Geophysical Journal International

Geophys. J. Int. (2016) **204,** 94–110 doi: 10.1093/gji/ggv411

GJI Geomagnetism, rock magnetism and palaeomagnetism

3-dimensional magnetotelluric inversion including topography using deformed hexahedral edge finite elements and direct solvers parallelized on symmetric multiprocessor computers – Part II: direct data-space inverse solution

M. Kordy,[1,2] P. Wannamaker,[2] V. Maris,[2] E. Cherkaev[1] and G. Hill[3]

[1]*Department of Mathematics, University of Utah.,* 155 *S* 1400 *E Room* 233, *Salt Lake City, UT* 84112-0090, *USA. E-mail: kordy@math.utah.edu*
[2]*Energy and Geoscience Institute, University of Utah,* 423 *Wakara Way, Suite* 300, *Salt Lake City, UT* 84108, *USA*
[3]*Gateway Antarctica, University of Canterbury, Private Bag* 4800 *Christchurch* 8140 *and Antarctica Scientific Limited, Unit 5E/39 Taranaki Street Wellington* 6011

Accepted 2015 September 23. Received 2015 September 22; in original form 2014 November 3

SUMMARY

Following the creation described in Part I of a deformable edge finite-element simulator for 3-D magnetotelluric (MT) responses using direct solvers, in Part II we develop an algorithm named HexMT for 3-D regularized inversion of MT data including topography. Direct solvers parallelized on large-RAM, symmetric multiprocessor (SMP) workstations are used also for the Gauss–Newton model update. By exploiting the data-space approach, the computational cost of the model update becomes much less in both time and computer memory than the cost of the forward simulation. In order to regularize using the second norm of the gradient, we factor the matrix related to the regularization term and apply its inverse to the Jacobian, which is done using the MKL PARDISO library. For dense matrix multiplication and factorization related to the model update, we use the PLASMA library which shows very good scalability across processor cores. A synthetic test inversion using a simple hill model shows that including topography can be important; in this case depression of the electric field by the hill can cause false conductors at depth or mask the presence of resistive structure. With a simple model of two buried bricks, a uniform spatial weighting for the norm of model smoothing recovered more accurate locations for the tomographic images compared to weightings which were a function of parameter Jacobians. We implement joint inversion for static distortion matrices tested using the Dublin secret model 2, for which we are able to reduce nRMS to ∼1.1 while avoiding oscillatory convergence. Finally we test the code on field data by inverting full impedance and tipper MT responses collected around Mount St Helens in the Cascade volcanic chain. Among several prominent structures, the north–south trending, eruption-controlling shear zone is clearly imaged in the inversion.

Key words: Numerical solutions; Inverse theory; Electrical properties; Magnetotellurics; Volcanic arc processes; Explosive volcanism.

1 INTRODUCTION

In Part I (Kordy *et al.* 2015), we have shown that moderately large 3-D magnetotelluric (MT) models including topography can be simulated accurately in practical run-times using a direct solver on a single-box, server-class multicore workstation with large RAM. The deformable mesh approach allows us to avoid expending many rows of cells to define just the topography as is done typically with finite differences, and which even then may not escape local electric field distortion (e.g. Liu *et al.* 2009; Stark *et al.* 2013). The MKL PARDISO library is effective on this platform, showing an overall scalability of 15 on 24 cores. For a mesh with 176 cells in x-direction, 176 cells in y-direction and 70 cells in z-direction (176x 176y 70z), 2000 source vectors (corresponding to 400 MT sites) could be solved in 2.5 times the time required for factorization, with total time for both under 2.5 hr. Meshes comparable to that could simulate site arrays of similar size to the Earthscope MT Transportable Array of the U.S. Pacific Northwest using this parallelized direct solver (Meqbel *et al.* 2014).

Here in Part II, we also use direct solvers exclusively to create a 3-D regularized inversion algorithm for MT data including topography, which we name HexMT. Due to its good convergence

properties, we pursue a Gauss–Newton formulation for the nonlinear, iterative parameter update, as have others (deGroot-Hedlin & Constable 1990; Key & Constable 2011; Grayver *et al.* 2013; Oldenburg *et al.* 2013). The number of parameters usually is significantly greater than the number of data for tomographic-style, regularized inversion. As noted by Siripunvaraporn *et al.* (2005a), inverse formulations using fewer parameters than data may suffer from a dependence of solution upon parametrization. One may also expect some lack of fit to data to occur if parameters are not defined optimally. On the other hand, tomographic inversions for MT data sets of a few hundred sites may require a number of parameters of order one million (e.g. Meqbel *et al.* 2014). Direct factorization of the reduced Hessian matrix in the traditional model-space definition (e.g. deGroot-Hedlin & Constable 1990; Sasaki 2001; Usui 2015), even using parallelization across multicore (Maris & Wannamaker 2010), is not practical for that scale of parametrization. As a result, researchers have tended to retain iterative solvers for the model update whether cast as Gauss–Newton or otherwise (e.g. Commer & Newman 2008; Zhdanov *et al.* 2011; Grayver *et al.* 2013; Schwarzbach & Haber 2013).

An alternative is to investigate the data-space formulation for solving the Gauss–Newton model update (Parker 1994; Siripunvaraporn & Egbert 2000; Siripunvaraporn *et al.* 2005a). This approach reduces the size of the matrix that needs to be inverted from $N_m \times N_m$ to $N_d \times N_d$ (m = model parameters, d = data), while the solutions in theory are identical. Consider an MT survey of 400 sites with 20 frequencies (four per decade say) and 12 data per frequency (four complex impedance and two complex tipper elements). The total data set size would be 96 000. As we show, this turns out to be a very manageable size of matrix to invert using direct solvers, particularly as parallelized across multicore symmetric multiprocessor (SMP) computers. Matrices twice this size in fact are not impractical, allowing data sets of more sites, greater bandwidth, or finer frequency sampling, with a fairly arbitrary number of model parameters.

This paper sets out with a brief overview of both model- and data-space approaches to solving the Gauss–Newton model update. Attention is paid to the mechanics of solving stably the normal system equation for a model gradient regularization functional. Runtime and scalability of the model update solver is investigated for multicore using different sized trial models. At this point it appears that model update solution time will remain significantly smaller than forward simulation run-time across all models with moderately fine parameter discretization. The inversion code is tested on several models. These include a simple conductive brick below a hill to show the strength of effect that topography can have on inversion models assuming a flat surface. Subsequently we examine a multiprism test model used as a community standard (Miensopust *et al.* 2013) and experiment with various regularization weighting schemes. Finally, we invert an extensive MT data set acquired over the volcano Mount St Helens (Hill *et al.* 2009) to show performance for a model where parameter number approaches one million.

2 FORWARD PROBLEM

The forward problem is described in detail by Kordy *et al.* (2015), touched on briefly here to define terms. We consider the MT problem in a domain Ω that includes the air and earth's subsurface. The Earth's surface is allowed to have topography. In order to calculate the MT response due to an arbitrary 3-D conductivity structure $\sigma = \sigma(x, y, z)$ we consider a hexahedral edge finite-element discretization of the equation for the secondary electric field E:

$$\int_\Omega \frac{1}{\mu}\nabla\times E\cdot\nabla\times M + \mathrm{i}\omega\int_\Omega \hat{\sigma}\,E\cdot M = \int_\Omega -\mathrm{i}\omega(\hat{\sigma}-\hat{\sigma}^p)E^p\cdot M \tag{1}$$

for $E, M \in \mathcal{H}_0(\nabla\times,\Omega)$, where ω is angular frequency, $\epsilon > 0$ is dielectric permittivity, μ is magnetic permeability, $\hat{\sigma} = \sigma + \mathrm{i}\omega\epsilon$ and $\hat{\sigma}^p = \sigma_p + \mathrm{i}\omega\epsilon$. E^p is the primary electric field, which is that of an arbitrarily 1-D earth of conductivity structure σ_p. We assume that $\sigma \approx \sigma_p$ close to the domain boundaries. The solution space is defined below:

$$\mathcal{H}_0(\nabla\times,\Omega) = \left\{ M{:}\Omega\to\mathbb{C}^3 : \int_\Omega (|M|^2 + |\nabla\times M|^2) < \infty, \; n\times M|_{\partial\Omega} = 0 \right\}. \tag{2}$$

The approximate solution to eq. (1) is obtained using edge elements. Secondary magnetic field H is calculated as

$$H = \frac{-\nabla\times E}{\mathrm{i}\omega\mu}. \tag{3}$$

The total field E^t, H^t is a sum of secondary and primary fields:

$$E^t = E + E^p, \quad H^t = H + H^p. \tag{4}$$

The MT response is obtained by finding Z, K such that

$$\begin{bmatrix} E^t_x \\ E^t_y \\ H^t_z \end{bmatrix} = \begin{bmatrix} Z_{xx} & Z_{xy} \\ Z_{yz} & Z_{yy} \\ K_{zx} & K_{zy} \end{bmatrix} \begin{bmatrix} H^t_x \\ H^t_y \end{bmatrix} \tag{5}$$

is satisfied no matter what is the polarization of the primary (E^p, H^p) plane wave.

3 GAUSS-NEWTON INVERSION PROCEDURE

For defining inversion terminology, we consider again fig. 1 of Part I (Kordy *et al.* 2015) where layers of hexahedral elements are deformed vertically to represent topography. This was efficient for the forward problem, but also will be for inversion. Although elements below the earth surface could be grouped to form a parameter, for maximal flexibility we usually consider each element as being a possible parameter linked through regularization in a tomographic inversion.

3.1 Description of the method

As is usual, the portion of the model domain Ω below the air–earth interface is split into N_m model parameters, which are disjoint regions with constant resistivity. Let $m = (m_1, \ldots, m_{N_m})$ be the vector of parameter $\log_{10}$ resistivity values. We work with $\log_{10}$ resistivity as this ensures that resistivity remains positive during inversion and makes the square norms of the reduced Hessian matrix columns more nearly equal in magnitude (Hohmann & Raiche 1988). There are N_d data points collected, denoted as $d = (d_1, \ldots, d_{N_d})$. As individual data values, we consider the real and imaginary parts of all entries in Z, K for all N_{rec} receivers, namely $N_d = 12N_{\mathrm{rec}}$ and for all frequencies. Let $e_1, \ldots, e_{N_d}$ be the vector of measurements errors, for which standard deviations s_i are known. By $F(m) \in \mathbb{R}^{N_d}$ we denote the response of the current model m, calculated by the forward code.

Denote B_d as a diagonal matrix with $\frac{1}{s_i^2}$ as entries. The inversion procedure seeks a model m such that the weighted data misfit

$$\|F(m)-d\|_{B_d}^2 = (F(m)-d)^T B_d (F(m)-d) = \sum_{i=1}^{N_d} \left(\frac{F_i(m)-d_i}{s_i}\right)^2 \quad (6)$$

is minimal, together with the constraint that some measure of roughness of the model m is limited. The roughness will be measured by

$$\|m-m_0\|_{B_m}^2 = (m-m_0)^T B_m (m-m_0) \quad (7)$$

where m_0 is a reference model and B_m is a symmetric non-negative definite matrix, so that $\|.\|_{B_m}$ is a seminorm. Often B_m is such that $\|m-m_0\|_{B_m} = \|\nabla(m-m_0)\|_{L_2}$, where ∇ denotes spatial gradient (in all three directions) of the $\log_{10}$ resistivity model. In the deformed mesh geometry we implement, the three directions in general are not purely perpendicular; one remains vertical while the other two lie along the variably deformed layer of elements.

Specifically, in the inversion we seek a model m that minimizes the functional

$$W(m) = \|F(m)-d\|_{B_d}^2 + \lambda \|m-m_0\|_{B_m}^2 \quad (8)$$

for some suitable value of $\lambda > 0$.

The Gauss–Newton is an iterative procedure that seeks a minimizer of (8). It starts with an initial guess m_1. Given a current model m_n, the Gauss–Newton scheme approximates the response $F(m)$ around m_n by the first-order Taylor expansion:

$$F(m) \approx F(m_n) + J\,[m-m_n] \quad (9)$$

where J is a $N_d \times N_m$ matrix of derivatives of F

$$J_{i,j} = \frac{\partial F_i}{\partial m_j}(m), \;\; i = 1,\ldots,N_d, \;\; j = 1,\ldots,N_m \quad (10)$$

whose computation we have described in Part I. If (9) is used, the functional (8) becomes quadratic and the minimizer m_{n+1} satisfies a linear equation:

$$[J^T B_d J + \lambda B_m](m_{n+1}-m_0) = J^T B_d \hat{d} \quad (11)$$

where $\hat{d} = d - F(m_n) + J[m_n - m_0]$.

The reduced Hessian matrix enclosed in square brackets in eq. (11) is dense, symmetric positive definite, and has dimension $N_m \times N_m$. This is the traditional model-space parameter update formulation. The numerical complexity of solving this equation using Cholesky decomposition is $O(N_m^3)$. This cubical growth eventually makes direct solution of the model-space Gauss–Newton scheme impractical for arbitrarily large model size N_m.

