submitted to *Geophys. J. Int.*

Using non-diagonal data covariances in geophysical inversion

Max Moorkamp[1] and Anna Avdeeva[2]

[1]*Ludwig-Maximilians-Universität München, Theresienstrasse 41, München,*

[2] *Complete MT Solutions Inc., Manotick, Canada*

SUMMARY

We present a new approach that allows for the inversion of quantities derived from the observed data using non-diagonal data covariance matrices. For example, we can invert approximations of apparent resistivity and phase instead of magnetotelluric impedance using this methodology. Compared to the direct inversion of these derived quantities, the proposed methodology has two advantages: i) It is fully compatible with the assumptions of least-squares optimization and thus avoids potential issues with bias when inverting quantities that are non-linear functions of the original data, ii) If an inversion algorithm allows for the specification of a full data covariance matrix, users can invert for arbitrary derived quantities by specifying the appropriate covariance matrix instead of having to rely on the inversion code to have implemented this feature. We discuss the theory of this approach and show an example using magnetotelluric data. However, the same method can be applied to other types of geophysical data, for example gravity gradient measurements.

Key words: magnetotellurics, inverse theory

1 INTRODUCTION

Measuring misfit, i.e. the discrepancy between observations and the predictions made by geophysical models, is a key element of data analysis and particularly inversion (e.g. Tarantola 2004). Depending on how we define our measure of misfit, we consider different models as adequate representations of the physical properties within the Earth (e.g. Farquharson & Oldenburg 1998). Defining misfit is closely related to estimates of uncertainty of the observed data and any measure of misfit explicitly or implicitly assumes certain confidence limits on each datum. It is only in the combination of a suitably defined misfit measure with some knowledge of data errors that we can make an informed judgement on the validity of a given model of the Earth (Mosegaard & Hansen 2016).

Despite their fundamental importance, data errors are notoriously difficult to estimate in geophysical applications. Usually, instrument noise of the equipment recording the field data is well known, i.e. the precision of A/D converters or the noise of recording amplifiers. However, environmental factors such as shaking of instruments or thermal drift are much more difficult to estimate and in most cases subsequent data processing such as the various corrections needed to analyse gravity data (e.g. Hwang et al. 2007), or picking arrivals teleseismic traveltime tomography (e.g. Di Stefano et al. 2006), for example, introduce significant uncertainties that escape rigorous error analysis. Where repeat data are available or the processing is based on stacking different events, we can use statistical methods to estimate uncertainty. For example, magnetotelluric impedance data are estimated from field observations through statistical processing (e.g. Egbert 2002; Chave & Thomson 2004) and a wealth of studies on the statistical properties of the estimates exists (e.g. Chave et al. 1987; Chave & Lezaeta 2007). Still, deviations from the underlying statistical models can cause incorrect estimates of data error (Eisel & Egbert 2001; Chave 2017). Even neglecting this fact, there is still a debate on how to incorporate the uncertainty estimates from statistical methods into practical inversion schemes (Tietze & Ritter 2013).

For these reasons, modelling and inversion of geophysical data is often performed using ad-hoc estimates of uncertainty. These are usually empirical values based on the practicioners assessment of the data and the adherence of certain vaguely defined desirable properties. As a consequence data are typically assumed to be independent and global error estimates for all data are often used. Such global estimates can be specified as absolute values, e.g. a picking error of 30 ms for all measurements in cross-hole tomography (Bregman et al. 1989), or as relative values, e.g. a few percent of each datum in DC resistivity inversions (e.g. Günther et al. 2006). In either case no distinction is made between different data points in terms of their quality. Even when individual error estimates for each datum are available, practical considerations often make it necessary to prescribe an error floor, i.e. a minimal error for all data, to reduce the influence of measurements with unrealistically small error estimates (e.g. Moorkamp et al. 2019). As most current inversion algorithms are based on a least-square misfit

function these error estimates translate into a variance for each datum, i.e. the data covariance matrix in the definition of the objective function (see Equation 1 below) is assumed to be diagonal.

Given these fundamental difficulties in obtaining good estimates of individual variances, it might seem surprising that we extend our analysis to misfit measures using non-diagonal covariance matrices. The off-diagonal entries of the covariance matrix, i.e. the interdependence of different data, can theoretically be estimated for different elements of the magnetotelluric impedance tensor at a given frequency (Chave 2017). However, we are not aware of any practical study that has used these estimates to date. Here we use non-diagonal covariance matrix to emulate the inversion of quantities derived from the original data, e.g. apparent resistivity and phase instead of impedance for magnetotelluric data. Such an inversion of derived quantities is sometimes preferred for practical reasons, e.g. to deal with effects of static distortion (e.g. Khoza et al. 2013).

Even though we focus here on magnetotelluric data, the presented methodology is completely general and can be used for any type of inversion. The same formulation can be used in the inversion of gravity gradient tensor data, for example when fitting tensor invariants instead of the elements of the tensor (Heath 2007).

2 LEAST-SQUARES MEASURES OF MISFIT

Most current geophysical inversion codes use a least squares measure of data misfit $\Phi(\mathbf{m})$ defined as

$$\Phi(\mathbf{m}) = (\mathbf{s}(\mathbf{m}) - \mathbf{d})^H \mathbf{C}_D^{-1} (\mathbf{s}(\mathbf{m}) - \mathbf{d}) . \tag{1}$$

Here $\mathbf{m}$ is the M-dimensional model vector for which we want to calculate the misfit, $\mathbf{s}(\mathbf{m})$ the n-dimensional vector of synthetic data calculated from that model, $\mathbf{d}$ the vector of observed data and $\mathbf{C}_D$ the $N \times N$-dimensional data covariance matrix which is symmetric for real valued observations and hermitian for complex valued data. H denotes the hermitian, i.e. the complex conjugate of the transpose matrix. Within the inversion, an optimization algorithm minimizes the misfit functional, usually combined with a measure of model complexity, until a local minimum is encountered or a threshold value for the misfit is reached. If the data errors are random and follow a Gaussian distribution, statistics gives us an objective criterion for satisfactory misfit, $\Phi(\mathbf{m}) = N$ (Tarantola 2004). Equivalently we can define the RMS misfit

$$R(\mathbf{m}) = \sqrt{\frac{\Phi(\mathbf{m})}{N}} \tag{2}$$

which indicates how many standard deviation on average the synthetic data differs from the observations. Thus mathematically any model that fits the observed data to an RMS value of 1 can be considered an adequate explanation for the observations from a statistical point of view. In practice,

additional soft criteria, e.g. geological plausibility, are used to select a preferred model from the infinite number of possible solutions (e.g. Rao et al. 2014).

For inversion we typically assume mutually independent data, so the covariance matrix has simple diagonal form

$$\mathbf{C}_D = \begin{pmatrix} \sigma_1^2 & 0 & \cdots & 0 \\ 0 & \sigma_2^2 & \cdots & 0 \\ \vdots & \vdots & \ddots & \vdots \\ 0 & 0 & \cdots & \sigma_n^2 \end{pmatrix}. \tag{3}$$

This is not only computationally convenient as the data covariance and its inverse can be readily calculated, but also reflects the difficulties in estimating representative errors as discussed above. As a consequence the variances σ_i^2 are rarely rigorously justified through quantitative analysis but a numerical reflection of the users confidence in each datum.

We now assume the more general case where we have a full covariance matrix with off-diagonal elements that express the interdependence of different data

$$\mathbf{C}_D = \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \cdots & \sigma_{1n} \\ \sigma_{12} & \sigma_2^2 & \cdots & \sigma_{2n} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{1n} & \sigma_{2n} & \cdots & \sigma_n^2 \end{pmatrix}. \tag{4}$$

All variances σ_i^2 are positive and $\mathbf{C}_D$ is hermitian and semi-positive-definite (Tarantola 2004), the individual covariances σ_{ij}, however, are not necessarily positive. If we assume that we do not have redundant data, i.e. the columns of $\mathbf{C}_D$ are linearly independent, the inverse weighting matrix $\mathbf{W} = \mathbf{C}_D^{-1}$ is guaranteed to exist and hermitian and positive-definite.

