A Reduced Order Approach for Probabilistic Inversions of 3D Magnetotelluric Data II: Joint inversion of MT and Surface-Wave Data

María Constanza Manassero[1,1], Juan Carlos Afonso[1,1], Fabio Iván Zyserman[2,2], Sergio Zlotnik[3,3], and Ilya Fomin[1,1]

[1]Macquarie University
[2]Facultad de Ciencias Astronómicas y Geofísicas, Universidad Nacional de La Plata and CONICET
[3]Laboratory of Numerical Analysis, Polytechnic University of Catalonia

November 30, 2022

Abstract

Joint probabilistic inversions of magnetotelluric (MT) and seismic data has great potential for imaging the thermochemical structure of the lithosphere as well as mapping fluid/melt pathways and regions of mantle metasomatism. In this contribution we present a novel probabilistic (Bayesian) joint inversion scheme for 3D MT and surface-wave dispersion data particularly designed for large-scale lithospheric studies. The approach makes use of a recently developed strategy for fast solutions of the 3D MT forward problem (Manassero et al., 2020) and combines it with adaptive Markov chain Monte Carlo (MCMC) algorithms and parallel-in-parallel strategies to achieve extremely efficient simulations. To demonstrate the feasibility, benefits and performance of our joint inversion method to image the conductivity, temperature and velocity structures of the lithosphere, we apply it to two numerical examples of increasing complexity. The inversion approach presented here is timely and will be useful in the joint analysis of MT and surface wave data that are being collected in many parts of the world. This approach also opens up new avenues for the study of translithospheric and transcrustal magmatic systems, the detection of metasomatised mantle and the incorporation of MT into multi-observable inversions for the physical state of the Earth's interior.

A Reduced Order Approach for Probabilistic Inversions of 3D Magnetotelluric Data II: Joint inversion of MT and Surface-Wave Data

M.C. Manassero[1], J.C. Afonso[1,2], F. Zyserman[3], S. Zlotnik[4]and I. Fomin[1]

[1]Australian Research Council Centre of Excellence for Core to Crust Fluid Systems/GEMOC, Department of Earth and Environmental Sciences, Macquarie University, Sydney, Australia.
[2]Centre for Earth Evolution and Dynamics, Department of Geosciences, University of Oslo, Norway.
[3]CONICET, Facultad de Ciencias Astronómicas y Geofísicas, Universidad de La Plata, Argentina.
[4]Laboratori de Càlcul Numèric, Escola Tècnica Superior d'Enginyers de Camins, Canals i Ports, Universitat Politècnica de Catalunya, Barcelona, Spain.

Key Points:

- We present a novel strategy to invert 3D magnetotelluric (MT) data together with other data sets in a fully probabilistic manner.
- We apply our method and perform the first joint probabilistic inversions of 3D MT and surface-wave dispersion data for imaging the electrical conductivity distribution in the lithosphere.
- We demonstrate the capability and applicability of our approach to include 3D MT data into joint probabilistic inversions for the physical state of the interior of the Earth.

Corresponding author: Maria Constanza Manassero, maria-constanza.manassero@mq.edu.au

Abstract

Joint probabilistic inversions of magnetotelluric (MT) and seismic data has great potential for imaging the thermochemical structure of the lithosphere as well as mapping fluid/melt pathways and regions of mantle metasomatism. In this contribution we present a novel probabilistic (Bayesian) joint inversion scheme for 3D MT and surface-wave dispersion data particularly designed for large-scale lithospheric studies. The approach makes use of a recently developed strategy for fast solutions of the 3D MT forward problem (Manassero et al., 2020) and combines it with adaptive Markov chain Monte Carlo (MCMC) algorithms and parallel-in-parallel strategies to achieve extremely efficient simulations. To demonstrate the feasibility, benefits and performance of our joint inversion method to image the conductivity, temperature and velocity structures of the lithosphere, we apply it to two numerical examples of increasing complexity. The inversion approach presented here is timely and will be useful in the joint analysis of MT and surface wave data that are being collected in many parts of the world. This approach also opens up new avenues for the study of translithospheric and transcrustal magmatic systems, the detection of metasomatised mantle and the incorporation of MT into multi-observable inversions for the physical state of the Earth's interior.

1 Introduction

Joint inversions of two or more geophysical data sets are common practice for imaging the Earth's interior and elucidating the physical state of the planet. When the inverted data sets have complementary sensitivities to the properties of interest, joint inversions can significantly reduce the ambiguity inherent in single-dataset inversions, achieve more stable solutions, increase identifiability of features and enhance model resolution. Perhaps more importantly, certain properties of the Earth's interior can only be revealed by combining observations from different techniques. An example is the bulk composition of the lithospheric mantle, which requires independent constrains on the bulk density (e.g. from gravity data sets) and shear-wave velocity (e.g. from surface-wave data). Recent discussions on the benefits and limitations of joint approaches for imaging the structure of the lithosphere and upper mantle can be found in e.g. Khan et al. (2006); J. Afonso et al. (2013a); J. C. Afonso, Moorkamp, & Fullea (2016) and Moorkamp (2017). The joint inversion of magnetotelluric (MT) with seismic data (e.g. Khan et al., 2006; Moorkamp et al., 2007; Gallardo & Meju, 2007; Jegen et al., 2009; Moorkamp et al., 2010; Vozar et al., 2014; Bennington et al., 2015; J. C. Afonso, Rawlinson, et al., 2016; Jones et al., 2017) is of particular interest as they offer complementary sensitivities to temperature, composition and fluid/melt content that are impossible to obtain with other data sets (e.g. Gallardo & Meju, 2007; Moorkamp et al., 2007; Jones et al., 2009; Moorkamp et al., 2010; Selway et al., 2019; J. C. Afonso, Rawlinson, et al., 2016; J. C. Afonso, Moorkamp, & Fullea, 2016). In the context of whole-lithosphere structure, both seismic (or seismic + gravity) and MT data can be used to put constrains on the background (or regional) thermal and mineralogical structure (e.g. Jones et al., 2009; S.-i. Karato & Wang, 2013; J. C. Afonso, Rawlinson, et al., 2016; J. C. Afonso, Moorkamp, & Fullea, 2016), but only MT is strongly sensitive to hydrogen content, minor conductive phases and/or small volumes of fluid or melt (S.-I. Karato, 1990; S.-i. Karato, 2006; R. Evans, 2012; Yoshino, 2010; Khan, 2016; Selway, 2014). Therefore, while both data sets should converge towards a consistent view of the background thermochemical structure, they will diverge in regions where the electrical conductivity of rocks is affected by factors other than temperature or bulk composition. This makes MT-seismic joint inversions a powerful means to detect fluid pathways in the lithosphere, (e.g. Selway & O'Donnell, 2019; R. L. Evans et al., 2019), including the locus of partial melting, ore deposits and hydrated (or metasomatized) lithologies. This unique potential of joint MT-seismic inversions has also given impetus to the acquisition of collocated MT and seismic data over large regions. Concrete examples are the MAGIC and EarthScope USArray in USA (www.usarray.org), the AusLAMP program and AusArray in Australia (www.ga.gov.au/eftf/minerals/nawa),

the IberArray (www.iberarray.ictja.csic.es/) in Europe and the Sinoprobe in China (www.sinoprobe.org). These programs are providing high-quality seismic and MT data with unprecedented resolution and coverage, allowing the pursue of large-scale 3D joint inversions for the physical state of the whole lithosphere and upper mantle.

The actual approach to the joint inversion of MT with seismic data is still a matter of much debate. While traditional deterministic methods are computationally efficient, they are not well prepared to deal with the inherent non-uniqueness of geophysical data sets, and MT data in particular (e.g Wait, 1962; Parker, 1971; Oldenburg, 1979; Mallick & Verma, 1979; Parker, 1980). They are also generally unstable with respect to measurement and/or modeling errors (thus requiring strong regularization) and ill-suited for global uncertainty analysis (e.g. J. C. Afonso, Moorkamp, & Fullea, 2016; Moorkamp, 2017). Probabilistic inversion methods represent an attractive alternative (Tarantola, 2005; Gregory, 2005; Mosegaard & Hansen, 2016) as they are less susceptible to the above-mentioned limitations and provide substantially more information on the parameters of interest via full probability distributions. In probabilistic or Bayesian approaches, the solution to the inverse problem is given by the so-called posterior probability density distribution (PDF) over the model parameter space. This PDF summarizes all the information about the unknown parameters and their uncertainties conditioned on the data and modeling assumptions. As such, it represents the most general solution to the inverse problem. In high-dimensional and/or non-linear problems with complex priors, the posterior PDF cannot be represented analytically and it needs to be sampled point-wise using e.g. Markov chain Monte Carlo (MCMC) algorithms (Mosegaard & Tarantola, 1995; Gilks et al., 1995; Tarantola, 2005; Gregory, 2005). This particular sampling-based approach to probabilistic inversions makes them less efficient than deterministic approaches, as they typically require the numerical solution of millions of forward problems. When the forward problems are computationally expensive, probabilistic approaches can be rendered impractical.

Joint probabilistic inversions of MT and seismic data have been successfully implemented by e.g Khan et al. (2006, 2008); J. Afonso et al. (2013a); J. C. Afonso et al. (2013b); Vozar et al. (2014) and Jones et al. (2017) in the context of 1D MT data only. For the cases of 2D and 3D MT data, however, the large computational cost of the MT forward problem has been the main impediment for pursuing probabilistic inversions, as the number of forward solutions required are typically on the order of $10^5 - 10^7$.

In recent years, various methods and strategies for reducing the cost of full forward solutions have been proposed (see reviews in Frangos et al., 2011; Peherstorfer et al., 2018). The general idea behind these methods is the construction of an approximation, called the *low-fidelity* or *surrogate* model, which can be used instead of, or combined with, the costly full forward or *high-fidelity* solution. Having a faster surrogate of the forward problem is beneficial in a number of contexts, but it is particularly attractive in the context of MCMC schemes used to estimate the posterior PDF in a probabilistic inversion (Christen & Fox, 2005; Cui et al., 2015; Florentin & Díez, 2012; Conrad et al., 2016; Galabert et al., 2019; Manassero et al., 2020; J. Zhang & Taflanidis, 2019). In traditional implementations, the surrogates are computed in an *offline* stage (prior to the probabilistic inversion) at specific locations within the parameter space called 'snapshots'. However, it has been recently shown (Cui et al., 2015; Yan & Zhou, 2019; J. Zhang & Taflanidis, 2019; Galabert et al., 2019; Manassero et al., 2020) that in the context of high- and ultra-high-dimensional probabilistic inversions, it is practically impossible to pre-explore the parameter space in an offline stage to create surrogates that will guarantee accurate solutions within the so far unknown high-probability regions. In these situations, an adaptive MCMC approach where the surrogate is refined *online* during the MCMC simulation is a more effective and efficient approach. A strategy to reduce the computational cost of the 3D MT forward solver and perform full probabilistic 3D MT inversions has recently been presented by Manassero et al. (2020). This novel strategy, called RB+MCMC,

combines i) an efficient parallel-in-parallel structure to solve the 3D MT forward problem, ii) a Reduced Basis Method to create fast and accurate surrogate models of the *high-fidelity* solution, and iii) adaptive strategies for both the MCMC algorithm and the surrogate model.

This paper builds on our previous work (Manassero et al., 2020) and presents the first joint inversion of 3D magnetotelluric and surface-wave data within the context of MCMC-driven, fully probabilistic inversions. Specifically, we focus on a realistic 3D mapping of the electrical conductivity structure of the lithosphere including the locus of deep thermochemical anomalies and fluid pathways. We adopt the RB+MCMC strategy to compute 3D MT surrogate models and propose complementary parameterizations to couple both data sets. Using realistic, whole-lithosphere synthetic models, we demonstrate the benefits and general capabilities of our method for 3D joint probabilistic inversions of MT with surface-wave data in particular, and with other data sets in general.

2 Bayesian Inversion

Within the context of Bayesian inference, the most general solution to the inverse problem is represented by a multi-dimensional probability density function (PDF) over the combined parameter-data space (cf. Tarantola & Valette, 1982; Gilks et al., 1995; Mosegaard et al., 2002; Gregory, 2005; Kaipio & Somersalo, 2006; Mosegaard & Hansen, 2016). This distribution is known as the *posterior* PDF and can be thought of as an objective measure of our best state of knowledge on the problem at hand. It is obtained as a conjunction of the available information on the model parameters ($\mathbf{m}$), the data ($\mathbf{d}$), and their uncertainties. In particular, the marginal posterior PDF over the model parameters, $P(\mathbf{m}|\mathbf{d})$, is formally given by

$$P(\mathbf{m}|\mathbf{d}) \propto \mathcal{L}(\mathbf{m})P(\mathbf{m}). \tag{1}$$

where $P(\mathbf{m})$ is a PDF encoding *a priori* information on the parameter space (what we know or believe about the unknown model parameters prior to considering the actual data) and $\mathcal{L}(\mathbf{m})$ is the so-called *likelihood* function, which describes the probability of obtaining the observed data $\mathbf{d}$ given $\mathbf{m}$. In general, $P(\mathbf{m}|\mathbf{d})$ will be non-linear and high-dimensional (and possibly multi-peaked), with no simple analytical description. When this is the case, unbiased approximations of $P(\mathbf{m}|\mathbf{d})$ are commonly obtained via Markov chain Monte Carlo (MCMC) methods (Gilks et al., 1995; Mosegaard & Tarantola, 1995; Tarantola, 2005; Gregory, 2005). These type of algorithms are designed to output Markov chains that have $P(\mathbf{m}|\mathbf{d})$ as their equilibrium distributions by repeatedly drawing models $\mathbf{m}_t$ and evaluating their posterior probability $P(\mathbf{m}_t|\mathbf{d})$. A large number of MCMC methods have been proposed in the literature, all with relative merits and drawbacks. We refer the reader to the excellent monographs by e.g. Tarantola & Valette (1982); Gilks et al. (1995); Gregory (2005); Calvetti & Somersalo (2007) and Mosegaard & Hansen (2016) for in-depth treatments of Bayesian and MCMC methods applied to inverse problems. In the following, we restrict ourselves to describing only the most relevant theoretical and computational aspects for our purposes.

2.1 The Likelihood Function

The construction of an appropriate likelihood function $\mathcal{L}(\mathbf{m})$ is a critical part of any Bayesian inference problem. $\mathcal{L}(\mathbf{m})$ is typically specified by the distribution of the data uncertainty, which includes both observational and modelization errors. In most cases, observational errors are relatively straightforward to model. Modelization errors, on the other hand, are more complex (and commonly ignored in most geophysical studies) to describe and typically involves exploratory assessments of both numerical errors - e.g. convergence analyses - and Monte Carlo estimates of the correlations between different data sets (see discussions and approaches in Gouveia & Scales, 1998; J. Afonso

et al., 2013a). In the convenient (and most popular) case where both observational and modelization errors can be assumed to be approximately Gaussian, the likelihood function takes the form:

$$\mathcal{L}(\mathbf{m}) \propto \Big(-\frac{1}{2}(\mathbf{g}(\mathbf{m}) - \mathbf{d})^t (\mathbf{C}_d + \mathbf{C}_T)^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d})^t \Big), \tag{2}$$

where $\mathbf{C}_d$ and $\mathbf{C}_T$ are the covariance matrices representing the data and theoretical (model) uncertainties, respectively, and $\mathbf{g}(\mathbf{m})$ denotes the data predicted by the *forward problem* for model $\mathbf{m}$. The term within the parenthesis in Eq. 2 is commonly referred to as the *misfit* of model $\mathbf{m}$.

In the case of joint inversions of uncorrelated observational data sets, the likelihood function can be written as the product of partial likelihoods:

$$\mathcal{L}(\mathbf{m}) = \prod \mathcal{L}_j(\mathbf{m}), \tag{3}$$

where $\mathcal{L}_j$ refers to the likelihood associated with the dataset $\mathbf{d}^j$. The assumption of independent observational data is well justified in most practical situations, an in particular in the MT+seismic case discussed in this paper, as different data sets are commonly gathered in separate surveys using different instrumentation. An important practical advantage of the factorization of the likelihood into partial likelihoods (Eq. 3) is that it makes it possible to adopt a Cascaded Metropolis (CM) approach (Tarantola, 2005; B. Hassani & Renaudin, 2013), which is typically more efficient than a standard Metropolis-Hastings algorithm applied to the total likelihood.

2.2 Cascaded-Metropolis Algorithm

The CM algorithm is particularly useful when the different data sets jointly inverted are uncorrelated, have complementary sensitivities to different aspects of the problem, and at least one of the forward solvers is more computationally demanding than the others. The basic idea is to apply a Metropolis criterion sequentially to each partial posterior (prior $\times$ partial likelihood), which becomes an updated prior in the evaluation of the subsequent partial posterior (e.g. B. Hassani & Renaudin, 2013; B. K. Hassani & Renaudin, 2018). The practical benefits of the above procedure are significant when the partial likelihoods are arranged in order of computational complexity or cost, as there is no need to compute expensive forwards for models that are rejected early in the sequence (see e.g. Tarantola, 2005, for further details).

