submitted to *Geophys. J. Int.*

A framework for 3D joint inversion of MT, gravity and seismic refraction data

Max Moorkamp, Björn Heincke, Marion Jegen, Alan W. Roberts and Richard W. Hobbs

27 September 2010

SUMMARY

We present a 3D joint inversion framework for seismic, magnetotelluric and scalar and tensorial gravity data. Using large-scale optimization methods, parallel forward solvers and a flexible implementation in terms of model parametrization allows us to investigate different coupling approaches for the various physical parameters involved in the joint inversion. Here we compare two different coupling approaches, direct parameter coupling where we calculate conductivities and densities from seismic slownesses and cross-gradient coupling, where each model cell has an independent value for each physical property and a structural similarity is enforced through a term in the objective function.

For both types of approaches we see an improvement of the inversion results over single inversions when the inverted datasets are generated from compatible models. As expected the direct coupling approaches results in a stronger interaction between the datasets and in this case better results compared to the cross-gradient coupling. In contrast, when the inverted magnetotelluric data is generated from a model that violates the parameter relationship in some regions but conforms with the cross-gradient assumptions, we obtain good results with the cross-gradient approach, while the direct coupling approach results in spurious features. This makes the cross-gradient approach the first choice for regions were a direct relationship between the physical parameters is unclear.

Key words: joint inversion – seismic tomography – magnetotellurics – gravity

1 INTRODUCTION

All geophysical inversion methods are fraught with the problem of non-uniqueness (e.g. Menke 1989; Muñoz & Rath 2006). Data acquired in a geophysical survey is usually restricted to the surface of the Earth or the shallow subsurface, often with relatively large spacing between measurement sites, and affected by noise. In such a situation different models can explain the observed data, creating ambiguity in their interpretation (Tarantola 2004). Applying regularization stabilizes the inversion and creates a model with certain characteristics but does not alleviate the underlying problem (Parker 1983).

Joint inversion approaches promise to reduce the set of acceptable models by combining several geophysical methods in a single inversion scheme and requiring the resulting model to explain all data simultaneously (Vozoff & Jupp 1975). Two factors help to achieve this goal: First, different methods have different resolving kernels and the null space for one type of data can be resolved by the other (Julia et al. 2000); second, the sources of noise and its impact on the data often differ so that adding another method can improve the results more than adding more data of the same type. Consequently joint inversion approaches have gained some attention recently (e.g Gallardo & Meju 2004; Linde et al. 2006; Chen et al. 2007; Colombo & Stefano 2007; Wagner et al. 2007).

The prize for the increased robustness is that we have to put in assumptions about the relationships between the different types of data. An apparently simple situation is where the models for all types of data are described by the same physical parameter. For example, Rayleigh waves and receiver functions are both usually modelled in terms of S-wave velocity and thus joint inversion of these two dataset seems straightforward. Receiver functions, however, sample the Earth in a different way from surface waves and we have to take care that both sample the same region (Julia et al. 2000) and assess in how far these two types of data are compatible (Moorkamp et al. 2010b).

When combining data that are sensitive to different physical parameters, e.g. seismic velocity and conductivity, we have to explicitly formulate a relationship between those physical parameters. Here we will focus on two approaches: Direct parameter relationships where we specify a functional relation between the parameters (Heincke et al. 2006; Jegen et al. 2009) and a structural constraint, the cross-gradient (Gallardo & Meju 2003), that enforces structural similarity

between the different models. Other approaches include using a parameter relationship as a constraint (Colombo & Stefano 2007) or using a more stringent structural coupling (Haber & Oldenburg 1997). Finally, statistical descriptions of the parameter relationships (Bosch 1999, 2004) allow for an accurate description of the variability and uncertainty of the assumed relationship. We are currently working on a Bayesian methodology that incorporates these and other uncertainties in order to investigate the reliability of joint inversion models (Roberts et al. 2010).

Direct parameter relationships provide a strong coupling between the datasets. The model is expressed in terms of a single physical parameter and all other physical parameters needed for solving the forward problems are calculated using an analytical relationship. Consequently the model update is influenced by the misfit of all datasets (see below for a more detailed description). This strong coupling promises a strong influence on the inversion process, but it is obvious that when the assumed relationship is violated this will distort the resulting model.

The cross-gradient approach makes few assumptions about the relationship of the different parameters. Within the inversion we minimize an objective function term that measures the structural similarity between the different models. This means that the cross-gradient criterion can be violated in parts of the model, but also the cross-gradient as such provides a relatively loose coupling (Gallardo et al. 2005). We can expect though that the improvement over a single inversion is not as strong as for the direct parameter coupling.

It is therefore necessary to tailor the joint inversion approach to the available a-priori information. In areas where the data suggests the possibility of a direct parameter relationship, utilizing this relationship promises improved resolution. In other areas it might not be possible to uniquely define such a relationship or its existence might be unknown or doubtful. In those cases the cross-gradient provides an adequate coupling strategy for joint inversion. In any case it is necessary to carefully assess the resulting models and compare the results from individual inversions with joint inversion results using different coupling strategies.

For this purpose we have developed a flexible 3D joint inversion framework that allows us to test different joint inversion approaches. We first discuss the different types of data that we can include in the joint inversion and the motivation for using each type of data. Then we give an

overview over the different parts of the joint inversion framework, the large scale optimization algorithms, objective function definition, model parametrization and forward and gradient calculation. Finally we use a simple synthetic example to compare two joint inversion approaches with single inversion results and discuss the properties of our joint inversion approach.

THE DATASETS

Before we describe the joint inversion algorithm we briefly discuss the different types of data we use for joint inversion. As the focus of the paper is on the inversion method we restrict ourselves to the most important properties of each method and the rationale for including this type of data in the joint inversion.

Seismic tomography based on first arrival travel times is one of the main methods to construct seismic velocity models of the subsurface (e.g., Zelt & Smith 1992). This method can be used both with active source data (e.g., Di Stefano & Chiarabba 2002) and passive seismological data (e.g., Ritter et al. 2001). In both cases it is necessary to identify the first arrival from a given source on the recorded seismograms. These first arrival picks then form the data that is used as input for the inversion. The solution to the forward problem can be obtained by solving the eikonal equation (Podvin & Lecomte 1991) or ray tracing through the velocity model (Zelt & Smith 1992). In regions with good ray coverage seismic tomography can provide relatively high resolution velocity models.

Scalar gravity data, i.e. the vertical component of gravitational acceleration U_z, are often used to quickly map the density distribution in a large area (e.g., Zanolla et al. 2006). Due to their limited resolution, gravity data are usually explained by forward modelling (e.g., Barrère et al. 2009) or constrained inversion (e.g., Welford & Hall 2007). In joint inversion schemes gravity can provide valuable additional information for the inversion by further constraining density that other data, for example surface waves, have only limited sensitivity to (Maceira & Ammon 2009) and by filling gaps in data coverage (Heincke et al. 2006; Vermeesch et al. 2009).

In addition to the first spatial derivative of the gravitational potential it is also possible to measure the elements of the second spatial derivatives that form the so-called full gravity tensor

Γ (FTG) (Li & Chouteau 1998). Again this type of data is often used to produce maps (Pedersen et al. 1990), but sometimes also inverted to produce models of the density distribution (Droujinine et al. 2007). Being a derivative of scalar gravity data it provides higher lateral resolution but the sensitivity kernels decay more quickly with depth. Thus FTG data are mostly sensitive to near surface structures and lateral contrasts near the surface.

Magnetotellurics (MT) is a passive source electromagnetic method. Assuming a plane electromagnetic wave impinging on the surface of the Earth and measuring the horizontal components of the resulting electric and magnetic fields, we can estimate the magnetotelluric impedance tensor $\mathbf{Z}$ whose elements purely depend on the conductivity distribution in the subsurface (Simpson & Bahr 2005). MT is used extensively to study the conductivity of the crust and upper mantle both on land (e.g. Brasse et al. 2009) and in marine environments (e.g. Baba et al. 2006; Jegen et al. 2009). In recent years it has also been used in marine exploration problems (Key et al. 2006; Constable et al. 2009) as it provides complementary information to seismic surveys. Particularly for sub-salt and sub-basalt problems, where seismic methods often have difficulty imaging the areas below the salt or basalt, MT can provide valuable information although with relatively lower resolution (Jegen et al. 2009).

Having the possibility to include these four types of data means that we cover the most commonly used parameters in geophysical surveys, seismic velocity, density and conductivity. Also, all types of data can be used on a large range of scales from near-surface studies through exploration problems to the scale of the crust and mantle.

2 JOINT INVERSION FRAMEWORK

Figure 1 shows an overview of the different parts of the joint inversion framework and the forward solvers. The framework is separated into different modules, each one with a specific responsibility. We implement each module such that only the minimum required information is exchanged between different parts and therefore we can exchange each module relatively easily to test different optimization methods, model parametrizations or forward solvers. Below we discuss each of the constituents of the joint inversion framework.

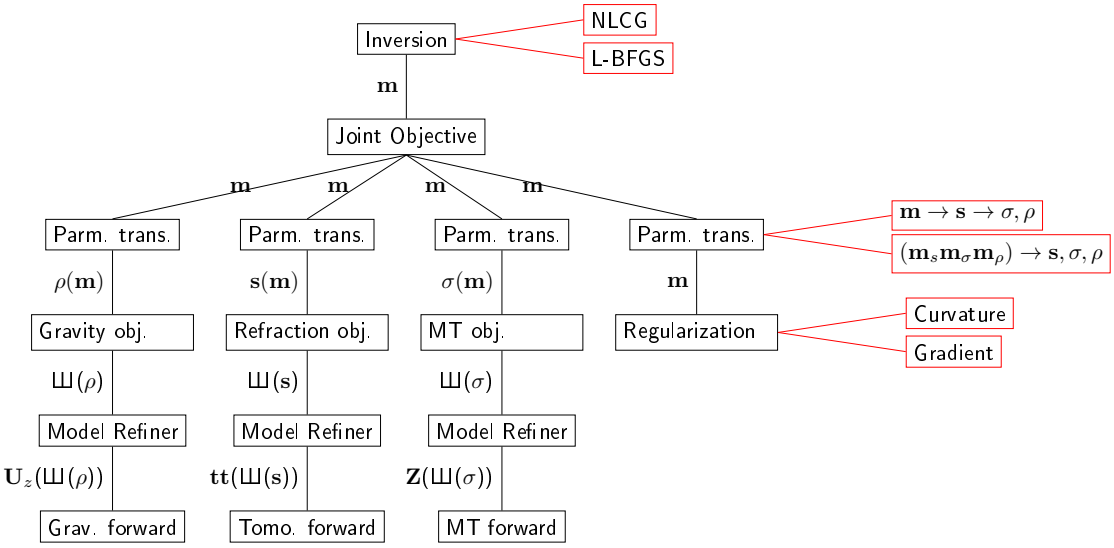

Figure 1. The structure of the joint inversion algorithm. Each module can be exchanged and configured with minimal interference with other parts of the algorithm. We show some of the options for various models in red boxes.

LARGE SCALE OPTIMIZATION

For efficient three-dimensional inversion using optimization methods that scale well to a large number of unknown parameters is essential. Two main factors determine the suitability of an optimization method for large scale inversion: the total computational cost and the memory requirements (Newman & Alumbaugh 2000). Often the main factor that determines the computational cost is the number of forward model calculations. Typically methods that require only few forward calculations, e.g. Gauss-Newton methods, consume large amounts of memory, while methods that require very little memory, e.g. steepest descent, have to perform a large number of forward calculations. We therefore have to find the optimum trade-off between these two factors.

For non-linear inversion Gauss-Newton methods usually require only few forward model calculations to reach a minimum misfit. For a problem with M model parameters and N data this method requires the calculation of the $N \times M$ sensitivity matrix $G_{ij} = \partial d_i / \partial m_j$ and the $M \times M$

approximation to the Hessian $\mathbf{G}^T\mathbf{G}$. For seismic tomography where $\mathbf{G}$ is sparse this approach is often employed (e.g. Zelt et al. 2006; Vermeesch et al. 2009). When the sensitivity matrix is fully occupied as for magnetotelluric and gravity inversion, storing a $M \times M$ matrix severely limits the maximum model size that can be handled with this approach. This can be to some degree remedied by using a data-space Gauss-Newton approach (Siripunvaraporn et al. 2005), but even then we have to store the $N \times M$ sensitivity matrix which for joint inversion with large amounts of data is not feasible.