To demonstrate how we can use non-diagonal covariance matrices in a practical inversion scheme, we use the example of magnetotelluric inversion. In magnetotellurics we estimate a complex, frequency dependent 2×2 impedance tensor

$$\mathbf{Z} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \tag{5}$$

and the corresponding variances from simultaneous measurements of electric and magnetic fields at the Earth's surface (e.g. Simpson & Bahr 2005; Chave & Jones 2012). Most published inversion approaches choose to neglect the tensorial nature of the magnetotelluric impedance and instead transform the 2×2 tensor at a given frequency into a 4-dimensional complex vector or a 8-dimensional real vector (e.g. Newman & Alumbaugh 2000; Farquharson & Oldenburg 2004; Abubakar et al. 2011; Moorkamp et al. 2011). The justification for this approach are the isomorphisms between the space of

2×2 complex matrices, the 4-dimensional complex vector space and the 8-dimensional real valued vector space. In a 4-dimensional complex vector space, the definition of misfit then takes the form

$$\Phi(\mathbf{m}) = \begin{pmatrix} \overline{Z^i_{xx} - S^i_{xx}} \\ \overline{Z^i_{xy} - S^i_{xy}} \\ \overline{Z^i_{yx} - S^i_{yx}} \\ \overline{Z^i_{yy} - S^i_{yy}} \\ \overline{Z^{i+1}_{xx} - S^{i+1}_{xx}} \\ \vdots \\ \overline{Z^N_{yy} - S^N_{yy}} \end{pmatrix}^T \begin{pmatrix} \frac{1}{\sigma^2_{xx,i}} & 0 & 0 & 0 & 0 & \cdots & 0 \\ 0 & \frac{1}{\sigma^2_{xy,i}} & 0 & 0 & 0 & \cdots & 0 \\ 0 & 0 & \frac{1}{\sigma^2_{yx,i}} & 0 & 0 & \cdots & 0 \\ 0 & 0 & 0 & \frac{1}{\sigma^2_{yy,i}} & 0 & \cdots & 0 \\ 0 & 0 & 0 & 0 & \frac{1}{\sigma^2_{xx,i+1}} & \cdots & 0 \\ \vdots & \vdots & \vdots & \vdots & 0 & \ddots & \vdots \\ 0 & 0 & 0 & 0 & 0 & \cdots & \frac{1}{\sigma^2_{yy,N}} \end{pmatrix} \begin{pmatrix} Z^i_{xx} - S^i_{xx} \\ Z^i_{xy} - S^i_{xy} \\ Z^i_{yx} - S^i_{yx} \\ Z^i_{yy} - S^i_{yy} \\ Z^{i+1}_{xx} - S^{i+1}_{xx} \\ \vdots \\ Z^N_{yy} - S^N_{yy} \end{pmatrix}, \tag{6}$$

where $i = 1, \ldots, N = F \cdot O$ is the index of the tensor element, F the number of frequencies and O the number of observation sites. S^i_{jj} are the synthetic impedance elements calculated from the current model $\mathbf{m}$. The variances are taken from statistical estimation procedures, possibly adjusted by prescribing an error floor, and are generally different for each element of the impedance tensor and each frequency (e.g. Newman & Alumbaugh 2000; Farquharson & Oldenburg 2004; Abubakar et al. 2011; Moorkamp et al. 2011). In the real valued vector formulation, the real and imaginary part of each impedance element are associated with separate vector components, but otherwise the misfit equation has the same structure as Equation 6. Typically the variance for the real and imaginary parts of each impedance element are assumed to be identical, as this is what most processing routines provide.

3 INVERTING FOR DERIVED QUANTITIES

We will now show how non-diagonal covariance matrices can be used to simulate the inversion of derived quantities. In magnetotellurics, for example, it is common practice to invert for apparent resistivity $\rho_a = \frac{1}{2\pi f \mu}|Z|^2$ and phase $\varphi = \tan^{-1}\left(\frac{Y}{X}\right)$ instead of complex impedance elements $Z = X + iY$. Here f is the frequency at which the impedance is measured and μ the magnetic permeability of the vacuum. In some cases impedance phase is less affected by noise or static shift effects than apparent resistivity and thus small error estimates for the phase but larger errors for apparent resistivity can be used in the inversion (e.g. Khoza et al. 2013). In other cases apparent resistivity is down-weighted compared to phase in order to allow inversion of marine MT data that is severely affected by coast effects (Worzewski et al. 2011). A number of inversion codes allow to fit apparent resistivity and phase (e.g. Mackie & Madden 1993). However, the residuals for both quantities are not normally distributed (Chave & Lezaeta 2007). The Gauss-Markov theorem guarantees that the estimates from least-squares

procedures are unbiased and consistent for gaussian errors on the observations, but deviations from these assumptions can lead to biased results (Jennrich 1969; Constable 1988).

However, for simple models and small errors the problem of bias seems to be neglible (Wheelock et al. 2015). Here we show an alternative way that preserves the statistical properties of the magnetotelluric impedance while simulating the fit to a derived quantity. Another advantage of this approach is that, once the general ability to specify the inverse covariance matrix has been implemented, it allows the users of the inversion code to invert for arbitrary derived quantities without changing the source code of the inversion. In our inversion algorithm we have chosen to implement the inverse covariance in terms of sparse matrices as typically we will not have covariances between many combinations of data. Furthermore, the user has to directly specify the inverse covariance, as this is needed throughout the algorithm. This simplifies the implementation tremendously as only a sparse matrix-vector multiplication has to added to the calculation of the objective function. It does put the burden of calculating the inverse to the user, but it can be readily achieved with open-source tools such as Maxima and Python and gives the usere the ability to invert for quantities that the developers might not have foreseen.

For a multi-variate problem

$$\mathbf{z}\left(\xi\right) = \left(z_1(\xi_1, \ldots, \xi_N), \ldots, z_N(\xi_1, \ldots, \xi_N)\right) \tag{7}$$

we can map the co-variance $\mathbf{C}_\xi$ specified with respect to the quantities $\xi_1, \ldots, \xi_N$ to a co-variance $\mathbf{C}_D$ with respect to the quantities $z_1, \ldots, z_N$ through

$$\mathbf{C}_D = \mathbf{J}\mathbf{C}_\xi\mathbf{J}^T \tag{8}$$

(Tellinghuisen 2001; Booker 2014). Here $\mathbf{J}$ is the matrix of partial derivatives

$$\mathbf{J} = \begin{pmatrix} \frac{\partial}{\partial\xi_1}z_1 & \cdots & \frac{\partial}{\partial\xi_N}z_1 \\ \cdots & & \cdots \\ \frac{\partial}{\partial\xi_1}z_N & \cdots & \frac{\partial}{\partial\xi_N}z_N \end{pmatrix}. \tag{9}$$

In the following we use magnetotelluric apparent resistivity and phase as first examples to demonstrate how we can use this in practice. Assuming the inversion fits the elements of the impedance tensor, we want it to behave as if it was fitting apparent resistivity and phase, i.e. we want to be able to specify variances for apparent resistivity and phase and map these variances into a covariance matrix for the impedance.

