Bayesian inversion of refraction seismic traveltime data

T. Ryberg and Ch. Haberland

Helmholtz Centre Potsdam GFZ German Research Centre for Geosciences, Telegrafenberg, Potsdam, Telegrafenberg, D-14473 Potsdam, Germany.
E-mail: trond@gfz-potsdam.de

SUMMARY

We apply a Bayesian Markov chain Monte Carlo (McMC) formalism to the inversion of refraction seismic, traveltime data sets to derive 2-D velocity models below linear arrays (i.e. profiles) of sources and seismic receivers. Typical refraction data sets, especially when using the far-offset observations, are known as having experimental geometries which are very poor, highly ill-posed and far from being ideal. As a consequence, the structural resolution quickly degrades with depth. Conventional inversion techniques, based on regularization, potentially suffer from the choice of appropriate inversion parameters (i.e. number and distribution of cells, starting velocity models, damping and smoothing constraints, data noise level, etc.) and only local model space exploration. McMC techniques are used for exhaustive sampling of the model space without the need of prior knowledge (or assumptions) of inversion parameters, resulting in a large number of models fitting the observations. Statistical analysis of these models allows to derive an average (reference) solution and its standard deviation, thus providing uncertainty estimates of the inversion result. The highly non-linear character of the inversion problem, mainly caused by the experiment geometry, does not allow to derive a reference solution and error map by a simply averaging procedure. We present a modified averaging technique, which excludes parts of the prior distribution in the posterior values due to poor ray coverage, thus providing reliable estimates of inversion model properties even in those parts of the models. The model is discretized by a set of Voronoi polygons (with constant slowness cells) or a triangulated mesh (with interpolation within the triangles). Forward traveltime calculations are performed by a fast, finite-difference-based eikonal solver. The method is applied to a data set from a refraction seismic survey from Northern Namibia and compared to conventional tomography. An inversion test for a synthetic data set from a known model is also presented.

Key words: Tomography; Controlled source seismology; Seismic tomography.

1 INTRODUCTION

Refraction seismic data sets have been widely used to construct images of the Earth's crust and uppermost mantle (Prodehl & Mooney 2012). Traveltimes from refracted seismic *P*- and *S*-wave first arrivals (diving waves and turning rays) and secondary arrivals (reflections) are used to infer the velocity structure of the subsurface. Typically waves from controlled sources are recorded along lines where seismic sensors are deployed. Since all sources and receivers are located at the surface, the resulting inversion problem is highly ill-posed and, as one of the consequences, structural resolution quickly decreases with depth. Trial-and-error-based methods utilizing traveltime calculations along rays have been used in early times to search for a single 'best-fitting' velocity model which would be in agreement with the observations. These methods, also because the number of traveltime picks was progressively increasing, have later been complemented by inversion techniques, which include tomographic techniques, full-waveform inversions, etc.

For several decades, traveltime tomography is a widely and successfully used inversion technique to investigate the Earth's internal structure. It is applied at all scales, from the local to the global scale (Romanowicz 2003), using signals from artificial sources and earthquakes (Rawlinson & Sambridge 2003; Rawlinson *et al*. 2010; Liu & Gu 2012). Traditionally, the traveltime values 'picked' from the observed waveform signals (of seismic waves emerging from artificial sources or from earthquakes) are inverted for the distribution of the seismic velocity (or slowness) in the subsurface, either along 2-D profiles or in 3-D volumes (Thurber & Aki 1987). Typically, formal inversion routines like damped least squares (DLSQ) or regularized inversions using conjugate gradient methods are invoked to solve the large number of linear equations (e.g. Thurber 1993; Zelt & Barton 1998). To allow for the use of these methods, the tomographic inversion problem is linearized, and the traveltime differences (residuals) between the observed data and synthetic traveltime data related to an initial (or previous) model are minimized in an iterative way (see e.g. Menke 1989). Usually, the subsurface

is parametrized by 2-D or 3-D cells of fixed sizes and shapes. Data distribution and desired spatial resolution are used to determine cell size (number) prior to the inversion.

Although these traditional methods have been very successfully applied for a long time, several issues remain unsatisfying. One of these issues is that the assessment of the quality of the solution and of the uncertainties of the models is not trivial and often only qualitatively feasible through the use of synthetic recovery tests, bootstrapping tests, evaluation of the resolution matrix, etc. Some inversion codes provide formal standard errors, however, they are often unintuitively small in poorly resolved model regions (see e.g. Evans & Achauer 1993; Evans *et al*. 1994). Furthermore, the distribution of data is typically not ideal, that is, traveltime data are spatially irregularly distributed, resulting in spatially varying resolution. In order to account partially for this issue, models with irregular meshes have been used (Bijwaard *et al*. 1998; Thurber & Eberhart-Phillips 1999; Sambridge & Rawlinson 2005). Attempts have been made to adjust the spatial density of the inversion mesh by the data itself (ray path sampling) during the inversion (Sambridge & Faletic 2003; Nolet & Montelli 2005), thus coping locally with varying resolution issues. Anyhow, probably the most severe problem is that the conventional inversion techniques search for a local minimum in the vicinity of a starting model and provide (only) a single 'final' model, that is, exploring the potential model space is—related to the inversion methods traditionally used—rather limited.

In addition to the issues with model parametrization (regular grids, fixed dimensions and grid spacings, etc.), the level of data fit, the level of smoothing and/or damping required to regularize the inverse problem, and other inversion-related parameters have somehow to be determined prior to the inversion. A different, sometimes arbitrary or subjective, choice of these parameters can have a significant impact on the final inversion result.

To overcome some of the disadvantages of the traditional methods presented above, the use of Monte Carlo (MC) searches has been proposed. Instead of applying an inversion method like DLSQ inversion, the model space (the velocity or slowness distribution in the subsurface) is randomly tested and well-fitting models are identified. The main advantages of MC methods are that they provide a suite of well-fitting models as well as estimates of the uncertainties of the obtained models. However, while a variety of MC algorithms (in particular genetic algorithms and simulated annealing) have been successfully applied to different geophysical problems [for example to waveform fitting; see Mosegaard & Sambridge (2002), Sambridge & Mosegaard (2002), and references therein], only a few attempts have been made to apply them to tomographic problems—particularly to surface-based refraction geometries (Pullammanappallil & Louie 1994; Weber 2000; Debski 2010, 2013; Bottero *et al*. 2016). This seems to be mainly due to the typically large number of model parameters, the large number of models necessary to be tested, and the usually 'expensive' traveltime calculation.

