Geophysical Journal International

*Geophys. J. Int.* (2014) **199,** 242–252 doi: 10.1093/gji/ggu258
GJI Marine geosciences and applied geophysics

# Magnetotelluric inversion based on mutual information

Eric Mandolesi and Alan G. Jones

*Dublin Institute for Advanced Studies, Geophysics Section, School of Cosmic Physics, 5 Merrion Square, Dublin 2, Ireland. E-mail: eric@cp.dias.ie*

Accepted 2014 July 3. Received 2014 July 2; in original form 2013 December 2

## SUMMARY

Joint inversion of different geophysical data sets is becoming a more popular and powerful tool, and it has been performed on data sensitive both to the same physical parameter and to different physical parameters. Joint inversion is undertaken to reduce acceptable model space and to increase sensitivity to model parameters that one method alone is unable to resolve adequately. We examine and implement a novel hybrid joint inversion approach. In our inversion scheme a model—the reference model—is fixed, and the information shared with the subsurface structure obtained from another method will be maximized; in our case conductivity structures from magnetotelluric (MT) inversion. During inversion, the joint probability distribution of the MT and the specified reference model is estimated and its entropy minimized in order to guide the inversion result towards a solution that is statistically compatible with the reference model. The powerful feature of this technique is that no explicit relationships between estimated model parameters and reference model ones are presumed: if a link exists in data then it is highlighted in the estimation of the joint probability distribution, if no link is required, then none is enforced. Tests performed verify the robustness of this method and the advantages of it in a 1-D anisotropic scenario are demonstrated. A case study was performed with data from Central Germany, effectively fitting an MT data set from a single station within as minimal an amount of anisotropy as required.

**Key words:** Inverse theory; Magnetotellurics; Europe.

## 1 INTRODUCTION

Abrupt, or relatively abrupt, changes in physical and chemical properties in the lithosphere are usually associated with Earth's structural discontinuities, with the primary plate tectonic discontinuities in the upper mantle above the 410 km discontinuity being the crust–mantle boundary (usually referred as the Mohorovicic discontinuity, or Moho, although the petrological crust–mantle boundary may not necessarily be the same as the seismically defined boundary, e.g. Giese *et al.* 1999; Janik *et al.* 2009) and the lithosphere–asthenosphere boundary (LAB; which has its own set of intricacies, see e.g. Eaton *et al.* 2009). These two boundaries are characterized by different fundamental changes: the Moho is characterized by a change in composition between the crust and the mantle from mafic to ultramafic compositions, whereas the LAB *sensu stricto* is governed by a rheological change between the rigid lithosphere and viscous asthenosphere (Barrell 1914; Parsons & McKenzie 1978), although different geophysical and geochemical proxies exist (e.g. Eaton *et al.* 2009). The cause of the LAB discontinuity in geophysical proxies is hotly debated, and is presumed to be the onset of partial melting or the rapid increase in water content, or both. These interfaces, as well as local lateral interfaces, are imaged with different accuracy and precision by different geophysics techniques, and are completely ignored by others; see recent overviews by Cook *et al.* (2010) for the Moho and Eaton *et al.* (2009) for the LAB.

This observation drove us to develop an inversion scheme for magnetotelluric (MT) data that is able to handle information obtained from other geophysical, or indeed geological or geochemical, data sets without any geometric link in the models obtained, as in for example Gallardo & Meju (2003, 2007), Moorkamp *et al.* (2007, 2010) and Roux *et al.* (2011), or any parametric physical link.

Anisotropy, the directional dependence of material properties, plays an essential role as an indicator of heterogeneity and petrological differentiation that characterize a certain region: anisotropy, its presence, absence or magnitude, particularly electrical and seismic anisotropy, constitutes an essential parameter for investigating lithospheric fabrics. Approximate agreement between geo-electric strike and seismic fast-axis direction in several regions, for example southern part of the Slave province (Eaton *et al.* 2004), central Australia (Simpson 2001), central Germany (Gatzemeier & Moorkamp 2005), the Kaapvaal craton (Hamilton *et al.* 2006) and the São Francisco craton (Padilha *et al.* 2006), suggests that a common origin may contribute to both electrical and seismic anisotropy. Usually, lattice-preferred orientation of olivine crystals is considered to be the source of seismic anisotropy in the upper mantle (e.g. Savage 1999), but laboratory measurements infer that electrical anisotropy caused by the orientation of olivine crystals cannot be the only source of the high levels of anisotropy observed in the field (Gatzemeier & Tommasi 2006).

While the source of upper-mantle electric anisotropy is still debated, its effect is usually recognizable in MT data either in form of distortion or by other peculiar indication (e.g. non-zero main diagonal elements in the impedance tensor with a null tipper vectors, indicative of a 1-D anisotropic environment). The qualitative agreement, as detailed above, suggests that an MT model can be constrained effectively by a seismic model, thereby reducing the non-uniqueness inherent the solution of a 1-D anisotropic MT inverse problem.

We note here that in 1-D for perfect data at all frequencies there exists only one model that fits the MT data (Bailey 1970), so MT is inherently unique and non-uniqueness is a function of data error, imprecision and insufficiency.

In order to reduce this non-uniqueness inherent the MT problem, we employ in our inversion scheme mutual information (MI), a quantity used in information theory that defines a distance metric between two images, as a functional link between two models. Maximizing MI has been used in object recognition, image registration to find and evaluate the alignment of model and image data (Viola & Wells 1997), and in medical tomography to compare and align tomographic maps obtained from different methods (Collignon *et al.* 1995), to name but three applications of many tens. Recently, Haber & Holtzman Gazit (2013) proposed the use of MI in joint inverse problem solution, but did not give any examples of its use.

The overarching idea that drives our inversion scheme is that tomographic models derived from the subsurface obtained by different methods are images whose relative distance apart can be measured and minimized. In detail, we invert MT impedance data keeping the resulting model close, in an MI sense, to a defined reference model of the same region. The introduced measure of distance is not only intuitive, but it respects all the properties a definition of distance is required to uphold.

In this scheme the electrical conductivity of the subsurface and the reference model are not forced to be similar in a geometrical sense, but rather they are interpreted as images or pictures of the subsurface and are kept close during the optimization path via the minimization of the measure of a distance between pictures defined. The closer the two pictures are the greater is the information shared about the source—the subsurface—that generates the models themselves.

Without the possibility to compare our algorithm results with analytical exact solutions, we performed both synthetic tests and a case study to investigate the applicability of the inverse scheme with real data from a well-studied region. With respect to the real world, we test our inversion scheme on MT data from Central Germany that have already been studied in detail. Since both velocity and electrical conductivity models are available for this region, we decided to use seismic models as the reference for our inversion scheme and compare our results with the MT models of this region available in literature. We could equally assume that the MT models are the reference model and invert the seismic data discovering models close, in the MI sense, to those reference MT models.

The studied region is characterized by two principal anisotropic layers, with the seismic fast-axis direction approximately aligned with the NE–SW and E–W directions related to the lithosphere and asthenosphere, respectively (Roux *et al.* 2011). The inversion of the MT data from the area produces an electrical conductivity model close—in a MI sense—to the seismic one, and a joint probability distribution that can be interpreted as a relation that links electrical conductivity and seismic velocity in the study area. Moreover the final model improves the compatibility of anisotropic electrical conductivity values characterizing the model with the values of rock samples measured in laboratory.

## 2 INVERSION METHOD AND DATA

The MT technique is broadly used in electrical tomography of the Earth at many scales, from 100s m to 100s km, and its characterizing equations, derived from the classic Maxwell's equations, are well known and solved in several different domains. The subject of this work is the solution of the MT inverse problem in a one-dimension anisotropic domain as an exemplar of the strengths and advantages of the MI approach to joint inversion. In this section, we present the setting of the problem from a mathematical perspective, as well as the synthetic and real data used for the algorithm tests, while in the Section 3 we will present the results of both synthetic tests and Central Germany data inversion.

### 2.1 Problem setting: the MI constraint

There already exist algorithms to invert MT 1-D anisotropic data (e.g. Abramovici & Shoham 1970; Pek & Santos 2006). The principal difference we introduce in our inversion scheme is the minimization of the distance with respect to a general reference model, formulated in term of MI (Haber & Holtzman Gazit 2013). MI is a fundamental quantity defined in information theory and is applied in several techniques. The MI feature used in the context of this paper is that it is possible define a distance measure between $n$-dimensional images via MI (*cf.* e.g. Modersitzki 2004). In its simplest form, the MI between two images $\bar{x}$ and $\bar{y}$ is defined as

$$I(\bar{x};\bar{y}) := \sum_{x\in\bar{x}}\sum_{y\in\bar{y}} p(x,y)\log_2\frac{p(x,y)}{p(x)p(y)}, \tag{1}$$

where $\bar{\star}$ is an array storing the image voxel values, $p(\star)$ is the marginal probability distribution of the random variable $\star$, and $p(\star_1, \star_2)$ the joint probability distribution of the two random variables $\star_1, \star_2$.