A single element Z of the magnetotelluric impedance tensor can be written in terms of apparent resistivity and phase as

$$Z = X + iY = \sqrt{2\pi\mu f\rho_a}\left(\cos\varphi + i\sin\varphi\right). \tag{10}$$

In situations where we want to express the variance and covariances of the impedance in terms of the variance of apparent resistivity and phase, we need the expression of the real and imaginary parts of Z in terms of these quantities, which is simply

$$X(\rho_a, \varphi) = \sqrt{2\pi\mu f \rho_a}\cos\varphi, \qquad Y(\rho_a, \varphi) = \sqrt{2\pi\mu f \rho_a}\sin\varphi. \tag{11}$$

We now consider only a single element Z of the impedance tensor and split the misfit calculation into a real vector that contains the real and imaginary parts of Z as separate elements, i.e. we examine a misfit functional of the form

$$\Phi(\mathbf{m}) = \begin{pmatrix}\mathrm{Re}(Z-S) & \mathrm{Im}(Z-S)\end{pmatrix}\begin{pmatrix}\sigma_{11} & \sigma_{12}\\ \sigma_{21} & \sigma_{22}\end{pmatrix}^{-1}\begin{pmatrix}\mathrm{Re}(Z-S)\\ \mathrm{Im}(Z-S)\end{pmatrix}. \tag{12}$$

This makes the analysis slightly easier and we will show below how to construct a full data covariance matrix for a practical inversion scheme with all impedance elements at a number of frequencies and sites.

In this case we have

$$\mathbf{J}_{\rho_a,\varphi} = \begin{pmatrix}\frac{\partial}{\partial\rho_a}X & \frac{\partial}{\partial\varphi}X\\ \frac{\partial}{\partial\rho_a}Y & \frac{\partial}{\partial\varphi}Y\end{pmatrix} = \begin{pmatrix}\frac{\pi\mu f\cos(\varphi)}{\sqrt{2\pi\mu f\rho_a}} & -\sqrt{2\pi\mu f\rho_a}\sin(\varphi)\\ \frac{\pi\mu f\sin(\varphi)}{\sqrt{2\pi\mu f\rho_a}} & \sqrt{2\pi\mu f\rho_a}\cos(\varphi)\end{pmatrix}. \tag{13}$$

So for each complex impedance element we have a covariance matrix of the form

$$\mathbf{C}_D = \mathbf{J}_{\rho_a,\varphi}\begin{pmatrix}\sigma^2_{rho} & 0\\ 0 & \sigma^2\varphi\end{pmatrix}\mathbf{J}^T_{\rho_a,\varphi} \tag{14}$$

$$= \pi\mu f\begin{pmatrix}\frac{\cos^2(\varphi)\,\sigma^2_{\rho a}}{2\,\rho_a} + 2\sin^2(\varphi)\,\rho_a\,\sigma^2_\varphi & \cos(\varphi)\,\sin(\varphi)\left(\frac{\sigma^2_{\rho_a}}{2\,\rho_a} - 2\rho_a\,\sigma^2_\varphi\right)\\ \cos(\varphi)\,\sin(\varphi)\left(\frac{\sigma^2_{\rho_a}}{2\,\rho_a} - 2\rho_a\,\sigma^2_\varphi\right) & \frac{\sin^2(\varphi)\,\sigma^2_{\rho a}}{2\,\rho_a} + 2\cos^2(\varphi)\,\rho_a\,\sigma^2_\varphi\end{pmatrix} \tag{15}$$

At first sight, the form of this co-variance matrix is confusing, as the expressions for each element contain the variances with respect to both apparent resistivity and phase in a non-trivial way. To better understand the properties of this matrix we calculate the eigenvalues and eigenvectors which are

$$\lambda_1 = \frac{\pi\mu f\sigma^2_{\rho_a}}{2\rho_a} \quad \mathbf{e}_1 = \begin{pmatrix}1\\ \tan(\phi)\end{pmatrix} \text{ and } \lambda_2 = 2\pi\mu f\rho_a\sigma^2_\phi \quad \mathbf{e}_2 = \begin{pmatrix}1\\ -\mathrm{cotan}(\phi)\end{pmatrix}. \tag{16}$$

Figure 1 shows a sketch of the situation in the complex plane. The variances for apparent resistivity and phase get mapped into an ellipse around the impedance estimate Z. In this illustration the eigenvalue for the eigenvector $\mathbf{e}_1$ is larger than $\mathbf{e}_2$, i.e. the variation in terms of apparent resistivity is relatively large and in terms of phase is relatively small. From equation 16 we can see how the magnitude of these two eigenvectors directly depends on the variances of apparent resistivity and phase and thus specifying a large variance in terms of phase and small variance for apparent resistivity will turn the elongated axis by 90 degrees.

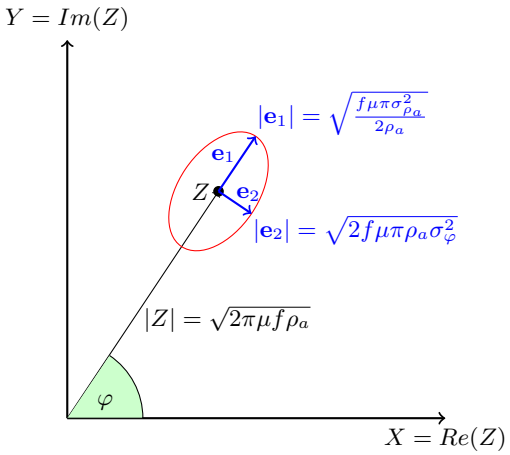

Figure 1. Mapping different variances for apparent resistivity ρ_a and phase φ onto the complex plane. The red ellipse around the impedance estimate Z shows the mapped confidence area as described by the covariance, we show the eigenvectors and the corresponding eigenvalues in blue.

For inversion we need the inverse of the covariance matrix

$$\mathbf{C}_D^{-1} = \frac{1}{2\pi\mu f\rho_a\sigma^2_{\varphi}\sigma^2_{\rho_a}} \begin{pmatrix} \sin^2(\varphi)\sigma^2_{\rho_a} + 4\cos^2(\varphi)\rho_a^2\sigma^2_{\varphi} & \cos(\varphi)\sin(\varphi)\left(4\rho_a^2\sigma^2_{\varphi} - \sigma^2_{\rho_a}\right) \\ \cos(\varphi)\sin(\varphi)\left(4\rho_a^2\sigma^2_{\varphi} - \sigma^2_{\rho_a}\right) & \cos^2(\varphi)\sigma^2_{\rho_a} + 4\sin^2(\varphi)\rho_a^2\sigma^2_{\varphi} \end{pmatrix}. \quad (17)$$

The 2×2 data covariance derived above, describes the inter-relationship between the real and imaginary parts of each impedance element. In this case, however, there is no dependence between different elements of impedance and thus the full covariance for an inversion with many elements, frequencies and sites can be built up from individual blocks of the smaller covariance matrices for each element. For magnetotelluric data the maximum size of such blocks would be 8×8 for quantities that depend on all elements of impedance at one frequency. We cannot currently think of practical applications where one would have dependencies between different frequencies and/or sites. As the inverse of a block matrix is the inverse of each block, such a scheme would require the inversion of many 8×8 matrices which is easily achievable on modern computer architectures.

3.1 Implementation in an inversion algorithm and comparison with standard approach

Incorporating this approach into an inversion algorithm is straightforward. For the definition of misfit all that needs to be done is to replace the inverse data covariance matrix, for example with the form shown in Equation 17 for apparent resistivity and phase. This change propagates to the calculation of the gradient of the objective function in a straightforward manner, viz.

$$\frac{\partial\Phi(\mathbf{m})}{\partial\mathbf{m}} = 2\mathbf{G}^T\mathbf{C}_D^{-1}\left(\mathbf{s}(\mathbf{m}) - \mathbf{d}\right), \quad (18)$$

where $\mathbf{G}$ is the matrix of partial derivatives with elements $G_{ij} = \frac{\partial s_i(\mathbf{m})}{\partial m_j}$. In most practical algorithms $\mathbf{G}$ is not formed explicitly, but the vector-matrix product is calculated using reciprocity (e.g. Mackie et al. 1993; Avdeeva et al. 2015).