Quasi-Newton and conjugate gradient methods only require the derivative $\nabla\Phi = \partial\Phi/\partial\mathbf{m}$ of the objective function Φ with respect to the model parameters $\mathbf{m}$. Two algorithms are commonly used in large scale inversion, the non-linear conjugate gradient method (NLCG) (Rodi & Mackie 2001; Commer & Newman 2009) or a limited memory quasi-Newton approach (L-BFGS) (Avdeev & Avdeeva 2009). Both methods are similar in the sense that they use information from previous iterations to improve the current model update. The main difference is that NLCG only uses the previous model update and gradient, while L-BFGS uses a number n_{cp} of correction pairs that can be chosen by the user. NLCG therefore only requires $2M$ extra storage, while L-BFGS requires $2n_{cp}M$ extra storage. As both algorithms have been described well in the literature (e.g. Rodi & Mackie 2001; Tarantola 2004; Nocedal 2006), we only outline the basic steps.

Given a starting model $\mathbf{m}_0$ we perform a number of optimization iterations until the value of the objective function Φ, the misfit between observed and synthetic data, reaches a certain threshold; we will discuss the details of the definition of misfit below. At each iteration i we calculate the misfit and the gradient of the objective function for the current model. Using NLCG or L-BFGS we transform the gradient to a search direction $\mathbf{p}_i$, determine an optimum step length μ using the line search procedure of Moré & Thuente (1994) and update the current model

$$\mathbf{m}_{i+1} = \mathbf{m}_i + \mu\mathbf{p}_i. \tag{1}$$

We then store the necessary information from the current iteration, check whether we have reached the target misfit and if not proceed to the next iteration.

The core inversion algorithm has no knowledge of the meaning of the inversion parameters $\mathbf{m}$, it only perturbs the vector according to the objective function and gradient values. This also im-

plies that the model parameters can assume any value including negative ones. This is problematic if the inversion parameters are seismic velocities or conductivities, for example, which are inherently positive. One solution is to use a constrained optimization method (e.g. Avdeev & Avdeeva 2009) that modifies the search direction or step length to keep the parameters within a specified range. We instead constrain the possible values of the physical parameters through appropriate transformations in the model parametrization which we will discuss below.

Objective function definition

The definition of the misfit in the objective function Φ determines which models we consider an appropriate explanation for the observed data. We follow the approach of Tarantola (2004) and define the misfit for a model $\mathbf{m}$ with respect to a dataset $\mathbf{d}$ as

$$\Phi_d(\mathbf{m}) = \left(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}\right)^T \mathbf{C}_d^{-1} \left(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}\right). \tag{2}$$

Here $\mathbf{g}(\mathbf{m})$ is the synthetic data from the forward calculation for the given model, $\mathbf{d}_{obs}$ the vector of observed data and $\mathbf{C}_d^{-1}$ the inverse of the data covariance matrix. Including the data covariance in the objective function reduces the influence of observations with large errors and equalizes the influence of data with similar relative errors irrespective of the magnitude of each datum. This latter property is important for MT data, where the impedances at high frequencies can be orders of magnitude larger than at low frequencies. In addition this type of misfit definition gives us an objective criterion for a sufficiently small data misfit, when $\Phi_d(\mathbf{m}) = N$, where N is the number of observations, we explain the data within the assumed observational errors (Tarantola 2004).

A geophysical inversion that only minimizes the misfit between observed and synthetic data is ill-posed in the sense that the resulting models show erratic parameter variations and do not resemble Earth structures (e.g., Constable et al. 1987). In some cases the inversion procedure even cannot proceed past the first few iterations and fails to produce an acceptable misfit. By including some form of regularization we can stabilize the inversion and produce smoothly varying models that allow us to identify coherent structures (e.g. Jupp & Vozoff 1975; Constable et al. 1987). This comes at the cost of increased data misfit, although often we can still explain the data within the observational errors.

The regularization functional we use for the joint inversion has the form

$$\Phi_{Reg}(\mathbf{m}) = \sum_i \alpha_i \left(\mathbf{m} - \mathbf{m}_0\right)^T \mathbf{W}_i^T \mathbf{C}_M^{-1} \mathbf{W}_i \left(\mathbf{m} - \mathbf{m}_0\right) + \beta \left(\mathbf{m} - \mathbf{m}_0\right)^T \mathbf{C}_M^{-1} \left(\mathbf{m} - \mathbf{m}_0\right). \quad (3)$$

We sum over the three axis directions $i = \{x, y, z\}$ and weight the contribution for each direction by a weight α_i. The matrices $\mathbf{W}_i$ are finite difference approximations to the first or second spatial derivative of the model parameters in the respective direction (e.g., Farquharson & Oldenburg 1998). Depending on the choice of $\mathbf{W}_i$ we seek a model that has a minimum parameter variation between adjacent cells or minimum curvature, respectively. $\mathbf{m}_0$ is an a priori model that we can use to keep the inversion result close to this reference model. We also include a term to minimize the total value of the model vector, although we usually keep the corresponding weight β small. Similarly to our definition of data misfit, we can include a diagonal model covariance matrix $\mathbf{C}_M$. Changing the entries of this matrix allows us to limit the variation of certain parts of the model, for example to fix known structures.

For the structural joint inversion approach we can have a number of cross-gradient terms (Gallardo & Meju 2003). These are used to enforce structural similarity between the different physical properties in the inversion domain. The cross-gradient objective function is defined as

$$\Phi_{Cross}(\mathbf{m}) = \left(\nabla\mathbf{m}_1 \times \nabla\mathbf{m}_2\right)^T \mathbf{C}_M^{-1} \left(\nabla\mathbf{m}_1 \times \nabla\mathbf{m}_2\right). \quad (4)$$

Here $\mathbf{m}_1$ and $\mathbf{m}_2$ correspond to the parts of the model vector $\mathbf{m}$ that correspond to two different physical properties in the joint inversion, for example seismic velocity and conductivity. This function vanishes if the spatial gradients of the two models point in the same direction, i.e. both physical properties change in the same direction, or one property remains constant. In these cases the magnitude of the change does not have an influence on the value of the cross gradient functional. Only when both properties change, but in different directions, the value of the functional is different from zero. This type of structural coupling is relatively weak and can be expected to hold within the Earth under very general conditions and is therefore a popular coupling method for joint inversion approaches (e.g. Gallardo & Meju 2007; Hu et al. 2009). With these different objective functions we obtain a joint inversion objective functional of the form

$$\Phi_{joint} = \sum_i \Phi_i. \quad (5)$$

Here the index i denotes the various data misfit, regularization and cross-gradient coupling terms. For the direct parameter coupling we have

$$\Phi_{joint} = \Phi_{d,seis} + \Phi_{d,grav} + \Phi_{d,FTG} + \Phi_{d,MT} + \Phi_{Reg}, \tag{6}$$

while for the cross-gradient coupling the full joint objective function is

$$\begin{aligned} \Phi_{joint} &= \Phi_{d,seis} + \Phi_{d,grav} + \Phi_{d,FTG} + \Phi_{d,MT} + \Phi_{Reg,s} + \Phi_{Reg,\rho} + \Phi_{Reg,\sigma} \\ &\quad + \Phi_{Cross,s/\rho} + \Phi_{Cross,s/\sigma} + \Phi_{Cross,\sigma/\rho}. \end{aligned} \tag{7}$$

Here the indices Reg, s, Reg, ρ and Reg, σ denote the regularization for the slowness, density and conductivity section of the model vector, respectively. $Cross, s/\rho$, $Cross, s/\sigma$ and $Cross, \sigma/\rho$ denote the cross-gradient terms between slowness and density, slowness and conductivity and conductivity and density sections of the model vector, respectively. We can achieve different weights between the various terms of the objective function by multiplying the respective covariances by a factor α. For the regularization terms this is equivalent to changing the lagrangian multiplier (e.g. Constable et al. 1987) used in other formulations.

Model parametrization

The way that we parametrize the inversion model determines the nature of the joint inversion to a large degree. In any case we divide a region of the subsurface into rectangular blocks with constant property values within each block. For the joint inversion with direct parameter coupling we assign a single property value to each block, e.g. seismic slowness, and calculate the other physical properties through empirical relationships. In contrast, for the joint inversion with structural coupling we assign a slowness, conductivity and density value to each block and achieve the coupling between the different methods through cross-gradient functionals.

In order to have maximum flexibility in terms of model parametrization and coupling and allow for future experiments with different strategies, we encapsulate the knowledge about the model parametrization in a separate module in our joint inversion implementation. The core inversion algorithm operates on a vector of generalized model parameters $\mathbf{m}$. The objective functions for each method in contrast expect the natural physical properties for the respective method, i.e. slowness

for seismic tomography, conductivity for MT and density for the two types of gravity data. We therefore have to define three transformation functions $\mathbf{s}(\mathbf{m})$, $\sigma(\mathbf{m})$ and $\rho(\mathbf{m})$ that translate between the generalized model parameters and the physical parameters. This type of implementation has the advantage that the coupling strategy for the joint inversion approach is separate from both the inversion routine and the objective function for each method and thus can be changed relatively easily. In terms of our joint inversion objective function it means that we have a definition of the form

$$\Phi_{Joint}(\mathbf{m}) = \Phi_{Tomo}(\mathbf{s}(\mathbf{m})) + \Phi_{MT}(\sigma(\mathbf{m})) + \Phi_{Grav}(\rho(\mathbf{m})) + \ldots, \tag{8}$$

here the dots indicate possible regularization and cross-gradient terms. We will now discuss this type of implementation in terms of the direct coupling and cross-gradient coupling to make it more clear.

For the direct coupling where we only have one value for each inversion cell we have to chose one physical property as the main inversion parameter, e.g. slowness. As mentioned above we want to restrict slowness to positive values and in most cases even to a specified range of slowness values, e.g. between $s_{min} = 10^{-4}$ s/m and $s_{max} = 0.005$ s/m. We therefore define a generalized parameter vector $\mathbf{m}$ and for the forward calculations transform it to slowness using

$$s_j(m_j) = s_{min} + \frac{1 + \tanh(m_j)}{2}\,(s_{max} - s_{min}) \tag{9}$$

for each element m_j of the generalized parameter vector. This way each element of $\mathbf{m}$ can vary throughout the whole numerical range, while the corresponding slowness values are guaranteed to be restricted between s_{min} and s_{max} (Commer & Newman 2008).

Once we have calculated the slowness values we can use the parameter relationships to calculate the corresponding conductivities and densities. Such parameter relationships can come from theoretical considerations of rock properties or simply from empirical laws derived from borehole data (Jegen et al. 2009). We use relationships between slowness s (in s/m) and conductivity σ (in S/m) and slowness and density ρ (in g/cm^3) of the form (Heincke et al. 2006)

$$\sigma = \exp\left(-as^2 - bs - c\right), \quad a = 2.31 \cdot 10^{-7}, \quad b = -5.79 \cdot 10^{-4}, \quad c = 0.124, \tag{10}$$

$$\rho = (1/s + 8500)/5000, \tag{11}$$

for our synthetic examples.

For the cross-gradient coupling the generalized model vector $\mathbf{m}$ contains three segments that correspond to slowness, conductivity and density, respectively, viz.

$$\mathbf{m} = (\mathbf{m}_s \quad \mathbf{m}_\sigma \quad \mathbf{m}_\rho)\,. \tag{12}$$

Each segment uses a transformation of the form of Equation 9 to reduce the range of possible parameter values, but with a different maximum and minimum value for each segment. To calculate synthetic data the corresponding segment of the model vector is passed to the appropriate forward modeling algorithm.

Forward modeling

We use parallel forward solvers to calculate first arrival travel times, MT impedances and scalar and tensor gravity data. The calculation of travel times is based on the code of Podvin & Lecomte (1991), but we can calculate solutions for different source positions in parallel. This type of parallelization is simple to implement, but with hundreds of source positions in a typical survey, scales to large clusters.

Both types of gravity data are calculated using a massively parallel approach that can run either on a standard CPU or on a graphics processing units (GPU). GPUs are particularly suited for problems where simple equations are applied to a large amount of data, such as the calculation of acceleration due to a density distribution. With our GPU based code we observe an acceleration of a factor of 30-40 compared to a single CPU (Moorkamp et al. 2010a).

For the MT impedance calculation we use the integral equation code of Avdeev et al. (1997) and parallelize the impedance calculation by frequency. Although this usually means that only 10-20 parallel calculations can be performed, we achieve good scaling of run time with frequency. An intrinsic parallelization of the forward code can utilize more processors but is difficult to achieve with good scaling (Avdeeva 2008) and out of the scope of this project.

