Geophysical Journal International

Geophys. J. Int. (2015) **202,** 584–603 doi: 10.1093/gji/ggv165
GJI Geomagnetism, rock magnetism and palaeomagnetism

Parallel three-dimensional magnetotelluric inversion using adaptive finite-element method. Part I: theory and synthetic study

Alexander V. Grayver

Institute of Geophysics, ETH Zürich, Sonneggstrasse 5, CH-8092 Zürich, Switzerland. E-mail: agrayver@erdw.ethz.ch

Accepted 2015 April 14. Received 2015 April 7; in original form 2014 December 1

SUMMARY

This paper presents a distributed magnetotelluric inversion scheme based on adaptive finite-element method (FEM). The key novel aspect of the introduced algorithm is the use of automatic mesh refinement techniques for both forward and inverse modelling. These techniques alleviate tedious and subjective procedure of choosing a suitable model parametrization. To avoid overparametrization, meshes for forward and inverse problems were decoupled. For calculation of accurate electromagnetic (EM) responses, automatic mesh refinement algorithm based on a goal-oriented error estimator has been adopted. For further efficiency gain, EM fields for each frequency were calculated using independent meshes in order to account for substantially different spatial behaviour of the fields over a wide range of frequencies. An automatic approach for efficient initial mesh design in inverse problems based on linearized model resolution matrix was developed. To make this algorithm suitable for large-scale problems, it was proposed to use a low-rank approximation of the linearized model resolution matrix. In order to fill a gap between initial and true model complexities and resolve emerging 3-D structures better, an algorithm for adaptive inverse mesh refinement was derived. Within this algorithm, spatial variations of the imaged parameter are calculated and mesh is refined in the neighborhoods of points with the largest variations. A series of numerical tests were performed to demonstrate the utility of the presented algorithms. Adaptive mesh refinement based on the model resolution estimates provides an efficient tool to derive initial meshes which account for arbitrary survey layouts, data types, frequency content and measurement uncertainties. Furthermore, the algorithm is capable to deliver meshes suitable to resolve features on multiple scales while keeping number of unknowns low. However, such meshes exhibit dependency on an initial model guess. Additionally, it is demonstrated that the adaptive mesh refinement can be particularly efficient in resolving complex shapes. The implemented inversion scheme was able to resolve a hemisphere object with sufficient resolution starting from a coarse discretization and refining mesh adaptively in a fully automatic process. The code is able to harness the computational power of modern distributed platforms and is shown to work with models consisting of millions of degrees of freedom. Significant computational savings were achieved by using locally refined decoupled meshes.

Key words: Inverse theory; Electrical properties; Electromagnetic theory; Geomagnetic induction.

1 INTRODUCTION

Magnetotellurics (MT) is an electromagnetic (EM) method of geophysics aimed at studying the electrical conductivity distribution in the interior of the Earth. To date, it is widely used on local, regional and even continental scales for a large variety of engineering and academic purposes (Berdichevskii & Dmitriev 2008; Chave & Jones 2012). The ultimate goal is to reconstruct the 3-D conductivity distribution given a limited number of EM field measurements. This requires solving an inverse problem. Despite the recent progress in solving these problems—thanks to the advances in computational science—it remains very challenging problem from the numerical and computational points of view (Newman 2014). This becomes particularly relevant for 3-D inverse problems which constitute the main focus of this study. Because of large number of unknowns, they are mostly formulated in form of a minimization problem and solved numerically using derivative-based optimization methods, such as conjugate gradient, quasi-Newton or Gauss–Newton (Nocedal & Wright 1999).

Since inverse problems are inherently ill-posed (Tikhonov & Arsenin 1977), their numerical solution remains an expert task and depends not only on quality of data and amount of *a priori* information at hand, but also on one's level of expertise. The latter introduces some degree of subjectivity due to the necessity to choose an initial model, its parametrization and a number of variables, such as regularization parameter, error floors and possibly other weighting coefficients (Tietze & Ritter 2013; Lindsey & Newman 2015), which optimal values are often not available or difficult to determine in practice. While appropriate methods for choosing these variables can be found in literature (Hansen 1998; Griesbaum *et al.* 2008; Egbert 2012), the problem of a model parametrization in geo-electromagnetic inverse problems remains largely unexplored.

In most situations, an appropriate model parametrization is not known *a priori*. Considering its large impact on inversion (Lindsey & Newman 2015), there is a demand for methods which help choose a suitable model parametrization. In finite-dimensional partial differential equation (PDE)-constrained optimization, discretization of the state, adjoint and parameter variables is required (Biegler *et al.* 2007; Bangerth 2008). For the specific case of MT inversion, state and adjoint variables are EM fields governed by Maxwell's equations with the plane wave and specific point dipole excitations, respectively (Pankratov & Kuvshinov 2010). Parameter variable is a quantity being recovered, that is electrical conductivity or its transformation (Kim & Kim 2011). In many inversion codes, the same pre-determined mesh is used to represent all three. This may result in overparametrization of the parameter variable or underparametrization of the state/adjoint variable. On the one hand, state and adjoint meshes have to be sufficiently fine in order to facilitate accurate EM responses. On the other hand, parameter spatial resolution is inherently limited. Therefore, if the same mesh is used to discretize a parameter variable, then more unknowns increase ambiguity. Larger ambiguity can only be handled by adding more data or strengthening regularization. One remedy for this is to decouple state, adjoint and parameter meshes. For instance, Commer & Newman (2008) and (Haber & Schwarzbach 2014) showed that such decoupling is extremely beneficial in handling large-scale EM inverse problems. Similarly, Günther *et al.* (2006) used decoupled tetrahedral meshes to gain efficiency in DC geo-electric inversion. Further, if some *a priori* information about an imaged target is known, *ad hoc* parametrizations can be used to substantially reduce the number of unknowns and hence ambiguity (Day-Lewis *et al.* 2007; Li *et al.* 2011; Commer *et al.* 2014).

Decoupling state, adjoint and parameter grids can be particularly advantageous in a combination with local mesh refinement on unstructured and non-conforming meshes. This offers an opportunity to carry out adaptive mesh refinement of the grids using individual criteria. The latter is an important aspect, attributed to the fact that different requirements are imposed on the meshes: a state and adjoint meshes have to provide accurate EM responses, whereas parameter mesh should adequately represent imaged subsurface structure. As a result, development of the automatic methods for obtaining such meshes would alleviate the burden of choosing a mesh in inverse problems and reduce ambiguity.

Existence of optimal error-estimators for Maxwell's equations (Bürg 2013), and efficient automatic adaptive refinement strategies and solvers (Ren *et al.* 2013; Grayver & Bürg 2014) make the problem of finding optimal state and adjoint variable meshes practically feasible. For MT inversion one typically deals with data at frequencies which span over many orders of magnitude. This prompts using individual meshes for each frequency and performing adaptive mesh refinement for them independently. Altogether, decoupling state, adjoint and parameter meshes allows reducing number of unknowns in inversion, whereas using goal-oriented adaptive local mesh refinement for each frequency independently ensures calculation of accurate MT responses within reasonable amount of computational time. Of note is the recent work by Key (2012) who used goal-oriented mesh refinement strategy (Key & Ovall 2011) for constructing optimal state variable meshes in his 2.5-D geo-electromagnetic inversion code.

Parameter variable mesh can be refined as well. In contrast to refinement of the state or adjoint variable meshes discussed in the previous paragraph, it is not immediately clear which refinement strategy shall be used for a parameter mesh. A good refinement strategy should eventually allow to converge to the true solution. There is, however, no unique solution to an ill-posed discrete inverse problem. Nevertheless, many authors addressed this problem, showing that the idea of adaptive mesh refinement in inverse problems is feasible. In medical imaging, Bangerth (2008) and Li *et al.* (2014) used adaptive finite-element method (FEM) to substantially reduce computational time and dimensionality of the problem. Similarly, Beilina *et al.* (2014) used a rigorously elaborated method for adaptive mesh refinement in inverse problems arising from time-domain wave scattering. Haber *et al.* (2007) discussed several practical aspects and demonstrated that adaptive meshes are efficient in the context of 3-D DC geo-electric inversion. These works share few common aspects. Specifically, the approaches suggest performing several inversion iterations until a local minimum is reached and no further progress is observed, then a parameter mesh is refined in regions where solution exhibits some irregularity. These regions are typically detected by inspecting gradient of a parameter variable or similar quantities.

Yet another important aspect is the initial mesh design. Since survey layouts are often not regular, data uncertainties vary among stations and due to rapid spatial attenuation of EM fields, model resolution varies in different parts of the model (Grayver *et al.* 2014). Using uniform grids in this case will likely result in unnecessarily large number of unknowns in inversion. This is particularly true for MT inversion with data sets usually consisting of measurements at a wide range of frequencies, resulting in substantially different spatial resolution lengths. Initial mesh should account for these factors. Local mesh refinement offers the means for addressing this problem. For instance, Schaefer *et al.* (2011) used adaptive meshes to perform global surface wave tomography. They refined mesh in regions with more rays and let the mesh to be coarse in those part, which are not well constrained by data. In full-waveform inversion, Fichtner *et al.* (2013) used locally refined meshes to account for uneven distribution of seismic stations.

The main goal of this paper is to develop and study automatic adaptive mesh refinement methods in the context of 3-D MT inversion. Using FEM, a consistent derivation is presented and a number of practical aspects related to the implementation of a parallel and distributed 3-D inversion code are covered. General formalism given in this work was adopted from the formulation presented by Bangerth (2008). Although in some ways similar, this work presents new distributed implementation of the adaptive FEM inverse solver tailored to geo-electromagnetic imaging applications. In addition, several different variations of the parameter mesh refinement are offered and tested on a series of relevant models. To author's knowledge, this is the first study that implements and demonstrates all these techniques in the context of geo-electromagnetic imaging.

2 FORMULATION OF THE PROBLEM

2.1 Forward modelling

2.1.1 Governing equations

Electromagnetic fields in 3-D settings are governed by the following complex-valued PDE equation augmented with the homogeneous Neumann boundary conditions

$$\nabla \times (\mu^{-1}\nabla \times \mathbf{E}^s) + \mathrm{i}\omega\sigma\mathbf{E}^s = -\mathrm{i}\omega(\sigma - \sigma^0)\mathbf{E}^0 \quad \text{in} \quad \Omega,$$
$$\mathbf{n} \times \mu^{-1}\nabla \times \mathbf{E}^s = 0 \quad \text{on} \quad \partial\Omega, \tag{1}$$

where $\Omega \subseteq \mathbb{R}^3$ is some bounded domain, ω angular frequency, $\mathbf{E}^s : \Omega \to \mathbb{C}^3$ a scattered complex-valued electric field and $\sigma : \Omega \to \mathbb{R}_+$ denotes real-valued electrical conductivity; μ is assumed to be magnetic permeability of free space, but in general is not limited to that choice; σ^0 is the conductivity of a 1-D background model; $\mathbf{E}^0$ known background field and $\mathbf{n}$ is an outward-pointing normal vector on $\partial\Omega$. The total field is calculated as $\mathbf{E} = \mathbf{E}^0 + \mathbf{E}^s$.

In this work, a real-valued analogue of the eq. (1) was solved. Namely, inserting $\mathbf{E}^s := \mathbf{E}^s_R + i\mathbf{E}^s_I$ into (1) and splitting the result, yields a system of two coupled real-valued equations

$$\nabla \times \left(\mu^{-1}\nabla \times \mathbf{E}^s_R\right) - \omega\sigma\mathbf{E}^s_I = \omega(\sigma - \sigma^0)\mathbf{E}^0_I \quad \text{in} \quad \Omega$$
$$\nabla \times \left(\mu^{-1}\nabla \times \mathbf{E}^s_I\right) + \omega\sigma\mathbf{E}^s_R = -\omega(\sigma - \sigma^0)\mathbf{E}^0_R \quad \text{in} \quad \Omega$$
$$\mathbf{n} \times \mu^{-1}\nabla \times \mathbf{E}^s_R = 0 \quad \text{on} \quad \partial\Omega$$
$$\mathbf{n} \times \mu^{-1}\nabla \times \mathbf{E}^s_I = 0 \quad \text{on} \quad \partial\Omega. \tag{2}$$

The magnetic field $\mathbf{H}$ is calculated from $\mathbf{E}$ by employing Faraday's law.

In MT, responses such as impedance tensor are typically inverted. Using notations $E_{x1}, E_{y1}, H_{x1}, H_{y1}$ and $E_{x2}, E_{y2}, H_{x2}, H_{y2}$ for horizontal components of the electric and magnetic fields for the two orthogonal polarizations, impedance tensor $\mathbf{Z}$ is expressed as (Berdichevskii & Dmitriev 2008)

$$\begin{pmatrix} E_{x1} & E_{x2} \\ E_{y1} & E_{y2} \end{pmatrix} = \begin{pmatrix} Z_{xx} & Z_{xy} \\ Z_{yx} & Z_{yy} \end{pmatrix} \begin{pmatrix} H_{x1} & H_{x2} \\ H_{y1} & H_{y2} \end{pmatrix}. \tag{3}$$

2.1.2 Discretization using finite elements

Eq. (2) was discretized using the FEM. To this end, domain Ω is represented using a mesh $\mathbb{T}^s$ consisting of non-overlapping hexahedral cells. Let an electric field $\mathbf{E}^s$ be from a space of 3-D vector functions with well-defined curl. The discrete version of this space is constructed using Nédélec finite elements (Monk 2003). After representing $\mathbf{E}^s$ in terms of Nedelec basis functions and multiplying the second equation in the system (2) by -1 to preserve symmetry, it can be written as a system of linear equations

$$\begin{pmatrix} \mathbf{C} & -\mathbf{M} \\ -\mathbf{M} & -\mathbf{C} \end{pmatrix} \begin{pmatrix} \mathbf{E}^s_R \\ \mathbf{E}^s_I \end{pmatrix} = \begin{pmatrix} \mathbf{s}_R \\ -\mathbf{s}_I \end{pmatrix}, \tag{4}$$

where $\mathbf{C}$, $\mathbf{M}$ and $\mathbf{s}$ are the discrete curl-curl, mass and right-hand side (RHS) terms. For a detailed derivation and verification of the numerical scheme, the reader is referred to Grayver & Bürg (2014) and (Grayver & Kolev 2015). Note that eqs (1)–(2) deal with continuous variables, whereas in eq. (4) discrete counterparts were used. For brevity, notation above omits this difference.