The data-space method (Parker 1994; Siripunvaraporn & Egbert 2000; Siripunvaraporn *et al.* 2005a) replaces eq. (11) with a linear equation having only N_d unknowns. For the moment we assume that B_m is invertible (which implies that B_m is positive definite and $\|.\|_{B_m}$ is a norm), the treatment of which we will revisit shortly. When (11) is left-multiplied by B_m^{-1}, we obtain:

$$B_m^{-1}[J^T B_d J + \lambda B_m](m_{n+1}-m_0) = B_m^{-1} J^T B_d \hat{d}$$

$$B_m^{-1} J^T (B_d J)(m_{n+1}-m_0) + \lambda(m_{n+1}-m_0) = B_m^{-1} J^T B_d \hat{d}$$

$$m_{n+1}-m_0 = B_m^{-1} J^T \frac{1}{\lambda} B_d [\hat{d} - J(m_{n+1}-m_0)]$$

This proves that

$$m_{n+1}-m_0 = B_m^{-1} J^T \beta \quad (12)$$

for some $\beta \in \mathbb{R}^{N_d}$. When (12) is plugged into (11) and the equation is left-multiplied by B_m^{-1}, we obtain an equation equivalent to (11):

$$B_m^{-1} J^T [B_d J B_m^{-1} J^T + \lambda I]\beta = B_m^{-1} J^T B_d \hat{d}. \quad (13)$$

This equation will be satisfied if β satisfies

$$[B_d J B_m^{-1} J^T + \lambda I]\beta = B_d \hat{d} \quad (14)$$

which is equivalent to

$$[J B_m^{-1} J^T + \lambda B_d^{-1}]\beta = \hat{d}. \quad (15)$$

The latter equation has a unique solution as $J B_m^{-1} J^T$ is symmetric nonnegative definite and B_d^{-1} is symmetric positive definite. The data-space Gauss–Newton method finds β, the solution to (15) and uses (12) to calculate a model update m_{n+1}.

The model update m_{n+1} in eq. (11) minimizes the inversion functional W using a linearized forward response F. However, the update may fail to decrease W if it is too far from current model m_n for the approximation in (9) to be accurate. A line search in the direction $m_{n+1} - m_n$ to minimize W (Nocedal & Wright 2006) can avoid this, but at the expense of multiple forward problems. One could use the Levenberg–Marquardt algorithm (Levenberg 1944) to keep the update m_{n+1} close to m_n by adding a term $\alpha\|m-m_n\|^2$ to W. This may be interpreted as choosing m_{n+1} that minimizes W for a trust region (Nocedal & Wright 2006) of models $\{m\colon \|m-m_n\|_2 \le \delta\}$ for some $\delta > 0$ than depends on α. The regularization term in W has a similar stabilizing effect, yet with a trust region defined as $\{m : \|m-m_0\|_{B_m} \le \delta\}$ for some $\delta > 0$ dependent on λ.

Apart from minimizing W, one also should select a value of λ to minimize the norm $\|m-m_0\|_{B_m}$ among models having a specific data misfit of X^2 (Parker 1994). A well-known method for this is Occam's inversion (Constable *et al.* 1987; deGroot-Hedlin & Constable 1990). As Occam also involves multiple forward calculations per iteration, we simply begin with a large value of λ and decrease it for subsequent iterations during the minimization. More precisely, λ that is used to obtain model m_{n+1} is set as:

$$\lambda = \mathrm{nRMS}(m_j)\cdot\kappa \quad (16)$$

From an initial value $\kappa_0 > 0$, parameter κ will be decreased by a factor of two if the normalized root mean square (nRMS) at an iteration does not fall by more than a user-specified amount (e.g. 10%). As a result in our inversions, the parameter λ steadily decreases and the model acquires increasing amounts of structure. This may be considered a form of cooling strategy (*cf.* Haber *et al.* 2007). The scaling with respect to nRMS is consistent with experience of Constable *et al.* (1987) where the optimal λ decreased as iterations proceeded and misfit improved. Procedures for determining λ warrant further investigation.

3.2 Computational considerations

In the data-space method, one has to invert B_m and apply it to J^T in order to calculate $J B_m^{-1} J^T$. In Siripunvaraporn & Egbert (2000), matrix B_m^{-1} was denoted C_m and called the model covariance matrix. It was not defined explicitly as a result of inverting a norm matrix B_m, but was treated as a natural matrix to consider for regularization.

The choice of a proper regularization functional $\|.\|_{B_m}$ is important, as minimizing the functional W in (8) is equivalent to minimizing the data misfit

$$(F(m)-d)^T B_d (F(m)-d) \quad (17)$$

subject to a condition on the model norm

$$\|m - m_0\|_{B_m} \leq \delta \tag{18}$$

where $\delta > 0$ depends on the choice of λ. The regularization functional we consider is L_2 norm of the gradient of the model,

$$\|m - m_0\|_{B_m} = \|\nabla(m - m_0)\|_{L_2}. \tag{19}$$

In order to use this functional in data-space method, matrix B_m must be inverted. However if (19) is used, B_m is non-negative definite and thus singular. To make it positive definite, we add a small number $\epsilon > 0$ to its diagonal before inverting. The functional is negligibly modified as then

$$\|v\|_{B_m} = \sqrt{\|\nabla v\|^2_{L_2} + \epsilon \sum_{i=1}^{N_m} v_i^2} \tag{20}$$

The nonzero value of ϵ tends to keep the inversion model slightly closer to the *a priori* model, which we do not view as a drawback. For very small values of ϵ, this effect is negligible.

To estimate the cost of calculation of $B_m^{-1}J^T$, we exploit the fact that matrix B_m has a nonzero pattern of a matrix coming from finite-difference approximation of a scalar Poisson equation over the inversion cell grid. Even if each inversion cell consists of only one element, the number of inversion cells will be no more than the number of vertices in the earth's subsurface. Thus B_m has less nonzero entries and less variables than the matrix used for the divergence correction described in Part I and the number of variables should be less. Even though the number of linear systems to be solved is $12 \times N_{\text{rec}}$, and for the divergence correction it was $5 \times N_{\text{rec}}$, here all variables are real valued. Thus if the solution library MKL PARDISO is used, the time of factorization of B_m and applying B_m^{-1} to J^T is expected to be less than the cost of applying the divergence correction, which takes a fraction of time of the forward problem. One concludes that the calculation of $B_m^{-1}J^T$ will not add a significant execution time to the inversion process no matter how large the grid is, as long as the direct solver is used for forward modeling.

For matrix multiplications like $J^T B_d J$ and the Cholesky factorization needed to solve eqs (11) and (15), we use the PLASMA library (Buttari *et al.* 2009; Baboulin *et al.* 2011). PLASMA is a linear algebra library for dense matrices, parallelized for shared memory machines (such as the SMP unit we use). It employs a matrix tiling approach (*cf.* Baboulin *et al.* 2005; Maris & Wannamaker 2010; Kordy *et al.* 2013), which reduces the time of transporting the matrix entries from RAM to CPU. The scalability of Cholesky factorization and matrix multiplication using PLASMA is presented in Fig. 1. The speedup using 24 cores is ∼17 and ∼19 for Cholesky factorization and matrix multiplication, respectively.

To assemble the model-space Gauss–Newton matrix (11), one has to evaluate $J^T B_d J$, which has numerical complexity $O(N_m^2 N_d)$. Solving the matrix as noted previously has complexity $O(N_m^3)$. For the data-space method on the other hand, to assemble the matrix in eq. (15) one has to evaluate $(B_m^{-1}J^T)$. By analysing the run-times for models 2–5 of Table 1, we estimate the numerical complexity to be $O(N_m^{1.13})$, $O(N_m^{1.5})$ and $O(N_d N_m^{1.18})$ for the reordering, factorization and the solution phases, respectively. Most of the time (over 98%) for the considered test cases is spent in the solution phase. With the direct solver approach, the time to calculate $(B_m^{-1}J^T)$ is always a small fraction of the cost of the forward problem. Further, $(B_m^{-1}J^T)$ is used to evaluate $J(B_m^{-1}J^T)$, which has numerical complexity $O(N_m N_d^2)$. Solving the equation

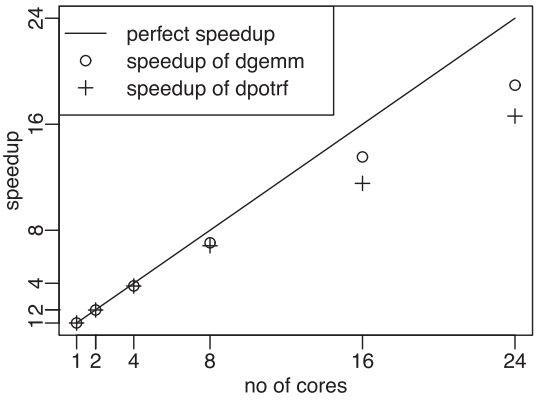

Figure 1. Speedup of PLASMA library for Cholesky factorization (dpotrf) and matrix multiplication (dgemm).

using Cholesky decomposition has complexity $O(N_d^3)$. As typically $N_d < N_m$, the computational cost associated with data-space method is less. The difference becomes more pronounced for larger MT surveys.

Example computation times for models listed in Table 1(a) are presented in Table 1(b). The time to solve the model-space Gauss–Newton eq. (11) increases rapidly with the model size and quickly gets impractical, reaching over 27 hr for the largest grid considered. On the other hand, the time to solve the data-space eq. (15) remains short, less than one minute for all the grids considered. In the case of the data-space method, more time consuming than solving eq. (15) is evaluating $J(B_m^{-1}J^T)$, which takes 1 hr for the largest mesh considered. Nevertheless, the corresponding evaluation of $J^T B_d J$ for model-space Gauss–Newton takes longer, more than seven times longer for mesh 5. Comparable to the time of calculation of $J(B_m^{-1}J^T)$ is the time of evaluation of $B_m^{-1}J^T$. However, as we expected, this time is less than that of applying the divergence correction (more than two times less), which in turn is almost 10 times less than the total time spent solving forward problem. Thus the advantage of the data-space Gauss–Newton approach over the model-space version for this application is clear.

Concerning RAM requirements, for model-space GN one needs to store the matrix J, which is $N_d \times N_m$ as well as the matrix in the eq. (11), which is a symmetric $N_m \times N_m$ matrix. In the data-space version, one needs to store matrix J and the matrix $B_m^{-1}J^T$ of the same size, but depending on implementation those may be saved on hard disk and parts of them may be read into RAM memory when needed. Also, one needs to store the matrix of eq. (15), which is a symmetric $N_d \times N_d$ matrix. One can see that as typically $N_d < N_m$ the memory requirements are smaller for data-space Gauss–Newton than for its model-space cousin.

In Table 1(c) we present the RAM memory requirements for the grids considered. The matrix J is $N_d \times N_m$ in size, but largely can be stored on hard disk and parts accessed as needed. For the largest mesh considered, the model-space GN update requires 413 GB, whereas the data-space update requires as little as 7.9 GB of RAM memory depending upon treatment of J.

3.3 Regularization norm weight

Up to this point, we have not specified details of the entries of matrix B_m, other than that it is a finite-difference representation of spatial gradients in the model parameter vector m. Several investigators have explored whether entries of B_m should also be weighted

Table 1. Run-times and memory use for forward modeling, model-space Gauss–Newton iteration and data-space Gauss–Newton iteration.

ID	FEM grid	inversion grid	edges no	recv no	freq no	N_d	N_m
1	$41x\ 41y\ 30z$	$39x\ 39y\ 19z$	143 120	81	13	12 636	28 899
2	$65x\ 65y\ 45z$	$63x\ 63y\ 29z$	550 400	121	15	21 780	115 101
3	$81x\ 81y\ 47z$	$79x\ 79y\ 31z$	896 960	64	15	11 520	193 471
4	$81x\ 81y\ 47z$	$79x\ 79y\ 31z$	896 960	196	15	35 280	193 471
5	$101x\ 101y\ 50z$	$99x\ 99y\ 34z$	1,489 800	256	15	46 080	333 234

(a) Statistics of test meshes used for run times testing.

ID	FP total	FP DC	GN J^TB_dJ	GN slv	DS $B_m^{-1}J^T$	DS JJ^T	DS slv	DS tot.
1	00:03:57	00:00:30	00:00:38	00:00:25	00:00:13	00:00:24	00:00:03	00:04:46
2	00:39:47	00:04:32	00:14:42	00:25:09	00:02:16	00:03:40	00:00:15	00:46:41
3	01:00:14	00:05:33	00:22:18	01:59:12	00:02:02	00:02:11	00:00:03	01:05:12
4	02:04:20	00:11:53	01:05:05	01:59:27	00:07:28	00:15:29	00:01:01	02:30:36
5	04:04:15	00:28:52	03:55:47*	10:00:49*	00:19:15	00:44:28	00:02:22	05:07:58

(b) Run times in format hh:mm:ss for meshes listed in table 1(a). "FP" denotes the total time of calculation of the forward problem (response $F(m)$ and Jacobian J) using MKL PARDISO. "FP DC" is the time spent on divergence correction using MKL PARDISO (fraction of "FP"). "GN J^TB_dJ" denotes the time spent on evaluation of J^TB_dJ using PLASMA. "GN slv" denotes time needed to solve Gauss-Newton equation (11) using PLASMA once the matrix $J^TB_dJ + \lambda B_m$ has been assembled. "DS $B_m^{-1}J^T$" denotes the time spent to calculate $B_m^{-1}J^T$ using MKL PARDISO. "DS JJ^T" denotes the time spent to evaluate $J(B_m^{-1}J^T)$ using PLASMA. "DS slv" denotes the time spent to solve equation (15) using PLASMA, once the matrix $JB_m^{-1}J^T + \lambda B_d^{-1}$ has been assembled. * denotes the estimated time, the calculation hasn't been done due to insufficient RAM memory. "DS tot." denotes the total time of one data space Gauss-Newton iteration. For the GN solve time, the estimate has been calculated by fitting a line for $\log(time)$ and $\log(N_m)$. The estimate for the matrix multiplication time has been calculated by fitting a linear model for $\log(time)$ as a function of $\log(N_m)$ and $\log(N_d)$. The calculations have been done on a 24-core workstation (four Intel Xeon E5-4607 v2 Hexa-core 2.60 GHz processors)

ID	N_m	N_d	FP	J	GN	DS
1	28 899	12 636	1.8 GB	2.7 GB	3.1 GB	0.6 GB
2	115 101	21 780	10.1 GB	18.7 GB	49.4 GB	1.8 GB
3	193 471	11 520	20.0 GB	16.6 GB	139.4 GB	0.5 GB
4	193 471	35 280	20.0 GB	50.9 GB	139.4 GB	4.6 GB
5	333 234	46 080	34.2 GB	114.4 GB	413.7 GB	7.9 GB

(c) RAM memory needed for matrices related to Gauss-Newton update using model-space and data-space for meshes listed in table 1(a). "FP" denotes RAM needed to calculate the forward problem. "J" denotes the memory needed to store the matrix J of size $N_m \times N_d$. "GN" denotes the memory needed to store the (symmetric) matrix $J^TB_dJ + \lambda B_m$ of size $N_m \times N_m$. "DS" denotes the memory needed to store the (symmetric) matrix $JB_m^{-1}J^T + \lambda B_d^{-1}$ of size $N_d \times N_d$.

according to influence (Jacobian) of their corresponding parameter (e.g. Zhdanov 2002; Yi *et al.* 2003). Here we present three different regularization functionals, the performance of which will be compared in numerical tests.

