with rotary-wing derived posterior conductivities. We also compare field results with induction log data where available. Finally, we estimate the information gain in each case via a covariate shift adaptation technique that has not been used before in geophysical work. Our findings have useful implications in AEM system selection, as well as in the design of better deterministic AEM inversion algorithms.

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

1 INTRODUCTION

Airborne electromagnetic (AEM) systems have been in operation since the 1940s (Palacky & West 1991), primarily as a 'bump-finding' tool when qualitatively reviewing data collected over large swathes of land. Since AEM data are acquired after pulsing the earth with an inducing, transmitted electromagnetic field, in accordance with the principles of electromagnetic induction, the earth response is 'anomalous' in the presence of electrically conducting bodies. Such anomalies have often been associated with base-metal ores or saline groundwater accumulations within resistive earth. However, qualitative approaches are unsatisfactory when trying to quantify the geoelectric structure of the earth, and such a quantification opens the doors to a far wider variety of applications (see Brodie 2010, for a historical review). To mention only a few, such applications include geotechnical investigations (e.g. Hodges 1999), mapping subtle subsurface conductivity contrasts (e.g. Worrall *et al.* 2001),

---

[*]Formerly at: Geoscience Australia, Symonston ACT 2609, Australia.

palaeovalley mapping (e.g. Eberle & Siemon 2006), hydrogeological investigations (e.g. Auken *et al.* 2017), national scale surveys for mapping subsurface architecture (e.g. Ley-Cooper *et al.* 2020) and mapping detailed river valley aquifer systems (e.g. Minsley *et al.* 2021b).

To estimate the subsurface geoelectric structure responsible for the recorded earth response, we need to convert the data from a time or frequency domain response to subsurface conductivity using Maxwell's equations and inversion theory (e.g. Parker 1994; Menke 2012). For meaningful interpretation of the inverted electrical conductivity models and their spatial variation in terms of buried geology, we require knowledge of model uncertainties. In principle, these uncertainties can be found by propagating the data uncertainties through the inversion process all the way to the model estimates. However, the electromagnetic inversion process is non-linear and unstable. Consequently, workhorse approaches such as Occam's inversion linearize the problem and iteratively produce the smoothest possible conductivity model compatible with the data noise (Constable *et al.* 1987). There are various similar deterministic inversion approaches which also produce one single estimate of the subsurface conductivity. They follow other regularization strategies and can also incorporate prior information to produce acceptable conductivity models which fit the geophysical data (e.g. Farquharson & Oldenburg 1998; Auken & Christiansen 2004). These methods can produce linearized estimates of uncertainty, which depend strongly on the assumptions made in the regularization, as well as the start model for inversion. With few exceptions (e.g. Kalscheuer *et al.* 2010), non linear uncertainties are difficult to obtain deterministically and are usually found with Bayesian approaches (e.g. Tarantola & Valette 1982; Mosegaard & Tarantola 1995; Tarantola 2005) requiring an explicit or implicit declaration of prior earth structure, followed by posterior sampling using Markov chain Monte Carlo (McMC, e.g. Minsley *et al.* 2002; Minsley 2011; Blatter *et al.* 2018; Ray & Myer 2019). An additional challenge for AEM inversion is that the acquisition system geometries may in some cases not be well known—leading to inaccuracies in the forward modelling of geophysical data from estimated conductivity models. As detailed by Brodie (2010), ambiguity in the transmitter or receiver position as well as their roll, pitch and yaw, will creep into the estimates of subsurface conductivity if they are not accounted for in the inversion process. This is especially true of fixed-wing AEM systems where the transmitter (Tx) is usually centred on the aircraft itself, and the receiver (Rx) 'bird' is towed some distance behind the aircraft. While we are not interested in transmitter or receiver rotations and positions in themselves, except as a quality check for the data and for safety purposes, they can be treated as nuisances during deterministic inversion as described by Brodie (2010). However, these geometry parameters trade-off amongst themselves, as well as with the subsurface conductivity since they all contribute to changes in the modelled electromagnetic field. These trade-offs make it even more difficult to correctly quantify the uncertainty associated with the electrical conductivity profiles of the subsurface.

In this study, through nuisance marginalization we obtain the posterior inverse model uncertainties of a technically mature, time-domain fixed-wing system. Using information theoretic principles based on the Kullback–Leibler divergence (henceforth abbreviated as KLD, Kullback & Leibler 1951), we rigorously compare the inverse uncertainty of the fixed-wing system with that of a technically mature, time-domain helicopter system. The KLD naturally leads to the formulation of the logarithmic score, (a strictly proper scoring rule, see Good 1952; Gneiting & Raftery 2007) as well as Bayesian

information gain (Lindley 1956; Chaloner & Verdinelli 1995; Ryan 2003; Valentine & Sambridge 2020). For the information gain computation, we use a covariate shift adaptation technique (Sugiyama *et al.* 2008b, a) that directly computes probability density ratios from posterior and prior samples. We believe that this has not been used before in near surface geophysics, and shows promise in other fields such as geostatistical learning and online learning from time-series (e.g. Hoffimann *et al.* 2021; Chen *et al.* 2021). Although there have been uncertainty analyses of AEM data using Bayesian methods (e.g. Minsley 2011; Hawkins *et al.* 2018; Blatter *et al.* 2018; Minsley *et al.* 2021a) and deterministic spatial resolution investigations (e.g. Bedrosian *et al.* 2016), we are not aware of studies that have carried out fixed-wing geometry nuisance marginalization, or compared the resulting subsurface uncertainties with those of a low flying helicopter system, while including deterministic inversions in the analysis.

While it is generally desirable to fly low and slow to achieve high signal-to-noise ratios, there are challenges to doing so which range from the technical to the practical. Pastoral activities and farm animals are disturbed by low flying aircraft, there are safety and pilot fatigue issues with flying low for extended periods of time, and for rotary-wing surveys there are logistical difficulties in ferrying and storing fuel drums within isolated acquisition areas (such as much of regional Australia). Through synthetic examples as well as with real AEM data collected over the same stretch of land, we show that the subsurface information content and inferred geological model interpretability of both fixed-wing and helicopter AEM data compare favourably with each other as well as upscaled induction log data.

2 FIXED WING AND HELICOPTER SYSTEMS

The details of casting the inverse problem to solve for system geometries in time domain systems can be found in chapters 2 and 5 of Brodie (2010). The most important geometry elements for a fixed-wing system were found to be the receiver pitch, as well as the Tx–Rx inline (horizontal) and vertical separations (Fig. 1). It was also ascertained that sufficient information for geometry estimation is not found within the AEM provider supplied secondary field alone. The removed nominal primary field must be added back into the secondary field, and the total field must be inverted for both geometry parameters as well as the earth conductivity model. In this work, instead of parametrizing the vertical Tx–Rx separation, we will focus on the height of the Rx bird (zRx), as we assume that the height of the Tx (i.e. aircraft height) is well known. For helicopter systems (Fig. 1) we will assume that the Tx loop frame is rigid and the Rx height relative to it is known and fixed, as is the case for many commercial systems. These assumptions are based on historical data stemming from our acquisition of hundreds of thousands of line-km using the systems described above.

A recent development for fixed-wing systems, is that we treat as data the amplitude of the joint *X*- and *Z*-component magnetic fields, that is

$$B_{\mathrm{amp}} = \sqrt{B_x^2 + B_z^2}, \tag{1}$$

where B_{amp} is a scalar. However, we note that B_{amp} is observed as a time-series at various 'channels' so both B_{amp} and B_x, B_z will be considered as vectors henceforth, in the linear algebra sense. For a layered Earth excited by a vertical magnetic dipole (i.e. a horizontal current loop transmitter), there are no azimuthal components to the
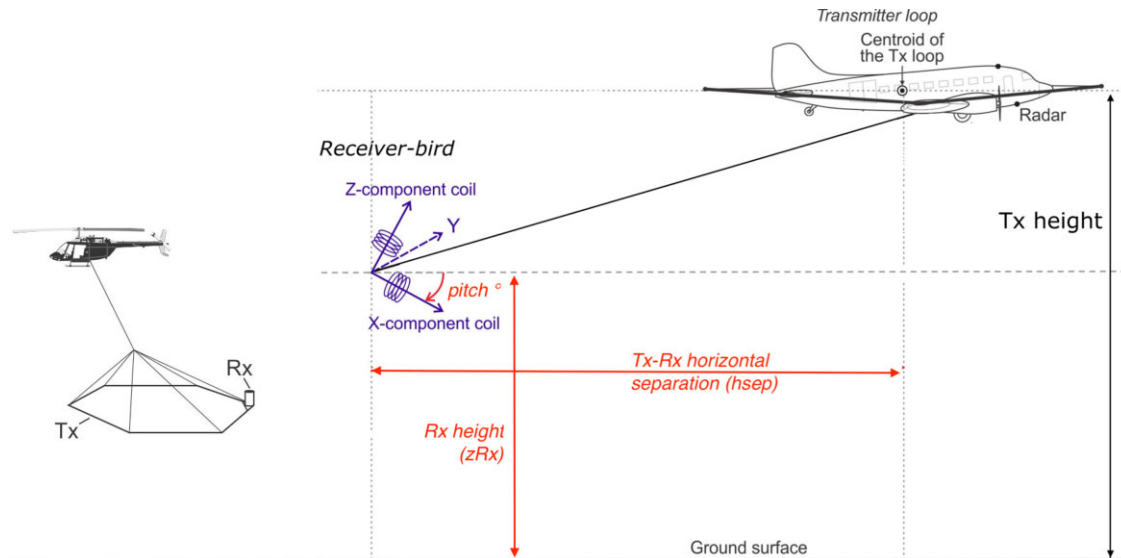
Figure 1. Helicopter (left-hand panel) and fixed-wing (right-hand panel) AEM systems. Helicopters typically fly lower and slower, and if the transmitter airframe is rigid, typically do not require geometry nuisance inversion. Fixed wing aircraft fly higher and faster, and the geometry nuisance parameters of importance have been marked in red.
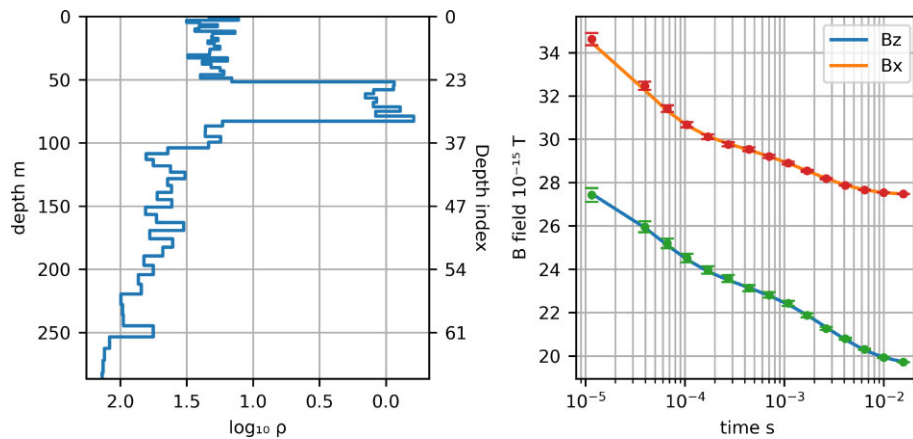

Figure 2. Noisy synthetics with nominal survey specifications and noise levels for a fixed-wing time-domain AEM system. The data correspond to the *X*- and *Z*-component magnetic fields recorded at each time channel. The synthetic model comprises 65 layers and is based on induction logs from an aquifer system in the Permian basin, TX, USA.

observed magnetic field (e.g. Loseth & Ursin 2007). As a consequence, many AEM providers do not usually provide *Y*-component data. The advantage of dealing with $\mathbf{B}_{amp}$ as opposed to $\mathbf{B}_x$ and $\mathbf{B}_z$ jointly, is that irrespective of the rotation of the *X* and *Z* coils in the *X*–*Z* plane (i.e. regardless of receiver pitch), the amplitude of the joint vector field remains invariant. This obviates the need to invert for the Rx pitch, reducing the number of unknowns in the nuisance estimation. Of course, this comes at the expense of subsurface information which the individual *X*- and *Z*-components provide in a conventional joint inversion. However, as we will see, there is not an appreciable difference with the resolving capabilities of the joint inversion, and the amplitude only inversions potentially remove troublesome conductivity artefacts at depth (Ley-Cooper & Brodie 2020). Using the theory of propagation of errors, assuming independence of the data errors in the *X*- and *Z*-components, the data error in $\mathbf{B}_{amp}$ at each time channel can be derived from (1) as:

$$\sigma_{B_{amp}} = \frac{1}{B_{amp}} \sqrt{B_x^2 \sigma_{B_x}^2 + B_z^2 \sigma_{B_z}^2}, \tag{2}$$

where σ_{B_x} and σ_{B_z} are the data errors in the *X*- and *Z*-components, respectively.

3 HIERARCHICAL BAYESIAN SAMPLING OF AEM NUISANCES AND EARTH PROPERTIES

We now return to the matter of trade-offs between system geometry and earth conductivity. Using Bayesian inference in a hierarchical setting (e.g. Gelman 2006), we can estimate distributions over parameters we are not interested in, to ensure that inferences over parameters of interest are unbiased. Nuisance estimation in this manner has a long and rich history in the Bayesian geophysical literature: for traveltime inversion to estimate data noise (Malinverno & Briggs 2004), in geoacoustics to estimate source waveforms and data noise (Mecklenbräuker & Gerstoft 2000; Dettmer *et al.* 2010), with receiver functions to parametrize the likelihood (Bodin *et al.* 2012), with magnetotellurics to estimate water column conductivities (Blatter *et al.* 2018), to illustrate a few examples. For AEM
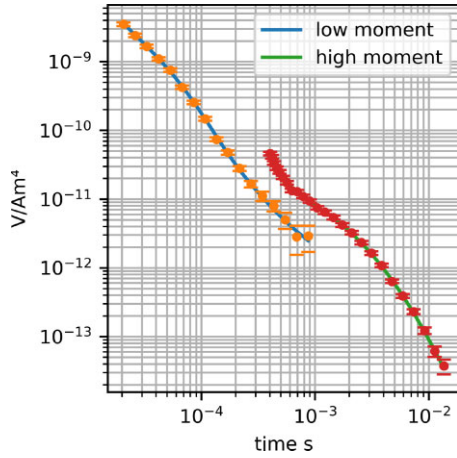
Figure 3. Noisy synthetics with nominal survey specifications and noise levels for a dual moment, time-domain helicopter system, for the same aquifer model as shown in Fig. 2. The data correspond to voltage in the *Z* receiver coil to measure changes in the magnetic field due to two independent exciting transients.

applications, Minsley (2011) has considered reported transmitter heights as uncertain and sampled them as a nuisance. However, ours is the first application we are aware of, which treats the rotations of the transmitter and receiver, as well as the transmitter–receiver horizontal and vertical separations, as nuisances to be sampled for a fixed-wing system. Possible reasons such literature is lacking are that handling passive and active frames of transmitter and receiver rotation is non-trivial (e.g. Fitterman & Yin 2004; Key & Lockwood 2010), modifying inline separations requires careful changes to pre-allocated Hankel filters for forward evaluations (see Key 2012, for details), and that a generic McMC inversion code which can handle an extensible number of nuisances is not easily available. As part of this work, we have extended the generic trans-dimensional Gaussian process (TDGP) McMC package (Ray & Myer 2019; Ray 2021) to handle AEM forward models based on the time domain AEM codes of Blatter *et al.* (2018) and the rotation formulations given in chapter 2 of Brodie (2010). As these nuisances and earth model parameters (EM conductivity) are sampled jointly by the McMC, all interparameter dependencies including non-linearities can be captured. By marginalizing over the posterior nuisance parameters, the posterior conductivities are freed from bias. Within the permissible prior ranges of the geometry nuisances, corresponding earth conductivities within their prior ranges are extensively sampled according to their likelihood as determined by corresponding data misfits. All of this is done through the McMC in accordance with Bayes' Theorem as detailed below.

3.1 Posterior sampling and parametrization

For observed data **d** and models θ_m it can be written that:

$$p(\theta_m|\mathbf{d}) \propto p(\mathbf{d}|\theta_m)p(\theta_m). \tag{3}$$

Reading from right to left, $p(\theta_m)$ is the prior probability of θ_m, which is known independent of the observations **d**. The prior importance of θ_m is re-weighted by carrying out an AEM sounding which shows how likely it is that θ_m fits the observations. The probability of fit is provided by the likelihood function $p(\mathbf{d}|\theta_m)$. The result of re-assessing or updating the prior importance of θ_m by the likelihood of θ_m provides the *posterior* probability of observing the model θ_m. This posterior probability is represented by the term $p(\theta_m|\mathbf{d})$.

