

Bootstrap resampling as a tool for uncertainty analysis in 2-D magnetotelluric inversion modelling

Sebastian Schnaidt^{1,2} and Graham Heinson²

¹Deep Exploration Technologies Cooperative Research Centre, Adelaide Airport SA 5950, Australia. E-mail: sebastian.schnaidt@adelaide.edu.au

²Electrical Earth Imaging Group, Department of Earth Sciences, School of Physical Sciences, University of Adelaide, Adelaide SA 5005, Australia

Accepted 2015 June 15. Received 2015 June 4; in original form 2014 December 8

SUMMARY

Uncertainty estimation is a vital part of geophysical numerical modelling. There exist a variety of methods aimed at uncertainty estimation, which are often complicated and difficult to implement. We present an inversion technique that produces multiple solutions, based on bootstrap resampling, to create a qualitative uncertainty measure for 2-D magnetotelluric inversion models. The approach is easy to implement, can be used with almost any inversion code, and does not require access to the inversion software's source code. It is capable of detecting the effect of data uncertainties on the model result rather than just analysing the effect of model variations on the model response. To obtain uncertainty estimates for an inversion model, the original data set is resampled repeatedly and alternate data set realizations are created and inverted. This ensemble of solutions is then statistically analysed to determine the variability between the different solutions. The process yields interpretable uncertainty maps for the inversion model and we demonstrate its effectiveness to qualitatively quantify uncertainty in synthetic model tests and a case study.

Key words: Instability analysis; Inverse theory; Magnetotellurics.

1 INTRODUCTION

Numerical inversion has become a standard approach in the interpretation of data from electromagnetic (EM) measurements. Models are a widely traded commodity in research as well as in the exploration industry. This requires them to be highly reliable to ensure correct interpretation, which can only be achieved if model shortcomings and the model parameter uncertainties are thoroughly understood.

The model appraisal tool bag contains a variety of different techniques and methods to quantify the influences of data noise, incorrect *a priori* assumptions and solution non-uniqueness on model resolution, model sensitivity and parameter uncertainties of EM models. To put the uncertainty estimation technique presented in this paper into perspective, we will first mention some examples of alternative methods and groups of methods to analyse different aspects of model uncertainty and discuss their advantages and shortcomings. Note that our list is in no way complete.

Parameter uncertainties can, for example, be analysed using the ‘model covariance matrix’ (MCM; Meju 1994; Alumbaugh & Newman 2000; Tarantola 2005; Menke 2012) or via full probabilistic Markov chain Monte Carlo (MCMC) inversion (Rosas-Carbajal *et al.* 2014), and parameter ranges can be estimated with ‘most-squares’ inversions (Jackson 1976; Meju & Hutton 1992; Kalscheuer *et al.* 2010). The sensitivity of a model can be assessed via the ‘Jacobian matrix’ (sensitivity matrix) and by

manual forward and inversion modelling tests of perturbed model results (see e.g. Becken *et al.* 2008; Thiel & Heinson 2010). Another method to test model features on their reliability is to use the model’s ‘null-space’ (Muñoz & Rath 2006), and model resolution can be estimated by analysing the ‘model resolution matrix’ (MRM; Jackson 1972; Meju 1994; Ramirez *et al.* 1995; Menke 2012). Below we will discuss these methods in more detail.

Monte Carlo methods (Metropolis & Ulam 1949) are commonly used in the determination of different kinds of uncertainty. They produce an ensemble of equally possible solutions that is used to calculate posterior probability distributions for the model parameters, which give direct indications of parameter uncertainty. As random searches are computationally expensive, usually MCMC methods (see e.g. Gamerman 1997) are used for numerical inversion modelling, instead of straight Monte Carlo searches. Rosas-Carbajal *et al.* (2014) employed a pixel-based, fully probabilistic MCMC inversion for 2-D plane wave EM data. To make the inversion computationally feasible, the model structure has to be constrained. Rosas-Carbajal *et al.* (2014) point out that the uncertainty estimates are to some extent determined by these constraints and that the constraints might lead to biased models. Other factors influencing the results are the model parameterization and the number of parameters, which seem to have great impact on the posterior probability distributions (Laloy *et al.* 2012; Linde & Vrugt 2013), as well as the important choice of the parameter priors necessary for the inversion. The result of an MCMC inversion strongly depends on the

used prior, as the resulting model can only reach regions of the model space that are contained in the prior. Hence, an insufficient prior can have adverse effects on the result.

Rosas-Carbajal *et al.* (2014) compared the uncertainty estimates of the MCMC inversion with results from a most-squares inversion (Jackson 1976; Meju & Hutton 1992), which has also been used by Kalscheuer *et al.* (2010) for joint inversion of direct current resistivity and radio-magnetotelluric (MT) data. Most-squares inversions are used to find the extremal values that the parameters of each model cell can adopt without altering the misfit beyond a specified threshold with respect to the best-fitting inversion model found by conventional means. This gives information about the feasible model ranges and model non-uniqueness.

MCM analysis (Meju 1994; Tarantola 2005; Menke 2012) is, for example, used by Alumbaugh & Newman (2000). They used a linearized MCM approach to assess EM models from 2-D and 3-D inversions. The MCM contains the covariances between all parameters and the diagonal of the matrix represents the parameter variances. Usually the square root of these values is plotted to produce an image of the parameter standard deviations, which gives an estimate of how the parameter errors are influenced by data noise and incorrect *a priori* assumptions. These are valuable information, but they have to be treated with care. Alumbaugh & Newman (2000) suggest that the reliability of the linearization, necessary to calculate the MCM for the highly non-linear EM problem, is difficult to determine. The need to linearize a non-linear problem might underestimate the error levels. The authors suggest that performing full non-linear inversions of multiple realizations of the model with different data and smoothing vectors could produce different results. However, this does not lessen the (at least qualitative) value of the information about the model quality that can be gained.

Alumbaugh & Newman (2000) also used a linearized approach to determine the MRM (Jackson 1972; Meju 1994; Ramirez *et al.* 1995; Menke 2012). As with the MCM, usually the diagonal of the MRM is evaluated to gain information about the model resolution. However, the calculation and evaluation of the MRM suffer from similar problems as the MCM.

The methods discussed so far mostly estimate model parameter uncertainties. A different group of techniques is used to analyse the sensitivity of the model and the reliability of specific model features. Muñoz & Rath (2006) use a null-space projection technique to test the non-uniqueness of MT models. The method has first been employed by Deal & Nolet (1996) and Rowbotham & Pratt (1997) on seismic tomography inversions and is based on perturbations of the best-fitting model generated with a standard inversion run. The parameters of the model to be analysed are perturbed by small amounts multiple times and those changes are projected onto the original model's null-space, producing multiple realizations of the model. Due to the projection of the changes onto the model's null-space, the model misfit is not altered. This corresponds to the exploration of the parameter space close to the original model. The resolution and reliability of model structures can then be assessed by comparing the differences and similarities of the inversion results of the new models. It is important to note that this technique does not make any statement as to how small changes to the data, for example those caused by noise, influence the model response. The method also has a few issues. First, the approach will only yield correct results if the chosen base model is close to the true model, which the whole process tries to determine. Second, the choice of setup parameters can greatly influence the outcome. Especially the null-space is greatly dependent on parameter choices and can vary significantly, which will influence the results.

A common method to determine the reliability and sensitivity of model structures is based on models that are manually perturbed and tested against the original model result via forward and inversion modelling (see e.g. Becken *et al.* 2008; Thiel & Heinson 2010). Its common use is founded in its relative simplicity and its comparatively easy implementation. To test structures and features of a model found by classic inversion, the model cells representing these features are changed to parameter values corresponding with the model background. These test models are then assessed by forward modelling or by using the new model, with the perturbed regions held fixed, as the starting model for additional inversions. Deterioration of the misfits of these new models would then indicate that the model features in question are constrained by the data. This checks to what extent the features in question are required by the data with certain error estimates.

In addition to model perturbation tests, Thiel & Heinson (2010) analyse the Jacobian matrix (sensitivity matrix), which provides information on the influence of small changes to a model parameter on the model response. Both the model perturbation method and the sensitivity matrix analysis suffer from the same shortcoming as the null-space projection approach discussed earlier; they only investigate how changes to the model influence the model response. In contrast the influence of perturbations to the data with regard to their error estimates cannot be inferred directly.

The methods described provide valuable information about the quality and reliability of models, but also have their shortcomings. They are often relatively complicated to implement, especially if one does not have access to the source code of the used inversion software.

In this paper, we present an approach to assess the effect of data errors on model uncertainty and reliability in 2-D MT inversion modelling using bootstrap resampling (Efron 1979, 1982). Like MCMC methods our approach creates multiple solutions. The exploration of the solution space is less comprehensive than with MCMC methods, which means the bootstrap approach yields qualitative rather than quantitative uncertainty estimates, while achieving higher computational efficiency. The model uncertainty is assessed with respect to a set of fixed starting parameters. The original data are resampled repeatedly, creating multiple realizations of the data set, which are each inverted to create a solution ensemble. This ensemble is statistically analysed to determine differences and similarities between the different solutions, allowing conclusions about the model quality. Our approach has the advantage of giving an indication of the effect of the data uncertainties on the model outcome, while being easy to implement. The method can be realized with almost any existing inversion code, as the availability of the inversion software's source code is not essential. Only a limited amount of new code has to be developed.

In the remainder of the paper we will describe our method, and especially bootstrap resampling, in more detail (Section 2), describe how the produced solution ensembles are analysed (Section 3), demonstrate the approach with a synthetic data example as well as a real world case study (Section 4), and finish the paper with a discussion of our findings (Section 5).