For an inversion algorithm that directly wants to fit a derived quantity, additional changes have to be made that are not particularly difficult, but require changes to the misfit and gradient calculation. Remembering that in our notation z_i are the original observables and ξ_i the quantities derived from them, the objective function becomes

$$\Phi_\xi(\mathbf{m}) = \left(\xi(\mathbf{s}(\mathbf{m})) - \xi(\mathbf{d})\right)^H \mathbf{C}_\xi^{-1} \left(\xi(\mathbf{s}(\mathbf{m})) - \xi(\mathbf{d})\right), \tag{19}$$

where $\xi(z_1, \ldots, z_N)$ is the function that describes the mapping between observables and derived elements and is the inverse of the function in Equation 7. For the gradient calculation we further need the derivative matrix $\mathbf{K}$ with elements $K_{ij} = \frac{\partial \xi_i}{\partial z_j}$. Then the gradient becomes

$$\frac{\partial \Phi_\xi(\mathbf{m})}{\partial \mathbf{m}} = 2\mathbf{G}^T\mathbf{K}^T\mathbf{C}_\xi^{-1} \left(\xi(\mathbf{s}(\mathbf{m})) - \xi(\mathbf{d})\right). \tag{20}$$

If the relationship between observables and derived quantities is linear, i.e.

$$\xi(\mathbf{z}) = \mathbf{K}\mathbf{z}, \tag{21}$$

we have $\mathbf{K} = \mathbf{J}^{-1}$ and thus Equation 20 can be written as

$$\begin{aligned}
\frac{\partial \Phi_\xi(\mathbf{m})}{\partial \mathbf{m}} &= 2\mathbf{G}^T\mathbf{K}^T\mathbf{C}_\xi^{-1} \left(\xi(\mathbf{s}(\mathbf{m})) - \xi(\mathbf{d})\right) && (22)\\
&= 2\mathbf{G}^T\mathbf{K}^T\mathbf{C}_\xi^{-1}\mathbf{K} \left(\mathbf{s}(\mathbf{m}) - \mathbf{d}\right) && (23)\\
&= 2\mathbf{G}^T\mathbf{J}^{-\mathbf{1}^T}\mathbf{C}_\xi^{-1}\mathbf{J}^{-\mathbf{1}} \left(\mathbf{s}(\mathbf{m}) - \mathbf{d}\right) && (24)\\
&= 2\mathbf{G}^T\mathbf{C}_D^{-1} \left(\mathbf{s}(\mathbf{m}) - \mathbf{d}\right). && (25)
\end{aligned}$$

So for a linear parameter transformation, our approach and the traditional approach of inverting transformed quantities are identical. However, if the parameter transformation is non-linear, there are some important differences. To begin with, the matrices $\mathbf{J}$ and $\mathbf{K}$ are no longer the inverse of each other. Furthermore, the derivatives in $\mathbf{J}$ are calculated around the observed data and are constant during the inversion while the derivatives in $\mathbf{K}$ are calculated around the synthetic data and change at each iteration.

3.2 Analysing bias

As alluded to in the introduction, inverting non-linear transformations of the observed data can introduce bias into the solution of the inverse problem (Tellinghuisen 2000). While this fact is established in the statistical literature, it is usually ignored in practical applications. Here we investigate for the

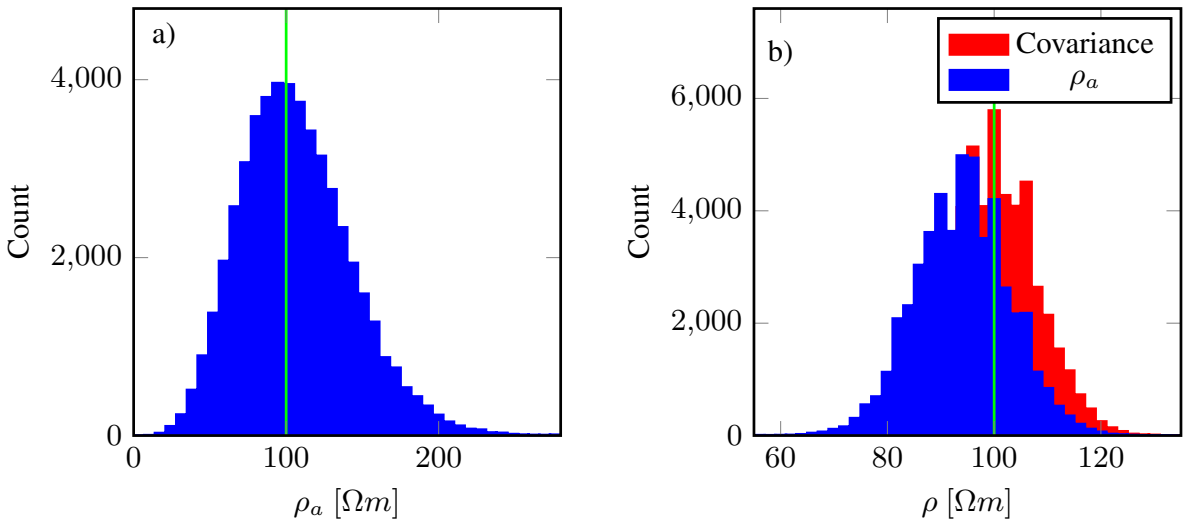

Figure 2. Histogram of the apparent resistivities calculated from the noisy realizations at a frequency of 100 Hz (a) and histogram of the resistivity estimates obtained from inversion of the noisy data (b). In both plots the true value of 100 Ωm is marked with a green line.

transformations established above to which degree this can pose a problem and whether it has significant impact in practice.

In order to illustrate the issue of bias, we use a simple example of retrieving the resistitivy of a homogeneous half-space with magnetotelluric measurements. We calculate synthetic impedances for a 100 Ωm half-space at 20 frequencies from $100 - 0.01$ Hz. We then generate 50,000 realizations of noisy data by adding random samples from a Gaussian distribution with zero mean and a standard deviation of $0.25|Z|$ to the real and imaginary parts of Z at each frequency. For each of these realizations we then find the best-fitting half-space by minimizing the objective function in Equation 1.

Figure 2a shows a histogram of apparent resistivity calculated from the noisy realizations of Z at a frequency of 100 Hz. We can identify a small upward bias of the histogram with a mean value of $106.4(2)\Omega m$ which matches the value predicted by equation 29 in Wheelock et al. (2015). Figure 2b shows a histogram of the half-space values resulting from the minimization of the objective function. Interestingly the histogram here is downward biased with a mean retrieved half-space conductivity of 94.05(4) Ωm. In contrast, the mean for the estimate based on impedance is $100.11(4)$. It is surprising and counter-intuitive that the inversion of the upward-biased apparent resistivity results in a downward biased estimate of the resistivity of the half-space. However, there is a simple explanation for this. When calculating the standard deviations for the noisy apparent resistivities we use the noisy impedances and apply linear error propagation

$$\sigma_\rho^2 = \frac{|Z|^2}{(\mu\pi f)^2}\sigma_Z^2 = \frac{2\rho}{\pi\mu f}\sigma_Z^2. \tag{26}$$