Each forward modeling algorithm has different requirements on the mesh to ensure numerical stability of the solution. All forward algorithms operate on rectilinear meshes where the model consists of cuboids of constant parameter value. However, the travel time code, for example, re-

quires all cuboids in the mesh to have the same size in all three dimensions and the error on the calculated travel time scales roughly with cell size. The error on synthetic impedances depends on the ratio of the mesh size to penetration depth, so deeper parts of the model can be discretized with a coarser grid, as they are only illuminated by long period data. Also, the calculation time for the integral equation code depends strongly on the number of cells in z-direction (depth), therefore it makes sense to use a coarse grid at depth for the MT forward calculations. Finally we might want to regularize the inversion by using large blocks in some areas to enforce constant parameter values there.

For these reasons, we separate the inversion grid from the forward modeling grid for each method. We can specify a different grid for each method and refine the inversion model to match that grid. We use a simple form of refinement where for the forward calculation we only can add new cell boundaries to the inversion mesh and always preserve the existing boundaries of the inversion grid. This way we do not have to interpolate any parameter values. This type of refinement is equivalent to resampling the model and we will denote this operation with $\text{Ш}(\mathbf{x})$.

Gradient calculation

The inversion algorithms require the gradient of the objective function with respect to the generalized model parameters. Given all the steps described above we have for the synthetic travel time data $\mathbf{t} = \mathbf{g}(\mathbf{m})$ in Equation 2

$$\mathbf{g}(\mathbf{m}) = \mathbf{g}\left(\text{Ш}(s(\mathbf{m}))\right). \tag{13}$$

Therefore the gradient of the objective function has to be calculated using the chain rule, for example for the seismic objective function we have

$$\frac{\partial \Phi_{seis}}{\partial \mathbf{m}} = \frac{\partial \mathbf{s}}{\partial \mathbf{m}} \frac{\partial \text{Ш}}{\partial \mathbf{s}} \frac{\partial \mathbf{t}}{\partial \text{Ш}} \frac{\partial \Phi_{seis}}{\partial \mathbf{t}}. \tag{14}$$

Here $\frac{\partial \mathbf{t}}{\partial \text{Ш}}$ is the sensitivity matrix for the seismic data with respect to the refined grid. As this matrix is sparse for the seismic data we can calculate this matrix separately by backtracking through the calculated traveltime field (Heincke et al. 2006), while for gravity and MT data we

directly compute the product with the last term using an adjoint approach (Plessix 2006; Avdeev & Avdeeva 2009).

The term $\frac{\partial \mathrm{III}}{\partial \mathbf{s}}$ describes how the refined grid changes when one of the parameters on the inversion grid changes. For our simple refinement schemes this is just a summation over the gradients in all cells in the refined grid that correspond to one cell in the inversion grid. Finally the term $\frac{\partial \mathbf{s}}{\partial \mathbf{m}}$ relates the changes of the objective function with respect to the physical parameters to the generalized model parameters. This term depends on the parameter relationship and in our case has a simple analytical form. Once we have calculated the gradient for each type of data and possible regularization and cross-gradient terms the gradient of the objective function becomes simply the sum of all gradients

$$\frac{\Phi_{Joint}(\mathbf{m})}{\partial \mathbf{m}} = \frac{\Phi_{Tomo}(\mathbf{m})}{\partial \mathbf{m}} + \frac{\Phi_{MT}(\mathbf{m})}{\partial \mathbf{m}} + \frac{\Phi_{Grav}(\mathbf{m})}{\partial \mathbf{m}} + \dots \tag{15}$$

We then use this gradient to update the inversion model with one of the large-scale optimization algorithms NLCG or L-BFGS.

3 SYNTHETIC EXAMPLES

We perform a number of tests on a simple synthetic model to evaluate the potential of the joint inversion approach. In particular, we want to see how our joint inversion performs in comparison with single inversions and how the two approaches of coupling the data impact on the joint inversion results. For the comparison between joint inversion and single inversion we use L-BFGS as an optimization method in all cases. We constrain the slowness between 10^{-4} s/m and 0.005 s/m, corresponding to velocities of $10,000$ m/s and 200 m/s, respectively.

A SIMPLE EXAMPLE

Our test model (Figure 2 consists of $20 \times 20 \times 10$ cubic cells with an edge length of 500 m. Within a layered background the model contains two rectangular anomalies that extend from the surface down to a depth of 2000 m. One anomaly has a 20% higher slowness than the corresponding background layer, while the other anomaly has a 20% lower slowness.

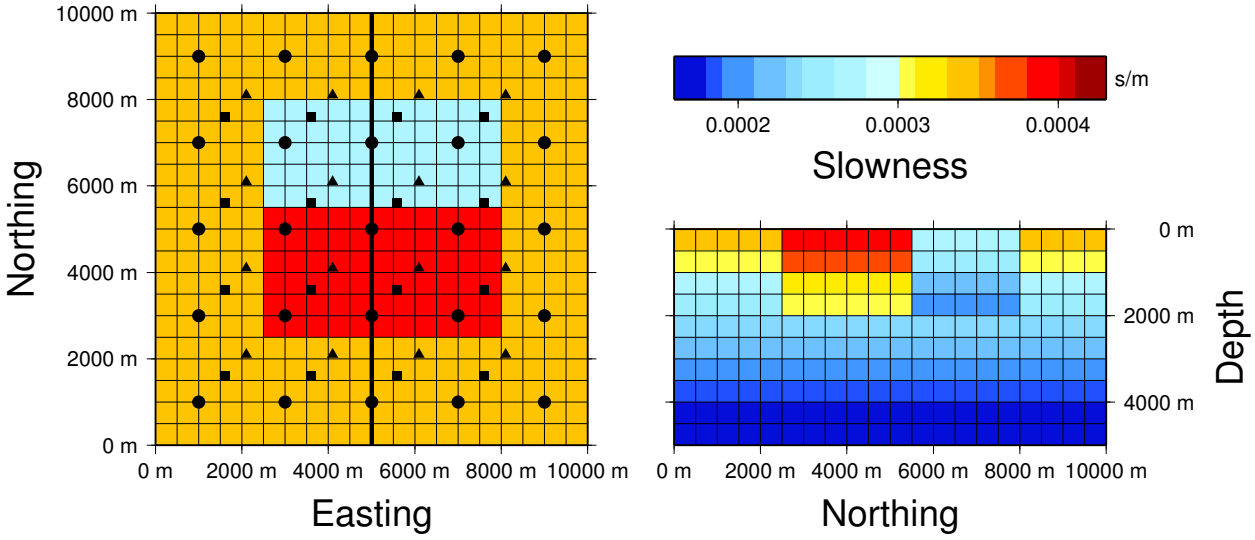

Figure 2. A plot of the true model. On the left we show a plot of the surface layer that shows the position of the two anomalies. On the right we show a vertical depth slice through the center of the two anomalies. The position of this vertical slice is marked by a thick line on the horizontal slice. We show the position of the measurement sites for which we generate synthetic data by circles (seismic receivers and sources), triangles (MT) and squares (gravity), respectively.

We construct corresponding density and conductivity models using the parameter relationships in Equations 10 and 11. From these three models we calculate synthetic travel times, scalar and tensorial gravity responses and MT impedances. We place 25 seismic sources and receivers on a regular grid on the surface with a distance of 2 km between each source in both horizontal directions. For all 600 source-receiver combinations we calculate synthetic travel times. For the gravity data the 16 measurement sites are arranged on a rectangular grid with a distance of 2 km in each direction. At each site we calculate both scalar gravity data as well as the full gravity tensor. Finally, we have 16 MT sites with a spacing of 2 km and calculate the for elements of the MT impedance tensor for 4 frequencies between 10 Hz and 0.1 Hz. We add Gaussian noise with a standard deviation of 20 ms to the synthetic travel times and 2% Gaussian noise to the MT and gravity data. These data with added noise form the input for all individual and joint inversion experiments described below. The combination of a small model with such a small dataset enables us to quickly perform various tests in order to investigate the properties of the two joint inversion approches.

Figure 3 shows the result of a single seismic tomographic inversion. The starting model is the

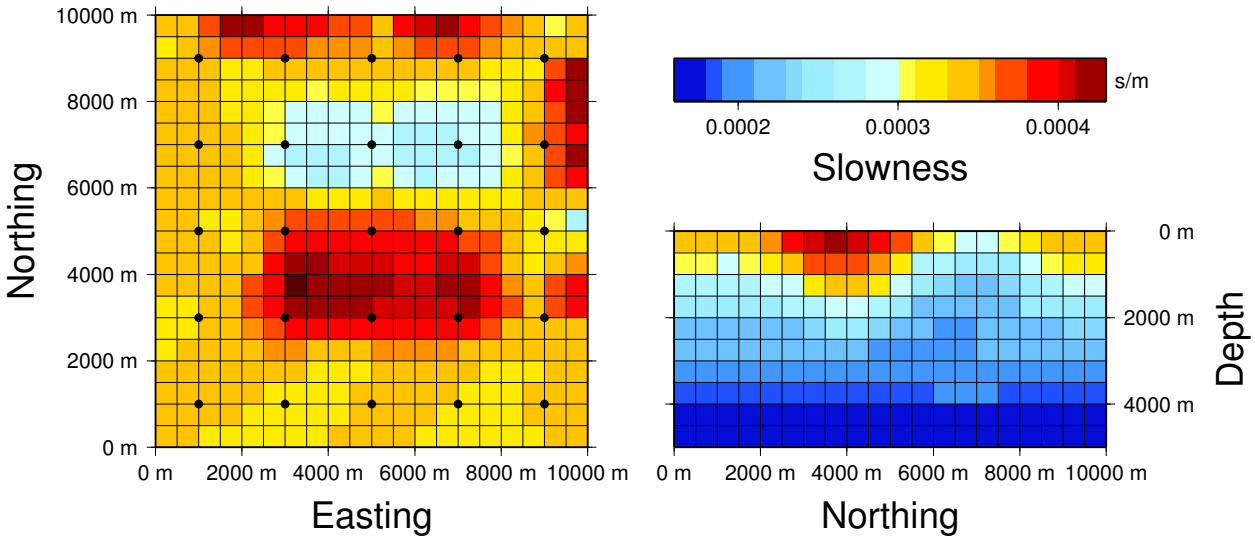

Figure 3. Tomography only model. We show the result of the tomographic inversion with the same scale as for the true model. For reference we also plot the position of seismic sources and receivers (circles).

true layered background without the two anomalies. We see that the inversion of seismic data alone recovers the general shape and position of the two anomalies. We overestimate the amplitude of the high slowness anomaly at the surface and do not fully recover the depth extent of both anomalies. Also, the transition between the two anomalies is blurred out and towards the borders of the model we observe some artifacts in the background. These are both effects of the regularization used in the inversion. Here we employ a curvature based smoothness functional. In addition to equalizing parameter variations it also permits constant spatial gradients and thus trends in the parameters tend to be extended into regions without data coverage. However, this property of permitting linear trends is useful for the inversion of seismic data, as we have increasing velocity with depth that otherwise would have to be facilitated by down-weighting the smoothing in the vertical direction. All in all the result of the single inversion is satisfactory for this simple model.