An entirely new approach to inversion problems is based on an MC-like investigation (Metropolis *et al*. 1953) of the model space using Markov chains within a Bayesian sampling framework (Shapiro & Ritzwoller 2002; Bodin & Sambridge 2009; Bodin *et al*. 2012a,b; Shen *et al*. 2013). Instead of using a fixed number of cells for the inversion, the dimension of the problem (number N of cells) is treated as an unknown and determined exclusively by the data themselves. The transdimensional or reversible jump Markov chain method (Green 1995; Sambridge *et al*. 2006) allows for transitions between models of different dimensions, thus adjusting the model dimension automatically to the data themselves (Bodin &
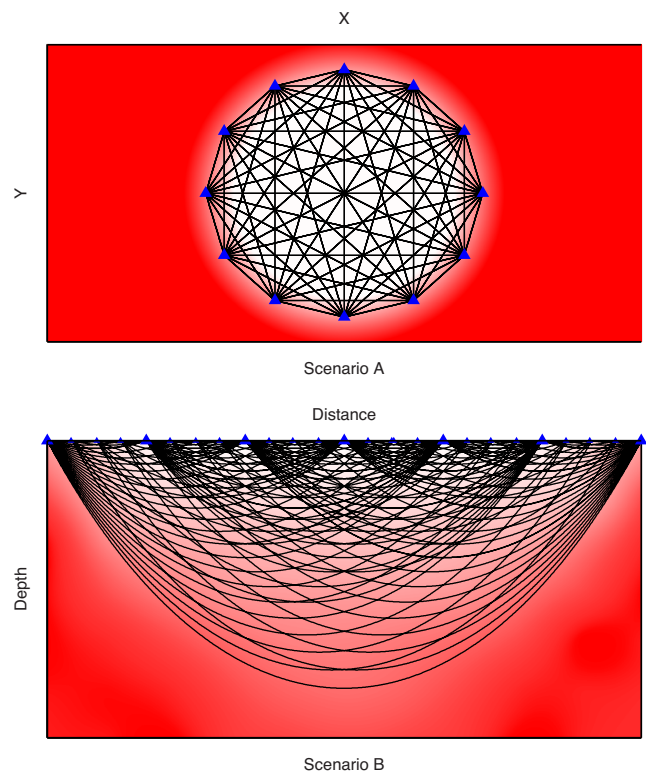

Figure 1. Sketch showing different tomographic scenarios depending on distribution of sources and receivers, colour coded according to ray density (a proxy to potential tomographic spatial resolution). Top: 'Scenario A' is representative of tomographic geometries with 'good' distribution of receivers and/or sources yielding even ray coverage with many crossing rays and—in turn—resulting in a compact region of good resolution (white). The transition region (light red) at the border to the unresolved region (red) is rather small. This scenario is typically found in studies of ambient noise derived group velocities of Rayleigh waves, medical computer tomographic scenarios (MRT) and—roughly—also in cross-hole seismic tomography. Bottom: 'Scenario B' depicts the ray distribution of a typical refraction style data set with sources and receivers situated along a line at the surface. This geometry is characterized by a rather small well-resolved region directly below the surface (white) and a wide gradual transition (light red) toward the unresolved deeper regions (red). Note many subparallel ray paths and only a few crossing rays in the deeper parts captured only by the far-offset recordings. Note that the ray paths for individual models investigated along the Markov chain, while still fitting the data very well, will significantly differ from the distribution shown here.

Sambridge 2009). For this inversion technique, the 'final' inversion result (reference solution) is derived by a superposition (averaging) of a large number of well-fitting models.

A natural extension of this approach treats the omnipresent data error (sometimes difficult to be quantified) as an extra, unknown variable, and consequently inverts for this parameter (Bodin *et al*. 2012a). For these so-called Hierarchical Bayes methods (Bodin *et al*. 2012a), the level of data noise (data uncertainty, i.e. typical traveltime data errors and forward modeling errors) directly affects the level of complexity (model dimension), that is, the inversion 'tries' to decompose the data into a part needed to explain the model and a residual one (actual noise or data uncertainty).

The transdimensional hierarchical tomography using Markov chain MC methods (McMC) has been successfully applied to 2-D traveltime tomographic data sets situated in the horizontal plane, see Fig. 1, as for example ambient noise derived group velocity analysis of Rayleigh waves (see e.g. Bodin & Sambridge
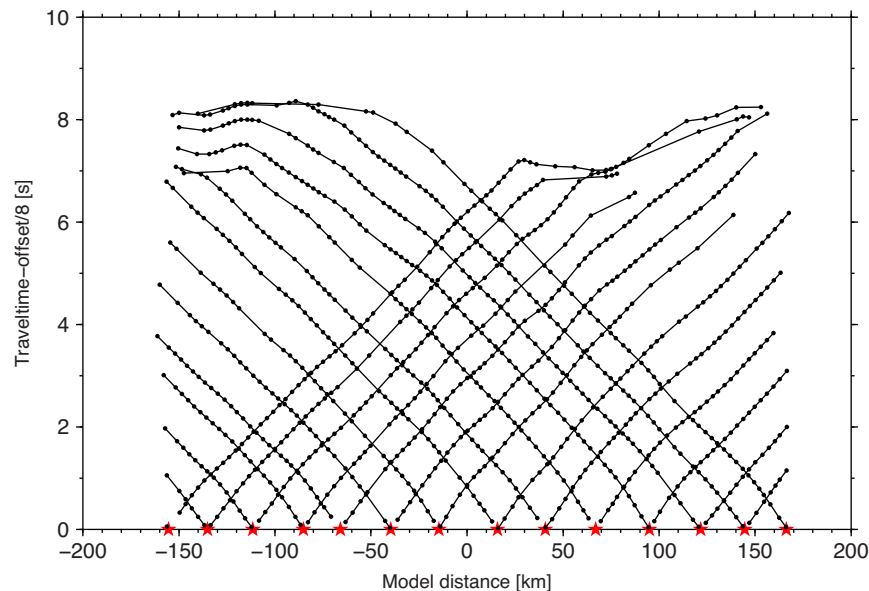
Figure 2. Reduced traveltime picks (black dots) for the real data set of Ryberg *et al*. (2015) for all 14 shots (red stars). All traveltime picks (1014) represent *P*-wave first arrivals (refracted phases).

2009; Bodin *et al*. 2012b), to derive pseudo-3-D models (see e.g. Young *et al*. 2013a,b) or applied to true 3-D problems (see e.g. Hawkins & Sambridge 2015; Burdick & Lekić 2017). In the following, we name this type of experiment geometry 'Scenario A' (see also Fig. 1, top). Agostinetti *et al*. (2015) applied it to a local earthquake tomography problem. In this paper, we apply and extend the McMC inversion technique to 2-D controlled source, refraction style (wide-angle) seismic traveltime data to derive the 2-D velocity distribution in the subsurface (along vertically oriented cross-sections). 2-D refraction style data sets consist of traveltimes of first arrivals recorded from several shots along a line of seismic receivers (typically a larger number). All sources and receivers are located at the Earth's surface and are distributed along a line. The first arrivals are associated to refracted or diving phases resembling arcuate rays whose course and depth penetration is critically controlled by the 2-D velocity distribution, primarily by the prevailing vertical velocity gradient. Fig. 1 (bottom) shows the general setup of this specific experiment geometry, in the following named 'Scenario B'. This tomographic problem is especially ill-posed, given the unfavourable distribution of sources and receivers. Typically, only the very shallow region below the sources and receivers is well constrained by data (traveltimes), while in the deeper part of the model 'crossing' rays are more or less absent, thus significantly reducing the resolution power at depth. In 'Scenario A' experiment geometries, the model space is roughly split into regions with rays passing through and regions with no ray coverage (not constrained by any data), and transitional regions in-between are rather small. In 'Scenario B' experiment geometries (refraction style data), most of the inversion space is of intermediate character, with varying, but generally smaller numbers of rays passing through (see Fig. 1, bottom).