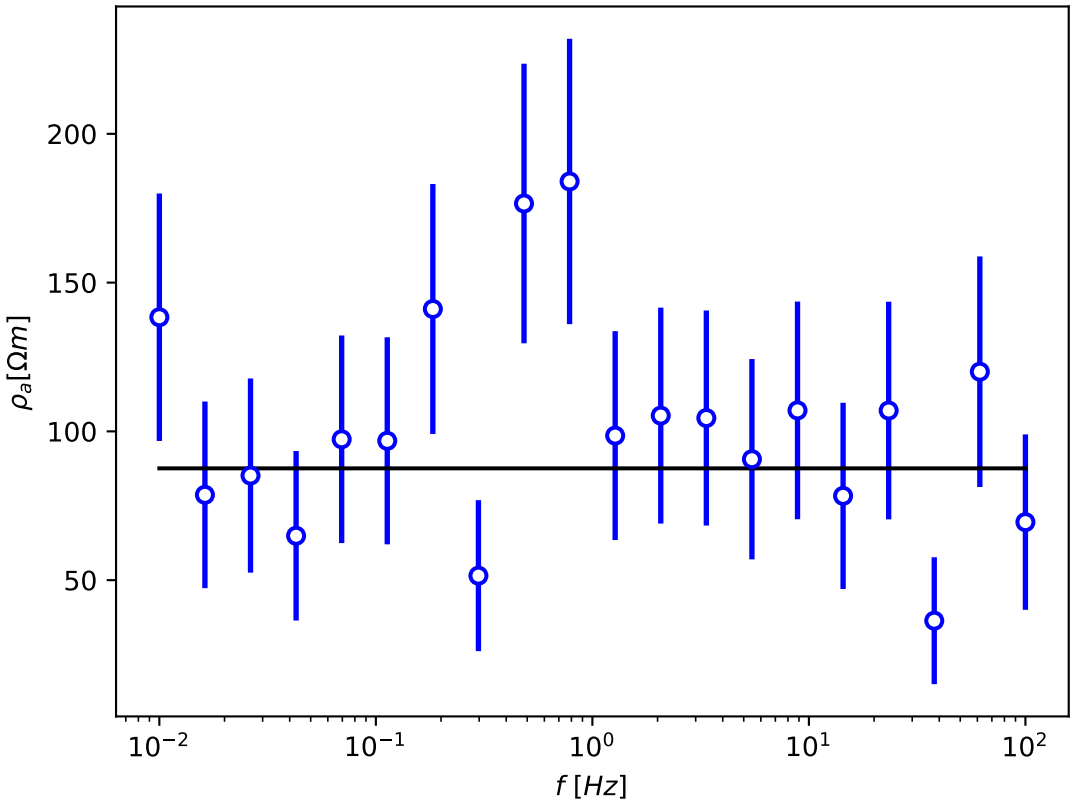

Figure 3. Apparent resistivity values with error bars for one realization of noise. The predicted line for the best fitting half-space is shown in black.

Thus even though all samples at a given frequency have been generated with the same variance σ_Z^2, those realizations that result in a higher than average apparent resistivity will have an increased associated variance in apparent resistivity and vice versa. The net result is that estimates with high apparent resistivity are down-weighted in the objective function compared to estimates with a low apparent resistivity and as a consequence the minimum of the misfit function is shifted towards lower estimates for the resistivity of the half-space.

Figure 3 illustrates this uneven weighting for different apparent resistivity values. Even though the relative error is identical for all frequencies, the propagated errors for apparent resistivity values greater than 100 Ωm is larger than for the values below 100 Ωm. As a result the line of best fit (black line) is biased towards smaller values.

When using our formulation with non-diagonal covariances and propagate the calculated errors for apparent resistivity and phase to impedance, the mean half-space conductivity recovered from all realizations is 100.28(5) Ωm. Strictly speaking our results indicate that the estimates from impedance and covariance are also biased as the true value deviates more than 3 standard deviations of the mean from the true value for impedance and 5 standard deviations of the mean for covariance. However,

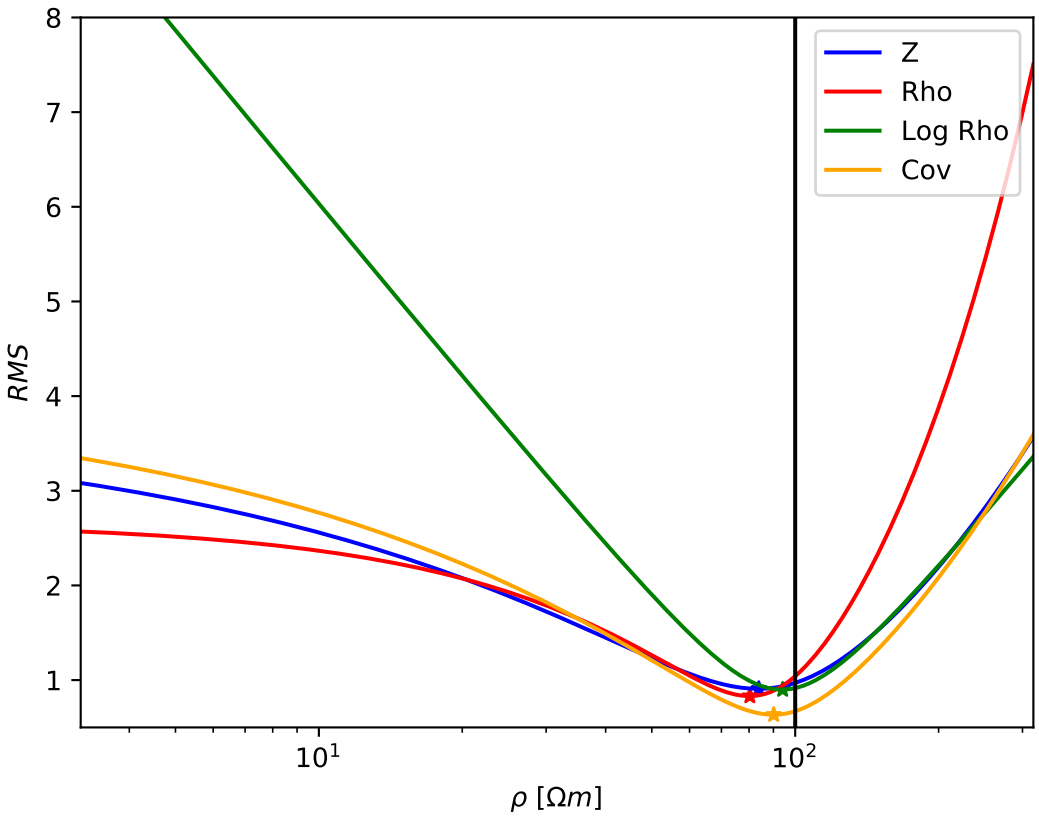

Figure 4. Dependency of the RMS misfit on the half-space resistivity for one realization of noise and different measures of misfit. The minimal misfit for each measure is marked with a star, the true value is marked with a black line.

our precision in the miminization of the objective function is also limited and the deviation of less than 0.3% from the true value is within the accuracy to which we can find that minimum. In contrast, the mean value based on the inversion of apparent resistivities is biased by more than 100 standard deviations of the mean and cannot be attributed to this.

Figure 4 shows the misfit as a function of half-space resistivity for one realization and four different measures of misfit. In addition to misfit measures based on impedance, apparent resistivity and our covariance measure, we also plot the dependency for the logarithm of apparent resistivity. As demonstrated by Wheelock et al. (2015) the misfit measure based on apparent resistivity saturates for small half-space resistivities, i.e. the misfit does not change significantly even when the resistivity changes by an order of magnitude. For this reason Wheelock et al. (2015) propose to use the logarithm of resistivity to avoid this phenomenon. When using a measure based on the covariance, the misfit increases even for resistivity values below 10 Ωm albeit at a lower rate than for the logarithmic measure. Overall, the behaviour of misfit for the covariance shows characteristics between inverting impedance and the logarithm of apparent resistivity.

Note that for this particular realization of noise, the minima of the objective functions for all four measures of misfit are close to each other and occur at half-space resistivities less than 100 Ωm. So the fact that the inversion of apparent resistivity is biased does not mean that in all cases the inversion result based on impedance or covariance will be superior. In fact, there will be realizations of noise where the inversion of apparent resistivity will yield the correct half-space resistivity and the inversion of impedance will give a shifted result. On average though theresults based on apparent resistivity will be too small and unbiased measures will give the correct answer.

4 PRACTICAL EXAMPLES

We demonstrate the practical properties of our approach using 1D inversion and two simple scenarios based on synthetic data calculated from a two-layered model (Figure 5). The true model consists of a 2 km thick layer with a resistivity of 100 Ωm underlain by a 1 Ωm half-space. The 1D inversion code is based on the algorithm described in (Avdeeva & Avdeev 2006) modified to accommodate the inverse data covariance matrix.