The basic procedure for the case of two forward operators is as follows: For a new sample $\mathbf{m}_t$, the first partial posterior $P_1(\mathbf{m}_t|\mathbf{d}) = \mathcal{L}_1(\mathbf{m}_t)P(\mathbf{m})$ is always computed using the computationally inexpensive forward. If $P_1(\mathbf{m}_t|\mathbf{d}) > P_1(\mathbf{m}_{t-1}|\mathbf{d})$, this first posterior becomes a prior in the evaluation of the second partial posterior which is now obtained from the expensive forward:

$$P_2(\mathbf{m}_t|\mathbf{d}) = \mathcal{L}_2(\mathbf{m}_t)P_1(\mathbf{m}_t|\mathbf{d}). \tag{4}$$

If $P_1(\mathbf{m}_t|\mathbf{d}) < P_1(\mathbf{m}_{t-1}|\mathbf{d})$, the algorithm randomly decides to evaluate $P_2(\mathbf{m}_t|\mathbf{d})$ or to reject the proposed moved with a probability $P = P_1(\mathbf{m}_t|\mathbf{d})/P_1(\mathbf{m}_{t-1}|\mathbf{d})$ of going to the second step. At the second step, the acceptance of the proposed move is computed as in the standard Metropolis-Hastings algorithm. In this work, $P_1(\mathbf{m}_t|\mathbf{d})$ and $P_2(\mathbf{m}_t|\mathbf{d})$ correspond to the surface-wave dispersion solver and the 3D MT solver, respectively (see details in Section 3).

We will also make use of the Adaptive Metropolis (AM) approach of Haario et al. (2001) to ameliorate the problem of choosing an optimal proposal before the start of the MCMC simulation and to obtain a more efficient sampling strategy of the parameter space that exploits correlations in the model parameters. We leave the presentation of this method to Section 5, where the general sampling strategy is discussed in detail.

3 Forward Problems

3.1 The Magnetotelluric Forward Problem

In this section, we introduce the 3D magnetotelluric (MT) forward problem, the finite-element high-fidelity solver and the RB+MCMC approach to compute surrogate solutions. The reader is referred to Douglas Jr et al. (1999, 2000) and Zyserman & Santos (2000) for an in-depth treatment of the theory behind the formulation of the 3D MT problem and to (Part I; Manassero et al., 2020) for a detailed description of the surrogate approach.

3.1.1 High-fidelity solver for the MT forward problem in 3D

Using the secondary field formulation of Douglas Jr et al. (1999, 2000) and the absorbent boundary conditions defined by Sheen (1997), the MT forward problem in 3D is defined as follows:

Find $\mathbf{E}$ and $\mathbf{H}$ such that

$$\begin{aligned} \sigma\mathbf{E} - \nabla\times\mathbf{H} &= -\mathbf{F} && \text{in } \Omega, && \text{(5a)}\\ i\omega\mu_0\mathbf{H} + \nabla\times\mathbf{E} &= 0 && \text{in } \Omega, && \text{(5b)}\\ (1-i)P_\tau a\mathbf{E} + \nu\times\mathbf{H} &= 0 && \text{on } \partial\Omega\equiv\Gamma, && \text{(5c)} \end{aligned}$$

where $\mathbf{E}$ is the electric field $[V/m]$; $\mathbf{H}$ is the magnetic field $[A/m]$; μ_0 is the magnetic permeability of free space $[Vs/Am]$; σ is the electrical conductivity $[S/m]$ of the medium $\Omega\in\mathbb{R}^3$ and $\Gamma\equiv\partial\Omega$ is the boundary of the domain Ω. a is defined as $a=(\sigma/2\omega\mu_0)^{1/2}$ and $P_\tau\boldsymbol{\varphi} = \boldsymbol{\varphi} - \nu(\nu\cdot\boldsymbol{\varphi})$ is the projection of the trace of any vector $\boldsymbol{\varphi}$ on Γ where ν is the unit outer normal to Γ.

High-fidelity numerical solutions to Eqs. 5 are sought via an optimized version of the finite element (FE) code developed by Zyserman & Santos (2000). In this optimized version, once the variational formulation of Eqs. 5 is discretized in terms of the FE shape functions, Eqs. 5 are converted into the following linear system of equations:

$$\mathbb{K}\mathbf{U} = \mathbf{F}, \tag{6}$$

where $\mathbb{K}^{N_{FE}\times N_{FE}}$ is a sparse and symmetric matrix (the so-called FE *stiffness matrix*) and N_{FE} is the number of degrees of freedom (usually very large). $\mathbf{F}^{N_{FE}\times 1}$ is the force vector and $\mathbf{U}^{N_{FE}\times 1}$ is a vector containing the unknown coefficients for the electric field in the whole domain. In MT, the numerical forward solution for a conductivity model requires the computation of two (typically orthogonal) components of the electromagnetic (EM) fields per frequency. Here, these components are referred to as $\mathbf{U}^{S^i}$ and $\mathbf{U}^{S^i_\perp}$, for a frequency i. Once these solutions are computed, their coefficients and the FE shape functions are used to derive the electric and magnetic fields in the whole domain and at the surface of the Earth (for comparison with the observed data). It is worth noting that although the EM fields that satisfy Eqs. 5 are the actual solution to the forward problem, we will refer to the vector $\mathbf{U}$ (either $\mathbf{U}^{S^i}$ or $\mathbf{U}^{S^i_\perp}$) as the *high-fidelity* solution to the forward problem.

As previously mentioned, the overall cost of computing the high-fidelity solution has been the main limitation preventing probabilistic inversions of 3D MT data. In the following section, we briefly describe the RB+MCMC strategy introduced in our previous paper (Manassero et al., 2020) to obtain fast and accurate approximations of the high-fidelity solutions.

3.1.2 Surrogate solutions: A Reduced Basis + MCMC approach

The RB+MCMC approach combines three main elements i) a Reduced Basis (RB) method to obtain fast approximations of the high-fidelity solution; ii) an MCMC algorithm that drives the sampling of the parameter space and iii) an efficient parallel-in-parallel structure to solve the 3D MT forward problem (for both the surrogate and high-fidelity solvers). The first level of parallelization is defined by frequency, i.e. different processors are in charge of computing the forward solution for different frequencies. The second level of parallelization includes a group of processors linked to each frequency which compute (when needed) the costly high-fidelity solutions using the parallel solver *MUMPS* (Amestoy et al., 2001, 2006).

The general idea behind RB approaches is to seek for surrogate solutions as projections onto a space of small dimensionality, referred to as the *reduced basis*. We generate a reduced basis space $\mathcal{V}_{\mathcal{RB}}$ per frequency and field orientation, with dimension $N_{RB} \ll N_{FE}$ and basis vectors $\mathbf{V}_j$. These *bases* are high-fidelity solutions of Eqs. 6 for specific realizations θ of the conductivity model, $\sigma(\mathbf{x}, \theta)$. In contrast to traditional RB approaches, these bases are not sampled in a pre-inversion stage, but rather during the MCMC inversion. In this way, each $\mathcal{V}_{\mathcal{RB}}$ is automatically updated (enriched) by adding new bases as needed during the evolution of the MCMC chain. This *online* enrichment approach circumvents the need of costly *offline* stages to build the reduced basis and increases the overall efficiency of the method (e.g. Manassero et al., 2020).

In the following, we summarize the main steps of the RB+MCMC procedure. Note that items $(i)-(iv)$ are implemented per frequency i and field orientation (S^i and $S^i_\perp$):

1. If there are bases available from an *offline* stage or from a preliminary probabilistic inversion, we load these bases as the initial basis matrix $\mathbb{V}_{\mathbb{RB}}$. Otherwise, we compute the high-fidelity solution of the starting model of the Markov chain and add it as a column vector in the initial $\mathbb{V}_{\mathbb{RB}}$.
2. For a new sample $\mathbf{m}_t = \sigma(\mathbf{x}, \theta)$, we first seek for a surrogate solution to the forward problem by solving
$$\mathbb{K}_{\mathbb{RB}}(\theta)\mathbf{a} = \mathbf{F_{RB}}(\theta) \tag{7}$$
for the the coefficients $\mathbf{a}(\theta)$; where $\mathbb{K}_{\mathbb{RB}}(\theta)^{N_{RB}\times N_{RB}} = {\mathbb{V}_{\mathbb{RB}}}^T \mathbb{K}(\theta)\mathbb{V}_{\mathbb{RB}}$ is the RB matrix,
$\mathbf{F_{RB}}(\theta)^{N_{RB}\times 1} = {\mathbb{V}_{\mathbb{RB}}}^T \mathbf{F}(\theta)$ is the RB force vector and ${\mathbb{V}_{\mathbb{RB}}}^{N_{FE}\times N_{RB}} = [\mathbf{V}_1, \mathbf{V}_2, ..., \mathbf{V}_{N_{RB}}]$ is the matrix of basis vectors of $\mathcal{V}_{\mathcal{RB}}$. The surrogate solution, $\mathbf{U_{RB}}(\theta)$, is then found as a linear combination of the basis vectors in $\mathcal{V}_{\mathcal{RB}}$ by substituting the coefficients $\mathbf{a}(\theta)$ into the following equation:
$$\mathbf{U_{RB}}(\mathbf{x}, \theta) = \sum_{j=1}^{N_{RB}} a_j(\theta)\mathbf{V}_j = \mathbb{V}_{\mathbb{RB}}\mathbf{a}(\theta). \tag{8}$$
Since the linear system of Eqs. 7 is of size $N_{RB} \ll N_{FE}$, its computational cost is only a small fraction of the time consumed in solving Eqs. 6.
3. The following relative error is computed to asses the accuracy of the surrogate (Quarteroni et al., 2015; Hesthaven et al., 2016):
$$\mathbf{R_{RB}} := \frac{||\mathbb{K}\mathbf{U_{RB}} - \mathbf{F}||}{||\mathbf{F}||}, \tag{9}$$
where $||\cdot||$ is the L_2 norm.
4. The surrogate solution is considered admissible if the $\mathbf{R_{RB}}$ verifies $\mathbf{R_{RB}} \leq \beta$ for a prescribed tolerance β.
5. If all the errors $\mathbf{R_{RB}}$ are smaller than β, we accept ${\mathbf{U_{RB}}}^{S^i}$ and ${\mathbf{U_{RB}}}^{S^i_\perp}$ as good approximations of the high-fidelity solution for all frequencies. In this case, the

corresponding approximate likelihood, $\overline{\mathcal{L}}_2(\mathbf{m}_t)$, is computed and the sample is either accepted or rejected according to the Metropolis-Hastings (MH) criterion.

6. In the case of any $\mathbf{R_{RB}} \gg \beta$, the high-fidelity FE solution for that frequency and component of the EM field is computed for $\mathbf{m}_t$ and added as a new basis vector to enrich the corresponding space $\mathcal{V}_{RB}$. Since the posterior probabilities of the proposed sample $\mathbf{m}_t$ and that of the current sample $\mathbf{m}_{t-1}$ are no longer comparable (i.e. they were computed with different solvers, FE and RB, respectively), we recompute the surrogate solution (and the associated likelihood) at sample $\mathbf{m}_{t-1}$ using the newly enriched RB space. If $\mathbf{m}_t$ is rejected by the MH criterion, a new trial $\mathbf{m}_t^*$ is proposed in the vicinity of $\mathbf{m}_t$ and its likelihood is computed with the newly enriched RB space. This new trial $\mathbf{m}_t^*$ is accepted/rejected according to a modified Metropolis ratio to account for the delayed rejection (i.e. two proposals) step (see e.g. Haario et al., 2006; Mira et al., 2001).

As explained in Manassero et al. (2020), the last step above is required to preserve the ergodicity of the algorithm, but it is not the only possible option. We refer the reader to our previous work (Manassero et al., 2020) for further details on the combined RB+MCMC approach and additional functionalities to improve the efficiency of the method (e.g. use of variable tolerances and Singular Value Decomposition of the basis).

3.2 The Surface-Wave Forward Problem

Surface waves provide one of most valuable data sets to study the lithospheric structure (e.g. Yang et al., 2008; Huang et al., 2009; J. Afonso et al., 2013a). One of the most common approaches involves the generation of dispersion curves or maps and the subsequent inversion of these curves for the velocity structure at depth. Here we compute dispersion curves as functions of 1D vertical velocity structures with a modified version of the forward code *disp96* (Herrmann & Ammon, 2002; J. C. Afonso et al., 2013b; J. C. Afonso, Rawlinson, et al., 2016) . We compute anelastic wave velocities (V_s and V_p) of mantle rocks as (J. C. Afonso et al., 2005, 2008, 2010):

$$V_s = V_{s0}(T,P)[1-(1/2)cot(\alpha\pi/2)Q_s^{-1}(T_o,T,P,d)], \quad (10)$$
$$V_p = V_{p0}(T,P)[1-(2/9)cot(\alpha\pi/2)Q_s^{-1}(T_o,T,P,d)], \quad (11)$$

where V_{s0} and V_{p0} are the unrelaxed, high-frequency (anharmonic) wave velocities at a given temperature (T) and pressure (P) (cf. J. C. Afonso et al., 2010). Without loss of generality, here we compute them as

$$V_{p0} = V_p^{ref} + \frac{\partial V_p}{\partial T}\Delta T + \frac{\partial V_p}{\partial P}\Delta P, \quad (12)$$
$$V_{s0} = V_s^{ref} + \frac{\partial V_s}{\partial T}\Delta T + \frac{\partial V_s}{\partial P}\Delta P, \quad (13)$$

where V_p^{ref} and V_s^{ref} are reference velocities at T_{ref} and P_{ref}; $\Delta T = T - T_{ref}$ and $\Delta P = P - P_{ref}$. The factor Q_s^{-1} is obtained as (Jackson et al., 2002; Jackson & Faul, 2010)

$$Q_s^{-1} = A\Big[\frac{T_o}{d}\exp(\frac{-E+VP}{RT})\Big]^{\alpha}, \quad (14)$$

where T_o is the oscillation period, d is grain size, E is the activation energy, V is the activation volume, α is an empirical exponent, A is a pre-exponential constant and R is the universal gas constant. Although more sophisticated/realistic approaches for computing anelastic seismic velocities are possible (e.g. Matas & Bukowinski, 2007; Khan et al., 2008; J. Afonso et al., 2013a; J. C. Afonso et al., 2013b; Vozar et al., 2014), the set represented by Eqs. 10-14 is sufficient for the goals of this paper.

4 Model Parameterization and Discretization

A key difficulty in the joint inversion of two or more disparate geophysical data sets is how to define the interdependence between model parameters in an internally consistent manner. For instance, if our goal was to jointly invert first arrivals of compressional waves (V_p) and gravity anomalies (a common approach in geophysics), we would need to answer the following question: how is V_p related to bulk density in our medium? A typical assumption in this case is considering a linear correlation between V_p and density (e.g. Birch, 1961, 1964; Feng et al., 1986; Yasar & Erdogan, 2004). While this is a popular and practical assumption, the actual relationship between V_p and density also depends on temperature, pressure and bulk composition (see e.g. J. Afonso et al., 2013a; Guerri et al., 2016). Several authors therefore distinguish between primary and secondary parameters (e.g. Bosch, 1999; Khan et al., 2006; J. Afonso et al., 2013a). The latter are the most commonly used in geophysical inversions and refer to those that enter the governing equations of the forward problems (e.g. V_p, density, electrical conductivity); the former are more fundamental in their nature and thus control the values of the secondary ones (e.g. temperature, porosity, pressure).

In the case of joint inversions of SW and MT data, the primary parameters controlling both the seismic velocities and electrical conductivity (σ) in the mantle are temperature (T), bulk major-element composition (C) and pressure P (e.g Jones et al., 2009; Fullea et al., 2011; R. Evans, 2012; Selway, 2014). Using empirically calibrated equations of state of the type $V_p(T, P, C)$, $V_s(T, P, C)$ and $\sigma(T, P, C)$, and thermodynamic constraints, we can establish direct relationships between the primary and secondary parameters (Bosch, 1999; Xu et al., 2000; Khan et al., 2006; Jones et al., 2009; Yoshino, 2010; Fullea et al., 2011). Since the electrical conductivity is also highly sensitive to hydrogen content, minor conductive constituents and localized melt/fluid pathways, we can explicitly write $\sigma(T, P, C, X)$, where X stands for any factor other than the bulk major-element composition of the rock. This distinction emphasizes the fact that although both seismic velocities and electrical conductivity can constrain the background T-P-C field, the electrical conductivity offers sensitivity to additional factors. The chosen model parameterization should thus be able to accommodate representative variations in both primary parameters (that simultaneously control V_p, V_s and σ) and those responsible for conductivity anomalies above the background values. At the same time, as in any other inverse geophysical problem, the choice of model parameterization needs to be based on the principles of i) flexibility, ii) parsimony, iii) parameter identifiability and iv) suitability for the intended use.