The likelihood function $p(\mathbf{d}|\theta_m)$ for Gaussian data noise can be written as:

$$\mathcal{L}(\theta_m) = p(\mathbf{d}|\theta_m) = \frac{1}{\sqrt{|2\pi\mathbf{C}_d|}} \exp\left(-\frac{1}{2}\Big[\mathbf{f}(\theta_m) - \mathbf{d}\Big]^t \mathbf{C}_d^{-1}\Big[\mathbf{f}(\theta_m) - \mathbf{d}\Big]\right), \tag{4}$$

where $[\mathbf{f}(\theta_m) - \mathbf{d}]$ is the residual vector of misfit for the model θ_m, between the forward model calculations for a given set of earth conductivities as well as nuisances, and the AEM data. The covariance matrix of data errors is given by $\mathbf{C}_d$. A Gaussian likelihood is generally justified by the application of stacking to increase signal-to-noise ratios in geophysics, and AEM is no exception. Stacking implies Central Limiting for the resulting noise estimates on the mean data, and hence the implication of Gaussianity for the likelihood function. The data error or noise model used throughout this work is based on the noise model given in Green & Lane (2003). The total noise is assumed to be due to two independent sources: multiplicative noise proportional to signal amplitude, as well as high-altitude noise measured away from the effects of Earth conductivity, added together in quadrature. In AEM parlance, the misfit measure ϕ_d is often used, with values close to 1 indicating a reasonable fit to within data noise. The more commonly used measure of RMS misfit in the broader geophysical EM literature is given by $\sqrt{\phi_d}$, where the likelihood and ϕ_d are related as:

$$p(\mathbf{d}|\theta_m) = \frac{1}{\sqrt{|2\pi\mathbf{C}_d|}} \exp\left(-\frac{n_{\text{data}}}{2}\phi_d(\theta_m)\right), \tag{5}$$

$$\Rightarrow \phi_d(\theta_m) = \frac{1}{n_{\text{data}}}\Big[\mathbf{f}(\theta_m) - \mathbf{d}\Big]^t \mathbf{C}_d^{-1}\Big[\mathbf{f}(\theta_m) - \mathbf{d}\Big], \tag{6}$$

or in other words, the χ^2 data error is ϕ_d times the number of data n_{data}.

To be explicit about the model parametrization, we separate θ_m into the earth resistivity θ_ρ and the geometry nuisance parts θ_n as follows:

$$\theta_m = [\theta_\rho, \theta_n]. \tag{7}$$

Resistivity is the inverse of electromagnetic conductivity, and to span the various orders of magnitude of earth resistivity we parametrize θ_ρ through the base-10 logarithm of linear resistivities. For sampling θ_ρ, we use a reversible jump sampler (Green 1995) or trans-dimensional (trans-D) McMC as it is often referred to in geophysics (Malinverno 2002; Sambridge *et al.* 2006; Bodin & Sambridge 2009; Dettmer *et al.* 2010; Ray & Key 2012; Gehrmann *et al.* 2015; Ghalenoei *et al.* 2021). However, instead of using a piecewise constant stair-step parametrization as is usually done for 1-D, we use a Gaussian process basis (GP) as described in (Ray & Myer 2019). In addition to TDGP being spatial-dimension agnostic (e.g. Blatter *et al.* 2021), we have found it to be particularly well suited to diffusion problems which require smooth parametrizations. Sharp changes can be well represented with two GP inference layers as detailed in Ray (2021). However, we have found that AEM data do not generally support such sharp changes in the Earth and we have opted to use the vanilla TDGP, that is trans-D McMC with birth and death of stationary GP nuclei (where stationarity implies an unchanging GP length scale over the model domain) in a reversible jump framework as described in Ray & Myer (2019). Specification of the prior for θ_ρ is exactly the same as shown in section 2.5.2 of Ray (2021) for a resistivity model and will not be repeated here. Priors for θ_n are uncorrelated and uniform, $\pm$ a few metres or degrees off from what is recorded during the AEM survey. Particulars of prior probabilities are provided in the applications section.
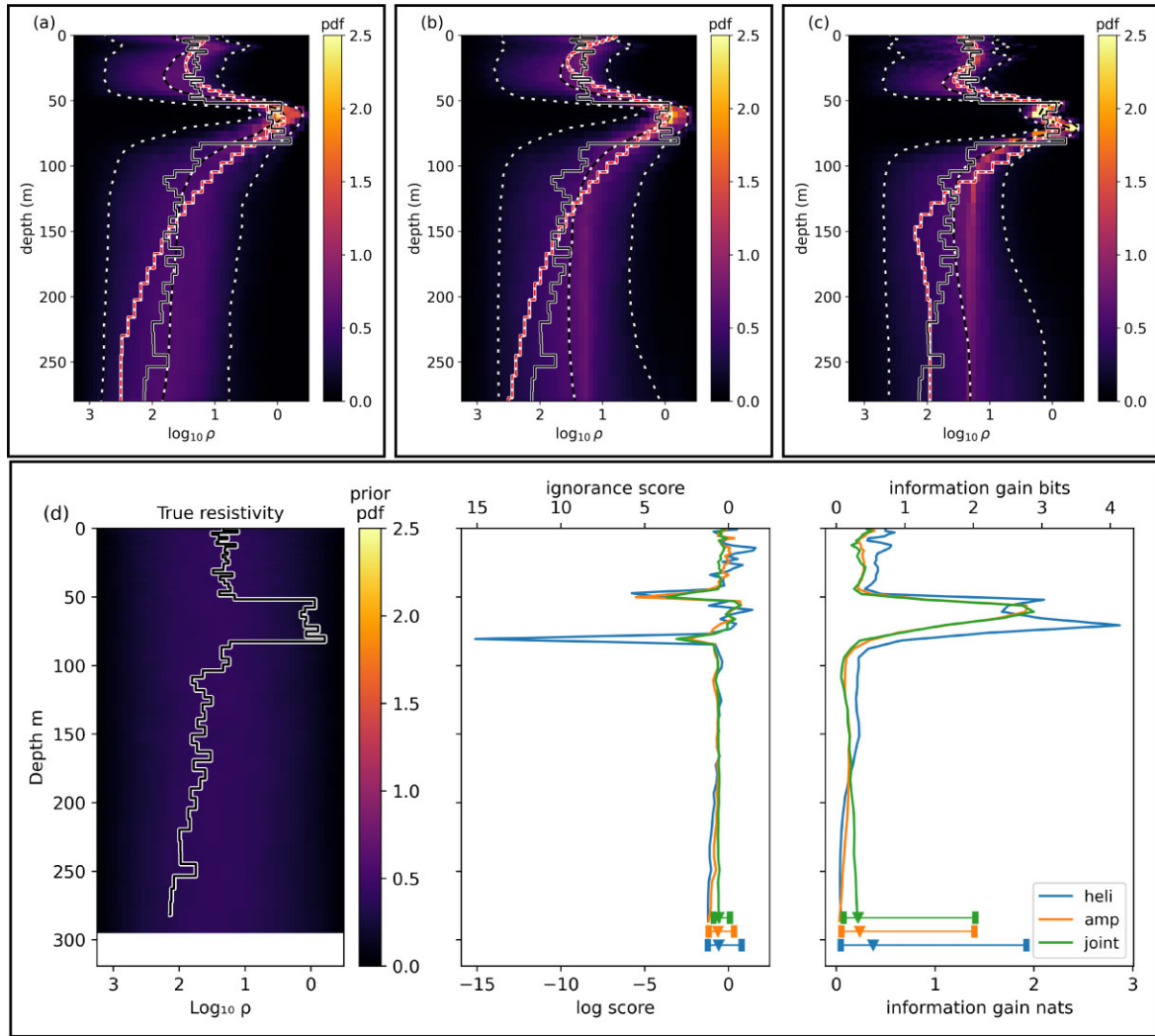
Figure 4. Synthetic sounding conductivity posteriors with depth for (a) Fixed wing: conventional B_x and B_z joint inversion, (b) Fixed-wing: inversion of **B** field amplitude in receiver X–Z plane and (c) Helicopter: dB_z/dt inversion. Hotter colours are more probable, and the true model has been overlain with a thick black line. The 5th, 50th and 95th percentiles of posterior conductivity are shown with dashed black-and-white lines. The deterministic inversion result has been plotted with a thick red-and-white line. (d) Information theoretic divergence analysis: True model with prior probabilities in the background (left-hand panel), logarithmic score for each AEM system/inversion and probabilistic inversion (centre panel) and Bayesian information gain (right-hand panel). High logarithmic scores and high information gain indicate better adherence to the truth and posterior resolution, respectively. For each inversion type, the summary plots at the bottom indicate 5th, 50th and 95th percentiles from scores and gains across all depths, with the filled triangle indicating the median.

The process of finding the posterior probability $p(\boldsymbol{\theta_m}|\mathbf{d})$ for various models $\boldsymbol{\theta_m}$ admissible by the prior is repeated until an ensemble of models representative of the posterior probability density function (PDF) $p(\boldsymbol{\theta_m}|\mathbf{d})$ is obtained. Sampling proportional to the posterior probability is carried out by using the following acceptance probability α to move from an Earth resistivity model vector or nuisance model vector $\boldsymbol{\theta}$ to proposed model $\boldsymbol{\theta}'$ in the McMC chain:

$$\alpha(\boldsymbol{\theta}'|\boldsymbol{\theta}) = \min\left[1, \left\{\frac{\mathcal{L}(\boldsymbol{\theta}')}{\mathcal{L}(\boldsymbol{\theta})}\right\}^{1/T}\right]. \tag{8}$$

We note here, that for a uniform prior over the number of nuclei and when proposing from the prior resistivities for birth and death, for all TDGP moves, (8) provides the acceptance probability. All the TDGP move proposal probabilities $q(\boldsymbol{\theta}, \boldsymbol{\theta}')$ are exactly the same as described in detail in Ray (2021). For symmetric fixed-dimensional

proposals, when proposing from the prior, and for uniform priors over the number of parameters, eq. (8) also holds for both fixed-dimensional or reversible jump Metropolis-Hastings–Green McMC (Metropolis *et al.* 1953; Hastings 1970; Geyer 2011). This is what we have used for sampling geometry nuisances and earth conductivities. Due to this choice of proposals, the move probability terms $q(.)$ never explicitly figure in calculation of the acceptance probability term α in our algorithm (e.g. Agostinetti & Malinverno 2010; Dosso *et al.* 2014). The exponent $1/T$ in (8) is an annealing factor for parallel tempering (Swendsen & Wang 1987) as described in Dettmer & Dosso (2012). Parallel tempering significantly accelerates the convergence to the posterior distribution (Dosso *et al.* 2012, 2014; Sambridge 2013) and is used by default in TDGP. The entire McMC algorithm encapsulated within a parallel tempering framework is described in Algorithm B within Appendix B.
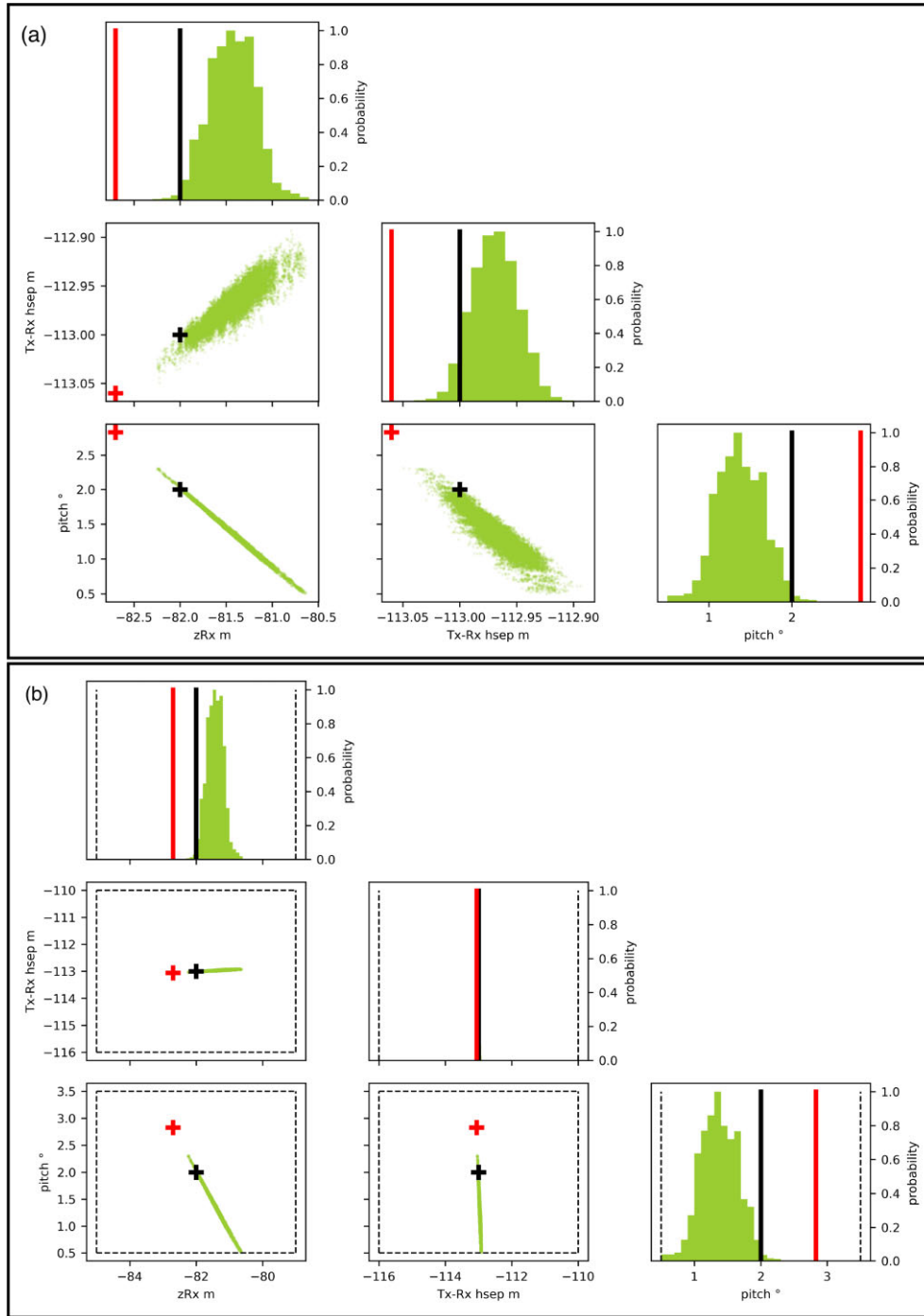
Figure 5. Nuisance geometry posterior scatter plots and histograms for fixed-wing AEM: conventional B_x and B_z joint inversion. (a) Detailed view, with red symbols indicating deterministic inversion estimates and black symbols denoting true values. (b) Zoomed out view, this time with dashed lines showing prior bounds.

4 SYNTHETIC STUDY: INFORMATION THEORY AND MARGINALIZATION OF GEOMETRY NUISANCES

While ground truth is in principle, the ultimate arbiter of the accuracy of a geophysical investigation method, a synthetic study at one sounding location, with noise levels, Tx–Rx geometries and flying heights typical of the systems under consideration is instructive. The transmitter waveforms, noise levels, nominal geometries

and heights are sourced from the Menindee calibration range AEM flights (Barlow 2019) that we will report on in a later section. As in the real data example, the fixed-wing AEM system and the helicopter AEM system compared are widely used and technically mature. Both have been found fit for purpose for various surveys we have undertaken (e.g. Ley-Cooper 2021, 2022) on the basis of competitive bidding from various commercial entities that have included, but not been limited to these two systems.
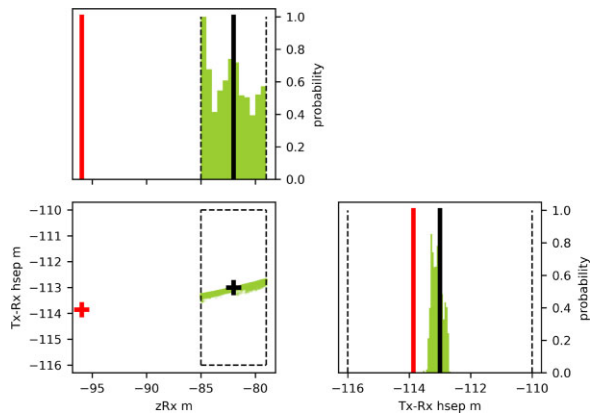
Figure 6. Nuisance geometry posterior scatter plots and histograms for fixed-wing AEM: Inversion of **B** field amplitude in receiver *X–Z* plane. Again, red and black symbols denote deterministic and true values, respectively. Prior bounds are shown with dashed lines. Compared to the joint inversion posteriors in Fig. 5, note that more of the prior space within the dashed lines has been occupied by the posterior samples.