2 METHOD

2.1 Bootstrap resampling

Bootstrapping (Efron 1979, 1982) is a resampling method used in statistics to calculate sample estimates by random sampling with

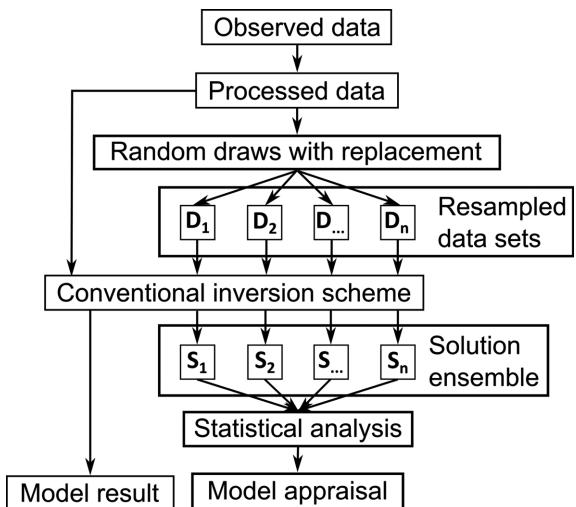


Figure 1. Workflow of the bootstrap resampling approach.

replacement. It is based on resampling a sample population of fixed size, creating multiple realizations, and calculating the statistical property of interest for all realizations, which allows for inference of statistical accuracy measures for the property, such as sample averages and variances. The strength of the bootstrapping method is that it is easy to implement and enables the estimation of any statistic, regardless of its complexity. This comes at the price of a higher computational cost, which nowadays is not much of a problem, as computational power is abundant.

The bootstrapping method was first introduced to the field of geophysics by McLaughlin (1988), who used it to estimate the uncertainty of seismic event magnitudes, and by Tichelaar & Ruff (1989), who used bootstrapping and other resampling techniques in a case study on the calculation of earthquake epicentre depths. Bootstrapping has since been used in seismology and seismic tomography on a regular basis for error and significance estimations for velocity data stacks (Flanagan & Shearer 1998; Gurrola & Minster 1998; Korenaga 2013, 2014), to estimate uncertainties on earthquake locations (Shearer 1997), in the analysis and inversion of velocity data (Koch 1992; Gurrola *et al.* 1994; Sacchi 1998; Choi *et al.* 2010) and to assess shear wave splitting errors (Sandvol & Hearn 1994). Bootstrapping is also used in magnetics, for example, in the analysis of magnetic susceptibility data (Tauxe *et al.* 1998), in error estimation for paleomagnetic data (Tauxe *et al.* 1991; Guyodo & Valet 1996; Shaar *et al.* 2011) and for evaluating geomagnetic dipole intensities (Jonkers 2007). In MT, the bootstrap method has not been used for modelling yet, but has been utilized in the estimation of MT transfer functions (Campanya *et al.* 2014; Chave 2014a; Neukirch & Garcia 2014) and the estimation of error of rotational invariant parameters (Marti *et al.* 2004; Chave 2014b).

Given an N -element data set $X = \{x_1, \dots, x_N\}$, k bootstrap realizations $X'_i, i = \{1, \dots, k\}$ are created by performing N random draws with replacement from X for each bootstrap data set. This means each new data set has the same number of elements as the original set and the bootstrap data sets can contain some elements multiple times and might not contain some of the original elements at all. As mentioned before, the statistic of interest is then evaluated for each bootstrap set producing k solutions, from which solution averages and variances can be calculated.

We apply the bootstrapping principle to 2-D MT data (Fig. 1). In our case, the properties of interest are the cell resistivity values of a

model resulting from a numerical inversion performed on the data sets. To perform these inversions we use *Ocram's* inversion for 2-D MT data (Constable *et al.* 1987; deGroot Hedlin & Constable 1990; see Section 2.2). A data element for the purpose of the bootstrap is equivalent to a data block containing the processed data for a specific evaluation frequency of a specific MT site. Each data block contains four elements: the values for the \log_{10} -scaled apparent resistivity $\rho^* = \log_{10}(\rho)$ and the phase ϕ for each the transverse electric (TE) mode and the transverse magnetic (TM) mode. By treating a whole data block as a contained unit rather than employing the bootstrap on the single data values, we avoid the creation of model artefacts, due to incomplete data for a certain frequency and site. Artefacts caused by the bootstrapping process have to be avoided, as they would be misinterpreted for uncertainty caused by the data.

The resampling process simulates repeated measurements and is executed in two stages. In the first stage a data block is drawn at random from a uniform distribution of the complete set of data blocks. For the second stage an error model with normally distributed data errors is assumed and Monte Carlo samples are created from the data blocks drawn in the first stage. The Monte Carlo samples are created by drawing from a normal distribution $x \sim \mathcal{N}(\mu, \sigma)$, with the value of the element (linear resistivity ρ or phase ϕ) as the expected value μ of the normal distribution and the error assigned to the particular data block element as the standard deviation σ of the distribution. By sampling from these distributions of the values, the created bootstrap sets are equally likely realizations as the original measured data set. These are inverted with the same starting models and the same starting parameters, to then assess the differences and similarities between the bootstrap models, which are indicators of model uncertainty. The starting parameters for each group of bootstrap samples have to be identical to make them comparable and the uncertainty estimates will to some degree be dependent on the choice of starting parameter.

Model areas that show good accordance across the different model results are considered well determined and areas of large variability exhibit elevated uncertainty. In the case of perfect, noise free data, all model results will be equal or at least very similar. Small variations would be caused by the fact that the samples are drawn with replacement and each bootstrap set is different, but even if certain data points are not drawn in a particular bootstrap set, the results would be very similar. This is due to the fact that in MT neighbouring frequencies contain similar information. Thus, most of the potentially missing information is contained in other data blocks. This is caused by the diffusive nature of the EM fields, which results in MT measurements representing averages over the volume of medium penetrated by the EM fields. Hence, Weidelt's dispersion relation (Weidelt 1972):

$$\phi(\omega_0) = \frac{\pi}{4} - \frac{\omega_0}{\pi} \int_{\rho}^{\infty} \log \left(\frac{\rho_a(\omega)}{\rho_0} \right) \frac{d\omega}{\omega^2 - \omega_0^2}, \quad (1)$$

which relates the apparent resistivity ρ_a to the impedance phase ϕ , predicts similar results for neighbouring frequencies $\omega = 2\pi f = 2\pi/T$.

There might be instances where a model includes regions that the data are not sensitive to, for example at great depth or far away from the sites. The resistivity distribution of these regions would greatly depend on the generally subjective choice of the starting model and other starting parameters, and would experience little to no change during the modelling and across all bootstrap realizations. This would lead to these regions being assigned high certainty even though the data do not have any sensitivity in these areas. To avoid this kind of misinterpretation, one should always

analyse to which depth the model is actually supported by the data and does not only consist of artefacts. One method to do this is the depth of investigation (DOI) index developed by Oldenburg & Li (1999). The DOI is assessed by running multiple inversions with different starting models and comparing the differences and similarities (Oldenburg & Li 1999; Marescot *et al.* 2003) to help avoid overinterpretation of the inversion result(s). Nevertheless, the influence from the starting model and other prior information, and therefore the potential for bias, increases with depth even above the depth threshold suggested by the DOI analysis.

2.2 Occam's inversion

Occam's inversion is a regularized smooth inversion algorithm, first employed for 1-D MT data by Constable *et al.* (1987) and later expanded by deGroot Hedlin & Constable (1990) to handle 2-D MT data. We chose this algorithm as it is still commonly used and is an accepted reference.

The algorithm uses a Tikhonov-type regularization (Tikhonov & Arsenin 1977) to regularize the inversion models by applying a penalty on the roughness of the model, seeking the smoothest model under the constraint that it explains the observed data to a degree specified by a desired data misfit Γ_*^2 . Model roughness is measured by the magnitude of the first or second derivative of the model resistivities. Assuming a 2-D structure with strike direction x the first derivative roughness of the M -element model vector \mathbf{m} in the lateral direction y and the vertical direction z is given by:

$$R_1 = ||\partial_y \mathbf{m}||^2 + ||\partial_z \mathbf{m}||^2. \quad (2)$$

By penalizing roughness, the simplest model that can fit the data to the desired misfit is sought. This ensures the principle of Ockham's Razor is followed, which states that the simplest solution to a problem is to be preferred. Thereby, unnecessarily complex models and the formation of model artefacts are avoided. Furthermore, the regularization stabilizes the inversion process.

During the inversion, the algorithm minimizes a constrained objective functional U that contains the model roughness term and a model misfit term:

$$U[\mathbf{m}] = \underbrace{||\partial_y \mathbf{m}||^2 + ||\partial_z \mathbf{m}||^2}_{\text{model roughness}} + \mu^{-1} \underbrace{[||\mathbf{W}(\mathbf{d} - F(\mathbf{m}))||^2 - \Gamma_*^2]}_{\text{model misfit}}, \quad (3)$$

where \mathbf{d} is the N -element data vector, F is the non-linear forward operator (Wannamaker *et al.* 1987), μ^{-1} is the Lagrange multiplier, and \mathbf{W} is the diagonal covariance weighting matrix, containing values inversely proportional to the standard deviation of the data errors.

The Lagrange multiplier μ^{-1} acts as a weighting parameter between the model roughness and the model misfit and is chosen automatically at each iteration to achieve the desired misfit.

If the desired misfit is not reached after the maximum number of iterations, it is common practice to rerun the inversion for a small amount of iterations using the last model result of the run as the starting model and using the achieved misfit increased by 5–10 per cent as the new desired misfit. This allows the algorithm to smooth out the model to find the simplest viable model and reduce model artefacts.

3 ENSEMBLE APPRAISAL

After an ensemble of model solutions has been created from a set of resampled data, the variations between these model solutions have

to be analysed to create uncertainty maps for the related ‘master-model’, which is based on the original data set.

We quantify uncertainty as the differences in the absolute cell value (resistivity) and the variations of the cell gradients across all models in the ensemble.

3.1 Variability in absolute value

Resistivities are usually considered a logarithmic parameter, as they generally show large variations. Hence, all calculations are performed with logarithmically scaled resistivities $\rho_i^*(y, z) = \log_{10}(\rho_i(y, z))$. The variation in absolute cell values are determined for each model cell by calculating the deviations of the cell resistivities $\rho_i^*(y, z)$ from the cell average $\bar{\rho}(y, z)$ over all k models of the ensemble.