With all of these in mind, and given our particular interest in lithospheric-scale imaging, we focus on a mixed parameterization of the conductivity distribution as the superposition of two contributions: a *background* conductivity related to the long-wavelength thermo-physical state of the lithosphere and an *anomalous* conductivity distribution associated with the presence of features such as fluid pathways, melt-rich regions, hydrogen-rich domains, anomalous mineral assemblages, etc. Following J. Afonso et al. (2013a); J. C. Afonso et al. (2013b), we choose the depth to the lithosphere-asthenosphere boundary (LAB) and the bulk mantle composition as the main model parameters to constraint the background velocity and conductivity structures. We discuss this paramaterization in more detail in Section 4.1. In order to account for smaller-scale conductivity anomalies superimposed on the background, we use a more standard paramaterization based on conductivity nodes. This paramaterization is only relevant to the MT forward problem and it is described in detail in Section 4.2. As shown in the numerical examples of Section 6, the advantage of using this combined parameterization is that a rapid convergence is achieved by using the LAB depths to constrain the first-order conductivity background at the beginning of the inversion. Once this first-order convergence has been achieved, the nodal values are used to locally modify the background to fit the smaller-scale features of the data.

Table 1: Parameters used in the computation of V_{s0} and V_{p0}.

T_{ref}	$800.0°C$
P_{ref}	0 Gpa
$\partial V_p/\partial T$	-5.1×10^{-4} (km/sC)
$\partial V_p/\partial P$	0.110 (km/sGPa)
$\partial V_s/\partial T$	$-3.3 \times 10{-}4$ (km/sC)
$\partial V_s/\partial P$	0.03 (km/sGPa)

4.1 Background parameterization

The 3D numerical model is made up of a collection of M_{col} columns (see Fig 1.b). Each individual column is characterized by its own LAB depth. Here, we identify the LAB with the depth to the $1250°C$ isotherm (cf. J. C. Afonso, Moorkamp, & Fullea, 2016). In order to obtain the background conductivity structure from the LAB structure, we first compute the thermal profile of each column by solving the steady-state heat transfer problem with Dirichlet boundary conditions at the surface (T_0=$10°C$) and bottom of the lithosphere (T_{LAB}=$1250°C$). For simplicity, but without loss of generality, we assume a linear temperature gradient between the LAB and 410 km depth, where the temperature is fixed at T_{410}=$1550°C$. This gradient is extrapolated to the bottom of the numerical domain (460 km). A pressure profile is also computed in each column using the following quadratic lithostatic-type approximation:

$$P(z) = 0.99 \times (4.4773 \times 10^{-3} z^2 + 3.2206 \times 10^4 z - 1.284278 \times 10^8), \tag{15}$$

where P is pressure in Pa and z is depth in meters.

As a further simplification, we assume a dry and homogeneous mantle composition with the following mineral modes: 56, 18.2 ,10.8 and 15 vol% for olivine, orthopyroxene, clinopyroxene and garnet, respectively. While more realistic/sophisticated approaches to map major-element composition into mineral phases should be used when working with real data (e.g. Khan et al., 2006; J. Afonso et al., 2013a; J. C. Afonso et al., 2013b; J. C. Afonso, Rawlinson, et al., 2016; Jones et al., 2017), this simplification does not affect the main results and conclusions of this paper. The electrical conductivity for each mineral phase is obtained using Eq. A3 with parameters specified in Table A1 and the bulk electrical conductivity (i.e. that of the mineral aggregate or rock) of each FE cell in the mantle is computed using the Hashin-–Shtrikman averaging scheme (Hashin & Shtrikman, 1962, 1963). In the numerical examples shown here, the resistivity in the crust (Moho at 49 km depth) is held constant and equal to $20,000\ \Omega m$.

For the surface-wave dispersion problem, each 1D column is further subdivided into 60 layers, each with constant density and wave velocities. The density of each layer is computed as a function of T and P values at the depth of its mid-point as follows:

$$\varrho(P,T) = \varrho_0 + 1 - \alpha(T - T_0) + \beta(P - P_0), \tag{16}$$

with $\varrho_0 = 3355\ kg/m^3$, $T_0 = 10°C$, $P_0 = 0\ Pa$, $\alpha = 3.6 \times 10^{-5}\ 1/°C$ and $\beta = 1.1 \times 10^{-11} 1/Pa$. For a particular layer, the V_p and V_s are obtained using Eqs. 10 and Eq. 14 with the following values: $A_v = 750 s^{-\alpha}\ \mu m^{\alpha}$, $\alpha = 0.26$, $E = 424\ kjmol^{-1}$, $V = 1.3 \times 10^{-5}\ m^3 mol^{-1}$ and grain size $d = 5.0\ \mu m$. Given the periods of interest for surface waves, we adopt $T_o = 50$ s in Eq. 14 (Liu et al., 1976; Lebedev & Van Der Hilst, 2008; Moorkamp et al., 2020). The values for the parameters used in Eqs. 12 and 13 are listed in Table 1 (after J. C. Afonso et al., 2010).

4.2 Node-based parameterization

Any conductivity anomaly that departs from the background is described with N_{nodes} nodes located within the numerical domain. In order to define the nodal locations (Fig. 8), the domain is first sub-divided into horizontal layers of variable thickness. The mid-points of these layers correspond to the nodal depths. Considering that bodies with dimensions smaller than the electromagnetic skin depth cannot be resolved by the MT data, the horizontal distance between different locations within each layer is chosen relative to the skin depth for the range of periods and apparent resistivities shown in the observed data (see for example Figs. 13). The parameters of interest to be retrieved by the inversion are the conductivity values of these nodes. During the probabilistic inversion, the nodal values are interpolated to each FE cell of the numerical domain via kriging interpolation (see e.g. Cressie, 1993; Omre, 1987; Williams & Rasmussen, 1996) using spatially varying correlation lengths (Section B1). Details about the implementation of the interpolation are given in Section Appendix B of the Appendix.

Intuitively, the range of anomalous conductivity values for the nodes should allow for positive and negative perturbations with respect to the background. However, as the electrical conductivity values can span several orders of magnitude, nodal values are typically obtained from proposal distributions defined in logarithmic scale (e.g. Jeffreys and log-normal distributions). Since the domain of the logarithmic function is the set of all positive real values, the sampled anomalous conductivity values (in linear scale) are always positive. In practice, this is not a limitation, as resistive structures (i.e. negative deviations from the background) are generally determined solely by changes in the thermophysical state (e.g. temperature and/or composition changes) whereas anomalous features of interest, such as presence of melt an/or fluid, hydrogen content, grain-boundary graphite films and interconnected sulfides produce positive conductivity anomalies (e.g. Selway, 2014; Hu et al., 2017).

5 Sampling Strategy

The sampling strategy is specifically tailored to take advantage of the differential sensitivities of the SW and MT data sets to the conductivity structure of the lithosphere. With this in mind, we subdivide the MCMC inversion into four main stages. The first stage aims to constrain the background conductivity associated with the first-order temperature structure defined by the LAB depths (if we were interested in inverting for bulk chemical composition, we would also sample this parameter). In the second stage, conductivity anomalies over the background start to be sampled. During these first two stages, we sample both the LAB depths and the conductivity nodes using a *metropolized-independent* sampler. Once enough information (i.e. enough samples) has been acquired for both sets of parameters, we incorporate adaptive strategies to efficiently sample the full parameter space during the third and fourth stages. We briefly describe each of these stages below.

5.1 First stage: focus on background fields

i Randomly select a column in the 3D domain using a *metropolized-independent sampler*.
ii Randomly propose an LAB depth for that column from its proposal distribution.
iii Re-compute the temperature and pressure profiles and update the conductivity and wave velocities ($\mathbf{m}_t$), as explained in Section 4.1.
iv Evaluate the first partial likelihood $P_1(\mathbf{m}_t|\mathbf{d})$ with the SW solver.
v Evaluate $P_2(\mathbf{m}_t|\mathbf{d})$ with probability $P = P_1(\mathbf{m}_t|\mathbf{d})/P_1(\mathbf{m}_{t-1}|\mathbf{d})$ using the MT forward solution:

(a) Seek for a surrogate RB solution to the 3D MT forward problem (Section 3.1.2).

(b) If $\mathbf{R_{RB}} < \beta$ for all frequencies, $\mathbf{m}_t$ is accepted or rejected according to the Metropolis-Hastings criterion.

(c) If any $\mathbf{R_{RB}} > \beta$, the high-fidelity FE solution is computed at $\mathbf{m}_t$. The RB surrogate is recomputed at $\mathbf{m}_{t-1}$ and the algorithm proposes a new move in the vicinity of $\mathbf{m}_t$ whose acceptance is evaluated with a Delayed Rejection criterion (Section 3.1.2).

5.2 Second stage: conductivity nodes begin to be sampled

When the number of MCMC steps reaches a predefined number of simulations (*LAB-stage*):

i Randomly chose a type of parameter to sample (i.e. LAB depths or nodes) at each MCMC step.

ii If chosen parameter = LAB, the algorithm follows the **first stage**.

iii If chosen parameter = conductivity nodes:

(a) Randomly select n_1 nodes at a time, with all nodes having the same probability of being chosen.

(b) Assign a random conductivity value to each node from their individual proposal distributions.

(c) Update the 3D conductivity model via kriging interpolation.

(d) $P_1(\mathbf{m}_t|\mathbf{d})$ remains unchanged, i.e. it only changes when a new LAB value is proposed.

(e) Evaluate $P_2(\mathbf{m}_t|\mathbf{d})$ with the MT solver following items (a)-(c) of the **first stage**.

5.3 Third stage: adaptive strategy for the LAB depths

When the number of MCMC steps reaches a predefined number of simulations (*LAB-adapt*):

i Compute a new multivariate Gaussian proposal distribution (via the Adaptive Metropolis algorithm of Haario et al. (2001)) using the history of the MCMC chains. This proposal now has information about spatial correlations in the LAB.

ii Randomly chose a type of parameter to sample (i.e. LAB depths or nodes) at each MCMC step.

iii If chosen parameter = LAB:

(a) Randomly select m columns at a time, with all columns having the same probability of being chosen.

(b) Propose a new sample for the selected LAB depths using the global multivariate Gaussian proposal.

(c) Follow items (iii)-(v) of the **first stage**.

iv If chosen parameter = conductivity nodes, the algorithm follows items (a)-(f) of the **second stage**.

5.4 Fourth stage: adaptive strategy for the conductivity nodes

When the number of MCMC steps reaches a predefined number of simulations (*nodes-adapt*):

i Compute a multivariate log-normal proposal distribution via the Adaptive Metropolis algorithm using the MCMC chains of all nodes.

ii Randomly chose a type of parameter to sample (i.e. LAB depths or nodes) at each MCMC step.

iii If chosen parameter = LAB, follow item (iii) of the **third stage**.

iv If chosen parameter = conductivity nodes:

(a) Randomly select n_2 nodes with a *metropolized-independent sampler*.

(b) Use the multivariate log-normal distribution to propose new conductivity values for the n_2 random nodes with probability $q(\cdot|\cdot)$ defined in Eq. C2.

(c) Follow items (c)-(f) of the **second stage**.

The first stage only needs a moderate number of models to significantly reduce the original range of possible LAB values. This rapid convergence is due to the strong combined sensitivity of SW and MT to the background field; it also allows the MCMC inversion to focus on the last three stages (i.e. on conductivity anomalies not related to the background T-P-C conditions) while still allowing a continuous improvement of the background field. Additional gain in convergence efficiency is obtained with adaptive sampling strategies applied to both LAB and conductivity nodes. The implementation of these strategies is almost imperative given the high-dimensionality of the problem. While more advanced sampling strategies (e.g, parallel tempering, differential evolution, autoregressive chains) can be implemented to further improve efficiency, we deliberately use this practical (and basic) four-step adaptive strategy to test our joint inversion algorithm under adverse circumstances.

6 Numerical Examples

In this section we consider two numerical examples of joint probabilistic inversion of SW and 3D MT data within the context of whole-lithosphere structure. The synthetic data correspond to two complex large-scale lithospheric models with dimensions 1600$\times$ 1600$\times$460 km (Figs. 1 and 8). In both cases, the computational domain is discretized with $40 \times 40 \times 20$ finite elements.

6.1 Synthetic Data

The MT synthetic data are the off-diagonal apparent resistivities and phases for Example 1 and the full impedance tensor for Example 2. Each dataset is computed for 12 periods between 3.2 and 10^4 seconds at 400 stations. The stations are located on a grid of 20×20 (Fig. 1.a) with an inter-station distance of 80 km. The data errors are assumed to be uncorrelated and normally distributed. In Example 1 we use a standard deviation of 12% for the apparent resistivities and 1.5 degrees for the phases, whereas in Example 2 the standard deviation is assumed as 5% of $\max(|Zxx|, |Zxy|)$ for the components Zxx and Zxy of the impedance tensor, and 5% of $\max(|Zyy|, |Zyx|)$ for the components Zyy and Zyx.

For the case of the SW, the synthetic data are the normal mode Rayleigh wave phase velocities for periods between 15 and 175 seconds, computed at the locations of the MT stations. We assume normally distributed data errors with a representative standard deviation (std) of 20% of the period for Example 1 (that is, $2 * std = 10m/s$ for 25s and $2 * std = 80m/s$ for 200s). For the second example, we consider a standard deviation of 1% of the velocity in meters, which is comparable to the data errors expected for real SW data in dense arrays (Moorkamp et al., 2010; Yang & Forsyth, 2006; Wang et al., 2020).

To minimise the so-called 'inversion crime' (Kaipio & Somersalo, 2006), we compute the actual synthetic data of the first example with a finer FE mesh than that used in the inversion. In the second example, while the models used during the inversion are

obtained via interpolation of the nodes' values, the MT data is generated with the true conductivity value for each FE cell. While this avoids the inversion crime, it also implies that a perfect data fit may not be achievable.

6.2 Data Misfits

The SW and MT misfits, ϕ_{SW} and ϕ_{MT}, are computed as

$$\phi_{SW} = -\frac{1}{2}\sum_{i=1}^{N_{sta}}\sum_{j=1}^{N_{per}}\left(\frac{g_{ij}-d_{ij}}{std_{ij}}\right)^2 \tag{17a}$$

$$\phi_{MT} = -\frac{1}{2\cdot N_{dat}}\sum_{i=1}^{N_{sta}}\sum_{j=1}^{N_{per}}\left(\frac{g_{ij}-d_{ij}}{std_{ij}}\right)^2 \tag{17b}$$

where N_{sta} and N_{per} are the number of stations and periods for each dataset; d_{ij} and g_{ij} correspond to the observed and computed data (with the MT or the SW forward) for station i and period j, and std_{ij} is the standard deviation for data d_{ij}. N_{dat} is the total number of MT data used for each station and frequency, e.g. $N_{dat} = 8$ when we invert for the real and imaginary parts of the full impedance tensor. The factor $1/N_{dat}$ is used here to assure that both data sets have similar absolute contributions in the joint inversion, i.e. that their weights in the overall misfit are not controlled by the number of data points in one of the data sets (e.g. Kalscheuer et al., 2013).

6.3 Example 1: Large-scale Thermal Lithospheric Structure

6.3.1 Model Setup

The inversion area is sub-divided into 18×18 columns (white squares in Fig. 1.b) of size $80 \times 80 \times 460$ km. Each column is comprised of $4 \times 4 \times 20$ FE cells. The model parameters are the depths to the LAB of the 324 columns within the inversion area, i.e. there is one model parameter per column. The true conductivity model is shown in Figs. 1 and is controlled by the subsurface thermal structure. The goal of this example is to assess the identifiability of the background conductivity distribution (via the recovery of the model parameters) from noisy MT and SW data. Accordingly, we only use the LAB parameterization in the first and third stages (Sections 5.1 and 5.3).

6.3.2 Prior and proposal distributions

The priors for the LAB depths are uniform distributions defined in a range of ±70 km, centered on the true value of each column. The proposals used in the first stage of the inversion are Gaussian distributions centered on the current sample with a standard deviation of 20 km. The proposal is adapted in the third stage and therefore it becomes a multivariate Gaussian distribution that reflects the spatial correlations between LAB values of all columns (see Section 5.3). The initial model (i.e. starting point of the MCMC inversion) has a flat LAB located at 180 km depth.