The fixed-wing system (Fig. 1) compared was TEMPEST (Lane *et al.* 2000), with the transmitter loop placed 120 m above the ground, and the Tx–Rx vertical separation was set to 38 m, that is the receiver antennae were placed at zRx = 82 m above the ground. The Tx–Rx horizontal separation or hsep was set to 113 m, with Rx pitch set to 2°. For the helicopter AEM comparison, we used a SkyTEM 312 system based on developments of the system originally described in Sorensen & Auken (2004). For the helicopter system (Fig. 1), the transmitter loop was placed 40 m above the ground, and the receiver was rigidly fixed at a vertical separation of 2 m above the transmitter, at a radial distance of 13 m from the loop centre, inline to the flight direction.

The synthetic earth model used for the comparative study is a realistic earth model for shallow hydrological investigations, based on variations we would expect in a well-log. It features 65 layers with many small conductivity contrasts, one large conducting 'layer' between 50 and 80 m depth and a resistive trend with depth. Forward modelling and noisy data for this earth model using both systems can be seen in Figs 2 and 3. We ask readers to note that from here on, we always have conductivity increasing to the right and resistivity increasing to the left of all our resistivity axes. Further, we represent resistivity in $\log_{10}$ units, in which a conductivity in $S\,m^{-1}$ or resistivity in Ωm are simply the negative of the other, being inverses. There are undoubtedly differences in the spectral content of the source waveforms for both systems, especially given that the helicopter system considered here is a 'dual moment' system with two exciting transients that are interleaved together after acquisition. It is also impossible to examine the uncertainty of inverted subsurface conductivity structures at all possible wavenumbers at all exciting frequencies. However, the utility of this synthetic study is twofold. First, it is not a closest-to-true model contest, as posterior uncertainties around both synthetic soundings will be inspected— lower uncertainties in conductivity at depth imply greater resolution. Secondly, it sets the stage for examining posterior subsurface uncertainty for real data with hundreds of soundings along a flight line with variable geology and established ground truth. However, we must make clear that practically speaking, we never have exact truth even when we have logged the earth within a drillhole. The effects of drilling mud, invasion, calibration errors and temporal (climatic and seasonal) changes between the acquisition of downhole induction log conductivities and the above surface EM data acquisition,

make exact comparison impossible. However, the surface geophysical data should still reflect well-constrained geology and property trends measured in the drillhole. We ask the reader to keep this in mind in the context of further references to ground truth or induction logs in the text. For those wishing to investigate an additional synthetic model, another example based on a real well-log featuring a thin near-surface conductor and a deeper resistor can be found in Appendix E.

Posterior uncertainties for both systems are shown in Figs 4(a)–(c). Note how the deterministic inversions almost always lie within the 90 per cent posterior credible intervals (CIs), the region between the 5 and 95 per cent posterior percentiles. Further, while some deterministic inversion results are closer to the true value, the posterior uncertainties paint a different picture of uncertainty. The uncertainty with depth, qualitatively given by the width of the CI, is not markedly different between the amplitude only and helicopter AEM posteriors. As expected, the posterior uncertainty for all three inversion types is least around the conducting 'layer' between 50 and 80 m depth. In fact, for this model, it would appear that the fixed-wing 'conventional' inversion does well overall, with the width of the CI remaining roughly similar over all depths except at the conductor. The helicopter AEM posterior however, seems to suggest that the top and bottom of the conductor can be separately resolved. It must be noted that we avoided inverse crime: the forward modelling grid for both the deterministic and probabilistic inversions was identical, but coarser (only 50 cells) than the true 65 cell modelling grid used to compute the synthetics. This was to emulate a measure of realism, as the true discretization of the earth is never known. As is the usual case, we overparametrized the modelling grid for the inversion, and an Occam regularization with a return to a resistive reference or 'prejudice' model (see Key 2009) was used in the deterministic inversions. For the probabilistic inversions, the prior parameter resistivity bounds for all the deterministic inversions were uniform between the extremal limits of the $\log_{10}$ conductivities shown in Figs 4(a)–(c). A correlation length of 2 thickness units was used, and the birth/death trans-D GP method was allowed to place a maximum of 40 GP nuclei uniformly between the top and bottom of the model. The resulting GP nuclei resistivity values are interpolated onto the 50 cell grid which is fine enough to model accurately the geophysical features of interest in a spatial dimension-agnostic manner (e.g. Blatter *et al.* 2021). The correlation ensures that any thickness unit has an exponentially decreasing correlation with interpolated resistivities two thickness units above and below it. Although it may appear that infinitesimally thin layers are desirable to model, they are largely unresolvable when using diffusive electromagnetic physics as has been described in detail by Ray & Myer (2019) and Ray (2021). The GP kernel used to impose the correlation is the sidelobe averse Ornstein–Uhlenbeck kernel, which is described in detail by Rasmussen & Williams (2006).

4.1 Information theoretic analysis of posterior conductivities

While we can gather a useful, qualitative idea of posterior uncertainty by inspecting Figs 4(a)–(c), we have also provided a rigorous and quantitative information theoretic analysis using Kullback–Leibler divergences (Fig. 4d). While the general geophysicist can follow the remainder of this section without specialized knowledge, we refer those interested in the details to Appendix D for a mathematical discussion with proofs. With this in mind, we look at two
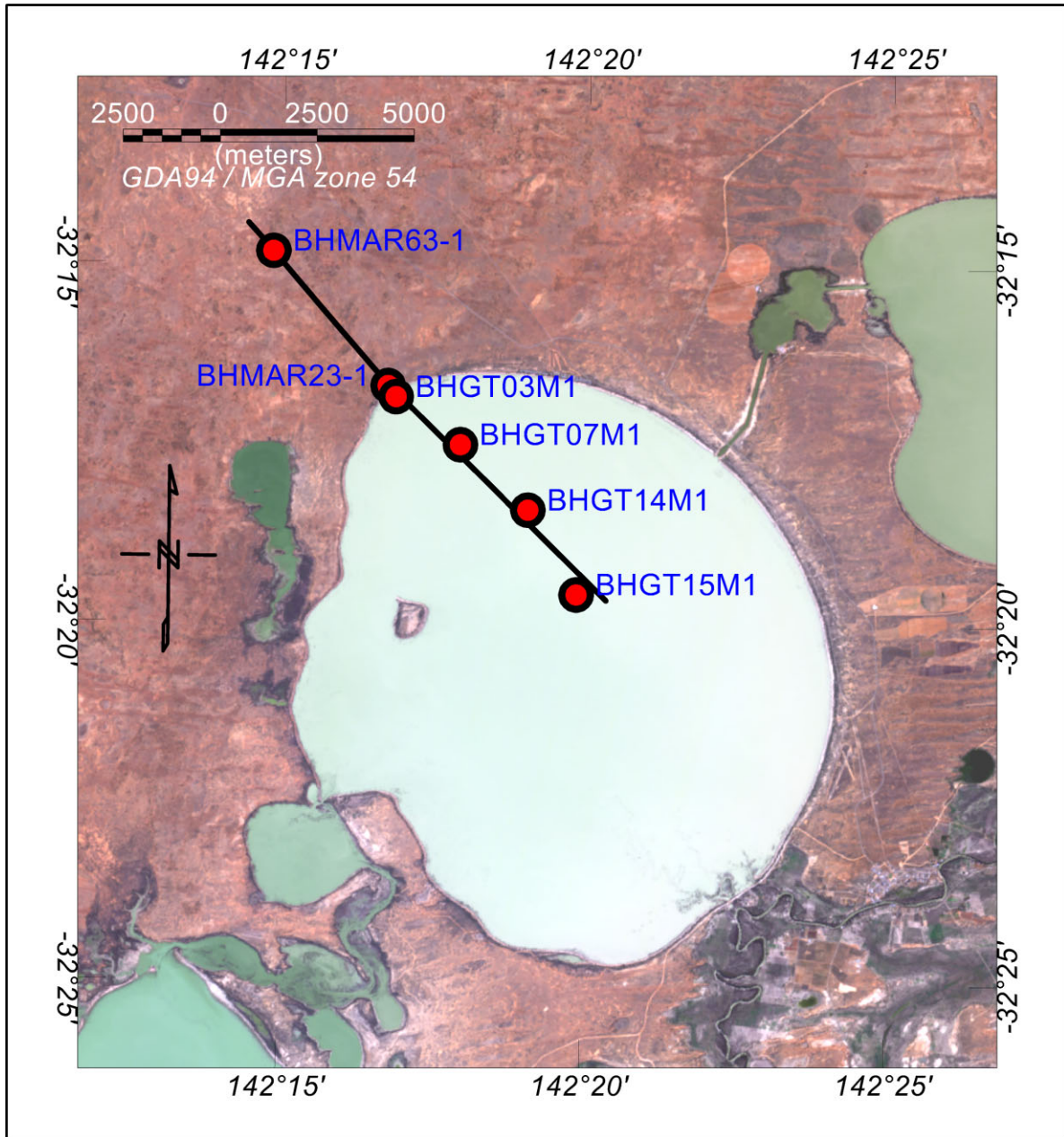
Figure 7. Test AEM line over Menindee Lakes, New South Wales, Australia, with logged boreholes shown in red. Note the transition from an arid Earth surface to the lake, midway along the test line.

divergences, as they are known, which can heuristically but not mathematically, be thought of as 'distances' between probability density functions (see Beier *et al.* 2002). The first such divergence leads to a strictly proper scoring rule (Gneiting & Raftery 2007), the logarithmic score, dating at least to Good (1952). In essence, for a forecast density $p(\cdot)$, say a posterior density, we assign a score $\log p(x)$ when the event x is actually observed. As shown with eqs (D9)–(D11), on average the highest score will be obtained for a forecast probability density that equals the true probability density of the observations, even if the true density is unknown. Naturally, such scoring systems have found heavy use in forecasting and allied decision theoretic fields such as meteorology, quantitative economics and finance, psychology and optimal energy usage (see Carvalho 2016, for a review). In near surface or exploration geophysics,

recent use of logarithmic scores can be found in Seillé & Visser (2020) for selecting an optimal likelihood function, and Friedli *et al.* (2022) for evaluating different McMC proposal schemes for a challenging high-dimensional inverse problem. This brings us to the second divergence, known as the Bayesian information gain, which represents our increased knowledge of the subsurface with a (usually) narrower posterior density than the prior density we began with. In other words, the information gain represents the diminishing overlap between the prior and posterior densities, a natural proxy for Bayesian resolution as shown by Blatter *et al.* (2018) for an Antarctic AEM survey. To examine these divergences we begin by showing the true model and prior resistivity probability density in the leftmost column of Fig. 4(d). As discussed in Ray & Myer (2019), though the McMC model parameter priors are uniform,
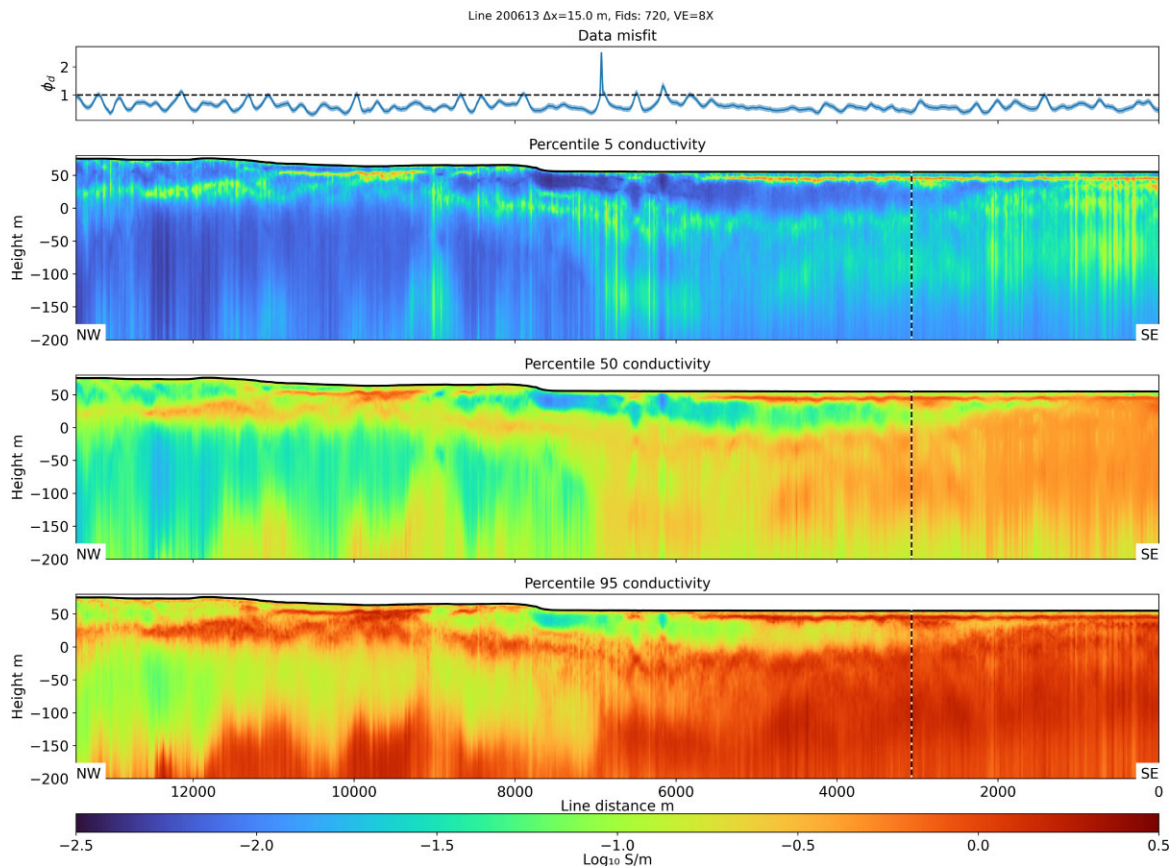
Figure 8. Menindee test range: Dual moment helicopter EM system inversion. The top row shows the mean sampled ϕ_d and one standard deviation. The next three rows show the 5th, 50th (median) and 95th percentiles of posterior conductivity. Where all three percentiles are similar, the posterior conductivities are more probable. The display of percentiles and their spread simplifies the task of assessing interpretive uncertainty, as discussed in detail in the text. The dashed vertical line is nearest to a borehole with induction logs to be examined later in the text.

the resulting resistivities interpolated by the GP parametrization are not. However, similar to a bounded uniform distribution, the interpolated prior resistivities do not have a focussed mode.

We follow Seillé & Visser (2020) and treat the posterior marginals as a forecast probability density, and the true model as the eventuating observation. We then fit a kernel density (Sheather & Jones 1991) $p_z(\cdot)$ to the marginal posterior samples at depth z and evaluate the logarithmic score $\log p_z(m_z)$ for the true $\log_{10}$ resistivity m_z at depth z. In the middle column of Fig. 4(d), we plot the logarithmic score for each system, with higher scores indicating better representation of the truth. There is however, a large amount of overlap summarized by the quantile plots at the bottom, showing the overall 5th, 50th and 95th percentiles of the score. Since there is only one score at each depth per inversion type, the summary percentiles are calculated from values across all depths. For decision theoretic problems, the forecast with the highest expected score is usually preferred (Diks *et al.* 2011). However, there is also a notable outlier score for the helicopter system which will skew the average score for that system downwards. This is because the posterior probability for all three systems at the bottom end of the true conductor is very small. However, it is not that the helicopter system does worse at localizing the conductor, there is a narrow high probability region only slightly to the left, within half a conductivity decade (1 decade = 1 log 10 unit of resistivity or conductivity) at the outlying depth of ∼80 m in Fig. 4(c). This problem with *locality* is a known issue

with the logarithmic score, and is discussed in detail by (Bröcker & Smith 2007). The negative of the logarithmic score is known as *ignorance*, and (near) zero prediction probability of the truth implies (high) infinite ignorance. This situation could easily occur as shown above, or for extremely resistive media and inductive EM methods, as these methods are not sensitive to highly resistive media. Barring the outlier, the helicopter inversion does slightly better up shallow, and the fixed-wing joint inversion scores better deeper.