2.1.3 Solver

Many factors affect numerical properties of the system (4), making it difficult to solve (Schwarzbach 2009; Ernst & Gander 2011). Grayver & Bürg (2014) presented a robust solver that can also scale to large problems. In this context, robustness means an ability of a method to preserve its performance for a wide range of model settings such as frequency, conductivity contrasts, locally-refined and/or stretched grids, whereas scalability allows taking advantage of modern distributed computational platforms when solving large-scale problems. Following this approach, system (4) is solved by using FGMRES method (Saad 2003) with the optimal block diagonal preconditioner

$$\mathbf{A} = \begin{pmatrix} \mathbf{B} & \mathbf{0} \\ \mathbf{0} & \mathbf{B} \end{pmatrix}, \tag{5}$$

where $\mathbf{B} = \mathbf{C} + \mathbf{M}$ is real and symmetric. Application of $\mathbf{A}^{-1}$ to a vector requires solution of two systems with system matrix $\mathbf{B}$. For large-scale problems, $\mathbf{B}^{-1}$ can be efficiently approximated by means of auxiliary space preconditioning (Hiptmair & Xu 2007; Kolev & Vassilevski 2009). For problems of moderate size and multiple RHS vectors, however, a direct solver outperforms auxiliary space preconditioner. For MT modelling, two solutions for orthogonal source polarizations and a dual problem (if the goal-oriented error estimator is used; see Section 2.1.4) are needed. Due to the homogeneous boundary conditions (1), the system matrix in eq. (4) remains the same for all RHS vectors. Therefore, direct solvers are efficient since they compute solutions for multiple RHS vectors virtually at the cost of a single matrix factorization. As a result, a hybrid preconditioner was used: for problems with number of degrees of freedom (DoFs) less than a prescribed threshold, a distributed direct solver was utilized for approximating $\mathbf{B}^{-1}$, whereas for larger problems the auxiliary space preconditioner was more efficient. In this work, this threshold was set to 900 000 DoFs based on the scalability analysis presented in Grayver & Bürg (2014).

2.1.4 Goal-oriented error estimator

To obtain accurate EM responses, adaptive mesh refinement has been used. Starting from a coarse mesh $\mathbb{T}^s_0$, eq. (2) was solved on a series of automatically refined meshes. For every subsequent mesh $\mathbb{T}^s_i$, a subset of cells with the most inaccurate solution is identified (Ainsworth & Oden 2000) and refined. For a numerical solution $\mathbf{F} = \mathbf{F}_R + i\mathbf{F}_I$ and all cells $K \in \mathbb{T}^s_i$, error indicators $\eta_K(\mathbf{F})$ are calculated as follows:

$$\eta_K(\mathbf{F})^2 := \eta_{R,K}(\mathbf{F})^2 + \eta_{J,K}(\mathbf{F})^2, \tag{6}$$

where residual-based and jump-based terms are given by

$$\eta_{R,K}(\mathbf{F})^2 := h_K^2 \left\| \nabla \times \left(\mu^{-1}\nabla \times \mathbf{F}\right) + \mathrm{i}\omega\sigma\mathbf{F} - \mathbf{s} \right\|^2_{L^2(K)}, \tag{7}$$

and

$$\eta_{J,K}(\mathbf{F})^2 := \frac{1}{2}\sum_{e=1}^{6} h_e \left(\left\| [\mathbf{n}_e \times (\mu^{-1}\nabla \times \mathbf{F})] \right\|^2_{L^2(e)} + \left\| [\mathbf{n}_e \cdot (\mathrm{i}\omega\sigma\mathbf{F} - \mathbf{s})] \right\|^2_{L^2(e)} \right). \tag{8}$$

Here, $h_K := \mathrm{diam}(K)$ is the diametre of the hexahedra K, e an interior face of K, $h_e := \mathrm{diam}(K)$ diametre of a face. Square brackets $[\cdot]$ denote the jump of the quantity across element boundaries. The physical meaning of these terms is discussed in Grayver & Bürg (2014).

To facilitate accuracy of EM fields at receiver positions, dual-weighted or goal-oriented error estimators (Ainsworth & Oden 2000; Bangerth & Rannacher 2003) can be used. These approaches help avoid refinement in regions where inaccuracy of a solution does not significantly affect EM fields at receivers, and were shown to be efficient in geo-electromagnetic modelling (Key & Ovall 2011; Ren *et al.* 2013). In this work, a goal-oriented error estimator similar to the one presented in Ren *et al.* (2013) was implemented. Along with numerical solution of eq. (2), a dual problem with fictitious sources placed at the receiver locations is solved. Denoting the dual solution by $\mathbf{E}^D$, the error indicator used to mark cells for refinement is given by

$$\eta_K^2 := \eta_K(\mathbf{E}^s)^2 \eta_K(\mathbf{E}^D)^2. \tag{9}$$

Finally, a fixed fraction $\theta \in (0, 1]$ of cells $\mathcal{T} \subseteq \mathbb{T}^s$ with the largest error indicators is selected. These cells are then refined through bisection in all three dimensions. The procedure is repeated until a predefined error level is attained or the number of maximum cycles $n_{\max}$ is reached, or number of degrees of freedom exceeds a prescribed threshold.

To derive MT transfer functions in 3-D settings, EM fields for two orthogonal source polarizations are calculated. Since forward modelling problem was formulated using homogeneous boundary conditions, system matrix and numerous related data structures remain identical for both polarizations as long as the mesh remains the same. To avoid time and memory overhead associated with the assembling and storing two matrices, the same mesh was used for both polarizations. Therefore, a subset of cells marked for refinement at every refinement cycle was formed as a union $\mathcal{T} = \mathcal{T}_1 \cup \mathcal{T}_2$ of the subsets identified for two polarizations individually.

2.2 Inversion

Inverse problem is formulated in form of a non-linear optimization problem

$$\min_{m \in \mathcal{W}} \mathcal{F}(m), \tag{10}$$

with

$$\mathcal{F}(m) := \frac{1}{2} \left\| f(m) - \mathbf{d}^{obs} \right\|^2_{\mathbf{C}_d^{-1}} + \frac{\beta^2}{2} \left\| \mathcal{L}(m - m_a) \right\|^2_{L_2(\Omega)}, \tag{11}$$

where $\mathbf{d}^{obs} \in \mathbb{R}^{N_d}$ is a vector of observed data, $\mathbf{C}_d$ data covariance matrix, $m \in \mathcal{W}$ parameter variable. In this work, m is given by the log-bounded transformation of the electrical conductivity $m = \nu(\sigma)$ (Kim & Kim 2011). Forward operator $f(m) : \mathcal{W} \to \mathbb{R}^{N_d}$ calculates MT responses for the given vector of parameters. $\mathcal{W} = \{v \in L_2(\Omega)\}$ is a space of piecewise-constant functions defined in domain Ω. Regularization term consists of the regularization parameter $\beta \geq 0$, *a priori* model $m_a \geq 0$, and $\mathcal{L} : \mathcal{W} \to \mathcal{W}$ is a Laplace-like operator. Regularization helps mitigate effects of poorly constrained data-term and narrow a class of feasible solutions.

To discretize eq. (11), let domain Ω be represented by a mesh $\mathbb{T}^m$ and $\mathbf{W}^m \subset \mathcal{W}$ be a space of piecewise-constant functions on it. $N_m = \dim(\mathbf{W}^m)$ is the number of cells in $\mathbb{T}^m$.

Usually, some cells of $\mathbb{T}^m$ are kept fixed in inversion (such as air, sea or padding cells). To account for this, a projector matrix $\mathbf{D} \in \mathbb{R}^{N_a \times N_m}$ is introduced with N_a being a number of active cells. The active set is a subset of cells from $\mathbb{T}^m$ which get updated. Matrix $\mathbf{D}$ maps a vector defined for all cells in $\mathbb{T}^m$ onto an active set of cells.

To find a solution of the eq. (10), Gauss–Newton method was utilized. This requires solving a system of linear equations for the nth model update:

$$\left(\mathbf{J}_n^T \mathbf{C}_d^{-1} \mathbf{J}_n + \beta^2 \mathbf{L}\right) \delta \mathbf{m}_n = -\mathbf{J}_n^T \mathbf{C}_d^{-1} (\mathbf{f}(\mathbf{m}_n) - \mathbf{d}^{obs}) - \beta^2 \mathbf{L}(\mathbf{m}_n - \mathbf{m}_a), \tag{12}$$

where $\mathbf{L} = \mathbf{D}\tilde{\mathbf{L}}\mathbf{D}^T$ is a finite-difference approximation of Laplace operator with homogeneous Neumann boundary conditions on active cells of $\mathbb{T}^m$. Details regarding its discretization on unstructured and non-conforming meshes are covered by Haber *et al.* (2007) and Schwarzbach & Haber (2013). Jacobian matrix $\mathbf{J}_n = \tilde{\mathbf{J}}_n \mathbf{D}^T \in \mathbb{R}^{N_d \times N_a}$ is the derivative of the discrete forward operator $\mathbf{f}(\mathbf{m}_n)$ with respect to the parameter variable on active cells. See Pankratov & Kuvshinov (2010) and Egbert & Kelbert (2012) for an elaborated description of its structure. New model is obtained as a sum of the current model and scaled update $\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha \delta \mathbf{m}_n$, where $0.1 < \alpha \leq 1$ is a step-length (Nocedal & Wright 1999). To allow the inversion to leave a local minimum, an increase in the objective function is accepted if a suitable $\alpha > 0.1$ cannot be found.

The regularization parameter β is calculated at every iteration as a ratio of the spectral radii of the data and regularization related operators:

$$\beta_n = \gamma \frac{\|\mathbf{C}_d^{-1/2} \mathbf{J}_n\|_2^2}{\|\mathbf{L}\|_2}, \tag{13}$$

where $0 \leq \gamma \leq 1$ is a user-defined relative scaling. See Grayver *et al.* (2013) for a discussion of this approach to calculate regularization parameter.

System (12) was solved using CGLS iterative solver (Hansen 1998). Thus, matrix $\mathbf{J}$ is never formed explicitly, rather matrix-vector products are calculated by solving N_f state and adjoint problems with N_f being the number of frequencies in a data set (Egbert & Kelbert 2012). System (12) is solved approximately using $N_{it} < N_a$ iterations, hence introducing additional regularization because of implicit filtering of eigenvectors corresponding to small eigenvalues (Haber 1997).

2.2.1 Decoupling state, adjoint and parameter meshes

As outlined in the introduction, decoupling state, adjoint and parameter meshes provides additional flexibility and opens the door for using specific adaptive mesh refinement criteria for state, adjoint and parameter variables.

In attempt to keep technical complexity feasible and computational time low, a number of requirements on meshes were imposed:

(i) State mesh $\mathbb{T}^s$ was used for solving corresponding adjoint problem. This significantly reduces amount of computations since all linear operators remain the same for both state and adjoint variables. Moreover, error indicator (9) used to refine state mesh is essentially a product of the estimated errors for the forward and adjoint variables, thus it aims at finding a mesh that reduces error for both variables simultaneously. This permits using the same mesh without significant accuracy penalties for either of problems.

(ii) At nth inversion iteration, a separate state meshes $\mathbb{T}^s_{n,i}$ $i = 1 \ldots N_f$ were constructed for each frequency and refined independently. This enables building optimal meshes for each frequency independently using automatic goal-oriented mesh refinement (see Section 2.1.4). This is particularly efficient since different frequencies can be calculated in parallel.

(iii) Any state mesh $\mathbb{T}^s_{n,i}$ can be obtained from the parameter mesh $\mathbb{T}^m_n$ by the hierarchical refinement. In other words, an initial coarse state mesh always coincides with the parameter mesh

and optionally gets refined to calculate EM fields accurately. Such hierarchical representation greatly simplifies implementation, especially in distributed memory environments.

Let us elaborate on the second point here. While building N_f state meshes may seem more expensive than using the same mesh for all frequencies, it is rather difficult to find a single mesh that would resolve EM fields at all frequencies equally well. Commonly, MT data sets incorporate responses at frequencies ranging over 6–8 orders of magnitude. Therefore, a single mesh would need to be refined everywhere where a solution for any frequency exhibits non-smooth behaviour (e.g. across conductivity contrasts). For different frequencies the rate of solution variability is different throughout the modelling domain. Therefore, it is highly unlikely that regions where the mesh would have to be refined, and degree of refinement will coincide for all frequencies. As a result, one either has to choose a very fine mesh incurring computational overburden or compromise accuracy. Since neither is favourable, using independent meshes for each frequency is an adequate alternative. The only downside of this decision is a slightly more complicated implementation due to extra bookkeeping. For the cases when densely spaced frequencies or a big number of controlled sources (e.g. in CSEM) are given, one can group them in order to reduce computational complexity (Key & Ovall 2011).

To solve eq. (12), products of a vector with sensitivity matrix $\mathbf{J}$ or its transpose are required. Due to the matrix-free formulation used in this work, these products entail solution of N_f state or adjoint equations. Recall that for ith frequency, mesh $\mathbb{T}_i^s$ is typically finer than the corresponding parameter mesh $\mathbb{T}^m$ for which the sensitivity matrix $\mathbf{J}$ has to be defined eventually. This requires constructing a transfer operator $\mathbf{P}_i : \mathbf{W}^m \rightarrow \mathbf{W}_i^s$ that takes a variable defined on $\mathbb{T}^m$ and maps it onto $\mathbb{T}_i^s$. Taking this into account and using block structure of the sensitivity matrix (Egbert & Kelbert 2012) allows to write

$$\mathbf{Jv} = \mathop{\Big\|}_{i=1}^{N_f} \mathbf{J}_i \mathbf{P}_i \mathbf{D}^T \mathbf{v}$$

$$\mathbf{J}^T \mathbf{u} = \sum_{i=1}^{N_f} \mathbf{D} \mathbf{P}_i^T \mathbf{J}_i^T \mathbf{u}, \tag{14}$$

where symbol $\|$ denotes vector concatenation. Sensitivity matrix for the ith frequency $\mathbf{J}_i$ is calculated on the mesh $\mathbb{T}_i^s$ and acts on the space $\mathbf{W}_i^s$. Structure of the corresponding transfer matrix $\mathbf{P}_i \in \mathbb{R}^{dim(\mathbb{T}_i^s) \times N_m}$ generally depends on a parameter variable space $\mathbf{W}$ and geometric relation between $\mathbb{T}_s^i$ and $\mathbb{T}^m$. Exploiting a piecewise-constant basis of $\mathbf{m}$ and the fact that $\mathbb{T}_i^s$ is obtained from $\mathbb{T}^m$ by hierarchical refinement, the values for cells of $\mathbb{T}^m$ which were refined in $\mathbb{T}_i^s$ are simply weighted sums of the values from all terminal children, rendering projection operator to be a sparse matrix with zeros and ones.

It should be understood that all objects introduced above also depend on an iteration number n. Unless unclear from the context, this index is omitted to avoid notation clutter.