As each bootstrap data set is different, the resulting inversion models will most likely have slightly differing misfit values. To avoid biasing the results we weight each solution with its misfit δ to calculate a weighted mean for each cell:

$$\bar{\rho}(y, z) = \frac{\sum_{i=1}^k \delta_i \cdot \rho_i^*(y, z)}{\sum_{i=1}^k \delta_i}. \quad (4)$$

The variance of the value of a specific model cell $\sigma_\rho^2(y, z)$, and thereby the standard deviation $\sigma_\rho(y, z)$, is given by

$$\sigma_\rho^2(y, z) = \frac{\left(\sum_{i=1}^k \delta_i \cdot (\rho_i^*(y, z) - \bar{\rho}(y, z))^2 \right) \cdot \sum_{i=1}^k \delta_i}{\left(\sum_{i=1}^k \delta_i \right)^2 - \sum_{i=1}^k \delta_i^2}. \quad (5)$$

The standard deviation $\sigma_\rho(y, z)$ of the resistivity values $\rho_i^*(y, z)$ represents the variability of the absolute cell value across the ensemble. To allow for a comparison of the standard deviation of different cells, the relative standard deviation $\sigma_{\rho,\text{rel}}(y, z) = \sigma_\rho(y, z)/\bar{\rho}(y, z)$ is displayed in the final uncertainty map, rather than the absolute standard deviation.

3.2 Variability in cell gradient

By only evaluating the variability of the absolute cell values we cannot distinguish between actual differences in magnitude and differences in structure. There might arise situations where the absolute cell values show high variability, but the different models show good structural agreement, which would indicate uncertainty in the absolute resistivity of anomalies but nevertheless robustness in the position and shape of structures.

To evaluate the structural similarity between the ensemble members, we assess the variability in the cell gradients. The model parameter vector \mathbf{m} , containing the \log_{10} -scaled resistivities, are analysed. For the 2-D case $\mathbf{m}(y, z)$ the gradient of the cell in the i th row and the j th column of the model mesh is defined as:

$$\nabla m_{i,j}(y, z) = \frac{\partial m_{i,j}}{\partial y} \hat{e}_y + \frac{\partial m_{i,j}}{\partial z} \hat{e}_z. \quad (6)$$

The derivatives in eq. (6) are estimated by first differences

$$\frac{\partial m_{i,j}}{\partial y} \approx \frac{1}{2} \left(\frac{m_{i,j} - m_{i,j-1}}{\Delta y_1} + \frac{m_{i,j+1} - m_{i,j}}{\Delta y_2} \right) \quad (7a)$$

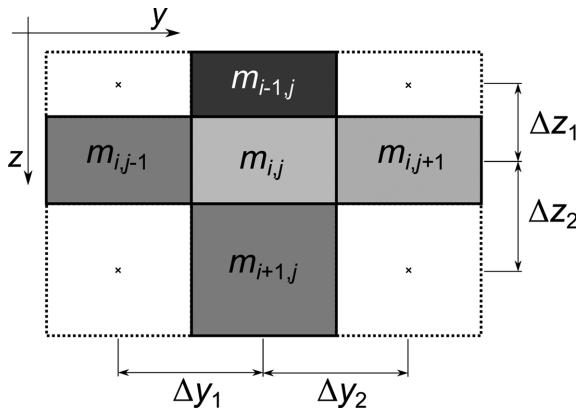


Figure 2. Definition of nearest neighbours for cell gradient calculation. Only adjacent cells are considered.

$$\frac{\partial m_{i,j}}{\partial z} \approx \frac{1}{2} \left(\frac{m_{i,j} - m_{i-1,j}}{\Delta z_1} + \frac{m_{i+1,j} - m_{i,j}}{\Delta z_2} \right). \quad (7b)$$

The notation is illustrated in Fig. 2, which displays the five-cell scheme that is used for the calculation of the first differences.

The cell gradients of two models are compared by calculating the cell cross-gradients $\mathbf{cg}(y, z)$ (cross product of the gradients)

$$\mathbf{cg}(y, z) = \nabla m_{i,j}^\alpha(y, z) \times \nabla m_{i,j}^\beta(y, z) \quad (8)$$

between the two models α and β , which becomes large if the difference in gradient direction is large and small when the difference in gradient direction is small.

Using the dot product, the angle of aperture of the two gradients θ can be calculated:

$$\begin{aligned} \nabla m_{i,j}^\alpha \cdot \nabla m_{i,j}^\beta &= \|\nabla m_{i,j}^\alpha\| \|\nabla m_{i,j}^\beta\| \cos \theta \\ \Rightarrow \theta &= \arccos \left(\frac{\nabla m_{i,j}^\alpha \cdot \nabla m_{i,j}^\beta}{\|\nabla m_{i,j}^\alpha\| \|\nabla m_{i,j}^\beta\|} \right). \end{aligned} \quad (9)$$

These angle differences are calculated for all possible pair combinations of the models in the solution ensemble and the average for each model cell is displayed as an uncertainty map. With k models there are

$$\binom{k}{2} = \frac{k!}{2!(k-2)!} = \frac{1}{2} \cdot k \cdot (k-1) \quad (10)$$

unique pair combinations.

4 RESULTS

To demonstrate our bootstrapping approach, we employ it on a synthetic model case, as well as real world data.

The number of bootstrap samples that can be used in the tests is mostly limited by the computing power available. One should evaluate as many samples as is reasonable, as even though more samples cannot increase the amount of information in the original data, more samples increase the reliability of the statistic calculated and reduce possible bias. Efron & Tibshirani (1986) performed tests on the ideal parameters for bootstrapping and came to the conclusion that $k \in [50, 200]$ bootstrap realizations are enough to yield meaningful results in most cases. They state, that as few as $k = 25$ realizations give acceptable results and that there is not much improvement observable for $k > 100$. During our testing we found that as little as 20 bootstrap sets are enough to yield qualitative

useful results, but we use 100 bootstrap samples in each of our tests, to ensure robust results.

4.1 Synthetic test case

For our synthetic test case we used a model similar to the synthetic models from Kalscheuer *et al.* (2010) and Rosas-Carbajal *et al.* (2014). It includes a high-resistivity structure covering a low-resistivity structure, as well as the opposite setup. The model is simple enough to serve as a controllable test environment, but is complex enough to predict the results of the bootstrapping techniques for a variety of modelling situations.

The model has a $100 \Omega\text{m}$ background resistivity (Fig. 3a). On the left side of the model a resistive block R_l ($1000 \Omega\text{m}$) is positioned 3.35 km above a conductive block C_l ($10 \Omega\text{m}$) and on the right side of the model a conductive block C_r ($10 \Omega\text{m}$) is positioned above a resistive block R_r ($1000 \Omega\text{m}$). The blocks on both sides have the same dimensions, with the top blocks starting at a depth of 3.1 km and being 13.3 km wide and 2.9 km high, and the bottom blocks starting at a depth of 9.35 km and being 8 km wide and 2.3 km high. The top blocks are 8 km apart and the bottom blocks are 13.3 km apart.

Forward data for 51 sites at a 1 km spacing for 42 frequencies in the range of 0.012 – 156.25 Hz were calculated with the finite element 2-D forward code by Wannamaker *et al.* (1987). This resulted in 2142 data blocks with apparent resistivities and phases for TE and TM mode. White noise with a standard deviation equivalent to 5 per cent of the linearly scaled resistivity amplitudes was added to the forward data. Under the standard assumption that the errors of the complex impedances are equal in the real and imaginary components the error of the phase σ_ϕ is related to the error of the resistivity σ_ρ by

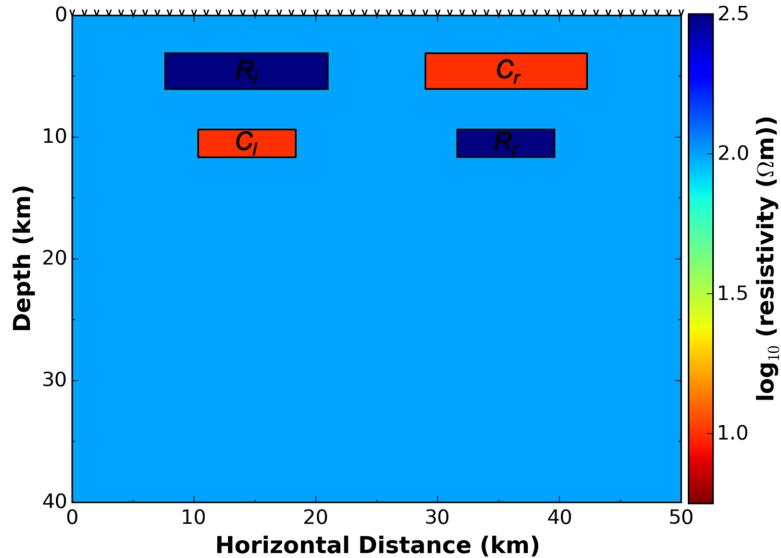
$$\sigma_\phi = \arcsin \left(\frac{1}{2} \frac{\sigma_\rho}{\rho} \right). \quad (11)$$

Accordingly, white noise with a standard deviation of $\sigma_\phi = \arcsin(\frac{1}{2} \cdot 0.05) \approx 1.43^\circ$ was added to the phase values. This is commonly (misleadingly) referred to as ‘5 per cent’ phase errors.

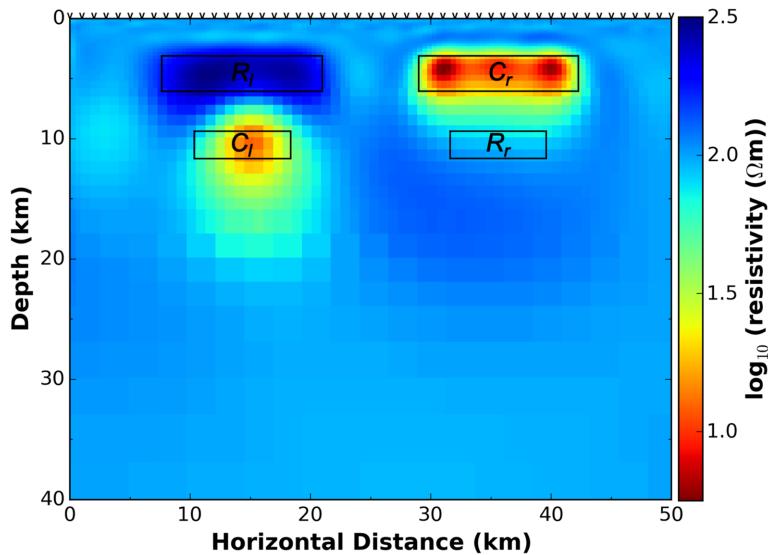
4.1.1 Conventional inversion

The complete data set was inverted with the Occam 2-D code (deGroot Hedlin & Constable 1990), described in Section 2.2, using a $100 \Omega\text{m}$ half-space as the starting model (Fig. 3b). A mesh different from the synthetic forward mesh was used for all inversions to avoid biasing the results.