In the rightmost column of Fig. 4(d) we show the Bayesian information gain, by calculating the KLD directly from the posterior and prior marginal samples at every depth. For this purpose, we use a covariate shift method (Sugiyama *et al.* 2008b), as detailed in Appendix D. Such methods are adept at quantifying a shift between samples from two probability densities, such as samples from the prior and the posterior, or when there are sudden changes in online streaming data (Chen *et al.* 2021). A recent use of sample based covariate shift methods for geostatistical transfer learning can be found in Hoffimann *et al.* (2021). The information gain focuses very minutely on the overlap (or the lack of it) between the posterior and prior probability densities. As shown in Appendix D, it is always positive and only zero if the prior and posterior densities are identical. To first order, the information gain is small when posterior widths are large, and large when posterior widths are narrow, such as within the conductor between 50 and 80 m depth. This is an information theoretic counterpart of deterministic sensitivity,
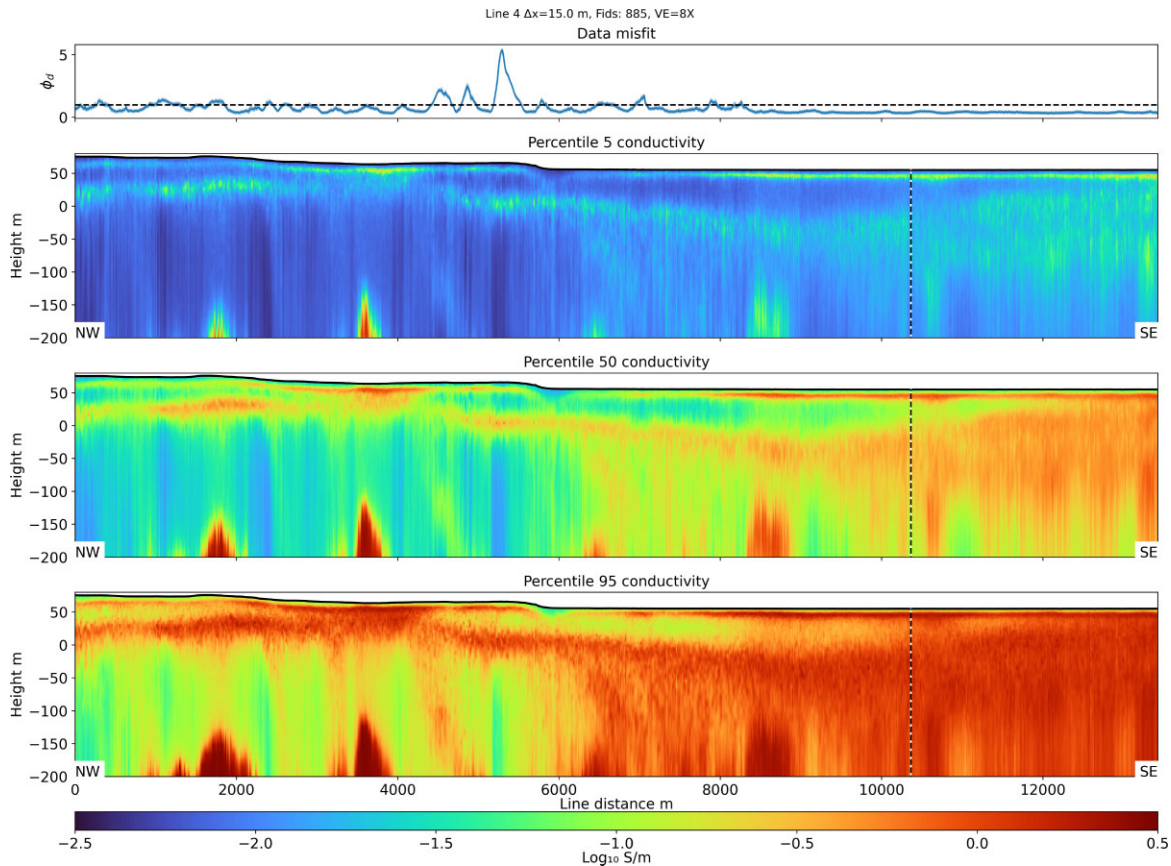
Figure 9. Menindee test range: Fixed-wing AEM system conventional joint B_x and B_z inversion. The top row shows the mean sampled ϕ_d and one standard deviation. The next three rows show the 5th, 50th (median) and 95th percentiles of posterior conductivity. Where all three percentiles are similar, the posterior conductivities are more probable. With the exception of the deep conductors at -200 m, the first-order geology of the lake system is very similar to what is shown by the helicopter EM system in Fig. 8. The dashed vertical line is nearest to a borehole with induction logs to be examined later.

as we know that TE mode inductive EM sources (Loseth & Ursin 2007) are sensitive to conductors. Examining in detail, we see that the information gain both above and below the conductor is small, but is slightly larger at shallow depths for all three inversion types. This is as we would expect, given signal-to-noise considerations at late times as well as conductive shielding effects in electric media. We would particularly like to draw attention to the fact that within the conducting body itself, unlike for the logarithmic score, the information gain remains large. This is due to the information gain not suffering from the aforementioned locality problem. In fact the helicopter system has a tight posterior distribution at ∼80 m depth within half a decade of the true value, accordingly it has the highest information gain.

In the absence of ancillary information, no inversion/system would clearly outmatch the others. In all cases, from the marginal posterior probabilities of resistivity with depth, we would interpret the following. Starting from the top: resistive geology, between 2.7 and 0.8 log$_{10}$ Ωm, followed by a conductor starting at ∼50 m depth. Owing to conductivity-thickness trade-offs the fixed-wing systems/inversions do not narrow down the conductor bottom well, while the helicopter estimate of the bottom is slightly too shallow. All three systems have high probability mass in the conductor between 50 and 80 m in the vicinity of 1 Ωm or 1 S m^{-1} (i.e. 0 in log$_{10}$). While we could point to information gain within the conductor being higher for the helicopter system—contextually speaking,

this gain from 2–3 to 4 bits is not significant. Each bit of information gain leads to reduction of half the prior probability mass (Pinkard & Waller 2022) or equivalently, doubles the concentration of probability mass in the posterior. Since 80 per cent of posterior conductivities are between 0.8 and -0.26 log$_{10}$ Ωm for the fixed-wing system inversions at ∼70 m, halving the posterior probability mass does not add significantly to inferred knowledge of a conductor. For all three inversion types, at this depth we are able to bracket a 0 log$_{10}$ Ωm conductor within less than half a decade, in a prior range spanning nearly 4 decades of resistivity. Underneath the conductor, all three inversions indicate a return to resistive geology in the 2.7 to 0.1 log$_{10}$ Ωm range, with the fixed-wing/joint inversion indicating a slightly narrower high probability region. Finally, the box plots showing the 5th, 50th and 95th percentiles of information gain for each system and their large overlap, are shown at the bottom of the rightmost panel in Fig. 4(d). Again, there is only one value of information gain at each depth for every inversion type, consequently the summary percentiles are calculated from values across all depths. As before, we see a large overlap and are able to confirm our earlier first order inspection of posterior quantiles with this analysis. This is useful to note since most geophysicists or geologists will not have computed information theoretic divergences readily available. Another useful measure of the overlap between probability densities is the Bhattacharyya distance and related coefficient (Bhattacharyya 1943), which has been used in geophysical
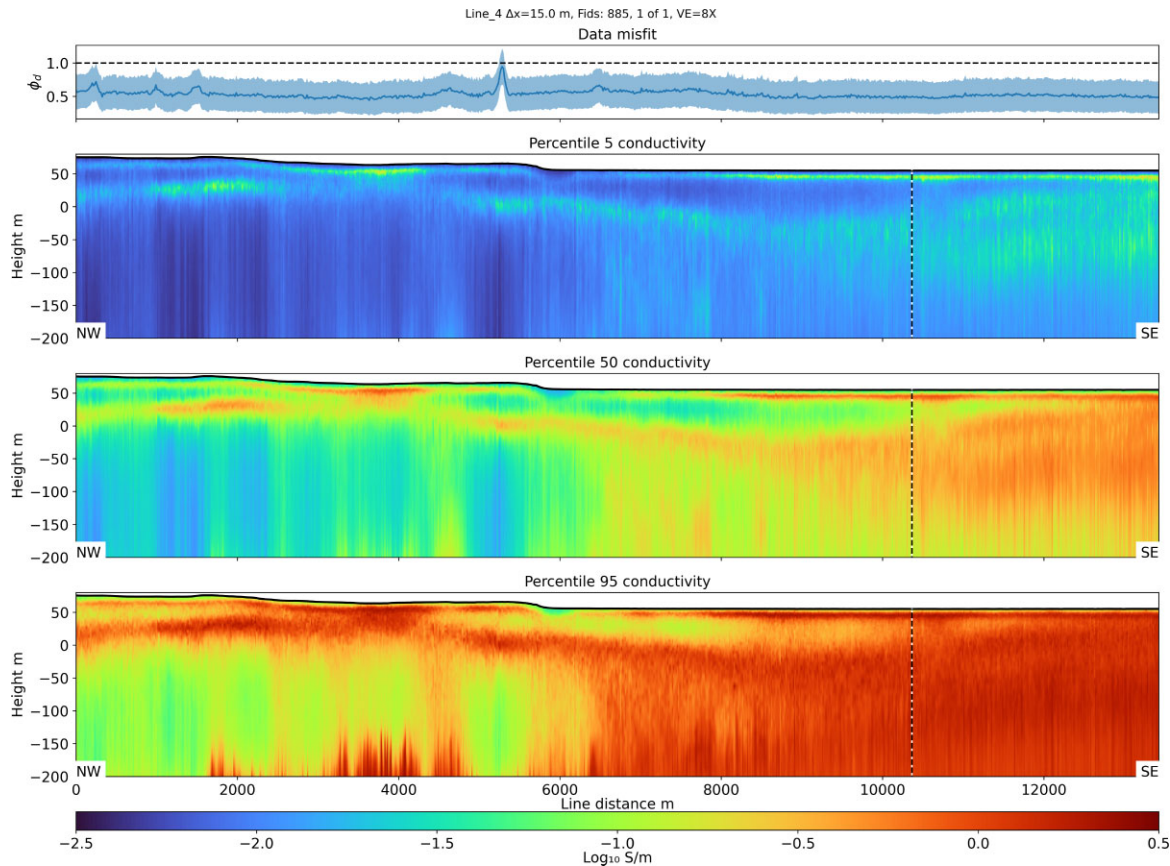
Figure 10. Menindee test range: Fixed-wing AEM system amplitude only inversion. The top row shows the mean sampled ϕ_d and one standard deviation. The next three rows show the 5th, 50th (median) and 95th percentiles of posterior conductivity. Again, with the exception of the deep conductors at -200 m in Fig. 9, the first-order geology of the lake system is very similar to what is shown by the helicopter EM system in Fig. 8. The dashed vertical line is nearest to a borehole with induction logs to be examined later.

work (Subašić *et al.* 2019), though it does not have a straightforward Bayesian interpretation.

In conclusion, if it were for this particular synthetic model, the choice of system from a technical standpoint is largely equivalent. We can say this as the CI widths are similar noting the variations discussed above, and neither the scoring rule nor information gain indicate without qualification, a superior system/inversion type.

4.2 Inversion details and nuisance sampling

All inversions converged to a rms value of 1. The probabilistic inversions for both kinds of fixed-wing inversion were run with 7 log-spaced parallel tempering chains, and for helicopter AEM inversion with 5, with a maximum annealing temperature of 2.5. A greater number of parallel chains are required for the fixed-wing inversions, as the inference problem with geometry nuisances is harder to sample. We had initially achieved stationarity well within 400 000 samples in the target McMC chain at $T = 1$. However, to sample the near-zero probability regions and establish stable score estimates that avoid infinite ignorance, we ran each inversion type for 1 000 000 samples, discarding the first fifth to preclude the possibility of biased inference. Within the legacy survey noise levels we have accumulated over the years, inverting for height has not made a significant difference to the conductivity model

and we do not do so here either. We surmise this is because the helicopter AEM system studied here, which we have observed in operation, has a rigidly mounted Rx coil. The aerodynamically stable Tx–Rx frame and its height are known well enough for this not to be inverted for. For conventional fixed-wing inversions (Fig. 4a), in addition to the earth conductivity, three more parameters need to be sampled (Fig. 5): the receiver pitch, the horizontal Tx–Rx separation (labelled Tx–Rx hsep) and the vertical Tx–Rx separation (equivalently, we have kept fixed the Tx height, and inverted the Rx height, labelled zRx). We make observations of note underneath:

In Fig. 5(a) the most probable nuisance model values and the truth do not coincide. Although other workers have encountered similar phenomena (e.g. Dettmer *et al.* 2015, most probable versus true values in their fig. 4), we decided to investigate further. We ran the Bayesian inversion for 200 000 more samples with the same noise realization as in Fig. 2, then did an independent run for 1 000 000 samples. The longer run and independent restarts with the same data as in Fig. 2 persistently produced histograms where the true nuisance values are in the tail region. Further, the posterior conductivities among these runs are virtually indistinguishable from Fig. 4(a). When compared with the prior extents (Fig. 5b), the posterior distributions do appear more generally in the vicinity of the true values. Given that the likelihood (Fig. 4) depends on the data noise—on running with a different random noise realization, we did indeed observe coincidence of true geometry values with
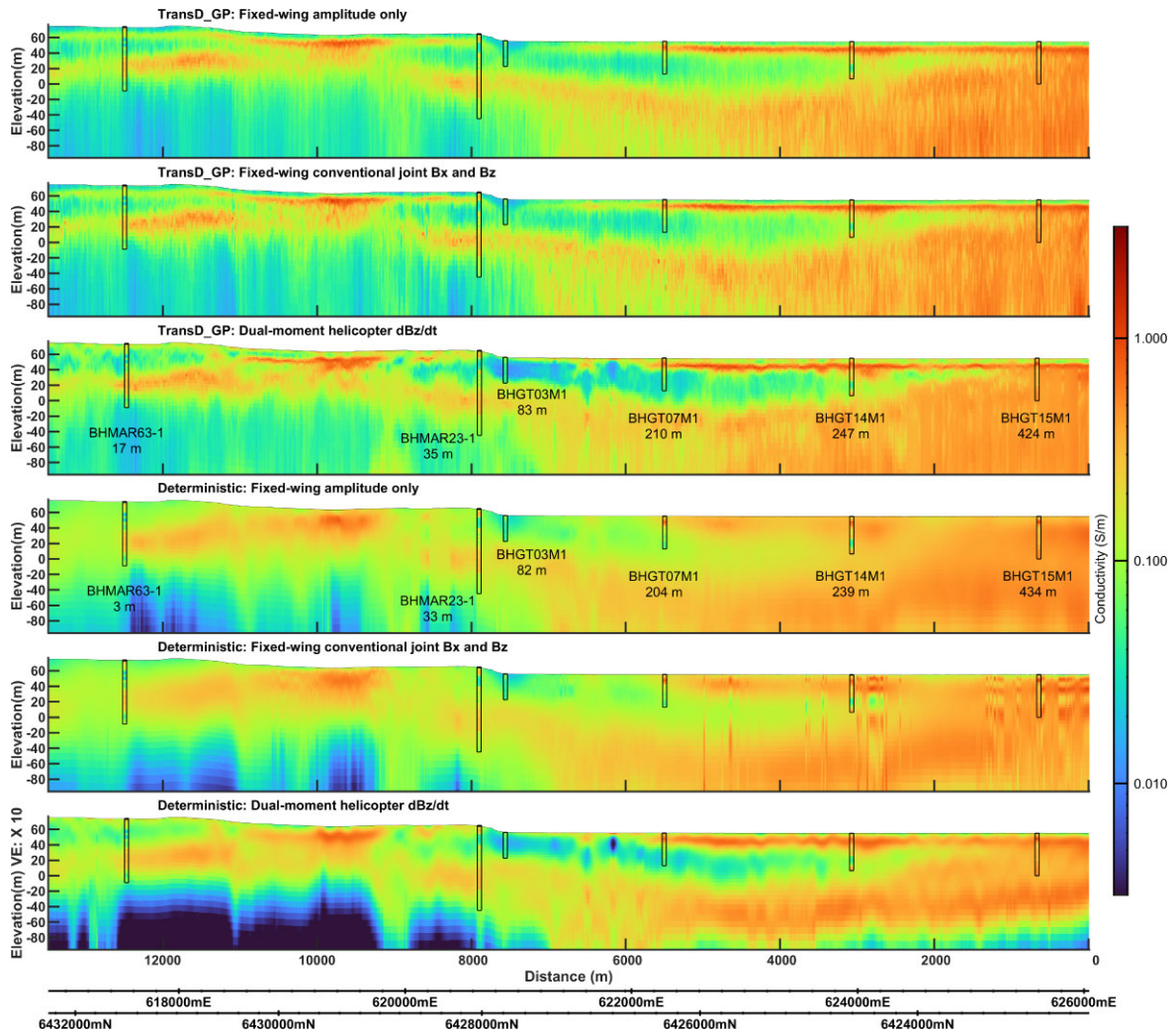
Figure 11. A comparison between probabilistic inversions (top three rows), deterministic inversions (bottom three rows) and induction logs (boxed, thin, tall rectangles on every row). The distance from the AEM flight line, together with the name of the well is shown once for fixed-wing and once for helicopter EM data. For example, the deepest logged borehole BHMAR23-1 is about 35 m away from both the helicopter and fixed-wing flights and 8 km from the southeastern end of the line. All rows show good general agreement with the wells, despite the logging and flying having occurred several years apart.

high probability regions, resulting in minor differences with the posterior conductivities presented in Fig. 4a).