In the Bayesian approach to inverse problems, probability density functions are used to find the optimal solution for an optimization problem. We prefer to clarify that in our approach the probability density functions are estimated via kernel estimation (Silverman 1986) and used to define a penalty function

$$\phi_I := [b - I(\bar{x};\bar{y})], \tag{2}$$

where $b$ is a constant value that makes $\phi_I \geq 0$, to be used in a Newton-type algorithm. As $I$ is used here as a measure of distance between images (*cf.* the Appendix for details), the presence of $b$ does not influence the shape of the objective function, but only on the relative weight of $\phi_I$ in the minimized objective function.

We opt for this choice in consideration of the extension of our scheme to two- and three-dimensions, environments in which the computation cost required by the Bayesian approach becomes barely affordable.

We thus use the kernel estimation method described in Silverman (1986), namely, the marginal probability distribution $p(u)$ is approximated by its estimation $p_s(u)$ that depends on the chosen kernel function $K(s,t)$, where $s > 0$ is a parameter that controls the width of the kernel function and $\int_{-\infty}^{+\infty} K(s,t)\mathrm{d}t = 1\ \forall s > 0$.

With the chosen method the probability density function is thus given by

$$p(x) \simeq p_s(x) = \frac{1}{\mathcal{N}s}\sum_{i=1}^{\mathcal{N}} K\left(\frac{x-\bar{x}_i}{s}\right) \tag{3}$$

and,

$$p(u) \simeq p_s(u) = \frac{1}{\mathcal{N}} \sum_{i=1}^{\mathcal{N}} K_H(u - \bar{u}_i), \tag{4}$$

where $\mathcal{N}$ is the number of classes used to discretize the random variable, $u = [x, y]^T$, $\bar{u}_i = [\bar{x}_i, \bar{y}_i]^T$ and

$$K_H = |H|^{-\frac{1}{2}} K(s, H^{-\frac{1}{2}}u), \text{ with } H = \begin{bmatrix} s^2 & 0 \\ 0 & s^2 \end{bmatrix}. \tag{5}$$

In this work we chose to follow Haber (2004, personal communication) and estimate the marginal and joint probability distributions using a kernel function $K$ defined as

$$K(s, t) := \begin{cases} \frac{4}{3s} \cos^4\left(\frac{\pi t}{2s}\right), & \text{if } -s < t < s \\ 0, & \text{otherwise.} \end{cases}, \tag{6}$$

In theory any symmetric function that integrates to the unity can be chosen as the kernel, but it is common practice to select a bell function that localizes the probability within bandwidth $s$. The kernel function used in this paper has the notable property to vanish outside the interval $[-s; +s]$, and has left and right derivatives with respect to $t$ equal and continuous in $\pm s$. Whereas these properties are not fundamental in the computation of the marginal probability distribution of a random variable, they become important in the computation of its derivative with respect to $t$. The adopted kernel function has thus the advantage of being continuous and derivable along its whole definition domain, with known derivative with respect to $t$. The knowledge of this analytical form of $\frac{\partial K}{\partial t}$ speeds up the computation of the Jacobian of the objective function used in the Newton-type minimization algorithm implemented.

The objective function we minimize is therefore

$$\phi(\bar{m}, \bar{R}) := \sum_i \left[\frac{d_i - F(\bar{m})_i}{\sigma_i}\right]^2 + \lambda_I^2 \phi_I^2(\bar{m}, \bar{R}), \tag{7}$$

where $\bar{m}$ is the vector storing the MT model parameters, $\bar{R}$ is the reference model, $\bar{d}$ is the data array, $\bar{\sigma}$ is the error array, $\bar{F}(\bar{\star})$ is the array of data predicted by the model $\bar{\star}$ and $\lambda_I$ is a fixed parameter that controls the trade-off between data fitting and minimization of MI.

### *2.1.1 An example*

Let us suppose we wish apply the described scheme to build an objective function to be used for an inversion that retrieves the values of the $x$ parameter to be constrained by the knowledge of the values of the $y$ parameter in the subsurface. The computation of $\phi_I$ requires the knowledge of the marginal probability distributions $p(x)$ and $p(y)$ and the joint probability distribution $p(x, y)$. Let us assume that $x$ may vary in certain interval $[x_m, x_M]$ with $x_m < x_M$. The number of classes used to discretize this interval is $\mathcal{N}$, so that the $i$th class is bounded by its limits $\left[\frac{i-1}{\mathcal{N}}; \frac{i}{\mathcal{N}}\right]$. It is now possible to estimate the three probability distributions required by substituting eq. (6) in eq. (5) and eq. (5) in eqs (3) and (4). With the knowledge (or an implementation) of the forward operator $F$ and the possibility to set values for $b$ and $\lambda_I$, the construction of the objective function $\phi(\bar{m} = x, \bar{R} = y)$ is concluded and the inverse problem set up is complete.

## 2.2 Problem setting: the inversion scheme

As eq. (7) is the sum of quadratic terms, we minimize it using the Levenberg–Marquardt algorithm (Marquardt 1963), a damped version of the classic Gauß–Newton algorithm, developed for quadratic programming. Damping is generally required both to stabilize the linear system and to avoid reaching a region of model space too far away with respect to the starting guess model $\bar{m}_0$. We define the vector

$$\bar{\Phi} = \begin{bmatrix} \Phi_1 \\ \Phi_2 \\ \dots \\ \Phi_N \\ \Phi_{N+1} \end{bmatrix}, \text{ with } \begin{cases} \Phi_i = \left[\frac{d_i - F(\bar{m})_i}{\sigma_i}\right], i = 1, \dots, N. \\ \Phi_{N+1} = \lambda_I \phi_I(\bar{m}, \bar{R}). \end{cases} \tag{8}$$

The system of equations we solved is then, at the $k+1$th iteration

$$\left[J(\bar{m}_k)^T J(\bar{m}_k) + \lambda I\right](\bar{m}_{k+1} - \bar{m}_k) = -J^T(\bar{m}_k)\bar{\Phi}(\bar{m}_k), \tag{9}$$

where

$$J(\bar{m}_k) = J_k = \left.\frac{\partial \bar{\Phi}}{\partial \bar{m}}\right|_{\bar{m}=\bar{m}_k}, \tag{10}$$

$\lambda$ is the damping factor, chosen by line search at each iteration, as suggested by Aster *et al.* (2005), and $I$ is the identity matrix. The $k+1$th model is thus expressed by

$$\bar{m}_{k+1} = \left[J_k^T J_k + \lambda I\right]^{-1} \left[-J_k^T \bar{\Phi}(\bar{m}_k)\right] + \bar{m}_k. \tag{11}$$

We chose, as the criterion to stop the iterative process, to define a threshold value $\varepsilon$ and we interrupt the iterative process as soon as $\|\bar{\Phi}(\bar{m}_{k+1})\| - \|\bar{\Phi}(\bar{m}_k)\| \leq \varepsilon$ [i.e. $\phi(\bar{m}_{k+1}, \bar{R}) - \phi(\bar{m}_k, \bar{R}) \leq \varepsilon$].

So far, the discussion relative to the inverse algorithm, as well as the one related to the MI constraint, has been general and applicable to a broad range of problems. In order to invert MT data, we need to specify the behaviour of $F(\star)$ in eq. (7). We chose to use the algorithm described by Pek & Santos (2002) in order to estimate the MT impedance tensor relative to each considered period. We thus predict four complex impedance tensor elements per frequency: four real parts and four imaginary parts.