Consider the infinite-dimensional problem and its response $F(m)$ given a spatially varying $\log_{10}$ resistivity model $m = m(\mathbf{r})$. Also consider the derivative $S(\mathbf{r})$ of F with respect to m satisfying:

$$F(m + \delta m) \approx F(m) + \int_\Omega S(\mathbf{r})\delta m(\mathbf{r})d\mathbf{r} \tag{21}$$

for a small change in model δm. The quantity $\|S(\mathbf{r})\|_2$ measures the sensitivity of the response F to the change of the conductivity at the point $\mathbf{r}$.

As the regularization functional $(m - m_0)^TB_m(m - m_0)$, we will consider L_2 norms of the gradient of m with a weight $\nu(\mathbf{r}) > 0$ defined as follows:

$$\|\nabla(m - m_0)\|^2_{L_2(\nu)} = \int_\Omega |\nabla(m - m_0)|^2\nu(\mathbf{r})d\mathbf{r} \tag{22}$$

Further, we consider three possible values for ν:

$$\begin{aligned}\nu(\mathbf{r}) &= \|S(\mathbf{r})\|_2\\ \nu(\mathbf{r}) &= 1\\ \nu(\mathbf{r}) &= \frac{1}{\|S(\mathbf{r})\|_2}\end{aligned} \tag{23}$$

Norm $\|\nabla(m - m_0)\|_{L_2(1)}$ uses no information about the influence of the inversion cell on the data. If norm $\|\nabla(m - m_0)\|_{L_2(\|S\|_2)}$ is used for regularization, smoothing is suppressed for parameter regions with low sensitivity allowing them to show additional structure (*cf.* Zhdanov 2002). Norm $\|\nabla(m - m_0)\|_{L_2(\frac{1}{\|S\|_2})}$ will smooth the model in regions with low sensitivity, using the reasoning that if we cannot detect the properties of a region well, we will make it similar to its surroundings. This is similar to the approach of Yi *et al.* (2003), although they make a rigorous evaluation of the parameter resolution matrix which is computationally intractible for the larger problems we consider here.

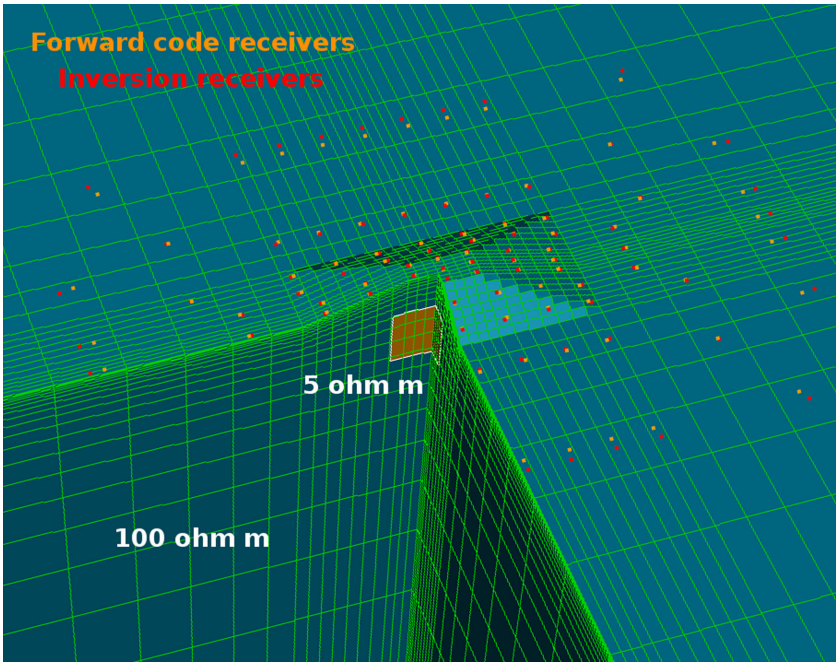

Figure 2. Central part of the inversion grid together with the receiver locations in three quadrants. Conductive brick is shown below the hill.

If the norms $\|m - m_0\|_{L_2(v)}$ and $\|\nabla(m - m_0)\|_{L_2(v)}$ are properly approximated on a discrete level, they will be mesh independent. An explanation of this claim and the details of the way we approximate the norms are provided in Appendix A.

4 SYNTHETIC INVERSION EXAMPLES

In this section, we present results of the inversion of synthetic MT data to evaluate algorithm performance under controlled conditions. As a measurement error we will use the value

$$\begin{aligned} e(Z_{ij}) &= \max\left\{3.5\%\tfrac{|Z_{xy}-Z_{yx}|}{2}\right\}, \quad i,j = x,y \\ e(K_{zj}) &= 0.03, \qquad\qquad\qquad\quad j = x,y \end{aligned} \tag{24}$$

We use $Z_{xy} - Z_{yx}$ because it is a rotational invariant and shows relative stability to data noise (see Groom & Bahr 1992). As the measurement error s_j (noted in (6)) for real or imaginary part of Z and K, we take the above value e. The data used in the inversion are calculated by the forward problem for the true conductivity model, with Gaussian noise having zero mean and standard deviation s added to the real and imaginary parts of Z and K.

To assess goodness of fit of a model response to the data, we use the nRMS, defined as:

$$\mathrm{nRMS}(m) = \sqrt{\frac{1}{N_d}\sum_{j=1}^{N_d}\left(\frac{d_j - F_j(m)}{s_j}\right)^2} \tag{25}$$

where d is the vector of our synthetic data, F is a vector of response of the model m and s is the measurement error vector.

4.1 Brick under a hill

Our first model is a brick under a hill in 100 Ωm background. The hill has dimensions 2000 and 4000 m in x- and y-directions at the bottom and 500 and 1000 m at the top. The hill is 450 m high. The object is placed below the hill with the top and the bottom of the object at 650 and 1600 m, respectively, below the top of the hill and its XY cross-section is a rectangle $[-328\text{ m}, 328\text{ m}] \times [-700\text{ m}, 700\text{ m}]$. We consider a conductive (5 Ωm) and resistive (2000 Ωm) object as well as no object at all. We compare the inversion that has the mesh conforming to the topography as in Fig. 2 to the mesh without topography (flat surface). Both meshes have the same location of cells in x- and y-directions and the same x and y coordinates of receivers. The only difference is the elevation of layers close to the earth surface.

We generated the data using a different grid than the one used for inversion. The forward code grid was $95x$, $95y$, $50z$ and extended to 18 km from the grid centre in x- and y-directions, 5.6 km above the earth's surface (air layer) and 12.5 km below the surface. The inversion grid was $41x$, $41y$ and $30z$. It extended 14 and 15 km from the centre of the grid in x- and y-directions, respectively. There were 106 receivers. The location of a receiver is always at the centre of the face of an element lying on the earth's surface, thus the location of the forward code receivers is slightly different than the inversion receivers. The inversion grid, together with the location of the brick and receivers in three of the four quadrants is presented in cutaway view in Fig. 2.

The data consisted of the impedance Z and the tipper K for 13 frequencies between 1 and 1000 Hz distributed evenly in $\log_{10}$ space. We added Gaussian error with standard deviation (24) to the forward data. The initial value of κ_0 started at the same value for all inversions. The starting and reference (*a priori*) models were set to 50 Ωm uniformly. The regularization functional used was $\|\nabla m\|_{L_2(1)}$.

Iteration history is presented in Fig. 3. The regularization norm $\|\nabla m\|_{L_2(1)}$ increases as the inversion proceeds and λ decreases in the effort to decrease nRMS. One can see that inversion with topography is able to achieve nRMS close to 1 in less than 3 iterations, whereas the inversion without topography is struggling to decrease nRMS below 1.6 even though the model norm is larger than in the case of inversion with topography.

We have plotted cross-sections of selected models for six inversions in Figs 4 and 5 for comparison. In all cases, the inversion with topography is able to recover a smoothed version of the original

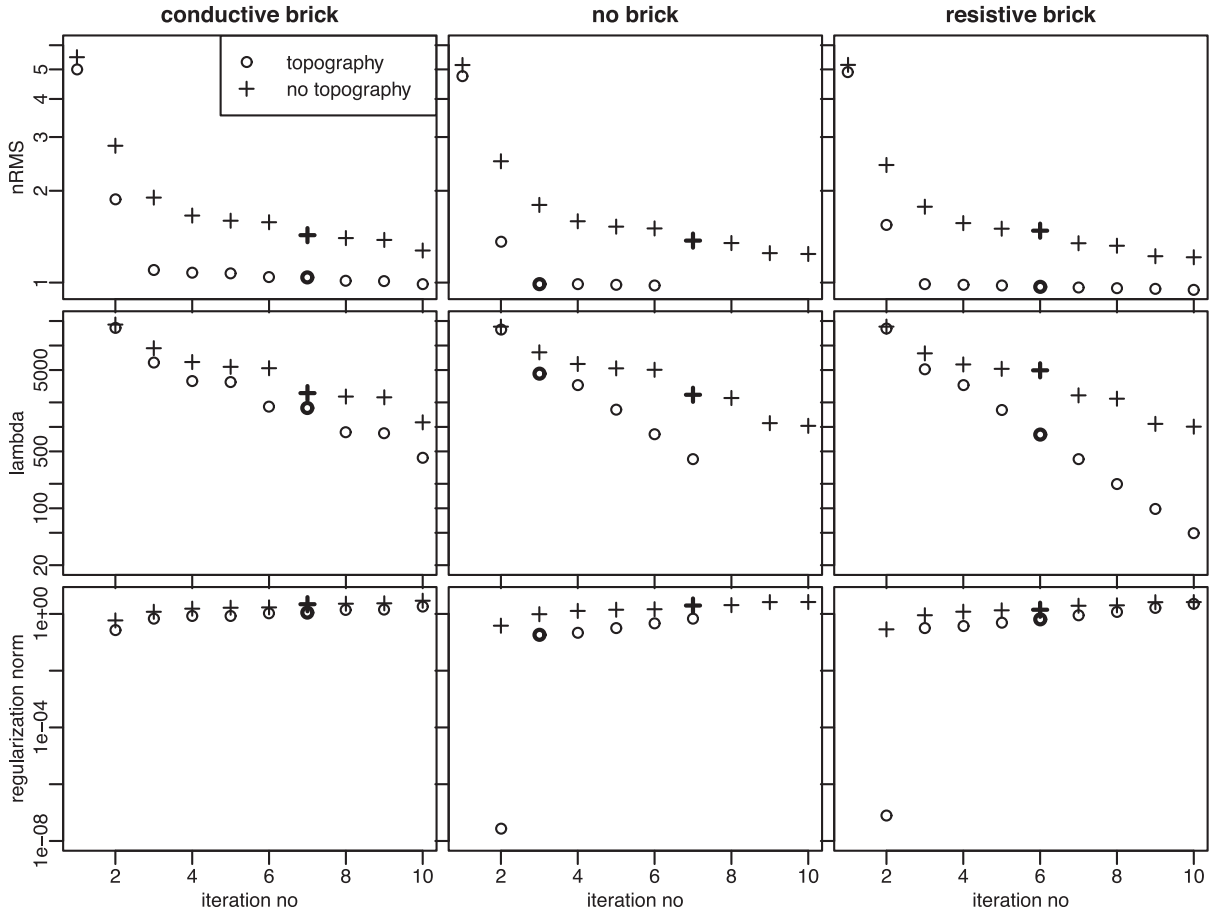

Figure 3. Inversion iteration history for model of bricks under a hill. nRMS(m_j), λ used to obtain model m_j, $\|\nabla m\|_{L_2(1)}$ as a function of iteration number j are plotted. The models plotted in Figs 4 and 5 are denoted by bold symbols.