The inversion of MT data alone shows the expected result for this inductive method (Figure 4). Compared to the seismic tomography model the transition between anomalies and background is more smooth. This obviously also depends on the amount of regularization but it is generally true for all models that do not show strong oscillations in parameter values that are a sign of underregularization. As before we recover the general shape and position of the anomalies and in the center the obtained parameters values match the true model well. Again we loose some

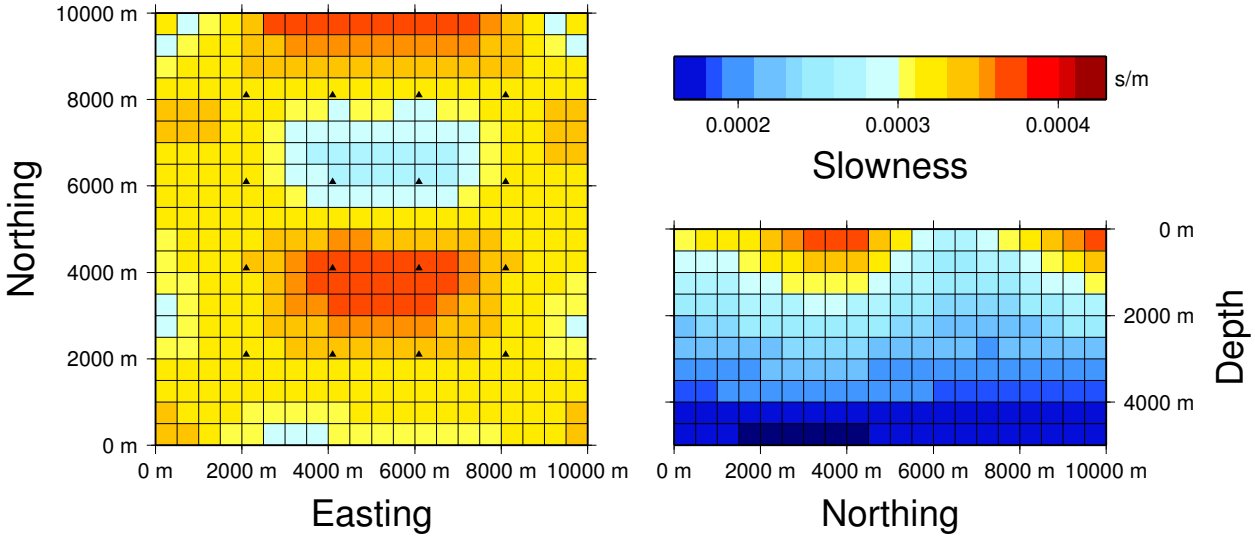

Figure 4. Final model obtained from single MT inversion. For easier comparison we plot the result of the MT inversion in terms of slowness using the true relationship between the test models.

resolution at depth, this is to some degree due to the fact that we only use a single frequency per decade.

The results of the inversion of the two types of gravity data are somewhat surprising. As shown in Figure 5 the inversion recovers the shape of the two anomalies at the surface very well. We have to remember though that for all inversions we start with the true layered background model and the inversion only has to recover the anomalous parameter values. For different starting models we

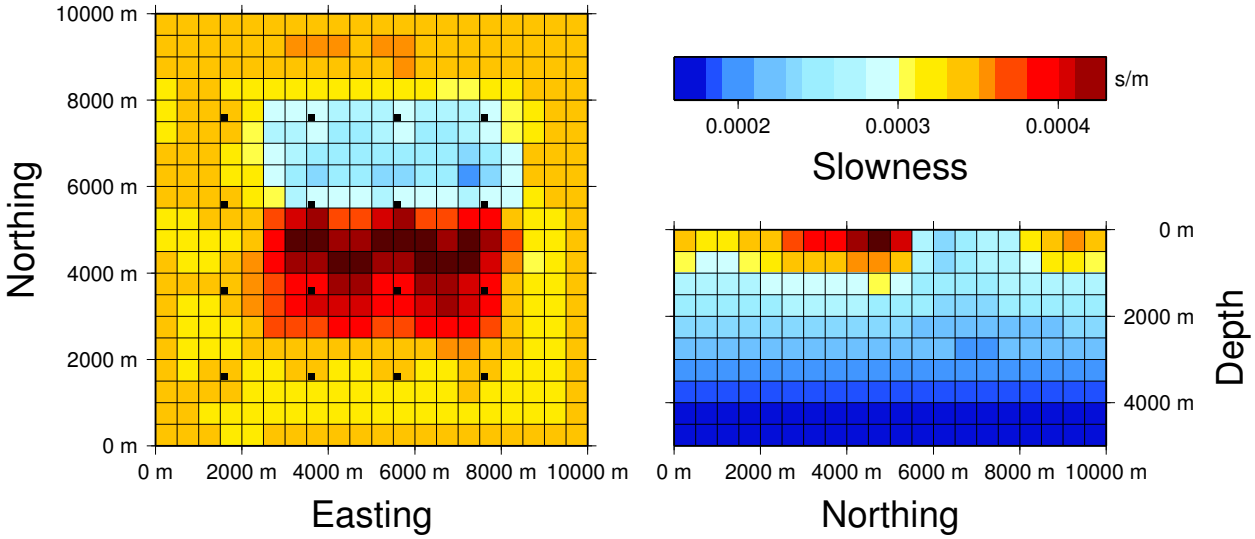

Figure 5. Final model obtained from gravity only inversion. For easier comparison we plot the result of the gravity inversion in terms of slowness using the true relationship between the test models.

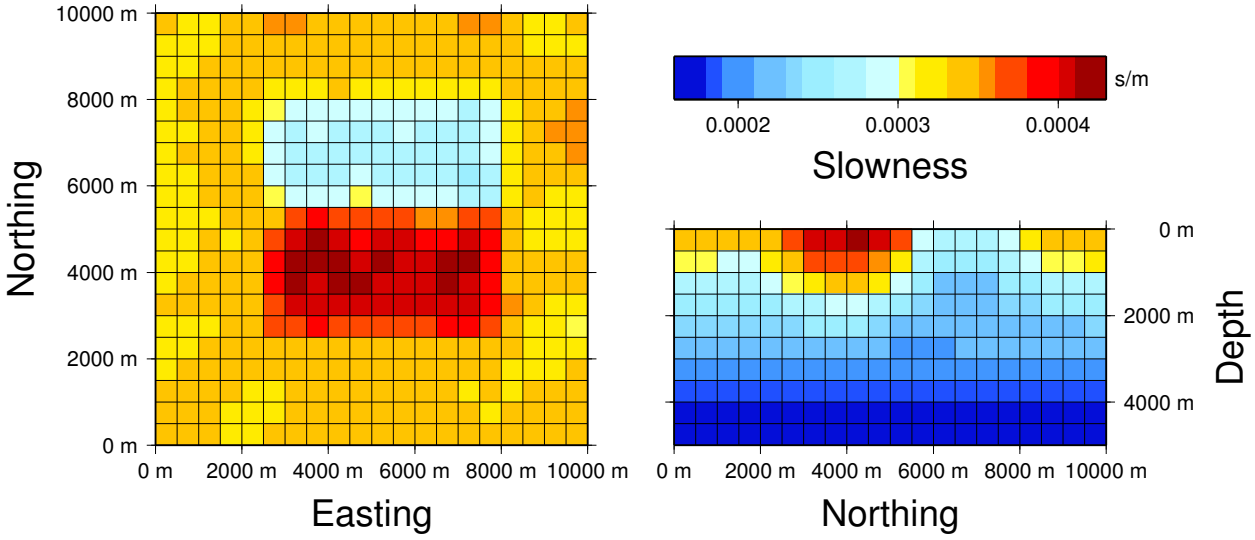

Figure 6. Final model obtained from joint inversion with direct parameter links.

expect the individual MT and seismic inversions to perform similarly but the gravity inversion to differ significantly because of lower resolution. Also the parameter values are not well estimated in cells without measurement sites and the good results are restricted to the surface layer. Already the second layer shows a much weaker signature of the anomalies before they disappear in the third layer.

For the joint inversion we use the same data with added noise as for the individual inversions. All data are weighted by their variance with no additional weighting between the datasets. We first present the results with direct parameter coupling. In this case the parameter relationships in the joint inversion exactly match the relationships that we use to generate the true models. This is the ideal situation for the joint inversion, as the different methods are strongly linked to each other and we can expect some improvement of the inversion results. For this example we use the same weight for the regularization term, however due to the additional data the effective weight of the regularization is slightly lower than for the individual inversions.

We plot the model resulting joint inversion with direct parameter coupling in Figure 6. For the surface layer we can see significant improvement over the inversion of the individual datasets alone. The shapes of both anomalies match the shape of the true anomalies and we also recover the slowness values of the anomalies with little scatter. In addition, the artifacts at the border of the model are strongly reduced. The vertical slice through the model shows that with depth

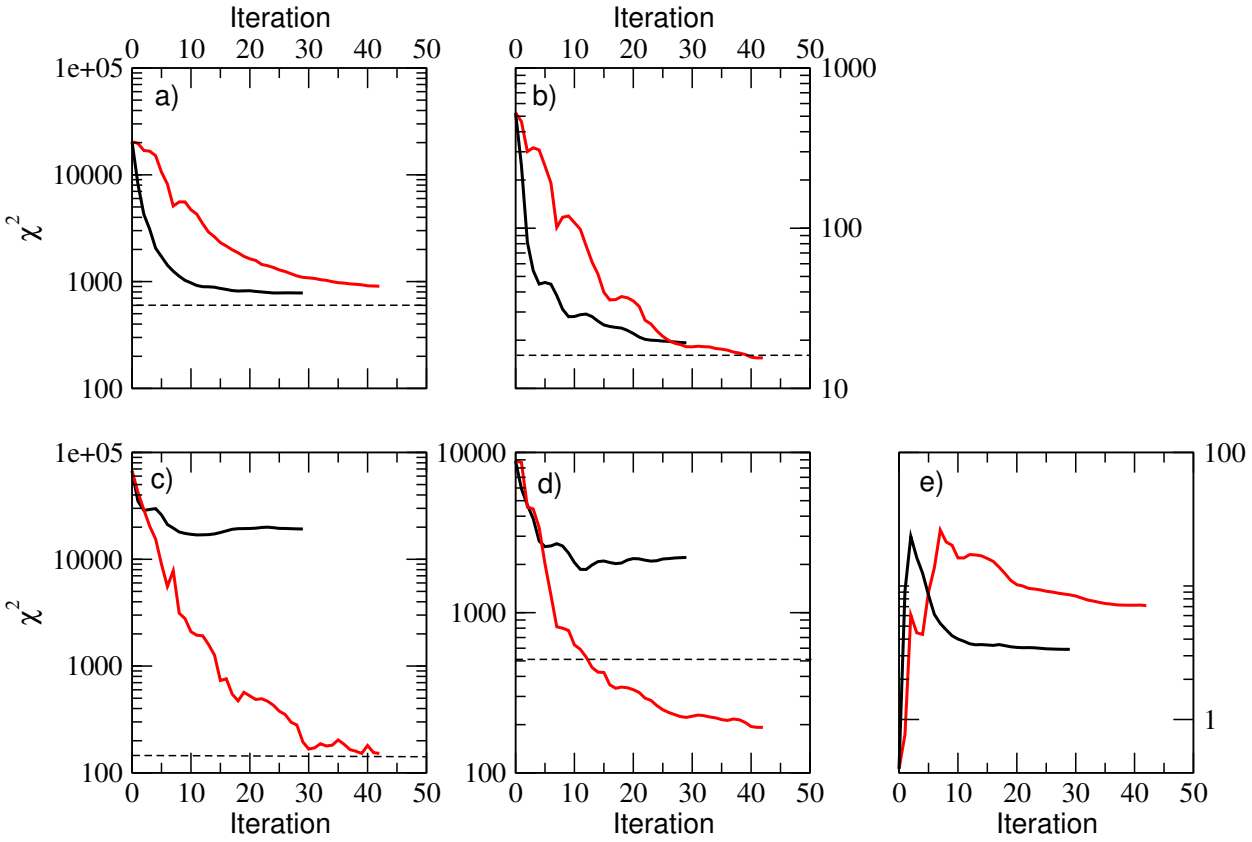

Figure 7. Convergence curve for the joint inversion compared to the seismic tomography inversion. We show the χ^2 misfit at each iteration for each type of data for the tomographic inversion (black) and joint inversion using direct parameter coupling (red). a) seismic data, b) scalar gravity data, c) tensor gravity data, d) MT data, e) regularization term. We also plot the level that corresponds to an RMS of 1 as a dashed line in each plot of data misfit.

the improvement becomes less pronounced. Below 2,000 m depth both inversion results are very similar. We have to consider though that gravity data has no direct depth resolution and that the MT data has only one impedance estimate per decade.

Comparing the convergence curves of the seismic tomography inversion (black) with the joint inversion (red) in Figure 7 illustrates the influence of the gravity and MT data. While the χ^2-misfit for the seismic data at final iterations is virtually identical, the misfit of the tensor gravity and MT datasets is consistently lower after the first iteration. It is interesting to see that the misfit for the scalar gravity data is higher for the initial iterations but then falls below the final misfit of the seismic tomography inversion. We can also see the complex interaction between the different types of data as the misfit for each type of data occasionally increases before it plunges down again.