In the local approximation, the Earth is modelled as a layered half–space, and each layer is characterized by two aggregated conductivities, a high electrical conductivity value $\sigma_h$ and a low electrical conductivity value $\sigma_l$ and the direction of the high conductivity value $\beta_s$. In general, it is possible to express the electrical conductivity tensor as a symmetric, positive defined tensor, but the MT problem in the anisotropic 1-D case is inherently singular (Pek & Santos 2002; Yin 2003). We thus inverted $8 \cdot n_f$ data, the four real and the four imaginary parts of the impedance tensor relative to each frequency, to retrieve the $n_l \cdot 3$ elements of the model, where $n_f$ is the number of considered frequencies and $n_l$ the number of layers (including the infinite basement on the bottom of the half-space) that describe, locally and in 1-D, the subsurface electrical conductivity. As the common anisotropic direction for EM and seismic model has been highlighted and there are no evidence of common sources between shear velocity and electrical conductivity that would lead to a parametric relationship between them, we decided to apply the MI constraint only to the anisotropic strike direction $\beta_s$, leaving the other parameters unconstrained.

The following sections are organized as:

Section 2.3: a description of the synthetic model and the reference models used for synthetic tests;

Section 2.4: a description of the data set used for the validation of this method in an appropriate region;

Section 3: the description of the result obtained by applying the described method to both the synthetic (Section 3.1) and the real (Section 3.2) data sets.

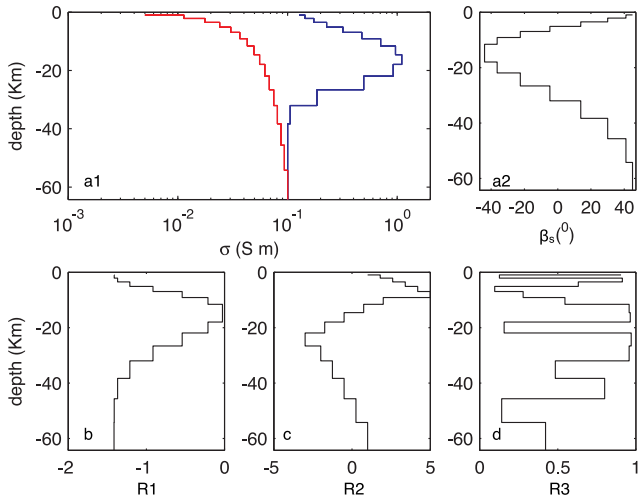


**Figure 1.** (a) Synthetic model high (red line) and low (blue line) conductivities (a1) and anisotropy direction (a2); (b) compatible reference model; (c) incompatible reference model and d) random reference model used in the synthetic tests. The reference models are here given in term of pure numbers, as the nature of the reference does not influence the algorithm.

### 2.3 Synthetic tests data

In Fig. 1 are reported the reference models we used for testing our approach as well as the synthetic model $\bar{m}_{\text{true}}$. We test the MI based constraint for three different cases:

(i) compatible reference model;
(ii) incompatible reference model and
(iii) randomly created reference model.

With these three different tests our intention is to ascertain if the impact of a wrong (consistent or inconsistent) reference model may severely influence the inversion by misdirecting the result towards an incorrect final model. We added noise to the synthetic data produced form $\bar{m}_{\text{true}}$ of the form of random Gaussian distributed error of 3 per cent of the amplitude of the highest tensor element relative to each frequency. As the aim of these tests is to check the behaviour of the algorithm, in replacement of the data, predicted and simulated, we report the first $N$ terms of the eq. (7). In all the tests performed we discretized the possible direction in eight values (i.e. $\mathcal{N} = 8$), the trade-off parameter $\lambda_I$ was set to 1 and the value of $b$ was set to 3.

### 2.4 Real data

The MT data used in this study were acquired in the Rhenish Shield region, Central Germany between 1997 June and September and 1998 October to December (Leibecker *et al.* 2002). From these surveys we analysed the data from station DIE, previously analysed and modelled by Roux *et al.* (2011). The location of the station and the topography of the surrounding region are shown on Fig. 2. Other stations in the area present the same characteristics, but are of inferior data quality (Roux *et al.* 2011). We considered data from station DIE at all periods within the period range 10–4100 s.

MT data are usually affected by distortion caused by small heterogeneities and local structure (Groom & Bailey 1989). Based on prior studies of this area (e.g. Leibecker *et al.* 2002; Roux *et al.*

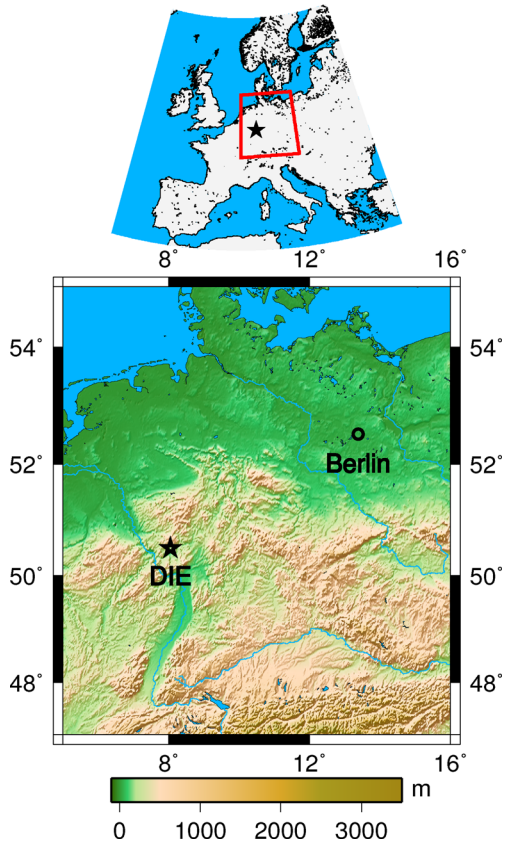


**Figure 2.** Topographic map of the studied area. The black star indicate the location of DIE station.

2011) we expect a 1-D anisotropic environment, both electrically and seismically. In order to model MT data in a 1-D anisotropic domain the determinable effects of galvanic charges on the electric field from local, small-scale, distorters have to be identified and removed. We pre-processed the data, as suggested in Jones (2012), to remove the determinable effects of galvanic distortion. As we use the same data, we refer to Jones (2012) for a complete discussion regarding the process undertaken to remove the determinable effects of galvanic charges on the electric field from local, small-scale, distorters. After the processing procedure, we obtained a set of 18 complex impedance tensors $\hat{\mathbf{Z}}(T)$, one per period, all with $\text{tr}(\hat{\mathbf{Z}}) = 0$; this formal property is expected in 1-D anisotropic environments (Pek & Santos 2002).

For the reference model, we chose one of the seismic models obtained by Roux *et al.* (2011). In that work, Roux *et al.* (2011) implemented a genetic algorithm (NAGA-II) as their optimization engine, and jointly inverted for a set of suitable and compatible seismic and resistivity anisotropic 1-D models. Our reference model is the one with the minimum RMS relative to the seismic problem, and is shown in Fig. 3.

[For information regarding the inversion technique used for the seismic data, see Roux *et al.* (2011) and references therein.]

We limit our MI constrain to the strike direction, making the assumption that the shear waves and the electrical anisotropy have a common source and thus that both directions share information about their common source. Thus, in eq. (7) the term $\phi_I(m, R)$ minimizes the distance between the one dimensional images given by the seismic fast-axis direction ($R$) and the electrical high-conductive direction ($m$).

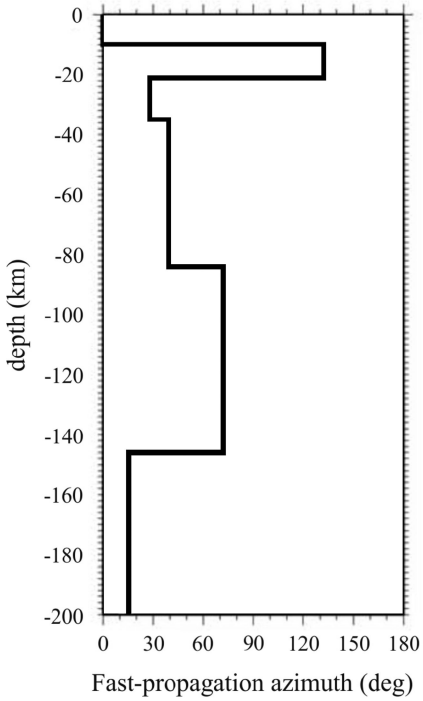


**Figure 3.** Seismic fast-axis direction from Roux *et al.* (2011; black line) used as reference model *R*.

## 3 RESULTS AND DISCUSSIONS

### 3.1 Synthetic tests

With these synthetic tests we aim to show both the effect of choice of different reference models on the retrieved model and to check that the (synthetic) data are reasonably well fit by the retrieved model. In Fig. 4, the terms of the objective function (7) relative to the data fit are reported. It is common in literature to report and display predicted and measured data, but for these synthetic tests we prefer