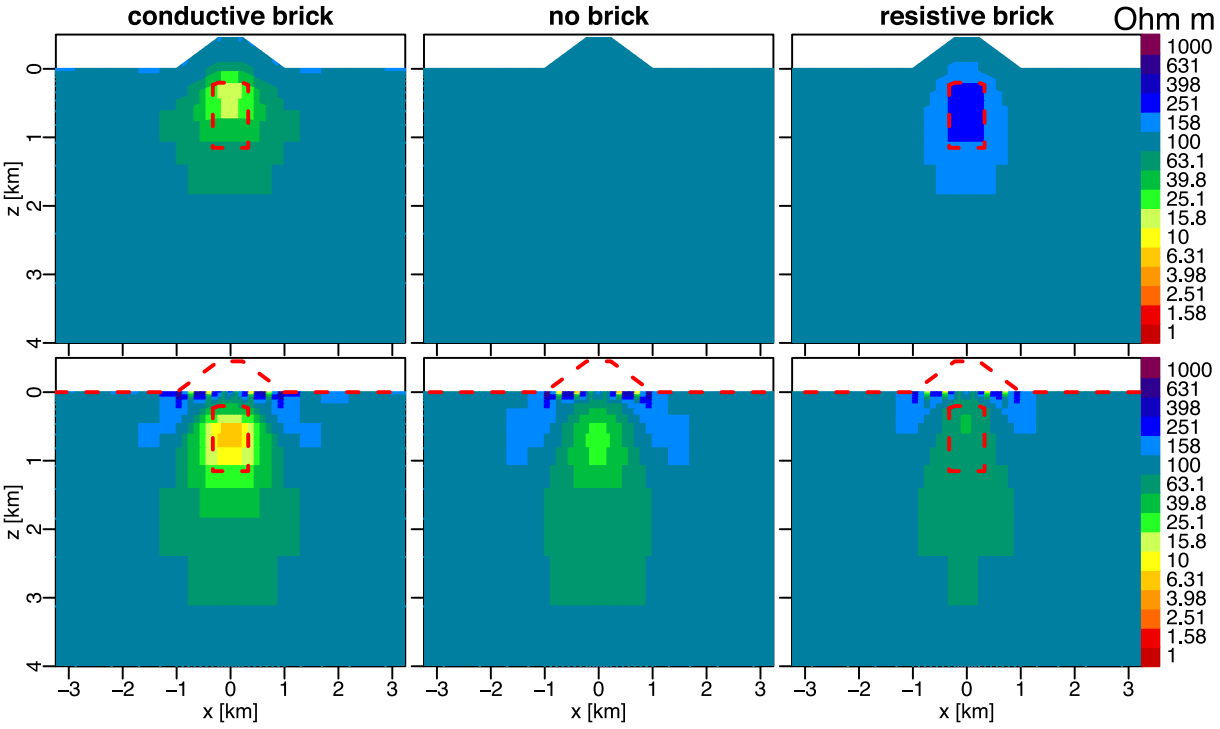

Figure 4. Inversion results for bricks under a hill along XZ cross-section at $y = 0$ km. Top row shows inversion with topography, bottom row the inversion without topography.

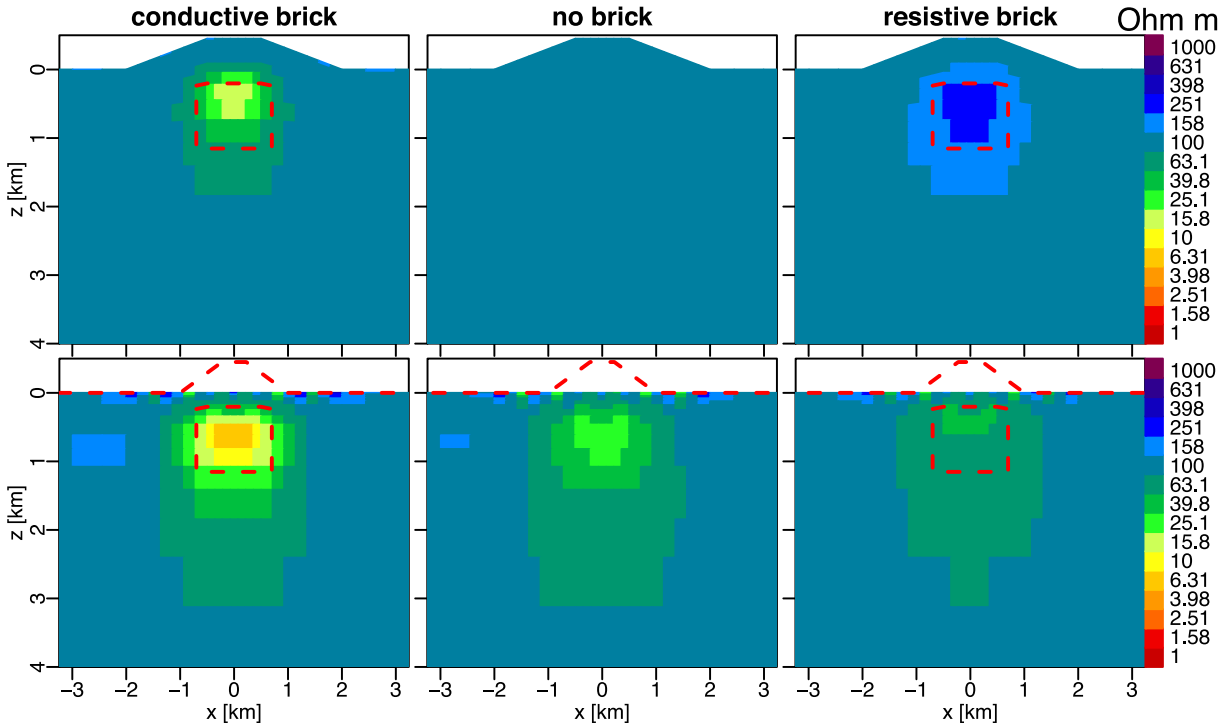

Figure 5. Inversion results for bricks under a hill along *YZ* cross-section at $x = 0$ km. Top row shows inversion with topography, bottom row the inversion without topography.

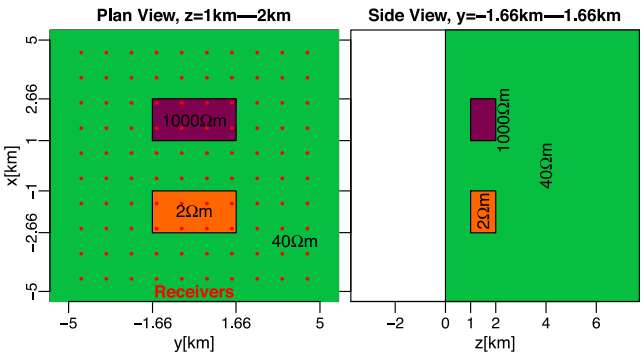

Figure 6. Sketch of two bricks model.

object (or no object in the no brick case). Inversion without topography puts a more conductive object below the ground to make-up for the absence of a hill. This occurs because the electric field is reduced by the hill as background electric current only partially flows upward into that volume (see TM mode results in Wannamaker *et al.* 1986). Even for the resistive brick forward data, the inversion without topography returns a (somewhat) conductive object. The inversion also creates an oscillatory region above the object (more apparent on *XZ* cross-section) that resembles the shape of a hill. These results emphasize the importance of including the topography in the inversion of MT data.

4.2 Simple two brick model

Our next synthetic model consists simply of two buried and separated bricks, one conductive (2 Ωm) and one resistive (1000 Ωm), in a 40 Ωm half-space (Fig. 6). With this model we examine the effect of inversion regularization weights on model characteristics.

The forward mesh consists of 58x 58y 45z elements. In the *XY* plane, the central 33 × 33 elements are square with sides equal to 0.333 km. Further from the centre, the element sizes grow gradually and extend 130 km from the centre of the grid. In the *Z*-direction, there were 34 elements below the surface and 11 elements in the air. The mesh extends to 100 km above the surface and 140 km below the surface. There are 10 × 10 receivers evenly distributed in *XY* plane, separated by 10 km. The forward response (impedance *Z* and tipper *K*) was generated for 31 frequencies evenly distributed in log space between 0.01 and 1000 Hz, which gives six frequencies per decade. We added Gaussian error with standard deviation (24) to the forward data.

The inversion mesh consists of 41x 41y 41z elements. In the *XY* plane, the central 24-by-24 elements are square with sides equal to 0.5 km. Further from the centre, the element sizes grow gradually and extend 135 km from the centre of the grid. In the *Z*-direction, there were 31 elements below the surface and 10 elements in the air. The mesh extends to 110 km above the surface and 140 km below the surface. Thus the forward and inversion meshes differ in discretization but have the same locations for the receivers, which are at the centre of elements faces in both cases. The inversion mesh is presented in Fig. 7.

For this model, we conducted inversions using different regularization functionals. We used (A5), (A2) and (A7) that give regularization functionals resembling $\|\nabla m\|^2_{L_2(\|S\|_2)}$, $\|\nabla m\|^2_{L_2(1)}$, and $\|\nabla m\|^2_{L_2(1/\|S\|_2)}$, respectively. The value of $\|S\|_2$ was confined to change within a factor of 10^4. More precisely, two values were found S_1 and S_2, such that $\frac{S_2}{S_1} = 10^4$ and the value of $\|S\|_2$ was truncated if it lies outside the interval $[S_1, S_2]$. Additionally, weights ν have been multiplied by a normalization constant, so that the average ν over the central part of the domain is the same in all cases. This allows us to use the same initial value of λ. The nRMS, λ and the regularization norm as a function of the iteration number

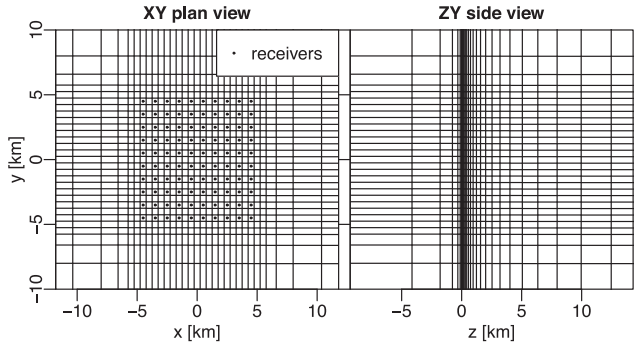

Figure 7. Cross-sections of the inversion grid for the two bricks model. Central part of the grid is shown.

are presented in Fig. 8. One can see that the nRMS values as a function of iteration number are almost the same for the different regularization schemes, and thus the amount of regularization is similar for all weightings.

The models calculated by the different inversion schemes at iteration 6 are presented in Fig. 9, with weights ν of $L_2(\nu)$ norms used for regularization to obtain those models plotted in Fig. 10. Generally speaking, the weight $\nu = \|S\|_2$ decreases with depth and the weight $\nu = \frac{1}{\|S\|_2}$ increases with depth. Thus, the effect of using $L_2(\|S\|_2)$ is to prolong the depth extent of the formed image to minimize the value of the regularization norm. For $L_2\left(\frac{1}{\|S\|_2}\right)$, the recovered objects tend to be compressed toward the surface for comparable reasons. In the case of $L_2\left(\frac{1}{\|S\|_2}\right)$, significant resistivity oscillations are apparent at shallow depths; one of their effects is to drive the background resistivity toward 25 Ωm rather than the true 40 Ωm because of the cell-scale heterogeneity formed under the receivers. Nevertheless, the nRMS values are all very close, underscoring the non-uniqueness inherent in this ill-posed inversion problem. We observed no systematic difference in the fit of the final models across the frequency range for the three regularizations. Results for other models might differ, however. Further challenges in establishing appropriate regularization may be expected for more complex settings.

4.3 DSM2 model

The Dublin MT Modeling and Inversion workshops have provided model results for the EM community to test newly developed simulation and imaging codes (see Miensopust *et al.* 2013). Here we consider inversion of the MT responses of the Dublin secret model 2 (DSM2) presented in Fig. 11. It is a flat-earth model with two contacting, shallow bricks in a four-layer earth. There are 144 MT receivers arranged in a uniform grid 12 × 12 with 7 km spacing.

The forward data, supplied by the workshop organizers, consist of the impedance tensor Z values only (no tipper) for 30 frequencies between 0.016 and 10 000 s evenly distributed in $\log_{10}$. Random galvanic distortion was applied to the responses by the organizers as described in Miensopust *et al.* (2013). Gaussian noise of 5% of the maximal impedance value also had been added to the distorted data set. This supplied error bound was treated as a standard deviation and was used for both real and imaginary parts of Z. The data from all sites and frequencies were used in our inversion.

The applied static distortion provides an opportunity for us to implement and test recent distortion removal procedures (Avdeeva *et al.* 2015), summarized in Appendix B. Initially, an inversion model is sought without distortion correction. This model is used as an initial guess to estimate a new, more stable model plus the static distortion matrices of the impedance Z. We invert the data using the $L_2(1)$ regularization functional.

We considered coarse and fine inversion meshes. The coarse mesh has two columns of parameters between sites in the central portion of the model whereas the fine mesh has five columns of parameters between sites. The purpose of the latter mesh is to test whether a fine discretization allows formation of small-scale shallow structure which can simulate the impedance galvanic distortion without having to solve explicitly for correction factors (*cf.* e.g. Meqbel *et al.* 2014).

Specifically, the coarse(fine) mesh consisted of 45x 45y 41z(78x 78y 50z) elements. In the XY plane, the central 23-by-23(58-by-58) elements are squares with sides = 3.5 km (1.4 km). Further from the centre, the element sizes grow gradually and extend 600 km from the centre of the grid. In the Z-direction, there were 31(38) elements below the surface and 10(12) elements in the air. The mesh extends to 300 km above the surface and 700 km below the surface. The central part of the coarse mesh is presented in Fig. 12.