2.2.2 *Initial parameter mesh design*

Model resolution is a function of data uncertainty, survey layout, conductivity distribution and governing PDE equations. Depending on these factors, some regions of the model are better constrained than others. Therefore, it is desirable for an initial parameter mesh $\mathbb{T}_0^m$ to reflect the actual model resolution in some way. One can design such mesh manually using some heuristic rules such as skin depth. This, however, may become a tedious task and result in a mesh that is still too fine or coarse. The goal of this section is to develop an automatic approach to choose an initial parameter mesh $\mathbb{T}_0^m$ such that it takes into account all mentioned factors.

Recall that for non-linear inverse problems, in the vicinity of a model $\mathbf{m}$, model resolution is characterized by the resolution matrix $\mathbf{R}_m$ (Menke 1984; Kalscheuer & Pedersen 2007). Before stating this matrix, first let

$$\mathbf{C}_d^{-1/2}\mathbf{J} = \mathbf{USV}^T = \begin{pmatrix} \mathbf{U}_r & \mathbf{U}_0 \end{pmatrix} \begin{pmatrix} \mathbf{S}_r & \mathbf{0} \\ \mathbf{0} & \mathbf{0} \end{pmatrix} \begin{pmatrix} \mathbf{V}_r^T \\ \mathbf{V}_0^T \end{pmatrix} \tag{15}$$

be the singular value decomposition (SVD) of the weighted Jacobian matrix. Here, r denotes the rank of the matrix, $\mathbf{S} \in \mathbb{R}^{N_d \times N_a}$ contains singular values on the main diagonal; matrices $\mathbf{U} \in \mathbb{R}^{N_d \times N_d}$ and $\mathbf{V} \in \mathbb{R}^{N_a \times N_a}$ represent left and right singular vectors, respectively, and are orthonormal. Further, let $\mathbf{C}_d^{-1/2}\mathbf{J}_r = \mathbf{U}_r\mathbf{S}_r\mathbf{V}_r^T$ be the SVD of the weighted Jacobian restricted to the range of the matrix. Then, the resolution matrix is given by

$$\mathbf{R}_m = \left(\mathbf{V}_r\mathbf{S}_r^2\mathbf{V}_r^T\right)^{-1} \mathbf{J}_r^T\mathbf{C}_d^{-1}\mathbf{J}_r = \left(\mathbf{V}_r\mathbf{S}_r^2\mathbf{V}_r^T\right)^{-1} \mathbf{V}_r\mathbf{S}_r^2\mathbf{V}_r^T = \mathbf{V}_r\mathbf{V}_r^T. \tag{16}$$

Note that this matrix does not involve regularization, thus specifying how model is constrained by data alone. Each row of this matrix is a point-spread function (Alumbaugh & Newman 2000; Grayver *et al.* 2014) quantifying how much a model parameter is influenced by all other model parameters. This information can be used to design an initial parameter mesh. Generally, calculation of this matrix is prohibitive. Nevertheless, since $\mathbf{J}_r^T\mathbf{C}_d^{-1}\mathbf{J}_r$ is a compact operator (Bui-Thanh & Ghattas 2012), its low-rank approximation can be build efficiently. Because resolution estimates are merely used as an indicator for mesh refinement, a relative quantity is sufficient. Therefore, the main diagonal of $\mathbf{R}_m$ was used as a measure of resolution (Vasco *et al.* 2003; Zhang & Thurber 2007).

Low-rank SVD can be obtained by computing $l \ll r$ singular vectors corresponding to the largest singular values. In this work, the following low-rank approximation of the main diagonal of $\mathbf{R}_m$ was computed

$$\mathrm{diag}(\mathbf{R}_m) \approx \mathbf{r}_m = \sum_{k=1}^{l} (\mathbf{v}_k)^2, \tag{17}$$

where singular vectors $\mathbf{v}_k$ were calculated using thick-restart Lanczos bidiagonalization algorithm (Wu & Simon 2000). Since singular vectors are unit vectors, elements of $\mathbf{r}_m$ range between zero and one with those closer to one indicating well-resolved model parameters.

Using these estimates, the following automatic initial mesh refinement algorithm was implemented:

(i) Choose a starting model $\mathbf{m}_0$ and an initial coarse mesh $\mathbb{T}_0^m$
(ii) Calculate $\mathbf{r}_m$ using eq. (17)
(iii) Select a subset of $\lfloor \epsilon N_a \rfloor$ $(0 < \epsilon < 1)$ maximal elements of $\mathbf{r}_m$ and refine corresponding cells of $\mathbb{T}_0^m$.
(iv) If necessary, go to Step (ii) and repeat refinement.

So far, the choice of subspace dimension l has not been discussed. Note that

$$(\mathbf{r}_m)_j = \left(\sum_{k=1}^{l} (\mathbf{v}_k)^2\right)_j \leq \left(\sum_{k=1}^{n} (\mathbf{v}_k)^2\right)_j \quad \forall n : l < n \leq r. \tag{18}$$

This means resolution gets enhanced as subspace dimension increases, that is more singular vectors are added into the sum (17).

In practice, factors such as data uncertainties and discretization errors limit the subspace dimension when calculating a model update (Alumbaugh & Newman 2000; Grayver *et al.* 2014). Therefore, it is reasonable to choose l to be on the same order as the subspace dimension used to solve eq. (12).

Several practical aspects need to be clarified. First of all, the introduced approach relies on a starting model $\mathbf{m}_0$ and the constructed mesh may vary for different models. However, a good starting model is vital when solving inverse problems using derivative-based methods and thus is required anyway. Following the conventional inversion workflow, several starting models should be tested. The other issue arises by recalling that the matrix $\mathbf{R}_m$ is valid only in some vicinity of the model $\mathbf{m}_0$ and estimated resolution is usually lower than the one given by non-linear analysis for conductive structures (Kalscheuer & Pedersen 2007). One could possibly overcome this issue by recalculating resolution estimates after several iterations and adjusting mesh accordingly or using adaptive mesh refinement as explained in the next section.

2.2.3 *Adaptive parameter mesh refinement*

Parameter variable vector $\mathbf{m}$ is updated iteratively by solving eq. (12) and performing line-search. To account for updating structure and better resolve emerging contrasts, parameter mesh can be refined in the neighborhood of points where parameter variable varies most. For this purpose, the following adaptive mesh refinement algorithm was implemented:

(i) Choose a starting model $\mathbf{m}_0$ and initial coarse parameter mesh $\mathbb{T}_0^m$ in the domain Ω.

(ii) *Optional:* Use algorithm from Section 2.2.2 to refine $\mathbb{T}_0^m$ on the basis of linearized model resolution estimates.

(iii) Construct initial state variable meshes $\mathbb{T}_i^s = \mathbb{T}_0^m$, $\forall i = 1..N_f$.

(iv) Calculate predicted data for the current model using adaptive goal-oriented mesh refinement (see Section 2.1.4). Check if any of the stopping conditions are met: $n \geq n_{max}$ or $\delta(\mathbf{f}(\mathbf{m}_n) - \mathbf{d}^{obs}) \leq \varepsilon$, where n_{max} is a maximum number of Gauss-Newton iterations, ε target data misfit. If yes, terminate inversion.

(v) Update model $\mathbf{m}_{n+1} = \mathbf{m}_n + \alpha\delta\mathbf{m}_n$ using Gauss-Newton algorithm (see equation 12).

(vi) Stop updating model $\mathbf{m}_n$ if optimization is stagnated and no further progress can be made on the current mesh. To detect this, check if step-length $\alpha < 0.1$ or sufficient reduction in misfit is absent for several iterations. Otherwise set $n = n + 1$ and go to Step (iv).

(vii) Calculate spatial derivative of the current model $\mathbf{g} = |\nabla(\mathbf{m}_n - \mathbf{m}_a)|$ and refine a fraction $\epsilon \in (0, 1)$ of cells for which $\mathbf{g}$ is the largest.

(viii) Construct new state meshes using refined parameter mesh obtained at the previous step and interpolate $\mathbf{m}_n$ on it. Set $n = n + 1$ and go to Step (iv).

While there is generally no guarantee that this strategy will converge to the true solution, it has multiple advantages when compared to the conventional inversion using a pre-determined mesh: (i) at the beginning the parameter mesh is coarse and iterations take less computational time; (ii) coarser mesh reduces ambiguity and acts as additional regularization; (iii) parameter mesh automatically accounts for gradually updating model structure; (iv) state/adjoint meshes are refined independently to deliver accurate EM responses without affecting number of unknowns in inversion.

In practice, a modification of the aforementioned approach may become necessary. Namely, after every mesh refinement, the parameter vector $\mathbf{m}_i$ can be reset to the initial guess $\mathbf{m}_0$ interpolated on the refined mesh. This means that once the parameter mesh gets refined, Gauss–Newton inversion is started anew. While at first glance this may seem as a loss of efficiency, such a decision is justified. If some artefacts appear on a coarser mesh due to the lack of spatial resolution, the inversion needs to suppress them after refinement. If these artefacts become too severe, they can potentially destroy convergence. In this case, it is better to start anew using a refined mesh even at the cost of a longer runtime. For a particular study, it is recommended to test both approaches and choose one that delivers a better result.

3 IMPLEMENTATION

The presented algorithm has been implemented in a distributed manner to harness the computational power of the modern clusters. The code is parallelized over frequencies and modelling domain. In other words, given N_f frequencies, the code can run on nN_f CPUs, where $n \geq 1$. For $n = 1$ this corresponds to parallelization over frequencies only. For $n > 1$, EM fields for each frequency were calculated using n Message Passing Interface (MPI) processes. In this case, the workload for each frequency is distributed approximately uniformly among n MPI processes by partitioning meshes into n subdomains. The outer inversion loop used all nN_f MPI processes. For the implementation of the forward modelling and distributed mesh handling deal.II and p4est libraries (Bangerth *et al.* 2011; Burstedde *et al.* 2011) were used. PETSc (Balay *et al.* 2013) and SLEPc libraries (Hernandez *et al.* 2005) were used for distributed linear algebra operations and low-rank SVD. For the distributed direct solver and auxiliary space preconditioning, MUMPS (Amestoy *et al.* 2006) and hypre (Kolev & Vassilevski 2009) libraries were employed. All third-party libraries and the forward modelling implementation were shown to handle very large problems with up to 10^8–10^9 unknowns, rendering implemented inversion scheme scalable and suitable for solution of large-scale inverse problems.

4 NUMERICAL EXPERIMENTS

The main purpose of this work is to introduce adaptive mesh refinement techniques for large-scale geo-electromagnetic inverse problems. In this section, a number of tests aim to demonstrate efficacy of the presented framework using 3-D MT inversion as an example. The results are divided into two parts as follows: first, the methodology for *a priori* parameter mesh design using linearized model resolution estimates as derived in Section 2.2.2 is tested, followed by a demonstration of the adaptive parameter mesh refinement algorithm from Section 2.2.3.

Several parameters were identical for all experiments. Electrical conductivity of the air layer was set to 10^{-8} S m^{-1}. Normalized relative residual was set to 10^{-8} when solving eq. (4). For linear problems originating in Jacobian matrix-vector products, slightly higher value of 10^{-6} was used. Experiments showed that these values keep computational time relatively low while preserving sufficient accuracy. To avoid inverse crime, all synthetic data were generated using different and much finer meshes than those in inversion. Full impedance tensor $\mathbf{Z}$ (Berdichevskii & Dmitriev 2008) was inverted. To solve system (12), 25 iterations were performed to reach a normalized residual of 10^{-3}.

4.1 Initial parameter mesh design

To gain some insight into how model resolution estimates can be used to steer mesh refinement, several simple tests were performed. For these tests, low-rank SVD with $l \geq 40$ in eq. (17) was calculated

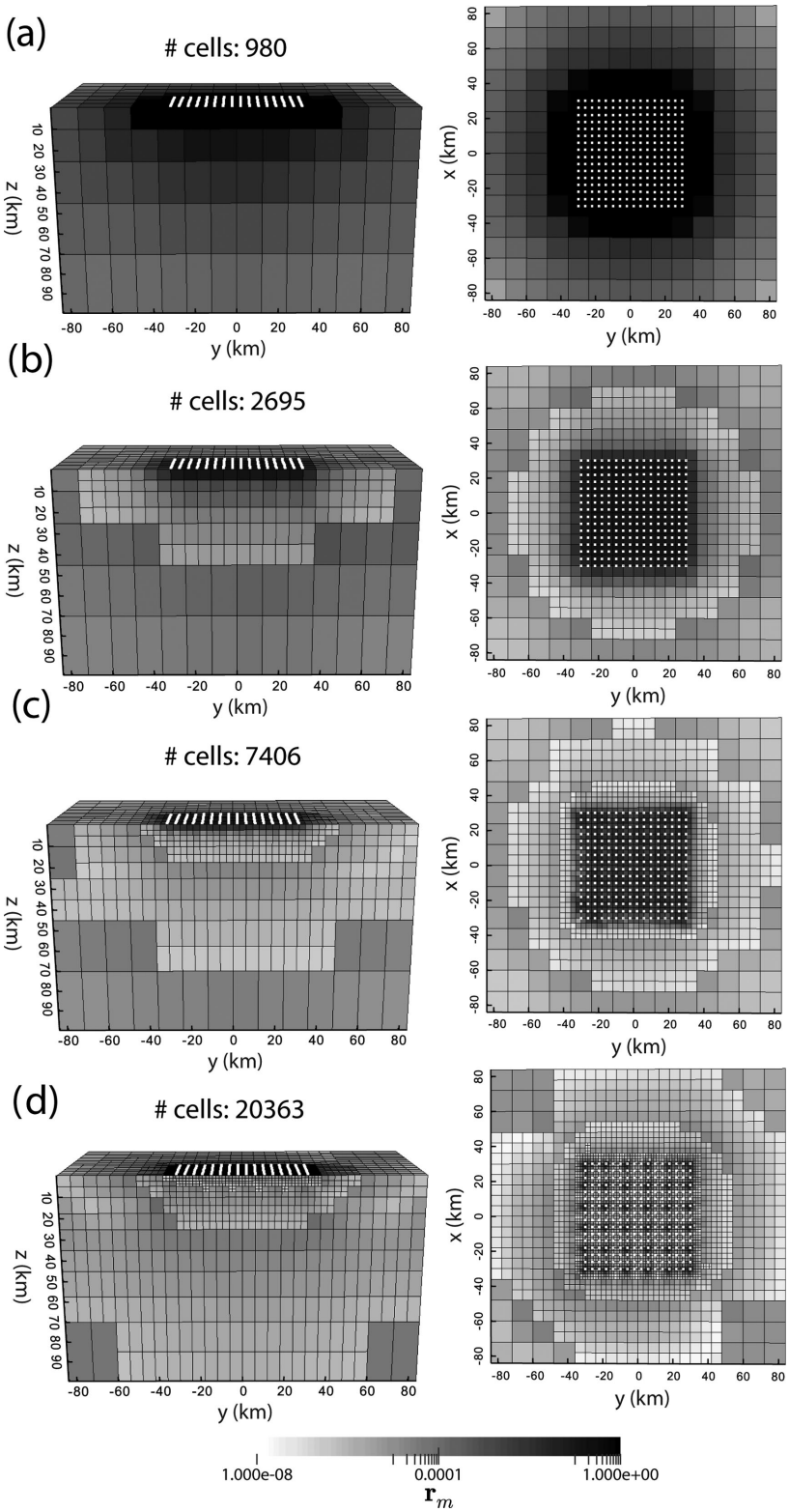

Figure 1. (a–d) Low-rank approximations of the main diagonal of the linearized model resolution matrix for a homogeneous halfspace of 0.01 S m^{-1} on a series of adaptively refined meshes using the algorithm from Section 2.2.2. At each subsequent cycle, 25 per cent of the cells with the largest resolution were refined. Positions of the receiver stations are denoted by the white rectangles. Vertical sections are plotted for $x = 0$ m.