6.3.3 Inversion results

We ran a total of 600,000 MCMC simulations using 2 processors (Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz processors) per frequency and variable RB tolerances of $\beta = 0.07$ for the first 50,000 MCMC steps and $\beta = 0.05$ for the rest of the simulation. Despite the small number of processors used, the joint inversion took only 25 hs, with a staggering average of 0.15 seconds per simulation. This means > 99.5 % gain in computational efficiency compared to the high-fidelity solution ($\sim$ 30 secs). For the same model, and using the same number and type of processors, the RB+MCMC inversion of MT data only (see Manassero et al., 2020) took $\sim$ 30 days (an average of 1.03 seconds per

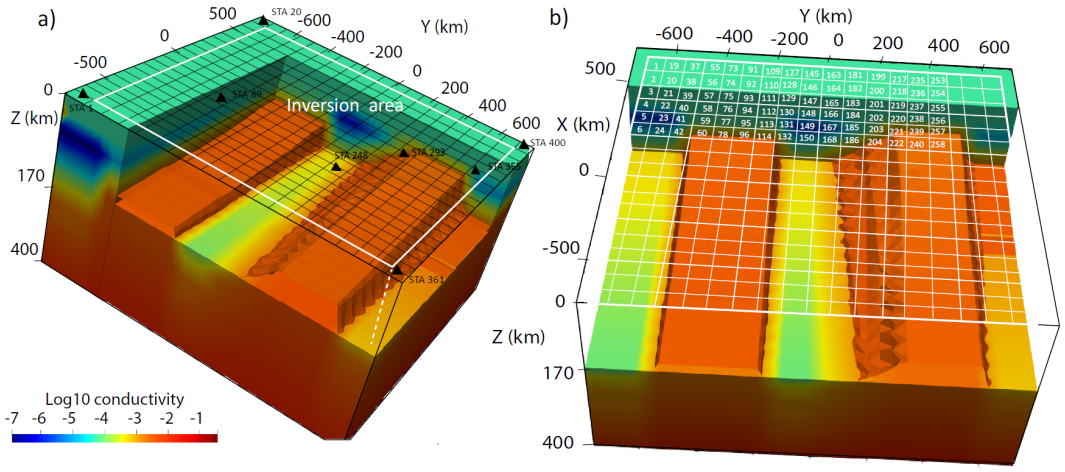

Figure 1. 3D rendering views of the true conductivity structure where the iso-surface of -2.8 log_{10} S/m is plotted as a reference. The white rectangle in (a) indicates the region used for the inversion. Panel (a) illustrates the 20x20 station-grid in black and eight of the 400 stations (black triangles). The model parameters are the depths to the LAB of 324 columns. Panel (b) displays the location of these columns (white small squares) and 96 column-parameters as a reference. The reader is referred to Section 4.1 for details on the parameterization.

MCMC iteration) and convergence was achieved after 2,500,000 MCMC simulations. This dramatic gain in efficiency of the joint inversion is due mainly to i) the implementation of the CM algorithm, ii) the use of adaptive MCMC strategies and iii) the high sensitivity of SW data to the background thermal structure.

The posterior PDFs of 60 of the 324 parameters are shown in Figure 2. The data PDFs for the dispersion curves at two illustrative stations and the data PDFs for MT at one station are shown in Figs. 3 and Figs. 4, respectively. Additional results can be found in the Supplementary material. The results clearly show that the posterior PDFs for all parameters are well behaved (i.e. single valued and approximately Gaussian) and include the true solution, which is always close to the peaks of the PDFs. The resulting uncertainties affecting the LAB values are comparable to those obtained in real inversions (e.g. J. C. Afonso, Moorkamp, & Fullea, 2016; A. Zhang et al., 2019). The data fit is excellent for both data sets (see Figs. 3 and 4).

The *maximum a posteriori* (MAP) and mean models are shown in Figs. 5, together with the 95% confidence intervals of the posterior PDFs. The root-mean-square (rms) values of the *maximum a posteriori* and mean conductivity models, as well as the rms for the LAB structure, are included in Table 2. As a comparison, we have also included the rms values obtained for the same model after the RB+MCMC inversion of 3D MT data only (see Manassero et al., 2020), which are considerable higher than those obtained with the joint inversion.

The evolution of the misfits for MT and SW data is shown in Fig. 6. The number of bases computed per frequency and field orientations are shown in Fig. 7. In all cases, a rapid increment in the basis size is observed during the first 100,000 simulations, which correlates with a rapid decrease in the overall misfits (Fig. 6). This rapid increment in the number of basis is the combination of two factors: i) the starting point of the inversion is far from the high probability region and ii) the initial proposal distribution is not optimal and of large variance. The MCMC algorithm thus samples a wide spectrum of models in its attempt to locate the best paths to the high probability re-

Table 2: Root-mean-square (rms) values of the mean and MAP conductivity and LAB models with respect to the true model. The rms values obtained after the RB+MCMC inversion of 3D MT data only are also included (extracted from Manassero et al., 2020).

	RMS conductivity (log_{10} Ωm)		RMS LAB depth (km)	
	Best Model	Mean Model	Best Model	Mean Model
Joint RB+MCMC	0.08	0.02	6.89	2.21
RB+MCMC	0.19	0.15	21.20	17.01

gions. During this exhaustive exploration, the moves or 'jumps' through the parameter space are large. Consequently, the resulting conductivity models are significantly different from each other and the surrogate needs to be constantly enriched in order to produce accurate solutions for all possible models.

After $\sim$ 150,000 MCMC steps, the basis size reaches a *plateau* (i.e. saturation of the surrogate) for all frequencies and orientations. This means that i) the chain has reached the high probability regions and ii) the RB surrogate is "rich enough" to be able to deliver accurate solutions within these regions (as only a small number of new bases are subsequently required). At this point, we could stop the adaptation or enrichment of the surrogate without compromising the accuracy of the final solution.

These results demonstrate that our RB+MCMC approach successfully solves the joint probabilistic inversion problem and retrieves the first order conductivity structure (and associated uncertainties) from noisy MT and SW data. Moreover, we demonstrate that the addition of the SW data increases the overall efficiency of the algorithm and significantly reduces the range of acceptable conductivity models compared to those obtained from the inversion of MT data only.

6.4 Example 2: Large-scale Lithospheric Structure with Conductivity Anomalies

6.4.1 Model setup

The true conductivity model (Fig. 8) includes the lithospheric model of Section 6.3 as a background (with an additional cut-off for resistivity values higher than 20,000 Ωm) plus three additional and localized conductive anomalies. There are 1155 conductivity nodes (black dots in Fig. 8.a) sparsely located within the inversion volume (1440×1440× 410 km), which is discretized into 324 columns. The vector of model parameters therefore contains 324 LAB values and 1155 nodal conductivity values. The conductivity value of each numerical cell is obtained by adding the background conductivity derived from the LAB structure (Section 4.1) and the anomalous conductivity obtained after interpolation of the nodal values (Section 4.2). The goal here is to assess the identifiability of the true conductivity structure, including background and conductivity anomalies, from noisy 3D MT and SW measurements.

6.4.2 Prior and proposal distributions

The prior and proposal distributions for the LAB parameters are the ones defined in Section 6.3.2. For the conductivity nodes, we use Gaussian prior distributions centered on the background conductivity value (in log-scale) with a standard deviation of 1.5 $log_{10}(S/m)$. This prior information behaves as a regularization term, i.e. it penalizes the introduction of anomalies that are not required by the data. The initial proposal distributions are log-normal (Eq. C1) centered on the current node value m^i_{t-1} with stan-

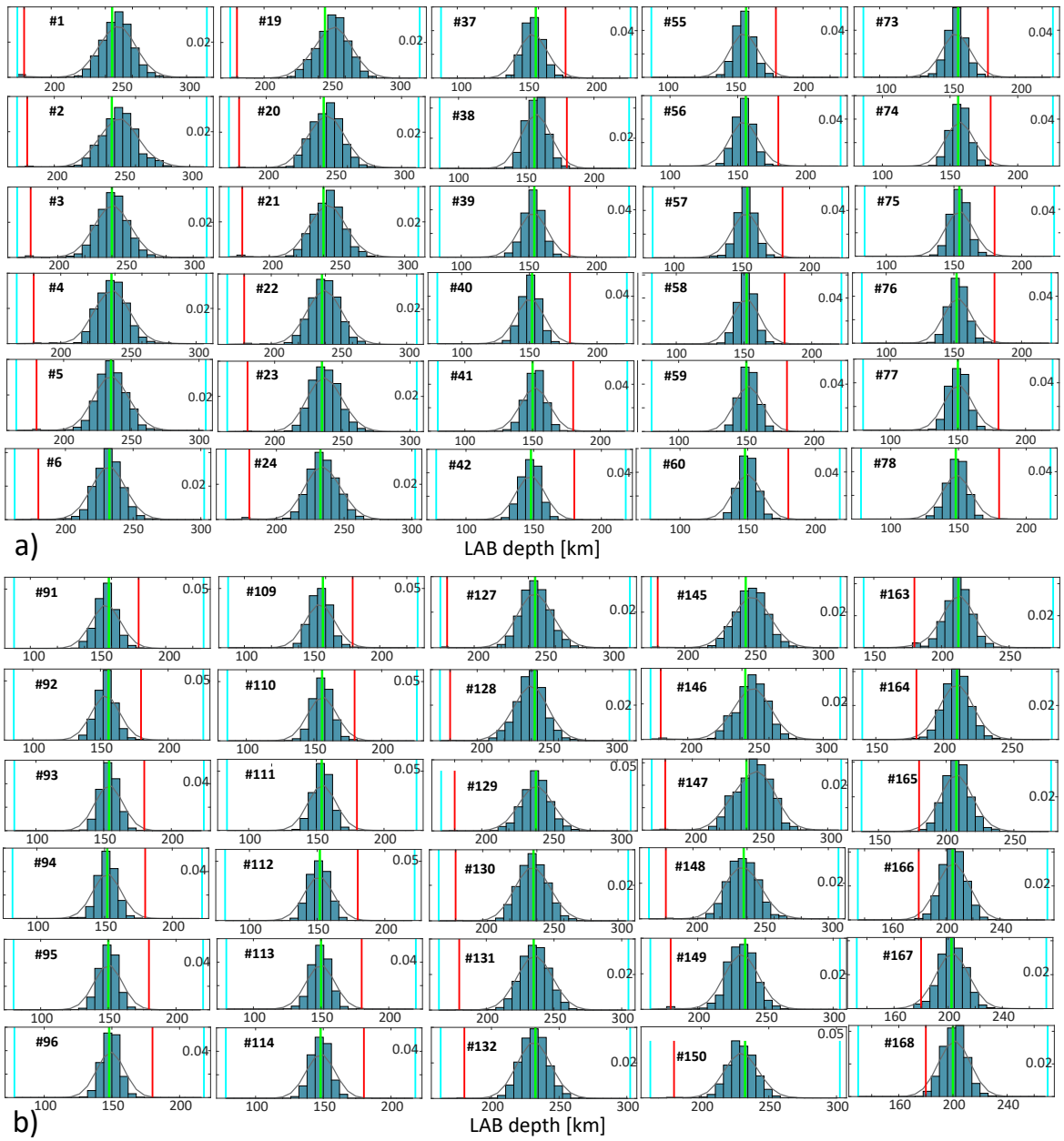

Figure 2. Marginal posterior PDFs (blue bars) of 60 model parameters obtained after 600,000 RB+MCMC simulations. The real value, starting value and prior bounds of each parameter are shown in green, red, and light blue vertical lines, respectively. The best Gaussian fits to the real PDFs given by the histograms are shown in black lines. The numbers within each panel correspond to the columns highlighted in Fig. 1.b.

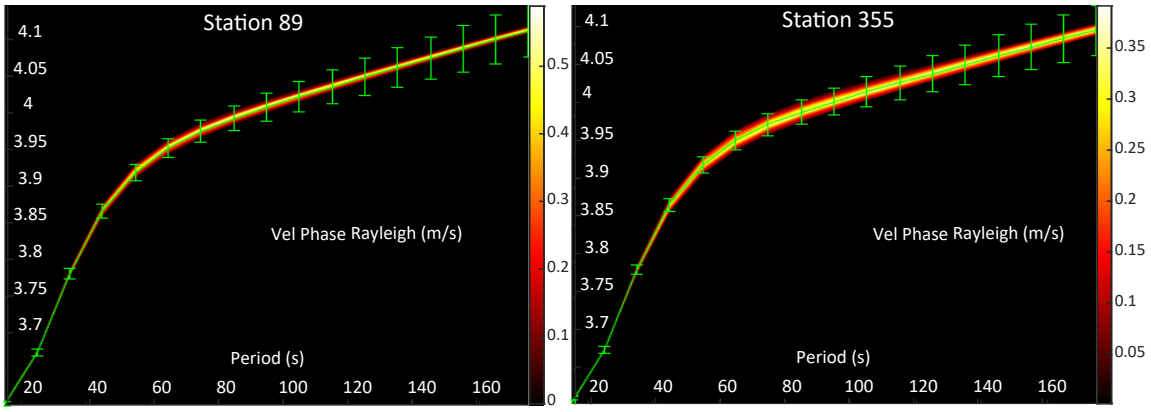

Figure 3. Posterior PDFs of Rayleigh dispersion curves for stations (a) 89 and (b) 355. Synthetic data and error bars are plotted in green. The location of the stations are shown in Fig. 1.a.

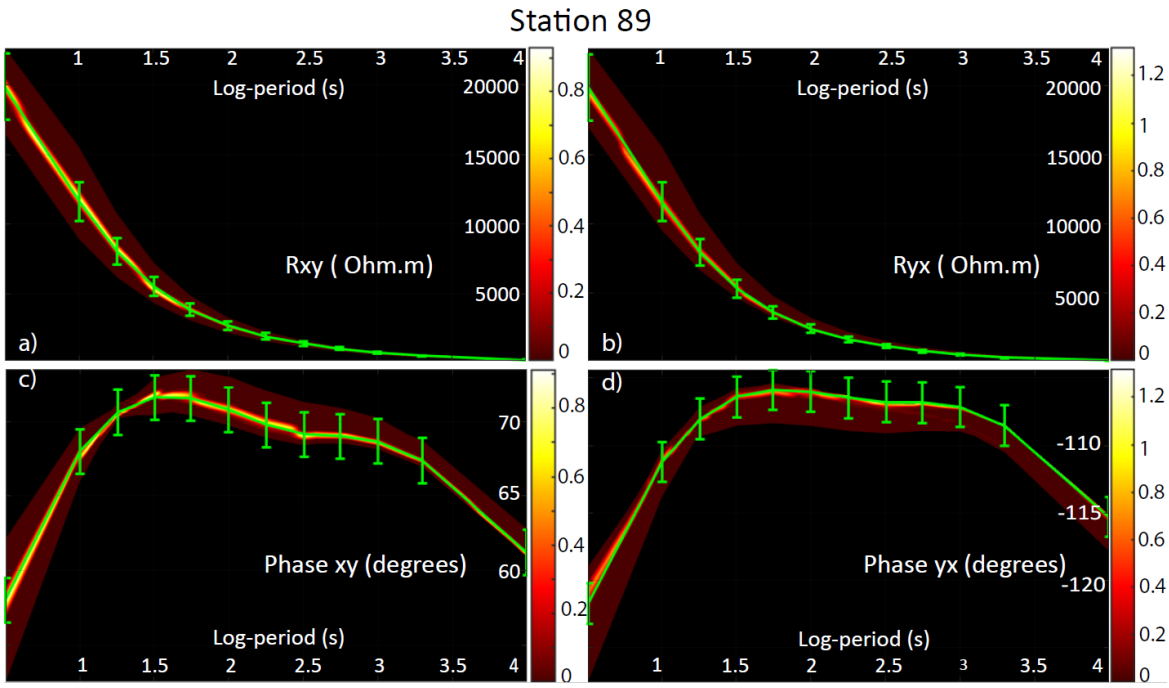

Figure 4. Posterior PDFs of MT data for station 89. Synthetic data and error bars are plotted in green. (a)-(b) Posterior PDFs of the off-diagonal apparent resistivities. (c)-(d) Posterior PDFs of the off-diagonal apparent phases. The location of the stations are shown in Fig. 1.a.