The inversion recovers structures R_l and C_r reasonably well. Both structures are smeared out, especially in the vertical direction. In case of the feature corresponding to structure R_l the smearing is most significant extending downwards from the left and right lower corners of the structure. The algorithm compensates for this by adjusting the resistivity of the features. The resistivity of the block R_l is underestimated by at least $685 \Omega\text{m}$ (69 per cent) at all corresponding cells, and the ends of structure C_r (horizontal position 30 – 32 km and 39 – 41 km) underestimate the true model by up to $5 \Omega\text{m}$ (50 per cent). There is very little indication of structure R_r , as it is shielded by the low-resistivity structure C_r . Structure C_l is also not well recovered. The centre of structure C_l is placed at roughly the correct depth, but the structure is greatly smeared out in the vertical direction and is too narrow in the horizontal direction. In addition to these shortcomings the model also shows a high number



(a) True synthetic model. The bodies R_l and R_r have a resistivity of 1000 Ωm , the bodies C_l and C_r have a resistivity of 10 Ωm , and the background has a resistivity of 100 Ωm . Scale has been limited to $10^{2.5}$ Ωm = 316 Ωm for easier comparison with the inversion results.



(b) Master inversion.

Figure 3. (a) True synthetic model and (b) the inversion result based on the data set forward calculated from the true model. Black 'v' mark the locations of the 51 simulated recording sites. The true locations of the four anomalous bodies R_l , C_l , C_r and R_r are marked with black boxes.

of small scale artefacts in the top 3 km which are most likely caused by the added noise.

An analysis of the DOI (Oldenburg & Li 1999) (not shown) confirms that all regions of the model are supported by the data.

4.1.2 Bootstrap inversions

To perform the bootstrap analysis the original data set was resampled as described in Section 2.1 and 100 bootstrap representations of the data were created. All bootstrap data sets were inverted using the same model mesh and the same starting model as the master-inversion. A maximum of 30 iterations per model were performed after which the target misfits of each individual model was adjusted to allow Occam to find the smoothest solution, and each model was run for a maximum of 10 additional smoothing iterations.

The average achieved misfit was $\bar{\delta} = 1.41$ (after adding 5 per cent smoothing buffer) and the final models had an average roughness of 15.6.

The resulting solution ensemble was analysed using eqs (4), (5) and (9), to calculate the uncertainty estimation maps (Figs 4a–f).

The average model (Fig. 4a) is similar to the master inversion, but exhibits some differences. These differences become evident in the map of the relative residuals between the master result and the average bootstrap result (Fig. 4b). The use of the relative residual as opposed to the absolute residual ensures comparability between the values for all cells.

Both models are generally very similar from a structural standpoint. The representation of feature C_r in the average model shows less extreme values than in the master inversion, but the true resistivity is still underestimated. Comparing other model regions of the

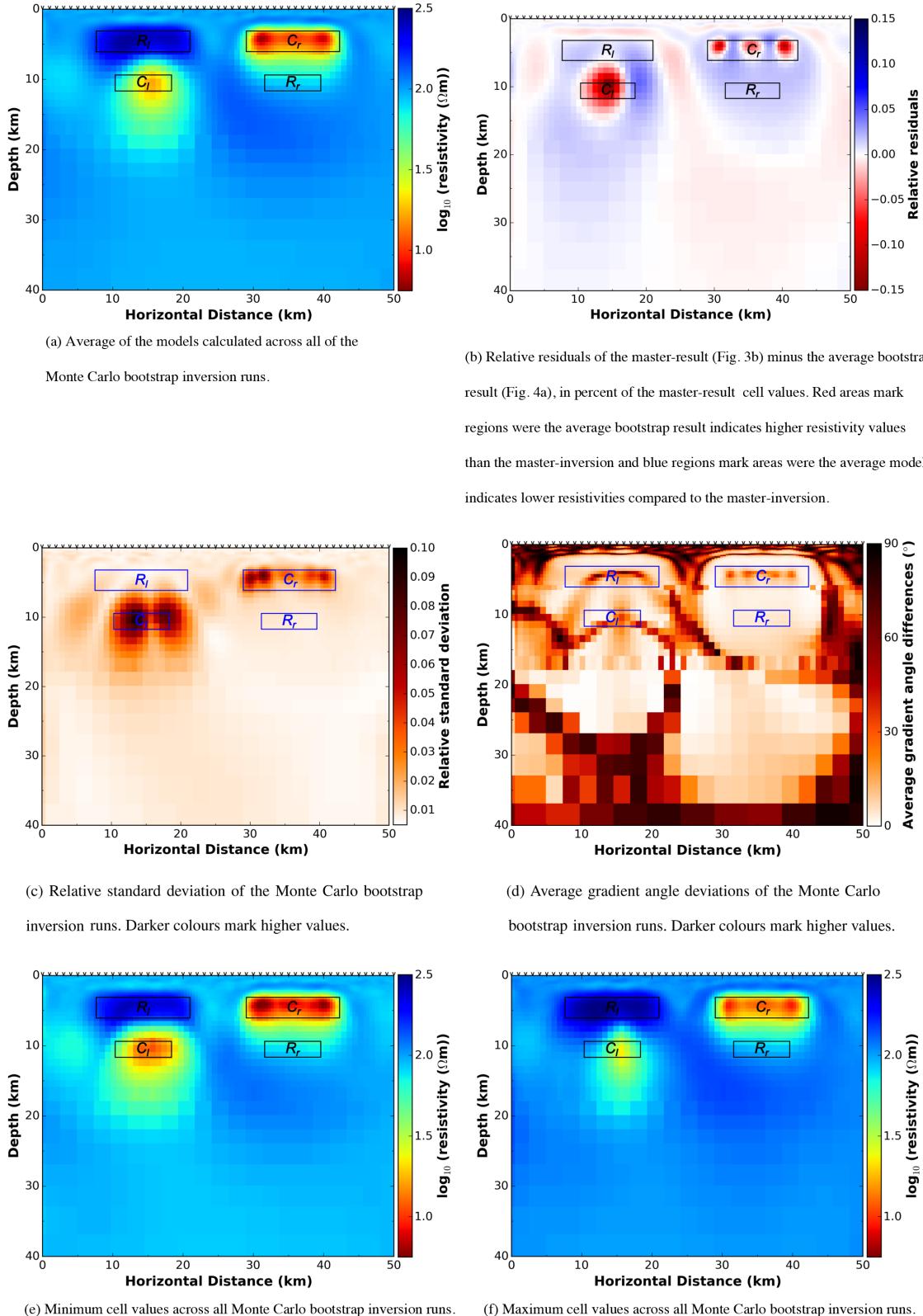


Figure 4. Monte Carlo bootstrap results for the synthetic test case data set. Black ‘v’ mark the locations of the 51 simulated recording sites. The true locations of the four anomalous bodies R_l , C_l , C_r and R_r are marked with black/blue boxes.

bootstrap average to the true model, it is also clear that the resistivity of structure R_l is still underestimated and the resistivity of structure C_l is overestimated. Furthermore, there is still no significant indication of structure R_r . This is not surprising as the bootstrapping

models do not include or add any new information and the structure is effectively shielded by the low-resistivity feature above.

The relative standard deviation shown in Fig. 4(c) is highest across the recovered low-resistivity feature corresponding to

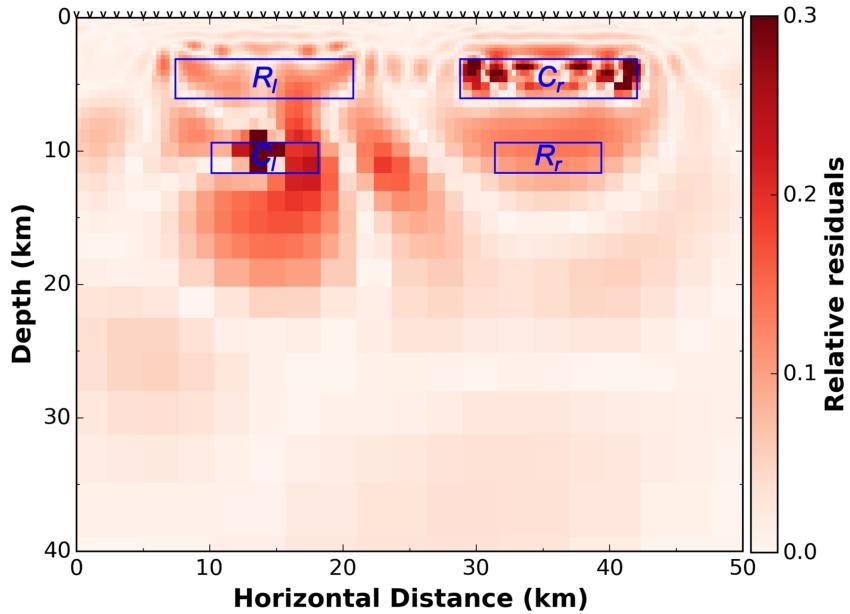


Figure 5. Relative residuals of a low-noise comparison run minus the master-result, in per cent of the low-noise cell values (absolute values of the relative residuals shown for easier comparison with relative standard deviation map Fig. 4c). Black 'v' mark the locations of the 51 simulated recording sites. The true locations of the four anomalous bodies R_l , C_l , C_r and R_r are marked with blue boxes.

structure C_r and around the feature correlating to structure C_l . These regions coincide with the lacking horizontal extension of the feature corresponding to C_l and the smearing effects around it, and the regions of resistivity underestimation of structure C_r . Further areas of elevated standard deviation are extending downwards from the left and right lower corners of structure R_l , as well as various small scale artefacts in the top 3 km of the model. This indicates that these regions show the largest variations in resistivity cell values across the 100 bootstrap solutions and therefore, exhibit the largest uncertainty. The higher variability associated with conductive features rather than resistive features is caused by the higher sensitivity of the MT method to conductive features (Chave & Jones 2012).