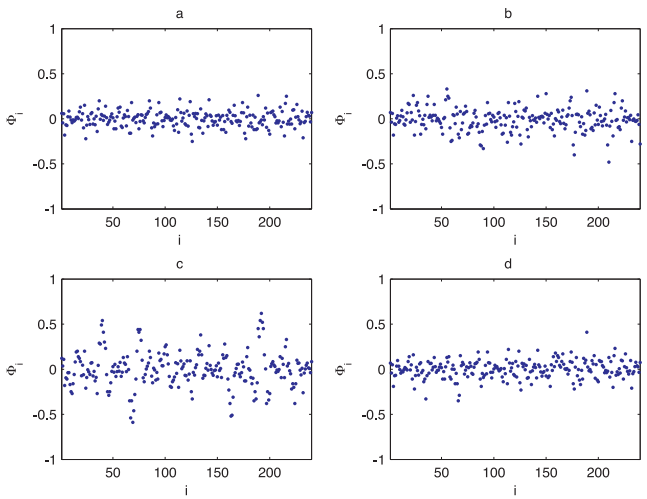


**Figure 4.** Objective function terms relative to the synthetic tests. In correspondence of the *i*th value of the *x*-axis is reported the value value of the *i*th term of the objective function as defined in eq. (8) (with $i \in [1, N]$). The graphics are relative to (a) unconstrained inverse problem; (b) inversion contained using the compatible reference model; (c) inversion constrained using the incompatible reference model and (d) inversion constrained using a random reference model.

**Table 1.** The rms relative to the synthetic tests.

| Reference type | Rms |
|---|---|
| No reference | 1.4 |
| Compatible | 1.9 |
| Incompatible | 3.0 |
| Random | 1.5 |

to show directly the objective function elements, as it is easier to compare these values with their expected value of 1. As can be seen, both from the graphics in Fig. 4 and from their root mean square reported in Table 1, the data fit is qualitatively similar. The unconstrained retrieved model predicts the data better (in the least square sense) than the constrained ones, as expected. Thus we can conclude that for these cases and using an isotropic half-space as the starting model for the algorithm, all the compatible reference model produce an acceptable data prediction, with a relevant difference in rms returned by the incompatible reference model.

Finally, in Fig. 5 are reported the results of the four inversions (the unconstrained one and the three constrained ones). The unconstrained solution, as the one constrained with a random reference model, presents an unstable profile, which is expected since MT inversion is a well-known example in the class of ill-posed inverse problems (Parker 1980). The constrained solutions are not dissimilar to the unconstrained ones, with the notable exception of the inversion constrained by the compatible reference model. These results demonstrate that the MI constraint term plays an active role in the inversion and prevents the final model from being too rough, without enforcing any similarity when the incompatible reference model is used.

### 3.2 Real data set from Central Germany

Notwithstanding the concerns about the use of linearized approaches found in, for example Tarantola (2006) and Moorkamp *et al.* (2007), we note that the validity for the use of a genetic algorithm (GA), as an engine for an optimization problem such as the one we aim to solve, is for problems that attempt to resolve model parameters numbering of the order of 30–40 (Mandolesi 2013, chap. 4). If the MT inverse problem requires a larger number of parameters, the GA loses its efficient convergence properties. Our goal is to work towards 2-D, and even 3-D, joint inversion, where the number of parameters becomes far too large for stochastic methods. Thus we need to explore methods that are more efficient, and we begin with the 1-D problem. As stated above, we discretize the model as described in Pek & Santos (2002), inverting for the high conductivity value $\sigma_{\mathrm{HI}}$, the low conductivity value $\sigma_{\mathrm{LO}}$ and the anisotropic strike direction $\beta_s$. A reminder again that here we have assumed that the anisotropy axes are horizontal and vertical, and that we cannot sense vertical anisotropy in 1-D in MT. Any deviation from this, such as dipping anisotropy, cannot, in any case, be resolved due to the non-uniqueness inherent the anisotropic MT 1-D problem (Yin 2003).

Assessing this non-uniqueness requires use of a Tikhonov regularization term in the objective function, in order to penalize models that are too rough (Constable *et al.* 1987). We added a quadratic term that penalizes the differences between the high and low conductivities, $\phi_A^2(\tilde{m}) = \sum_i (\log \sigma_{\mathrm{HI}i} - \log \sigma_{\mathrm{LO}i})^2$, to penalize models that exhibit high total (summed) anisotropy, and a term to introduce

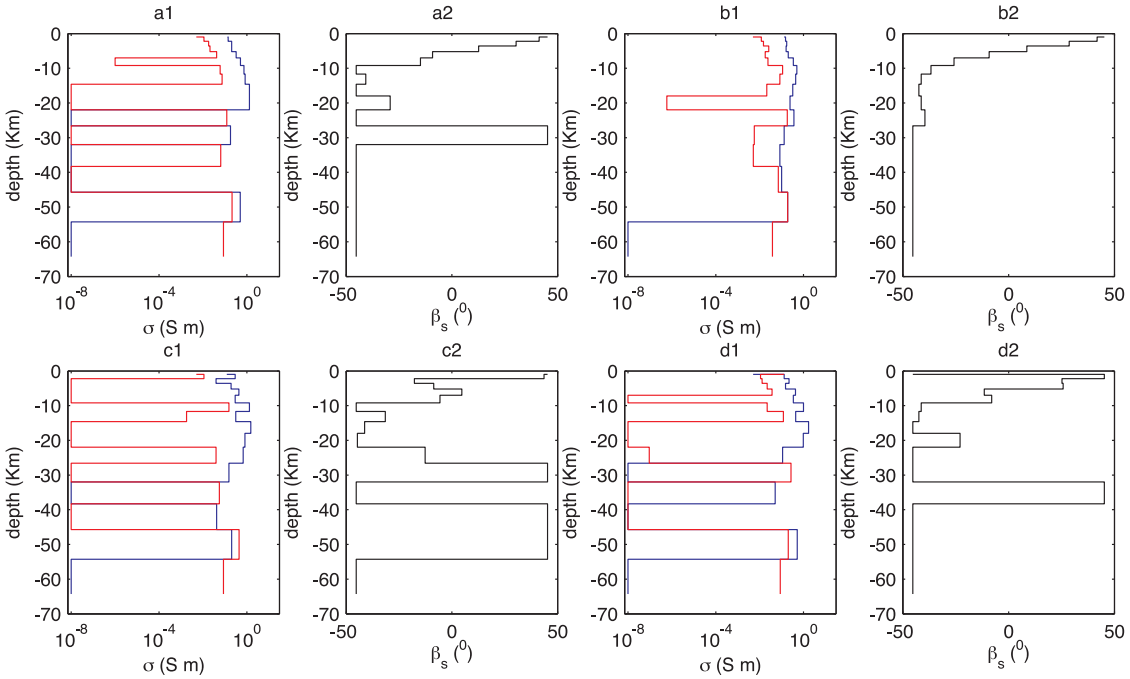


**Figure 5.** Inversion results of the synthetic tests. These profiles are relative to (a) unconstrained test conductivity (a1) and anisotropy direction (a2); (b) compatible reference model test conductivity (b1) and anisotropy direction (b2); (c) incompatible reference model test conductivity (c1) and anisotropy direction (c2) and (d) random reference model test conductivity (d1) and anisotropy direction (d2).

a first-order Tikhonov regularization between the layer-to-layer conductivities, $\lambda_T^2\phi_T^2(\bar{m}) = \lambda_T^2 L^2 \bar{m}^2$ where $L$ is the roughening matrix

$$L = \begin{bmatrix} -1 & 1 & & \\ & -1 & 1 & \\ & & \cdots & \\ & & -1 & 1 \\ & & & -1\ 1 \end{bmatrix}, \tag{12}$$

and the product $L \cdot \bar{m}$ is a finite-difference approximation that is proportional to the first derivative of $\bar{m}$. Trade-off between data-fitting, regularization and the distance from the reference model is controlled via appropriate coefficients, selected *ad hoc* via a trial-and-error process, during the inversion.