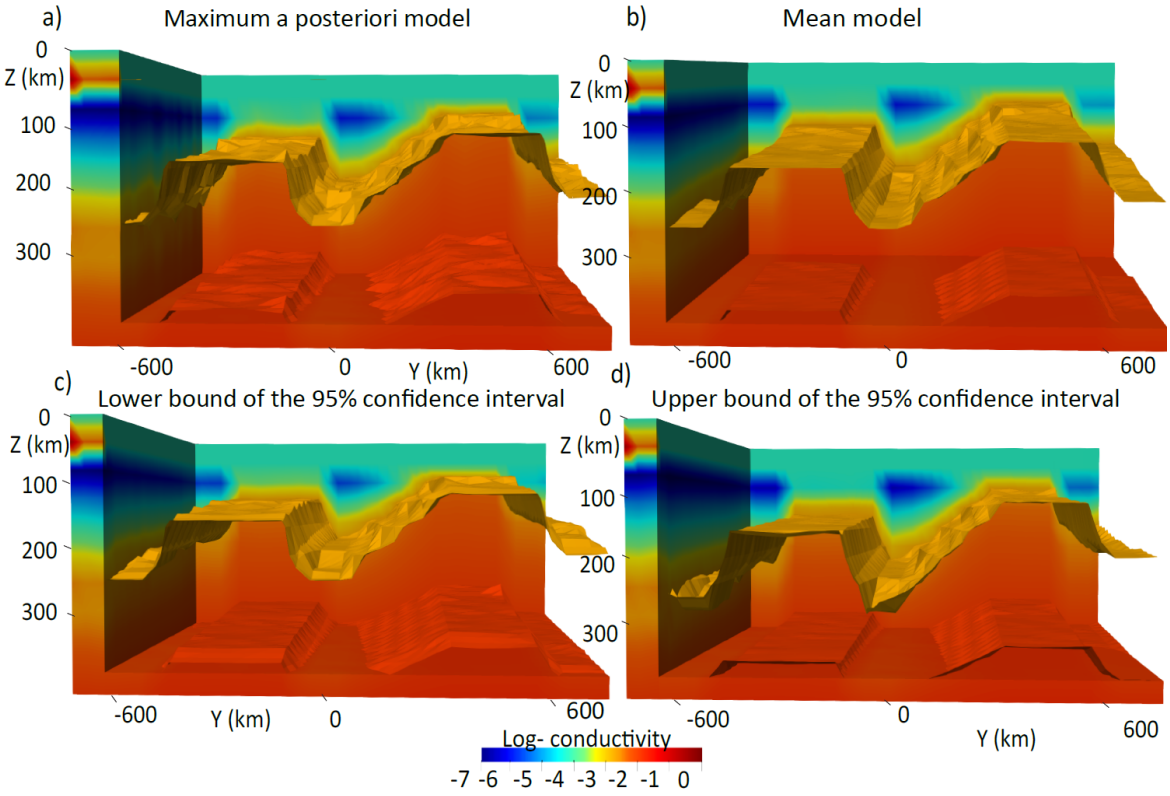

Figure 5. Conductivity structures corresponding to the (a) *maximum a posteriori* (best-fitting) model; (b) mean model; and conductivity models corresponding to the lower (c) and upper bound (d) of the 95% confidence interval of the posterior PDF obtained after 600,000 MCMC steps. The iso-surfaces of -2.8 and -2 log_{10} S/m are plotted as a reference.

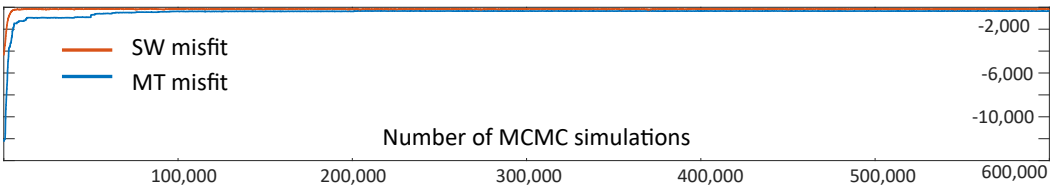

Figure 6. Data misfits for the dispersion curves (red line) and MT (blue line) for each one of the 600,000 RB+MCMC steps.

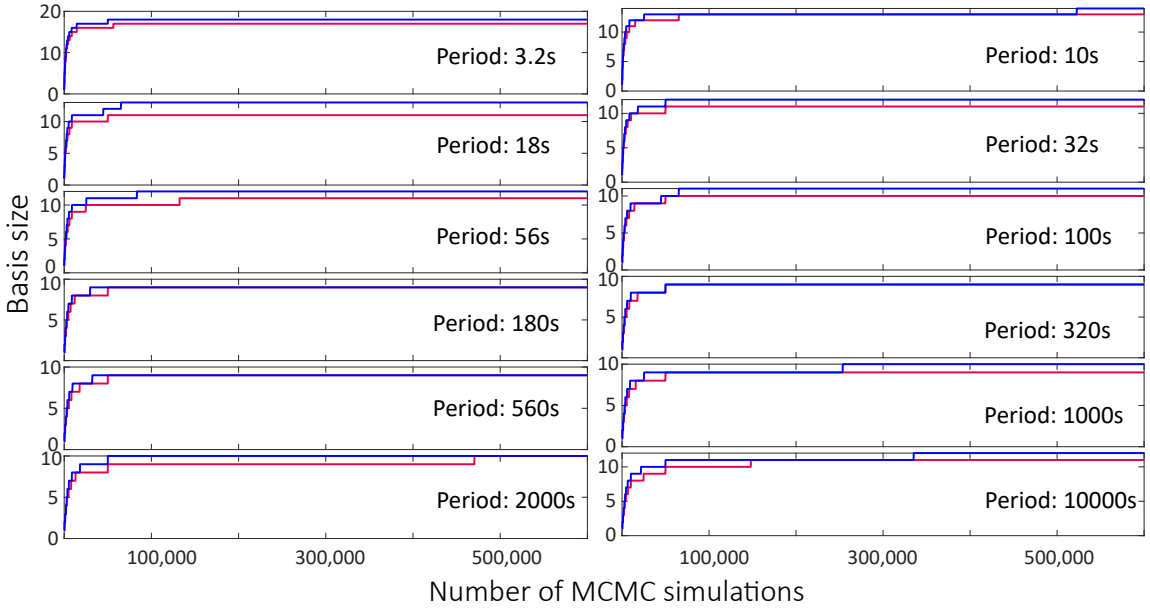

Figure 7. Basis size as a function of the MCMC simulations for different frequencies and field orientations ($S_\perp$ mode in blue and S mode in red).

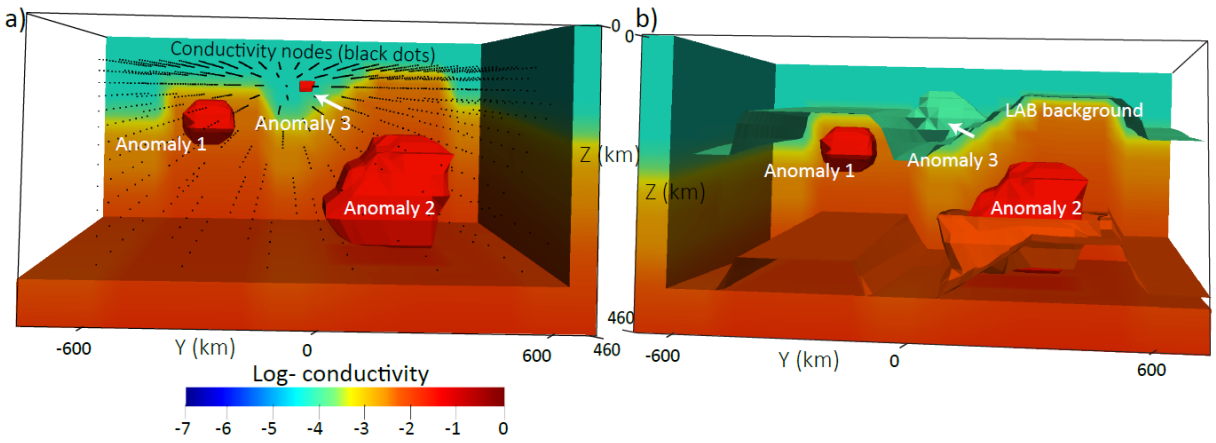

Figure 8. 3D rendering views of the true conductivity structure. Conductivity anomalies are highlighted in both (a) and (b) panels. Black dots in (a) indicate the position of the node-parameters within the inversion volume. Panel (a) shows the iso-surface corresponding to -1.5 log_{10} (S/m), whereas iso-surfaces of -2.15, -1.5 and -4 log_{10} (S/m) are shown in (b).

dard deviations of 0.9 $log_{10}(S/m)$. During the fourth stage, we use an adapted multivariate log-normal distribution centered on the current sample (see Section 5.4). The starting conductivity model is the same as that used in Example 1 and contains no conductivity anomalies.

The first stage was set to 3,000 steps, where we sample LAB depths one column at a time. Once the second stage starts, the algorithm randomly decides to sample the LAB depth of one column or the conductivity values of n_1 =2 nodes. The multivariate proposal for the LAB (start of the third stage) is computed when the chains achieve 250,000 samples and it is adapted every 100,000 LAB samples during the the rest of the inversion. During this third stage, we propose conductivity values of n_1 =2 random nodes or LAB depths of m=6 random columns (from the adapted multivariate proposal distributions; see Section 5.3). The multivariate log-normal proposal distribution for the nodes is computed when their chain reaches 400,000 samples (start of the fourth stage) and it is subsequently adapted every 100,000 steps. During this stage we randomly select n_2 =10 nodes or m=6 columns at a time (see Section 5.4).

6.4.3 Inversion results

We ran a total of 1,000,000 MCMC simulations for 12 frequencies using 2 processors (Intel(R) Xeon(R) CPU E5-2680 v3 @ 2.50GHz) per frequency. The tolerances used were $\beta = 0.068$ for the first 150,000 steps and $\beta = 0.058$ for the remaining of the simulation. Again, even with modest computational resources, the inversion took only 14 days with an average of 1.2 seconds for each simulation. This represents a time reduction of $\sim 96\%$ for each forward computation. The difference in computational time compared to those presented in our previous paper (Manassero et al., 2020) is due to the current implementation of kriging-type interpolation, which is faster and more stable than the Shepard's interpolation. We also note that the average time spent in each simulation is higher than those in Example 1. This is mainly due to the higher number of bases (Fig. 14) required in order to explain the complexity of this 3D model.

The MAP and mean models are shown in Figs. 9 together with the 95% confidence intervals of the posterior PDF. The background conductivity structure and the location and volume of the conductivity anomalies are well resolved. Depth slices from the 95% confidence intervals, MAP and mean models are shown in Figs. 10. In this figure we also include depth slices from five random subsets from the posterior, each obtained as the mean of 10 randomly chosen models form the entire collection of conductivity models making up the posterior PDF. As expected, features that are well resolved by the inversion are persistent in all subsets, whereas poorly resolved features show higher variability among subsets. The identifiability of the background structure is also illustrated in Figs. 11 and Figs. S4-S10 of the Supplementary Material, where we show that the true LAB depths are close to the mean value of the marginal posterior PDFs for all parameters. The sizes of the basis per frequency and the SW-MT data misfits for each of the 1,000,000 steps are shown in Figs. 14 and 15, respectively, and show a similar pattern to those in Fig. 6.

It is worth noting that contrary to what we would expect from an inversion of MT data alone (see results in e.g. Manassero et al., 2020; Rosas-Carbajal et al., 2013), model variability decreases with depth. The reason for this is the tighter constrains that the SW data puts on the background thermochemical structure. Compared with the results presented in Manassero et al. (2020), the joint inversion of MT with SW data highly reduces model uncertainty and increases model resolution.

Examples of the posterior PDFs of SW and MT data are shown in Figs. 12 and 13; additional posterior PDFs are shown in the Supplementary Material. All of the dispersion data points are contained within one standard deviation of the posterior PDFs. This is also the case for the great majority of the MT data, although a poor data fit (or

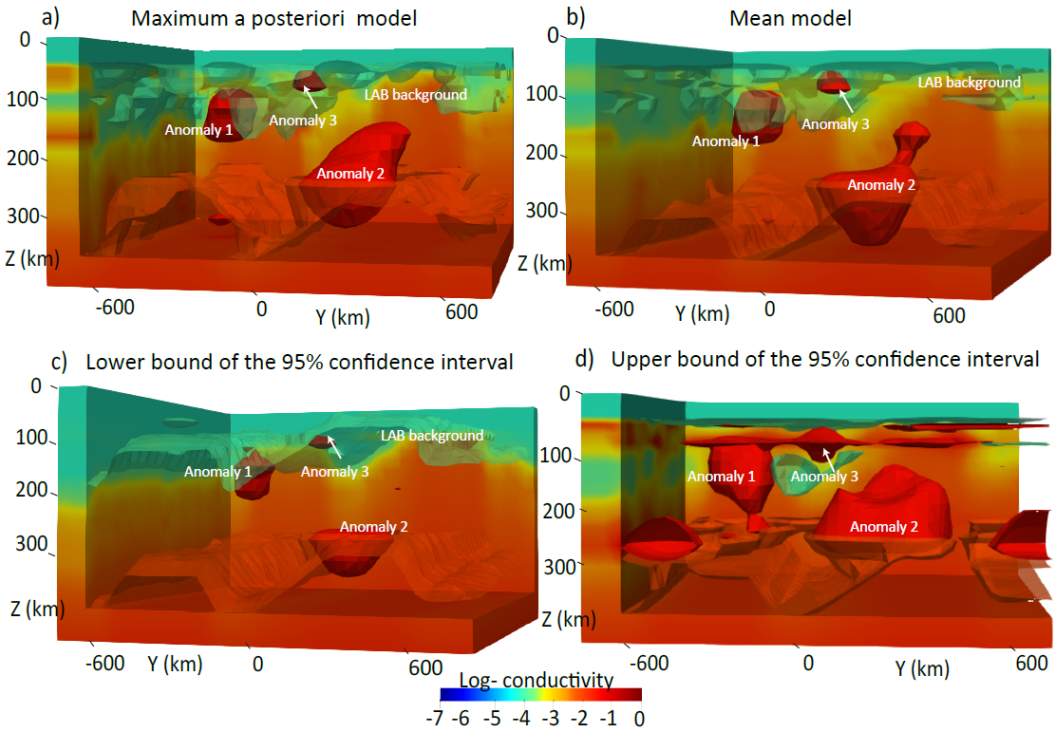

Figure 9. 3D rendering views of the conductivity structure corresponding to the (a) *maximum a posteriori* model; (b) mean model; (c) lower and (d) upper bound of the 95% confidence interval of the posterior PDF obtained after 1,000,000 MCMC steps. The iso-surfaces of -1.5, -2.15 and -4 log_{10} S/m are plotted as a reference. The background structure and the conductivity anomalies are highlighted in all panels.

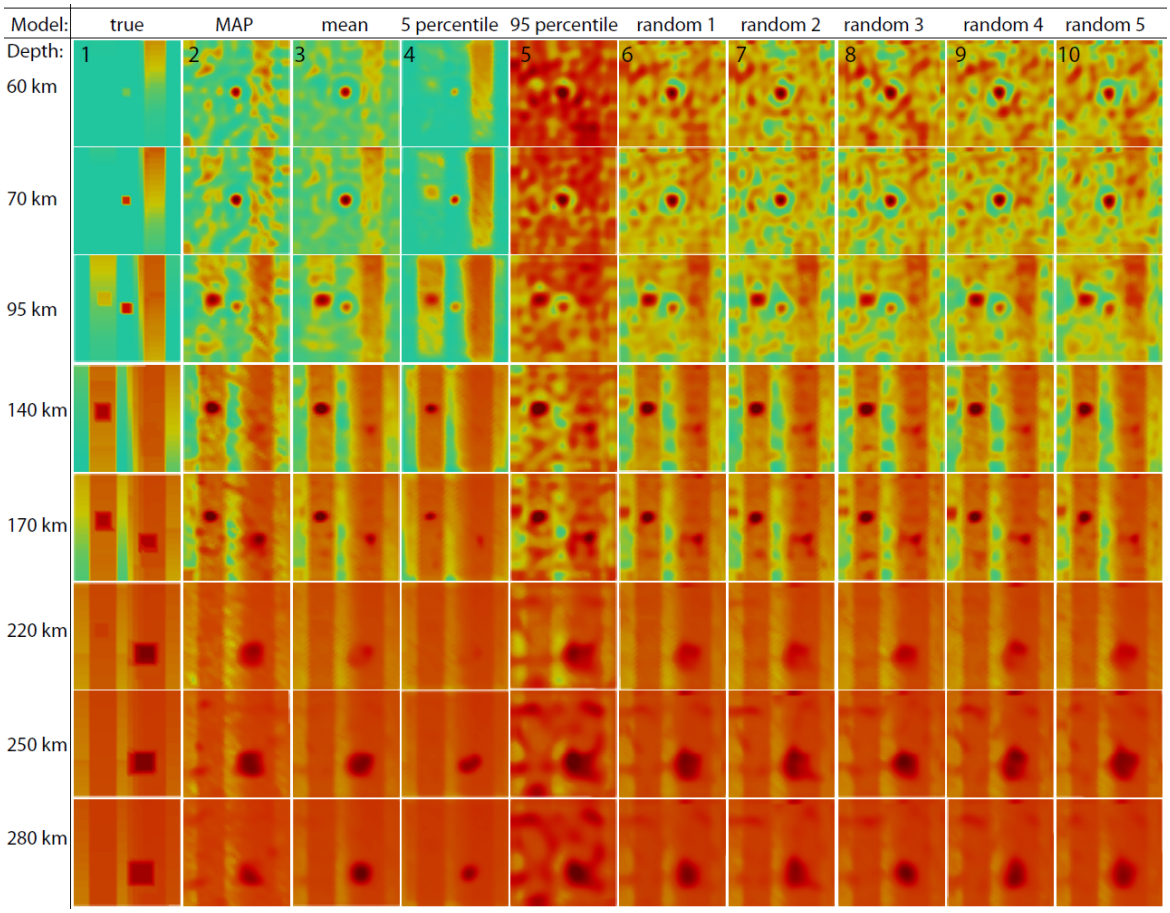

Figure 10. Columns (1)-(5): depth slices from the (1) true model; (2) MAP model, (3) mean and conductivity models corresponding to (4) the 5% percentile and (5) the 95% percentile of the posterior PDF. Columns (6)-(10): depth slices for five mean models computed with 10 random samples of the posterior PDF. Selected depths are shown on the left of the figure.