2 EXPERIMENTAL SETUP AND FIELD DATA SET

In this paper, we present an inversion method and its application to a typical, real refraction data set along from an onshore seismic refraction experiment at the eastern prolongation of the Walvis

Ridge into Africa. This seismic experiment, which was carried out in 2010/2011 and aimed to study the continental break-up and creation of the South Atlantic ocean, consisted of a 320 km long, coast-parallel refraction profile in Northern Namibia. For the experiment shots from 14 boreholes were used as seismic sources. The explosions were recorded along the profile with 100 autonomous seismic data loggers recording at 100 sps (samples per second) using short-period (4.5 Hz eigenfrequency), vertical component geophones (see Ryberg *et al*. 2015 for details). Fig. 2 shows the traveltime picks of the refracted *P* phases used in this study.

3 MODEL PARAMETRIZATION

The 2-D models are described by a set of unstructured points $p_i = (x_i, z_i, v_i)$ with $0 < i < N$, located in the x–z plane (N is number of model nodes; x is horizontal coordinate; z is depth and v is seismic velocity or slowness). For the interpolation between these model points (i.e. to generate a fine grid/mesh to calculate traveltimes), we follow and evaluate two approaches, one based on Voronoi cells (with constant velocities within the cell), and one based on a triangulated mesh (with constant gradients within the cells).

In the Voronoi case, the velocity (or slowness) value at each x,z-point of the regular grid $v(x, z)$ is set to the value of the nearest point p of the irregular model (see Fig. 3, left). The generation of the Voronoi-based velocity grid is quite simple (e.g. no interpolation, no special treatment of the model edges) and fast algorithms exist to convert Voronoi meshes to regular grids (Sambridge & Gudmundsson 1998).

The generation of the velocity grid based on a triangulated mesh requires the triangulation of the irregular model p_i and the interpolation within the triangles. We use the *triangle* code (Shewchuk, 1996, 2002) which provides the Delaunay triangulation of a given set of points p_i. In order to yield a full coverage of the area of interest and to achieve a concave hull, we added four artificial points at the corners of the regular model (velocity values set to the value of the nearest point p) before the triangulation. For the generation of the regular grid $v(x, z)$, we first have to know in which triangle a particular gridpoint is located, and then to interpolate within this
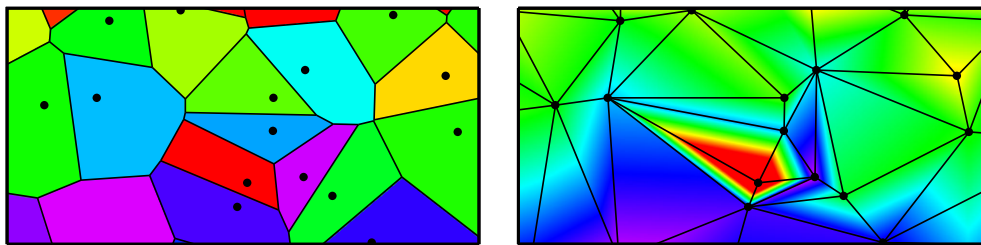
Figure 3. Examples for a Voronoi tesselation model (with constant slowness cells; left) and triangulated mesh (with interpolation within triangles; right) for the same set of model points (black circles). Note the absence of sharp slowness contrasts in the triangulated mesh model.

triangle. The efficient search of the encircling triangle is performed by the *walking triangle* or *trifind* algorithm (Lawson 1977; Lee & Schachter 1980; Sambridge *et al*. 1995) without the need to check all triangles. The interpolation within a particular triangle is done by barycentric interpolation well known from computer graphics (e.g. Möbius 1827). Fig. 3 (right) shows the velocity distribution based on the triangulation. Using these two methods (either Voronoi-cell-based or triangulation-based) fine and regular 2-D-grids/meshes are efficiently generated from the unstructured point models p_i.

4 FORWARD PROBLEM

The resulting regular grid (either based on Voronoi constant slowness cells or on triangulated mesh and interpolation, see above) is then used in a finite-difference (FD) eikonal solver (Podvin & Lecomte 1991) to calculate the first arrival traveltimes for all source and receiver pairs. This traveltime estimation using an eikonal solver is very efficient and no ray tracing is needed. Since traveltimes are calculated on a regular grid, bilinear interpolation between the encircling gridpoints is used to calculate the arrival times at the specific receiver positions. Eventually, the root-mean-square (rms) value of the differences between the measured and calculated traveltime values of a particular model is estimated.

For our applications to synthetic and real data, we used a grid spacing (in *x*- and *z*-directions) of 1 km, resulting in FD grids of 320×80 in size. We extensively tested the potential influence of the forward grid size by using sparser and finer grids. We found no significant differences between the derived reference models for forward grid sizes of 0.25, 0.5, 1.0, 2.0 and 4 km other than those introduced by the per se randomness of the MC technique. Even when using a sparse 4 km forward grid, were we would expect forward traveltime errors caused by the sparse model parametrization when using the eikonal solver, a reference model which did not differ from those with finer forward grids could be inverted for. To avoid propagation (and eventually inversion) of waves above the Earth's surface, we replaced the gridded model above the surface by low velocities (V_{air}) before calculating traveltimes.

5 HIERARCHICAL BAYESIAN APPROACH

We mainly follow the hierarchical transdimensional Bayes algorithm proposed by Bodin *et al*. (2012a,b) by studying multiple, non-interacting Markov chains. We start the Markov chains by choosing a randomly initialized model, then iteratively proceeding with the evolution algorithm. Every step of the Markov chain involves the following steps: we propose a new model based on the current model by (1) changing (with a probability of 1/5) the slowness of a randomly picked cell, or (2) changing the position (move) of a cell,

(3) changing the noise parameter, or (4) adding a new cell (birth) or (5) deleting a randomly chosen cell (death). The choice of the new values for the first three steps is based on the values for the current model (position, slowness, or data noise) which are changed accordingly to a Gaussian probability distribution centred at the current value. The Gaussian probability distributions are characterized by appropriate standard deviations s_x and s_z, s_s and s_n for horizontal and vertical moves, cell slowness and data noise, respectively. We did not allow cells to move outside the model boundaries and restricted the slowness values to be within s_{min} and s_{max}. Note that these values should be chosen carefully, so that the posterior distribution will not be truncated by these limits. We assume minimal prior knowledge and a 'nearly' uninformative prior by choosing a uniform prior distribution with relatively wide bounds (i.e. Bodin *et al*. 2012b; Shen *et al*. 2013; Young *et al*. 2013a,b; Pachai *et al*. 2014). More details regarding the implementation of the Markov chains can be found in Bodin *et al*. (2012a).

Traveltimes are estimated for a newly proposed model (see above, Sections 3 and 4), then the misfit is used to determine the likelihood of the new model. According to the acceptance criterion of Bodin & Sambridge (2009) or Moosegard & Tarantola (1995), the new model is randomly accepted or rejected. To improve the acceptance rate and model space sampling we used the delayed rejection technique (DR; Tierney & Mira 1999; Mira 2001). The proposed new model (or retained current one in the case of rejection) then acts as a starting model for the next iteration. By reiterating this step, we produce a chain of models (Markov chain). The first part of this chain (burn-in phase) is discarded until stationarity of random model space sampling is achieved. After this period, the chain of models is asymptotically distributed according to the posterior distribution, thus realizing a Metropolis sampling algorithm (Gallagher *et al*. 2009).