The inverted models are presented in Fig. 13. Inverting only for $\log_{10}$ resistivity on the coarse mesh with no distortion correction yields a model with nRMS of 4.2, with little further improvement by relaxing the regularization factor (see Fig. 14). Subsequently, inverting also for the distortion matrices obtains a model with nRMS of 1.1. The latter model achieves generally smoother resistivity structure with values closer to the true values, especially in the deeper structure, than does the former model. For the coarse model,

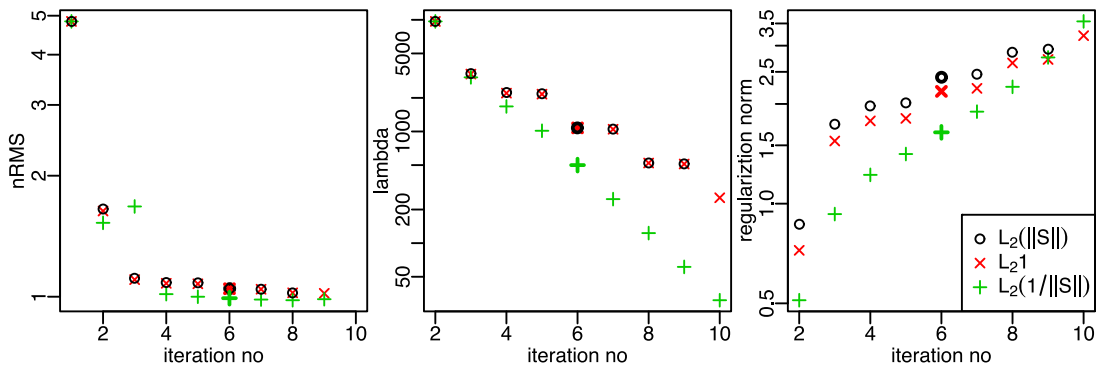

Figure 8. Two bricks model iteration history. nRMS(m_j), λ used to obtain model m_j as well as $\|m_j - m_0\|_{B_m}$ as a function of iteration number j are plotted. The model number 6, plotted on Fig. 9 is denoted by a bold symbol.

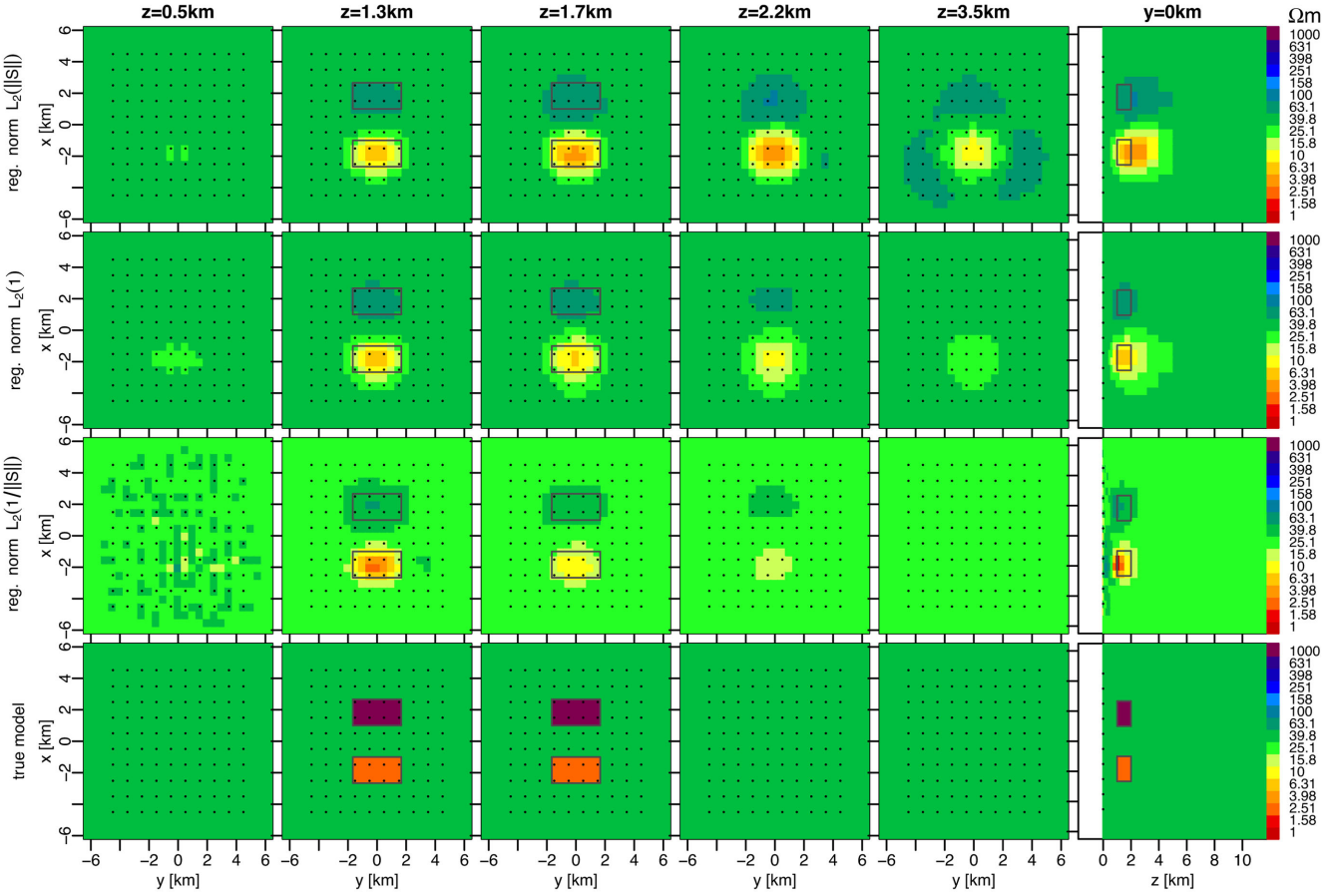

Figure 9. Models calculated by inversions of synthetic responses of two bricks using regularization functionals $\|\nabla m\|^2_{L_2(\|S\|_2)}$, $\|\nabla m\|^2_{L_2(1)}$ and $\|\nabla m\|^2_{L_2\left(\frac{1}{\|S\|_2}\right)}$. In each case, the model obtained at iteration 6 is plotted.

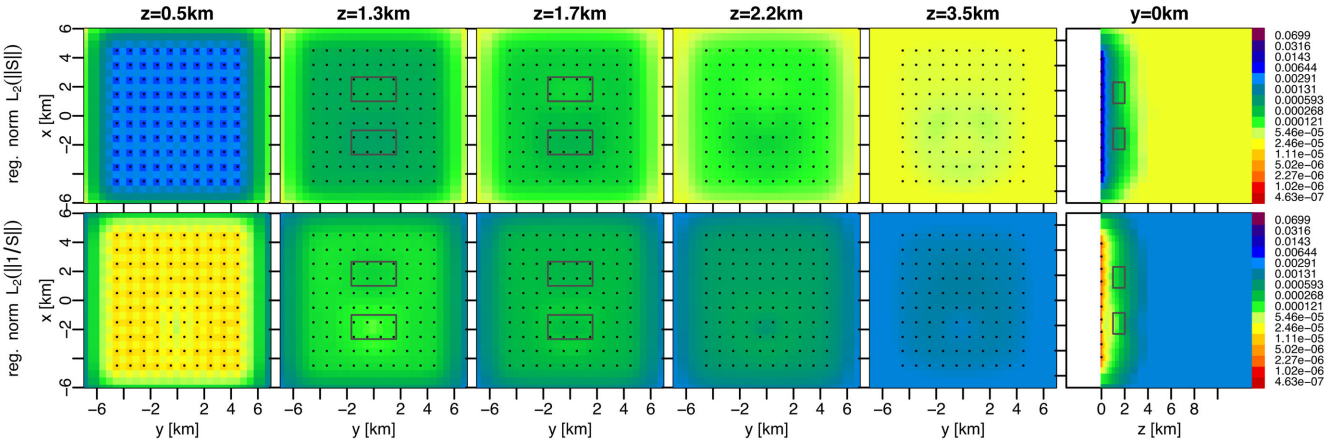

Figure 10. Weight ν of $L_2(\nu)$ norm used to obtain model number 6 for regularization schemes used for two bricks model. For $L_2(1)$ regularization, the weight is constant so is not plotted.

there is some scatter in the norm of distortion matrices versus iteration. This presumably is a result of small regularization ($\tau = 0.01$). Further investigation is warranted as to when and to what degree of regularization, distortion should be estimated through the iteration history.

When the fine mesh inversion for $\log_{10}$ resistivity only is considered, the resulting model has nRMS of 2.2, significantly less than the similar model obtained on a coarse mesh. The fine mesh inversion is able to represent some of the distortion by small-scale variability of $\log_{10}$ resistivity in the vicinity of the receivers, at shallow depths. Nevertheless, the fine mesh inversion for $\log_{10}$ resistivity including the distortion estimation provides a smoother model with a smaller nRMS of 1.1 (see Fig. 14). Here we see smoother behaviour in the estimated distortion versus iteration. From this result we suggest

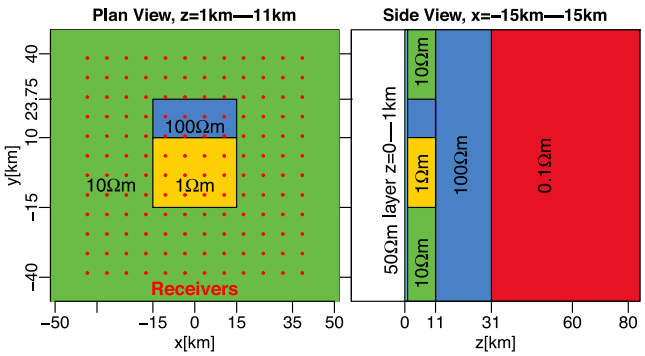

Figure 11. Sketch of DSM2 model.

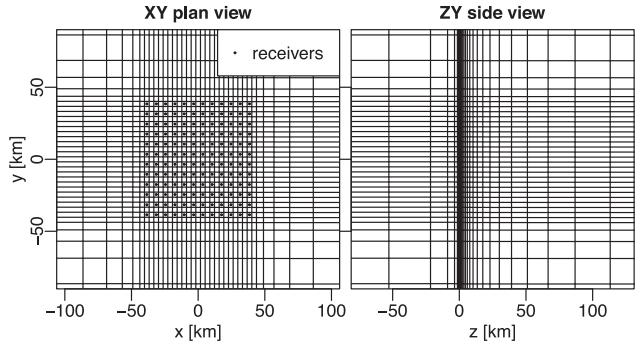

Figure 12. Cross-sections of the central part of the coarse inversion grid for DSM2 model.

that distortion matrices should be considered in tensor impedance inversion even for fine discretizations. However, we also advocate that fine discretization be used to the extent practical to ensure that non-galvanic variations at the highest frequencies are accommodated by the smallest scale mesh structure.

5 FIELD INVERSION EXAMPLES

5.1 Mount St Helens

Finally, we examine the MT field data set collected by Hill *et al.* (2009) from the north-central Cascade volcanic environment in Washington State, USA, to demonstrate the ability of our solution to handle moderately large models with topography. There are 82 soundings primarily clustered over the recently active Mount St Helens volcano, but with 14 of the sites extending in a nearly E-W profile past the north side of Mount Adams (Fig. 15). This gives us the opportunity also to compare 3-D inversion of profile data (e.g. Siripunvaraporn *et al.* 2005b) with 2-D inversion results. We invert the complex tensor impedance Z and tipper K for 20 frequencies log-uniformly distributed from 100 through 0.0018 Hz.

The mesh consists of $111x$ $167y$ $62z$ elements in total (see Fig. 16). This requires storage of 500 GB which fills the capacity of our particular workstation. Over the large central section including the two volcanoes, horizontal dimensions of the elements were in the 500 × 600 to 500 ×1000 m range typically. Around this region the element sizes grew gradually, covering a total area of 375 km × 425 km. In the z-direction, there are 50 elements below the ground and 12 elements in the air. The elements at the earth's surface have a thickness of 80 m (at mesh edge) and grow gradually to reach an elevation of 250 km above the surface and a depth of 220 km below the ground. Topography for the area was obtained from ASTER GDEM data (downloaded from http://gdex.cr.usgs.gov/gdex/), a product of METI and NASA. We did not attempt to include the Pacific Ocean nearly 200 km to the west, as that distance is significantly larger than the depth range of interest here (<100 km). A rim of one element around the side edges and bottom of the mesh was excluded from the inversion and fixed to be a 1-D (flat) initial model. Thus, there are 109 × 165 × 49 = 881 265 inversion parameters in the Mount St Helens model. However, in data-space formulation, the rank of the model update matrix is only 19 680. This may seem like a large model to handle 82 MT sites, but that is a result of particular site distribution. In principle, many more MT sites could be placed in this model mesh with additional computational cost only being more Jacobian source reductions and a larger (though still modest) data-space model update matrix.

The inversion was run without distortion matrix estimation for 11 iterations, with iteration history shown in Fig. 17. Data error floors as given in eq. (23) were adopted. The starting model was a 100 Ωm half-space, the same as considered by Hill *et al.* (2009), and the starting nRMS was ∼11.5. Run-time on the 24-core workstation was ∼13 hr per iteration, which was by far dominated by the forward and Jacobian calculations over the 20 frequencies. Model 11 has nRMS of 1.2, which is considered a good fit so that distortion correction should yield little improvement and was not carried out.

Model cross-section and plan views are presented in Figs 18 and 19, and can be compared to the original results of Hill *et al.* (2009). The cross-section overall bears a close resemblance to the 2-D inversion of Hill et al, which emphasized the nominal TM mode (relative to profile orientation) of data. Steep low resistivity is seen in the middle crust directly under Mount St Helens, presumably related to recent eruptive processes, of which more will be discussed shortly. This gives way at depths >20 km to broad, quasi-horizontal low resistivity between the two volcanoes, which we attribute to lower crustal magmatic underplating and high-temperature fluid release. Shallow, very low resistivity overlies the deep crustal conductor approaching Mount Adams which may reflect in part the presence of graphitic metasediments associated with a suture between the Siletz terrane and former North American margin (the southern Washington Cascades conductor or SWCC of Stanley *et al.* 1996), although this interpretation is non-unique and not without controversy (Egbert & Booker 1993; Hill *et al.* 2009). A large resistive body extending to >15 km depth lies between the Mount St Helens and Mount Adams and could be correlated with earlier Western Cascades intrusive rocks (see Wannamaker *et al.* 2014).