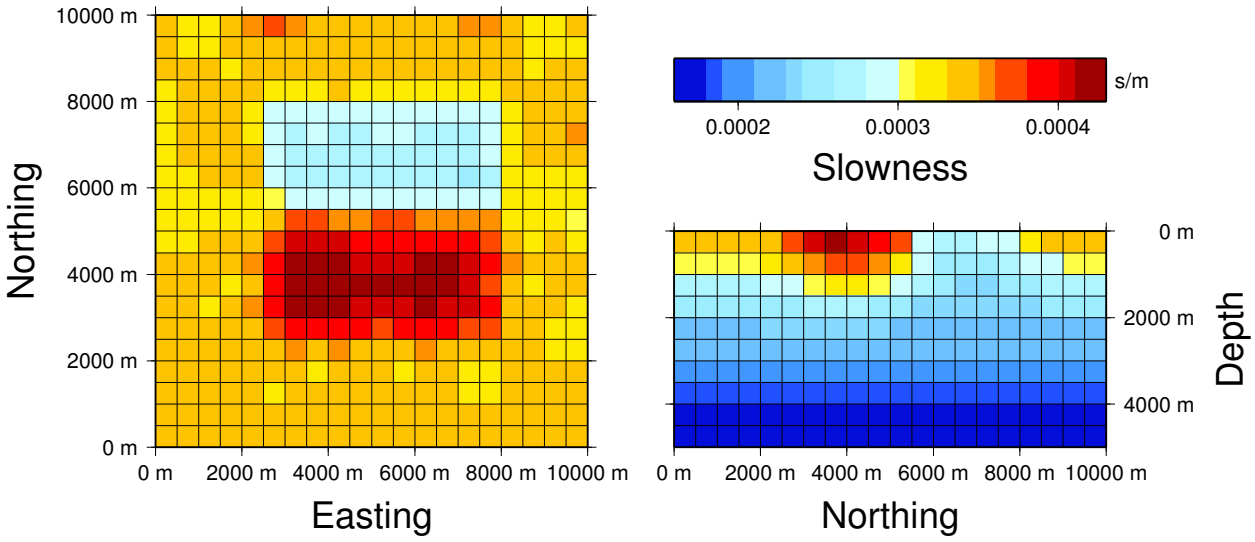

Figure 8. Final model obtained from joint inversion with cross-gradient links.

Although the tomographic inversion also reduces the misfit of the other datasets, the joint inversion reaches a final misfit that is a factor 2-100 lower than for the seismic tomography inversion. The joint inversion stops after 42 iterations as it cannot find a suitable step size that minimizes the objective function significantly. The total run time in this case is ≈ 40 minutes on a standard Core 2 Q6600 quad core desktop computer with 4 GB of main memory. Most of this time is consumed by the MT forward and gradient calculations.

For the joint inversion using the cross-gradient approach we need to specify the weights for three cross-gradient terms and three regularization terms, one for each of the physical parameters slowness, density and conductivity. We obtain a separate model for each parameter, however the cross-gradient terms ensure that there is structural similarity between these three models. Due to the interaction between data misfit objective, cross-gradient objective and regularization, the value of the weight for the regularization term is not directly comparable to the weight for the individual inversion. We therefore scale the covariance for the three regularization terms so that the misfit for the cross-gradient inversion becomes comparable to the joint inversion with direct parameter coupling. This ensures that the models remain comparable. In this case the values for the covariance scaling factors are $\alpha_{Reg,s} = 50$, $\alpha_{Reg,\sigma} = 50$ and $\alpha_{Reg,\rho} = 1$.

Figure 8 shows the inversion result for seismic slowness using cross-gradient coupling. The result looks similar to the direct parameter coupling although there is some additional scatter at

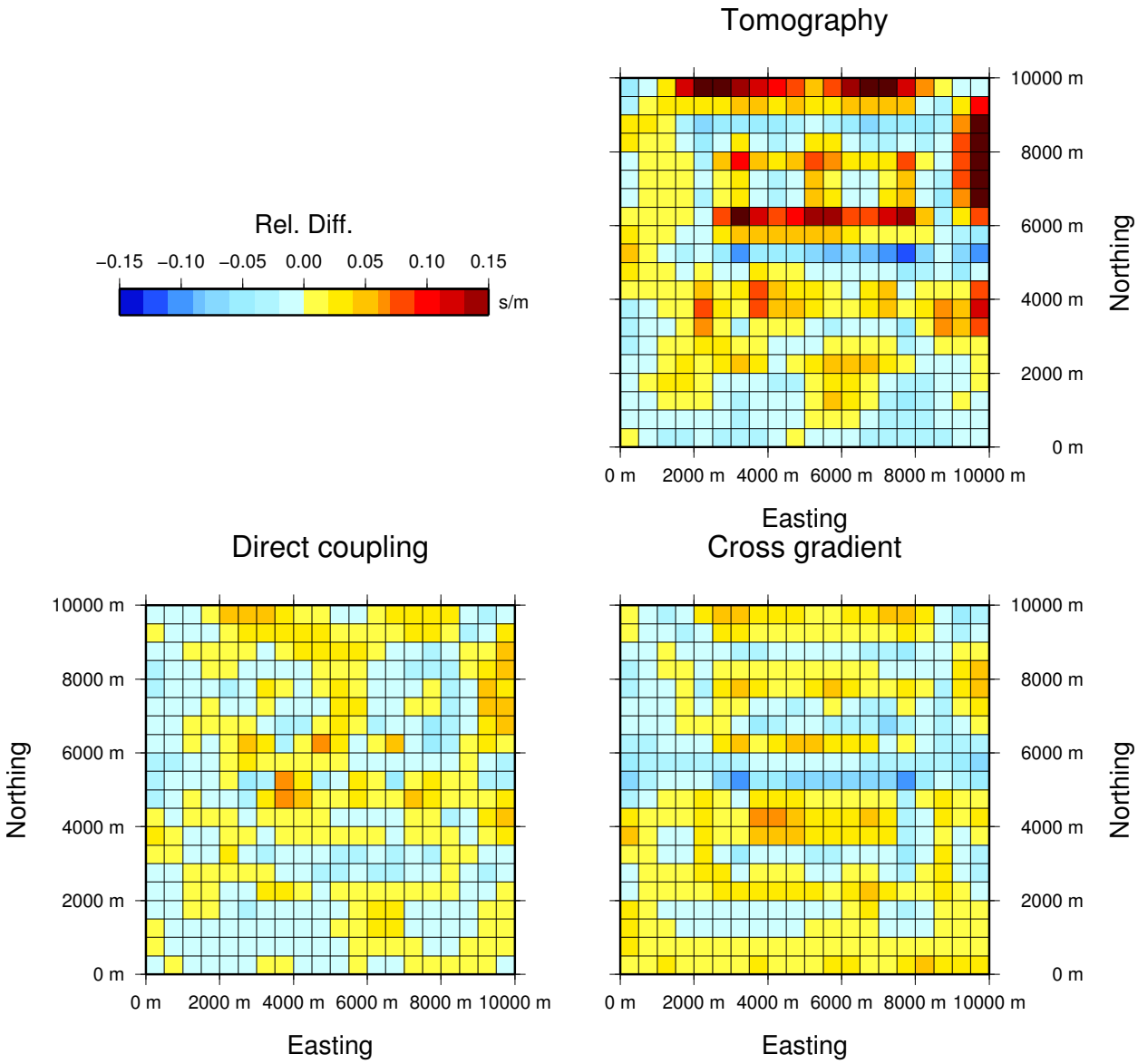

Figure 9. Relative difference $(s^{true} - s)/s^{true}$ between inversion results and true model for the slownesses. For each inversion we plot the relative difference for the surface layer.

the borders of the two anomalies. As before we can see some improvement over the individual inversions for the surface layer, but at depth all results are relatively similar. The colour scale used to plot the inversion result is a compromise between representing the whole range of parameter variations while keeping the anomalies and small fluctuations visible. To improve the visibility of fluctuations from the true model we plot the relative difference between the inversion results and the true model for the surface layer in Figure 9 and for a vertical slice across the center of the model in Figure 10.

Deviations from the true model are relatively similar for both types of joint inversion. The max-

imum relative deviation from the true model is $\pm 10\%$. Particularly at the boundaries of the model we can see larger deviations that are due to reduced data coverage. We can identify larger deviations from the true model for the cross-gradient case at the transition between the two anomalies, but throughout the rest of model the inversion results of both joint inversions fluctuate around the true model with positive and negative deviations equally distributed. In comparison the individual seismic inversion shows much larger deviations from the true model demonstrating the improvement we can obtain with joint inversion in comparison to the single inversions. This observation is confirmed by the misfit between the true model and the three inversion results,

$$\chi^2 = \sum_i \left(\frac{s_i^{true} - s_i^{inv}}{s_i^{true}} \right)^2 . \tag{16}$$

Here the summation goes over all model cells and s^{inv} and s^{true} denote the slowness in each cell for the inversion result and the true model respectively. For the tomographic inversion we obtain $\chi^2_{tomo} = 5.3$, while for the cross-gradient inversion $\chi^2_{cross} = 4.8$ and for the direct coupling joint inversion $\chi^2_{direct} = 3.5$.

Obviously the joint inversion results depend on the chosen weighting between the data, the regularization, and, if present, the cross-gradient functionals. Particularly for the cross-gradient approach we need to specify various weights that all impact on the final model. We therefore investigate the interaction between the different functionals for this approach further.

Figure 11 shows how variations in the regularization of the conductivity model change the obtained velocity model with the cross-gradient approach. Both models are coupled through a cross-gradient term, so as long as the coupling is strong the smoothness of the MT model will be reflected in the seismic model. From these plots it appears that any value for $\alpha_{Reg,\sigma}$ between 10 and 100 produces reasonable results. Below $\alpha_{Reg,\sigma} = 10$ we start to see erratic behaviour, while for $\alpha_{Reg,\sigma} > 100$ the resulting models look overly smooth. Similarly, the misfit for the seismic data (not shown) increases with increasing regularization of the conductivity model. It therefore appears that for our chosen weight for the cross-gradient term $\alpha_{Cross,s/\sigma} = 10^7$ the regularization of conductivities has a direct impact on the seismic data. We will now see how changing the weight of this term influences the inversion results.

Again we use the initial cross-gradient result as a basis for our experiment and vary the values

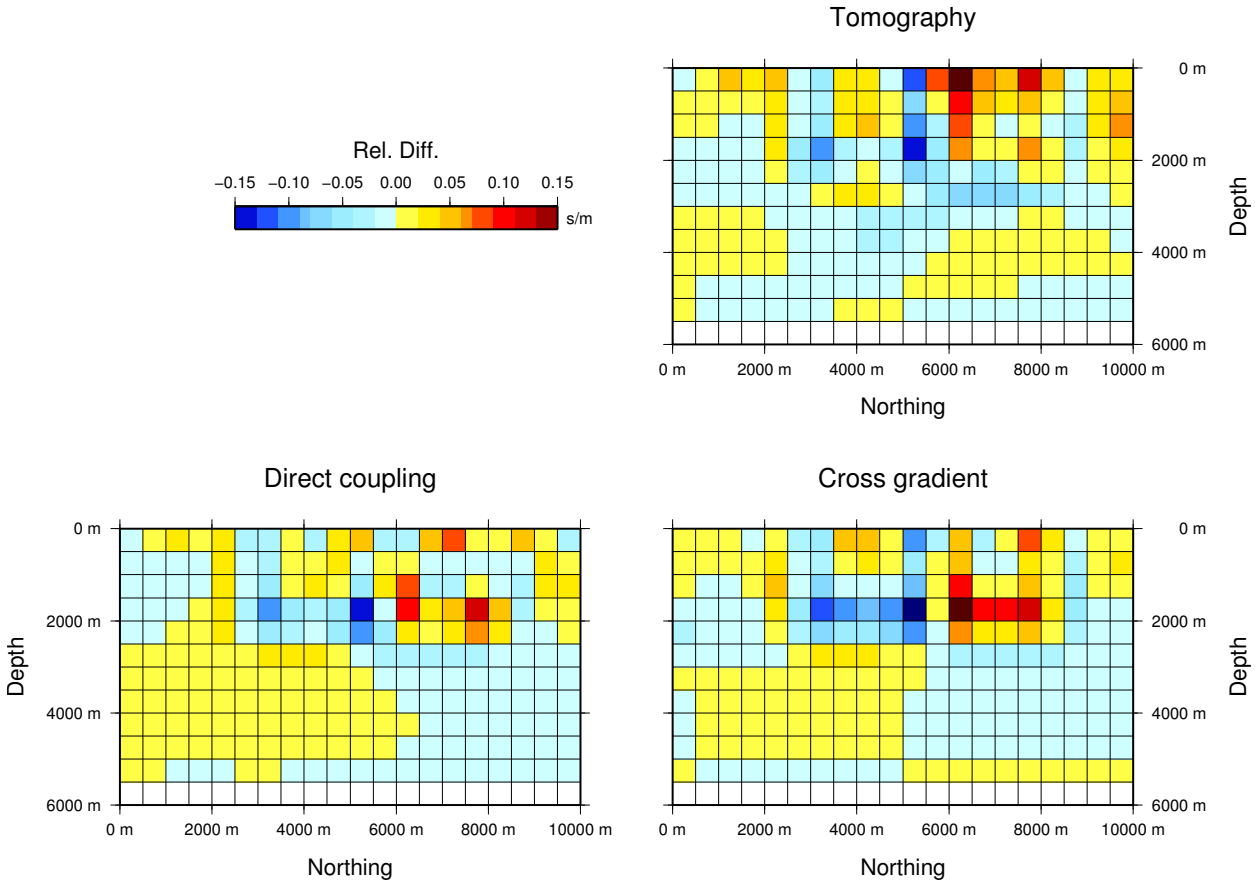

Figure 10. Relative difference $(s^{true} - s)/s^{true}$ between inversion results and true model for the slownesses. For each inversion we plot the relative difference along a vertical slice through the center of the model.