6 REFERENCE SOLUTION AND ERROR MAP

As the result of the Markov chain calculation, a large number of models fitting the data set well are generated. Each such individual model is usually coarse and looks very 'ungeological'—general examples of these models are for example presented in Fig. 3. Fig. 4 shows the distribution of posterior values for the slowness, the number of cells and the inverted data noise of the well-fitting models (in the post burn-in phase).These models can be converted into a regular grid with a grid-specific spacing (see Section 3). From these gridded models, statistical properties like the average, standard deviation, median, etc., can be constructed locally at every model position (following Bodin & Sambridge 2009; Bodin *et al*. 2012a; Young *et al*. 2013a,b; Burdick & Lekić 2017; Galetti *et al*. 2017). Typically, the average model is treated as the reference solution and the standard deviation is interpreted as a measure of
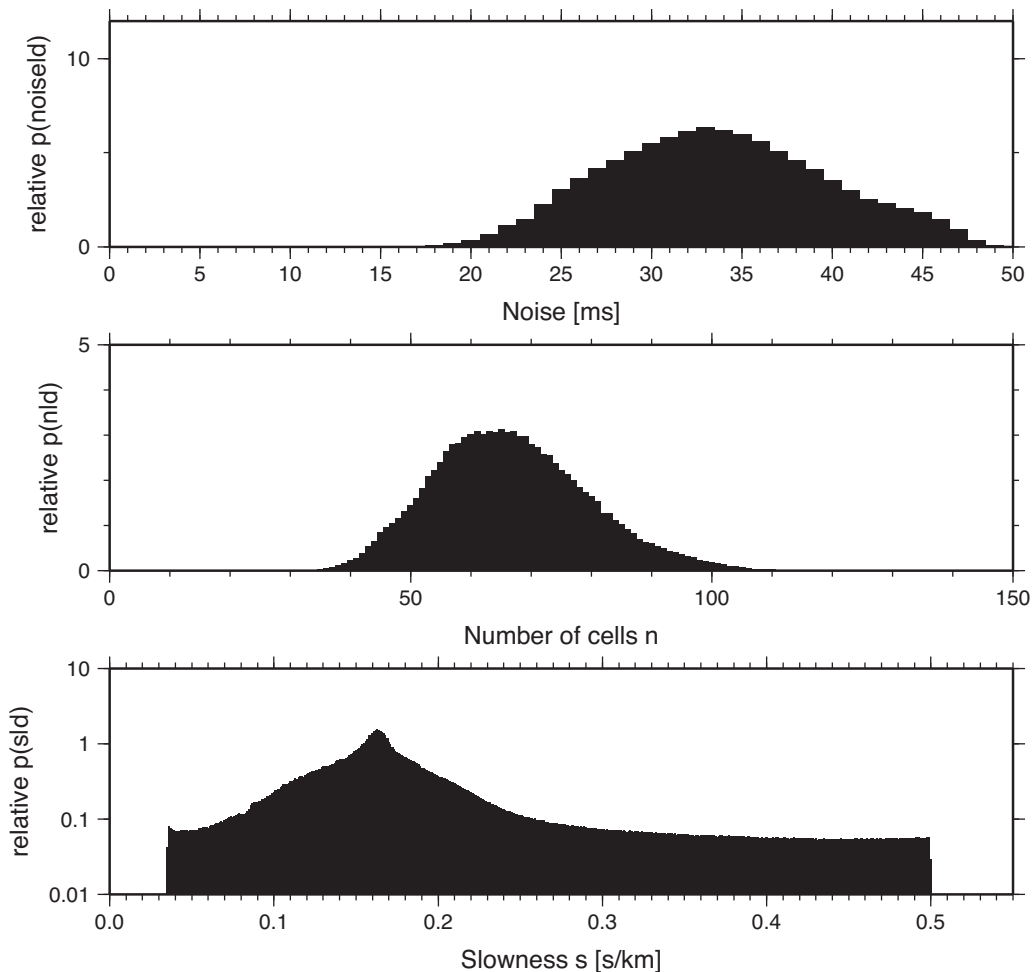
Figure 4. Posterior distribution of slowness values (bottom) at every cell node for the entire model, number of cells (middle) and inverted noise (top) for 1000 Markov chains after the burn-in phase. Note the wide distribution of slowness values, actually a composite of slowness values for regions of highly varying ray coverage. The traveltime data set is compatible with models described by 70 cells on average, while containing ∼0.03 s of traveltime noise.

the model error (uncertainty and resolution). In the following, we refer to this averaged model as the 'reference model'. We averaged the models in slowness space and later converted them to velocities.

For 'Scenario A' tomographic experiments, the conventional averaging procedure (using all models after the burn-in phase) generally works well. In model regions with good ray coverage [and thus in spatial regions having an influence on the traveltimes under consideration (data set)], a reference model and its error map can be recovered. In model regions without ray coverage (i.e. in spatial regions not having an influence on the traveltimes), the recovered model and error map are the mean and the standard deviation of the prior, respectively.[1] In the case of Voronoi cell parametrization, the mean is $(s_{max} - s_{min})/2$ and the standard deviation is $(s_{max} - s_{min})/\sqrt{12}$. For 'Scenario A' type experimental geometries, the model space is typically split into a part with ray coverage (constrained by data) and a part without ray coverage (not constrained by data at all).

For 'Scenario B' tomographic experiments, most of the rays are diving (refracted or turning) rays, and a simple averaging of the Markov chain models will not be suitable to derive a reference solu-

tion. The lower corners of the model are obviously not constrained by any data (no ray coverage), and thus the reference model is a simple function of s_{min} and s_{max}. In the regions close to the receivers and sources, the reference model is well constrained by data, similar to 'Scenario A' type experiments. The dominating space in between has a transitional character ranging from areas of good to very poor ray coverage with a rather complex distribution. In this transitional region, the averaged model is more or less biased by the prior distribution. The complexity of the slowness distribution at every spatial model point is mainly caused by the fact that a large number of good-fitting models with a corresponding small misfit have equivalent ray paths which do not cover the deeper part of the model at all. Thus, any slowness value from the prior distribution in the deeper part of the model will produce a model with a small data misfit, leading finally to a complex, non-Gaussian posterior distribution which consists of a blend of the prior distribution and a contribution constrained by the data. The reason for this behaviour is mainly related to the strong non-linearity of refraction seismic inversion problems: especially in the deeper part of the model characterized by a—on average—poor ray coverage the depth penetration of rays for an individual model strongly and non-linearly depends on the specific velocity model itself. Slight slowness changes (introduced by the model exploration along a Markov chain) can have a large effect on the actual, model-specific depth penetration of rays. As