We are extremely sensitive to the Tx–Rx separation, but receiver pitch and receiver height trade-off near linearly. This intuitively makes sense—if the pitch decreases (antenna axis along flight line tilts upwards) this could be compensated by the antenna origin being translated closer to the ground. However, this implies that we cannot resolve both receiver height and pitch, and perhaps the information contained in one ought to be enough for the inverse problem—with a suitable rotation of the nuisance coordinate axes. Unfortunately, this rotation of the axes is data dependent and we need to do an initial sampling run to estimate a principle components rotation. However, we have found that it is indeed much more efficient to sample along such rotated parameter axes as shown by Yardim *et al.* (2006) for radio-refractivity inversion and Dosso & Dettmer (2011) for geoacoustic inversion. An efficient alternative for sampling nuisances could be pseudomarginal methods and their correlated variant (see Andrieu & Roberts 2009; Friedli *et al.* 2022, for details).

Though the nuisance prior bounds are based on what we would expect for errors from the onboard inertial measurement unit (IMU) sensor and variability within a flightline—only a very small portion

of this prior volume is of posterior importance (i.e. a large part of the boxed region with dashed lines in Fig. 5b is empty model space). Finally, the posterior sampling surface is quite rugged (i.e. fat tails with sharp jagged dropoffs) as can be seen from the zoomed in crossplots and marginals for both pitch and zRx in Fig. 5(a). Most real data McMC AEM inversions converge to stationarity well within 200 000 samples, but in order to draw robust conclusions from synthetics we have massively oversampled as described above. Parallel and high performance computing (HPC) considerations during sampling have been laid out in Appendix C

For an amplitude only inversion (Fig. 4b) of the same data, we do not need to estimate the Rx pitch and the posterior nuisances are shown in Fig. 6. Immediately, we see that the fraction of prior volume required to solve the problem is greater and the posterior surface is less rugged, denoting that the posterior is easier to sample. While the deterministic estimate of the receiver height nuisance parameter is outside the prior range, the deterministic conductivity model, which is ultimately the earth feature of interest, is within the 90 per cent CI as can be seen from Fig. 4(b). Undoubtedly, tweaking the regularization and constraints for the deterministic nuisances can lead to 'better' estimates of conductivity. Since
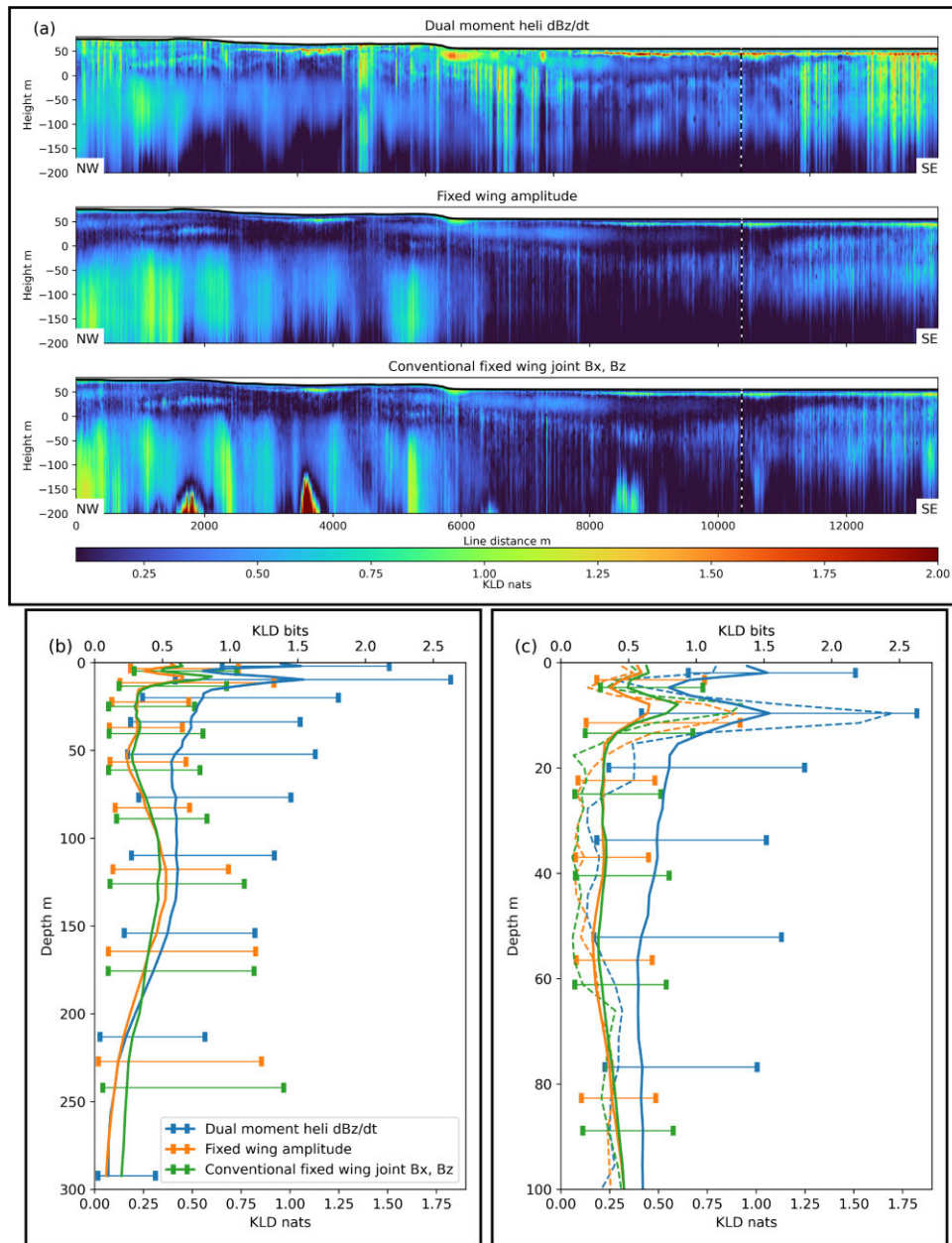
Figure 12. Information gain at the Menindee test site for a) all system/inversion combinations. (b) The median information gain with depth across the surveys, with summary bars showing the 5th and 95th percentiles at every depth. Since geology can be quite variable at the same depth, the interval between the 5th and 95th percentiles simply reflects this variability with depth. (c) The same as (b) but zoomed into the location nearest to BHGT14M1 with dashed lines indicating the information gain at that sounding location.

we do not know the true earth model for real data scenarios, we have not opted for such tweaking to keep the synthetic exercise meaningful.

5 MENINDEE CALIBRATION LINE, NEW SOUTH WALES: COMPARISON WITH BOREHOLES AND INFORMATION GAIN

In the Broken Hill region of New South Wales, Australia, lie Menindee Lakes. Over one of these lakes we operate an AEM testing range. A 12-km-long flight line lies partly over arid ground and partly over a shallow ephemeral lake, co-incident with or very near boreholes with induction logs (Fig. 7). As will be shown later, downhole conductivities from these logs provide a useful comparison of inversion results with ground truth. This in turn allows us to assess within the limits of temporal (seasonal or climatic) variation, the accuracy with which different AEM systems image the subsurface. It is unusual for the same survey line to be systematically flown repeatedly by vendors (see Minsley *et al.* 2021a, for another example of overflown lines), especially in the presence of well-constrained geology and induction logs. Hence the Menindee test range has become a valuable proving ground for AEM technology. This holds true for testing mechanical features and electronic
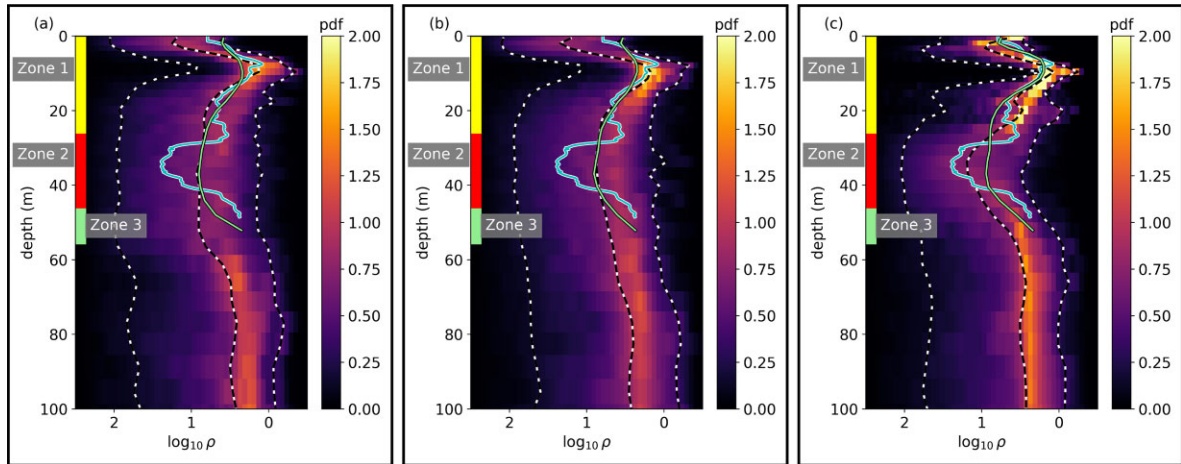
Figure 13. Posterior distributions of conductivity with depth at borehole BHGT14M1 for (a) conventional fixed-wing B_x and B_z joint inversion, (b) fixed-wing amplitude only inversion and (c) helicopter dB_z/dt inversion. The conductivity log has been overlain in cyan. Note the good agreement between the upscaled well-log (thin green line) with the variation in the credible interval as well as with the the median posterior conductivity. The well and the CIs both support a return to conductive media with depth as well as division into three zones discussed in the text.
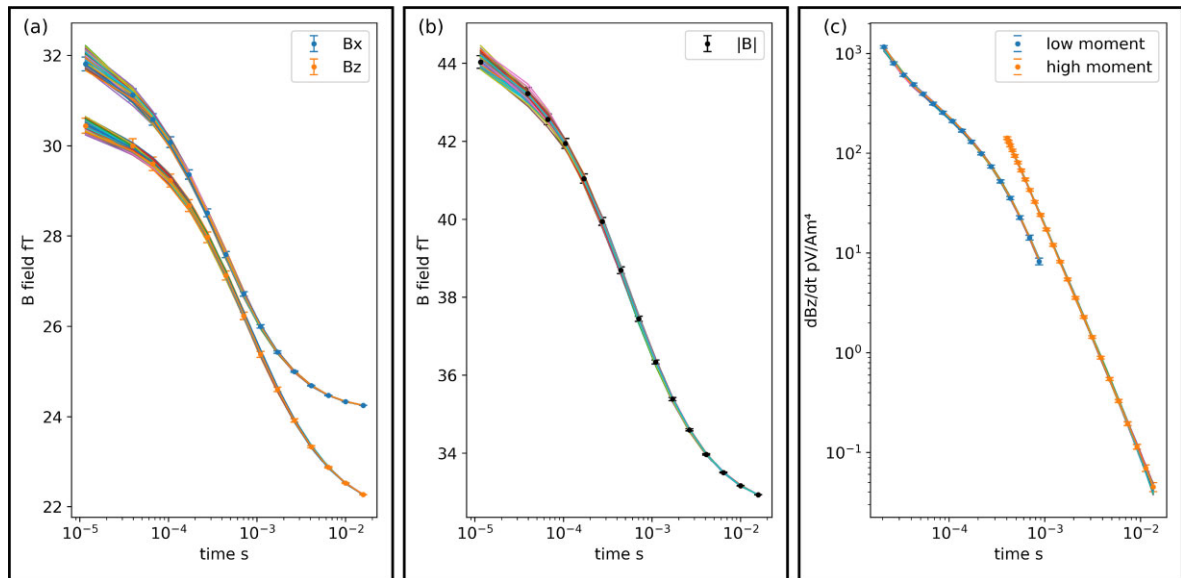

Figure 14. Data fits for 100 randomly selected posterior models at well BHGT14M1 for (a) conventional fixed-wing B_x and B_z joint inversion and (b) fixed-wing amplitude only inversion and (c) helicopter dB_z/dt inversion.

instrumentation of the AEM systems, as well as for inversion codes and modelling theory.

In Fig. 8, we show the results from probabilistic inversions of all helicopter EM soundings inverted independently along the test line. At each sounding, McMC was carried out for 400 000 samples on multiple parallelly tempered interacting chains, with the first 80 000 samples discarded in the burn-in phase. The first row from the top shows the mean square misfit or ϕ_d, with a reasonable fit to data given by the dashed line at $\phi_d = 1$. The second, third and fourth rows from the top show the 5th, 50th and 95th percentile posterior conductivities with depth at every sounding. Similar to the synthetic studies shown in Fig. 4, wherever the percentiles show similar values, the CIs are narrow and hence imply greater posterior certainty. Conversely, a large spread from red to blue among the three percentile images indicate a broad CI and greater posterior *un*certainty. In exactly the same format as just described, in Figs 9 and 10, we show the results for fixed-wing AEM data flown over the test line. The closest location to a well studied borehole which intersects a known near surface regional conductor has been marked in all three figures for closer inspection later in the text. First of all, it is apparent from Figs 8–10 that all three inversions show remarkably similar posterior uncertainties, with exactly the same priors on conductivity, using the same geometry constraints for fixed-wing data as in the synthetic studies, all the while using measured high-altitude noise levels for the test flights. From the northwest to the southeast, as we descend into the lake bed, the lake clearly shows up in all percentiles as a resistive structure relative to its surroundings. To the southeast, in the near surface, there appears to be a layer of clays which show up as conductive. It must be noted that the helicopter system was flown in 2015 and the fixed-wing system in 2017. Differences in the posterior conductivity percentiles between AEM systems, especially in the shallow tens of metres could be due to differences in the subsurface water saturation in these years.

A major point of difference between all three inversions is the presence of relatively certain, deep conductors shown by the joint inversion of fixed-wing data at −200 m relative to the Australian Height Datum (Fig. 9). Notably, these deep features are missing from Fig. 10 which uses the same input AEM data as the joint inversion. The presence of these features cannot be validated as there are no induction logs in their vicinity. Whether the helicopter system and the amplitude-only inversions fail to see these deep conductors, or whether they are artefacts, we are unable to say at this point. A more detailed examination of the fixed-wing nuisances is carried out in Appendix A.

5.1 Comparison with borehole conductivities and deterministic inversion results

Since it is difficult to show probabilistic results and induction log conductivities all together, we have taken the following approach: We display in each panel of Fig. 11 the median model (i.e. 50th percentile) from the probabilistic inversions. For the same data, in the remaining rows we show the results from a deterministic Occam's inversion. Please note that the comparison is only made in the shallow region as the boreholes do not extend deeper. The depth averaged induction-log conductivities have also been displayed on every panel. While comparing AEM bulk conductivities with downhole induction-log conductivities we need to keep in mind that the footprint of the AEM system is far larger. Therefore, it is sufficient that the long-wavelength trends from the induction-log match the inverted AEM conductivities (see Davis & Hauser 2020, and Appendix E for a detailed comparison workflow). We should also note that the wells were logged in different years from the AEM data acquisition. These caveats notwithstanding, there is good agreement between the probabilistic median model, deterministic inversions and the downhole conductivity logs, for all systems. The fact that the deterministic inversion results are smoother than the median conductivity section for all systems can be attributed to the fact that the Occam method produces the smoothest model compatible with the data. The fixed-wing Occam inversions are smoother than their helicopter Occam counterparts, while the difference between systems is not so apparent in the probabilistic inversions. We conjecture that this is due to the all-at-once nature of the updating scheme for geometry nuisances and conductivity values in the non-linear deterministic inversion (see Brodie 2010, for details). This presents itself as an opportunity to improve the existing update and/or regularization schemes that are currently in use for fixed-wing data.