For one test dataset we shift the apparent resistivities at Periods $>$ 10 s (left two panels in Figure 5) and assume large errors for apparent resistivity in the inversion. In the other case (right panels in Figure 5) we set the phase constant at long periods and assume large errors for phase. If we invert the data with shifted apparent resistivity with small errors for both phase and apparent resistivity, the inversion tries to find a compromise between matching the conflicting information from the two quantities (green line) as expected. As our starting model consists of 26 layers with a resistivity of 100 Ωm, we match the high-frequency part of both curves, but have a significant discrepancy at longer periods. If we down-weight the apparent resistivity by prescribing a relative error of 200%, we correctly reproduce the phase curve within the specified error, but the inversion ignores the shift in apparent resistivity at long-periods (red line). It is well known that in 1D impedance phase has no sensitivity to the absolute value of resistivity but permits a large range of resistivity-thickness combinations (e.g. Oldenburg 1979). The good match with the unshifted apparent resistivities (black line) in this case is only due to the fact that we used the correct resistivity for the upper layer in the starting model. Of course, prescribing relative errors of 200% means that we completely ignore the apparent resistivity information and in practice data with such a severe shift could not be trusted to have reliable phase information. The main purpose of this experiment is to demonstrate the feasibility of our approach in general. How to choose the relative weighting depends on the dataset under investigation and its specific properties.

To demonstrate how to deal with problematic phase data, we use the same original synthetic data and set the phase at periods $>$ 100 s to a constant value (right panels in Figure 5). We now use a

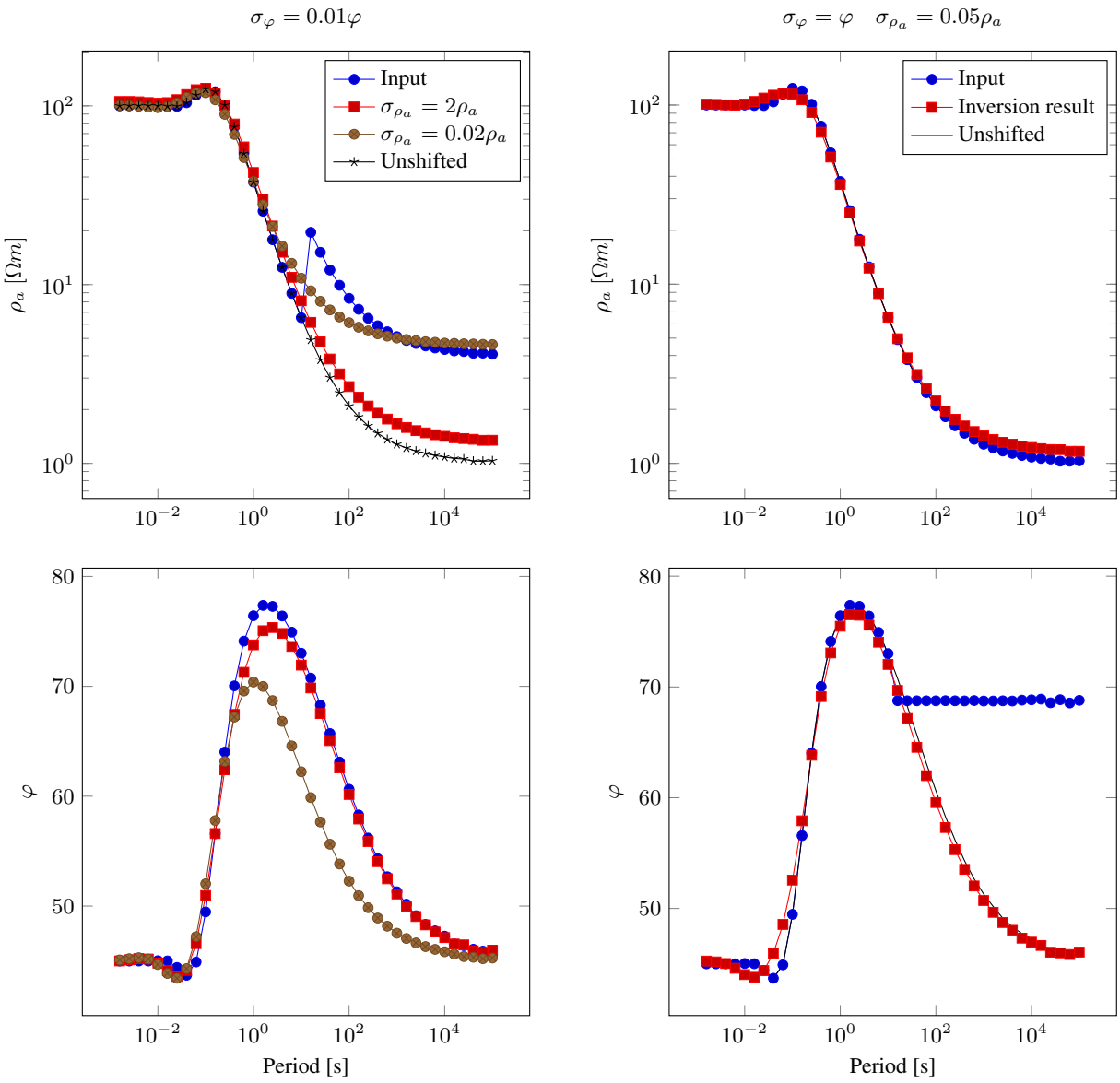

Figure 5. An example of using covariances to fit apparent resistivity and phase only. We feed synthetic data (blue symbols) with severely distorted apparent resistivity (left) and phase (right) into the inversion. When using covariances that simulate large error bounds on apparent resistivty (red line left) the inversion ignores the distorted information and fits the undistorted phase. When using large error bounds on phase (red line right), the phase information is ignored and we fit only apparent resistivity.

relative error of 100% for the phase and 5% for apparent resistivity. We can see that the inversion now honours the apparent resistivity information, but ignores the phase. As these two do not provide independent information, but are related by a dispersion relationship (Yee & Paulson 1988), fitting apparent resistivity results in a phase that also fits the unmodified data (black line). For both the

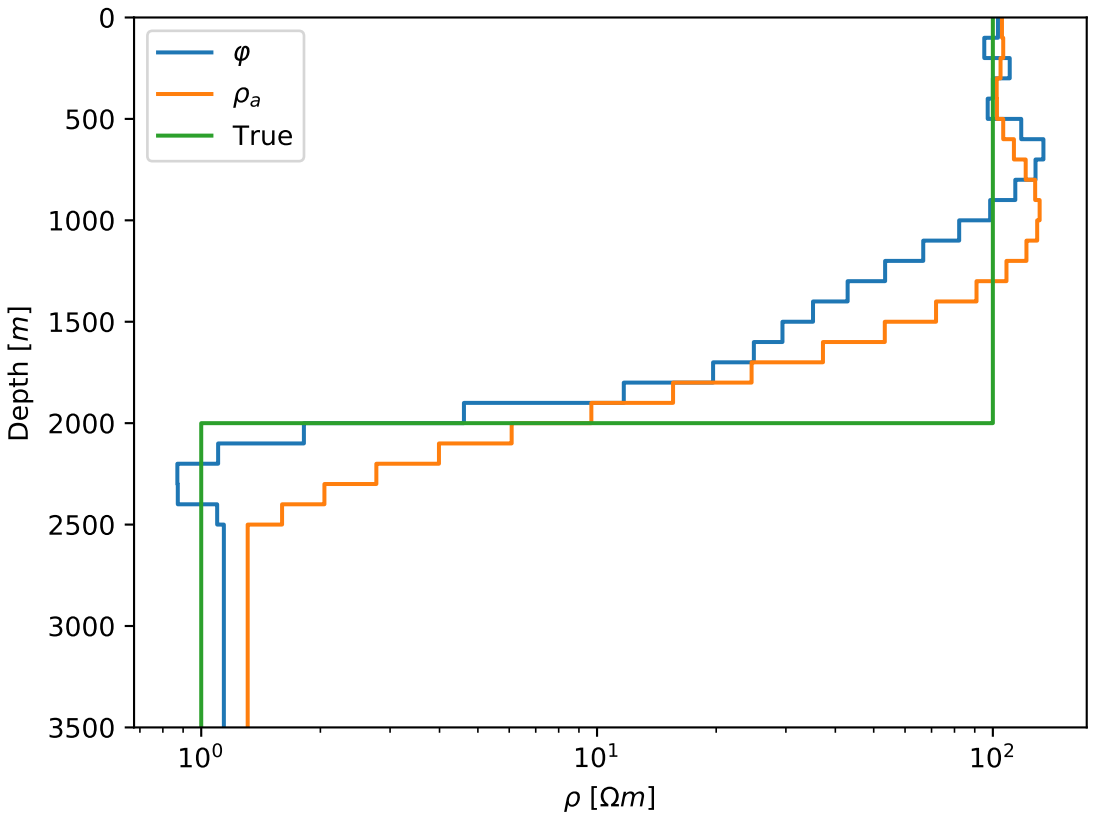

Figure 6. Comparison of the inversion results with the true model when inverting the shifted phase data (φ) and the shifted apparent resistivity (ρ_a) with large error bounds emulated through the covariance.

shifted apparent resistivity and the shifted phase, the inversion models match the true model to the degree expected for magnetotellurics as shown in Figure 6.