Summarizing, the relevant objective function is

$$\phi(\bar{m}, \bar{R}) := \sum_i \left[\frac{d_i - F(\bar{m})_i}{\sigma_i}\right]^2 + \lambda_I^2\phi_I^2(\bar{m}, \bar{R}) + \lambda_A^2\phi_A^2(\bar{m}) + \lambda_T^2\phi_T^2(\bar{m}). \tag{13}$$

In order to set the number of layers, we performed an $F$-test on results from an isotropic inversion validating the hypothesis that a certain number, $n$ layers, is adequate to fit the data. Starting from $n = 2$ logarithmically spaced layers, we successively attempt to fit data progressively better by doubling the number of layers at each iteration. We accepted the hypothesis that $n = 32$ layers are adequate to fit the DIE data with a significance level of 0.05. This test ensures that a finer discretization of the subsurface domain, using 64 logarithmically spaced layers instead of 32, does not improve the $\chi^2$ statistic more than 5 per cent. Similarly, 16 layers is insufficient. In Fig. 6 the rms against number of layers related to an isotropic inversion of the data set is reported.

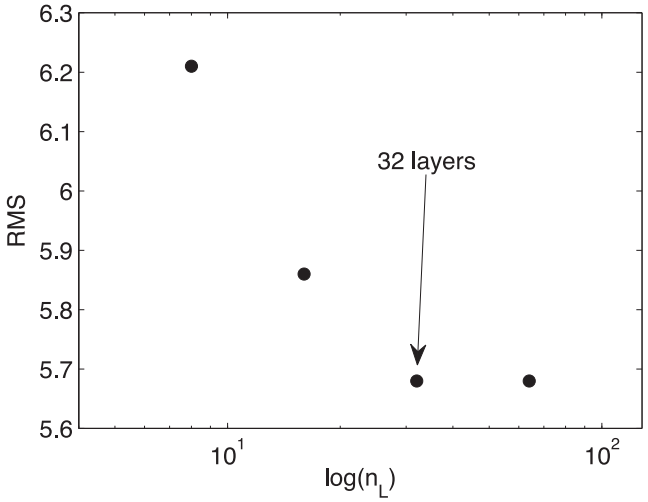


**Figure 6.** The rms versus number of layers. The arrow indicates the point that represents, in the presented scheme, the best trade-off between data fit and model complexity.

The errors relative to the model parameters were estimated via the linearized formula (Aster *et al.* 2005):

$$\sigma_m = \left(\sqrt{\frac{\sum_i r_i^2}{p-q}}\right)^2 \cdot \mathrm{diag}[\mathrm{Cov}(\bar{m}^*)], \tag{14}$$

where is the $p$ number of data and $q$ number of parameters, $r_i = G(\bar{m}^*)_i - d_i$ is the $i$th residual, Cov is the estimated covariance matrix:

$$\mathrm{Cov}(\bar{m}^*) \approx \left[J(\bar{m}^*)^T J(\bar{m}^*)\right]^{-1}, \tag{15}$$

and $J(\bar{m}^*)$ the Jacobian matrix computed respect to the final model $\bar{m}^*$ .

In the previous studies of this area, taken without the constraint of *MI*, problems relating to the minimum value of the horizontal electrical conductivity ratio $\frac{\sigma_{HI}}{\sigma_{LO}}$ needed to fit the data have been highlighted (i.e. the resistivity in resistive direction). In particular, it is important to define the minimum acceptable level of electrical anisotropy in the asthenosphere to compare with petrological and seismological studies.

Constrained by the reference model, our resulting model, shown in Fig. 7, presents three anisotropic regions, similar to the reference model itself. Inside the 95 per cent confidence interval, a difference of more than one order of magnitude between the high conductivity and low conductivity is never required. In Fig. 8 the retrieved model relative to the top 50 km is shown.

Interpreting the anisotropy as an indicator of physical and chemical property changes, it is possible to infer physical or chemical boundaries in the areas in which they arise. Thus, the Moho interface is resolved between 24 and 29 km depth, inside the layer that exhibits the highest anisotropy, and the LAB, the next important boundary, is between 85 and 99 km in depth, again in a highly anisotropic region. Both of these are in excellent agreement with the values found in independent seismic and MT studies of the area (*cf.* Gatzemeier & Moorkamp 2005; Bischoff 2006) and consistent

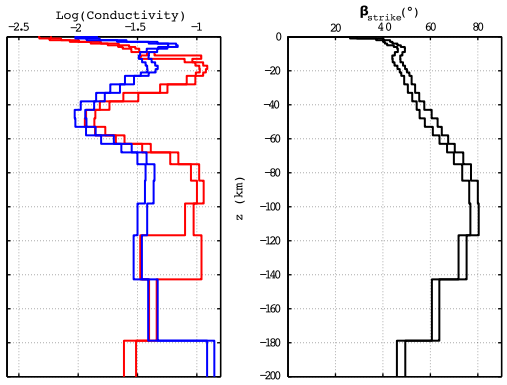


**Figure 7.** Electrical conductivity model from DIE station in Central Germany, relative to the complete discretization of the domain. The left-hand panel present the logarithm of conductivities ± the relative error. Blue and red are relative to the two different aggregate conductivities. The right-hand panel present the anisotropic strike direction ± the direction error. A detail of the upper 50 km of this model is presented in Fig. 8.

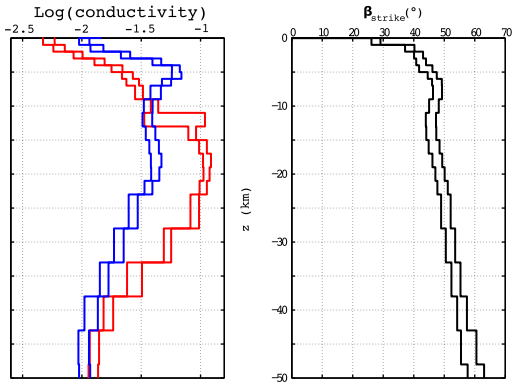


**Figure 8.** Electrical conductivity model from DIE station in Central Germany, relative to the top 50 km. The left-hand panel present the logarithm of conductivities ± the relative error. Blue and red are relative to the two different aggregate conductivities.The right-hand panel present the anisotropic strike direction ± the direction error.

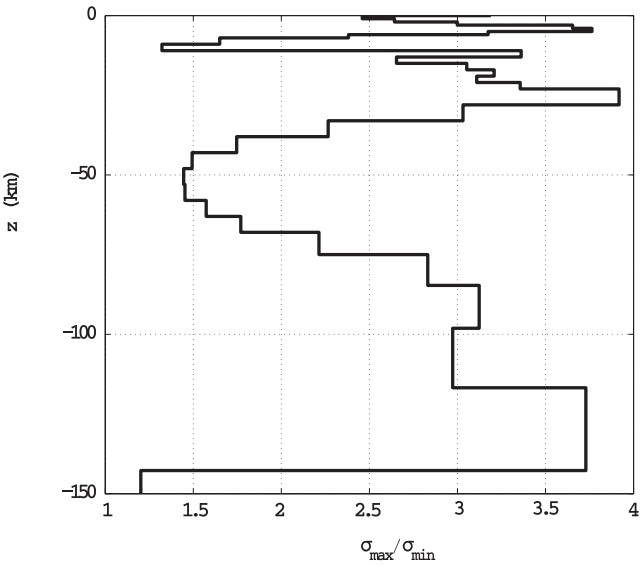


**Figure 9.** Maximum anisotropy rate $\frac{\sigma_{HI}+\mathrm{var}(\sigma_{HI})}{\sigma_{LO}-\mathrm{var}(\sigma_{HI})}$ computed from the model in Fig. 7.

with Phanerozoic Europe eLAB values (Koria 2007; Jones *et al.* 2010).

The anisotropy direction in the crust is resolved in the interval of $(50 \pm 2)^\circ$, in good agreement with the absolute plate motion (APM) direction of $[50^\circ - 55^\circ]$ for Central Europe determined by Gripp (1990), and $(78 \pm 2)^\circ$ at the depth of about 100 km, which is oddly at a high angle to APM, as remarked by prior authors (Gatzemeier & Moorkamp 2005; Roux *et al.* 2011). These anisotropy directions roughly correspond to direction of NE–SW in the crust and E–W in the upper mantle. Interpretation of these directions is beyond the scope of this paper.

The strongest anisotropy in the model is approximately at a depth of 70 km $< z <$ 150 km, and the maximum anisotropy rate, estimated by ratio $\frac{\sigma_{HI}+\mathrm{var}\sigma_{HI}}{\sigma_{LO}-\mathrm{var}\sigma_{LO}}$, in which var is the variance of the considered model parameters, reported in Fig. 9.