Another point of importance for geological interpretation when examining Figs 8–11, is that in the northwest half of the line the posterior percentiles indicate a return to resistive geology with depth as supported by the depth averaged conductivities in BHMAR63-1 (Fig. 11). However, this is not clear from the deterministic inversion (bottom three rows of Fig. 11) since the inverted models in this region could be returning to the resistive background model at depth and there is no good way to tell from the deterministic inversion alone. Similarly, within the lake itself to the southeast, borehole BHGT14M1 indicates a return to conductive formations at the bottom, as do all the deterministic inversions. Although the deterministic inversions shown here do not do so, a gradient-based inversion can prefer conductivities at $0\log_{10} \Omega$m due to minimum norm updates, and can skew conductive underneath shallow conductors due to the absence of a resistive prejudice model. A probabilistic approach can preclude this scenario as well, if posterior probability

mass moves towards conducting (e.g. near BHGT14M1). We look next in detail at borehole BHGT14M1, after examining the Bayesian information gain at all sounding locations for context.

Fig. 12(a) shows the Bayesian information gain for all surveys and inversions at all locations, calculated using the methods detailed in Appendix D. On average, the helicopter system shows a higher information gain nearer the surface, and the fixed-wing/joint inversion shows higher gain as we go deeper (Fig. 12b). All systems and inversions show increasing information gain in the near surface, where conductors are inferred. Keeping in mind the discussion in Section 4.1 on the synthetic example, we now examine the information gain (Fig. 12c) and resolving capability of all systems in the near surface at BHGT14M1 (Fig. 13), where there is a known strong conductor within the first 25 m, with well established knowledge of the aquifer system.

From Fig. 13 it is apparent that for all three probabilistic inversions, the posterior percentiles of conductivity follow the same trends as the induction log shown with a cyan trace. As mentioned earlier, owing to the temporal changes that may have occurred in subsurface water content, as well the difference in induction logger and AEM system footprints, comparing more than the long-wavelength trends is inadvisable. To facilitate qualitative comparison, the borehole conductivity has been upscaled to the resolution of the AEM data taking into consideration both the AEM system physics as well as the prior correlation length, as detailed in Appendix E, using the methodology of Davis & Hauser (2020). The upscaled log takes into account different physics and data types, but uses the same priors on resistivity and a depth length scale of 2 units. This is the same length scale as is used in the real data inversions, and the upscaled log is shown with a green line in each case (Figs 13a–c). The induction log (cyan) intersects at least three distinct geological zones, which can be interpreted alongside the lithological log and groundwater salinity measurements from different intervals in the sequence. Zone 1 from ∼3 to 29 m coincides with variable conductivity saturated clays of the Willotia Formation and upper Calivil Formation. Zone 2 from 25 to 49 m is of low conductivity down to about 40 m depth (reaching ∼1.4 $\log_{10} \Omega$m) as it intersects stacked sand facies with fresh water from the Calivil aquifer, below which the resistivity decreases gradually to ∼0.4 $\log_{10} \Omega$m due to the presence of brackish water. Although induction log data was not acquired in the underlying Renmark formation (Zone 3), this unit is known to be more conductive than the overlying Calivil formation. While the helicopter data do indeed narrow down the posterior range at ∼10 m by about 1 bit as evidenced by both Figs 12(c) and 13(c), this could be due to temporal variation in near surface conductivity. More importantly at this depth, with all three inversion types, 0.1 $\log_{10} \Omega$m is bracketed at most within an 80 per cent credible width of half a conductivity decade—giving good indication of a conductor. To summarize, the posterior resistivity from all three inversion types are consistent with and well supported by the borehole data as well as known hydrogeology. Fig. 14 assures us that the fits to AEM data at this spatial location nearest the borehole for all inversions are good and that the posterior conclusions drawn are sound.

6 CONCLUSIONS

From our study over the Menindee test site in New South Wales, Australia, as well as from synthetic experiments, we conclude the following: both the low flying helicopter and higher flying fixed-wing time domain AEM systems studied contain comparable

subsurface information. This information largely reflects the true geoelectrical profile—evidenced from the behaviour of inverted posterior conductivities and information theoretic divergences with depth, as well as the compatibility of conductivity CIs with long wavelength trends in upscaled induction logs. Prevalent deterministic fixed-wing AEM inversions are carried out using the nuisance updating strategies laid out in Brodie (2010), and given the similarities in the probabilistic inversion posteriors for all systems (Figs 8, 9 and 10), there is a possibility that this updating strategy could be improved upon. If this possibility can be realized, then the last three rows of Fig. 11 from the top will look as similar to each other as do the first three rows. The implications of such an improvement are that the fixed-wing AusAEM dataset (https://www.eftf.ga.gov.au/ausaem) covering nearly 60 per cent of Australia can be re-inverted to provide greater detail at little extra cost compared to the acquisition of this massive dataset (Ley-Cooper & Brodie 2020). Finally, we formally present in this work the amplitude only inversions for fixed-wing systems, and note that they appear to remove artefacts in the inversion process. However, we are unclear if the features removed could also be considered targets worthy of further geophysical investigation, especially if they fit to within noise in the joint inversion.

Computational considerations for McMC based inversion have been made in Appendix C. Our experience is that the cost of compute is a small fraction of the survey acquisition costs including mobilization charges. This is especially true if we carry out deterministic inversions for all soundings, followed by probabilistic inversions at a decimated set of locations. This ensures we can gather the posterior probabilistic detail to properly characterize the geological framework of interest and make land, water or resource utilization decisions under uncertainty (e.g. Symington *et al.* 2020).

ACKNOWLEDGMENTS

DATA AVAILABILITY

All software used in this work is under active development. Software and examples can be freely cloned from Geoscience Australia's official GitHub repositories:

https://github.com/GeoscienceAustralia/HiQGA.jl
https://github.com/GeoscienceAustralia/GA-AEM

REFERENCES

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

Andrieu, C. & Roberts, G.O., 2009. The pseudo-marginal approach for efficient Monte Carlo computations, *Ann. Stat.,* **37**(2), 697–725.

Auken, E. & Christiansen, A.V., 2004. Layered and laterally constrained 2D inversion of resistivity data, *Geophysics,* **69**(3), 752–761.

Auken, E., Boesen, T. & Christiansen, A.V., 2017. A review of airborne electromagnetic methods with focus on geotechnical and hydrological applications from 2007 to 2017, *Adv. Geophys.,* **58**, 47–93.

Barlow, M., 2019. Geoscience Australia: The year in review, *Preview,* **2019**(201), 9–10.

Bedrosian, P.A., Schamper, C. & Auken, E., 2016. A comparison of helicopter-borne electromagnetic systems for hydrogeologic studies, *Geophys. Prospect.,* **64**(1), 192–215.

Beier, P., Burnham, K.P. & Anderson, D.R., 2002. Model selection and inference: a practical information-theoretic approach, *Model Selection and Inference: A Practical Information-Theoretic Approach,* 2nd edn, Springer-Verlag.

Bezanson, J., Karpinski, S., Shah, V.B. & Edelman, A., 2012. *Julia: A Fast Dynamic Language for Technical Computing,* pp. 1–27, preprint (arXiv:1209.5145). https://doi.org/10.48550/arXiv.1209.5145.

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

Bhattacharyya, A., 1943. On a measure of divergence between two statistical populations defined by their probability distributions, *Bull. Calcutta Math. Soc.,* **35**, 99–109.

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., Ray, A. & Key, K., 2021. Two-dimensional Bayesian inversion of magnetotelluric data using trans-dimensional Gaussian processes, *Geophys. J. Int.,* **226**(1), 548–563.

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

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117**(B2), doi:10.1029/2011JB008560.

Bröcker, J. & Smith, L.A., 2007. Scoring probabilistic forecasts: the importance of being proper, *Wea. Forecast.,* **22**(2), 382–388.

Brodie, R.C., 2010. Holistic inversion of airborne electromagnetic data, *PhD thesis*, pp. 121–127, The Australian National University.

Carvalho, A., 2016. An overview of applications of proper scoring rules, *Decision Anal.,* **13**(4), 223–242.

Chaloner, K. & Verdinelli, I., 1995. Bayesian experimental design: a review, *Stat. Sci.,* **10**(3), 409–435.

Chen, Y., Liu, S., Diethe, T. & Flach, P., 2021. Continual density ratio estimation in an online setting, *Neural Informat. Process. Syst.,* **35**, https://doi.org/10.48550/arXiv.2103.05276.

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.

Cui, S. & Luo, C., 2016. Feature-based non-parametric estimation of Kullback–Leibler divergence for SAR image change detection, *Remote Sens. Lett.,* **7**(11), 1102–1111.

Davis, A. & Hauser, J., 2020. Blocking borehole conductivity logs at the resolution of above-ground electromagnetic systems, *Geophysics,* **85**(2), E67–E77.

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

Dettmer, J., Dosso, S.E. & Holland, C.W., 2010. Trans-dimensional geoacoustic inversion, *J. acoust. Soc. Am.,* **128**(6), 3393–3405.

Dettmer, J., Molnar, S., Steininger, G., Dosso, S.E. & Cassidy, J.F., 2012. Trans-dimensional inversion of microtremor array dispersion data with hierarchical autoregressive error models, *Geophys. J. Int.,* **188**(2), 719–734.

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

Diks, C., Panchenko, V. & Van Dijk, D., 2011. Likelihood-based scoring rules for comparing density forecasts in tails, *J. Econometr.,* **163**(2), 215–230.

Dosso, S.E. & Dettmer, J., 2011. Bayesian matched-field geoacoustic inversion, *Inverse Problems,* **27**(5), doi:10.1088/0266-5611/27/5/055009.

Dosso, S.E., Holland, C.W. & Sambridge, M., 2012. Parallel tempering for strongly nonlinear geoacoustic inversion., *J. acoust. Soc. Am.,* **132**(5), 3030–3040.

Dosso, S.E., Dettmer, J., Steininger, G. & Holland, C.W., 2014. Efficient trans-dimensional Bayesian inversion for geoacoustic profile estimation, *Inverse Problems,* **30**(11), doi:10.1088/0266-5611/30/11/114018.

Earl, D.J. & Deem, M.W., 2005. Parallel tempering: theory, applications, and new perspectives, *Phys. Chem. Chem. Phys.,* **7**(23), 3910–3916.

Eberle, D. & Siemon, B., 2006. Identification of buried valleys using the BGR helicopter-borne geophysical system, *Near Surf. Geophys.,* **4**, 125–133.

Farquharson, C.G. & Oldenburg, D.W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophys. J. Int.,* **134**(1), 213–227.

Fisher, R.A. & Yates, F., 1938. *Statistical Tables: For Biological, Agricultural and Medical Research,* Oliver and Boyd.

Fitterman, D.V. & Yin, C., 2004. Effect of bird maneuver on frequency-domain helicopter EM response, *Geophysics,* **69**(5), 1203–1215.

Friedli, L., Linde, N., Ginsbourger, D. & Doucet, A., 2022. Lithological tomography with the correlated pseudo-marginal method, *Geophys. J. Int.,* **228**(2), 839–856.

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.

Gelman, A., 2006. Multilevel (hierarchical) modeling: what it can and cannot do, *Technometrics,* **48**(3), 432–435.

Geyer, C., 2011. Introduction to MCMC methods, in *Handbook of Markov Chain Monte Carlo,* 1st edn, pp. 3–48, eds Gelman, A., Jones, G., Brooks, S. & Meng, X.-L., Chapman & Hall/CRC.

Geyer, C.J., 1991. Markov chain Monte Carlo maximum likelihood, in *Proceedings of the 23rd Symposium on the Interface,* American Statistical Association, New York, pp. 156–163.

Ghalenoei, E., Dettmer, J., Ali, M.Y. & Kim, J.W., 2021. Gravity and magnetic joint inversion for basement and salt structures with the reversible-jump algorithm, *Geophys. J. Int.,* **227**(2), 746–758.

Gneiting, T. & Raftery, A.E., 2007. Strictly proper scoring rules, prediction, and estimation, *J. Am. Stat. Assoc.,* **102**(477), 359–378.

Good, I.J., 1952. Rational decisions, *J. R. Stat. Soc., B,* **14**(1), 107–114.

Green, A. & Lane, R., 2003. Estimating noise levels in AEM data, *ASEG Extend. Abstr.,* **2003**(2), 1–5.

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

Hastings, W.K., 1970. Monte Carlo sampling methods using Markov chains and their applications, *Biometrika,* **57**(1), 97–109.

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

Hodges, G., 1999. A world of applications for helicopter electromagnetics to environmental and engineering problems, in *Symposium on the Application of Geophysics to Engineering and Environmental Problems 1999,* pp. 899–907, Environmental & Engineering Geophysical Society.

Hoffimann, J., Zortea, M., de Carvalho, B. & Zadrozny, B., 2021. Geostatistical learning: challenges and opportunities, *Front. Appl. Math. Stat.,* **7**(July), 1–15.

Kalscheuer, T., de los Ángeles García Juanatey, M., Meqbel, N. & Pedersen, L.B., 2010. Non-linear model error and resolution properties from two-dimensional single and joint inversions of direct current resistivity and radiomagnetotelluric data, *Geophys. J. Int.,* **182**(3), 1174–1188.

Key, K., 2009. 1D inversion of multicomponent, multifrequency marine CSEM data: methodology and synthetic studies for resolving thin resistive layers, *Geophysics,* **74**(2), doi:10.1190/1.3058434.

Key, K., 2012. Is the fast Hankel transform faster than quadrature?, *Geophysics,* **77**(3), F21–F30.

Key, K. & Lockwood, A., 2010. Determining the orientation of marine CSEM receivers using orthogonal Procrustes rotation analysis, *Geophysics,* **75**(3), doi:10.1190/1.3378765.

Kullback, S. & Leibler, R.A., 1951. On information and sufficiency, *Ann. Math. Stat.,* **22**(1), 79–86.

Lane, R., Green, A., Golding, C., Owers, M., Pik, P., Plunkett, C., Sattel, D. & Thorn, B., 2000. An example of 3D conductivity mapping using the TEMPEST airborne electromagnetic system, *Explor. Geophys.,* **31**(2), 162–172.

Ley-Cooper, A. & Brodie, R., 2020. AusAEM: imaging the near-surface from the world's largest airborne electromagnetic survey, in *Exploring for the Future: Extended Abstracts,* pp. 1–4, eds Czarnota, K., Roach, I., Abbott, S., Haynes, M., Kositcin, N., Ray, A. & Slatter, E., Geoscience Australia.

Ley-Cooper, A.Y., 2021. Exploring for the future AusAEM eastern resources corridor : 2021 airborne electromagnetic survey, Tech. rep., Geoscience Australia, Symonston.

Ley-Cooper, A.Y., 2022. AusAEM – WA, Murchison Airborne electromagnetic survey blocks, Tech. rep., Geoscience Australia, Symonston.

Ley-Cooper, A.Y., Brodie, R.C. & Richardson, M., 2020. AusAEM: Australia's airborne electromagnetic continental-scale acquisition program, *Explor. Geophys.,* **51**(1), 193–202.

Lindley, D.V., 1956. On a measure of the information provided by an experiment, *Ann. Math. Stat.,* **27**(4), 986–1005.

Loseth, L.O. & Ursin, B., 2007. Electromagnetic fields in planarly layered anisotropic media, *Geophys. J. Int.,* **170**(1), 44–80.

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

Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *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), doi:10.1190/1.1778243.

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

Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press.

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

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

Minsley, B.J., Esfahani, A., Deszcz-pan, M., Brodie, R.C. & Survey, U. S.G., 2002. A Bayesian approach to the interpretation of airborne electromagnetic surveys: quantifying data errors, model assessment, and lithology classification, in *Proceedings of the 6th International AEM Conference & Exhibition,* October 2013, cp-383-00021, European Association of Geoscientists & Engineers.

Minsley, B.J., Foks, N.L. & Bedrosian, P.A., 2021a. Quantifying model structural uncertainty using airborne electromagnetic data, *Geophys. J. Int.,* **224**(1), 590–607.

Minsley, B.J., Rigby, J.R., James, S.R., Burton, B.L., Knierim, K.J., Pace, M. D.M., Bedrosian, P.A. & Kress, W.H., 2021b. Airborne geophysical surveys of the lower Mississippi Valley demonstrate system-scale mapping of subsurface architecture, *Nat.. Commun. Earth Environ.,* **2**(1), 1–14.

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

Palacky, G.J. & West, G.F., 1991. Airborne electromagnetic methods, in *Electromagnetic Methods in Applied Geophysics,* Vol. 2, Application, Parts A and B, Society of Exploration Geophysicists.