The deviations in gradient direction (Fig. 4d) indicate variation in position and shape of structures. Large variations are indicated in the top 3 km of the model along the whole profile. Furthermore, variations are shown at the top of the recovered features corresponding to structures R_l and C_r , and in the middle of these features, running along the long axis, indicating the point where the gradient direction reverses over the structures. In addition to the line of gradient reversal, elevated levels of gradient deviations are indicated for the area extending down from the right corner of structure R_l and for an area extending down from a point about 2 km to the right of the lower left corner of the same structure. For structure C_l the line of gradient direction reversal is less well defined. The model gradients also vary significantly at depths greater than 20 km, which is most likely caused by the lack of structure in those regions, so that even small differences result in large changes in gradient direction.

Some additional information about the range of possible cell values is given by the analysis of the minimum values across all bootstrap models for each cell (Fig. 4e) and the maximum values respectively (Fig. 4f). These maps display the extreme values for each individual cell and do not represent viable models and can therefore not be expected to achieve misfits comparable to the master inversion. In our synthetic test case the range between minimum and maximum cell values is small. Structurally, the two plots do not differ greatly from each other or from the master and average results, and also give no real indication of the shielded structure R_r .

As the bootstrap approach is mainly sensitive to model errors caused by data noise, we performed an inversion on a control data set that has almost no noise added to it (a small noise component is added for stability purposes). By comparing the inversion result of the low noise data set with the master inversion and calculating the residuals between the two (Fig. 5), we can analyse which model errors were likely caused by the added noise, and not by the inversion algorithm or other influences the bootstrap method is not sensitive to. Viewing the residual map in Fig. 5, it is apparent that the biggest differences caused by the added noise are in the areas of structure C_r and around it, structure R_l , and structure C_l , as well as under it. In addition, a variety of small scale artefacts are observable in the top 3 km of the model. Comparing the residuals with the map of the relative standard deviation, it is obvious that most of these areas are indicated as having elevated variability. The only exception is structure R_l , where the standard deviation shows little indication for increased uncertainty.

Assuming no knowledge of the true model, the combined interpretation of the calculated average model, the standard deviation and the average gradient direction differences would indicate that the various small scale features in the top 3 km are artefacts caused by data noise and that the amplitude of the resistivity values of the features correlated to the true structures C_r and C_l are subject to elevated uncertainty. Furthermore, the standard deviation map indicates that the lateral extension of the feature belonging to C_l is unsure. Most other parts of the model, especially at depths of 20 km and greater, show low relative standard deviation and are rated as well determined. It has to be noted that the maximum skin depth of this set of MT data is about 50 km and there may be model regions that are not fully constrained by the data even before this depth.

4.2 Delamerian MT transect data

To demonstrate the application to real data we conducted a case study on a data set from the Delamerian MT transect (Fig. 6). This data set consists of data from 68 broad-band MT sites, with a

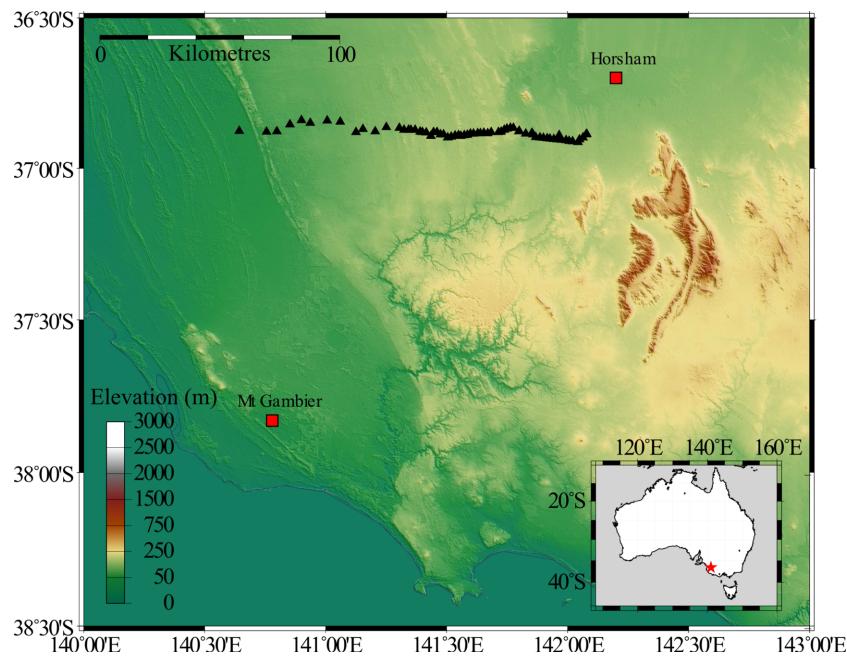


Figure 6. Map of the station locations of the Delamerian MT transect (\blacktriangle).

sampling frequency of 1000 Hz, collected in 2010 and 2012 along the L193 AuScope 09GA-SD1 southern Delamerian seismic line. For the 14 western-most sites the site spacing averages to about 5 km, and the 54 sites on the eastern side of the profile have a site spacing of about 1.5 km. The data cover a frequency range of 0.012–156.25 Hz, and have been corrected for a geoelectric strike of 49° west of north. The data were notch filtered for 50 Hz power line noise (including harmonics). Data blocks with phase tensor skew values of $-5^\circ > \beta > 5^\circ$ (Caldwell *et al.* 2004) were removed from the data set for violating the 2-D assumption. After the omission of these data points a total of 1667 data blocks remained. For a full geological interpretation see Robertson *et al.* (2015).

4.2.1 Conventional inversion

As with the synthetic data, the complete Delamerian MT transect data were inverted with the Occam 2-D code (deGroot Hedlin & Constable 1990) to create the master-inversion (Fig. 7). A $100 \Omega\text{m}$ homogeneous half-space was used as the starting model. We used the same error floors as Robertson *et al.* (2015), they were set to 10 per cent for the phase and 20 per cent for the apparent resistivities. Following the definition in Section 4.1 this corresponds to a phase error of $\sigma_\phi \approx 2.87^\circ$. By setting the error floor of the phase lower than the error floor of the resistivities, the effects of possible static shift on the model outcome are reduced. The inversion was run for 50 iterations, plus an additional ten smoothing iterations and achieved a misfit of $\delta = 1.93$ (including 5 per cent smoothing buffer) and a roughness value of 217.3. These values are comparable to the values of $\delta = 1.95$ and a roughness of 197 reached by Robertson *et al.* (2015). The small differences are likely due to a slightly higher smoothing buffer used in Robertson *et al.* (2015), which has a negligible influence on the model result, as demonstrated by the fact that the model from Robertson *et al.* (2015) is virtually identical to the master-result generated here.

The master-model covers a wide range of resistivity values, spanning 1–10 000 Ωm , and the main features of the master model are

three low-resistivity structures C_1 , C_2 and C_3 , as well as four high-resistivity features R_1 , R_2 , R_3 and R_4 (Fig. 7). Anomalies C_2 and C_3 range from 1 to $100 \Omega\text{m}$ in resistivity, and anomaly C_1 has a minimum resistivity of about $37 \Omega\text{m}$. Anomaly C_1 is located at the very eastern edge of the model, connected to the model's padding cells. The data density for that area of the profile is sparse and the feature should be interpreted carefully, as it might be a boundary effect.

In addition to the main features the model shows low-resistivity structures F_1 , F_2 and F_3 , marked by arrows in Fig. 7, which were interpreted as faults serving as fluid pathways by Robertson *et al.* (2015).

4.2.2 Bootstrap inversions

As with the synthetic test case, 100 resampled data sets were drawn from the original data set to perform the bootstrap analysis, using the same model mesh and starting model as the master inversion. As with the master-inversion a maximum of 50 iterations were performed, followed by a maximum of ten smoothing iterations. The bootstrap inversion achieved misfits ranging from $\delta = 2.1$ to $\delta = 2.4$ with an average misfit of $\bar{\delta} = 2.2$ (after adding 5 per cent smoothing buffer). The average model roughness achieved was 186.4.

The average bootstrap model (Fig. 8a) shows almost all of the same main features as the master inversion, but differs in details. The average model shows a more homogeneous lower crust, that exhibits lower resistivities as is evident from the residuals between the master model and the average model (Fig. 8b). The high-resistivity feature R_4 , which can be observed in the master inversion is not evident in the average model. Compared to the master result, features C_2 and C_3 show a smoother resistivity distribution and the boundary structure C_1 exhibits a higher resistivity closer to the background. Structures R_1 and R_3 extend less deep than in the master result, which coincides with the overall reduced resistivity of the lower crust. The interpreted fault structures F_1 and F_2 are clearly evident and still show a connection going through to the surface. Most

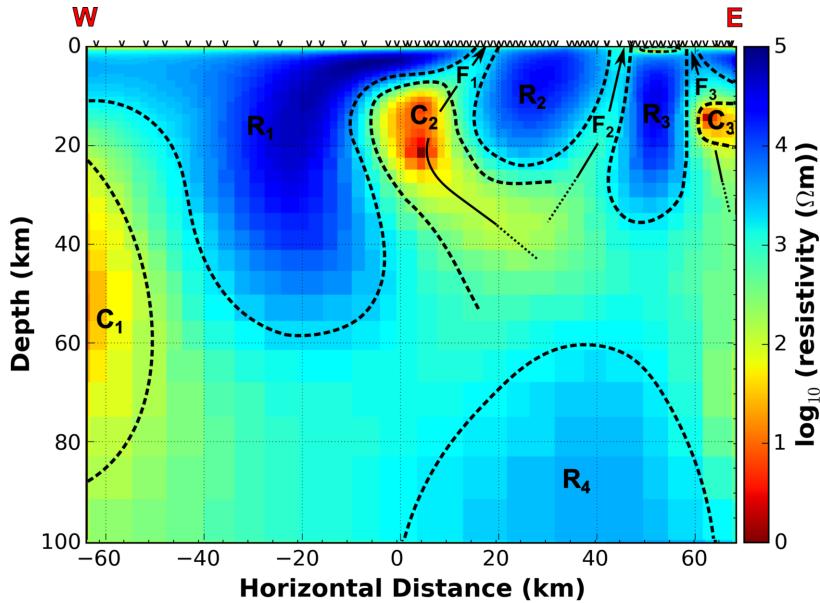


Figure 7. Master-inversion (all data points) of Delamerian MT Transect data set. Black 'v' mark the locations of the 68 broad-band MT recording sites. The interpreted locations of the model features R_1 – R_4 and C_1 – C_3 are marked by black dashed lines. The interpreted fault structures F_1 – F_3 and their shape are marked with black arrows.

of structure F_3 is still evident in the average model, but feature R_3 seems to connect to the right boundary of the model, and the connection of F_3 to the surface is no longer clearly evident.