for all three cross-gradient terms that were $\kappa = \alpha_{Cross,s/\rho} = \alpha_{Cross,\sigma/\rho} = \alpha_{Cross,\sigma/s} = 10^7$ above between 10^5 and 10^9. Figure 12 shows a plot along the surface for all three parameters involved in the joint inversion. We observe that the results are similar for coupling values between $\kappa = 10^6$ and $\kappa = 10^8$. The three inversion results obtained with this parameter range all show a good recovery of the two anomalies with little scattering or excessive smoothing. The model with a weight of $\kappa = 10^5$ for the three cross-gradient terms in comparison is visibly smoother, while for a value of $\kappa = 10^9$ we start to see erratic parameter variations for slowness, density and conductivity. Interestingly the model with the largest coupling is not only more erratic but also has a larger misfit for all datasets compared to the other models. This indicates that for such high values the inversion is mostly occupied with decreasing the cross-gradient term at the expense of the data misfit terms and the regularization. Conversely a lower weight for the cross-gradient

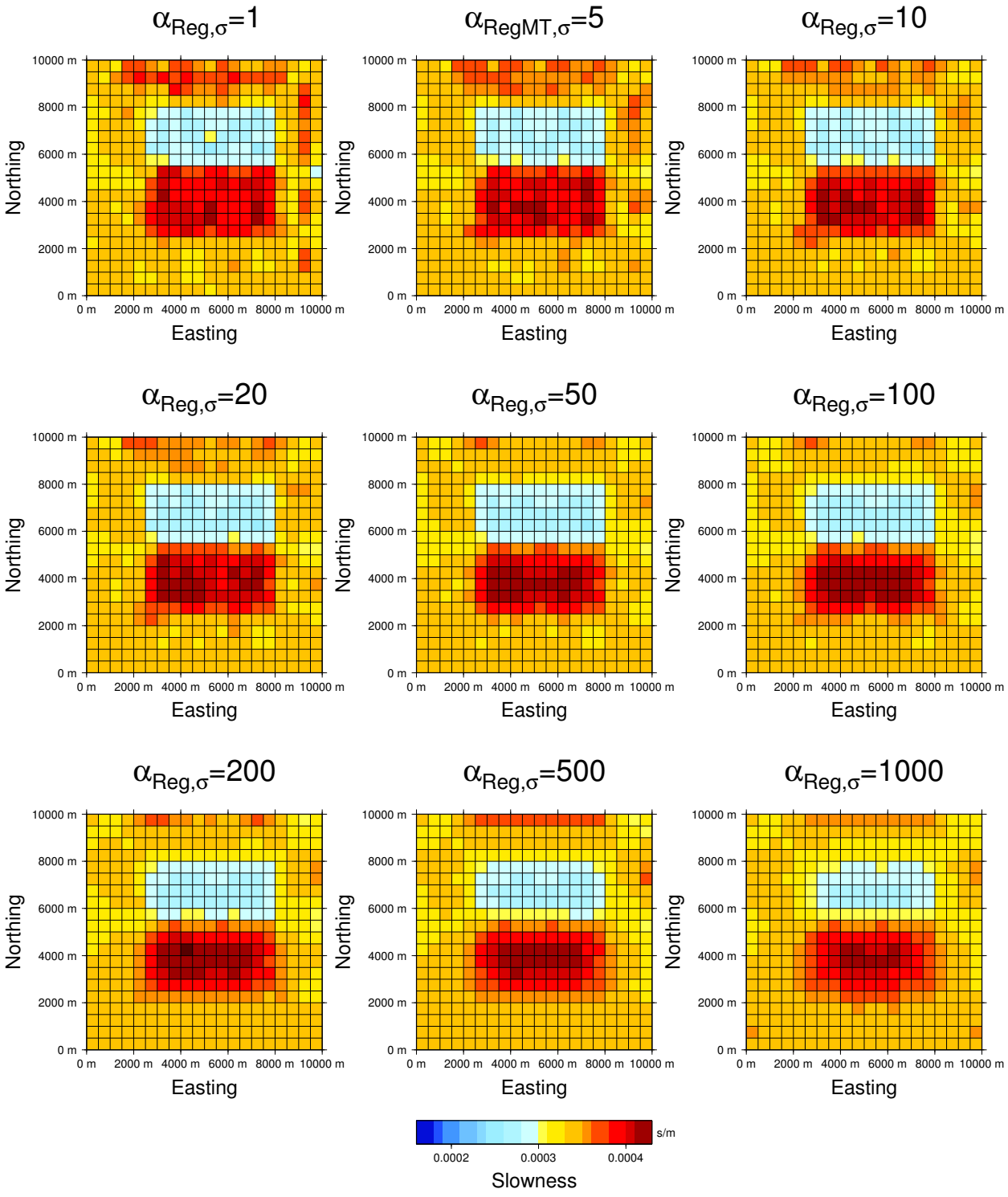

Figure 11. The impact of MT regularization on slowness in the cross-gradient inversion. We plot the seismic model from the cross-gradient inversion for different values of the weight for the regularization of the conductivities $\alpha_{Reg,\sigma}$. This demonstrates how the different physical parameters interact through cross-gradient coupling.

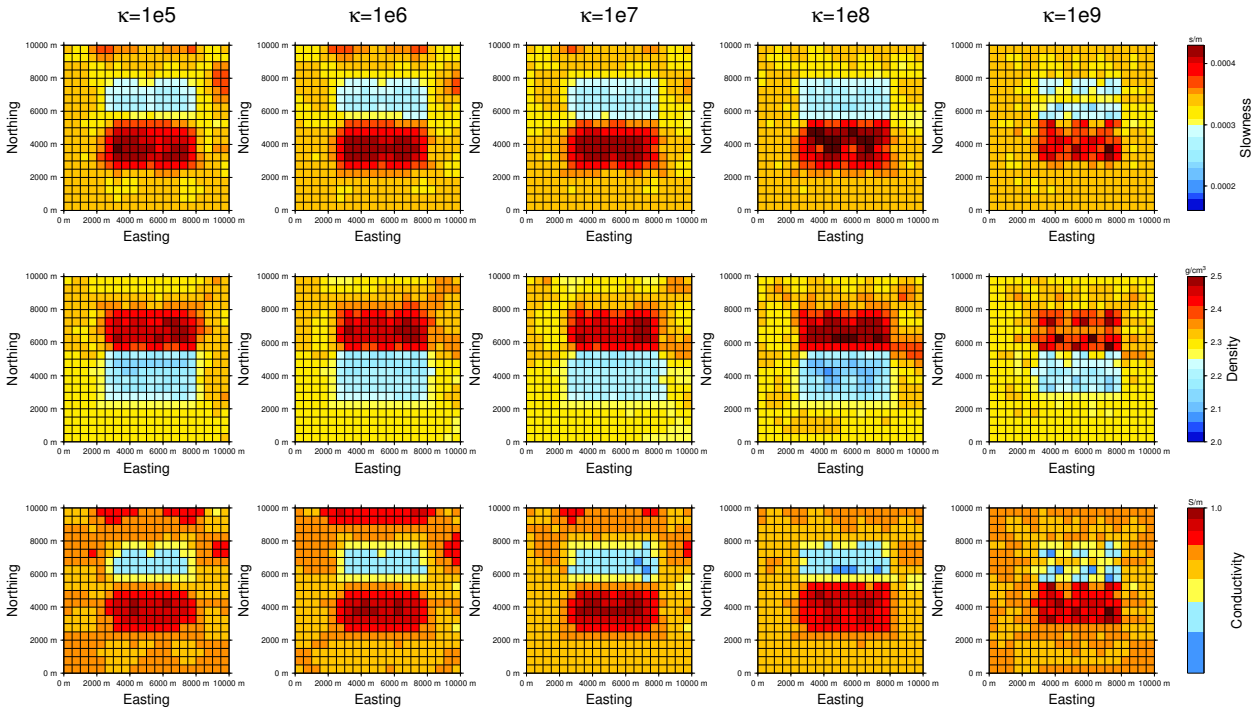

Figure 12. The impact of the weight of the cross-gradient coupling term $\kappa = \alpha_{Cross,s/\rho} = \alpha_{Cross,\sigma/\rho} = \alpha_{Cross,\sigma/s}$. Higher values increase the coupling and therefore the similarity between the models for different physical parameters. We keep all other weights constant.

terms allows the regularization to act more strongly as indicated by the smooth variations of the inversion results.

Finally we examine how the direct parameter coupling and the cross-gradient approach behave when conductivity and slowness do not have exactly the same spatial structure. We now combine the seismic and gravity datasets used above with synthetic MT data calculated from the incompatible model shown in Figure 13. Here we multiplied the conductivities for the right half of the two anomalies by a factor of two. Although we only plot the surface layer we apply this multiplication for the whole depth range of the anomalies between the surface and 2000 m depth. The site distribution, frequencies and noise characteristics are exactly as before for the test with the compatible models and we call this the incompatible test case.

Figure 14 shows the joint inversion result using the direct parameter coupling with exactly the same weights as for the compatible case. As expected the deviation from the assumed parameter relationship is mapped into spurious anomalies. These occur not only directly in the regions where the true models do not match the relationship, but also in the surrounding regions. The joint in-

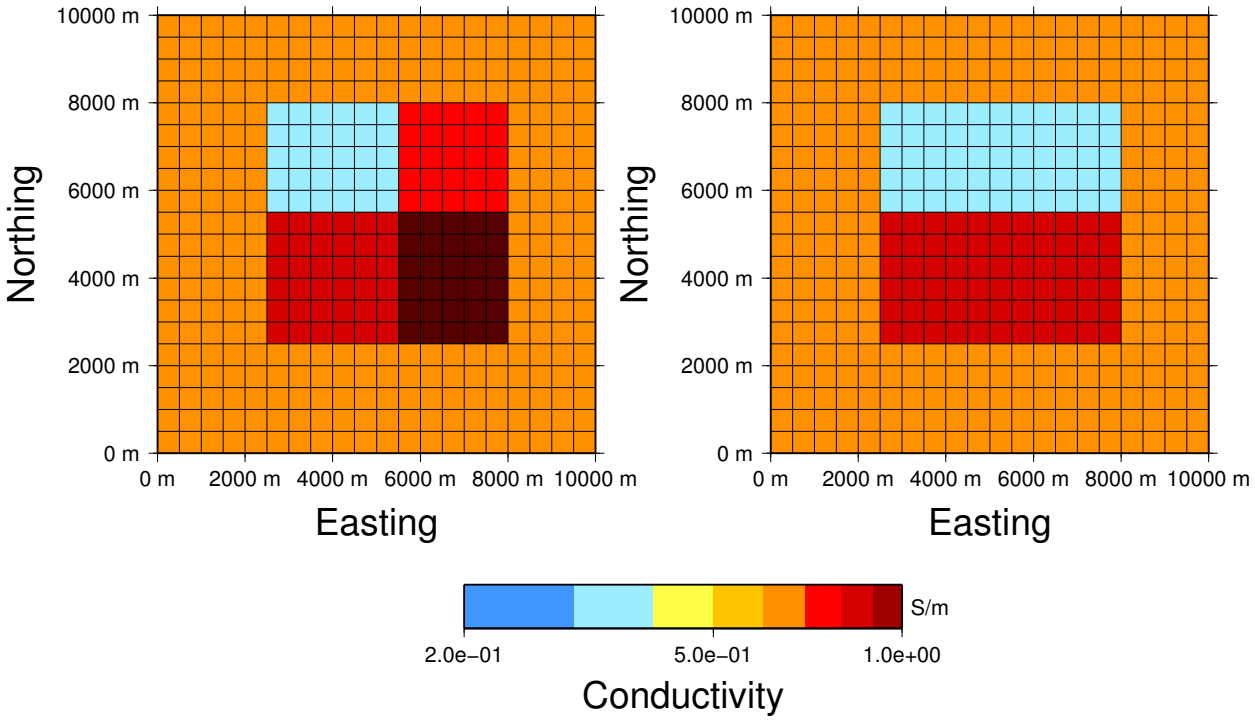

Figure 13. The incompatible conductivity model (left) that we use for the following experiment and for comparison the compatible conductivity model used above (right).