Parker, R.L., 1994. *Geophysical Inverse Theory,* Princeton Univ. Press.

Pinkard, H. & Waller, L., 2022. A visual introduction to information theory, *Lecture Notes Electr. Eng.,* **785**, 517–558.

Rasmussen, C.E. & Williams, C. K.I., 2006. *Gaussian Processes for Machine Learning,* MIT Press.

Ray, A., 2021. Bayesian inversion using nested trans-dimensional Gaussian processes, *Geophys. J. Int.,* **226,** 302–326.

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. & Myer, D., 2019. Bayesian geophysical inversion with transdimensional Gaussian Process machine learning, *Geophys. J. Int.,* **217,** 1706–1726.

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

Ray, A., Greene, M., Edwards, M. & Cardona, R., 2022. *Methods and systems for calibrating depth in a well to seismic data in a subsurface volume of interest.* Patent.

Roulston, M.S. & Smith, L.A., 2002. Evaluating probabilistic forecasts using information theory, *Mon. Wea. Rev.,* **130**(6), 1653–1660.

Ryan, K.J., 2003. Estimating expected information gains for experimental designs with application to the random fatigue-limit model, *J. Comput. Graph. Stat.,* **12**(3), 585–603.

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. Transdimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167**(2), 528–542.

Seillé, H. & Visser, G., 2020. Bayesian inversion of magnetotelluric data considering dimensionality discrepancies, *Geophys. J. Int.,* **223**(3), 1565–1583.

Sheather, S.J. & Jones, M.C., 1991. A reliable data-based bandwidth selection method for kernel density estimation, *J. R. Stat. Soc., B,* **53**(3), 683–690.

Smith, L.A., Suckling, E.B., Thompson, E.L., Maynard, T. & Du, H., 2015. Towards improving the framework for probabilistic forecast evaluation, *Clim. Change,* **132**(1), 31–45.

Sorensen, K.I. & Auken, E., 2004. SkyTEM—a new high-resolution helicopter transient electromagnetic system, *Explor. Geophys.,* **35**(3), 194–202.

Subašić, S., Agostinetti, N.P. & Bean, C.J., 2019. Estimating lateral and vertical resolution in receiver function data for shallow crust exploration, *Geophys. J. Int.,* **218**(3), 2045–2053.

Sugiyama, M., Nakajima, S., Kashima, H., Von Bünau, P. & Kawanabe, M., 2008a. Direct importance estimation with model selection and its application to covariate shift adaptation, in *Proceedings of the Advances in Neural Information Processing Systems 20 Conference (NIPS 2007),* pp. 1–8.

Sugiyama, M., Suzuki, T., Nakajima, S., Kashima, H., Von Bünau, P. & Kawanabe, M., 2008b. Direct importance estimation for covariate shift adaptation, *Ann. Inst. Stat. Math.,* **60**(4), 699–746.

Sugiyama, M., Suzuki, T. & Kanamori, T., 2012. *Density Ratio Estimation in Machine Learning,* Cambridge Univ. Press.

Sugiyama, M., Liu, S. & Christoffel, M., 2013. Direct divergence approximation between probability, *J. Comput. Sci. Eng.,* **7**(2), 99–111.

Swendsen, R.H. & Wang, J.S., 1987. Nonuniversal critical dynamics in Monte Carlo simulations, *Phys. Rev. Lett.,* **58**(2), 86–88.

Symington, N., Ray, A., Harris-Pascal, C., Tan, K.P., Ley-Cooper, Y. & Brodie, R.C., 2020. Groundwater salinity estimation using borehole and AEM data: a framework for uncertainty analysis, in *EFTF Extended Abstracts,* August, Commonwealth of Australia (Geoscience Australia), doi:10.11636/135242.

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

Tarantola, A. & Valette, B., 1982. Inverse problems= quest for information, *J. Geophys.,* **50,** 159–170.

Valentine, A.P. & Sambridge, M., 2020. Gaussian process models—I. A framework for probabilistic continuous inverse theory, *Geophys. J. Int.,* **220**(3), 1632–1647.

Weijs, S.V., van Nooijen, R. & van de Giesen, N., 2010. Kullback-leibler divergence as a forecast skill score with classic reliability-resolution-uncertainty decomposition, *Mon. Wea. Rev.,* **138**(9), 3387–3399.

Worrall, L., Whitaker, A., Lane, R. & Meyers, J., 2001. Exploring through cover – the integrated interpretation of high resolution aeromagnetic, airborne electromagnetic and ground gravity data from the Grant's Patch area, Eastern Goldfields Province, Archaean Yilgarn Craton. Part A: mapping geology using airb, *ASEG Extended Abstracts,* **2001**(1), 1–4.

Yardim, C., Gerstoft, P. & Hodgkiss, W.S., 2006. Estimation of radio refractivity from radar clutter using Bayesian Monte Carlo analysis, *IEEE Trans. Antenn. Propagat.,* **54**(4), 1318–1327.

APPENDIX A: NUISANCES FOR FIXED-WING INVERSIONS OVER MENINDEE

The inverted conductivities as well as acquisition geometry nuisances are shown in this section, both for the conventional fixed-wing joint inversion (Fig. A1) and for the same input AEM data, the amplitude only inversion (Fig. A2). The fact that the inverted nuisance values (blue lines and shaded blue regions) do not always overlap with the measured IMU-provided values (orange) is not surprising, as the IMU readings could be inaccurate, or it is possible that the inversion has found some trade-off as was shown in the synthetic examples. What is important however, is that the inversion be parametrized to allow for geometry nuisance inference, as in their absence, the data residual would be propagated incorrectly into the inverted conductivities.
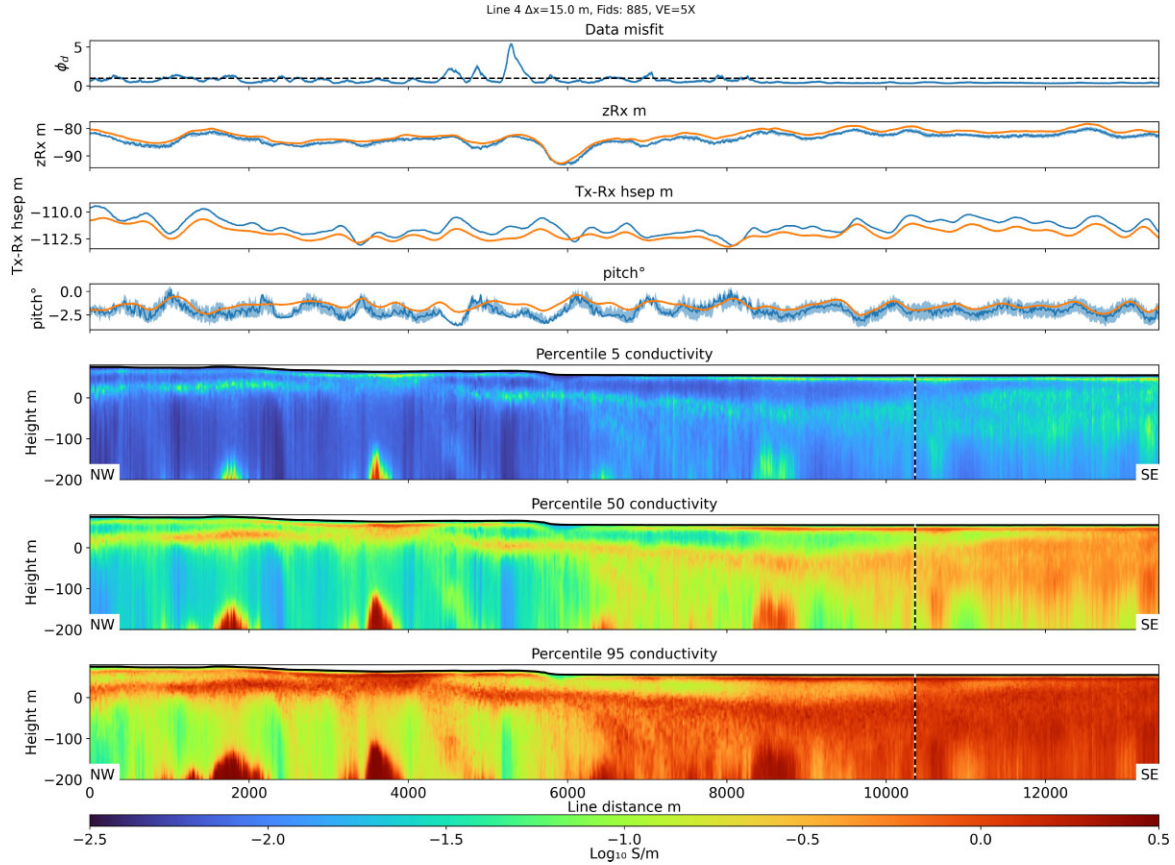

Figure A1. Summary of posterior nuisances for conventional fixed-wing B_x and B_z joint inversion over Menindee. The top row shows the mean sampled ϕ_d and one standard deviation. The next three rows show in blue the mean sampled receiver height, Tx–Rx horizontal separation and receiver pitch, as well as their one standard deviation spread. Orange values indicate what has actually been measured during the survey. The last three rows, as before, show the 5th, 50th (median) and 95th percentiles of posterior conductivity. Over the deep conductors at -200 m, there do not appear to be any sudden changes in geometry during the flight which may otherwise cause artefacts.

APPENDIX B: TRANS-DIMENSIONAL ALGORITHM WITH NUISANCE SAMPLING AND PARALLEL TEMPERING

As shown by the pseudocode provided in Algorithm B, one step of our algorithm encapsulates a reversible jump or trans-dimensional step (starting at Line 4), followed by a nuisance sampling step (starting at Line 11), contained within a parallel tempering loop (Lines 3-20), followed by a parallel tempering swap (Lines 21–28).

Details of the trans-dimensional birth and death moves for resistivity models are exactly the same as described for stationary Gaussian processes (Ray & Myer 2019; Ray 2021). The nuisance sampler uses ordinary Metropolis–Hastings (Metropolis *et al.* 1953; Hastings 1970) for the fixed-wing geometry nuisance parameters which are $n_{\text{nuisances}}$ in number.

Parallel tempering is used to exchange information between interacting McMC chains to escape local misfit minima (i.e. likelihood maxima). Temperatures or models are exchanged at the end of each McMC step using the following Metropolis–Hastings criterion (Swendsen & Wang 1987; Geyer 1991; Earl & Deem 2005; Dettmer *et al.* 2012; Sambridge 2013):

$$\alpha_{swap}(a, b) = \min \left[1, \left\{ \frac{\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{m}_b})}{\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{m}_a})} \right\}^{1/T_a} \left\{ \frac{\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{m}_a})}{\mathcal{L}(\boldsymbol{\theta}_{\boldsymbol{m}_b})} \right\}^{1/T_b} \right]. \tag{B1}$$
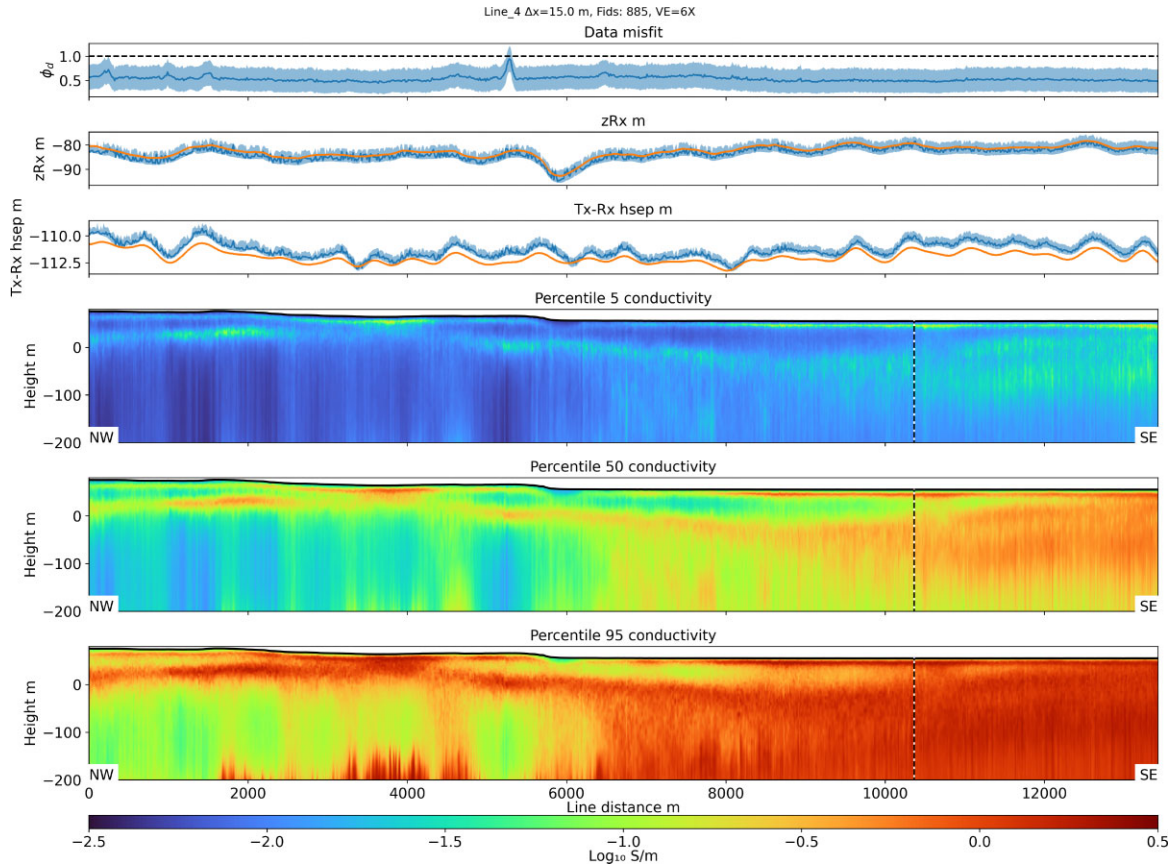
Figure A2. Summary of posterior nuisances for fixed-wing amplitude only inversion over Menindee. The top row shows the mean sampled ϕ_d and one standard deviation. The next two rows show the sampled receiver height and Tx–Rx horizontal separation, as well as their one standard deviation spread. Orange values in these rows indicate what has actually been measured during the survey. The final three rows, as before, show the 5th, 50th (median) and 95th percentiles of posterior conductivity. The deep conductors at -200 m from the fixed-wing amplitude only inversion do not appear in these posterior summaries.
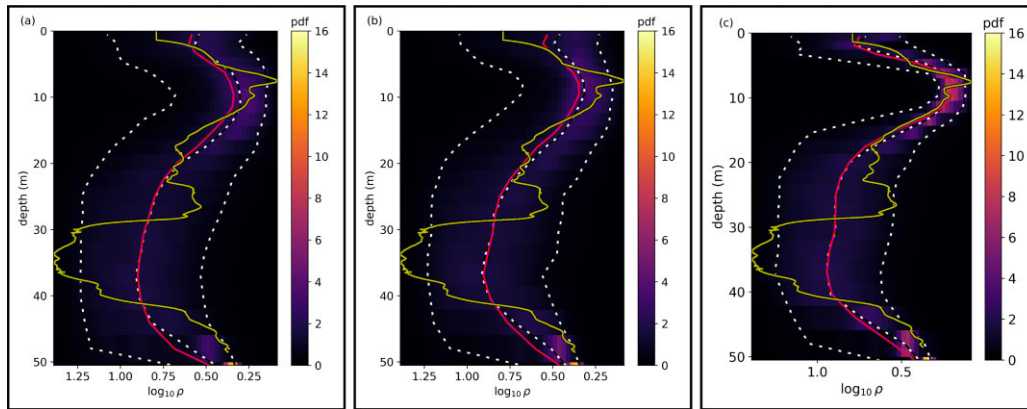

Figure A3. Posterior conductivities obtained after forward modelling the well log conductivities (yellow line) at borehole BHGT14M1 at Menindee lakes for (a) joint B_x, B_z inversion, (b) amplitude only B field and (c) helicopter dB_z/dt data. As in the main text, the dashed lines represent the 5, 50 and 95 percent posterior percentiles of conductivity. In each case, the posterior median has been used as the upscaled log for qualitative comparison in Fig. 13.

For a description of why swapping models for escaping local likelihood maxima using (B1) is effective, see section 3.2 of Blatter *et al.* (2018). For computational efficiency, temperatures are exchanged during inter-process communication to achieve the exact same effect as swapping models. No likelihood computations are required in the calculation of the swap probability in (B1), as they have already been evaluated in the preceding *j* loop on Line 3 of Algorithm B.