Overall, our results from station DIE are in good agreement with models obtained by past studies in this area (e.g. Leibecker *et al.* 2002; Roux *et al.* 2011), with the significant enhancement that our modelling vastly improves the constraint on acceptable maximum electrical anisotropy by defining, for the first time, the resistivity in the resistive direction in the asthenosphere. In detail the maximum electrical anisotropy predicted by our model—in the 95 per cent confidence interval, computed as $1.96\sqrt{\mathrm{var}(\star)}$—is never higher than 4. This value is in agreement with estimates of likely observable anisotropy based on values measured in laboratory for olivine (Gatzemeier & Tommasi 2006), and far lower than that of 125–250:1 in Gatzemeier & Moorkamp (2005) and lower than that of about 10:1 in Roux *et al.* (2011) (note that we modelled distortion-corrected responses, not the original responses, which may account for these different anisotropic factors). In Fig. 10 the retrieved model is compared with results from Roux *et al.* (2011) and the constraint on resistivity in resistive direction is evident. The data fit is illustrated in Fig. 11.

The chosen discretization of the domain prevents optimum fitting of high frequency impedances, where data relative to the longest periods are precisely predicted by the model. The process of removing distortion from the raw data produces non-negligible errors in the resulting impedances. Therefore, propagation of errors via the usual linearized formula is not possible for these impedances, preventing

us from modelling apparent resistivities and phases, more commonly displayed in the MT community. The achieved rms = 0.24 in combination with its expected value $E[\text{rms}] = 1$ highlights that the parametric errors relative to the impedances are probably overestimated [see e.g. Chave & Jones (1997) for a comparison of parametric and jackknife error estimates].

Finally, we examine the estimated joint probability distribution of the seismic $V_s$ fast-axis azimuth angle and high electrical conductivity anisotropic strike angle shown in Fig. 12. The joint probability distribution has been used in by other authors as an indicator of compatible models (*cf.* e.g. Bedrosian *et al.* 2007; Muñoz *et al.* 2010), with the difference that in this work it takes an active role in constraining the model space.

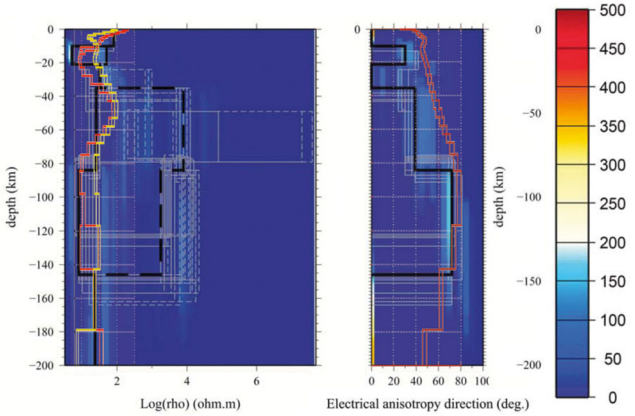


**Figure 10.** Retrieved model (yellow and red) compared with results from Roux *et al.* (2011) model A (black solid and dashed line). It is evident the constraint on resistivity in resistive direction. Colours in background and grey (solid and dashed) lines show the model distribution as retrieved in Roux *et al.* (2011).

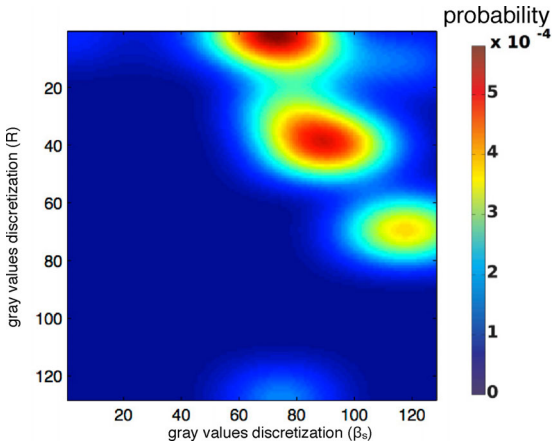


**Figure 12.** Computed joint probability distribution of anisotropic electric strike direction ($x$-axis) and azimuthal seismic fast-axis direction ($y$-axis). This function have been estimated fixing $s = 1.0$. The values on the axis correspond to the discretization—in this case in 128 values—of the values that define the MT strike directions and the ones that depicts the reference model **R** shown in Fig. 3. Thus the probability that a value $R^\star : \frac{\pi \cdot n}{128} \leq R^\star \leq \frac{\pi \cdot (n+1)}{128}$ in figure **R** correspond to a certain value $m^\star : \frac{\pi \cdot i}{128} \leq m^\star \leq \frac{\pi \cdot (i+1)}{128}$ is given by the value of $p_s(n, i)$.

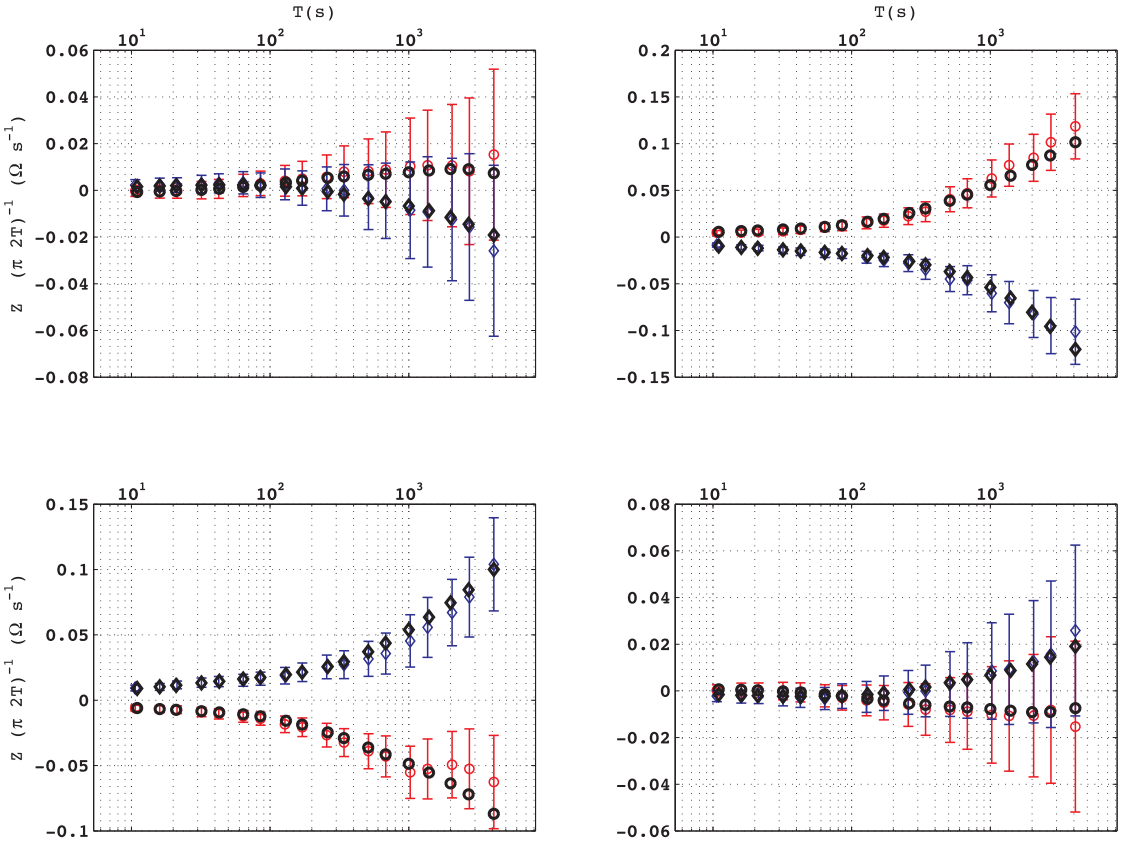


**Figure 11.** Measured and predicted data for DIE station. Red circles represent the real parts, blue diamonds represent imaginary parts, black circles and diamonds are the predicted data for real and imaginary parts, respectively. The top left-hand panel is relative to $Z_{xx}$, the top right to $Z_{xy}$, the bottom left to $Z_{yx}$ and the bottom right to $Z_{yy}$.

The estimated joint probability distribution shows a pattern that suggests overall that similar values of seismic fast-axis direction correspond to the high conductivity directions. In detail, the three high probability areas are approximately aligned, suggesting a constant offset between the considered directions.

These considerations reinforce the support for the hypothesis that the most conductive electric direction and the fast-axis seismic direction have causes that are, at least partially, in common. Given that the pattern appears non-continuous, we reject the hypothesis of a direct, functional relation between seismic and electrical anisotropy, while the clear trend suggests that a possible relation between the sources of the anisotropy. Clustering in the joint probability distribution is due to the use of a layered reference model.