The map of the relative standard deviation across the bootstrap models (Fig. 8c) shows values of more than 50 per cent in some regions of the model. Elevated variability is indicated for the centres of C_2 and C_3 . Other regions of elevated variability are the area around C_1 and the area under C_2 , extending downwards from the tip of R_1 .

From the display of the average gradient angle differences (Fig. 8d) we can infer the stability of position and extension of structures. Down to a depth of about 10 km the resistive structures R_1 , R_2 and R_3 show well defined and localized lines of gradient direction reversal, indicating that the position of these features is well defined in those depths. Below 10 km the points of gradient reversal are less well defined indicating a degree of uncertainty in the position of these structures. The line of gradient direction reversal of R_3 cuts across the indicated fault F_3 and extends all the way to the boundary of the model. A similar behaviour can be observed for F_1 , which is intersected by the line of gradient direction reversal of R_1 , but in this case F_1 is not cut off completely. Apart from that, F_1 shows low levels of gradient angle differences, indicating its position is well determined. This is not the case for F_2 , which shows elevated gradient angle differences, denoting elevated uncertainty in the position of this fault. Other areas of significant gradient angle differences are structure C_1 , the entire western side of the model from a depth of about 70 km down, and the whole eastern side of the model from a depth of about 40 km downwards. This eastern area of elevated gradient angle differences also extends along C_2 , marking a poorly defined line of gradient direction reversal for this feature, and along the right flank of the resistive structure R_1 , extending up to a depth of about 20 km. In contrast, the position of the lower edge and the left flank of R_1 and the lower edges of R_2 and R_3 seem well defined.

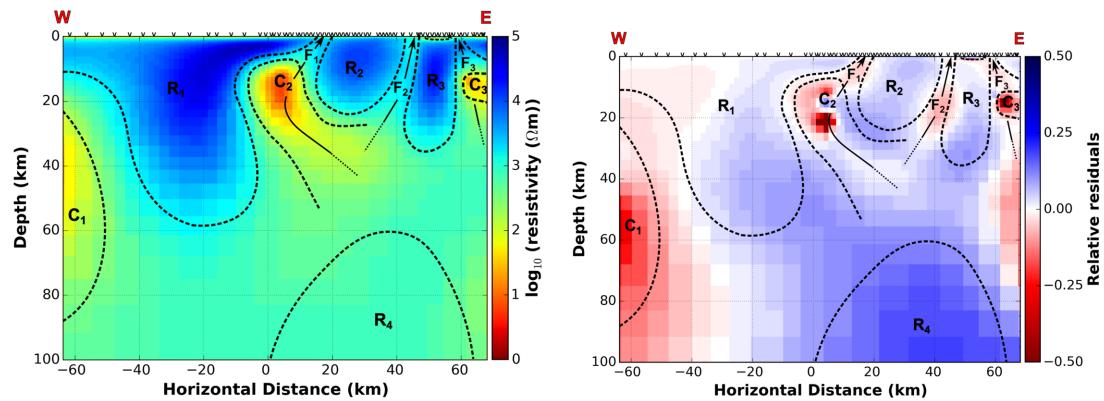
The high uncertainty in the position and extent of R_1 and C_2 is also evident in the display of the lowest and highest cell values reached across all bootstrap models (Figs 8e and f). The two maps show

large differences. Most main features are still evident in both maps, except for boundary feature C_1 , which is no longer distinguishable from the background in the display of the maximum reached resistivity values. The fault structures are still visible in both maps, except for the fact that the fault feature F_3 is not connected to the surface in the maximum value display. Apart from these features, the biggest differences evident between the minimum and maximum value displays are the areas between and under R_1 and C_2 . In the minimum resistivities map, the main body of R_1 extends to a depth of only 20–30 km and a large low-resistivity region is present under R_1 and C_2 that extends from 25 km downwards. On the maximum resistivities map R_1 is notably wider and extends into the area of the western side of C_2 .

Taking into account the additional information from the bootstrap analysis, the master model can be reinterpreted (Fig. 9). The bootstrap results show indications that the existence of features C_1 and R_4 is very uncertain and there is a high probability that they are artefacts. Most other features show a high degree of robustness. Only the extension of fault structure F_3 to the surface is subject to high uncertainty. The area between and under R_1 and C_2 is also not well determined, especially the transition between the two features and the downwards extension of R_1 are uncertain, in position as well as absolute resistivity value. Furthermore, features C_2 and C_3 are likely to have a smoother resistivity distribution than indicated by the master model, with less extreme values towards low resistivities.

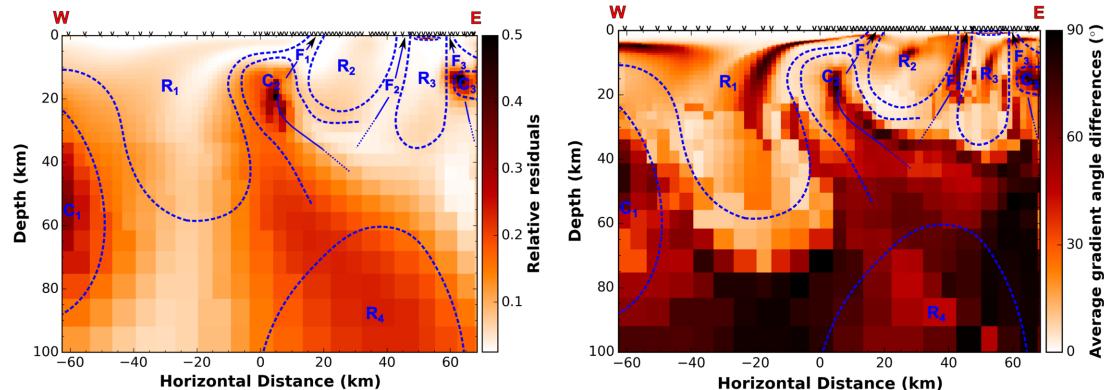
5 DISCUSSION

Our bootstrapping method for the appraisal of inversion models yields estimates for a variety of different parameters. The standard bootstrapping result-package consists of the calculated average model, the residuals between average model and master result, the relative standard deviation, the gradient angle differences, and the minimum and maximum cell values. These should all be interpreted together, when analysing the results of the bootstrap process to ensure that the correct conclusions are drawn.



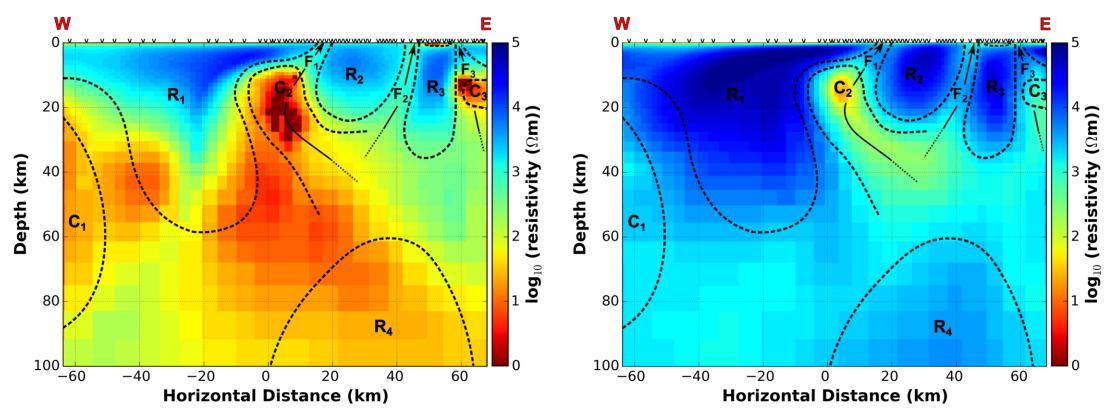
(a) Average of the models calculated across all of the Monte Carlo bootstrap inversion runs.

(b) Relative residuals of the master-result (Fig. 7) minus the average bootstrap result (Fig. 8a), in percent of the master-result cell values. Red areas mark regions where the average bootstrap result indicates higher resistivity values than the master-inversion and blue regions mark areas where the average model indicates lower resistivities compared to the master-inversion.



(c) Relative standard deviation of the Monte Carlo bootstrap inversion runs. Darker colours mark higher values.

(d) Average gradient angle deviations of the Monte Carlo bootstrap inversion runs. Darker colours mark higher values.



(e) Minimum cell values across all Monte Carlo bootstrap inversion runs. (f) Maximum cell values across all Monte Carlo bootstrap inversion runs

Figure 8. Monte Carlo bootstrap results for the Delamerian MT transect data set, derived from inversions performed on 100 bootstrap realizations of the original data set. Black 'v' mark the locations of the 68 broad-band MT recording sites. The interpreted locations of the model features R_1-R_4 and C_1-C_3 are marked by black/blue dashed lines. The interpreted fault structures F_1-F_3 and their shapes are marked with black/blue arrows.

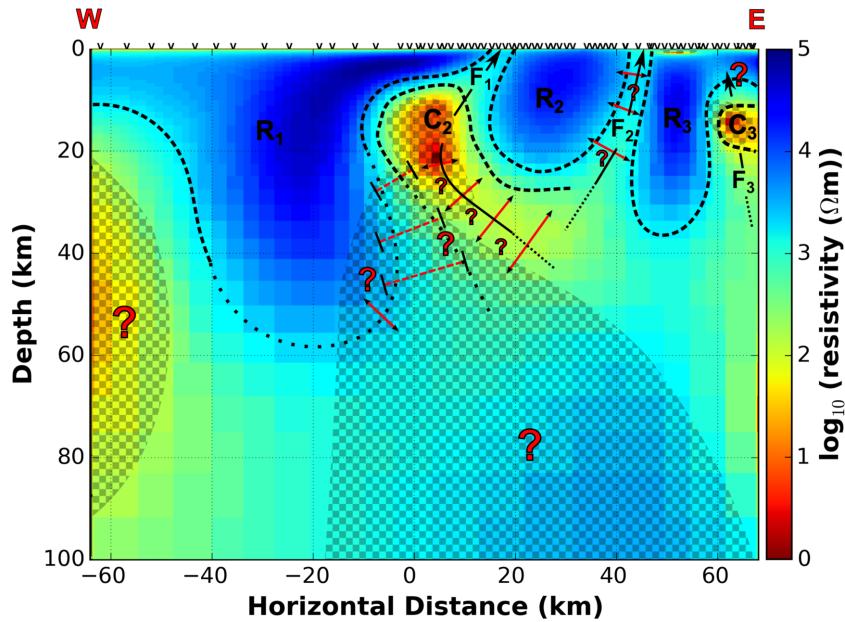


Figure 9. Reinterpretation of the master-model for the Delamerian MT transect data, taking the bootstrap results into consideration. Black ‘v’ mark the locations of the 68 broad-band MT recording sites. The interpreted locations of the model features R_1 – R_3 and C_2 – C_3 are marked by black dashed lines. Dotted borders mark parts of feature edges that are not well determined. The interpreted fault structures F_1 – F_3 and their shape are marked with black arrows. Red arrows ($<\!\!-\!\!>$) mark the uncertainty in the location of a feature and red bars ($\mid\!\!-\!\!-$) mark the uncertainty of the width of a model feature or area. Areas overlain with a checker pattern have high uncertainty in the amplitude of the cell resistivity values and red question marks (?) are general markers for areas of uncertainty.