using thick-restart Lanzcos bidiagonalization algorithm. It is important to use an algorithm with some restart and/or reorthogonalization strategy since accuracy of the classic Lanzcos bidiagonalization algorithm in finite precision arithmetic deteriorates rapidly (see Zhang & Thurber (2007) and references therein). Accuracy of the calculated low-rank SVD decomposition can be controlled by using *a posteriori* error bounds (Saad 2011). Numerous tests showed that the error of 0.1 per cent was sufficient to get reasonable resolution

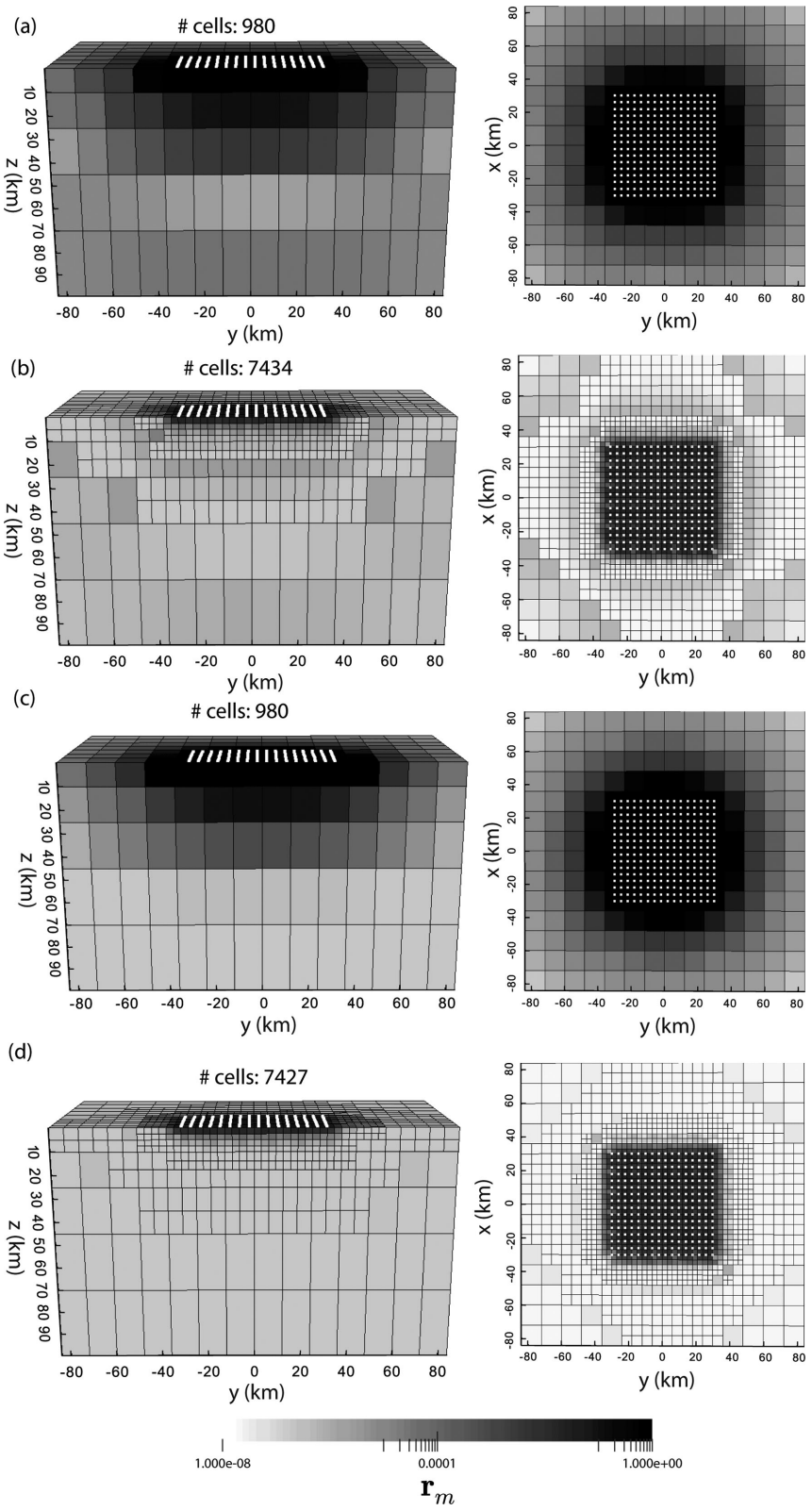

Figure 2. (a, c) Low-rank approximations of the main diagonal of the linearized model resolution matrix for a homogeneous half-spaces of 0.1 and 1 S m^{-1}, respectively. (b, d) After two cycles of adaptive refinement using the algorithm from Section 2.2.2 (*cf.* Fig. 1c). Positions of the receiver stations are denoted by the white rectangles. Vertical sections are plotted for $x = 0$ m.

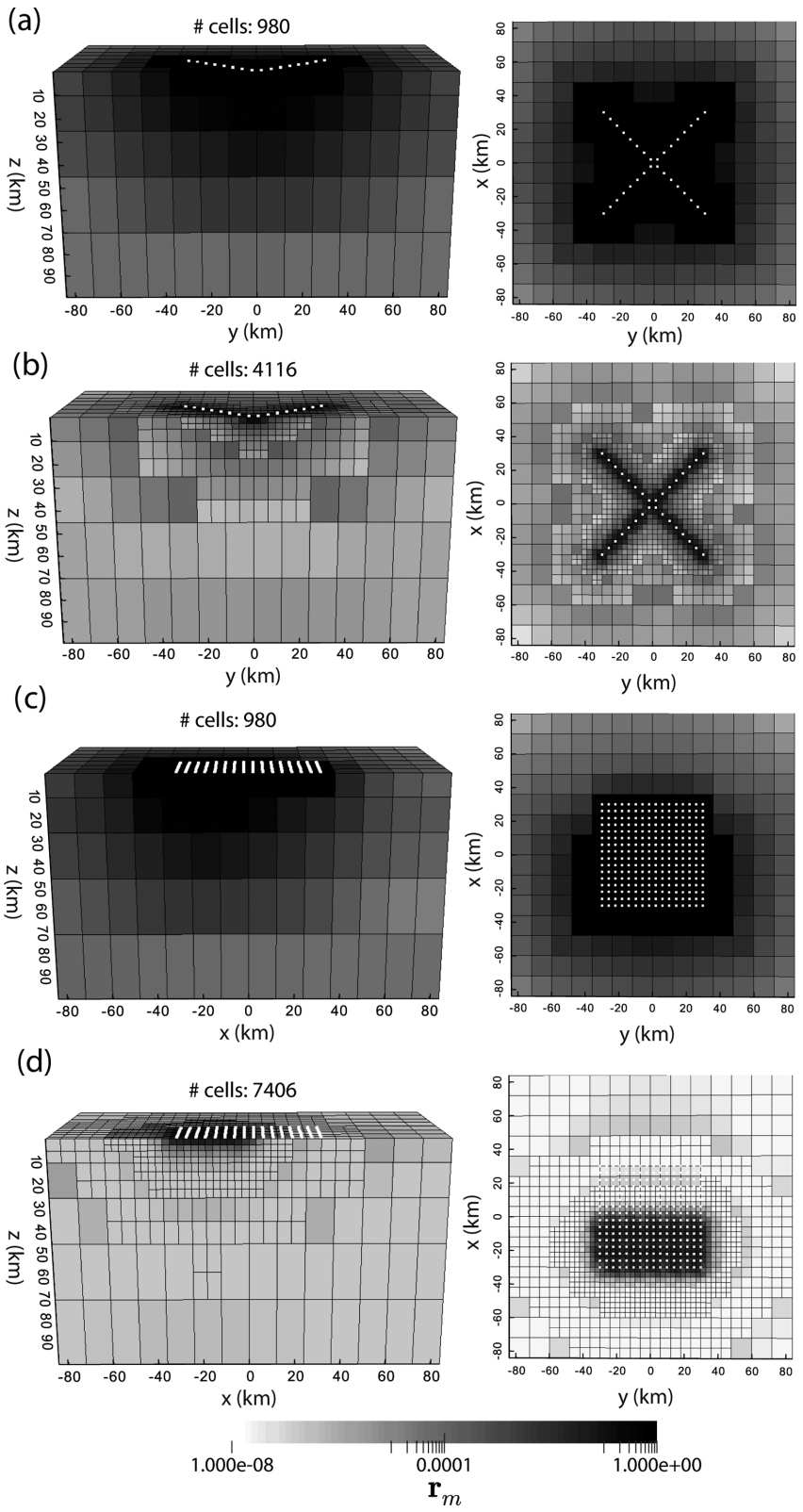

Figure 3. (a and b) Resolution estimates for the homogeneous half-space and non-uniform site distribution on the initial and adaptively refined meshes. (c and d) Same as (a and b), but for the uniform 3-D survey layout. In this case, however, error floors of 30 per cent were set for the half of the receivers located to the north of $x = 0$, whereas error floors of 3 per cent were used for the second half. Positions of the receiver stations are denoted by the white rectangles.

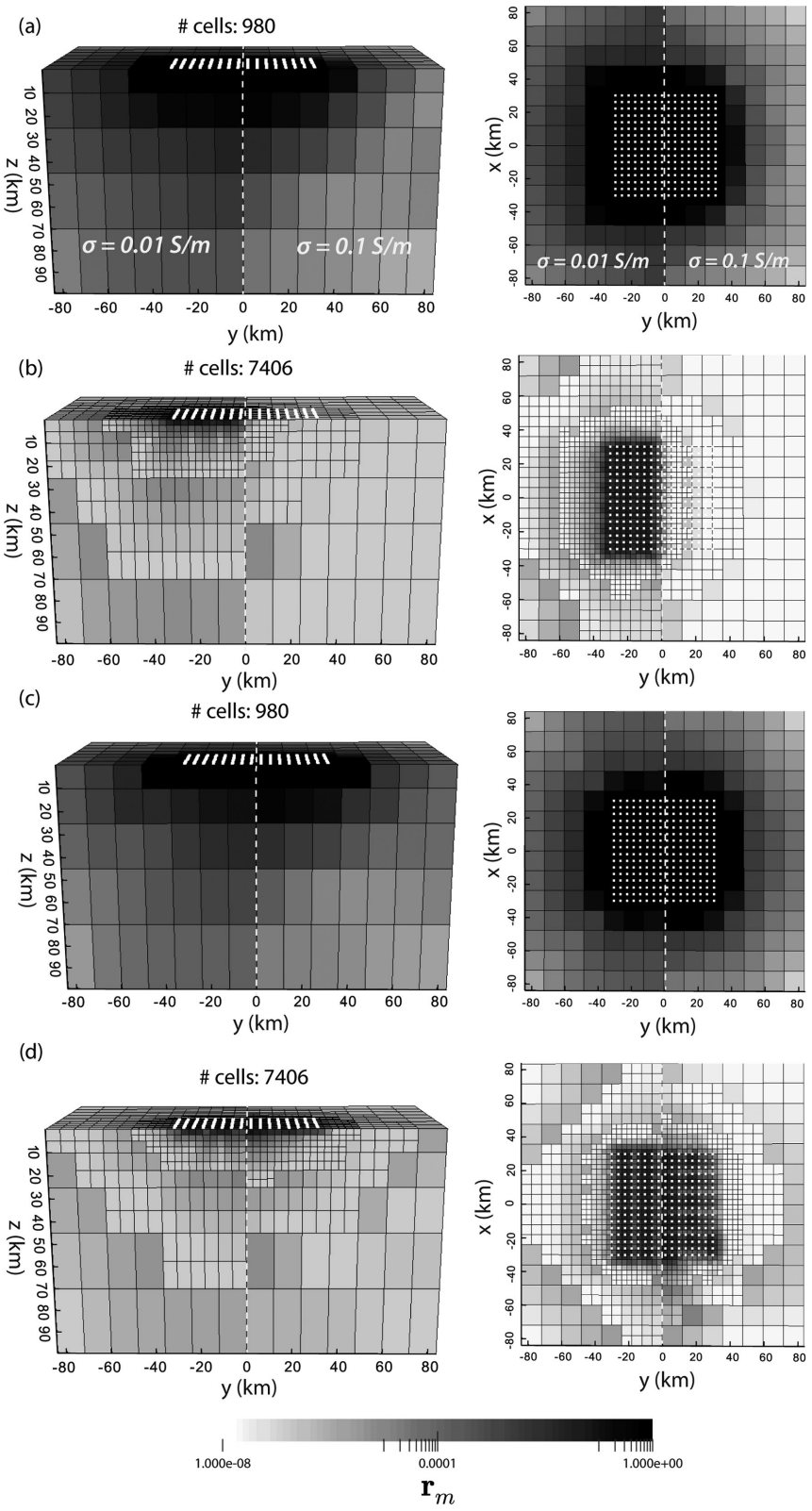

Figure 4. (a and b) Resolution estimates for the quarter-space using full impedance tensor on the initial and adaptively refined meshes. (c and d) Same as (a and b), but with vertical magnetic transfer functions included. Positions of the receiver stations are denoted by the white rectangles. Conductivity and boundary of the vertical contrast are given in (a).