## 4 CONCLUSIONS

A new approach to inverting MT data is proposed in this paper. This method is based on the concept of maximizing the MI between the conductivity map of the subsurface and a map relative to a reference quantity. To achieve this goal, the joint probability distribution between conductivity and the reference quantity is computed, and its entropy, defined as in information theory, is minimized. We applied the proposed inversion method to a previously studied data set from Central Germany, obtaining results that support the most up-to-date interpretation of the subsurface structure in the studied area with the crucial key addition of constraining the amount of electrical anisotropy in the asthenosphere.

The optimal removal of distortion from the data generates high statistical uncertainties in impedance tensor elements, and this effect produces a final rms one half an order of magnitude smaller than that expected. Nevertheless, the final model results are robust. Moreover, the proposed inversion scheme provides a bounded constraint on the resistivity in the resistive direction, effectively showing that the use of a model from an independent data set as the reference has the property of reducing the lack of sensitivity in resistive direction of the MT method.

The joint probability distribution of the high conductivity strike direction with the fast-axis azimuthal direction presents a clustered pattern that provides further support for the hypothesis that the two phenomena are generated by a concurrent source. Meanwhile, the presence of an unaligned probability cluster suggests that there are sources of electric anisotropy that do not affect the seismic fast axis direction. A more accurate analysis is necessary to link these observations to petrological–geophysical properties of the crust and the upper mantle. Future work should focus on both 2-D and 3-D inverse problems, providing the increased number of parameters that characterizes these approach a better base in order to estimate the joint probability distribution, improving in this way the integration of different data.

Finally, interpretation of the highly unusual E–W direction for both the seismic fast axis and the high conductivity direction in the asthenosphere, given that the APM is NE–SW, has been further established. The interpretation of this phenomenon is though beyond the scope of this paper.

## ACKNOWLEDGEMENTS

This work was undertaken under the auspices of Science Foundation Ireland grant Anisotropy of the Continental Lithosphere: Evidence of past tectonics or of active deformation? (SFI 07/RFP/GEOF759) to AGJ. Some of it was accomplished while the authors were at the University of Montpellier 2 as a guest of Professor Andrea Tommasi continuing on a collaboration initiated by an Ireland-France Ulysses travel grant in 2010 to AGJ and Tommasi. The authors wish to acknowledge and thank the financial support of SFI, IRCSET and the University of Montpellier 2 Region Languedoc–Roussillon. Discussions with Estelle Roux, Sergei Lebedev, Eldad Haber, Josef Pek, Andrea Tommasi and Doug Oldenburg contributed this work. We thank Eldad Haber for his source codes to compute MI, and Josef Pek for the contribution he gave us in understanding the contribution of anisotropy in the electromagnetic problem.

## REFERENCES

Abramovici, F. & Shoham, Y., 1977. Inversion of anisotropic magnetotelluric data, *Geophys. J. Int.,* **50,** 55–74.

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

Bailey, R.C., 1970. Inversion of the geomagnetic induction problem, *Proc. R. Soc. Lond., A,* **315,** 185–194.

Barrell, J., 1914. The strength of the Earth's crust, *J. Geol.,* **1914,** 655–683.

Bedrosian, P.A., Maercklin, N., Weckmann, U., Bartov, Y., Ryberg, T. & Ritter, O., 2007. Lithology-derived structure classification from the joint interpretation of magnetotelluric and seismic models, *Geophys. J. Int.,* **170,** 737–748.

Bischoff, M., 2006. Lower crustal anisotropy in central Europe deduced from dispersion analysis of love and rayleigh waves, *Geophys. Res. Abstr.,* **8,** 10010.

Collignon, A., Maes, F., Delaere, D., Vandermeulen, D., Suetens, P. & Marchal, G., 1995. Automated multi-modality image registration based on information theory, *Inform. Process. Med. Imag.,* **3**(6), 263–274.

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.

Cook, F.A., White, D.J., Jones, A.G., Eaton, D.W.S., Hall, J. & Clowes, R.M., 2010. How the crust meets the mantle: lithoprobe perspectives on the Mohorovičić discontinuity and crust-mantle transition, *Can. J. Earth Sci.,* **47,** 315–351.

Chave, A.D. & Jones, A.G., 1997-06. Electric and magnetic field galvanic distortion decomposition of BC87 data, *J. Geomag. Geoelectr.,* **49**(6), 767–789.

Eaton, D.W., Jones, A.G. & Ferguson, I.J., 2004. Lithospheric anisotropy structure inferred from collocated teleseismic and magnetotelluric observations: Great Slave Lake shear zone, northern Canada, *Geophys. Res. Lett.,* **31,** L19614, doi:10.1029/2004GL020939.

Eaton, D.W., Darbyshire, F., Evans, R.L., Grütter, H., Jones, A.G. & Yuan, X., 2009. The elusive lithosphere-asthenosphere boundary (LAB) beneath cratons, *Lithos,* **109,** 1–22.

Gallardo, L.A. & Meju, M.A., 2003. Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data, *Geophys. Res. Lett.,* **30**(13), 1658, doi:10.1029/2003GL017370.

Gallardo, L.A. & Meju, M.A., 2007. Joint two-dimensional cross-gradient imaging of magnetotelluric and seismic traveltime data for structural and lithological classification, *Geophys. J. Int.,* **169,** 1261–1272.

Gatzemeier, A. & Moorkamp, M., 2005. 3D modelling of electrical anisotropy from electromagnetic array data: hypothesis testing for different upper mantle conduction mechanisms, *Phys. Earth planet. Inter.,* **149**(3-4), 225–242.

Gatzemeier, A. & Tommasi, A., 2006. Flow and electrical anisotropy in the upper mantle: finite-element models constraints on the effects of olivine crystal preferred orientation and microstructure, *Phys. Earth planet. Inter.,* **158**(2-4), 92–106.

Giese, P., Scheuber, E., Schilling, F., Schmitz, M. & Wigger, P., 1999. Crustal thickening processes in the Central Andes and the different natures of the Moho-discontinuity, *J. South Am. Earth Sci.,* **12,** 201–220.

Gripp, A.E., 1990. Current plate velocities relative to the hotspots incorporating the Nuvel-1 global plate motion model, *Geophys. Res. Lett.,* **17,** 1109–1112.