All Markov chains were run at log-spaced temperatures between 1 and 2.5. Details of setting a temperature ladder can be found in Dettmer *et al.* (2012) and Ray *et al.* (2013). The helicopter AEM inversions required five temperatures and no nuisance sampling, whereas the fixed-wing inversions used 7 temperatures with nuisance sampling. Larger numbers of temperatures are required to sample rugged likelihoods. Posterior inference is carried out only from models that are at $T = 1$ after sorting the chains by temperature.

---

1 Initialise McMC chains with stationary Gaussian process resistivity model $\boldsymbol{\theta}_{\rho_j}$ and nuisance properties model $\boldsymbol{\theta}_{n_j}$ for temperatures T_j where $j = 1, 2, ..., nTemps$. If nuisances are to be sampled, set *samplenuisance* to **true**.

2 **for** $i \leftarrow 1$ **to** $nSteps$ **do**

3   **for** $j \leftarrow 1$ **to** $nTemps$ **do**

4   Select $type$ from [**birth, death, fixed dimension**] with probability $[\frac{1}{3}, \frac{1}{3}, \frac{1}{3}]$

5 $\boldsymbol{\theta}_{\rho_j}[i] \leftarrow \boldsymbol{\theta}_{\rho_j}[i-1]$

6 $\boldsymbol{\theta}'_{\rho_j} \sim q_{type}(\boldsymbol{\theta}'_{\rho_j} | \boldsymbol{\theta}_{\rho_j}[i])$

7 $u \sim U(0, 1)$

8   **if** $u < \alpha_j^{type}(\boldsymbol{\theta}'_{\rho_j} | \boldsymbol{\theta}_{s_j}[i])$ and $p(\boldsymbol{\theta}'_{\rho_j}) > 0$ **then**

9 $\boldsymbol{\theta}_{\rho_j}[i] \leftarrow \boldsymbol{\theta}'_{\rho_j}$

10   **end**

11   **if** *samplenuisances* **then**

12   Select $type$ from **nuisance moves** with probability $\frac{1}{n_{\text{nuisances}}}$

13 $\boldsymbol{\theta}_{n_j}[i] \leftarrow \boldsymbol{\theta}_{n_j}[i-1]$

14 $\boldsymbol{\theta}'_{n_j} \sim q_{type}(\boldsymbol{\theta}'_{n_j} | \boldsymbol{\theta}_{n_j}[i])$

15 $u \sim U(0, 1)$

16   **if** $u < \alpha_j^{type}(\boldsymbol{\theta}'_{n_j} | \boldsymbol{\theta}_{n_j}[i])$ and $p(\boldsymbol{\theta}'_{n_j}) > 0$ **then**

17 $\boldsymbol{\theta}_{n_j}[i] \leftarrow \boldsymbol{\theta}'_{n_j}$

18   **end**

19   **end**

20   **end**

21   **for** $a \leftarrow nTemps$ **to** 2 **do**

22 $b \sim U(1, a); b \in \mathbb{I}$

23   **if** $a \neq b$ **then**

24   **if** $u < \alpha_{swap}(a, b)$ **then**

25   swap T_a and T_b

26   **end**

27   **end**

28   **end**

29 **end**

Algorithm B: Pseudocode for trans-dimensional McMC with nuisance sampling, and parallel tempering exchanges though a Fisher-Yates shuffle (Fisher & Yates 1938).

APPENDIX C: HIGH PERFORMANCE COMPUTING CONSIDERATIONS

It can be seen from Algorithm B that the nuisance sampler does require likelihood evaluation (i.e. a forward call). Thus sampling the fixed-wing system posterior requires twice as many forward calls as it does to sample the helicopter system posterior. For the real data examples, one sampling step required 70-80 ms for the fixed-wing system, and 40–45 ms for the helicopter system. This evaluation time is dominated by the forward evaluation for the 52 layer model. This is fortunate since, if each McMC sample took less time than a few ms, the overhead in inter-process communication per McMC step would make the parallel computation inefficient. Parallel tempering is not an embarrassingly parallel algorithm, owing to the requirement that information be exchanged between interacting chains. This can be seen from Algorithm B on Line 3, where an independent McMC sampling step is made at each temperature, with the exchange of temperatures occurring on Line 21 of the algorithm. Care must be taken in the computer implementation of this exchange step to transfer the bare minimum information possible, in order to the make the inter-process communication efficient.

The real data examples required 200 000 McMC samples for each chain in parallel tempering, though we extended each sampling run to 400 000 samples. For each sounding, 200 000 samples required approximately 4.5 hr for the fixed-wing system on 7 + 1 CPUs, and 2.5 hr for the helicopter system on 5 + 1 CPUs. The +1 indicates the manager process for inter-CPU communication and each remaining CPU was

assigned to McMC computation at a specified temperature. A significant improvement has been made in the HPC implementation of parallel tempering in this work compared to our earlier work in Blatter *et al.* (2018), which was based on the parallel tempering implementation of Ray *et al.* (2013). As noted by Blatter *et al.* (2018), one single-moment helicopter AEM sounding required 5 d to invert probabilistically. In our current work, soundings were batched in parallel such that 120 soundings for the fixed-wing system, and 160 soundings for the helicopter system were carried out at the same time using 960 CPUs. To be explicit, an entire batch of 160 soundings were probabilistically inverted within 2.5 hr for the helicopter system using 960 CPUs. The full set of 720 helicopter soundings over the Menindee test range required five batches and a total of 12.5 hr. If more CPUs are made available, then a greater number of soundings can be parallelly inverted in a batch, and fewer numbers of batches will be required in total. Similar calculations can be made for the fixed-wing system. On average, the fixed-wing posterior sampling results require twice as much time with the same computational resources, due to nuisance sampling needing an additional forward evaluation—though this can perhaps be avoided by reusing precomputed Hankel transform evaluations (future work).

The Bayesian inversion including the forward modeller code is written purely in Julia (v1.6, tested on v1.7-1.8). The deterministic inversion code including the forward modeller is written in modern C++. The parallel computation in this study used 48-core @3.2 GHz Intel Xeon Cascade Lake nodes in the NCI's *Gadi* cluster.

APPENDIX D: INFORMATION THEORETIC BACKGROUND

D1 Relative entropy and information gain

Uncertainty analysis in a Bayesian sense naturally leads to the formulation of information gain (Lindley 1956), in terms of the difference in information when comparing the posterior and prior distributions. This is provided by the relative entropy between the two densities, known as the Kullback–Leibler divergence (Kullback & Leibler 1951). The KLD is not symmetric between the densities that differ, and to avoid confusion, for information gain we will use the definition of the KLD for prior density $q(x)$ and posterior density $p(x)$ as defined in eq. (22) of Valentine & Sambridge (2020):

$$D_{\mathrm{KL}}(p||q) = \int p(x) \log \frac{p(x)}{q(x)} \mathrm{d}x. \tag{D1}$$

When this is computed with the natural logarithm as given in (D1), the information gain units are known as 'nats' and a change of base to 2 (i.e. multiplying (D1) by $\log_2 e$) changes the units to bits. As we will prove shortly, the KLD is zero if and only if the densities p and q are identical, otherwise it is always an unbounded positive quantity. We can gain a useful perspective on the KLD by looking at it as the expectation of the log ratio of the posterior to the prior when x is distributed according to the posterior,

$$D_{\mathrm{KL}}(p||q) = \mathbf{E}_{x \sim p(x)} \left[\log \frac{p(x)}{q(x)} \right]. \tag{D2}$$

Eq. (D2) sets the stage for both the estimation of the KLD as information gain from samples, and showing how the KLD leads to the derivation of the logarithmic score, a *strictly proper* scoring rule (Gneiting & Raftery 2007).

D2 Density ratio estimation for Bayesian information gain

If the densities p and q, respectively, can be parametrically estimated from their samples, it may be possible to analytically compute the KLD. This is the approach followed by (Friedli *et al.* 2022) in assessing how well different methods approximate a known, Gaussian posterior for a linear inverse problem with a Gaussian prior. However, calculation of the KLD is subject to the usual difficulties (Cui & Luo 2016) when calculating entropy, if parametric approaches are not suitable, as in our AEM inversion case. Histogram binning for p and q may be problematic as this will depend on the number and location of histogram bins. Another possibility is separately estimating densities for p and q and then carrying out numerical or Monte Carlo integration, but the derived quantity p/q is unstable and may be be difficult to use. In cases where a density ratio p/q is the quantity of interest (as in (D2)), it is advantageous to resort to direct methods for density ratio estimation (see Sugiyama *et al.* 2012, for a comprehensive review). A recent example requiring direct density ratio estimation in the geosciences can be found in Hoffimann *et al.* (2021), for geostatistical transfer learning. In the particular case of the Kullback-Leibler divergence, the Kullback–Leibler Importance Estimation Procedure (KLIEP, Sugiyama *et al.* 2008a) can be used effectively for direct divergence estimation as detailed by Sugiyama *et al.* (2013). As can be seen from (D2), the KLD is given by the expected value under the posterior density, of the logarithm of the density ratio of the posterior to the prior. This is done in a three step process. First, with McMC realizations of prior and posterior samples, KLIEP is used to fit a linear combination of Gaussian kernels directly to the ratio $r(x) = \frac{p(x)}{q(x)}$. In a second step, $r(x)$ is computed from the KLIEP estimated density ratio function when x is simulated from stored posterior samples. Finally, the mean of $\log r(x)$ provides the KLD as Bayesian information gain according to (D2). The KLIEP procedure selects the Gaussian kernel bandwidth to approximate r through cross-validation, and a constrained, convex optimisation problem is solved to obtain the density ratio estimate (see Sugiyama *et al.* 2008a, for details). Sampling the prior is inexpensive as it requires no likelihood evaluations (i.e. forward calls), and the posterior samples must be obtained as usual to solve the AEM Bayesian inverse problem.

D3 Logarithmic scoring rule

Scoring rules are used to evaluate the quality of probabilistic forecasts and have found heavy use in weather prediction and econometrics. A comprehensive review can be found in Gneiting & Raftery (2007). A scoring rule is proper, if incorrect forecast densities are expected

to receive lower scores than the true forecast density, and it is strictly proper, if the maximum expected score attained is unique. While they have not been used much in near surface geophysics, we note recent applications in Seillé & Visser (2020) for selecting the best likelihood function for Bayesian inversion, and in Friedli *et al.* (2022) for evaluating the performance of different Monte Carlo methods. We shall now prove that the logarithmic score, which dates back as far as Good (1952), is related to the KLD, by first showing that the KLD is always greater than or equal to zero. As (D2) provides the general definition of $D_{KL}(p||q)$, not just for priors and posteriors, we can write for *any* two valid probability densities p and q:

$$- D_{KL}(p||q) = \mathbf{E}_{x \sim p(x)}\left[\log \frac{q(x)}{p(x)} \right]. \tag{D3}$$

Using the inequality $1 + \log u \leq u$, with exact equality at 1, we can substitute $u \to \frac{q}{p}$ and take expectations with respect to p. We can thus write:

$$1 + \mathbf{E}_{x \sim p(x)}\left[\log \frac{q(x)}{p(x)} \right] \leq \mathbf{E}_{x \sim p(x)}\left[\frac{q(x)}{p(x)} \right], \tag{D4}$$

$$1 + \mathbf{E}_{x \sim p(x)}\left[\log \frac{q(x)}{p(x)} \right] \leq \int p(x)\frac{q(x)}{p(x)}dx. \tag{D5}$$

Now using the fact that $\int q(x)dx = 1$ and using (D3) in (D5), we can say that

$$1 - D_{KL}(p||q) \leq 1, \tag{D6}$$

thus proving that

$$D_{KL}(p||q) \geq 0; \text{with exact equality only at } p(x) = q(x) \ \forall x. \tag{D7}$$

The above inequality is known as Gibb's inequality and establishes with no loss of generality that the KLD is always greater than zero unless the two densities are identical. Now if we were to consider the logarithmic scoring rule, and p and q are the true and forecast densities respectively, then we can formulate the Kullback–Leibler Information Criterion (see Diks *et al.* 2011) for q as:

$$D_{KL}(p||q) = \int p(x)\left[\log p(x) - \log q(x) \right]dx \geq 0, \tag{D8}$$

$$\Rightarrow \int p(x)\log q(x)dx \leq \int p(x)\log p(x)dx \ \forall x. \tag{D9}$$

Inequality (D9) establishes that for the logarithmic score

$$S(q, x) = \log q(x), \tag{D10}$$

$$\int p(x)S(q, x)dx \leq \int p(x)S(p, x)dx \ \forall x. \tag{D11}$$

The integral on the left hand side of (D11) is the expectation of the incorrect forecast score over all possible values $x \sim p(x)$ where $p(x)$ is the true density. This quantity is always less than or equal to the average of the true forecast score, which is the integral on the right hand side of (D11). This shows that the requirement of propriety is satisfied by the logarithmic scoring rule. Since the maximum expected forecast score is reached uniquely for $q(x) = p(x) \ \forall x$, the logarithmic scoring rule is also strictly proper.

D4 Summary of key information theoretic points

The KLD or relative entropy is a very useful information theoretic tool to discriminate between two probability densities (e.g. MacKay 2003). When defined with p as posterior and q as prior, the KLD as given in (D1) represents the information gain after carrying out our geophysical experiment. The KLD can be calculated at every location in the earth, showing the information gain at all depths considered in the geophysical inversion (e.g. Blatter *et al.* 2018; Valentine & Sambridge 2020). In the context of assigning scores to posterior predictions, if p is the true forecast distribution and q is the posterior, then $S(q, x) = \log q(x)$ represents a strictly proper score. If we were to treat the posterior resistivity as a prediction of the truth, then at every location in the earth (D10) provides a strictly proper score, with higher values representing better performance. For an interpretation in terms of ignorance, given by the negative of the score defined in (D10), we refer readers to Roulston & Smith (2002) as well as Weijs *et al.* (2010).

It must be noted that for assessing posterior uncertainty, in the domain of the physical sciences, this is done in terms of the width of quantities of interest. Broader ranges of probable values indicate lower certainty. In this regard, we have found the KLD for information gain to be a more robust indicator of subsurface property resolution in terms of higher values indicating lower overlap between a narrow posterior and broad prior. The logarithmic score is very sensitive to probability modes, and severely penalises narrow posteriors if they have missed true values. This is a source of debate; see section 4, page 366 of Gneiting & Raftery (2007) and section 5, page 387 of Bröcker & Smith (2007). In fact, Friedli *et al.* (2022) report that only a certain fraction of parameters in their geophysical simulations have true values that fall within

the range of the posterior samples. This is important to keep in mind as low probability regions will score very poorly (or even have infinite ignorance $=-\log 0$) and can bias the average score for a particular posterior density. While other non-local scores exist, they can be relatively insensitive to the width of probability modes (Smith *et al.* 2015) and are not considered here. In sum, both logarithmic scores and information gain are useful analysis tools derived from appropriate usage of the KLD. However, information gain can be estimated when true parameter values are not exactly known (i.e. most of the earth) for real data. Criteria such as mean absolute error could be used to quantify closeness to the truth, but again this is not as generally useful as information gain for similar reasons. Further, information gain is not sensitive to true values falling adjacent to high posterior probability regions—a phenomenon which is common in non-linear geophysical inverse problems.

APPENDIX E: WELL-LOG UPSCALING

In order to populate the static (or dynamic) properties of a reservoir (flow) model, or for comparison with inversion results from surface geophysical data, borehole logs are often upscaled or adjusted to be compatible with the surface data. Due to their parsimonious nature, trans-dimensional methods have been found to be highly suited to this task. Using such methods, seismic well ties to adjust surface migration imaging results with downhole sonic logs have been successfully automated (Ray *et al.* 2022) and conductivity logs have been represented at the resolution of AEM systems (Davis & Hauser 2020). For the AEM problem we have followed the latter approach as it is directly applicable to our case. Their thesis is elegant yet simple, that the best representation of the well log at the scale of above-ground geophysical systems, is through a central tendency of synthetic posterior conductivities. The synthetic posterior is obtained after inverting the forward modelled borehole conductivity log using the transmitter characteristics of the overflying AEM system. Using the measured survey data noise and within the prior bounds given by borehole conductivities, this approach takes into account both the Earth filter due to AEM forward physics, as well as any assumptions made in the inversion such as choice of prior length scale. The results of modelling the wells and sampling the posterior conductivities per AEM system are provided in Fig. A3. We have taken the median posterior conductivity, and used it to represent the long-wavelength features of the borehole log in Fig. 13.