---

[1] Here and in the following, we resort to rays for clarity, however, please note that we are not calculating rays in our inversion scheme.
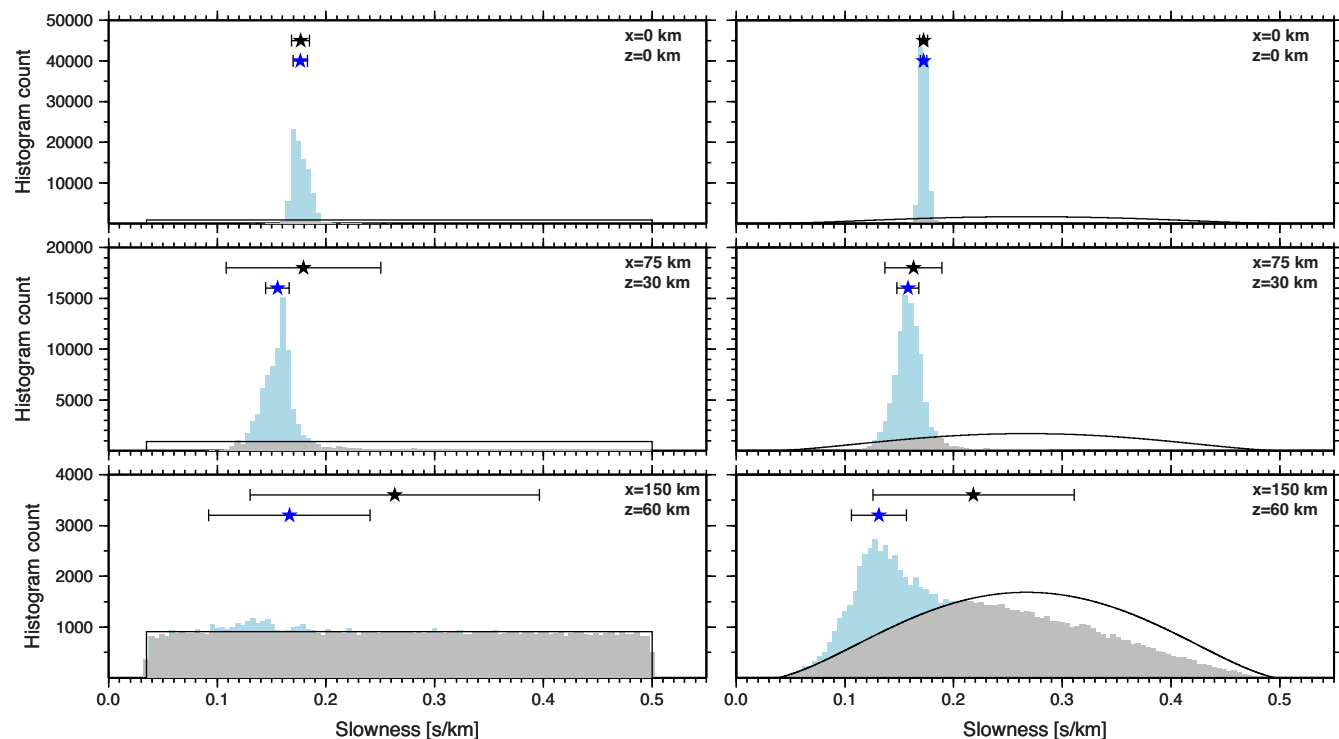
Figure 5. Posterior distribution of recovered slowness values (from ~100 000 models for the same traveltime data set) at different model locations for Voronoi mesh (left) and triangulation based models (right). The top, middle and bottom rows represent model regions with very good, poor and very low/no ray coverage, respectively. Note the different abscissa scales. The slowness distribution is shown as histogram plots (grey and light blue). Black lines show the prior distribution (equally distributed for Voronoi mesh models, complex for triangulated ones). The black stars and bars represent the plain averages (and their standard deviations). Blue stars and bars show the averages (and standard deviations) of the slowness values 'exceeding' the prior distribution (light blue part of the histogram). Note that the two lower panels are located in a region which is typically assumed to be unresolved (no ray coverage in the reference model).

a consequence, regions with, on average, poor ray coverage will sometimes not be illuminated by rays, and thus a posterior slowness distribution is recovered which is identical to the prior. Sometimes rays do propagate into this region, thus contributing information to the posterior distribution which contains a data-driven part. Note, that after the burn-in phase all of these models have a small data misfit. As a result, we obtain models with a depth dependent and complex, i.e. not necessarily Gaussian, slowness distribution. In general, we see a tendency from a nearly Gaussian posterior slowness distribution in regions well covered by rays (shallow parts of the model), through complex distributions in poorly covered regions (intermediate depth) to simple, and identical to the prior distributions in regions not covered by rays (deepest regions of the model), see Fig. 5. These non-Gaussian posterior slowness distributions (bimodal, multimodal, or more complex), for which the derivation of the reference models by simple averaging fails, have been observed in regions of poor ray coverage by Burdick & Lekić (2017). Especially the potential bias by the prior distribution when extracting the reference solution makes the choice of a suitable prior important, since we do not want to have the final model dominated by the prior information (Bodin *et al.* 2012b). Therefore, we assume a 'nearly' uninformative prior by choosing a uniform prior distribution with relatively wide bounds, which includes slownesses values (velocities) not observed in real rocks.

To be able to extract the 'data-driven' part of the posterior distribution, we suggest to replace the simple averaging procedure by a procedure which analyses the posterior (at every gridded model location). In regions poorly covered by rays, the slowness distribution will be biased by the prior distribution in a way that conven-

tional averaging to derive a reference solution will not work. As we mentioned, the posterior at any given model point, consisting of n slowness values, can be assumed to be a mixture of the prior and a data constrained part. By analysing only those data points, which 'exceed' (light blue part of the posterior distribution in Fig. 5), the theoretical expectation for a prior (black line in Fig. 5), we will be able to recover a modified reference model which is, to a high degree, only constrained by data, even in regions of poor ray coverage. For instance, in the case of Voronoi cell parametrization, we determine the average and calculate the standard deviation at every model point only from those slowness values which exceed the expectation for the corresponding prior, i.e. equal distribution between s_{min} and s_{max}. For the case of the triangulate model parametrization, an equal prior distribution of slowness at the model points p_i does not result in an equal prior distribution at arbitrary model gridpoints due to the interpolation involved. Therefore, we calculated the prior distribution at model gridpoints numerically by generating and gridding a large number of models with random nodes p_i (position and slowness), since no obvious analytical solution for calculating the prior distribution exists. Therefore, we approximated the prior distribution of the regridded slowness by a polynomial of degree 8 for further reference. We tested different polynomial degrees and found degree 8 to be sufficiently accurate to approximate the prior distribution. The averaging procedure of the 'residual' models finally results in an average model (reference solution) and its (locally varying) standard deviation, with the latter one assumed to be an approximation to the model error or resolution (Bodin & Sambridge 2009). Fig. 5 shows the posterior distribution of slowness values for ~100 000 good-fitting models at three locations
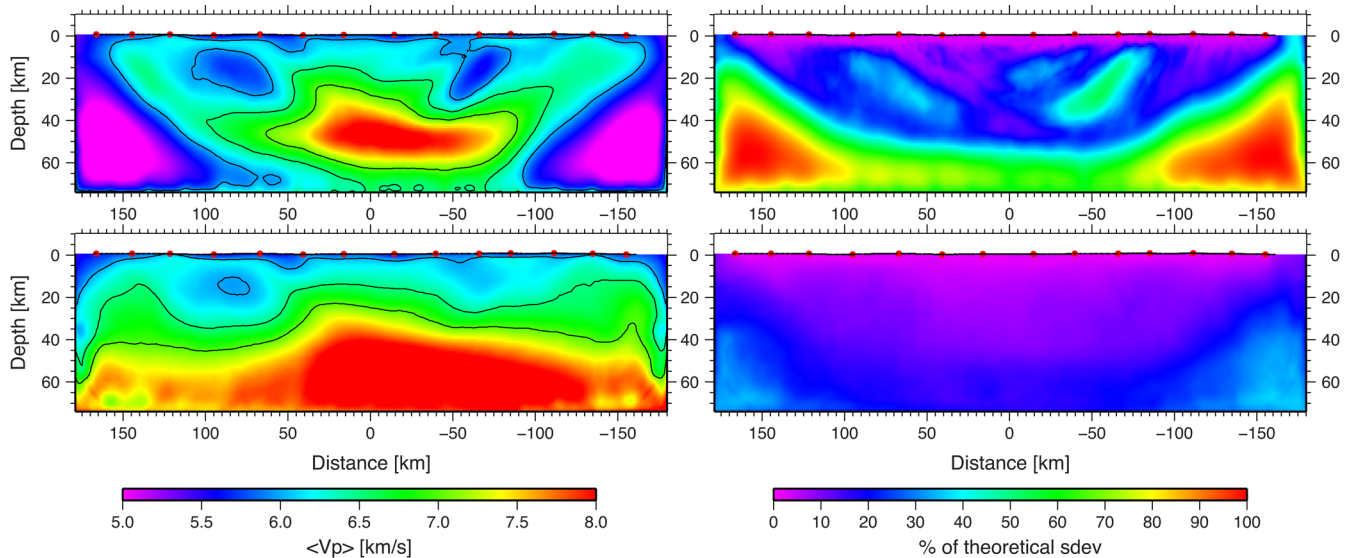
Figure 6. Left: comparison of conventional averaging (top) and modified averaging (bottom) for the entire model after slowness to velocity conversion. Note that the conventional averaging is strongly biased in regions of poor to no ray coverage. Right: comparison of standard deviation (error map) scaled with respect to the theoretically expected values derived by the conventional (top) and the modified averaging procedure (bottom).