version is creating these structures in order to fit all datasets simultaneously. We can identify the limited reliability of this model through the significantly increased misfit for both the MT and the

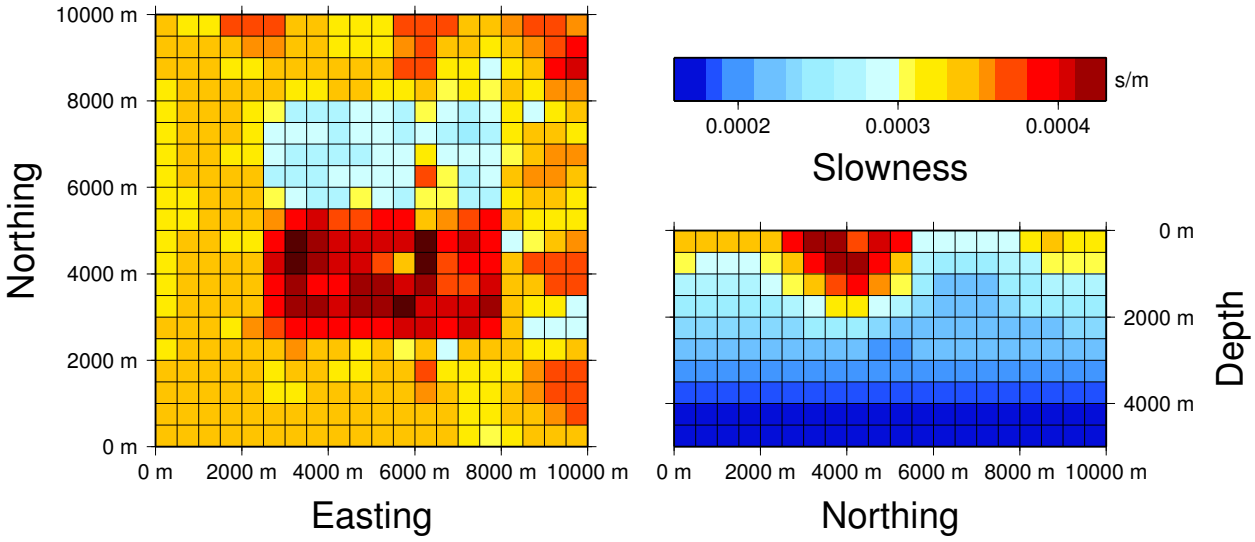

Figure 14. The slowness model from joint inversion with direct parameter coupling for the incompatible dataset. We can see how the discrepancy between the different data is mapped into spurious anomalies.

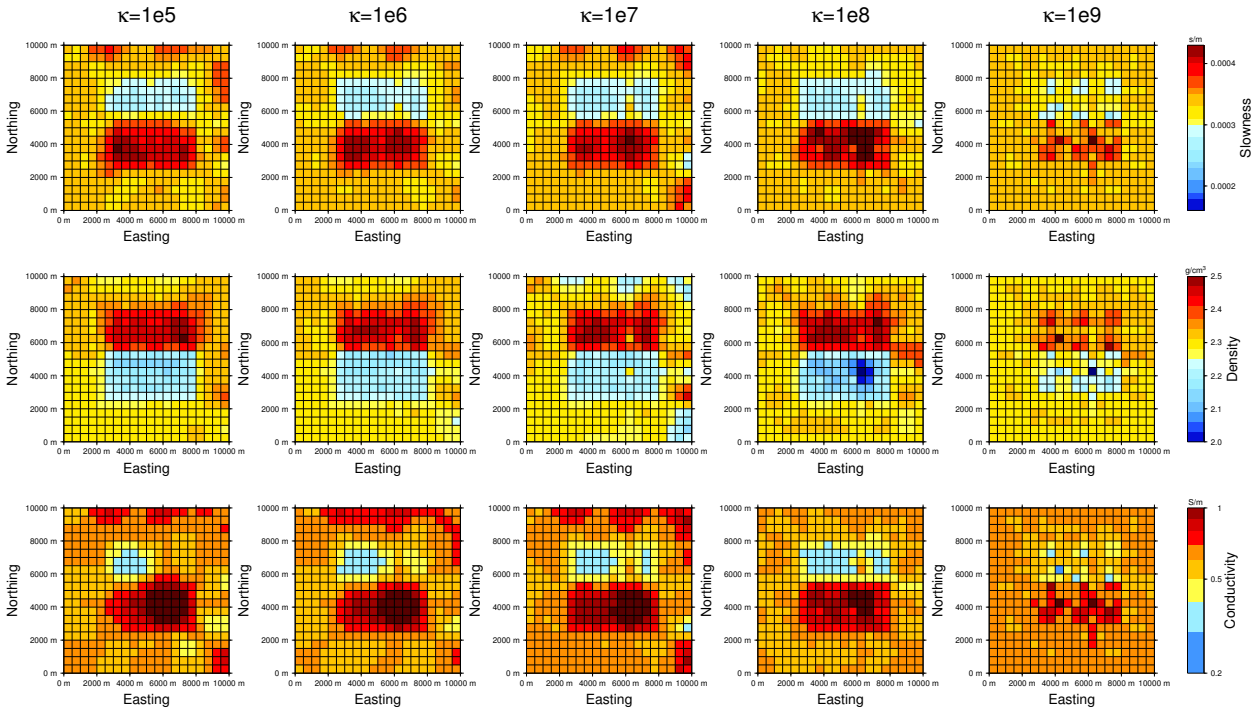

Figure 15. The impact of the weight of the cross-gradient coupling term $\kappa = \alpha_{Cross,s/\rho} = \alpha_{Cross,\sigma/\rho} = \alpha_{Cross,\sigma/s}$ for the incompatible test case. As before, higher values increase the coupling and therefore the similarity between the models for different physical parameters. We keep all other weights constant.

FTG data, the first arrival times and scalar gravity data in contrast are fitted to a similar level as for the compatible case.

For the cross-gradient inversion of the incompatible dataset we perform a similar experiment as before and vary the weight of the cross-gradient term to analyze its impact on the results (Figure 15). As before a coupling weight of $\kappa = 10^9$ reduces the influence of the regularization and results in strong fluctuations of the parameter values in all three models. For all other values of κ we obtain reasonable results. The seismic and gravity anomalies are more or less continuous as in the true models while we can identify a partitioning of the MT model. This effect becomes more pronounced with reduced coupling as this allows the models to be more different from each other. We consider the results for $\kappa = 10^7$ the best although the results for smaller values are similar. Interestingly, in theory we should be able to reconstruct the true model regardless of the weight for the cross-gradient term, as the difference between the true conductivity model and the other two models is chosen so that the cross-gradient functional for the true model vanishes. However, the regularization produces additional gradients with directions depending on data coverage and

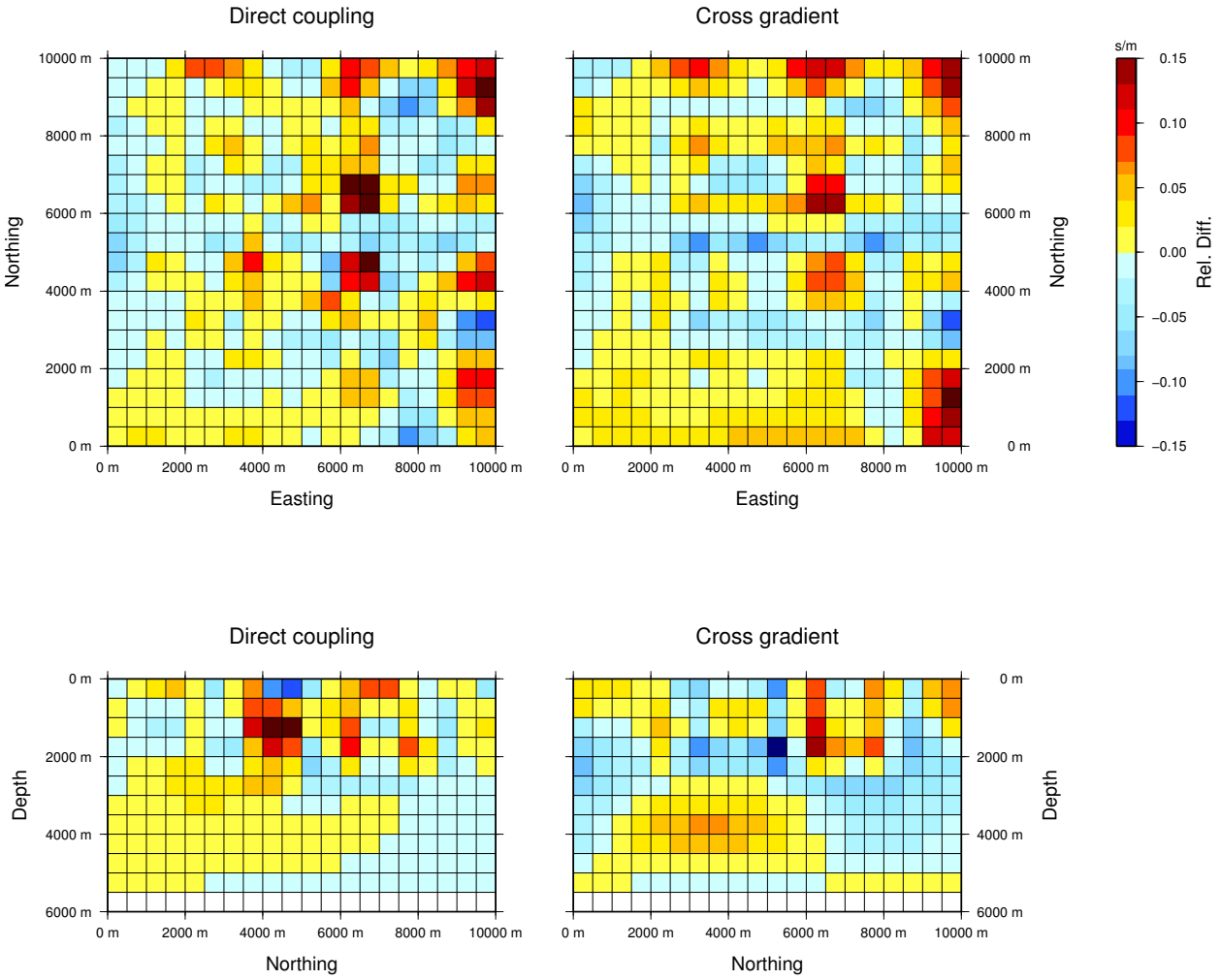

Figure 16. Relative difference $(s^{true} - s)/s^{true}$ between inversion results and true model for the incompatible case and for $\kappa = 10^7$. For each inversion we plot the relative difference for the surface layer (upper row) and a vertical slice through the center of the model (lower row).

resolution characteristics for each dataset. This then results in a complex interaction between the data misfit terms, regularization and cross-gradient terms and an imperfect fit to the data.

For our preferred cross-gradient model with $\kappa = 10^7$ and the direct parameter coupling we again plot the relative difference to the true model for the slowness model. This plot confirms our previous observations: For the direct coupling we obtain strong deviations from the true model mostly where the parameter relationship is violated, but also in the surrounding regions. For the cross-gradient coupling the deviations are smaller. Also, the strongest deviations are located at the transition between the anomalies where the regularization acts strongest and at border of the inversion domain where we have limited data coverage. This demonstrates that in cases where

there is doubt about the structural coherence for the different parameters in the joint inversion, cross-gradient coupling is an adequate approach.

4 CONCLUSIONS

We have presented first joint inversion results for a simple test model and examined the impact of various factors on the inversion results. The flexible structure of the joint inversion framework allows us to change the model parametrization and number of objective functions in order to investigate different coupling approaches. Obviously, the simple geometry of our synthetic test model is not representative for a possible geological target and the inversion domain is small for a 3D geometry. However, this model is well suited to investigate some of the basic properties of joint inversion.