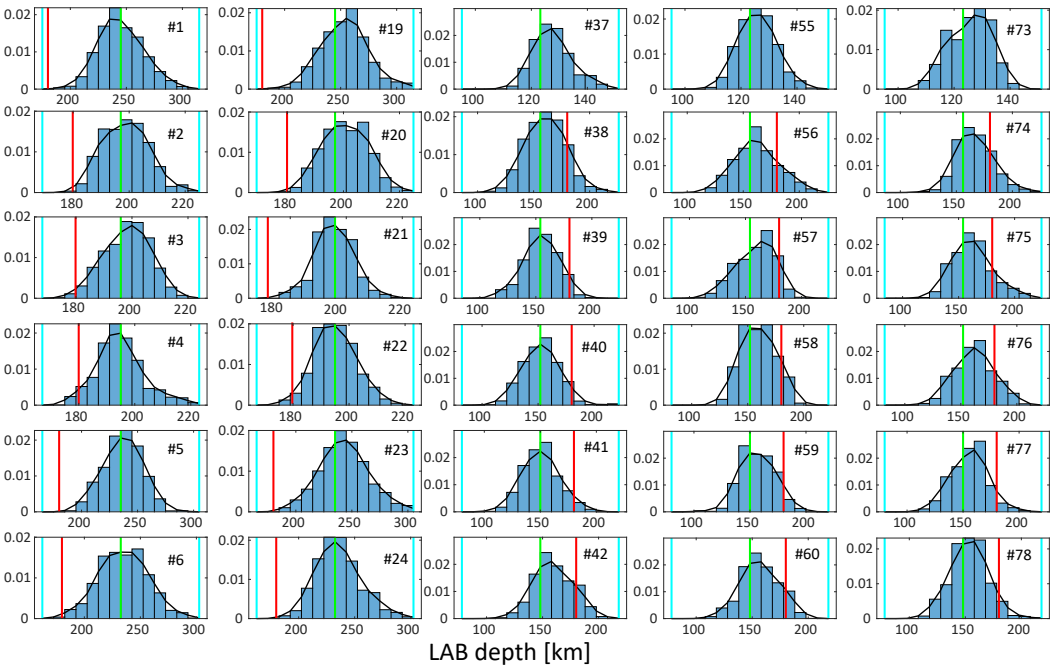

Figure 11. Marginal posterior PDFs (blue bars) of 30 LAB depths obtained after 1,000,000 RB+MCMC steps. The real value, starting value and prior bounds of each parameter are shown in green, red, and light blue vertical lines, respectively. The best Gaussian fits to the real PDFs given by the histograms are shown in black lines. The numbers within each panel correspond to the columns highlighted in Fig. 1.b.

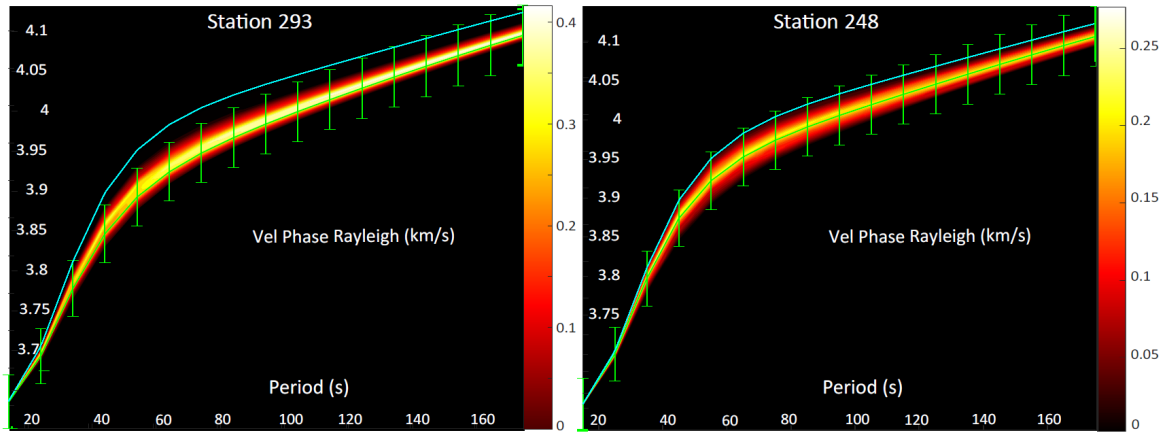

Figure 12. Posterior PDFs of Rayleigh dispersion curves for stations (a) 293 and (b) 248. Synthetic data and error bars are plotted in green and the computed data for the initial model is plotted in blue.. The location of the stations are shown in Fig. 1.a.

bias) is observed in some stations. As mentioned in Section 6.1, the MT synthetic data is computed with the true conductivity model (Fig. 8), whereas the conductivity models used in the actual inversion are derived from the interpolation of nodal values. This discrepancy or inadequacy between models is the main reason of the poorer data fit seen at some stations (e.g. Smith, 2013).

The results from this example demonstrate that the joint probabilistic inversion of wave dispersion and MT data i) is a practical option with modest computational resources, ii) succeeded in identifying the true LAB and conductivity structures (background plus anomalies) and iii) produced well behaved posterior distributions and global measures of uncertainty and correlations between model parameters. In addition, when compared to the results in Manassero et al. (2020) for the inversion of MT alone, the current joint inversion provides better resolution to both the background and anomalies.

7 Discussion

Here we discuss some practical aspects relevant to real inversions and possible technical improvements of the algorithm.

7.1 Inverting for Crustal Structure and Bulk Mantle Composition

Since our main interest is the deep lithospheric structure, we have so far assumed constant properties and thickness for the crust. This is similar to assuming that we have good prior information, for example, from a previous deterministic inversion or from previous independent studies. A similar idea was applied to a real joint inversion by Jones et al. (2017). However, if the interest is also in the shallow crust, we could simply discretize the crust with layers of constant properties (e.g. bulk density, Vs, Vp and σ) within each 1D column and include them in the vector of model parameters. Perhaps a more efficient possibility is to define crustal nodes with associated σ, density, Vs, and Vp and interpolate these values to each numerical cell within the crust, similar to what we do in the mantle for σ. We are currently working on these schemes and results will be presented in a forthcoming publication.

Throughout this work we have also considered a constant major-element composition in the mantle. In practical applications with emphasis on the general structure of

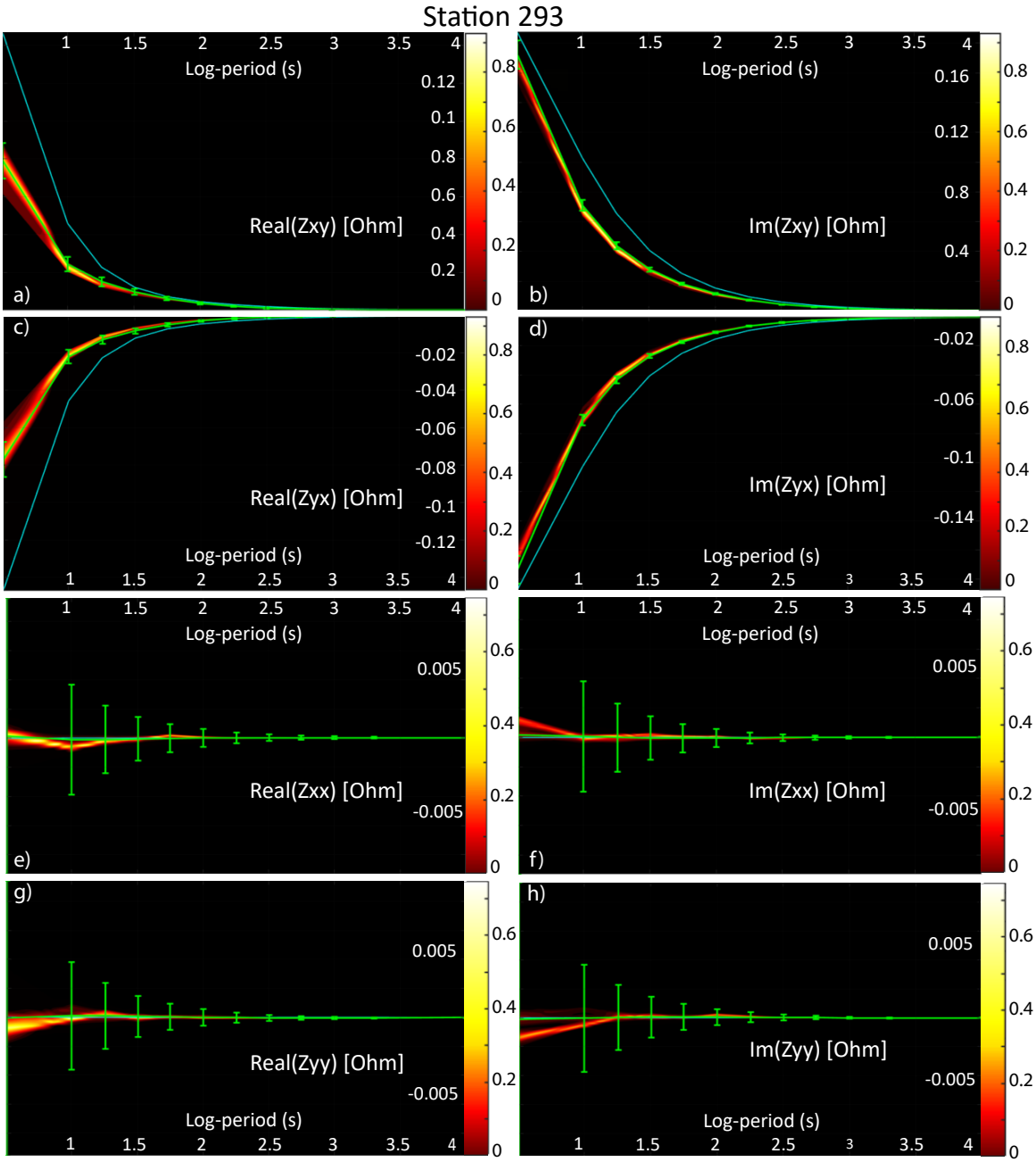

Figure 13. Posterior PDFs of MT data for station 293. Synthetic data and error bars are plotted in green and the computed data for the initial model is plotted in blue. (a)-(d) Posterior PDFs of the real and imaginary parts of the off-diagonal components of the impedance tensor. (e)-(h) Posterior PDFs of the real and imaginary parts of the diagonal components. The location of the station is shown in Fig. 1.a.

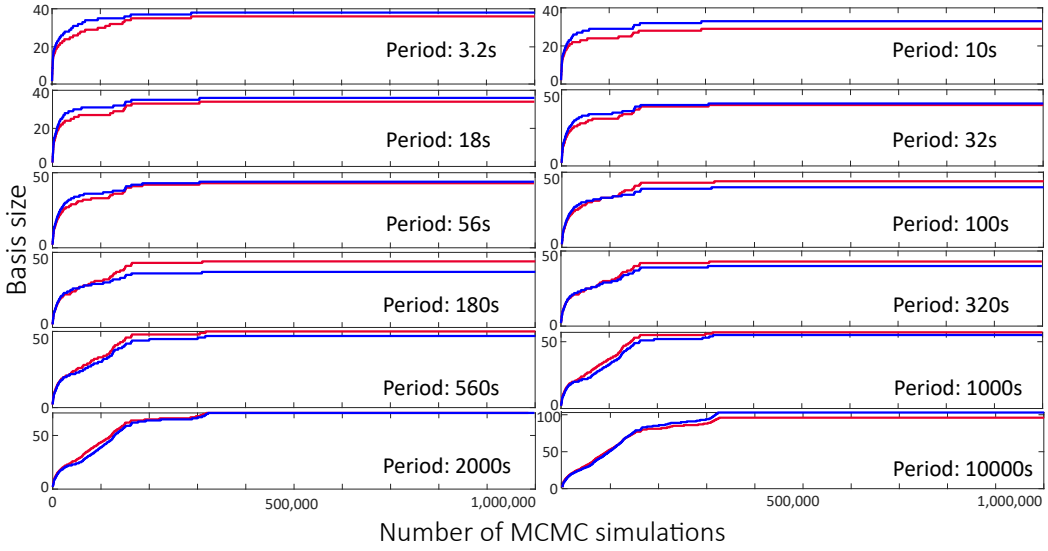

Figure 14. Basis size as a function of the RB+MCMC steps for different frequencies and field orientations ($S_\perp$ mode in blue, and S mode in red).

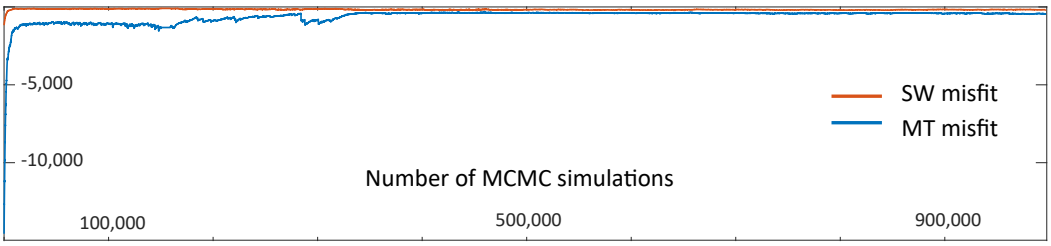

Figure 15. Data misfits for the dispersion curves (red line) and MT (blue line) for each of the 1,500,000 RB+MCMC steps.

the lithosphere/upper mantle, this is not a major issue. This is because the sensitivity to bulk major-element composition is of second-order compared to other factors such as temperature and fluid content. Moreover, in order to increase the sensitivity to bulk mantle composition, the inclusion of gravity and/or geoid anomalies into the inversion becomes a requirement (J. Afonso et al., 2013a; J. C. Afonso et al., 2013b; J. C. Afonso, Rawlinson, et al., 2016). This is left for future work.

7.2 Parameterization and Efficiency of the Algorithm

The current parameterization is specifically tailored to constrain the first-order conductivity background and to locally accommodate smaller-scale anomalies. This parameterization also allows for considerable model variance/flexibility, as it is capable of approximating any conductivity structure, and it favors a rapid convergence at the beginning of the inversion. There are, however, two main drawbacks: i) it is almost impossible to know a priori the minimum number of parameters necessary to retrieve the true model; ii) the algorithm can be inefficient if the number and location of the nodes are not optimal, as an over-parameterization of the model can seriously compromise the convergence of the MCMC algorithm, whereas an under-paramaterization can introduce spurious features in regions where the conductivity nodes are far from each other (since the kriging-like interpolation produces unreal values where poor or none information from the surrounding nodes is available; see Appendix B and B1).

In practice, these issues are addressed by running preliminary inversions (similar to what is done in deterministic inversions with the variance-resolution trade-off diagram; Menke (2018)). A more efficient approach would be to implement trans-dimensional algorithms (e.g. Ray & Myer, 2019; Brodie & Jiang, 2018; Bodin & Sambridge, 2009) to identify the minimum dimensionality of the model (i.e. parsimony), as required by the data. In particular, the combination of the kriging interpolation (also known as Gaussian process regression) with a trans-dimensional algorithm is a promising approach to tackle the problem of potential under/over-parameterization (e.g. Ray & Myer, 2019).

7.3 Ergodicity of the Algorithm

While the sampling strategy described in Section 5 brings in a number of important benefits to the joint RB+MCMC inversion, the first stage (with focus on constraining the first-order temperature structure) can potentially affect the ergodicity of the chain. Given that we deliberately chose to sample one set of parameters (i.e. the LAB depths) for a pre-defined number of MCMC steps, the chain is precluded from reaching the states of the first stage once the second stage starts (i.e. it is not irreducible). As demonstrated in the examples, only a small number of iterations are necessary in the first stage for the LAB's chains to approach the high-probability region of the posterior PDFs. We therefore note that a sufficient condition to ensure the overall ergodicity (e.g. Meyn & Tweedie, 2012) and correct convergence of the sampler requires the *burn-in* period to be larger than the total number of steps in the first stage (*LAB-stage*).

8 Conclusions

We presented a novel, MCMC-driven probabilistic joint inversion of 3D magnetotelluric (MT) and surface-wave (SW) dispersion data for imaging the electrical conductivity and velocity structures of the whole lithosphere and sublithospheric upper mantle. The method is based on i) an efficient parallel-in-parallel structure to solve the 3D MT forward problem, ii) the combination of a reduced order, MCMC-driven strategy to compute fast and accurate surrogate solutions to the 3D MT forward problem, iii) adaptive strategies for both the MCMC algorithm and the surrogate and iv) an efficient dual parameterization to couple both data sets.