The steep low resistivity directly under Mount St Helens in Fig. 18 is similar to that in the flat-earth 3-D inversion model of Hill *et al.* (2009) although the most anomalous portion does not extend to quite as shallow a depth as that in Hill's. This may in part be explained by the conical edifice of the volcano inducing additional depression of the electric field as discussed with Fig. 4. A second, somewhat lesser conductor in the 4–9 km depth range appears just west of the first one, which is more subtly expressed in the model of Hill *et al.* These two conductors correspond reasonably well to the two high-scattering bodies displayed in east–west section by DeSiena *et al.* (2014, their fig. 6e). In plan view at 7 km depth (Fig. 19), we see that this steep conductor is strongly linear in a nearly N-S direction and is associated with the Mount St Helens shear zone (MSZ) passing through the volcanic edifice (Weaver *et al.* 1987; Lagmay *et al.* 2000). Clear representation of this structure in our model we believe may be due to inclusion of the tipper elements in the inversion, as the tipper shows a subtle reversal on the west flank of the volcano (Hill *et al.* 2009, also see our Supplemental Material

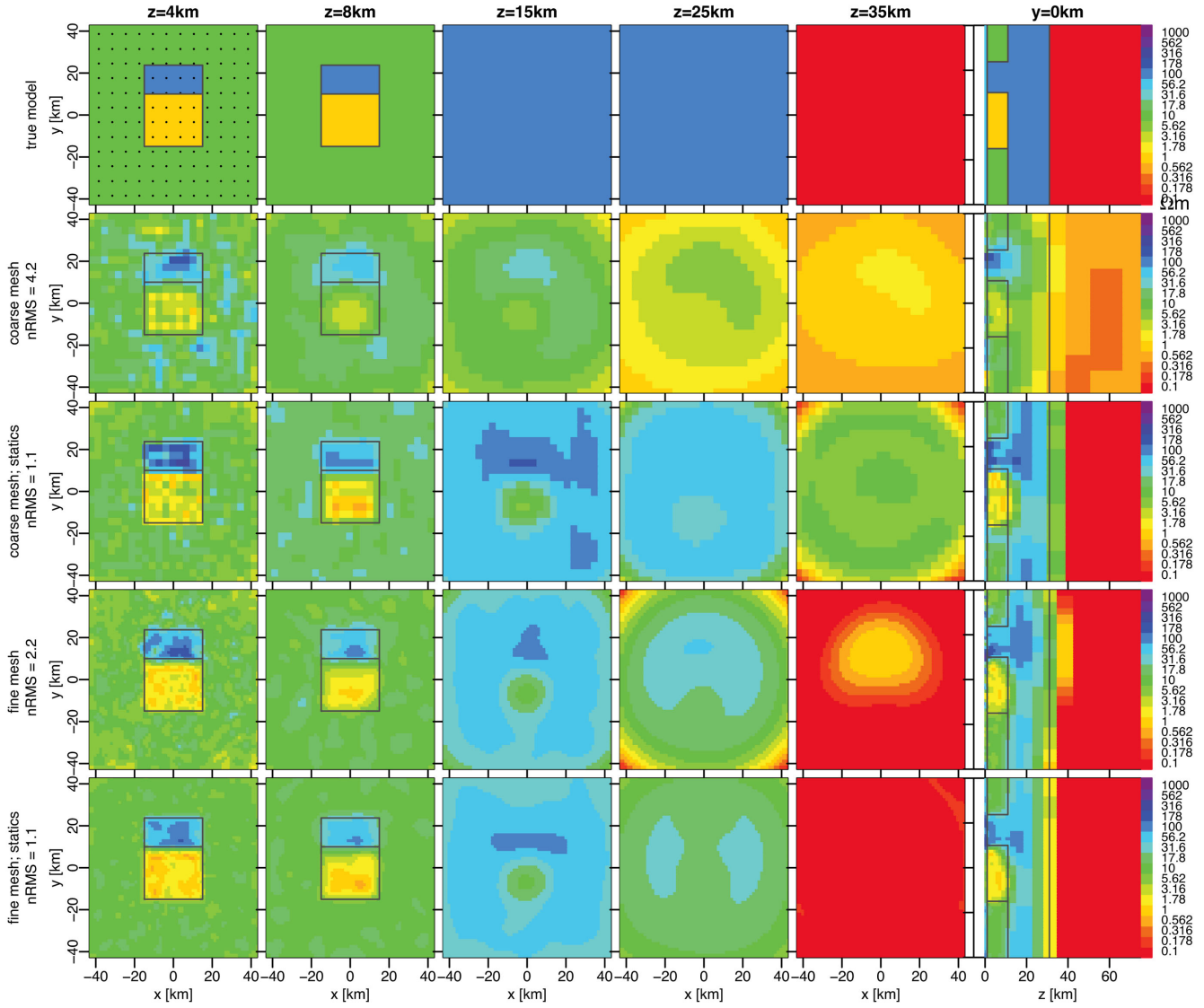

Figure 13. Comparison of inversion models obtained for DSM2 model. The second and the third rows contain the result of inversions on a coarse mesh. The fourth and the fifth rows show results of inversions on a fine mesh. The third and the fifth rows, denoted by 'statics', show results of inversions for $\log_{10}$ resistivity and the static distortion matrices. The top row shows the true model.

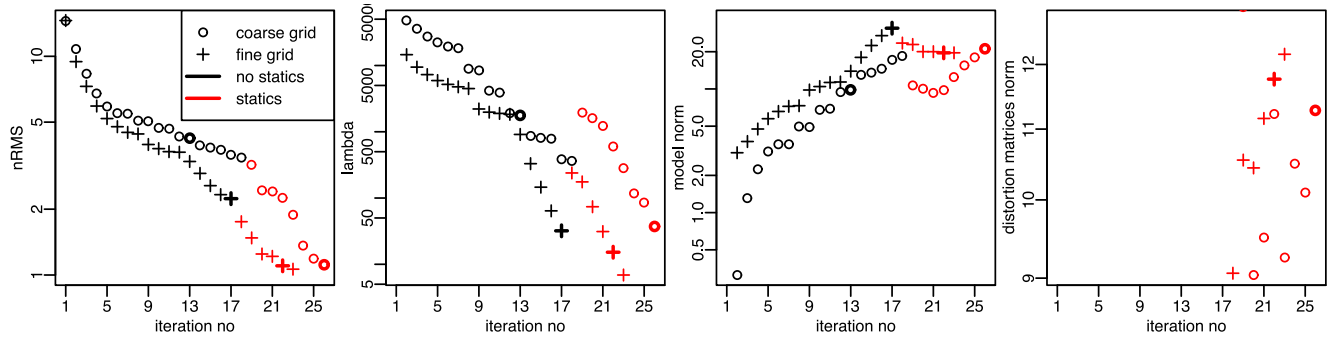

Figure 14. Iteration history for DSM2 model, for coarse and fine meshes. Initial inversion without static distortion is shown in black. Subsequent inversion with distortion matrix estimation is plotted in red. Bold symbols denote models shown in Fig. 13.

section). The second, subsidiary conductor flanks the shear zone nearby to the west.

The large resistor east of Mount St Helens confines the large conductor further east to be in the Mount Adams area, providing better resolution than prior 3-D images based just on regional tipper data (Egbert & Booker 1993). The NNW-SSE limits of the resistor cannot be considered as well-resolved, however, without site coverage. At lower crustal depths (27 km in Fig. 19), resistivity under

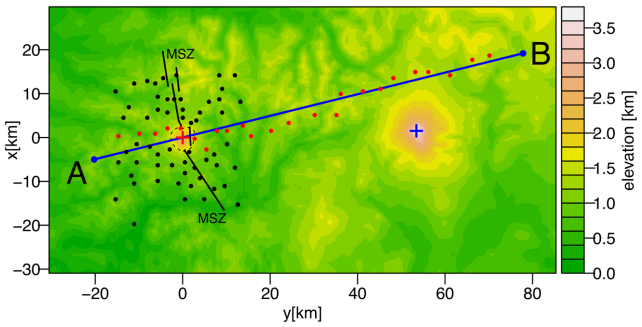

Figure 15. Mount St Helens inversion model. Contour plot of topography of the central part of the domain is presented. Coordinate (0, 0) corresponds to the location of Mount St Helens peak, marked by a red cross. Blue cross denotes Mount Adams. Blue line denotes profile A–B used in Fig. 18. MT receiver locations are marked by black and red dots. Red dots denote receivers used in the 2-D inversion of Hill *et al.* (2009). Mount St Helens shear zone (MSZ) after Lagmay *et al.* (2000).

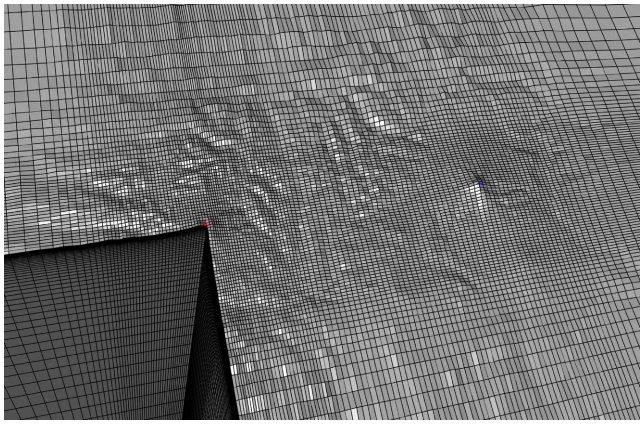

Figure 16. Central part of the mesh for the Mount St Helens inversion model. Blue and red crosses denote Mount Adams and Mount St Helens peaks, respectively.

Mount St Helens decreases from west to east as in Hill *et al.* (2009) and the low is somewhat elongate toward the south–southeast. On the other hand, low deep crustal resistivity under Mount Adams expands to the north. It is tempting to assign this geometry to an offset in lower crustal magmatic underplating associated with the E-W offset in the Cascade volcanic chain at this latitude. However, such conjecture should await better resistivity structural constraints from further 3-D MT coverage both north and south of the current data set.

Finally, in Fig. 20, we show a 3 Ωm isosurface within the Mount St Helens model viewed from the east with conductive material around Mount Adams excluded. It represents the conductive upper reaches of the MSZ. The top of the main surface is only slightly undulatory along its north–south extent. A few steep, narrow conductors projecting upward mainly to the north of the mount itself could represent local fluidized damage zones but do not obviously correspond to a central volcanic conduit (e.g. Musumeci *et al.* 2002).

6 CONCLUSIONS

As other researchers are finding as well, direct solutions to various aspects of the diffusive EM inversion problem are becoming

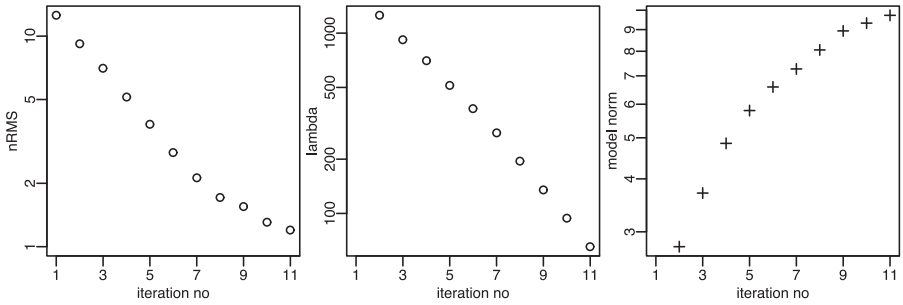

Figure 17. Values of nRMS, λ and model norm as a function of iteration number for the Mount St Helens inversion.

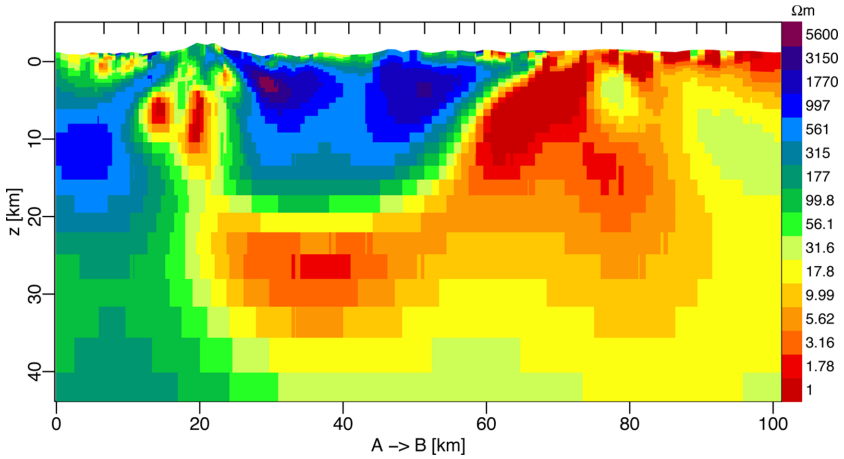

Figure 18. Cross-section of Mount St Helens inversion at iteration 11 along profile A–B marked on Fig. 15. Black ticks at the top denote the locations of receivers plotted in red on Fig. 15 that were used in the 2-D inversion of Hill *et al.* (2009).