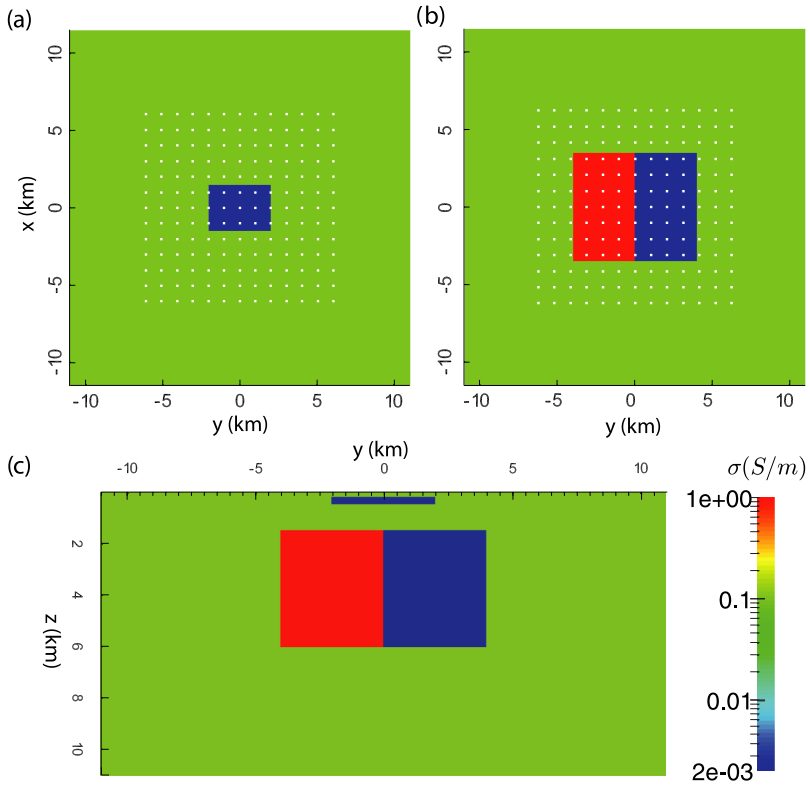

Figure 5. Horizontal sections through the synthetic model at (a) $z = 350$ m, (b) $z = 2500$ m and (c) vertical section for $x = 0$. White rectangles indicate receiver positions.

estimates. In fact, the value of 0.1 per cent was the upper bound. In all considered cases, majority of the singular vectors had much smaller error. Finally, full impedance tensor at all receiver stations and 24 frequencies uniformly distributed on a logarithmic scale in the frequency range of 10^{-3}–10^{3} Hz was used.

Fig. 1 displays a series of adaptively refined meshes using the algorithm from Section 2.2.2. The main diagonal of the linearized model resolution matrix $\mathbf{r}_m$ was approximated for a homogeneous half-space of 0.01 S m^{-1} and 256 receiver stations shown with white rectangles. Starting from the coarse mesh with 980 cells, 25 per cent cells ($\epsilon = 0.25$) with the largest resolution were refined at every subsequent step. Since every refined cell creates eight new cells and becomes inactive itself, value of $\epsilon = 0.25$ results in roughly 2.5 times more cells upon each refinement. Accordingly, after two refinement steps, a mesh with 7406 cells (Fig. 1c) was obtained with the smallest cell size approximately equal to the distance between stations. The last refinement step shown in Fig. 1(d) indicates that singular vectors become rather localized and therefore the obtained mesh is likely overparametrized. Clearly, resolution decays rapidly as the distance from receivers increases, indicating that the model in these regions is poorly constrained by the data. Therefore, it is sufficient to have coarser mesh there. These intuitive expectations are well fulfilled by the algorithm. It is important to reiterate that the linearized resolution estimates are lower than realistic ones and taking into account a potential need to compensate for small near-surface heterogeneties, one may still prefer a finer mesh.

Resolution estimates $\mathbf{r}_m$ depend on an underlying conductivity model. To investigate this dependence, Figs 2(a and b) and Figs 2(c and d) illustrate mesh refinement for the homogeneous halfspaces of 0.1 and 1 S m^{-1} , respectively. As anticipated, resolution drops with increasing conductivity of the model because EM fields attenuate faster. However, resolution patterns among three models exhibit many similarities and since a fixed number of cells is refined at every step, the final meshes look similar and have virtually the same number of cells. Therefore, for this refinement strategy, the dependence of the refined mesh on an initial model is moderate. It is a pattern of the resolution distribution that matters most. For real data with no good starting model available this behaviour is advantageous. On the other hand, if one wants to adjust mesh more to the actual values of the calculated resolution estimates $\mathbf{r}_m$, a different refinement strategy can be used. For instance, a minimal subset of cells, which account for a prescribed fraction of the total quantity (i.e. $\sum_i r_m^i$) can be selected and refined at every step (Bangerth 2002). Considering that every real data set creates new challenges, it is difficult to say in advance which strategy will perform better.

So far, idealistic site distribution and constant data uncertainty were used. Since this is hardly realizable in practice and the model resolution matrix highly depends on both survey layout and data uncertainties, two more scenarios were studied. Figs 3(a and b) show resolution estimates for a homogeneous half-space of 0.1 S m^{-1} and the corresponding adaptively refined mesh for a non-uniform distribution of stations. As expected, refined mesh accommodates survey layout properly by keeping a coarser mesh in regions without stations. In addition to receiver layout and conductivity distribution, resolution is also affected by data uncertainties via data covariance matrix in the eq. (15). Generally, higher uncertainties in data deteriorate resolution. Figs 3(c and d) display resolution estimates on an initial coarse and adaptively refined meshes for a 3-D survey where receivers located to the north and south of $x = 0$ were assigned error floors of 30 per cent and 3 per cent, respectively. Clearly, receivers with higher uncertainties exhibit lower resolution and hence the mesh remains coarse in their neighbourhood.

In case an initial model is not a homogeneous half-space, resolution estimates will reflect contained structures and gear mesh

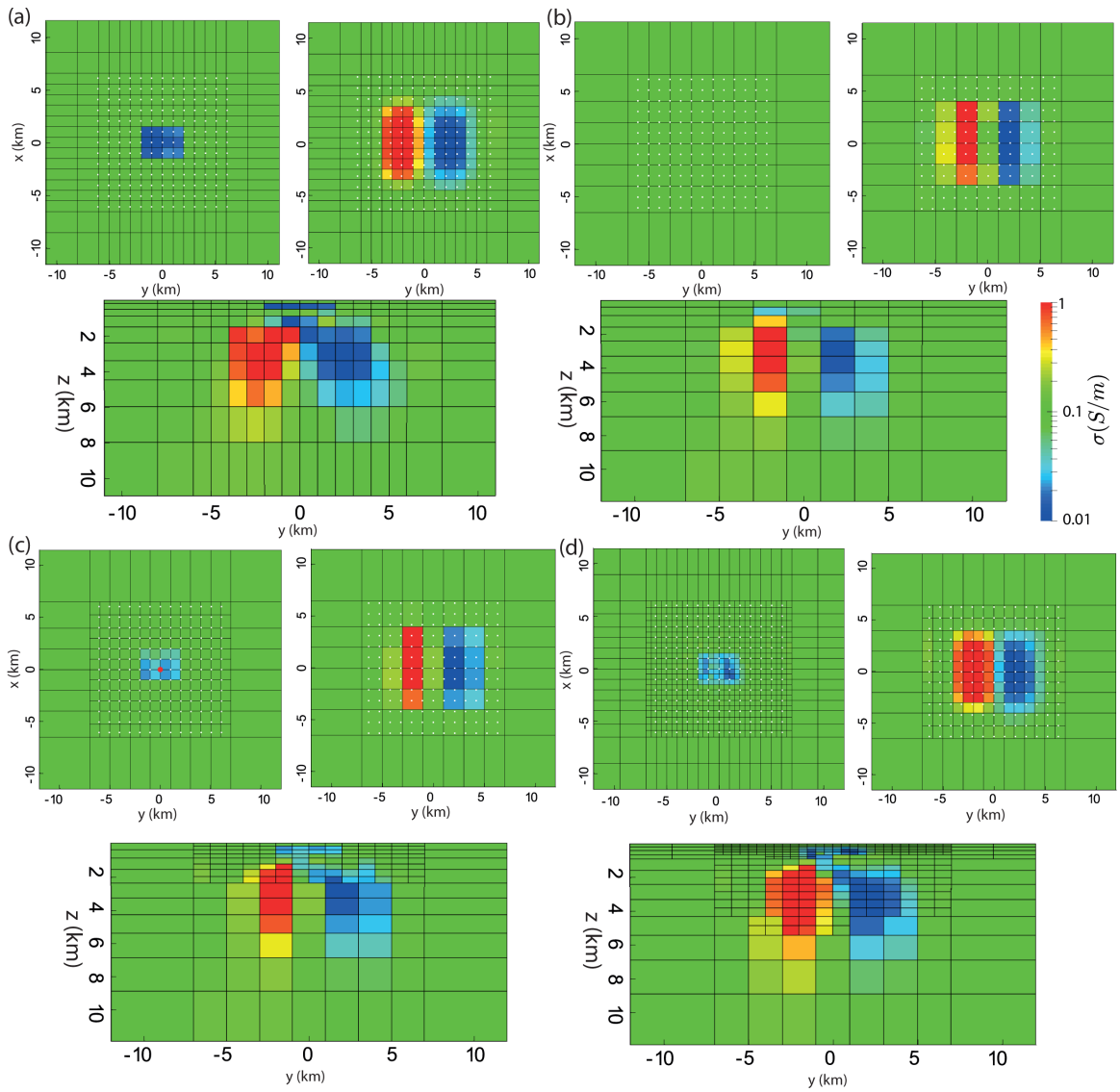

Figure 6. (a) Final model for the mesh with cell boundaries matching edges of the true objects. (b) Final model for the coarse mesh with cell boundaries not coincident with the edges of the true objects. (c and d) Final models for the meshes which were adaptively refined using resolution estimates. Numbers of active cells (i.e. parameter unknowns) are 2100, 420, 1596, 6062 for (a–d), respectively. Horizontal sections are shown at $z = 350$ and $z = 2500$ m.

refinement accordingly. Figs 4(a and b) show resolution estimates for the initial and adaptively refined meshes using a quarter-space model. When using impedance tensor only, the more resistive part gains higher resolution estimates. Because of the sharp vertical conductivity contrast, this model produces large vertical magnetic transfer functions (VTF). Including them in calculations produced meshes shown in Figs 4(c and d). Interestingly, VTF changes resolution distribution and now area around conductivity contrast gains higher resolution which affects mesh refinement accordingly.

Finally, it is worth noting that the linearized model resolution matrix is based on the sensitivity, which in turn is an integral quantity. Therefore, resolution is proportional to a cell volume. Other conditions being equal, bigger cells get higher resolution estimates as can be inferred from the results shown in Figs 1–4.

To demonstrate the utility of the algorithm, inversion of the model shown in Fig. 5 was performed. The model consists of the homogeneous background of conductivity 0.1 S m^{-1} and three rectangular objects: shallow 3 × 4 × 0.3 km object of conductivity 0.002 S m^{-1} located at the depth of 0.2 km and two deeper objects of conductivity 1 and 0.002 S m^{-1}, each of size 7 × 4 × 4.5 km located at the depth of 1.5 km. Data at 169 receivers were generated for 16 frequencies evenly spaced on the logarithmic scale in the range of 0.001–100 Hz. This results in 21 632 real-valued data measurements for inversion. Two percent random Gaussian noise was added to the data. Lower and upper bounds of the conductivity were set such that $0.0001 \leq \sigma \leq 10$ throughout the imaged volume. Error floors of $0.01|Z_{xy}Z_{yx}|^{1/2}$ were adopted. State and adjoint equations were solved on automatically refined meshes using goal-oriented error estimator. At every iteration, four refinement cycles with $\theta = 0.1$ were performed to reduce initial error approximately 10 times (see Section 2.1.4 for more details). The starting model was the homogeneous half-space of 0.1 S m^{-1}.

Fig. 6(a) illustrates final model obtained for a parameter mesh in which cell boundaries coincide with the edges of the true anomalous objects. The quality of the obtained image is very good with both positions and conductivities of the anomalous objects well resolved. The inversion was run using 48 MPI processes, corresponding to three MPI processes per frequency. The parameter mesh for this test

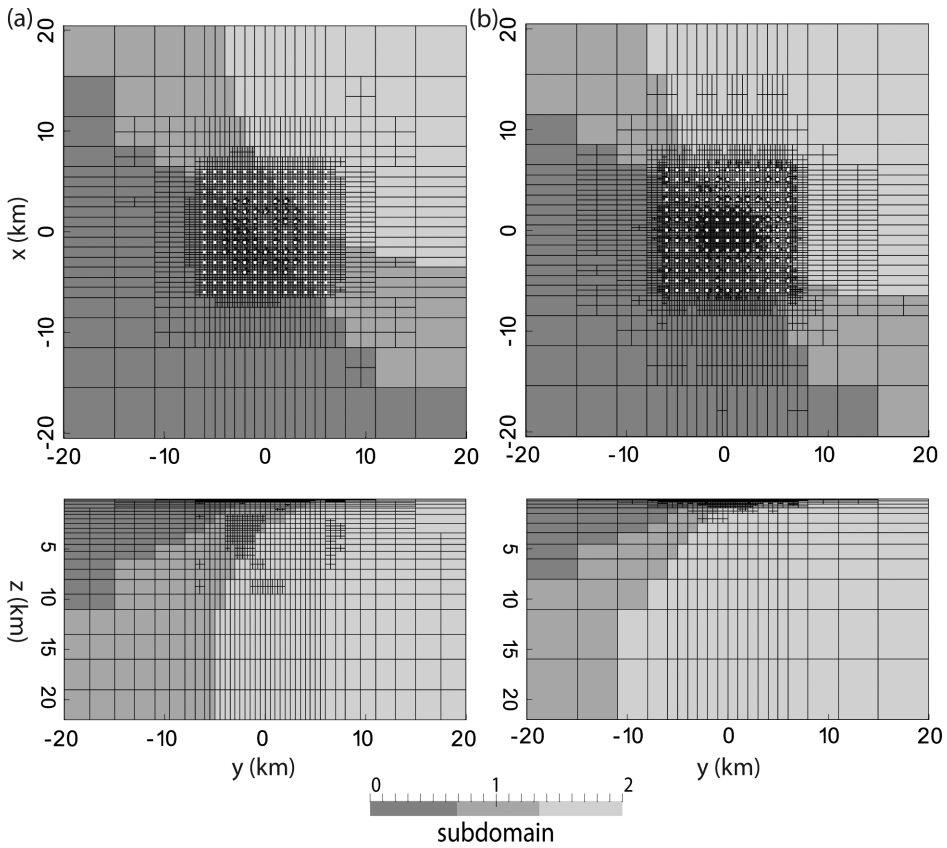

Figure 7. Adaptively refined state meshes for the model shown in Fig. 6(a) for the frequencies of (a) 0.001 and (b) 100 Hz. Colour depicts distributed mesh partitioning. White rectangles denote receiver positions.

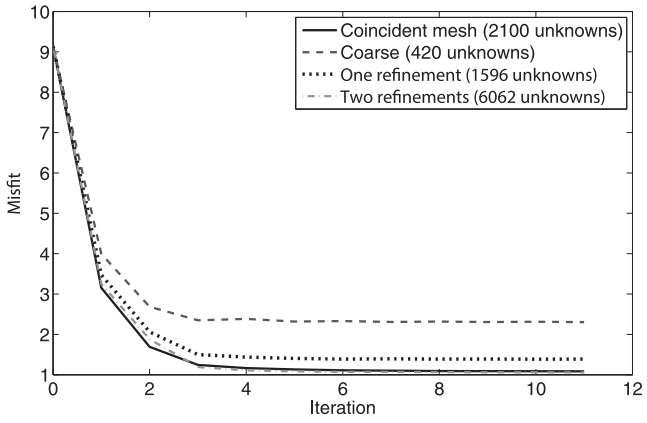

Figure 8. Data misfit versus iteration number for the inversion runs shown in Fig. 6. Note corresponding numbers of unknowns (i.e. active cells in the parameter mesh) in the legend.