The feasibility, potential and efficiency of our algorithm to solve the joint inverse problem are demonstrated with two realistic whole-lithosphere examples. In both cases, we obtain staggering gains in computational efficiency (>96%) compared to a traditional MCMC implementation. Average times per MCMC step are of the order of 1 sec, even when using modest computational resources. We also show that the inclusion of SW data and a simple Cascade-Metropolis algorithm resulted in drastic improvements in computational efficiency and quality of the recovered models compared to the RB+MCMC inversion of MT data only (Manassero et al., 2020).

The model parameterization takes advantage of the differential sensitivities of MT and SW dispersion data to different aspects of the problem by using two sets of parameters. The first set corresponds to LAB depths, which control the large-scale (background) conductivity/velocity structure. The second set corresponds to conductivity nodes inside the model, which control the small-scale conductivity anomalies. An additional advantage of using this parameterization is that a rapid convergence is achieved by using the LAB depths to constrain the first-order conductivity/velocity background at the beginning of the inversion. Once this first-order convergence has been achieved, the nodes are used to locally modify the background to fit the smaller-scale features of the data.

Finally, we note that proposed method is general enough to incorporate more advanced MCMC algorithms (e.g. tras-dimensional schemes, parallel tempering, differential evolution), additional model parameters (e.g. bulk mantle composition) and other forward operators (e.g. gravity anomalies).

Appendix A Mapping Thermochemical Parameters to Electrical Conductivity

The temperature dependence of electrical conductivity can be described with an Arrhenius-type Equation:

$$\sigma = \sigma_0 \exp\Big(\frac{-\Delta H}{k_B T}\Big), \tag{A1}$$

where σ_0 is the so-called pre-exponential factor, T $[K]$ is absolute temperature and k_B $[eV/K]$ the Boltzmann's constant. ΔH $[eV]$ is the pressure-dependent activation enthalpy, defined as

$$\Delta H = \Delta E + P\Delta V, \tag{A2}$$

where P is the pressure $[GPa]$, ΔE and ΔV are the activation energy and activation volume, respectively.

The main bulk conduction mechanisms in mantle minerals are ionic conduction, small polaron (hopping) conduction and proton conduction (e.g. Yoshino, 2010). Each mechanism follows an Arrhenius-type equation with particular activation energies depending on their charge mobility. These three conduction mechanisms can be integrated in a model for the electrical conductivity of mantle rocks as a function of pressure, temperature, water content, and composition (via Fe content) for each mineral phase (see also Yoshino et al., 2009; Fullea et al., 2011):

$$\sigma = \sigma_0 \exp\Big(\frac{-\Delta H(X_{Fe}, P)}{k_B T}\Big) + \sigma_{0i} \exp\Big(\frac{-\Delta H_i}{k_B T}\Big) + \sigma_p \tag{A3a}$$

$$\sigma_p = f(C_w) \exp\Big(\frac{-\Delta H_{wet}(C_w)}{k_B T}\Big), \tag{A3b}$$

$$-\Delta H(X_{Fe}, P) = a + bX_{Fe} + cX_{Fe}^2 + dX_{Fe}^3 + eX_{Fe}^4 + fX_{Fe}^5 + P\Delta V, \tag{A3c}$$

Table A1: Parameters used to compute mantle conductivity

Phase	σ_0	σ_{0i}	a	b	c	d	e	f	ΔV	ΔH_i	X_{Fe}
Olivine	2.70	4.73	1.64	0.246	-4.85	3.26			0.68	2.31	0.10
Opx	3.0		1.90	-2.77	2.61	-1.09					0.107
Cpx	3.25		2.07	-2.77	2.61	-1.09					$5.84e^{-2}$
Garnet		4.96	2.60	-15.33	80.40	-194.6	202.6	-75.0			0.168

where σ_0, σ_{0i} $[S/m]$ and $f(C_w)$ are the small polaron, ionic and proton pre-exponential factors, respectively, ΔV $[cm^3/mol]$ is activation volume, ΔH, ΔH_i $[eV]$ and ΔH_{wet} are activation enthalpies and X_{Fe} is the bulk Fe content in wt%.

The first term in the right-hand side of Equation A3a describes the contribution from small polaron conduction. As mentioned above, the activation enthalpy for this process depends on the iron content and pressure. This dependence is represented by a polynomial on X_{Fe} (Eq. A3c) plus a term that depends on pressure (the coefficients a, b, c, d, e, f are determined experimentally). The second term of Equation A3a represents ionic conduction at high temperature and the third term (σ_p) represents the proton conduction due to the presence of "water" (hydrogen diffusion). $f(C_w)$ and ΔH_{wet} are functions of the water content C_w $[wt\%]$ and they are obtained from laboratory experiments. The reader is referred to Fullea et al. (2011) and Pommier (2014) for a summary on results from different laboratories.

Appendix B Kriging Interpolation

Kriging, or Gaussian process regression, is one of the most common methods for spatial interpolation (see e.g. Cressie, 1993; Rasmussen, 1997; Williams & Rasmussen, 1996; Omre, 1987; M. Gibbs & MacKay, 1997; M. N. Gibbs, 1998). The main idea is to predict (or interpolate) the value of a function Z at m locations from n observations by computing average spatial weights (W). In simple kriging, these weights are derived using a known covariance function c between observations (given by the matrix K_{obs}) and between the observations and the m estimation locations (given by the covariance matrix K_{loc}):

$$W = K_{obs}^{-1} \cdot K_{loc}, \tag{B1}$$

where $K_{obs} = \begin{pmatrix} c(x_1^{obs}, x_1^{obs}) & ... & c(x_1^{obs}, x_n^{obs}) \\ ... & ... & .... \\ c(x_n^{obs}, x_1^{obs}) & ... & c(x_n^{obs}, x_n^{obs}) \end{pmatrix}$ and $K_{loc} = \begin{pmatrix} c(x_1^{obs}, x_1^{loc}) & ... & c(x_1^{obs}, x_m^{loc}) \\ ... & ... & .... \\ c(x_n^{obs}, x_1^{loc}) & ... & c(x_n^{obs}, x_m^{loc}) \end{pmatrix}$.

The interpolation (or estimated value) at the m locations is then given by $Z^{loc} = W \cdot Z^{obs}$, where Z^{obs} is the vector containing the n observations.

The covariance function c can take any form with the only constrain that it must generate a non-negative definite covariance matrix. A common form is given by (e.g. M. Gibbs & MacKay, 1997):

$$c(\mathbf{x}_m, \mathbf{x}_n) = \theta_1 exp\Big(-\frac{1}{2} \sum_l \frac{(x_m^l - x_n^l)^2}{r_l^2} \Big) + \theta_2, \tag{B2}$$

where x_n^l is the l component of $\mathbf{x}_n$. θ_1 and θ_2 are hyperparamaters, where θ_1 represents the overall vertical scale relative to the mean field and θ_2 gives the vertical uncertainty.

r_l is the correlation or scale length and it characterizes the distance in the direction l over which the value of Z varies significantly. It should be noted that since the spatial weights (W) depends on the covariance function c, the interpolated values at the m locations also depends on the chosen form for c.

B1 Spatially varying length scales

The covariance function of Eq. B2 assumes that the correlation length (r_l) is fixed in each direction (l) and location ($\mathbf{x}$). In the most general case, however, assuming a fixed r_l might lead to a simplistic and poor representation of the conductivity model. We, therefore, use a positive definite covariance function with spatially variable correlation lengths (M. Gibbs & MacKay, 1997; M. N. Gibbs, 1998):

$$c(\mathbf{x}_m, \mathbf{x}_n) = \theta_1 \prod_l \left(\frac{2 r_l(\mathbf{x}_m) r_l(\mathbf{x}_n)}{r_l^2(\mathbf{x}_m) + r_l^2(\mathbf{x}_n)} \right)^{1/2} exp\Big(- \sum_l \frac{(x_m^l - x_n^l)^2}{r_l^2(\mathbf{x}_m) + r_l^2(\mathbf{x}_n)} \Big) \tag{B3}$$

where $r_l(\mathbf{x})$ is an arbitrary parameterized function of position $\mathbf{x}$ defined in $[-1,1]^2 \times [0,1]$. The form of $r_l(\mathbf{x})$ as a function of the scaled coordinates (x, y, z) used in Examples 1 and 2 in the main text is shown in Procedure 1. This covariance function has the property that the variance is independent of $\mathbf{x}$ and equal to θ_1. Since a change in θ_1 will produce changes in the vertical scale in the whole domain (see previous section), the inclusion of θ_1 as an additional parameter of the inversion can (potentially) benefit the efficiency and convergence of the algorithm. The implementation of θ_1 as an hyper-parameter of the inversion is left for future work.

Appendix C Log-normal proposal distributions

The log-normal distribution (Gaussian in log-scale) used in the second stage is defined as:

$$y(m_t^i) = \frac{1}{\sqrt{2\pi} m_t^i s} \exp\left(- \frac{\ln(m_t^i) - \mu}{2 s^2}^2 \right), \tag{C1}$$

where $y(m_t^i)$ is the proposed value for node i, and μ and s are the mean and standard deviation in log-scale.

In Section 5 we have chosen to define a multivariate Gaussian proposal of dimension $N_{nodes} \times N_{nodes}$, where N_{nodes} is the number of conductivity nodes in the model. Since the nodes' conductivity values can span several orders of magnitude, the Gaussian proposal is defined in log-scale but we evaluate its probability $q(\cdot|\cdot)$ in linear scale, i.e. a multivariate log-normal PDF centered at the current state $\mathbf{m}_{t-1}$ with covariance Σ:

$$q(\mathbf{m}_t|\mathbf{m}_{t-1}) = \frac{1}{(2\pi)^{\frac{N_{nodes}}{2}} (\det \Sigma)^{\frac{1}{2}} \prod_{j=1}^{N_{nodes}} m_t^j} \exp\left[-\frac{1}{2} (\ln(\mathbf{m}_t) - \ln(\mathbf{m}_{t-1}))^t \Sigma^{-1} (\ln(\mathbf{m}_t) - \ln(\mathbf{m}_{t-1})) \right], \tag{C2}$$

where $\mathbf{m}_t$ is the proposed value for all nodes and $\mathbf{m}_{t-1}$ is the current sample.

Algorithm 1 Definition of $r_l(\mathbf{x})$ as a function of the scaled coordinates (x, y, z).

procedure $r_l(x)$
  if $z >= 0.9$ **then**
    $r_3 = 0.5$
    $r_2 = r_1 = 0.4$
  else if $z < 0.9$ and $z >= 0.8$ **then**
    $r_3 = 0.45$
    $r_2 = r_1 = 0.35$
  else if $z < 0.8$ and $z >= 0.7$ **then**
    $r_3 = 0.4$
    $r_2 = r_1 = 0.3$
  else if $z < 0.7$ and $z >= 0.6$ **then**
    $r_3 = 0.38$
    $r_2 = r_1 = 0.28$
  else if $z < 0.6$ and $z >= 0.5$ **then**
    $r_3 = 0.35$
    $r_2 = r_1 = 0.25$
  else if $z < 0.5$ and $z >= 0.4$ **then**
    $r_3 = 0.33$
    $r_2 = r_1 = 0.23$
  else if $z < 0.4$ and $z >= 0.3$ **then**
    $r_3 = 0.3$
    $r_2 = r_1 = 0.2$
  else if $z < 0.34$ and $z >= 0.2$ **then**
    $r_3 = 0.28$
    $r_2 = r_1 = 0.2$
  else if $z < 0.2$ **then**
    $r_3 = 0.2$
    $r_2 = r_1 = 0.18$
  end if
end procedure

Acknowledgments

We acknowledge that data were not used, nor created for this research. We thank Farshad Salajegheh for providing part of his Matlab codes for plotting results. Special thanks to Kate Selway and Anandaroop Ray for their suggestions at different stages of this work. The 3D rendering views were created using ParaView (Ahrens et al., 2005).

MCM thanks support from an International Macquarie Research Excellence Scholarship (iMQRES). MCM and JCA acknowledge support from ARC Grant DP160103502, ARC Linkage Grant LP170100233, the ARC Centre of Excellence Core to Crust Fluids Systems (http://www.ccfs.mq.edu.au) and the Centre for Earth Evolution and Dynamics, Geoscience Australia and the European Space Agency via the "3D Earth - A Dynamic Living Planet". FZ acknowledges support from CONICET through grant PIP 112-201501-00192. SZ has been funded by the Spanish Ministry through grant DPI2017-85139-C2-2-R, by Catalan government through grant 2017-SGR-1278 and by the EU's Horizon 2020 research and innovation programme under the Marie Sklodowska-Curie grant agreement No 777778.

References

Afonso, J., Fullea, J., Griffin, W., Yang, Y., Jones, A., Connolly, J., & O'Reilly, S. (2013a). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. I: A priori petrological information and geophysical observables. *Journal of Geophysical Research: Solid Earth*, *118*(5), 2586–2617.

Afonso, J. C., Fernandez, M., Ranalli, G., Griffin, W., & Connolly, J. (2008). Integrated geophysical-petrological modeling of the lithosphere and sublithospheric upper mantle: Methodology and applications. *Geochemistry, Geophysics, Geosystems*, *9*(5).

Afonso, J. C., Fullea, J., Yang, Y., Connolly, J., & Jones, A. (2013b). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. II: General methodology and resolution analysis. *Journal of Geophysical Research: Solid Earth*, *118*(4), 1650–1676.

Afonso, J. C., Moorkamp, M., & Fullea, J. (2016). Imaging the lithosphere and upper mantle: Where we are at and where we are going. In N. L. M. Moorkamp P. Lelievre & A. Khan (Eds.), *Integrated imaging of the earth: Theory and applications* (pp. 191–218). John Wiley & Sons.

Afonso, J. C., Ranalli, G., & Fernàndez, M. (2005). Thermal expansivity and elastic properties of the lithospheric mantle: results from mineral physics of composites. *Physics of the Earth and Planetary Interiors*, *149*(3-4), 279–306.

Afonso, J. C., Ranalli, G., Fernàndez, M., Griffin, W. L., O'Reilly, S. Y., & Faul, U. (2010). On the Vp/Vs–Mg# correlation in mantle peridotites: Implications for the identification of thermal and compositional anomalies in the upper mantle. *Earth and Planetary Science Letters*, *289*(3-4), 606–618.

Afonso, J. C., Rawlinson, N., Yang, Y., Schutt, D. L., Jones, A. G., Fullea, J., & Griffin, W. L. (2016). 3-D multiobservable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle: III. Thermochemical tomography in the Western-Central US. *Journal of Geophysical Research: Solid Earth*, *121*(10), 7337–7370.

Ahrens, J., Geveci, B., & Law, C. (2005). Paraview: An end-user tool for large data visualization. *The visualization handbook*, 717–731.

Amestoy, P. R., Duff, I. S., L'Excellent, J.-Y., & Koster, J. (2001). A fully asynchronous multifrontal solver using distributed dynamic scheduling. *SIAM Journal on Matrix Analysis and Applications*, *23*(1), 15–41.

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

136–156.

Bennington, N. L., Zhang, H., Thurber, C. H., & Bedrosian, P. A. (2015). Joint inversion of seismic and magnetotelluric data in the Parkfield Region of California using the normalized cross-gradient constraint. *Pure and Applied Geophysics*, *172*(5), 1033–1052.

Birch, F. (1961). Composition of the earth's mantle. *Geophysical Journal International*, *4*(Supplement_1), 295–311.

Birch, F. (1964). Density and composition of mantle and core. *Journal of geophysical research*, *69*(20), 4377–4388.

Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*(3), 1411–1436.

Bosch, M. (1999). Lithologic tomography: From plural geophysical data to lithology estimation. *Journal of Geophysical Research: Solid Earth*, *104*(B1), 749–766.

Brodie, R., & Jiang, W. (2018). Trans-dimensional Monte Carlo inversion of short period magnetotelluric data for cover thickness estimation. *ASEG Extended Abstracts*, *2018*(1), 1–7.

Calvetti, D., & Somersalo, E. (2007). *An introduction to Bayesian scientific computing: ten lectures on subjective computing* (Vol. 2). Springer Science & Business Media.

Christen, J. A., & Fox, C. (2005). Markov chain Monte Carlo using an approximation. *Journal of Computational and Graphical statistics*, *14*(4), 795–810.

Conrad, P. R., Marzouk, Y. M., Pillai, N. S., & Smith, A. (2016). Accelerating asymptotically exact MCMC for computationally intensive models via local approximations. *Journal of the American Statistical Association*, *111*(516), 1591–1607.

Cressie, N. (1993). *Statistics for spatial data* (Revised Edition ed.). Wiley-Interscience.

Cui, T., Marzouk, Y. M., & Willcox, K. E. (2015). Data-driven model reduction for the Bayesian solution of inverse problems. *International Journal for Numerical Methods in Engineering*, *102*(5), 966–990.

Douglas Jr, J., Santos, J. E., & Sheen, D. (2000). A nonconforming mixed finite element method for Maxwell's Equations. *Mathematical Models and Methods in Applied Sciences*, *10*(04), 593–613.