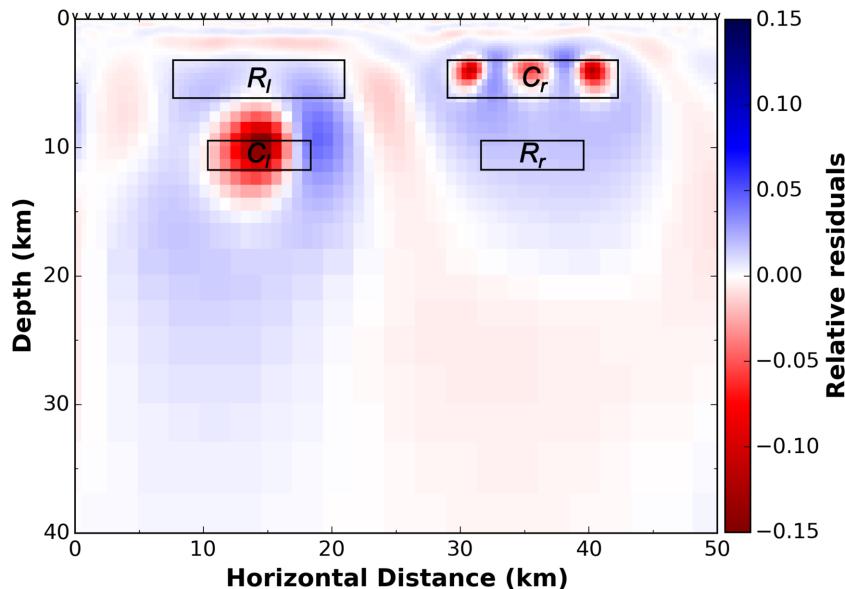


Figure 10. Differences between the residuals of the bootstrap average and the true model and the residuals of the master-model with the true model (all based on the synthetic forward mesh (*cf.* Section 4.1)). Blue colours mark areas where the average bootstrap solution is closer to the true model than the master-inversion, red colours mark areas where the master result is closer to the true model. Black ‘v’ mark the locations of the 51 simulated recording sites. The true locations of the four anomalous bodies R_l , C_l , C_r and R_r are marked with black boxes.

Tests with a synthetic model show that the average bootstrap model exhibits less artefacts compared to the master-inversion, and that it is closer to the true model regarding the actual cell values for most parts of the model, as indicated by the blue regions in Fig. 10, which compares the residuals between true model and a master-solution to the residuals between the true model and the average bootstrap model for the synthetic case (for easy comparison all models based on the synthetic forward mesh). However, the average

model is not an improvement over the master-result in all regions of the model, as indicated by the red model regions. The average model can be used to infer the quality of the master-inversion with regards to location and gradient distribution of structures, but has to be treated carefully when assessing the magnitude of the model parameters as it is not unambiguous.

We interpret the average model as being a more realistic representation. This is also apparent for the real world example, as

the average model shows less artefacts and less model cells with unrealistically extreme parameter values.

The most important indicators for the model uncertainty are the relative standard deviation and the average gradient angle differences maps, which allow for the identification of model areas with elevated variability. The standard deviation maps are especially sensitive to variations in absolute resistivity values, whereas the gradient angle difference maps indicate variations in structure shape and position; particularly by indicating the points of gradient direction reversal. Localized and well defined lines of gradient reversal are an indication for little variation, and therefore little uncertainty in the position of a structure. They are especially useful in the assessment of fault structures, as demonstrated in the analysis of the Delamerian data set.

Zones of low gradient angle differences around structures that are common to most model representations mark the minimum extend of a feature. The gradient angle differences have to be interpreted with care if they occur in large regions of uniform resistivity distribution, especially if they coincide with low standard deviation values of the cell parameters: areas with no structure cause even small cell value differences to result in large changes in gradient direction, which could be mistaken for high uncertainty. A weighting of the angle differences by the average gradient amplitude can correct for this, but would greatly complicate the interpretation.

The magnitude of the standard deviation does not necessarily indicate the magnitude of the error of the model cell value, but the areas of increased standard deviation are an indicator for the location of areas with increased uncertainty in the amplitude of the cell value. A DOI study should be performed before interpreting the bootstrapping results to avoid wrongly identifying model regions, that the data is not sensitive to, as well determined.

In addition to the uncertainty estimates, valuable extra information is provided by mapping the extremal values of each model cell. The bootstrap method supplies only very limited information about the non-uniqueness of the model, but these plots give an important insight into the range of values that can be reached and can indicate the minimal/maximal feasible extent of an anomaly. The value of these plots is very evident in the analysis of the Delamerian MT transect data, as they highlighted areas of high uncertainty in the lower crust and give an indication of the variability of the location and extent of major anomalies. It is important to note that these maps represent the most extreme values reached by each individual cell across all bootstrap realizations. These maps do not represent feasible models and would not fit the data to the same level as the optimal model. Extremal models, which fit the data to a similar level as the optimal solution, are for example provided by the most-squares method (Jackson 1976; Meju & Hutton 1992; Kalscheuer *et al.* 2010; Rosas-Carbajal *et al.* 2014).

Both MCMC methods (Metropolis & Ulam 1949; Gamerman 1997; Rosas-Carbajal *et al.* 2014) and bootstrapping create solutions ensembles and assess posterior distributions. MCMC methods comprehensively sample the solution space and, hence, can yield quantitative uncertainties based on Bayesian statistics. On the other hand, the bootstrap approach samples the solution space less exhaustively and provides only qualitative uncertainty estimates, but is computationally more efficient. The easier implementation of the bootstrap method is also an advantage over both the MCM analysis (Meju 1994; Alumbaugh & Newman 2000; Tarantola 2005; Menke 2012) and the model resolution matrix analysis (Jackson 1972; Meju 1994; Ramirez *et al.* 1995; Alumbaugh & Newman 2000; Menke 2012).

The commonly used manual model perturbation technique (Becken *et al.* 2008; Thiel & Heinson 2010) is attractive in its simplicity, but only investigates how model perturbations influence the model response, which is also a drawback of the analysis of the sensitivity matrix (Thiel & Heinson 2010), and the null-space projection method (Deal & Nolet 1996; Rowbotham & Pratt 1997; Muñoz & Rath 2006). Bootstrap resampling requires the calculation of more models, but is otherwise comparatively simple to implement as the model perturbation method, and has the advantage of providing direct information about how data perturbations map into the model.

In this paper we have presented conclusions from bootstrap runs with 100 resampled data sets each, but tests show that as few as 20 bootstrap models are enough to draw preliminary conclusions. We recommend the use of at least 30–50 model to achieve robust results.

In various forms the bootstrap approach makes a qualitative statement about the uncertainty of a model. This is valuable information, as the only measure of model quality provided by commonly used algorithms like Occam (deGroot Hedlin & Constable 1990) is usually a single scale value. Furthermore, the bootstrap method can provide this information without requiring access to the inversion algorithm's source code and without additional code development.

The exact representations of the uncertainties will vary with the employed inversion algorithm, but the qualitative result should remain the same. Nevertheless, this has to be verified with synthetic tests before the approach is used on real data with a different algorithm. We used the approach with the Occam 2-D inversion code as it is widely used and an accepted reference. The bootstrapping approach does not require any additional assumptions. However, the results are influenced by the assumptions required by the inversion algorithm. For example, tests showed that excessive smoothing greatly reduces the variability in the solution ensemble, leading to very low uncertainty estimates. Also different starting models can influence the magnitude of the estimates. It should be noted that the bootstrap uncertainty estimates are always with respect to the used parameters. As for most inversion approaches, the ideal parameter configurations have to be identified through tests before the main inversion.

The approach requires the calculation of several tens of inversion models, but by calculating all models in parallel, the increased computational cost is dispersed. Hence, for 2-D inversions, the execution is only marginally more time consuming than for a single inversion.

The bootstrap approach appears to be working better on the real world example than on the synthetic test model. This cannot be verified as we do not know the true model underlying the Delamerian transect data, and might be a misinterpretation. It could be due to higher data noise levels in the real world data, which in turn cause more variability in the bootstrap models. Another possible cause is that the simulated noise in the synthetic example may not exactly replicate the effects of real world noise. As is common practice, we used white noise for the synthetic data, but the noise on real world MT data can be more complex and might not always be normally distributed (Garth & Poor 1994; Manoj & Nagarajan 2003). Bootstrap resampling may be more effective on data with real world noise. The second stage of the resampling process would need to be adjusted in cases where the noise model deviates considerably from the normal distributed noise assumed here. One such example would be spatially correlated noise. Situations like that might also require larger number of bootstrap samples to achieve reliable estimates. These questions require further research.

6 CONCLUSION

We have presented an uncertainty estimation approach for inversion modelling of 2-D MT data sets based on bootstrap resampling. We have demonstrated its effectiveness in detecting noise-induced uncertainty by testing it on synthetic and real data examples. The approach is suited for qualitative estimation of errors introduced into the model by data noise and for the analysis of model feature robustness. The method can generally be used with any available inversion algorithm, as it does not require access to the inversion algorithm's source code, and it is easy to implement. It is therefore a viable tool for model uncertainty estimation.