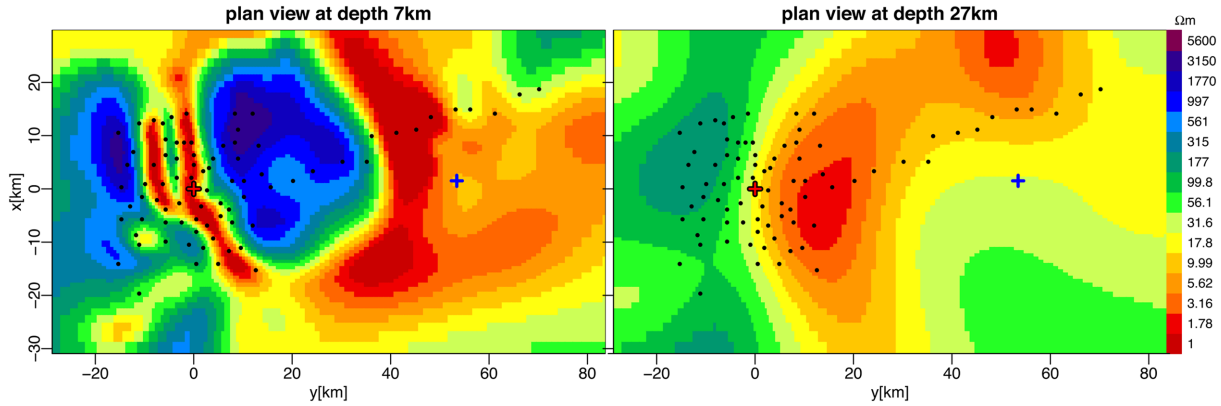

Figure 19. Plan views of Mount St Helens inversion model 11. Receivers' locations marked by black dots, Mount St Helens peak by red cross and Mount Adams peak by blue cross.

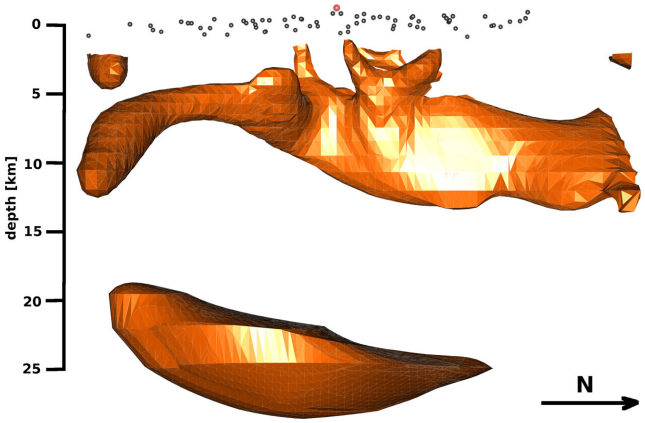

Figure 20. 3-D view from the East onto the 3 Ωm isosurface below Mount St Helens. A region (−25 km, 25 km) × (−24 km, 19 km) around St Helens only is shown, so the conductors below Mount Adams are not seen. Only structures deeper than 2 km are shown. Red circle denotes Mount St Helens peak, black circles denote locations of the receivers around the mountain.

increasingly practical. Here we have shown that direct solvers can effectively handle the Gauss–Newton update for inverse problems approaching one million parameters with parallelization on multicore SMP workstations and large RAM if the model update is formulated in data space. In this case, the limiting computational cost both in run-time and memory is the forward problem (including Jacobians), which can be computed effectively using MKL PARDISO. Finite-element models of order 176 × 176 × 70 elements fill half the memory of a workstation with 0.5 TB RAM but such meshes can, for example, fit large MT data sets of 400 sites with six columns of parameters between sites in both *x*- and *y*-directions with padding of nearly 30 expanding element columns around the mesh edges. We have not experienced system conditioning problems due to high element aspect ratios with our direct solutions.

Single-box SMP workstation capabilities continue to progress, with platforms holding up to 4 TB RAM available at the time of this writing. The overall scalability of MKL PARDISO is good on multicore shared memory systems and may be expected to do well on machines with more cores. Another option could be to construct a distributed cluster whose nodes were large-RAM multicore machines such as we employed herein each devoted to a different response frequency, although at considerably greater hardware investment. Finally, we find that the deformable hexahedral mesh framework lends a predictability to mesh design and performance of libraries such as MKL PARDISO that should offset concerns that the geometries of simulation with such a mesh may not be as arbitrary as is possible with assemblies of tetrahedra. More research is warranted, however, into optimal inversion regularization functionals and methods for guaranteeing rapid convergence.

ACKNOWLEDGEMENTS

We acknowledge the support of this work from U.S. Department of Energy under contract DE-EE0002750 and the U.S. National Science Foundation under grant AES-1443522 to PW. EC acknowledges the partial support of the U.S. National Science Foundation through grants ARC-0934721 and DMS-1413454. GH acknowledges the support of both an MIPRS and MGS award through Monash University and the support of the Royal Society of New Zealand Marsden Fund award ASL-1301. Presentation and content of the paper were improved substantially following the remarks of three anonymous reviewers. The submission was handled by Associate Editor Ute Weckmann.

REFERENCES

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

Baboulin, M., Giraud, L. & Gratton, S., 2005. A parallel distributed solver for large dense symmetric systems: applications to geodesy and electromagnetism problems, *Int. J. High Perform. Comput. Appl.,* **19**(4), 353–363.

Baboulin, M., Becker, D. & Dongarra, J., 2011. A parallel tiled solver for dense symmetric indefinite systems on multicore architectures, *Technical Rep.*, University of Tennessee Computer Science, ICL-UT-11-07.

Buttari, A., Langou, J., Kurzak, J. & Dongarra, J., 2009. A class of parallel tiled linear algebra algorithms for multicore architectures, *Parallel Computing*, **35**, 38–53.

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

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.

deGroot-Hedlin, C. & Constable, S., 1990. Occam's inversion to generate smooth, two-dimensional models from magnetotelluric data, *Geophysics,* **55**(12), 1613–1624.

DeSiena, L., Thomas, C., Waite, G., Moran, S. & Klemme, S., 2014. Attenuation and scattering tomography of the deep plumbing system of mount st. helens, *J. geophys. Res.: Solid Earth,* **119**(11), 8223–8238.

Egbert, P.G.D. & Booker, J.R., 1993. Imaging crustal structure in southwestern Washington with small magnetometer arrays, *J. geophys. Res.,* **98,** 15 967–15 985.

Grayver, A.V., Streich, R. & Ritter, O., 2013. Three-dimensional parallel distributed inversion of CSEM data using a direct forward solver, *Geophys. J. Int.,* **193,** 1432–1446.

Groom, R.W. & Bahr, K., 1992. Corrections for near surface effects: decomposition of the magnetotelluric impedance tensor and scaling corrections for regional resistivities: a tutorial, *Surv. Geophys.,* **13**(4–5), 341–379.

Haber, E., Oldenburg, D.W. & Shekhtman, R., 2007. Inversion of time domain three-dimensional electromagnetic data, *Geophys. J. Int.,* **171**(2), 550–564.

Hill, G.J., Caldwell, T.G., Heise, W., Chertkoff, D.G., Bibby, H.M., Burgess, M.K., Cull, J.P. & Cass, R.A.F., 2009. Distribution of melt beneath Mount St Helens and Mount Adams inferred from magnetotelluric data, *Nat. Geosci.,* **2,** 785–789.

Hohmann, G.W. & Raiche, A.P., 1988. Inversion of controlled-source electromagnetic data, in *Electromagnetic Methods in Applied Geophysics*, Vol. 1, pp. 469–503, ed. Nabighian, M.N., Soc. Expl. Geophys., Tulsa, OK.

Key, K. & Constable, S., 2011. Coast effect distortion of marine magnetotelluric data: insights from a pilot study offshore northeastern Japan, *Phys. Earth planet. Inter.,* **184,** 194–207.

Kordy, M., Maris, V., Wannamaker, P. & Cherkaev, E., 2013. 3-D edge finite element solution for scattered electric field using a direct solver parallelized on an SMP workstation, in *5th International Symposium on Three-Dimensional Electromagnetics,* Sapporo, May 7–9, p. 4.

Kordy, M., Wannamaker, P., Maris, V., Cherkaev, E. & Hill, G.J., 2015. 3-dimensional magnetotelluric inversion including topography using deformed hexahedral edge finite elements and direct solvers parallelized on SMP computers, Part I: forward problem and parameter Jacobians, *Geophys. J. Int.,* doi:10.1093/gji/ggv410.

Lagmay, A. M.F., van Wyk de Vries, B., Kerle, N. & Pyle, D.M., 2000. Volcano instability induced by strike-slip faulting, *Bull. Volc.,* **62,** 331–346.

Levenberg, K., 1944. A method for the solution of certain non-linear problems in least squares, *Q. appl. Math.,* **2,** 164–168.

Liu, J., Brio, M. & Moloney, J., 2009. Overlapping Yee FDTD method on nonorthogonal grids, *J. Sci. Comput.,* **39**(1), 129–143.

Maris, V. & Wannamaker, P.E., 2010. Parallelizing a 3-D finite difference MT inversion algorithm on a multicore PC using OpenMP, *Comput. Geosci.,* **36**(10), 1384–1387.

Meqbel, N.M., Egbert, G.D., Wannamaker, P.E., Kelbert, A. & Schultz, A., 2014. Deep electrical resistivity structure of the northwestern U.S. derived from 3-D inversion of USArray magnetotelluric data, *Earth planet. Sci. Lett.,* **402,** 290–304.

Miensopust, M.P., Queralt, P., Jones, A.G. & the 3-D Modelers, 2013. Magnetotelluric 3-D inversion—a review of two successful workshops on forward and inversion code testing and comparison, *Geophys. J. Int.,* **193,** 1216–1238.

Musumeci, C., Gresta, S. & Malone, S.D., 2002. Magma system recharge of Mount St. Helens from precise relative hypocenter location of microearthquakes, *J. geophys. Res.: Solid Earth (1978–2012),* **107**(B10), ESE 16-1–ESE 16-9.

Nocedal, J. & Wright Stephen, J., 2006. *Numerical Optimization,* Springer.

Oldenburg, D.W., Haber, E. & Shekhtman, R., 2013. Three dimensional inversion of multi-source time domain electromagnetic data, *Geophysics,* **78**(1), E47–E57.

Parker, R.L., 1994. *Geophysical Inverse Theory, Princeton University Press*, Princeton, 386 pp.

Sasaki, Y., 2001. Full 3-D inversion of electromagnetic data on PC, *J. appl. Geophys.,* **46,** 45–54.

Schwarzbach, C. & Haber, E., 2013. Finite element based inversion for time-harmonic electromagnetic problems, *Geophys. J. Int.,* **193,** 615–634.

Siripunvaraporn, W. & Egbert, G., 2000. An efficient data-subspace inversion method for 2-D magnetotelluric data, *Geophysics,* **65,** 791–803.

Siripunvaraporn, W., Egbert, G., Lenbury, Y. & Uyeshima, M., 2005a. Three-dimensional magnetotelluric inversion: data-space method, *Phys. Earth planet. Inter.,* **150,** 3–14.

Siripunvaraporn, W., Egbert, G. & Uyeshima, M., 2005b. Interpretation of two-dimensional magnetotelluric profile data with three-dimensional inversion: synthetic examples, *Geophys. J. Int.,* **160**(3), 804–814.

Stanley, W.D., Johnson, S.Y., Qamar, A.Y., Weaver, C.S. & Williams, J.M., 1996. Tectonics and seismicity of the southwest Washington Cascade Range, *Bull. seism. Soc. Am.,* **86,** 1–18.

Stark, M.A., Soyer, W., Hallinan, S. & Watts, M.D., 2013. Distortion effects on magnetotelluric sounding data investigated by 3d modeling of high-resolution topography, *Geotherm. Res. Counc. Trans.,* **37,** 521–527.

Usui, Y., 2015. 3-D inversion of magnetotelluric data using unstructured tetrahedral elements: applicability to data affected by topography, *Geophys. J. Int.,* **202**(2), 828–849.

Wannamaker, P., Stodt, J. & Rijo, L., 1986. Two-dimensional topographic responses in magnetotelluric modeling using finite elements, *Geophysics,* **51,** 2131–2144.

Wannamaker, P.E., Evans, R.L., Bedrosian, P.A., Unsworth, M.J., Maris, V. & McGary, R.S., 2014. Segmentation of plate coupling, fate of subduction fluids, and modes of arc magmatism in Cascadia, inferred from magnetotelluric resistivity, *Geochem. Geophys. Geosyst.,* **15**(11), 4230–4253.

Weaver, C.S., Grant, W.C. & Shemeta, J.E., 1987. Local crustal extension at Mount St. Helens, Washington, *J. geophys. Res.: Solid Earth (1978–2012),* **92**(B10), 10 170–10 178.

Yi, M.J., Kim, J.H. & Chung, S.H., 2003. Enhancing the resolving power of least-squares inversion with active constraint balancing, *Geophysics,* **68**(3), 931–941.

Zhdanov, M.S., 2002. *Geophysical Inverse Theory and Regularization Problems,* Elsevier.

Zhdanov, M.S., Wan, L., Gribenko, A., Cuma, M., Key, K. & Constable, S., 2011. Large-scale 3-D inversion of marine magnetotelluric data: case study from the Gemini prospect, Gulf of Mexico, *Geophysics,* **76,** F77–F87.

SUPPORTING INFORMATION

Additional Supporting Information may be found in the online version of this paper:

Figure 1. Real and imaginary induction vectors for the measured data and the prediction by the model obtained in the Mount St Helens inversion for five frequencies between 1 Hz and 0.01 Hz. Parkinson convention is used.

Figure 2. nRMS for each component of the MT response and for each receiver for the final model of Mount St Helens inversion.

Table 1. Table of nRMS as a function of frequency for the starting model and the final model for Mount St Helens inversion. (http://gji.oxfordjournals.org/lookup/suppl/doi:10.1093/gji/ggv411/-/DC1).