in the model for good, poor and very low ray coverage. Note that the posterior slowness distribution at a given model location may contain values which are beyond the limits of the corresponding wave propagation velocities of real rocks. These models with 'unrealistic' rock velocities still have reasonably small traveltime misfits and contribute to the derivation of a reference (average) model and the error map. Fig. 6 compares the reference solutions and error maps using the conventional and modified averaging techniques, showing that the suggested procedure improves the reference solution significantly.

7 STARTING A MARKOV CHAIN

When using the McMC method to explore the model space, one still needs a model to start with. Ideally the choice of this starting model should not have any influence on the model space exploration and—in turn—the results. We investigated different starting models, that is, models with only one cell with a velocity (slowness) matching the average slowness of the data, single cell models with a random slowness, models with huge numbers of cells (up to 1000) of fixed or random slowness, large or small data errors (including completely unrealistic values) etc. In all cases, the evolution of the Markov chain quickly converged to models whose dimension, slownesses and data errors match the data very well. Of course, when starting with completely random models (dimension and slowness), the burn-in phase was extended, compared to more 'realistic' starting models.

We also tested the influence of the choices for the hard boundaries for slowness and data uncertainty (noise), limiting the search space for those values. Significantly changing those values did not have any influence on the reference solution, as long as the values stay sufficiently away from values for good-fitting models (see Section 5). We investigated the influence of the standard deviations (i.e. average length of proposed step for noise, spatial moves and value changes), and, again, found no influence. Of course, the acceptance rate decreased if quite unrealistic values were chosen.

Since no influence of the starting model as well as of the slowness and noise boundaries on the final reference solution could be

observed, we used a 150-cell, constant slowness model for all further inversions, the standard deviation for cell moves was 10 km horizontally (s_x) and 5 km vertically (s_z). The standard deviation for slowness s_s was 0.01 s km^{-1}, the noise deviation s_n was set to 0.01 s. We tested different s_x, s_z, s_s and s_n values, in- or decreased by 50 per cent and found no significant difference for the derived reference models. Since we assumed a 'nearly' uninformative prior we chose a uniform prior distribution with the relatively wide bounds for s_{min} and s_{max} between 0.035 and 0.5 s km^{-1}, thus being far beyond values for rocks in the study area. Again, we tested s_{min} and s_{max} values increased or decreased by 50 per cent and found no significant difference for the derived reference models, and thus no dependence on s_{min} and s_{max}.