Douglas Jr, J., Santos, J. E., Sheen, D., & Ye, X. (1999). Nonconforming galerkin methods based on quadrilateral elements for second order elliptic problems. *ESAIM: Mathematical Modelling and Numerical Analysis*, *33*(4), 747–770.

Evans, R. (2012). Conductivity of Earth materials. In J. A. Chave A. (Ed.), *The magnetotelluric method, theory and practice* (pp. 50–95). Cambridge: Cambridge Univ. Press New York.

Evans, R. L., Benoit, M. H., Long, M. D., Elsenbeck, J., Ford, H. A., Zhu, J., & Garcia, X. (2019). Thin lithosphere beneath the central Appalachian Mountains: A combined seismic and magnetotelluric study. *Earth and Planetary Science Letters*, *519*, 308–316.

Feng, R., Yan, H., & Zhang, R. (1986). Fast inversion method and corresponding programming for 3d potential field. *Acta Geol Sin*, *4*(3), 390–402.

Florentin, E., & Díez, P. (2012). Adaptive reduced basis strategy based on goal oriented error assessment for stochastic problems. *Computer Methods in Applied Mechanics and Engineering*, *225*, 116–127.

Frangos, M., Marzouk, Y., Willcox, K., & van Bloemen Waanders, B. (2011). Surrogate and reduced-order modeling: a comparison of approaches for large-scale statistical inverse problems. In O. G. M. H. D. K. B. M. Y. M. L. T. B. v. B. W. L. Biegler G. Biros & K. Willcox (Eds.), *Computational methods for large-scale inverse problems and quantification of uncertainty* (p. 266-290). John Wiley & Sons.

Fullea, J., Muller, M., & Jones, A. (2011). Electrical conductivity of continental lithospheric mantle from integrated geophysical and petrological modeling: Application to the Kaapvaal Craton and Rehoboth Terrane, southern Africa. *Journal of Geophysical Research: Solid Earth*, *116*(B10), 94–105. doi: doi:10.1029/2011JB008544

Galabert, O., Zlotnik, S., Afonso, J. C., & Diez, P. (2019). Ultra-fast Stokes flow solvers for geophysical-geodynamic inverse problems and sensitivity analyses based on reduced order modeling. *Manuscript in review*.

Gallardo, L. A., & Meju, M. A. (2007). Joint two-dimensional cross-gradient imaging of magnetotelluric and seismic traveltime data for structural and lithological classification. *Geophysical Journal International*, *169*(3), 1261–1272.

Gibbs, M., & MacKay, D. J. (1997). Efficient implementation of gaussian processes.

Gibbs, M. N. (1998). *Bayesian gaussian processes for regression and classification* (Unpublished doctoral dissertation). Citeseer.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman and Hall/CRC.

Gouveia, W. P., & Scales, J. A. (1998). Bayesian seismic waveform inversion: Parameter estimation and uncertainty analysis. *Journal of Geophysical Research: Solid Earth*, *103*(B2), 2759–2779.

Gregory, P. (2005). *Bayesian Logical Data Analysis for the Physical Sciences: A Comparative Approach with Mathematica® Support*. Cambridge University Press.

Guerri, M., Cammarano, F., & Tackley, P. J. (2016). Modelling earth's surface topography: decomposition of the static and dynamic components. *Physics of the Earth and Planetary Interiors*, *261*, 172–186.

Haario, H., Laine, M., Mira, A., & Saksman, E. (2006). DRAM: efficient adaptive MCMC. *Statistics and computing*, *16*(4), 339–354.

Haario, H., Saksman, E., Tamminen, J., et al. (2001). An adaptive Metropolis algorithm. *Bernoulli*, *7*(2), 223–242.

Hashin, Z., & Shtrikman, S. (1962). A variational approach to the theory of the effective magnetic permeability of multiphase materials. *Journal of applied Physics*, *33*(10), 3125–3131.

Hashin, Z., & Shtrikman, S. (1963). A variational approach to the theory of the elastic behaviour of multiphase materials. *Journal of the Mechanics and Physics of Solids*, *11*(2), 127–140.

Hassani, B., & Renaudin, A. (2013). The cascade bayesian approach for a controlled integration of internal data, external data and scenarios.

Hassani, B. K., & Renaudin, A. (2018). The cascade bayesian approach: Prior transformation for a controlled integration of internal data, external data and scenarios. *Risks*, *6*(2), 47.

Herrmann, R. B., & Ammon, C. J. (2002). Computer programs in seismology: Surface waves, receiver functions and crustal structure. *St. Louis University, St. Louis, MO*.

Hesthaven, J. S., Rozza, G., & Stamm, B. (2016). *Certified reduced basis methods for parametrized partial differential Equations*. Springer.

Hu, H., Dai, L., Li, H., Hui, K., & Sun, W. (2017). Influence of dehydration on the electrical conductivity of epidote and implications for high-conductivity anomalies in subduction zones. *Journal of Geophysical Research: Solid Earth*, *122*(4), 2751–2762.

Huang, Z., Li, H., Zheng, Y., & Peng, Y. (2009). The lithosphere of north china craton from surface wave tomography. *Earth and Planetary Science Letters*, *288*(1-2), 164–173.

Jackson, I., & Faul, U. H. (2010). Grainsize-sensitive viscoelastic relaxation in olivine: Towards a robust laboratory-based model for seismological application. *Physics of the Earth and Planetary Interiors*, *183*(1-2), 151–163.

Jackson, I., Fitz Gerald, J. D., Faul, U. H., & Tan, B. H. (2002). Grain-size-sensitive seismic wave attenuation in polycrystalline olivine. *Journal of Geophysical Research: Solid Earth*, *107*(B12), ECV–5.

Jegen, M. D., Hobbs, R. W., Tarits, P., & Chave, A. (2009). Joint inversion of marine magnetotelluric and gravity data incorporating seismic constraints: Preliminary results of sub-basalt imaging off the Faroe Shelf. *Earth and Planetary Science Letters*, *282*(1-4), 47–55.

Jones, A. G., Afonso, J. C., & Fullea, J. (2017). Geochemical and geophysical constrains on the dynamic topography of the Southern African Plateau. *Geochemistry, Geophysics, Geosystems*, *18*(10), 3556–3575.

Jones, A. G., Evans, R. L., & Eaton, D. W. (2009). Velocity–conductivity relationships for mantle mineral assemblages in archean cratonic lithosphere based on a review of laboratory data and Hashin–Shtrikman extremal bounds. *Lithos*, *109*(1-2), 131–143.

Kaipio, J., & Somersalo, E. (2006). *Statistical and computational inverse problems* (Vol. 160). Springer Science & Business Media.

Kalscheuer, T., Bastani, M., Donohue, S., Persson, L., Pfaffhuber, A. A., Reiser, F., & Ren, Z. (2013). Delineation of a quick clay zone at smørgrav, norway, with electromagnetic methods under geotechnical constraints. *Journal of Applied Geophysics*, *92*, 121–136.

Karato, S.-I. (1990). The role of hydrogen in the electrical conductivity of the upper mantle. *Nature*, *347*(6290), 272.

Karato, S.-i. (2006). Remote sensing of hydrogen in earth's mantle. *Reviews in Mineralogy and Geochemistry*, *62*(1), 343–375.

Karato, S.-i., & Wang, D. (2013). Electrical conductivity of minerals and rocks. *Physics and Chemistry of the Deep Earth*, *5*, 145–182.

Khan, A. (2016). On Earth's mantle constitution and structure from joint analysis of geophysical and laboratory-based data: An example. *Surveys in Geophysics*, *37*(1), 149–189.

Khan, A., Connolly, J., & Olsen, N. (2006). Constraining the composition and thermal state of the mantle beneath europe from inversion of long-period electromagnetic sounding data. *Journal of Geophysical Research: Solid Earth*, *111*(B10).

Khan, A., Connolly, J., & Taylor, S. (2008). Inversion of seismic and geodetic data for the major element chemistry and temperature of the Earth's mantle. *Journal of Geophysical Research: Solid Earth*, *113*(B9), B09308. doi: 10.1029/2007JB005239

Lebedev, S., & Van Der Hilst, R. D. (2008). Global upper-mantle tomography with the automated multimode inversion of surface and S-wave forms. *Geophysical Journal International*, *173*(2), 505–518.

Liu, H.-P., Anderson, D. L., & Kanamori, H. (1976). Velocity dispersion due to anelasticity; implications for seismology and mantle composition. *Geophysical Journal International*, *47*(1), 41–58.

Mallick, K., & Verma, R. (1979). Time-domain electromagnetic sounding—computation of multi-layer response and the problem of equivalence in interpretation. *Geophysical Prospecting*, *27*(1), 137–155.

Manassero, M. C., Afonso, J. C., Zyserman, F., Zlotnik, S., & Fomin, I. (2020). A Reduced Order Approach for Probabilistic Inversions of 3D Magnetotelluric Data I: General Formulation. *Geophysical Journal International*, *223*(3), 1837–1863.

Matas, J., & Bukowinski, M. S. (2007). On the anelastic contribution to the temperature dependence of lower mantle seismic velocities. *Earth and Planetary Science Letters*, *259*(1-2), 51–65.

Menke, W. (2018). *Geophysical data analysis: Discrete inverse theory.* Academic press.

Meyn, S. P., & Tweedie, R. L. (2012). *Markov chains and stochastic stability.*

Springer Science & Business Media.
Mira, A., et al. (2001). On Metropolis-Hastings algorithms with delayed rejection. *Metron*, *59*(3-4), 231–241.
Moorkamp, M. (2017). Integrating electromagnetic data with other geophysical observations for enhanced imaging of the earth: a tutorial and review. *Surveys in Geophysics*, *38*(5), 935–962.
Moorkamp, M., Fullea, J., Aster, R., & Weise, B. (2020). Inverse methods, resolution and implications for the interpretation of lithospheric structure in geophysical inversions.
Moorkamp, M., Jones, A., & Eaton, D. (2007). Joint inversion of teleseismic receiver functions and magnetotelluric data using a genetic algorithm: Are seismic velocities and electrical conductivities compatible? *Geophysical Research Letters*, *34*(16).
Moorkamp, M., Jones, A., & Fishwick, S. (2010). Joint inversion of receiver functions, surface wave dispersion, and magnetotelluric data. *Journal of Geophysical Research: Solid Earth*, *115*(B4).
Mosegaard, K., & Hansen, T. M. (2016). Inverse methods: Problem formulation and probabilistic solutions. *Integrated Imaging of the Earth: Theory and Applications, Geophysical Monograph*, *218*, 9–27.
Mosegaard, K., & Tarantola, A. (1995). Monte Carlo sampling of solutions to inverse problems. *Journal of Geophysical Research: Solid Earth*, *100*(B7), 12431–12447.
Mosegaard, K., Tarantola, A., et al. (2002). Probabilistic approach to inverse problems. *International Geophysics Series*, *81*(A), 237–268.
Oldenburg, D. W. (1979). One-dimensional inversion of natural source magnetotelluric observations. *Geophysics*, *44*(7), 1218–1244.
Omre, H. (1987). Bayesian kriging—merging observations and qualified guesses in kriging. *Mathematical Geology*, *19*(1), 25–39.
Parker, R. L. (1971). The inverse problem of electrical conductivity in the mantle. *Geophysical Journal International*, *22*(2), 121–138.
Parker, R. L. (1980). The inverse problem of electromagnetic induction: existence and construction of solutions based on incomplete data. *Journal of Geophysical Research: Solid Earth*, *85*(B8), 4421–4428.
Peherstorfer, B., Willcox, K., & Gunzburger, M. (2018). Survey of multifidelity methods in uncertainty propagation, inference, and optimization. *SIAM Review*, *60*(3), 550–591.
Pommier, A. (2014). Interpretation of magnetotelluric results using laboratory measurements. *Surveys in Geophysics*, *35*(1), 41–84.
Quarteroni, A., Manzoni, A., & Negri, F. (2015). *Reduced basis methods for partial differential Equations: an introduction* (Vol. 92). Springer.
Rasmussen, C. E. (1997). *Evaluation of gaussian processes and other methods for non-linear regression* (Unpublished doctoral dissertation). University of Toronto Toronto, Canada.
Ray, A., & Myer, D. (2019). Bayesian geophysical inversion with trans-dimensional Gaussian process machine learning. *Geophysical Journal International*, *217*(3), 1706–1726.
Rosas-Carbajal, M., Linde, N., Kalscheuer, T., & Vrugt, J. A. (2013). Two-dimensional probabilistic inversion of plane-wave electromagnetic data: methodology, model constraints and joint inversion with electrical resistivity data. *Geophysical Journal International*, *196*(3), 1508–1524.
Selway, K. (2014). On the causes of electrical conductivity anomalies in tectonically stable lithosphere. *Surveys in Geophysics*, *35*(1), 219–257.
Selway, K., & O'Donnell, J. (2019). A small, unextractable melt fraction as the cause for the low velocity zone. *Earth and Planetary Science Letters*, *517*, 117–124.

Selway, K., O'Donnell, J., & Özaydin, S. (2019). Upper mantle melt distribution from petrologically constrained magnetotellurics. *Geochemistry, Geophysics, Geosystems*, *20*(7), 3328–3346.

Sheen, D. (1997). Approximation of electromagnetic fields: Part I. Continuous problems. *SIAM Journal on Applied Mathematics*, *57*(6), 1716–1736.

Smith, R. C. (2013). *Uncertainty quantification: theory, implementation, and applications* (Vol. 12). Siam.

Tarantola, A. (2005). *Inverse problem theory and methods for model parameter estimation* (Vol. 89). siam.

Tarantola, A., & Valette, B. (1982). Inverse problems= quest for information. *Journal of geophysics*, *50*(1), 159–170.

Vozar, J., Jones, A. G., Fullea, J., Agius, M. R., Lebedev, S., Le Pape, F., & Wei, W. (2014). Integrated geophysical-petrological modeling of lithosphere-asthenosphere boundary in central Tibet using electromagnetic and seismic data. *Geochemistry, Geophysics, Geosystems*, *15*(10), 3965–3988.

Wait, J. R. (1962). Theory of magnetotelluric fields. *J. Res. NBS D*, *66*(5), 509–541.

Wang, K., Lu, L., Maupin, V., Ding, Z., Zheng, C., & Zhong, S. (2020). Surface wave tomography of northeastern tibetan plateau using beamforming of seismic noise at a dense array. *Journal of Geophysical Research: Solid Earth*, *125*(4), e2019JB018416.

Williams, C. K., & Rasmussen, C. E. (1996). Gaussian processes for regression. In *Advances in neural information processing systems* (pp. 514–520).

Xu, Y., Shankland, T. J., & Poe, B. T. (2000). Laboratory-based electrical conductivity in the Earth's mantle. *Journal of Geophysical Research: Solid Earth*, *105*(B12), 27865–27875.

Yan, L., & Zhou, T. (2019). Adaptive multi-fidelity polynomial chaos approach to bayesian inference in inverse problems. *Journal of Computational Physics*, *381*, 110–128.

Yang, Y., & Forsyth, D. W. (2006). Rayleigh wave phase velocities, small-scale convection, and azimuthal anisotropy beneath southern california. *Journal of Geophysical Research: Solid Earth*, *111*(B7).

Yang, Y., Ritzwoller, M. H., Lin, F.-C., Moschetti, M., & Shapiro, N. M. (2008). Structure of the crust and uppermost mantle beneath the western united states revealed by ambient noise and earthquake tomography. *Journal of Geophysical Research: Solid Earth*, *113*(B12).

Yasar, E., & Erdogan, Y. (2004). Correlating sound velocity with the density, compressive strength and young's modulus of carbonate rocks. *International Journal of Rock Mechanics and Mining Sciences*, *41*(5), 871–875.

Yoshino, T. (2010). Laboratory electrical conductivity measurement of mantle minerals. *Surveys in Geophysics*, *31*(2), 163–206.

Yoshino, T., Matsuzaki, T., Shatskiy, A., & Katsura, T. (2009). The effect of water on the electrical conductivity of olivine aggregates and its implications for the electrical structure of the upper mantle. *Earth and Planetary Science Letters*, *288*(1-2), 291–300.

Zhang, A., Afonso, J. C., Xu, Y., Wu, S., Yang, Y., & Yang, B. (2019). The deep lithospheric structure of the junggar terrane, nw china: Implications for its origin and tectonic evolution. *Journal of Geophysical Research: Solid Earth*, *124*(11), 11615–11638.

Zhang, J., & Taflanidis, A. A. (2019). Accelerating MCMC via Kriging-based adaptive independent proposals and delayed rejection. *Computer Methods in Applied Mechanics and Engineering*, *355*, 1124–1147.

Zyserman, F. I., & Santos, J. E. (2000). Parallel finite element algorithm with domain decomposition for three-dimensional magnetotelluric modelling. *Journal of*

Applied Geophysics, *44*(4), 337–351.