To complete our analysis we investigate how the specified covariance in impedance space maps back into apparent resistivity and phase. It is clear that because of the non-linear mapping, the confidence ellipse in impedance only approximates the situation of independent errors in apparent resistivity and phase. Figure 7 shows how the acceptable region in impedance space (left) maps into apparent resistivity and phase (right). Here we construct the impedance covariance matrix assuming standard deviations $\sigma_\varphi = 0.1\varphi$ and $\sigma_{\rho_a} = 0.05\rho_a$ for apparent resistivity and phase, respectively (left) and calculate how points within the elliptical region map back into apparent resistivity and phase (right). Compared to the red boxed region that describes the originally specified confidence area, we can see some differences. In particular, the remapped confidence area does not allow extreme apparent resistivity and phase combinations. Generally the covariances describe the confidence area well, especially when considering that in practical inversion we will use such confidence intervals for derived quantities to overcome certain deficiencies in the data and not because we have statistically firm reasons for it.

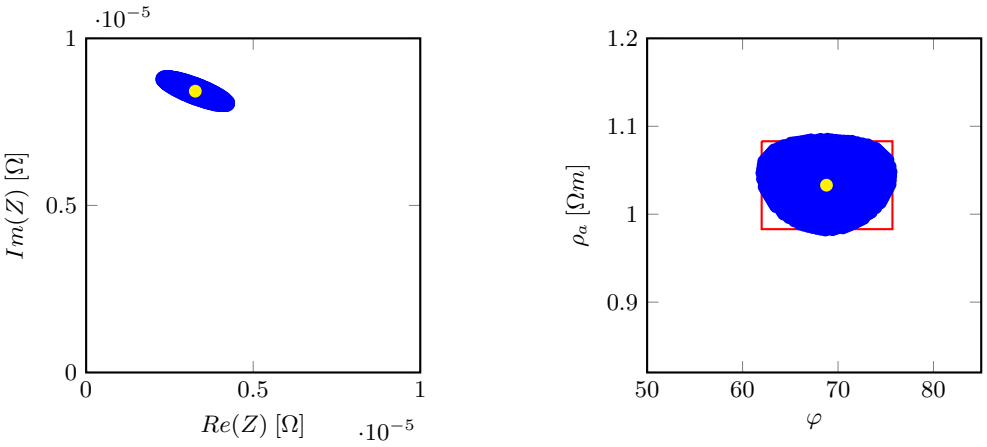

Figure 7. The true mapping of the confidence area specified by the covariance. We use relative errors of $\sigma_\varphi = 0.1\varphi$ and $\sigma_{\rho_a} = 0.05\rho_a$ to calculate the covariance matrix for impedance. We show all points around the impedance estimate (yellow) that fall within the error bounds (blue region left). We then map the points inside this region back into apparent resistivity and phase (blue dots right). Compared to the originally specified error bounds (red box), we approximate the intended region well.

Our approach is not limited to apparent resistivity and phase, but can be applied to virtually any transformation of the impedance tensor. As a final example we show how this can be achieved for the Berdichevsky invariant

$$Z_B = \frac{Z_{xy} - Z_{yx}}{2}. \tag{27}$$

This simple transformation also demonstrates the equivalence for linear transformations, discussed above. As we need to express Z_{xy} and Z_{yx} in terms of our new parameters we need a second complex variable,

$$Z_F = \frac{Z_{xy} + Z_{yx}}{2} \tag{28}$$

as an auxiliary quantity. In this case we can write

$$Z_{xy} = Z_B + Z_F \text{ and } Z_{yx} = Z_F - Z_B. \tag{29}$$

The 4×4 inverse of the covariance matrix corresponding to the off-diagonal elements of impedance is

$$\mathbf{C}^{-1} = \begin{pmatrix} \frac{\sigma_F^2+\sigma_B^2}{4\sigma_B^2\sigma_F^2} & 0 & -\frac{\sigma_F^2-\sigma_B^2}{4\sigma_B^2\sigma_F^2} & 0 \\ 0 & \frac{\sigma_F^2+\sigma_B^2}{4\sigma_B^2\sigma_F^2} & 0 & -\frac{\sigma_F^2-\sigma_B^2}{4\sigma_B^2\sigma_F^2} \\ -\frac{\sigma_F^2-\sigma_B^2}{4\sigma_B^2\sigma_F^2} & 0 & \frac{\sigma_F^2+\sigma_B^2}{4\sigma_B^2\sigma_F^2} & 0 \\ 0 & -\frac{\sigma_F^2-\sigma_B^2}{4\sigma_B^2\sigma_F^2} & 0 & \frac{\sigma_F^2+\sigma_B^2}{4\sigma_B^2\sigma_F^2} \end{pmatrix} \overset{\sigma_F^2 \gg \sigma_B^2}{\approx} \frac{1}{4\sigma_B^2} \begin{pmatrix} 1 & 0 & -1 & 0 \\ 0 & 1 & 0 & -1 \\ -1 & 0 & 1 & 0 \\ 0 & -1 & 0 & 1 \end{pmatrix} \tag{30}$$

Here the approximation in Equation 30 is made for the case that the covariance σ_F^2 for the auxiliary

quantity is much larger than the covariance for the Berdichevsky invariant. Conversely, if $\sigma_B^2 = \sigma_F^2 = \frac{1}{2}\sigma_Z^2$, we can see that the inverse data covariance matrix becomes diagonal with all elements on the diagonal equal to $\frac{1}{\sigma_Z^2}$.

The first case ($\sigma_F^2 \gg \sigma_B^2$) corresponds to the case where we invert for the invariant instead of the individual impedances. As we have shown above, the result is equivalent to an objective function formulated in terms of the invariant. In the second case ($\sigma_B^2 = \sigma_F^2$) we invert two quantities that are linear combinations of the original data and this is equivalent to inverting the original data with identical variances as expected.

5 DISCUSSION AND CONCLUSIONS

We have shown a simple methodology to use non-diagonal covariance matrices to emulate the inversion of quantities calculated from the original data. To illustrate the approach we have used magnetotelluric apparent resistivity and phase and the Berdichevsky invariant. However, the approach is completely general and it can easily be applied to other quantities such as phase tensors (Caldwell et al. 2004) in MT or invariants of the gravity gradient tensor. Even though we have to specify a large data covariance matrix and calculate its inverse, the approach is computationally feasible as in practice the matrix will be sparse and have a large number of small independent blocks that can be inverted independently.