Our results demonstrate that when a direct parameter relationship exists and is utilized in the joint inversion we obtain the largest improvement over single inversion results. Still, cross-gradient coupling, which makes only few assumptions about the relationship between different datasets, results in a considerable improvement. Cross-gradient coupling should therefore be the preferred approach when the existence of a direct relationship between the different physical parameters is doubtful. This conclusion is supported by the experiments where the conductivity structure does not match the structure of the other anomalies. In this case the joint inversion with cross-gradient constraints still produces good results, while the direct parameter coupling results in a number of spurious artifacts. In addition, for the inversion of real data it is rarely possible to fully describe the relationship between different physical parameters by analytical functions. Disregarding deviations from the assumed relationship in the inversion will then result in inversion artifacts and underestimating uncertainties for the model parameters. A possible solution is to include a statistical property relation description in the inversion (Bosch 2004) that allows for deviations from the assumed relationship.

One difficulty with cross-gradient inversion of three different physical parameters is the large numbers of weights that need to be specified. Although our experiments suggest that the exact values of the weights are not critical, the acceptable range for the MT regularization was

$\alpha_\sigma = 10 - 100$ and for the cross-gradient terms $\kappa = 10^5 - 10^7$, we observe a complex interaction between the different terms of the objective function. It is therefore necessary to perform various joint inversions with different weights. This is, however, advisable in any case in order to be able to assess the validity of the joint inversion results. For the compatible test case there was little variation of the anomalies with cross-gradient weight. In all cases there were two continuous anomalies and only the transition to the background changed. For the incompatible case we see how with increasing weight for the cross-gradient term the conductivity anomalies that were split before start to join, an indication of the incompatibility. Furthermore the misfit that we can achieve with the joint inversion compared to single inversion is another important indicator of compatibility.

5 ACKNOWLEDGMENTS

This work was funded by Chevron, ExxonMobil, Nexen, RWE Dea, Shell, Statoil and Wintershall within the JIBA consortium. We thank M. Bosch and an anonymous reviewer for helpful suggestions that improved the clarity of the manuscript. P. Podvin kindly made his eikonal solver publicly available and D. Avdeev provided his 3D MT forward code. A. Avdeeva supported us with implementing the gradient calculation for the MT part of the joint inversion.

REFERENCES

Avdeev, D. & Avdeeva, A., 2009. 3d magnetotelluric inversion using a limited-memory quasi-newton optimization, *Geophysics*, **74**(3), F45–F57.

Avdeev, D. B., Kuvshinov, A. V., Pankratov, O. V., & Newman, G. A., 1997. High-performance three-dimensional electromagnetic modeling using modified neumann series. wide-band numerical solution and examples, *J. Geomagn. Geoelectr.*, **49**(11, 12), 1519–1539.

Avdeeva, A., 2008. *Three-dimensional Magnetotelluric Inversion*, Ph.D. thesis, National University of Ireland.

Baba, K., Chave, A. D., Evans, R. L., Hirth, G., & Mackie, R. L., 2006. Mantle dynamics beneath the East Pacific Rise at 17degS: Insights from the Mantle Electromagnetic and Tomography (MELT) experiment, *Journal of Geophysical Research (Solid Earth)*, **111**(B10), 2101.

Barrère, C., Ebbing, J., & Gernigon, L., 2009. Offshore prolongation of Caledonian structures and basement characterisation in the western Barents Sea from geophysical modelling, *Tectonophysics*, **470**, 71–88.

Bosch, M., 1999. Lithologic tomography: From plural geophysical data to lithology estimation, *Journal of Geophysical Research*, **104**, 749–766.

Bosch, M., 2004. The optimization approach to lithological tomography: Combining seismic data and petrophysics for porosity prediction, *Geophysics*, **69**, 1272–1282.

Brasse, H., Kapinos, G., Mütschard, L., Alvarado, G. E., Worzewski, T., & Jegen, M., 2009. Deep electrical resistivity structure of northwestern Costa Rica, *Geophysical Research Letters*, **36**, L02310.

Chen, C. W., Rondenay, S., Weeraratne, D. S., & Snyder, D. B., 2007. New constraints on the upper mantle structure of the Slave Craton from Rayleigh wave inversion, *Geophysical Research Letters*, **34**, L10301.

Colombo, D. & Stefano, M. D., 2007. Geophysical modeling via simultaneous joint inversion of seismic, gravity, and electromagnetic data: Application to prestack depth imaging, *The Leading Edge*, **26**(3), 326–331.

Commer, M. & Newman, G. A., 2008. New advances in three-dimensional controlled-source electromagnetic inversion, *Geophysical Journal International*, **172**, 513–535.

Commer, M. & Newman, G. A., 2009. Three-dimensional controlled-source electromagnetic and magnetotelluric joint inversion, *Geophysical Journal International*, **178**, 1305–1316.

Constable, S., Key, K., & Lewis, L., 2009. Mapping offshore sedimentary structure using electromagnetic methods and terrain effects in marine magnetotelluric data, *Geophysical Journal International*, **176**, 431–442.

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.

Di Stefano, R. & Chiarabba, C., 2002. Active source tomography at Mt. Vesuvius: Constraints for the magmatic system, *Journal of Geophysical Research (Solid Earth)*, **107**, 2278.

Droujinine, A., Vasilevsky, A., & Evans, R., 2007. Feasibility of using full tensor gradient (FTG) data for detection of local lateral density contrasts during reservoir monitoring, *Geophysical Journal International*, **169**, 795–820.

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.

Gallardo, L. A. & Meju, M. A., 2003. Characterization of heterogeneous near-surface materials by joint 2D inversion of dc resistivity and seismic data, *Geophysical Research Letters*, **30**(13), 1658.

Gallardo, L. A. & Meju, M. A., 2004. Joint two-dimensional DC resistivity and seismic travel time inversion with cross-gradients constraints, *Journal of Geophysical Research (Solid Earth)*, **109**(B18), B03311.

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

Gallardo, L. A., Meju, M. A., & Pérez-Flores, M. A., 2005. A quadratic programming approach for joint image reconstruction: mathematical and geophysical examples, *Inverse Problems*, **21**, 435–452.

Haber, E. & Oldenburg, D. W., 1997. Joint inversion: A structural approach, *Inverse Problems*, **13**, 63–77.

Heincke, B., Jegen, M., & Hobbs, R., 2006. Joint inversion of mt, gravity and seismic data applied to sub-basalt imaging, *SEG Technical Program Expanded Abstracts*, **25**(1), 784–789.

Heincke, B., Maurer, H., Green, A. G., Willenberg, H., Spillmann, T., & Burlini, L., 2006. Characterizing an unstable mountain slope using shallow 2D and 3D seismic tomography, *Geophysics*, **71**, B241–256.

Hu, W., Abubakar, A., & Habashy, T. M., 2009. Joint electromagnetic and seismic inversion using structural constraints, *Geophysics*, **74**, R99–R109.

Jegen, M. D., Hobbs, R. W., Tarits, P., & Chave, A., 2009. Joint inversion of marine magnetotelluric and gravity data incorporating seismic constraints. Preliminary results of sub-basalt imaging off the Faroe Shelf, *Earth and Planetary Science Letters*, **282**, 47–55.

Julia, J., Ammon, C. J., Herrmann, R. B., & Correig, A. M., 2000. Joint inversion of receiver function and surface wave dispersion observations, *Geophysical Journal International*, **143**(1), 99–112.

Jupp, D. L. B. & Vozoff, K., 1975. Stable Iterative Methods for the Inversion of Geophysical Data, *Geophysical Journal*, **42**, 957–976.

Key, K. W., Constable, S. C., & Weiss, C. J., 2006. Mapping 3D salt using the 2D marine magnetotelluric method: Case study from Gemini Prospect, Gulf of Mexico, *Geophysics*, **71**, B17–B27.

Li, X. & Chouteau, M., 1998. Three-dimensional gravity in all space, *Surveys in Geophysics*, **19**, 339–368.

Linde, N., Binley, A., Tryggvason, A., Pedersen, L. B., & Revil, A., 2006. Improved hydrogeophysical characterization using joint inversion of cross-hole electrical resistance and ground-penetrating radar traveltime data, *Water Resources Research*, **42**, 12404.

Maceira, M. & Ammon, C. J., 2009. Joint inversion of surface wave velocity and gravity observations and its application to central Asian basins shear velocity structure, *Journal of Geophysical Research (Solid Earth)*, **114**(B13), B02314.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, vol. 45 of **International Geophysics Series**, Academic Press.

Moorkamp, M., Jegen, M., Roberts, A., & Hobbs, R., 2010a. Massively parallel forward modeling of scalar and tensor gravimetry data, *Computers & Geosciences*, **36**(5), 680 – 686.

Moorkamp, M., Jones, A. G., & Fishwick, S., 2010b. Joint inversion of receiver functions, surface wave dispersion and magnetotelluric data, *Journal of Geophysical Research*, **115**, B04318.

Moré, J. J. & Thuente, D. J., 1994. Line search algorithms with guaranteed sufficient decrease, *ACM Trans. Math. Softw.*, **20**(3), 286–307.

Muñoz, G. & Rath, V., 2006. Beyond smooth inversion: the use of nullspace projection for the exploration of non-uniqueness in mt, *Geophysical Journal International*, **164**, 301–311.

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

Nocedal, J. und Wright, S., 2006. *Numerical Optimization*, Springer.

Parker, R. L., 1983. The magnetotelluric inverse problem, *Geophysical Surveys*, **6**, 5–25.

Pedersen, L. B., Rasmussen, T. M., & Dyrelius, D., 1990. Construction of component maps from aeromagnetic total field anomaly maps, *Geophysical Prospecting*, **38**, 795–804.

Plessix, R. E., 2006. A review of the adjoint-state method for computing the gradient of a functional with geophysical applications, *Geophysical Journal International*, **167**(2), 495–503.

Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophysical Journal International*, **105**, 271–284.

Ritter, J., Jordan, M., Christensen, U. R., & Achauer, U., 2001. A mantle plume below the eifel volcanic fields, germany, *Earth and Planetary Science Letters*, **186**, 7–14.

Roberts, A. W., Hobbs, R. W., Goldstein, M., Max Moorkamp, B. H., & Jegen, M., 2010. Emulation: A bayesian tool for joint inversion, *SEG Technical Program Expanded Abstracts*, **accepted**.

Rodi, W. & Mackie, R. L., 2001. Nonlinear conjugate gradients algorithm for 2-D magnetotelluric inversion, *Geophysics*, **66**, 174–187.

Simpson, F. & Bahr, K., 2005. *Practical Magnetotellurics*, Cambridge University Press.

Siripunvaraporn, W., Egbert, G., Lenbury, Y., & Uyeshima, M., 2005. Three-dimensional magnetotelluric inversion: data-space method, *Physics of the Earth and Planetary Interiors*, **150**, 3–14.

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

Vermeesch, P. M., Morgan, J. V., Christeson, G. L., Barton, P. J., & Surendra, A., 2009. Three-dimensional joint inversion of traveltime and gravity data across the Chicxulub impact crater, *Journal of Geophysical Research (Solid Earth)*, **114**(B13), B02105.

Vozoff, K. & Jupp, D. L. B., 1975. Joint Inversion of Geophysical Data, *Geophysical Journal*, **42**, 977–991.

Wagner, D., Koulakov, I., Rabbel, W., Luehr, B., Wittwer, A., Kopp, H., Bohm, M., & Asch, G., 2007. Joint inversion of active and passive seismic data in Central Java, *Geophysical Journal International*, **170**, 923–932.

Welford, J. K. & Hall, J., 2007. Crustal structure of the Newfoundland rifted continental margin from constrained 3-D gravity inversion, *Geophysical Journal International*, **171**, 890–908.

Zanolla, C., Braitenberg, C., Ebbing, J., Bernabini, M., Bram, K., Gabriel, G., Götze, H., Giammetti, S., Meurers, B., Nicolich, R., & Palmieri, F., 2006. New gravity maps of the Eastern Alps and significance for the crustal structures, *Tectonophysics*, **414**, 127–143.

Zelt, C. A. & Smith, R. B., 1992. Seismic traveltime inversion for 2-D crustal velocity structure, *Geophysical Journal International*, **108**, 16–34.

Zelt, C. A., Azaria, A., & Levander, A., 2006. 3D seismic refraction traveltime tomography at a groundwater contamination site, *Geophysics*, **71**, H67–H78.