$\mathbb{T}^m$ has only 2100 active cells and remained fixed throughout inversion. Refined state and adjoint meshes $\mathbb{T}^s$ contained more cells. In Fig. 7, these meshes for the lowest and highest frequency and the final model from the Fig. 6(a) are shown. They were obtained using automatic adaptive goal-oriented error estimator from Section 2.1.4. These meshes have approximately 70 000 cells each, resulting in a system of eqs (4) with roughly 500 000 unknowns. In Fig. 7, colour indicates different subdomains of the partitioned mesh. For a given frequency, each of the three subdomains was assigned to a single MPI process such that the workload is approximately uniformly distributed for all 16 frequencies. As anticipated, refined meshes differ significantly. Namely, for higher frequency of 100 Hz refinement is mostly concentrated in the near-surface part. This is attributed to the fact that EM fields for higher frequencies attenuate faster and hence deeper parts of the model do not affect EM responses at the receivers.

In the previous test, cell edges coincided with boundaries of the true objects. Even though this mesh is different from the one that was used to generate data and inverse crime is thus avoided, the inversion gets *a priori* information that is generally unavailable when working with real data. In order to avoid this situation, a coarse mesh in which cell edges do not match with object boundaries was taken instead. Fig. 6(b) shows the final model on this mesh with 420 active cells. While deep objects were recovered with some success, the shallower resistor is completely missed. To improve the image quality, a single cycle of the adaptive refinement using resolution estimates obtained with conservatively chosen $l = 150$ was applied to the coarse mesh and the final model on the refined mesh is shown in Fig. 6(c). The refined mesh has 1596 active cells and enables inversion to resolve shallow object reasonably well, although deeper objects are still poorly imaged. Performing one more refinement step produces a mesh with 6062 active cells and Fig. 6(d) shows the final model. Now both the shallow and deep objects are well resolved. While this mesh results in more parameter unknowns than the one from Fig. 6(a), it was obtained automatically, starting from a very coarse mesh without making any assumptions on potential objects' geometry and positions. Fig. 8 summarizes progress of the data misfit over iterations for all four inversion runs. Finer parameter meshes deliver models with better data fit due to additional degrees of freedom.

Table 1 summarizes computational resources required for the inversion shown in Fig. 6(d) with an inversion run on a struc-

Table 1. Computational time and memory required per inversion iteration for a uniformly and adaptively refined meshes.

	Uniformly refined	Adaptively refined
# of parameter cells	98 560	8766
# of active cells	32 760	6062
Memory per core (Gb)	2.2	0.2
Timer per iteration (min)	70	6

tured parameter mesh that is uniformly refined two times in the subsurface. Observed significant reductions in CPU time and memory are due to the difference in parameter space dimensions which influence size of the corresponding state and adjoint problems.

For demonstration of the mesh refinement algorithm, previous inversions were run using the true background as a starting model. However, like the mesh in which cell edges coincide with boundaries of the true objects, the true background is unavailable in practice, too. Therefore, two more tests were performed with starting models being homogeneous half-spaces of 0.2 and 0.01 S m^{-1} for the same data set. As in the previous tests, two adaptive mesh refinement cycles were performed using resolution estimates based on the corresponding starting models. Fig. 9 displays the final conductivity models. For both cases, inversion converged in few iterations. As could be anticipated, when starting from more conductive model, the conductor gets resolved better, whereas starting from more resistive background enables resolving resistive objects better. Overall, quality of the image remained good despite the starting models differed from the true background. More important here is that the mesh refinement algorithm performed well for both cases by delivering meshes which were suitable to resolve subsurface structure.

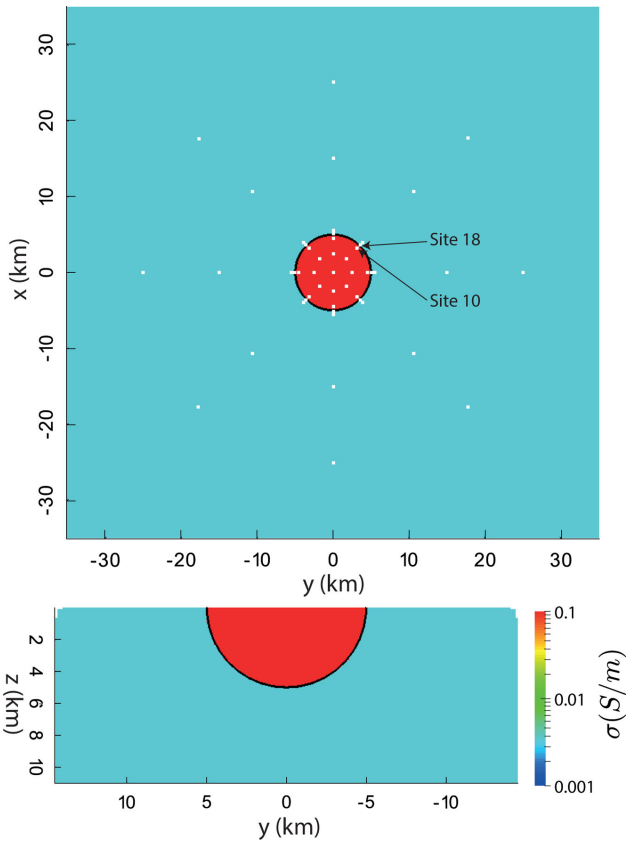

Figure 10. Dublin Test Model 2. A plain view (top) and a section view (bottom) show a 0.1 S m^{-1} hemisphere of the radius 5 km embedded in the 0.003(3) S m^{-1} half-space. Positions of the receivers are shown with white rectangles. Receiver 10 and 18 are indicated with arrows.

4.2 Adaptive mesh refinement

While initial mesh design based on resolution estimates offers a very useful tool in practice, the adaptive mesh refinement may further improve models when working with real-world data. The main reason for this is that initial and true models may have substantially different complexities. Even if an initial mesh was refined using resolution estimates, it might be incapable of capturing all features of the true model sufficiently well. Adaptive parameter mesh refinement strategy derived in Section 2.2.3 helps alleviate this limitation. For demonstration, Dublin Test Model 2 (DTM2) was chosen (Miensopust *et al.* 2013). The true model and receiver stations distribution are shown in Fig. 10. Spherical objects pose a challenge for

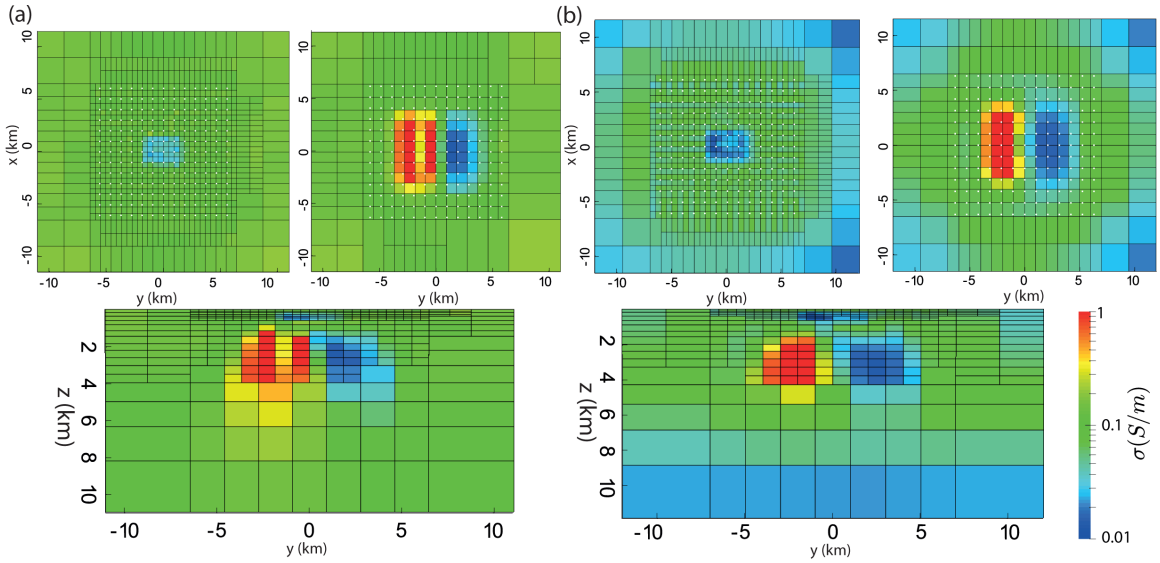

Figure 9. Final models for the meshes which were adaptively refined using resolution estimates. Starting models were homogeneous half-spaces of 0.2 and 0.01 S m^{-1} for (a) and (b), respectively. Horizontal sections are extracted at $z = 350$ and $z = 2500$ m.

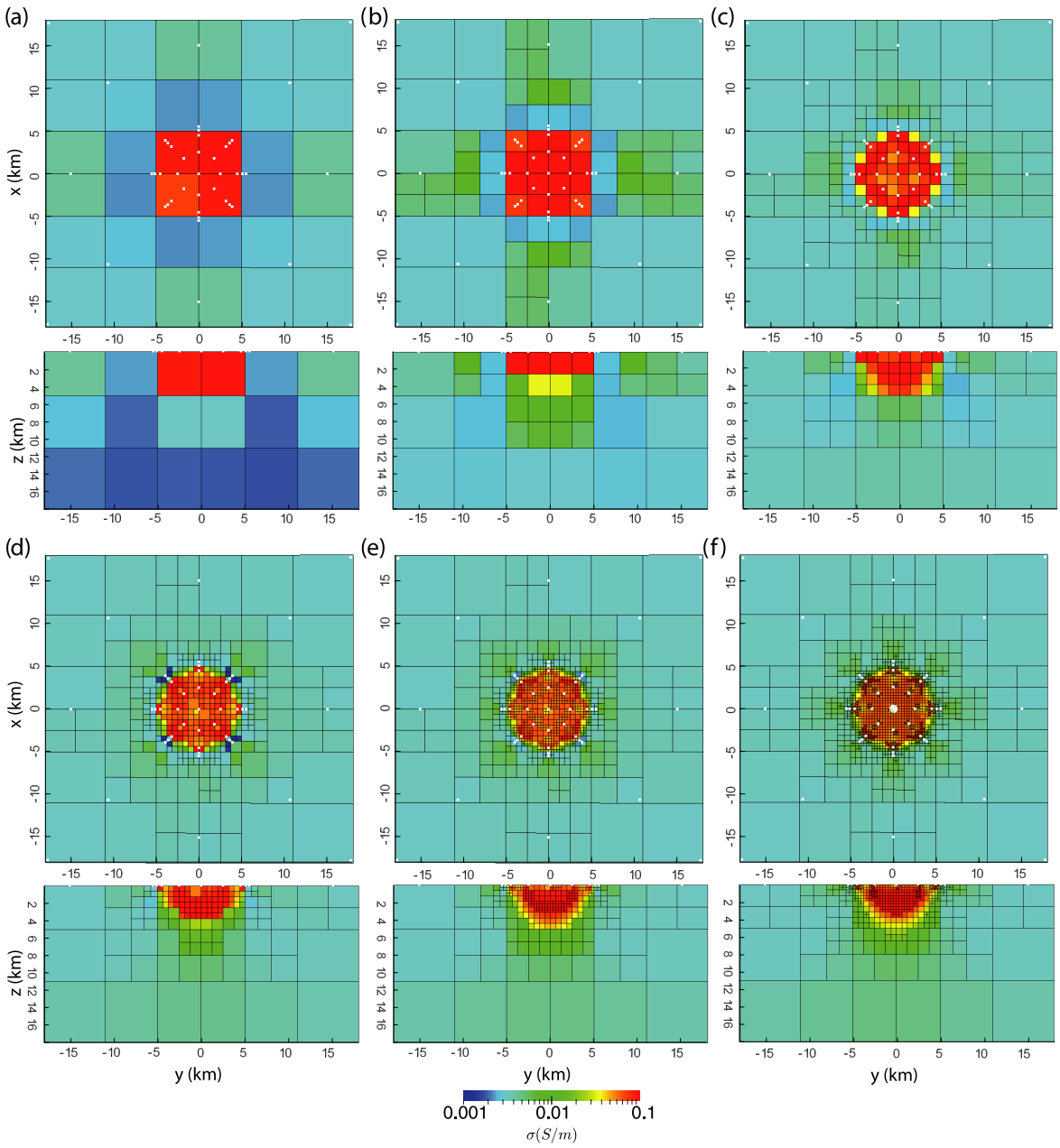

Figure 11. Inversion of DTM2 using automatic adaptive parameter mesh refinement. Plane views at the surface and section views along the centre show conductivity models and the corresponding parameter meshes at the iterations where stagnation in inversion was detected. The numbers of active cells (i.e. unknowns in inversion) are 108, 255, 724, 1837, 4721, 12162 for (a–f), respectively. Positions of receivers are shown with the white rectangles.

conventionally used cell-based model parametrizations. One could use unstructured meshes to represent such geometry very accurately with a moderate number of finite elements, but in inversion it is not known *a priori* which mesh to choose. Therefore, even codes which support unstructured meshes will struggle. The easiest solution is to choose a uniformly fine parameter mesh. However, this results in over-parametrization and increases computational demands as was discussed earlier.

Data for DTM2 model was calculated at 49 receivers (shown with white rectangles in Fig. 10) and 24 periods ranging from 0.01 to 5600 seconds using a refined mesh with 543 332 elements. In total, 9408 real data values were used in inversion. Random two percent Gaussian noise was added to the data. The inversion was started from a homogeneous halfspace of 300 Ωm using a coarse parameter mesh consisting of only 108 active cells. Once it is detected that no further progress can be made on a current parameter mesh, it was refined in the neighbourhood of points where parameter varies most (see Section 2.2.3 for details). For this study, 20 per cent of cells with the largest parameter variations were refined. Conductivity lower and upper bounds were set such that $0.0005 \leq \sigma \leq 1$. State and adjoint equations were solved on automatically refined meshes using goal-oriented error estimator. Four refinement cycles with $\theta = 0.1$ were performed at every iteration to reduce initial error 10 times.