ACKNOWLEDGEMENTS

The work has been supported by the Deep Exploration Technologies Cooperative Research Centre whose activities are funded by the Australian Government's Cooperative Research Centre Programme. This is DET CRC Document 2014/607.

We also thank Kate Robertson, the Geological Survey of Victoria and the University of Adelaide for providing us with the Delamerian MT Transect data, and thank Lars Krieger for his technical advice.

REFERENCES

- Alumbaugh, D.L. & Newman, G.A., 2000. Image appraisal for 2-D and 3-D electromagnetic inversion, *Geophysics*, **65**(5), 1455–1467.
- Becken, M., Ritter, O., Park, S.K., Bedrosian, P.A., Weckmann, U. & Weber, M., 2008. A deep crustal fluid channel into the San Andreas fault system near Parkfield, California, *Geophys. J. Int.*, **173**, 718–732.
- Caldwell, T.G., Bibby, H.M. & Brown, C., 2004. The magnetotelluric phase tensor, *Geophys. J. Int.*, **158**(2), 457–469.
- Campanyà, J., Ledo, J., Queralt, P., Marcuello, A. & Jones, A.G., 2014. A new methodology to estimate magnetotelluric (MT) tensor relationships: estimation of local transfer-functions by combining interstation transfer-functions (ELICIT), *Geophys. J. Int.*, **198**(1), 484–494.
- Chave, A.D., 2014a. Magnetotelluric data, stable distributions and impropriety: an existential combination, *Geophys. J. Int.*, **198**(1), 622–636.
- Chave, A.D., 2014b. On the statistics of magnetotelluric rotational invariants, *Geophys. J. Int.*, **196**(1), 111–130.
- Chave, A.D. & Jones, A.G., 2012. *The Magnetotelluric Method: Theory and Practice*, Cambridge Univ. Press.
- Choi, H., Byun, J. & Seol, S.J., 2010. Automatic velocity analysis using bootstrapped differential semblance and global search methods, *Explor. Geophys.*, **41**(1), 31–39.
- 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.
- Deal, M.M. & Nolet, G., 1996. Nullspace shuttles, *Geophys. J. Int.*, **124**, 372–380.
- deGroot Hedlin, C. & Constable, S., 1990. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data, *Geophysics*, **55**(12), 1613–1624.
- Efron, B., 1979. Bootstrap methods: another look at the jackknife, *The Annals of Statistics*, **7**(1), 1–26.
- Efron, B., 1982. *The Jackknife, the Bootstrap and Other Resampling Plans*, SIAM.
- Efron, B. & Tibshirani, R., 1986. Bootstrap methods for standard errors, confidence intervals, and other measures of statistical accuracy, *Statistical Science*, **1**(1), 54–77.
- Flanagan, M.P. & Shearer, P.M., 1998. Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors, *J. geophys. Res.*, **103**(B2), 2673–2692.
- Gamerman, D., 1997. *Markov Chain Monte Carlo: Stochastic Simulation for Bayesian Inference*, Chapman and Hall/CRC.
- Garth, L.M. & Poor, H.V., 1994. Detection of non-Gaussian signals: a paradigm for modern statistical signal processing [and prolog], *Proc. IEEE*, **82**(7), 1061–1095.
- Gurrola, H. & Minster, J.B., 1998. Thickness estimates of the upper-mantle transition zone from bootstrapped velocity spectrum stacks of receiver functions, *Geophys. J. Int.*, **133**(1), 31–43.
- Gurrola, H., Minster, J.B. & Owens, T., 1994. The use of velocity spectrum for stacking receiver functions and imaging upper mantle discontinuities, *Geophys. J. Int.*, **117**, 427–440.
- Guyodo, Y. & Valet, J.-P., 1996. Relative variations in geomagnetic intensity from sedimentary records: the past 200,000 years, *Earth planet. Sci. Lett.*, **143**(1–4), 23–36.
- Jackson, D.D., 1972. Interpretation of inaccurate, insufficient and inconsistent data, *Geophys. J. R. astr. Soc.*, **28**(2), 97–109.
- Jackson, D.D., 1976. Most squares inversion, *J. geophys. Res.*, **81**(5), 1027–1030.
- Jonkers, A.R.T., 2007. Bootstrapped discrete scale invariance analysis of geomagnetic dipole intensity, *Geophys. J. Int.*, **169**(2), 646–658.
- Kalscheuer, T., de los Angeles Garcia 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**, 1174–1188.
- Koch, M., 1992. Bootstrap inversion for vertical and lateral variations of the S-wave structure and the VP/VS-ratio from shallow earthquakes in the Rhinegraben seismic zone, Germany, *Tectonophysics*, **210**(1–2), 91–115.
- Korenaga, J., 2013. Stacking with dual bootstrap resampling, *Geophys. J. Int.*, **195**(3), 2023–2036.
- Korenaga, J., 2014. Teleseismic migration with dual bootstrap stack, *Geophys. J. Int.*, **196**(3), 1706–1723.
- Laloy, E., Linde, N. & Vrugt, J.A., 2012. Mass conservative three-dimensional water tracer distribution from MCMC inversion of timelapse GPR data, *Water Resour. Res.*, **48**, W07510, doi:10.1029/2011WR011238.
- Linde, N. & Vrugt, J., 2013. Distributed soil moisture from crosshole ground-penetrating radar using Markov chain Monte Carlo simulation, *Vadose Zone Journal*, **12**(1), 0101, doi:10.2136/vzj2012.0101.
- Manoj, C. & Nagarajan, N., 2003. The application of artificial neural networks to magnetotelluric time-series analysis, *Geophys. J. Int.*, **153**(2), 409–423.
- Marescot, L., Loke, M.H.D., Chapellier, Delaloye, R., Lambiel, C. & Reynard, E., 2003. Assessing reliability of 2D resistivity imaging in mountain permafrost studies using the depth of investigation method, *Near Surface Geophysics*, **1**, 57–67.
- Marti, A., Queralt, P. & Roca, E., 2004. Geoelectric dimensionality in complex geological areas: application to the Spanish Betic Chain, *Geophys. J. Int.*, **157**, 961–974.
- McLaughlin, K.L., 1988. Maximum-likelihood event magnitude estimation with bootstrapping for uncertainty estimation, *Bull. seism. Soc. Am.*, **78**(2), 855–862.
- Meju, M.A., 1994. *Geophysical Data Analysis: Understanding Inverse Problem Theory and Practice*, Society of Exploration Geophysicists.
- Meju, M.A. & Hutton, V.R.S., 1992. Iterative most-squares inversion: application to magnetotelluric data, *Geophys. J. Int.*, **108**(3), 758–766.
- Menke, W., 2012. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press.
- Metropolis, N. & Ulam, S., 1949. The Monte Carlo method, *J. Am. Stat. Assoc.*, **44**, 335–341.
- Muñoz, G. & Rath, V., 2006. Beyond smooth inversion: the use of nullspace projection for the exploration of non-uniqueness in MT, *Geophys. J. Int.*, **164**, 301–311.
- Neukirch, M. & Garcia, X., 2014. Nonstationary magnetotelluric data processing with instantaneous parameter, *J. geophys. Res.*, **119**, 1634–1654.
- Oldenburg, D.W. & Li, Y., 1999. Estimating depth of investigation in DC resistivity and IP surveys, *Geophysics*, **64**(2), 403–416.
- Ramirez, A.L., Daily, W.D. & Newmark, R.L., 1995. Electrical resistance tomography for steam injection monitoring and control, *Journal of Environmental and Engineering Geophysics*, **1**, 39–52.
- Robertson, K.E., Taylor, D.H., Thiel, S. & Heinson, G.S., 2015. Magnetotelluric evidence for serpentinisation in a Cambrian subduction zone beneath

- the Delamerian Orogen, southeast Australia, *Gondwana Research*, **28**(2), 601–611.
- Rosas-Carbajal, M., Linde, N., Kalscheuer, T. & Vrugt, J.A., 2014. Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data, *Geophys. J. Int.*, **196**, 1508–1524.
- Rowbotham, P.S. & Pratt, R.G., 1997. Improved inversion through use of the null space, *Geophys. Prospect.*, **62**, 869–883.
- Sacchi, M.D., 1998. A bootstrap procedure for high-resolution velocity analysis, *Geophysics*, **63**(5), 1716–1725.
- Sandvol, E. & Hearn, T., 1994. Bootstrapping shear-wave splitting errors, *Bull. seism. Soc. Am.*, **84**(6), 1971–1977.
- Shaar, R., Ron, N., Tauxe, L., Kessel, R. & Agnon, A., 2011. Paleomagnetic field intensity derived from non-SD: testing the Thellier IZZI technique on MD slag and a new bootstrap procedure, *Earth planet. Sci. Lett.*, **310**(3–4), 213–224.
- Shearer, P.M., 1997. Improving local earthquake locations using the $\| \cdot \|_1$ norm and waveform cross correlation: application to the Whittier Narrows, California, aftershock sequence, *J. geophys. Res.*, **102**(B4), 8269–8283.
- Tarantola, A., 2005. *Inverse Problem Theory and Methods for Model Parameter Estimation*, SIAM.
- Tauxe, L., Kylstra, N. & Constable, C., 1991. Bootstrap statistics for paleomagnetic data, *J. geophys. Res.*, **96**(B2), 11 723–11 740.
- Tauxe, L., Gee, J.S. & Staudigel, H., 1998. Flow directions in dikes from anisotropy of magnetic susceptibility data: the bootstrap way, *J. geophys. Res.*, **103**(B8), 17 775–17 790.
- Thiel, S. & Heinsohn, G., 2010. Crustal imaging of a mobile belt using magnetotellurics: an example of the Fowler Domain in South Australia, *J. geophys. Res.*, **115**, B06102, doi:10.1029/2009JB006698.
- Tichelhaar, B.W. & Ruff, L.J., 1989. How good are our best models? Jack-knifing, bootstrapping, and earthquake depth, *EOS, Trans. Am. geophys. Un.*, **55**(12), 1613–1624.
- Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-posed Problems*, John Wiley and Sons.
- Wannamaker, P.E., Stodt, J.A. & Rijo, L., 1987. A stable finite-element solution for two-dimensional magnetotelluric modeling, *Geophys. J. R. astr. Soc.*, **88**, 277–296.
- Weidelt, P., 1972. The inverse problem of geomagnetic induction, *Z. Geophys.*, **38**, 257–289.