Groom, R.W. & Bailey, R.C., 1989. Decomposition of magnetotelluric impedance tensors in the presence of local three-dimensional galvanic distortion, *J. Geophys. Res.,* **94,** 1913–1925.
Haber, E. & Holtzman Gazit, M., 2013. Model fusion and joint inversion, *Surv. Geophys.,* **34**(5), 675–695.
Hamilton, M.P. *et al.*, 2006. Electrical anisotropy of South African lithosphere compared with seismic anisotropy from shear-wave splitting analyses, *Phys. Earth planet. Inter.,* **158,** 226–239.
Hansen, P.C. & O'Leary, D.P., 1993. The use of the l-curve in the regularization of discrete ill-posed problems, *SIAM J. Scient. Comput.,* **14**(6), 1487–1503.
Janik, T., Grad, M. & Guterch, A. the Carpathians Working Group, 2009. Seismic structure of the lithosphere between the East European Craton and the Carpathians from the net of CELEBRATION 2000 profiles in SE Poland, *Geol. Quart.,* **53,** 141–156.
Jones, A.G., 2012. Distortion decomposition of the magnetotelluric impedance tensors from a one-dimensional anisotropic Earth, *Geophys. J. Int.,* **189,** 268–284.
Jones, A.G., Plomerova, J., Koria, T., Sodoudi, F. & Spakman, W., 2010. Europe from the bottom up: a statistical examination of the central and northern European lithosphere-asthenosphere boundary from comparing seismological and electromagnetic observations, *Lithos,* **120,** 14–29.
Koria, T., 2007. How is the European lithosphere imaged by magnetotellurics?, *Surv. Geophys.,* **28,** 238–272.
Leibecker, J., Gatzemeier, A., Honig, M., Kuras, O. & Soyer, W., 2002. Evidence of electrical anisotropic structures in the lower crust and the upper mantle beneath the rhenish shield, *Earth planet. Sci. Lett.,* **202**(2), 289–302.
Mandolesi, E., 2013. Inversion of magnetotelluric data in an anisotropic domain, *PhD thesis,* National University of Ireland, Galway.
Marquardt, D.W., 1963. An algorithm for least-squares estimation of nonlinear parameters, *J. Soc. Indust. Appl. Math.,* **11,** 431–441.
Modersitzki, 2004. *Numerical Methods for Image Registration,* Oxford Univ. Press.
Moorkamp, M., Jones, A.G. & Eaton, D.W., 2007. Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: are seismic velocities and electrical conductivities compatible?, *Geophys. Res. Lett.,* **341,** doi:10.1029/2007GL030519.
Moorkamp, M., Jones, A.G. & Fishwick, S., 2010. Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data, *J. geophys. Res.: Solid Earth,* **115**(B14), B04318, doi:10.1029/2009JB006369.
Muñoz, G., Bauer, K., Moeck, I., Shulze, A. & Ritter, O., 2010. Exploring the gro schnebeck (germany) geothermal site using a statistical joint interpretation of magnetotelluric and seismic tomography models, *Geothermics,* **39**(1), 35–45.
Padilha, A.L., Vitorello, Í., Pádua, M.B. & Bologna, M.S., 2006. Lithospheric and sublithospheric anisotropy beneath central-southeastern Brazil constrained by long period magnetotelluric data, *Phys. Earth planet. Inter.,* **158,** 190–209.
Parker, R.L., 1980. The inverse problem of electromagnetic induction—existence and construction of solutions based on incomplete data, *J. geophys. Res.,* **85,** 4421–4428.
Parsons, B. & McKenzie, D., 1978. Mantle convection and the thermal structure of the plates, *J. geophys. Res.,* **83,** 4485–4496.
Pek, J. & Santos, F.A.M., 2002. Magnetotelluric impedances and parametric sensitivities for 1-D anisotropic layered media, *Comput. Geosci.,* **28,** 939–950.
Pek, J. & Santos, F.A.M., 2006. Magnetotelluric inversion for anisotropic conductivities in layered media, *Phys. Earth planet. Inter.,* **158,** 139–158.
Roux, E., Moorkamp, M., Jones, A.G., Bischoff, M., Endrun, B., Lebedev, S. & Meier, T., 2011. Joint inversion of long-period magnetotelluric data and surface-wave dispersion curves for anisotropic structure: application to data from central germany, *Geophys. Res. Lett.,* **38**(5), L05304, doi:10.1029/2010GL046358.
Savage, M.K., 1999. Seismic anisotropy and mantle deformation: what have we learned from shear wave splitting?, *Rev. Geophys.,* **37**(1), 65–106.
Shannon, C.E., 1948. A Mathematical theory of communication, *Bell Syst. Tech. J.,* **27,** 379–423, 623–656.
Silverman, B.W., 1986. *Density Estimation for Statistics and Data Analysis,* Vol. 37, Chapman and Hall, 120 pp.
Simpson, F., 2001. Resistance to mantle flow inferred from the electromagnetic strike of the Australian upper mantle, *Nature,* **412,** 632–635.
Tarantola, A., 2006. Popper, Bayes and the inverse problem, *Nat. Phys.,* **2,** 492–494.
Tourassi, G.D., Frederick, E.D., Markey, M.K., Carey, J. & Floyd, E., 2001. Application of the mutual information criterion for feature selection in computer-aided diagnosis, *Med. Phys.,* **28**(12), 2394–2402.
Viola, P. & Wells, W.M. III, 1997. Alignment by maximization of mutual information, *Int. J. Comp. Vision,* **24**(2), 137–154.
Yin, C., 2003. Inherent nonuniqueness in magnetotelluric inversion for 1D anisotropic models, *Geophysics,* **68**(1), 138–146.

## APPENDIX A: ENTROPY AND MUTUAL INFORMATION

In 1948 Shannon (1948) published two papers that are regarded as the seed of the so-called 'information theory'. Notwithstanding the importance of this science branch on its own, in this section we present the features we used in the solution of the MT inverse problem.

Our interest in information theory stems from the ability to define a metric between images. In fact, if interpreted as images of the subsurface, tomographic maps are all depicting from the same source—the Earth interior—with each image carrying different information; thus, part of the information is shared between multiple images (*cf.* e.g. Modersitzki 2004).

In information theory, the 'self-information' that an event $A$ may happen with probability $p_A$ is indicated by $I(A)$ and is defined as:

$$I(A) := A \to \log \frac{1}{p_A} = -\log p_A, \tag{A1}$$

if the logarithmic function is base 2, $I(A)$ is measured with 'bin of information'. Other bases of measurement will be not considered here. Moreover—as considered by information theory—the 'entropy' of a random variable $X$, $H(X)$, is the average of the self-information $I(x_i)$ of its possible values $(x_1, x_2, ..., x_n)$:

$$H(X) := E[I(x_i)] = \sum_{i=1}^{n} p(x_i) \log_2 \frac{1}{p(x_i)}, \tag{A2}$$

which follows from the definition $H(X) \geq 0 \forall X$.

Intuitively, the relationship between a variable's entropy and information can be imagined as follows: the more uncertain a random variable is, the more information can be gained through its measurement.

Dealing with two random variables $X$ and $Y$, there are two other interesting quantities to be considered. The 'conditional entropy' $H(X|Y)$ is the entropy of the variable $Y$ conditioned by $X$ taking a certain value $x$, and is the average of $H(Y|X = x)$, over each value $X$ can assume.

$$\begin{aligned} H(X|Y) &:= \sum_{i=1}^{n} p(x_i) H(Y|X = x_i) \\ &= \sum_{i=1}^{n} p(x_i) \sum_{j=1}^{m} p(y_j|x_i) \log_2 \frac{1}{p(y_j|x_i)} \\ &= -\sum_{i=1}^{n} \sum_{j=1}^{m} p(x_i, y_j) \log_2 p(y_j|x_i) \\ &= \sum_{i,j} p(x_i, y_j) \log_2 \frac{p(x_i)}{p(x_i, y_j)}, \end{aligned} \tag{A3}$$

where $p(x, y)$ indicates the joint probability distribution and $p(x|y)$ the conditional probability distribution of $X$ and $Y$. The joint entropy is defined as

$$H(X, Y) := \sum_x \sum_y P(x, y) \log_2 P(x, y), \text{ with } \lim_{P(x,y)\to 0} P(x, y) \log_2 P(x, y) = 0, \tag{A4}$$

being $x$ and $y$ values of $X$ and $Y$ respectively, and $P(x, y)$ the probability that jointly $X = x$ and $Y = y$.

It is possible define the *mutual information* (MI) $I(X; Y)$ as

$$I(X; Y) := \sum_{y\in Y} \sum_{x\in X} p(x, y) \log_2 \frac{p(x, y)}{p(x)p(y)}, \tag{A5}$$

where $p(x, y)$ is the joint probability distribution function of $X$ and $Y$, and $p(x)$ and $p(y)$ are the marginal probability distribution functions of $X$ and $Y$, respectively. MI can be expressed equivalently as

$$\begin{aligned} I(X; Y) &= H(X) - H(X|Y) \\ &= H(Y) - H(Y|X) \\ &= H(X) + H(Y) - H(X, Y) \\ &= H(X, Y) - H(X|Y) - H(Y|X). \end{aligned} \tag{A6}$$

MI measures the information shared by $X$ and $Y$. Alternatively, it measures how much the knowledge of one variable can provide about the other. For instance, if $X$ and $Y$ are independent then $p(x, y) = p(x)p(y)$ and MI $= 0$, while if X=Y then $I(X; X) = H(X)$ the information about $X$ given from the knowledge of $X$ itself is exactly its entropy.

The metric distance between $X$ and $Y$ defined by MI, called 'variation of information', is defined as:

$$\mathcal{D}(X, Y) := H(X, Y) - I(X; Y) \tag{A7}$$

$$= H(X) + H(Y) - 2I(X; Y). \tag{A8}$$

Variation of information respects all the properties a distance metric is required to uphold. In particular:

(i) $\mathcal{D}(X, Y) \leq \mathcal{D}(X, Z) + \mathcal{D}(Z, Y)$;
(ii) $\mathcal{D}(X, Y) \geq |\mathcal{D}(X, Z) - \mathcal{D}(Z, Y)|$;
(iii) $\mathcal{D}(X, Y) \geq 0 \; \forall X, Y$;
(iv) $\mathcal{D}(X, Y) = \mathcal{D}(Y, X)$.

Bearing these relation in mind, it is possible to define a geophysical model as a picture. In fact, if the model is discretized in cells, each cell can be seen as a pixel/voxel of an image.