8 CONVERGENCE ASSESSMENT

Fig. 7 shows the evolution of the data misfit and dimensionality (number of cells) for 1000 Markov chains for the inversion applied to the real traveltime data set (Fig. 2). The starting model consisted of 150 randomly distributed, constant slowness cells. The average slowness of the entire traveltime data set was used as the starting slowness. Only every 200th model of every individual Markov chain is displayed. While the early models still have a high rms misfit and high dimensionality, both values quickly decrease during the evolution of the Markov chains. After 200 000 steps, we assumed that the burn-in phase (red and blue in Fig. 7) was completed, resulting in low rms misfits and a stationary dimensionality (number of cells). Tests with a longer burn-in phase did not change the reference model significantly, and thus stationarity of the Markov chains was achieved. For a small number of the Markov chains, the rms misfit stays at higher levels, typically associated with models of low complexity (dimension) (Fig. 7, lower panel). Instead of manually excluding those models from deriving a reference model, we limited the averaging to the 90 per cent best-fitting, post burn-in models (red in Fig. 7). It is interesting to note that these 90 per cent best-fitting models are characterized by a somewhat higher dimensionality (proxy for the model complexity).
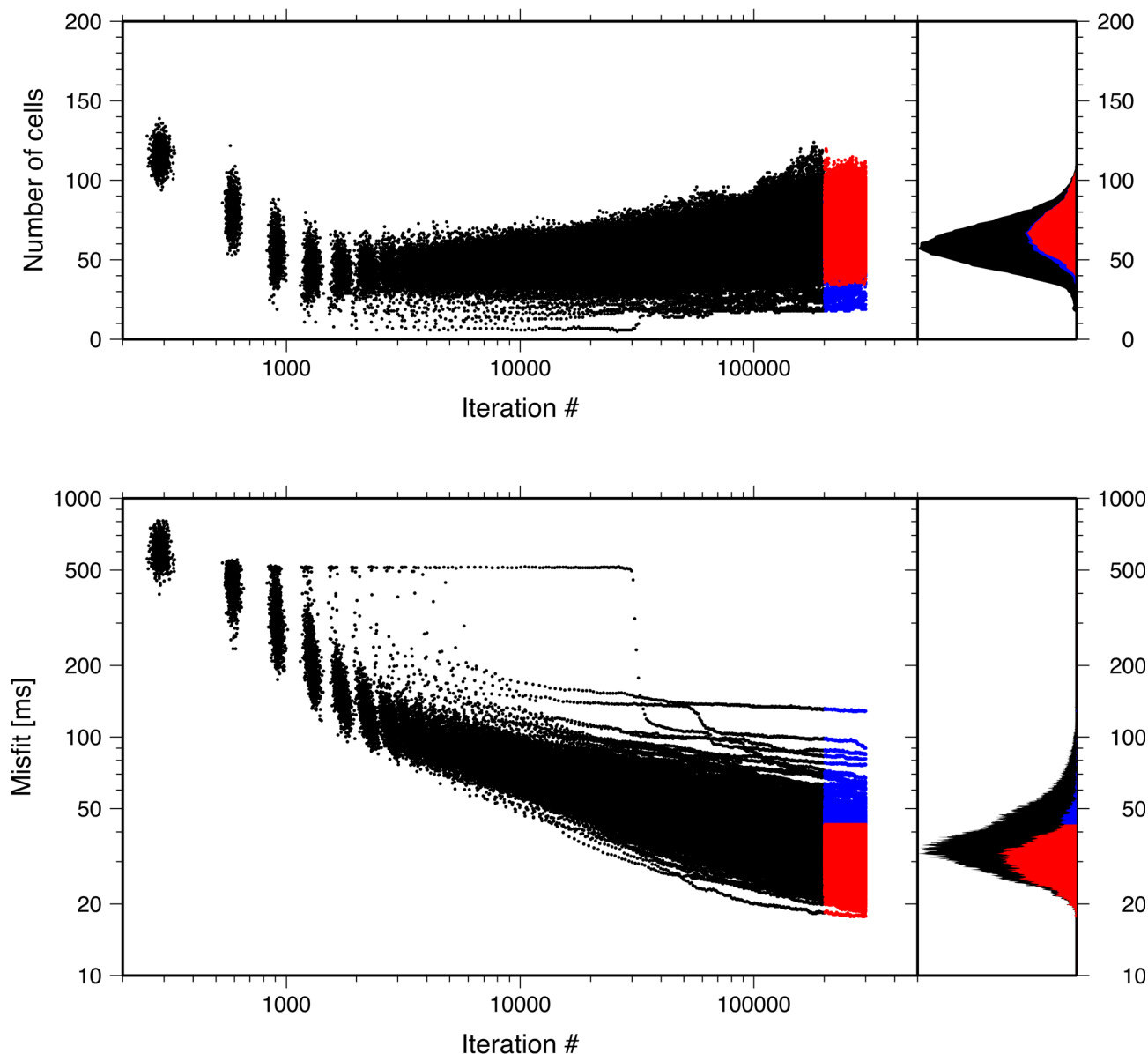
Figure 7. Convergence of 1000 Markov chains. Shown is the distribution of the data misfit (bottom) and model dimension (number of cells, top) during the evolution along the chains. Black dots stand for pre-burn-in models. Blue and red points show model values after the burn-in phase, with red points representing the 90 per cent best-fitting models. Relative histogram plots of the distribution of data misfit and model dimension are added at the right-hand side. The best-fitting models are typically characterized by higher model dimensionality. The forward problem (traveltime calculations) was solved for more than 3.5×10^8 models. Note the log scale for the iteration number and data misfit.

Conventionally, the McMC inversion is applied to multiple Markov chains (for instance Bodin *et al*. 2012a), thus accelerating the computation by taking advantage of parallel computing hardware environments. We observed that analysing the post burn-in models of an individual Markov chain, although producing a large number of good-fitting models, only led to a somewhat locally limited search of the model space. The reference solution of a single Markov chain shows that the model space was not sufficiently explored. For our data example (Fig. 2), we found that the analysis of 100–1000 Markov chains was sufficient for exhaustive model space exploration. Thus, it is not only computationally convenient to explore the model space by multiple Markov chains at the same time, but it seems to be important for sufficiently good model space

exploration. We found that for our data example at least 100 Markov chains should be explored to achieve a stable reference solution.

Our approach of calculating new traveltimes for the entire data set after a model update along a Markov chain is computationally expensive. Given the performance of the eikonal solver we used, in combination with available computer resources, data sets of a typical refraction seismic experiment can be inverted without requiring excessive hardware resources. With a typical runtime of ∼0.04 s on a present day CPU for a complete evaluation of an individual model (regular grid generation, traveltime and misfit calculations), the overall runtime for 1000 Markov chains with 300 000 models each stays below 4000 CPU hours. Since the problem can be perfectly parallelized on a computer cluster, a complete inversion run for
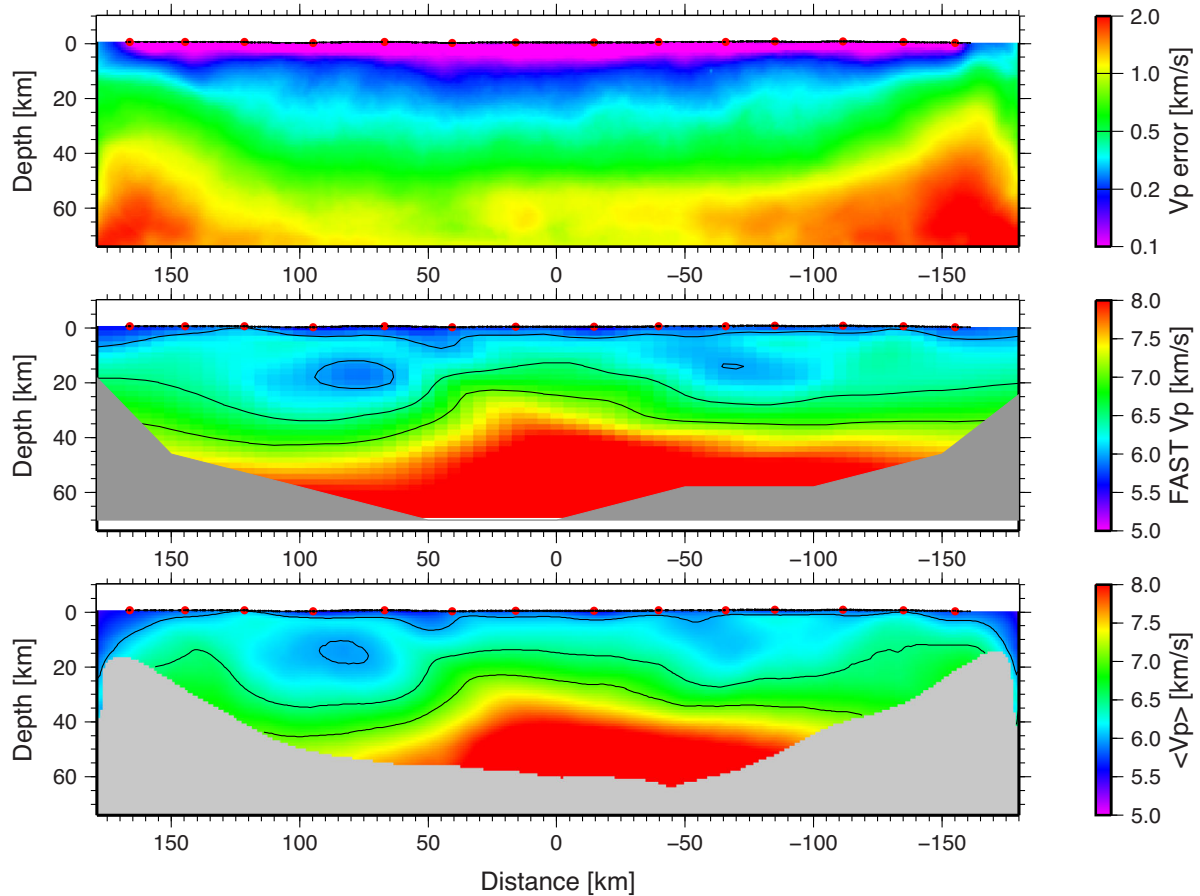
Figure 8. Comparison of McMC inversion results (bottom) with conventional inversion (middle; after Ryberg *et al.* 2015). The top panel shows the distribution of the standard deviation (error map) for the McMC method. As expected, regions of low errors are confined to the topmost part (<20 km depth) of the model. The lower crust and upper mantle is only poorly resolved. Grey areas indicate regions of no ray coverage (middle panel) or standard deviations which exceed a given threshold (50 per cent of the theoretical expectation, see Fig. 6 top right) similar to Galetti *et al.* (2017).

a typical refraction seismic data set can be done within several hours.