Plots in Figs 11(a–e) show conductivity models at the iterations at which inversion could no longer make any progress and therefore

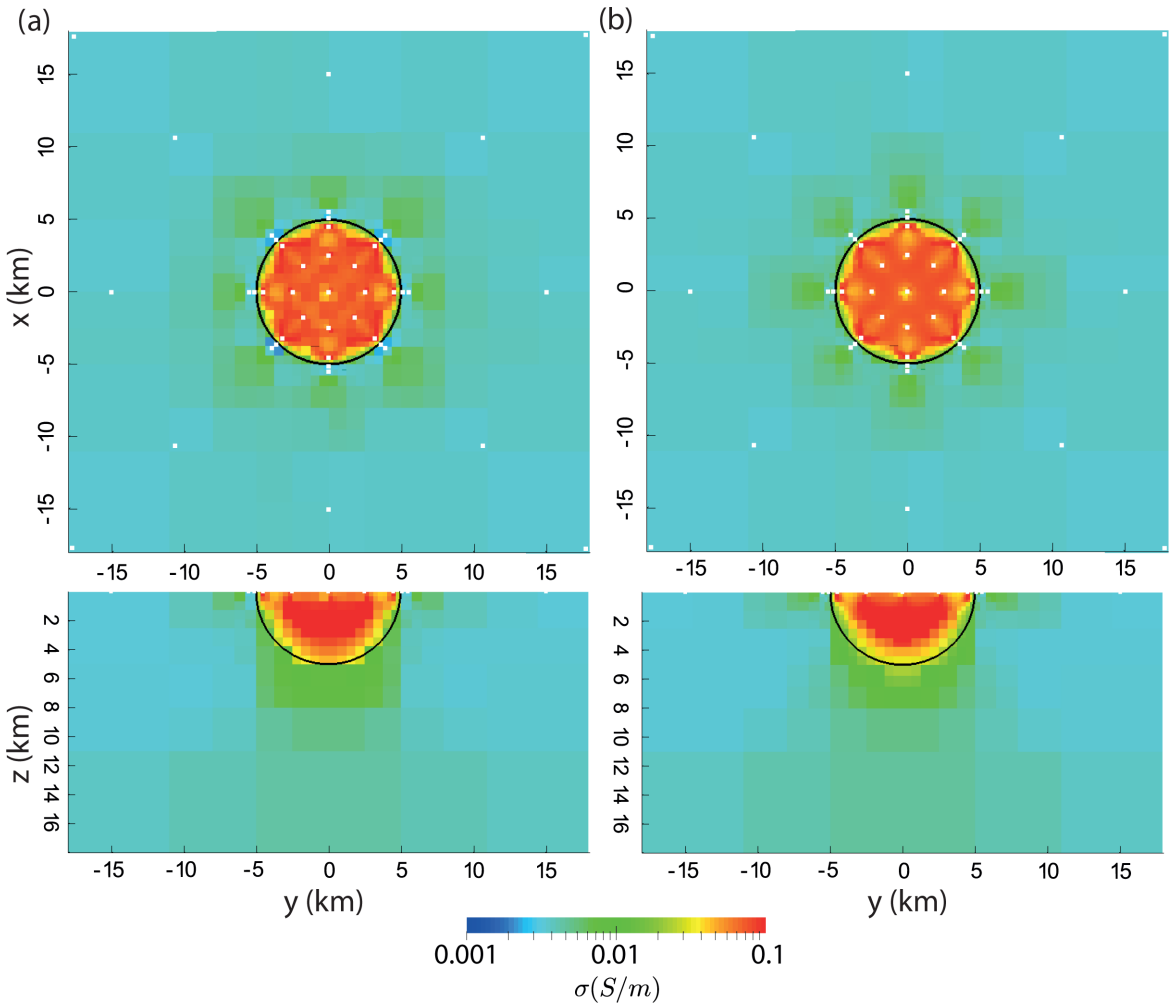

Figure 12. Final conductivity models for the parameter meshes with (a) 4721 and (b) 12 162 active cells, and the rms of 1.2 and 1.16, respectively. Location of the true hemisphere is outlined using the black line. Positions of receivers are shown with the white rectangles.

mesh has been refined. Note that the conductivity model from the Fig. 11(e) corresponds to the 35th iterations and produces an rms of 1.2, but inversion still runs further and after one more refinement a model which delivers rms of 1.16 is obtained (Fig. 11f). These models without grid lines are displayed in Fig. 12. Already for the model with only 4721 active cells, the hemisphere is resolved remarkably well (*cf.* with inversion results in figure 8 in Miensopust *et al.* (2013)). Taking into account that this model was obtained by a fully automatic approach starting from a very coarse mesh with 108 active cells, the adaptive mesh refinement for this inverse problem appeared to be very efficient.

Fig. 13 illustrates data misfit and corresponding number of unknown parameter values (i.e. active cells) versus iteration. Inspection of this plot indicates that as inversion gets stuck, a current mesh is refined and minimization algorithm is then capable to make further progress. Surprisingly, already on the coarsest mesh, numerical scheme reduced data misfit eight times. After that, it took inversion much longer to resolve smaller scale features of the object. Generally, these observations are in a very good agreement with other related studies (Haber *et al.* 2007; Bangerth 2008; Beilina *et al.* 2014; Li *et al.* 2014).

Fig. 14 displays apparent resistivities and phases for all impedance tensor components for two sites indicated with arrows on the Fig. 10 (*cf.* figure 9 in Miensopust *et al.* 2013). Overall, good fit has been achieved for all components and frequencies with slight deterioration for diagonal elements. The latter is anticipated taking into account that diagonal elements are more challenging to fit (Berdichevskii & Dmitriev 2008).

Adding more active cells in inversion results in better resolved small-scale structures. With this, state meshes also contain more cells and generally more unknowns were generated to model EM responses accurately. This is visible in the Fig. 15, which shows numbers of degrees of freedom for the state meshes at each inversion iteration. Since independent adaptively refined meshes are used for each frequency, there is no fixed number. Instead, numbers for all frequencies are plotted with circles. This figure also plots timer per iteration. As can be seen, first iterations proceed much faster since much fewer degrees of freedom were used. During the last iterations, the number of unknowns in state and adjoint problems is around one million. The total runtime for this test was eight hours when using 48 CPU cores.

As can be identified from Fig. 15, number of degrees of freedom may vary significantly for different frequencies. Furthermore, these meshes typically have very different refinement patterns. For instance, Fig. 16 depicts state meshes from 30th inversion iteration after four refinement cycles using goal-oriented error estimator were performed. In contrast to the low frequency mesh, for higher frequency of 100 Hz, regions that are far from

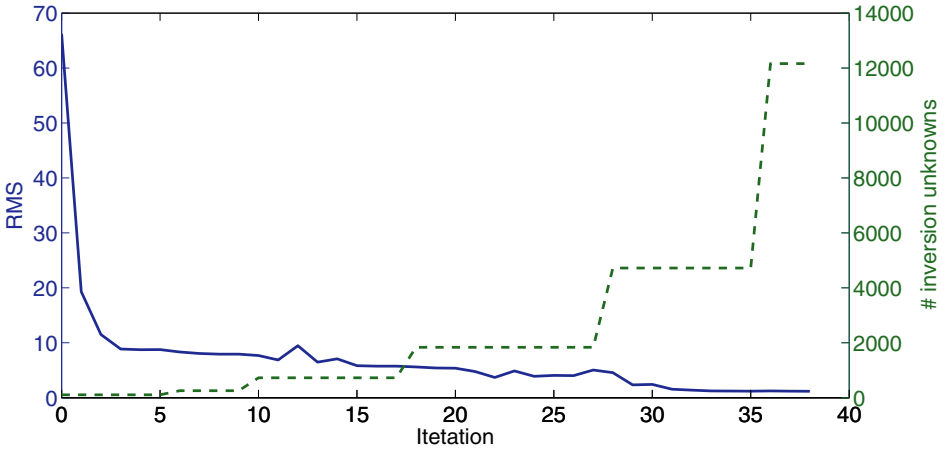

Figure 13. Data misfit (solid line) and number of inversion unknowns (dash-dot line) versus iteration number for DTM2 model.

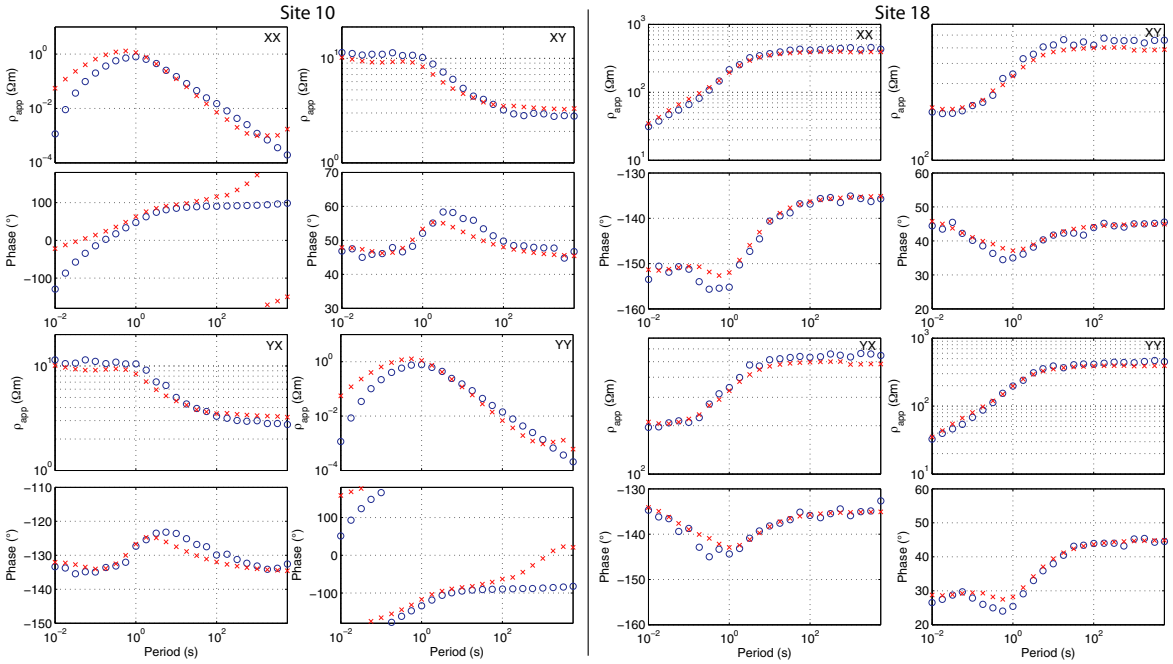

Figure 14. Comparison of the noise-contaminated forward (circles) versus inversion (crosses) responses for DTM2 model at two observation sites indicated in Fig. 10 by arrows.

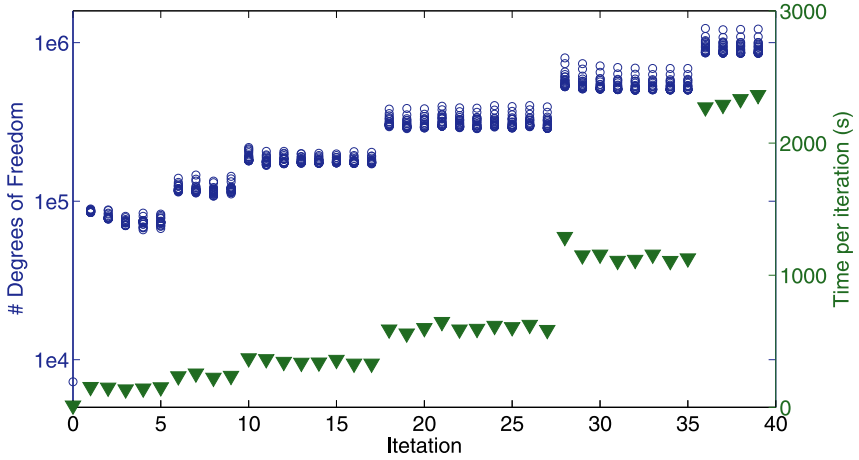

Figure 15. Number of Nédélec degrees of freedom on the adaptively refined state/adjoint meshes at each inversion iteration for DTM2 (circles) and timer per iteration in seconds (triangles). Since independent meshes are used for all frequencies, each circle denotes the value for one of 24 frequencies.

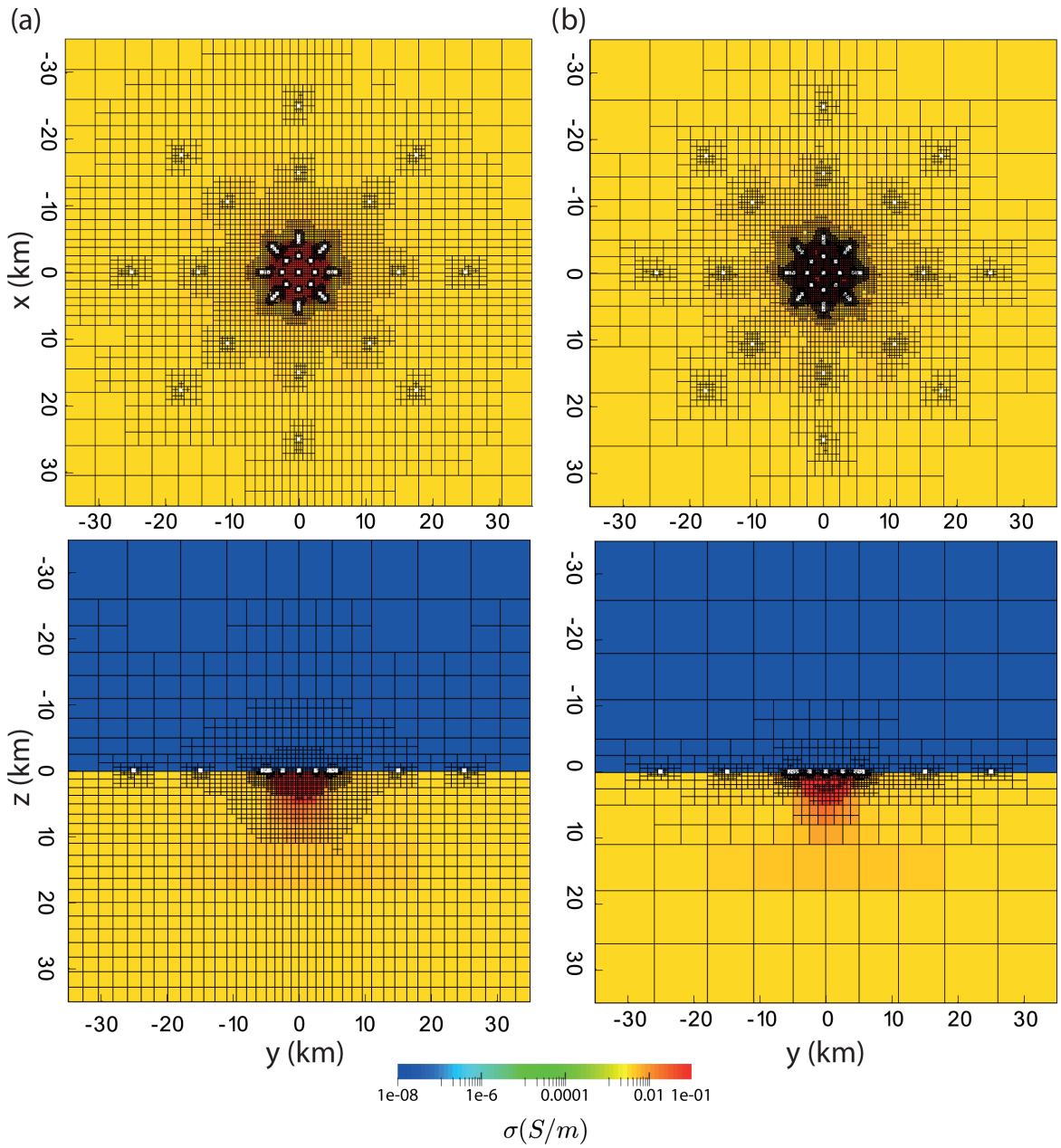

Figure 16. Adaptively refined state meshes at the 30th inversion iteration for the frequencies of (a) 0.001 and (b) 100 Hz. Plane views (top) are shown at the surface. Section views (bottom) are given for $x = 0$. Positions of receivers are shown with the white rectangles.

receivers remained coarse since they do not affect accuracy at receivers due to higher attenuation rates of the corresponding EM fields.