Using apparent resistivity as an example, we have shown that inverting quantities calculated from non-linear functions of the observations can result in biased model estimates. This bias is statistically significant and thus relevant from a theoretical point of view. However, it can be argued that in practice the bias is small enough to have little to no effect on the interpretation of the results. The apparent resistivies shown in Figure 3 show errorbars of $\pm 20\Omega m$ and the resulting bias for the estimated half-space resistivity is only $6\Omega m$. Given such noisy data, any experienced practitions would only use the resulting model as an order of magnitude estimate and would not try to discern variations of tens of Ωm. Even in these situations it is of advantage though to use an unbiased inversion approach particularly if we want to move towards statistical evaluation of the results and determine model uncertainties. These issues become even more prevalent with quantities such as the phase tensor that is calculated as a ratio of random variables and where it has been shown that care needs to be taken to calculate uncertainties (Chave 2013; Booker 2014).

Our experiments with inverting shifted data have shown that the emulation of the derived quantities works well and approximates the specified uncertainty bounds sufficiently for practical applications. Furthermore, using non-diagonal covariances the user is not limited by the quantities envisaged during code development, but can emulate arbitrary quantities once the inversion algorithm supports the

specification of a covariance. Given how easy it is to implement this scheme in practical inversion methods, we believe that it should become a standard feature.

ACKNOWLEDGMENTS

We would like to thank the editor Gary Egbert, as well Alexander Grayver and an anonymous reviewer for helpful comments that improved the quality of the manuscript. This work was funded by the German Research Foundation, DFG under grant MO 2265/4-1.

REFERENCES

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

Avdeeva, A. & Avdeev, D., 2006. A limited-memory quasi-newton inversion for 1d magnetotellurics, *Geophysics*, **71**(5), G191–G196.

Avdeeva, A., Moorkamp, M., Avdeev, D., Jegen, M., & Miensopust, M., 2015. Three-dimensional inversion of magnetotelluric impedance tensor data and full distortion matrix, *Geophysical Journal International*, **202**(1), 464–481.

Booker, J. R., 2014. The magnetotelluric phase tensor: a critical review, *Surveys in Geophysics*, **35**(1), 7–40.

Bregman, N., Bailey, R., & Chapman, C., 1989. Crosshole seismic tomography, *Geophysics*, **54**(2), 200–215.

Caldwell, T. G., Bibby, H. M., & Brown, C., 2004. The magnetotelluric phase tensor, *Geophysical Journal International*, **158**, 457–469.

Chave, A. & Jones, A., 2012. *The magnetotelluric method*, Cambridge University press, Cambridge, UK.

Chave, A. & Lezaeta, P., 2007. The statistical distribution of magnetotelluric apparent resistivity and phase, *Geophysical Journal International*, **171**(1), 127–132.

Chave, A. D., 2013. On the statistics of magnetotelluric rotational invariants, *Geophysical Journal International*, **196**(1), 111–130.

Chave, A. D., 2017. Estimation of the magnetotelluric response function: the path from robust estimation to a stable maximum likelihood estimator, *Surveys in Geophysics*, **38**(5), 837–867.

Chave, A. D. & Thomson, D. J., 2004. Bounded influence magnetotelluric response function estimation, *Geophysical Journal International*, **157**, 988–1006.

Chave, A. D., Thomson, D. J., & Ander, M. E., 1987. On the robust estimation of power spectra, coherences and transfer functions, *Journal of Geophysical Research*, **92**(B1), 633–648.

Constable, C., 1988. Parameter estimation in non-gaussian noise, *Geophysical Journal International*, **94**(1), 131–142.

Di Stefano, R., Aldersons, F., Kissling, E., Baccheschi, P., Chiarabba, C., & Giardini, D., 2006. Automatic

seismic phase picking and consistent observation error assessment: application to the italian seismicity, *Geophysical Journal International*, **165**(1), 121–134.

Egbert, G. D., 2002. Processing and interpretation of electromagnetic induction array data, *Surveys in Geophysics*, **23**, 207–249.

Eisel, M. & Egbert, G. D., 2001. On the stability of magnetotelluric transfer function estimates and the reliability of their variances, *Geophysical Journal International*, **144**, 65–82.

Farquharson, C. G. & Oldenburg, D. W., 1998. Non-linear inversion using general measures of data misfit and model structure, *Geophysical Journal International*, **134**, 213–227.

Farquharson, C. G. & Oldenburg, D. W., 2004. A comparison of automatic techniques for estimating the regularization parameter in non-linear inverse problems, *Geophysical Journal International*, **156**, 411–425.

Günther, T., Rcker, C., & Spitzer, K., 2006. Three-dimensional modelling and inversion of dc resistivity data incorporating topography II. Inversion, *Geophysical Journal International*, **166**(2), 506–517.

Heath, P. J., 2007. *Analysis of potential field gradient tensor data : forward modelling, inversion and near - surface exploration*, Ph.D. thesis, University of Adelaide.

Hwang, C., Hsiao, Y.-S., Shih, H.-C., Yang, M., Chen, K.-H., Forsberg, R., & Olesen, A. V., 2007. Geodetic and geophysical results from a taiwan airborne gravity survey: Data reduction and accuracy assessment, *Journal of Geophysical Research*, **112**(B4).

Jennrich, R. I., 1969. Asymptotic properties of non-linear least squares estimators, *The Annals of Mathematical Statistics*, **40**(2), 633–643.

Khoza, T. D., Jones, A. G., Muller, M. R., Evans, R. L., Miensopust, M. P., & Webb, S. J., 2013. Lithospheric structure of an archean craton and adjacent mobile belt revealed from 2-d and 3-d inversion of magnetotelluric data: Example from southern Congo craton in northern Namibia, *Journal of Geophysical Research: Solid Earth*, **118**(8), 4378–4397.

Mackie, R. L. & Madden, T. R., 1993. Three-dimensional magnetotelluric inversion using conjugate gradients, *Geophysical Journal International*, **115**, 215–229.

Mackie, R. L., Madden, T. R., & Wannamaker, P. E., 1993. Three-dimensional mangnetotelluric modeling using difference equations - Theory and comparisons to integral equation solutions, *Geophysics*, **58**, 215–226.

Moorkamp, M., Heincke, B., Jegen, M., Roberts, A. W., & Hobbs, R. W., 2011. A framework for 3-D joint inversion of MT, gravity and seismic refraction data, *Geophysical Journal International*, **184**, 477–493.

Moorkamp, M., Fishwick, S., Walker, R. J., & Jones, A. G., 2019. Geophysical evidence for crustal and mantle weak zones controlling intra-plate seismicity–the 2017 botswana earthquake sequence, *Earth and Planetary Science Letters*, **506**, 175–183.

Mosegaard, K. & Hansen, T. M., 2016. *Inverse Methods*, pp. 7–27, John Wiley & Sons, Inc.

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

Oldenburg, D. W., 1979. One-dimensional inversion of natural source magnetotelluric observations, *Geo-*

physics, **44**(7), 1218–1244.

Rao, C., Jones, A. G., Moorkamp, M., & Weckmann, U., 2014. Implications for the lithospheric geometry of the Iapetus suture beneath Ireland based on electrical resistivity models from deep-probing magnetotellurics, *Geophysical Journal International*, **198**(2), 737–759.

Simpson, F. & Bahr, K., 2005. *Practical magnetotellurics*, Cambridge University press, Cambridge, UK.

Tarantola, A., 2004. *Inverse Problem Theory*, SIAM, 1st edn.

Tellinghuisen, J., 2000. Bias and inconsistency in linear regression, *The Journal of Physical Chemistry A*, **104**(50), 11829–11835.

Tellinghuisen, J., 2001. Statistical error propagation, *The Journal of Physical Chemistry A*, **105**(15), 3917–3921.

Tietze, K. & Ritter, O., 2013. Three-dimensional magnetotelluric inversion in practice - the electrical conductivity structure of the san andreas fault in central california, *Geophysical Journal International*.

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

Worzewski, T., Jegen, M., Kopp, H., Brasse, H., & Taylor Castillo, W., 2011. Magnetotelluric image of the fluid cycle in the Costa Rican subduction zone, *Nature Geoscience*, **4**(2), 108–111.

Yee, E. & Paulson, K. V., 1988. Concerning dispersion relations for the magnetotelluric impedance tensor, *Geophysical Journal*, **95**(3), 549–559.