Please note: Oxford University Press is not responsible for the content or functionality of any supporting materials supplied by the authors. Any queries (other than missing material) should be directed to the corresponding author for the paper.

APPENDIX A: APPROXIMATION OF REGULARIZATION NORMS

Here, we present how we approximate norms $\|\nabla(m-m_0)\|_{L_2(1)}$,$\|\nabla(m-m_0)\|_{L_2(\|S\|_2)}$,$\|\nabla(m-m_0)\|_{L_2\left(\frac{1}{\|S\|_2}\right)}$. For simplicity we will write m instead of $m-m_0$.

Assume that the discrete m is a representation of an infinite-dimensional model, that is, a function $m(\mathbf{r})$ defined for every location $\mathbf{r}$ in the subsurface part of Ω. If the discrete norms $\|m\|_{L_2(v)}$ and $\|\nabla m\|_{L_2(v)}$ are defined as the norms of the infinite-dimensional model, they have the property that they are mesh independent. If one considers two different meshes with two different discrete

representations of the same infinite-dimensional model $m(\mathbf{r})$, the two norms on different meshes will be equal, as they are equal to the norm of the infinite-dimensional model.

In the case of the norm $\|m\|_{L_2(v)}$ one can think of the infinite-dimensional model $m(\mathbf{r})$ being piecewise constant, that is, $m(\mathbf{r}) = m_j$ whenever $\mathbf{r}$ is inside an inversion cell V_j. In such a case, if a mesh is refined and the value of the model at an inversion cell is passed on to the subcells that the cell is split to, then the infinite-dimensional model remains the same. In the case of the norm $\|\nabla(m)\|_{L_2(v)}$ one cannot think of the infinite-dimensional model being piecewise constant as the gradient of such a model does not exist (the gradient is not a square integrable vector field). The infinite-dimensional model should be continuous. In 1-D one could think of a model that is piecewise linear between the centres of inversion cells.

First we will consider norms $\|m\|_{L_2(1)}$, $\|m\|_{L_2(\|S\|_2)}$, $\|m\|_{L_2\left(\frac{1}{\|S\|_2}\right)}$. To approximate them, we will take norm of the form:

$$\|m\|^2_{B_m} = m^T B_m m \tag{A1}$$

with appropriate matrix B_m, where $m = (m)_{j=1}^{N_m}$ is a vector of $\log_{10}$ resistivities of inversion cells.

If one takes B_m to be a diagonal matrix with entries w_j equal to volumes of inversion cells:

$$w_j = \#V_j = \int_{V_j} d\mathbf{r} \tag{A2}$$

then one obtains a model norm $\|m\|_{B_m}$ that is equal to the $L_2(1)$ norm of the piecewise constant model $m(\mathbf{r})$:

$$\begin{aligned}\|m\|^2_{B_m} &= \sum_{j=1}^{N_m} m_j^2 B_m(j,j) = \sum_{j=1}^{N_m} m_j^2 \#V_j \\ &= \sum_{j=1}^{N_m} \int_{V_j} m(\mathbf{r})^2 d\mathbf{r} = \int_{\Omega} m(\mathbf{r})^2 d\mathbf{r} \\ &= \|m\|^2_{L_2(1)}\end{aligned} \tag{A3}$$

Consider the derivative S of infinite-dimensional problem defined at (21). Assuming that the discretization of the domain is fine enough so that the finite-dimensional approximation of the problem is close to the infinite-dimensional problem, using F for finite-dimensional response, one could write that the jth column of Jacobian matrix J is:

$$J_{\cdot j} = \frac{\partial F}{\partial m_j} = \int_{V_j} S(\mathbf{r}) d\mathbf{r} \tag{A4}$$

where V_j is a j-th inversion cell. If we assume that the inversion cell V_j is small enough that $S(\mathbf{r}) \approx S_j = \text{const}$ for $\mathbf{r} \in V_j$, then a sensitivity of inversion cell V_j is obtained:

$$\begin{aligned} w_j &= \sqrt{\sum_{i=1}^{N_d} J_{ij}^2} = \sqrt{J_{\cdot j}^T J_{\cdot j}} = \|J_{\cdot j}\|_2 = \left\| \int_{V_j} S(\mathbf{r}) d\mathbf{r} \right\|_2 \\ &\approx \|\#V_j S_j\|_2 = \#V_j \|S_j\|_2 = \int_{V_j} \|S_j\|_2 d\mathbf{r} \\ &\approx \int_{V_j} \|S(\mathbf{r})\|_2 d\mathbf{r} \end{aligned} \tag{A5}$$

If we define B_m to be a diagonal matrix with w_j as entries, then

$$\begin{aligned}\|m\|^2_{B_m} &= \sum_{j=1}^{N_m} m_j^2 w_j \approx \sum_{j=1}^{N_m} m_j^2 \int_{V_j} \|S(\mathbf{r})\|_2 d\mathbf{r} \\ &= \int_{\Omega} \|m(\mathbf{r})\|_2^2 \|S(\mathbf{r})\|_2 d\mathbf{r} \\ &= \|m\|^2_{L_2(\|S\|_2)}.\end{aligned} \tag{A6}$$

The regularization norm is approximately equal to the weighted L_2 norm with $\|S\|_2$ as a weight. Note also that to calculate $w_j = \|J_j\|_2$ one does not need to know the cell volume, only Jacobian matrix J is used. This regularization was considered in Zhdanov (2002) (see eq. 3.89).

The third weight we consider is defined as

$$w_j = \frac{(\#V_j)^2}{\sqrt{J_{\cdot j}^T J_{\cdot j}}} \approx \frac{(\#V_j)^2}{\#V_j \|S_j\|_2} = \#V_j \frac{1}{\|S_j\|_2} \tag{A7}$$

The corresponding norm of the model is approximately

$$\begin{aligned}\|m\|^2_{B_m} &= \sum_{j=1}^{N_m} m_j^2 w_j \approx \sum_{j=1}^{N_m} m_j^2 \#V_j \frac{1}{\|S_j\|_2} \\ &\approx \sum_{j=1}^{N_m} m_j^2 \int_{V_j} \frac{1}{\|S(\mathbf{r})\|_2} d\mathbf{r} \\ &= \int_{\Omega} \|m(\mathbf{r})\|_2^2 \frac{1}{\|S(\mathbf{r})\|_2} d\mathbf{r} \\ &= \|m\|^2_{L_2\left(\frac{1}{\|S\|_2}\right)}\end{aligned} \tag{A8}$$

The norm is approximately equal to the weighted L_2 norm with $\frac{1}{\|S\|_2}$ as a weight. This norm will suppress model change in regions with low sensitivity, using the reasoning that if we cannot detect the properties of a region well, we will make it similar to its surroundings. This is similar to the approach of Yi *et al.* (2003).

To get an approximation of a norm of the model gradient, rather than of the model, so a norm that resembles $\|\nabla m\|_{L_2(v)}$ rather than $\|m\|_{L_2(v)}$ in essence one should think of an infinite-dimensional continuous model $m(\mathbf{r})$ that is represented by the discrete model m and calculate the $L^2(v)$ norm of its gradient. For simplicity though, we do not pursue this approach and we use a simple finite-difference approximation calculated as if the topography was not present and the element layers were horizontal. Assume that each inversion cell consists of one finite element. Air layers as well as one layer of elements close to the boundary $\partial\Omega$ are not used in the inversion. One layer close to the boundary has a fixed conductivity equal to the conductivity of the 1-D layered Earth primary model. This is done to prevent us from having a non-zero source term J^{imp} very close to the boundary. As a result the inversion cells can be addressed using three indices $i_x = 1, \ldots, n_x, i_y = 1, \ldots, n_y, i_z = 1, \ldots, n_z$, where the total number of inversion cells is $N_m = n_x n_y n_z$. Matrix B_m is such that

$$\begin{aligned}\|m\|^2_{B_m} = &\sum_{i_x=2}^{n_x} \sum_{i_y=1}^{n_y} \sum_{i_z=1}^{n_z} \tilde{w}^x_{i_x,i_y,i_z} \left(\frac{m_{i_x,i_y,i_z} - m_{i_x-1,i_y,i_z}}{x_{i_x,i_y,i_z} - x_{i_x-1,i_y,i_z}} \right)^2 \\ &+ \sum_{i_x=1}^{n_x} \sum_{i_y=2}^{n_y} \sum_{i_z=1}^{n_z} \tilde{w}^y_{i_x,i_y,i_z} \left(\frac{m_{i_x,i_y,i_z} - m_{i_x,i_y-1,i_z}}{y_{i_x,i_y,i_z} - y_{i_x,i_y-1,i_z}} \right)^2 \\ &+ \sum_{i_x=1}^{n_x} \sum_{i_y=1}^{n_y} \sum_{i_z=2}^{n_z} \tilde{w}^z_{i_x,i_y,i_z} \left(\frac{m_{i_x,i_y,i_z} - m_{i_x,i_y,i_z-1}}{z_{i_x,i_y,i_z} - z_{i_x,i_y,i_z-1}} \right)^2\end{aligned}$$

where

$$\tilde{w}^x_{i_x,i_y,i_z} = \frac{w_{i_x-1,i_y,i_z} + w_{i_x,i_y,i_z}}{2}$$

$$\tilde{w}^y_{i_x,i_y,i_z} = \frac{w_{i_x,i_y-1,i_z} + w_{i_x,i_y,i_z}}{2}$$

$$\tilde{w}^z_{i_x,i_y,i_z} = \frac{w_{i_x,i_y,i_z-1j} + w_{i_x,i_y,i_z}}{2} \quad \text{(A9)}$$

and

$$\left[x_{i_x,i_y,i_z}, y_{i_x,i_y,i_z}, z_{i_x,i_y,i_z}\right]$$

is the location of the centre of mass of the inversion cell denoted by i_x, i_y, i_z.

Using the procedure described above with w given by (A5), (A2) or (A7) one gets norms of model m resembling $\|\nabla m\|_{L_2(\|S\|_2)}$, $\|\nabla m\|_{L_2(1)}$ and $\|\nabla m\|_{L_2\left(\frac{1}{\|S\|_2}\right)}$, respectively. Those norms are used for regularization and the inversion results are compared.

APPENDIX B: INVERSION FOR STATIC DISTORTION MATRICES

We present the inversion for the impedance static distortion matrices similar to the approach of Avdeeva *et al.* (2015). Shallow conductivity structure causes a static distortion of the impedance such that

$$Z^{\text{obs}}_k(\omega) = C_k Z_k(\omega) \quad \text{(B1)}$$

where Z_k is the impedance without the shallow conductivity structure and Z^{obs}_k is the impedance with the shallow conductivity structure. Matrix $C_k \in \mathbb{R}^{2\times 2}$ is real valued and not dependent on frequency, yet different for each receiver $k = 1, \ldots, N_{\text{rec}}$ (see Avdeeva *et al.* 2015).

In the inversion procedure, apart from calculating the unknown model $m = (m_j)_{j=1}^{N_m}$ of $\log_{10}$ resistivities, we invert also for real valued matrices $C = (C_k)_{k=1}^{N_{\text{rec}}}$, one for each receiver location.

The forward problem response $F(m)$, defined by (5), is modified by applying (B1) to obtain $F(m, C)$. The regularized functional to be minimized changes from (8) by adding squares of Frobenius norms $\|.\|_F$ of the difference between distortion matrices C_k and identity matrix I, yielding:

$$\tilde{W}(m, C) = (F(m, C) - d)^T B_d (F(m, C) - d) + \lambda(m - m_0)^T B_m (m - m_0) + \tau \sum_{k=1}^{N_{\text{rec}}} \|C_k - I\|_F^2 \quad \text{(B2)}$$

Note that if we define $\tilde{N}_m = N_m + 4N_{\text{rec}}$,

$$\begin{aligned}
\tilde{m} = (\tilde{m}_k)_{k=1}^{\tilde{N}_m} = \; & (m_1, \ldots, m_{N_m}, \\
& C_{1,xx}, C_{1,yx}, C_{1,yx}, C_{1,yy}, \ldots, \\
& C_{N_{\text{rec}},xx}, C_{N_{\text{rec}},yx}, C_{N_{\text{rec}},yx}, C_{N_{\text{rec}},yy}) \\
\tilde{m}_0 = (\tilde{m}_{k,0})_{k=1}^{\tilde{N}_m} = \; & (m_{1,0}, \ldots, m_{N_m,0}, \\
& 1, 0, 0, 1, \ldots, \\
& 1, 0, 0, 1)
\end{aligned}$$

$$\tilde{B}_m = \begin{bmatrix} B_m & 0 \\ 0 & \frac{\tau}{\lambda} I \end{bmatrix}$$

Then (B2) may be written in the form similar to (8):

$$\tilde{W}(\tilde{m}) = (F(\tilde{m}) - d)^T B_d (F(\tilde{m}) - d) + \lambda(\tilde{m} - \tilde{m}_0)^T \tilde{B}_m (\tilde{m} - \tilde{m}_0)$$

Jacobian $\tilde{J}$ of the forward response $F(\tilde{m})$ may be easily obtained from J using chain rule. As a result one can apply Gauss–Newton and data-space Gauss–Newton procedure similarly to the case of inversion for m only.

Similarly to Avdeeva *et al.* (2015), we use $\tau = 0.01$. Note that this value of τ is very small, giving almost no regularization for distortion matrices term in (B2). Yet it is enough to obtain good models, if only the starting model is not far from the true model. It is our experience so far that using the starting model that was obtained in the inversion without the distortion matrix yields good results. On the other hand, if one starts from a half-space that is far from the true model, the iteration may not converge to a plausible model. In this case, we have seen the matrices C converge to 0.