5 CONCLUSIONS

Large-scale geo-electromagnetic inversion based on adaptive FEM has been presented. In contrast to the conventional 3-D inversion workflow where a model parametrization is chosen *a priori*, this study investigated the use of automatic adaptive mesh refinement techniques for state, adjoint and parameter variables. Decoupling state, adjoint and parameter meshes gave an opportunity to choose them independently and use different refinement criteria. This avoids unnecessary fine inversion models without compromising accuracy of the calculated PDE solutions and was realized in a very efficient manner by using hierarchical mesh refinement. To capture different spatial characteristics of the EM fields over a wide range of frequencies, state meshes for all frequencies have been refined independently by using the fully automatic goal-oriented error estimator that was capable to deliver accurate solutions for forward and adjoint problems at low computational cost. The presented experiments show that computational savings can be significant when using adaptively refined meshes. These savings come primarily from opportunity to use smaller state/adjoint problems. While derivative-based optimization methods have little convergence penalty associated with additional parameter variables, performance of stochastic algorithms such as MCMC depends on parameter space dimension a lot. Therefore, applications requiring uncertainty

quantification will additionally benefit from representing parameter variables using coarser meshes.

Adaptive mesh refinement was shown to offer an elegant tool to mitigate—at least to some extent—the problem of a model parametrization. Refinement strategy based on model resolution estimates produces meshes taking into account survey layouts, specifics of transfer functions being inverted, frequency content and data uncertainty. Low-rank approximation of the linearized model resolution matrix was used to derive these estimates. Even with an adaptively constructed initial mesh, some more refinement can be desirable to adjust the mesh according to an emerging subsurface conductivity structure. This was done by using adaptive parameter mesh refinement based on spatial derivative of the updating parameter. Good results were obtained for DTM2 model in which a hemisphere object was well imaged starting from a very coarse mesh. The presented results were found to be in a good agreement with similar approaches from other applications of inverse imaging.

Although this work has covered multiple important aspects, there is a potential to further improvement. For instance, other criteria for initial mesh design which reduce dependence on an initial model or encompass nonlinearity of the inverse operator may appear to have some advantages. This creates a challenge, however, since such criteria need to be computationally tractable for 3-D problems. Further, while the decision to refine mesh by inspecting the spatial derivative of the parameter variable is indeed adequate, it lacks rigorous theoretical background at the moment. In other words, it is not guaranteed that by using this criterion, inversion converges to the right solution. More theoretical work is therefore necessary in the future. Application of this numerical scheme to the real data will certainly serve as a good check for the presented approach. Results on this will be reported in the next part.

ACKNOWLEDGEMENTS

Author thanks Alexey Kuvshinov for multiple constructive discussions about this work and Wolfgang Bangerth for insightful suggestions regarding efficient implementation. Comments and suggestions by the editor G. Egbert, C.G. Farquahrson and an anonymous reviewer helped improve quality of the manuscript.

REFERENCES

Ainsworth, M. & Oden, J.T., 2000. *A Posteriori Error Estimation in Finite Element Analysis,* Vol. 37, John Wiley & Sons.

Alumbaugh, D.L. & Newman, G.A., 2000. Image appraisal for 2-D and 3-D electromagnetic inversion, *Geophysics,* **65**(5), 1455–1467.

Amestoy, P.R., Guermouche, A., L'Excellent, J.-Y. & Pralet, S., 2006. Hybrid scheduling for the parallel solution of linear systems, *Parallel Comput.,* **32**(2), 136–156.

Balay, S. *et al.*, 2013. PETSc Users Manual Revision 3.4.

Bangerth, W., 2002. Adaptive finite element methods for the identification of distributed parameters in partial differential equations, *PhD thesis,* Diss., University of Heidelberg, Heidelberg, Germany.

Bangerth, W., 2008. A framework for the adaptive finite element solution of large-scale inverse problems, *SIAM J. Scientific Comput.,* **30**(6), 2965–2989.

Bangerth, W. & Rannacher, R., 2003. *Adaptive Finite Element Methods for Differential Equations,* Springer-Verlag.

Bangerth, W., Burstedde, C., Heister, T. & Kronbichler, M., 2011. Algorithms and data structures for massively parallel generic adaptive finite element codes, *ACM Trans. Math. Software,* **38**(2), 1–30.

Beilina, L., Thánh, N.T., Klibanov, M.V. & Malmberg, J.B., 2014. Reconstruction of shapes and refractive indices from backscattering experimental data using the adaptivity, *Inverse Prob.,* **30**(10), 105007, doi:10.1088/0266-5611/30/10/105007.

Berdichevskii, M.M.N. & Dmitriev, V.I., 2008. *Models and Methods of Magnetotellurics,* Springer-Verlag.

Biegler, L.T., Ghattas, O., Heinkenschloss, M., Keyes, D. & van Bloemen Waanders, B., 2007. *Real-Time PDE-Constrained Optimization,* Vol. 3, Society for Industrial and Applied Mathematics.

Bui-Thanh, T. & Ghattas, O., 2012. Analysis of the Hessian for inverse scattering problems: I. Inverse shape scattering of acoustic waves, *Inverse Probl.,* **28**(5), 055001, doi:10.1088/0266-5611/28/5/055001.

Bürg, M., 2013. Convergence of an automatic *hp*-Adaptive finite element strategy for Maxwell's equations, *Appl. Numer. Math.,* **72,** 188–206.

Burstedde, C., Wilcox, L.C. & Ghattas, O., 2011. `p4est`: Scalable algorithms for parallel adaptive mesh refinement on forests of octrees, *SIAM J. Sci. Comput.,* **33**(3), 1103–1133.

Chave, A.D. & Jones, A.G., 2012. *The Magnetotelluric Method: Theory and Practice,* Cambridge Univ. Press.

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

Commer, M., Kowalsky, M.B., Doetsch, J., Newman, G.A. & Finsterle, S., 2014. MPiTOUGH2: a parallel parameter estimation framework for hydrological and hydrogeophysical applications, *Comput. Geosci.,* **65**(0), 127–135, TOUGH Symposium 2012.

Day-Lewis, F., Chen, Y. & Singha, K., 2007. Moment inference from tomograms, *Geophys. Res. Lett.,* **34**(22), doi:10.1029/2007GL031621.

Egbert, G.D., 2012. Hybrid conjugate gradient-Occam algorithms for inversion of multifrequency and multitransmitter EM data, *Geophys. J. Int.,* **190**(1), 255–266.

Egbert, G.D. & Kelbert, A., 2012. Computational recipes for electromagnetic inverse problems, *Geophys. J. Int.,* **188,** 251–267.

Ernst, O.G. & Gander, M.J., 2011. Why it is difficult to solve Helmholtz problems with classical iterative methods, *Numer. Anal. Multiscale Prob.,* **83,** 325–361.

Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y. & Villaseñor, A., 2013. Multiscale full waveform inversion, *Geophys. J. Int.,* **194**(1), 534–556.

Grayver, A.V. & Bürg, M., 2014. Robust and scalable 3-D geo-electromagnetic modelling approach using the finite element method, *Geophys. J. Int.,* **198**(1), 110–125.

Grayver, A.V. & Kolev, T.V., 2015. Large-scale 3D geo-electromagnetic modeling using parallel adaptive high-order finite element method, *Geophysics,* in press.

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**(3), 1432–1446.

Grayver, A.V., Streich, R. & Ritter, O., 2014. 3D inversion and resolution analysis of land-based CSEM data from the Ketzin CO2 storage formation, *Geophysics,* **79**(2), E101–E114.

Griesbaum, A., Kaltenbacher, B. & Vexler, B., 2008. Efficient computation of the Tikhonov regularization parameter by goal-oriented adaptive discretization, *Inverse Prob.,* **24**(2), 025025, doi:10.1088/0266-5611/24/2/025025.

Günther, T., Rücker, C. & Spitzer, K., 2006. Three-dimensional modelling and inversion of DC resistivity data incorporating topography—II. Inversion, *Geophys. J. Int.,* **166**(2), 506–517.

Haber, E., 1997. Numerical strategies for the solution of inverse problems, *PhD thesis,* The University of British Columbia, British Columbia, Canada.

Haber, E. & Schwarzbach, C., 2014. Parallel inversion of large-scale airborne time-domain electromagnetic data with multiple OcTree meshes, *Inverse Prob.,* **30**(5), 055011, doi:10.1088/0266-5611/30/5/055011.

Haber, E., Heldmann, S. & Ascher, U., 2007. Adaptive finite volume method for distributed non-smooth parameter identification, *Inverse Prob.,* **23**(4), 1659, doi:10.1088/0266-5611/23/4/017.

Hansen, P.C., 1998. *Rank-Deficient and Discrete Ill-Posed Problems: Numerical Aspects of Linear Inversion,* Society for Industrial and Applied Mathematics.

Hernandez, V., Roman, J.E. & Vidal, V., 2005. SLEPc: a scalable and flexible toolkit for the solution of eigenvalue problems, *ACM Trans. Math. Software,* **31**(3), 351–362.

Hiptmair, R. & Xu, J., 2007. Nodal auxiliary space preconditioning in H(curl) and H(div) spaces, *SIAM J. Numer. Anal.,* **45**(6), 2483–2509.

Kalscheuer, T. & Pedersen, L.B., 2007. A non-linear truncated SVD variance and resolution analysis of two-dimensional magnetotelluric models, *Geophys. J. Int.,* **169**(2), 435–447.

Key, K., 2012. Marine EM inversion using unstructured grids: a 2D parallel adaptive finite element algorithm, in *2012 SEG Annual Meeting,* Society of Exploration Geophysicists.

Key, K. & Ovall, J., 2011. A parallel goal-oriented adaptive finite element method for 2.5-D electromagnetic modelling, *Geophys. J. Int.,* **186**(1), 137–154.

Kim, J.H. & Kim, Y., 2011. A unified transformation function for lower and upper bounding constraints on model parameters in electrical and electromagnetic inversion, *J. Geophys. Eng.,* **8,** 21–26.

Kolev, T.V. & Vassilevski, P., 2009. Parallel auxiliary space AMG for H(curl) problems, *J. Comput. Math.,* **27**(5), 604–623.

Li, M., Abubakar, A., Habashy, T. & Zhang, Y., 2011. Inversion of controlled-source electromagnetic data using a model-based approach, *Geophys. Prospect.,* **58,** 455–467.

Li, T., Isaacson, D., Newell, J.C. & Saulnier, G.J., 2014. Adaptive techniques in electrical impedance tomography reconstruction, *Physiol. Measure.,* **35**(6), 1111–1124.

Lindsey, N.J. & Newman, G.A., 2015. Improved workflow for 3D inverse modeling of magnetotelluric data: examples from five geothermal systems, *Geothermics,* **53,** 527–532.

Menke, W., 1984. *Geophysical Data Analysis: Discrete Inverse Theory,* Academic Press.

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

Monk, P., 2003. *Finite Element Methods for Maxwell's Equations,* Oxford University Press.

Newman, G.A., 2014. A review of high-performance computational strategies for modeling and imaging of electromagnetic induction data, *Surv. Geophys,* **35**(1), 85–100.

Nocedal, J. & Wright, S.J., 1999. *Numerical Optimization,* Springer-Verlag.

Pankratov, O. & Kuvshinov, A., 2010. General formalism for the efficient calculation of derivatives of EM frequency-domain responses and derivatives of the misfit, *Geophys. J. Int.,* **181**(1), 229–249.

Ren, Z., Kalscheuer, T., Greenhalgh, S. & Maurer, H., 2013. A goal-oriented adaptive finite-element approach for plane wave 3-D electromagnetic modelling, *Geophys. J. Int.,* **194**(2), 700–718.

Saad, Y., 2003. *Iterative Methods for Sparse Linear Systems,* 2nd edn, Society for Industrial and Applied Mathematics.

Saad, Y., 2011. *Numerical Methods for Large Eigenvalue Problems,* 2nd edn, Society for Industrial and Applied Mathematics.

Schaefer, J., Boschi, L. & Kissling, E., 2011. Adaptively parametrized surface wave tomography: methodology and a new model of the European upper mantle, *Geophys. J. Int.,* **186**(3), 1431–1453.

Schwarzbach, C., 2009. Stability of finite element solutions to Maxwell's equations in frequency domain, *PhD thesis,* TU Bergakademie Freiberg, Freiberg, Germany.

Schwarzbach, C. & Haber, E., 2013. Finite element based inversion for time-harmonic electromagnetic problems, *Geophys. J. Int.,* **193**(2), 615–634.

Tietze, K. & Ritter, O., 2013. Three-dimensional magnetotelluric inversion in practice—the electrical conductivity structure of the San Andreas Fault in Central California, *Geophys. J. Int.,* **195**(1), 130–147.

Tikhonov, A.N. & Arsenin, V.Y., 1977. *Solutions of Ill-Posed Problems,* V. H. Winston and Sons.

Vasco, D.W., Johnson, L.R. & Marques, O., 2003. Resolution, uncertainty, and whole Earth tomography, *J. geophys. Res.: Solid Earth,* **108**(B1), ESE–9.

Wu, K. & Simon, H., 2000. Thick-restart Lanczos method for large symmetric eigenvalue problems, *SIAM J. Matrix Anal. Appli.,* **22**(2), 602–616.

Zhang, H. & Thurber, C., 2007. Estimating the model resolution matrix for large seismic tomography problems based on Lanczos bidiagonalization with partial reorthogonalization, *Geophys. J. Int.,* **170**(1), 337–345.