9 VORONOI VERSUS TRIANGULATE MODEL PARAMETRIZATION

Ideally the model parametrization should not have an impact on the reference solution. We observed that the traveltime misfit of the averaged models (reference solution) differs depending on what type of model parametrization was used. When visually inspected, the averaged models for both parametrizations look similar. The misfit of the reference solution when using triangulated models (and barycentric interpolation within the triangles) compared to typical individual post burn-in models was only slightly higher (45 versus typically 35 ms for individual models). When using constant slowness Voronoi cells for model parametrization, we observed a significantly higher misfit for the reference solution (>50 ms). We think that this is caused by the way, we calculate the forward problem. Voronoi-based models have per definition sharp slowness contrasts across individual Voronoi cells. Since the forward solver calculates first arrivals only, a significant number of predicted traveltimes correspond to head waves (interface waves). The calculation of the reference solution involves an averaging procedure, that is, the occurrence of sharp contrasts, which potentially produce head waves, is strongly decreased. The smoother reference solution is less

prone to generate head waves, thus explaining the occurrence of an elevated misfit when compared to individual models (which incorporate potential head waves). For refraction style data set inversion, we would therefore give preference to a model parametrization by triangulation and interpolation within the triangles.

10 RESULTS OF REAL-DATA ANALYSIS AND RECOVERY TEST

Fig. 8 shows the reference model and its error map derived by the McMC method, both eventually converted to seismic velocity, compared to the model derived by conventional tomography (Ryberg *et al.* 2015). Generally, both models show a good agreement. Especially, the shallower parts (<10 km depth) show a remarkable coincidence. The mid-crustal low-velocity anomalies differ somewhat: this is probably caused by too small damping of the regularized inversion model. The high-velocity, lower crustal anomaly (a region of potentially poor resolution) is imaged with similar shape and magnitude. A major advantage of the McMC method is that it provides information on the uncertainties of the recovered model. In the very shallow part of the model, down to a few km depth, the velocity uncertainty is around 0.1 km s^{-1}, increasing to 0.2 km s^{-1} down at 20 km depth. The deeper regions have significant larger velocity uncertainties (>0.5 km s^{-1}), because they are only poorly constrained by data.
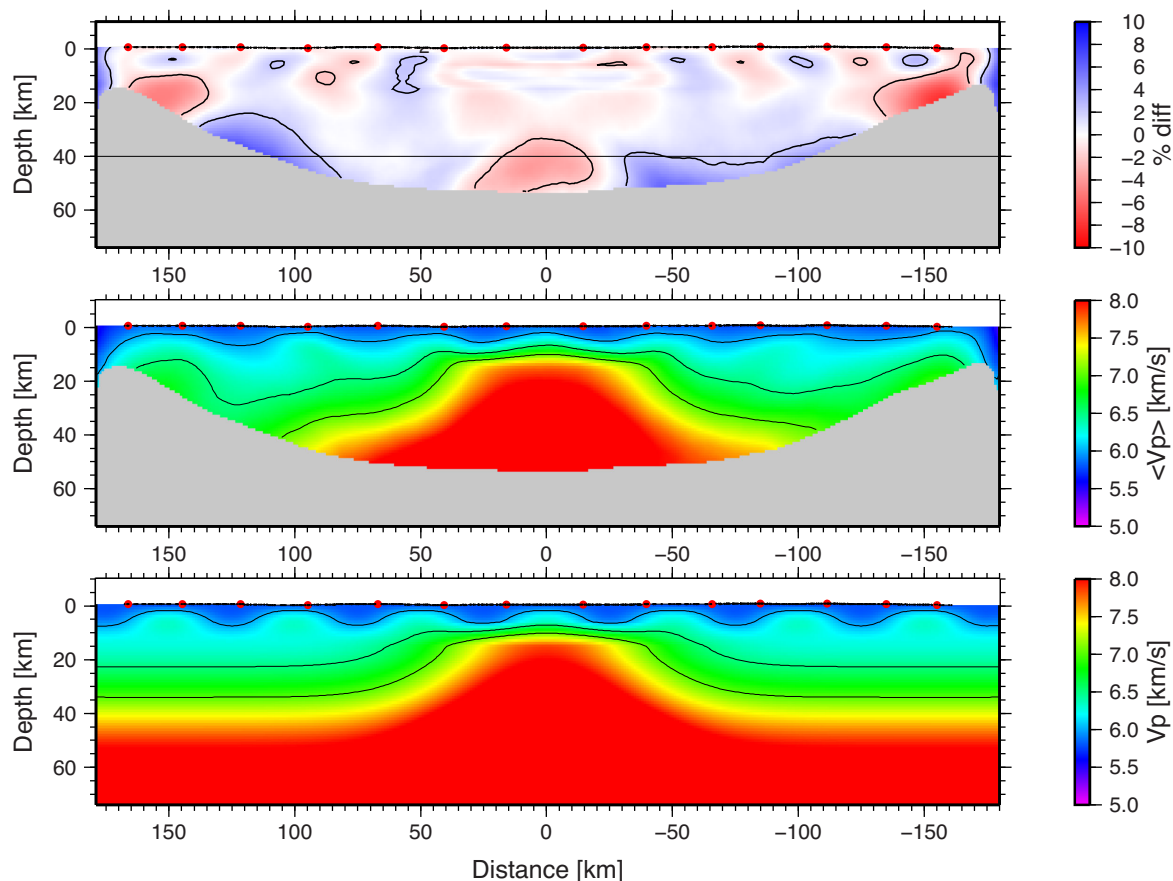
Figure 9. Results of model recovery test. Synthetic traveltimes were calculated for the model (bottom), time jitter (noise) added and inverted (middle). The differences between the input and inverted model (top) are generally below 2 per cent.

Comparing the inversion results for a traveltime data set by the two different methods, although resulting in quite similar models, is limited since we do not know the true crustal model. To test the ability of the McMC method to recover true models, we performed a synthetic test. We generated synthetic traveltimes for the same shot and receiver geometry as in the real experiment. We modeled traveltimes for 14 shots recorded at 100 receiver locations. The number of traveltime picks was identical to the real data set of Ryberg *et al.* (2015). The synthetic model consisted of a smoothed, 1-D version of the final model (Ryberg *et al.* 2015) to ensure similar depth penetration (ray coverage). This model was overlaid by upper crustal high- and low-velocity anomalies and a high-velocity lower crustal body. We added random time jitter (30 ms) to the synthetic data set to simulate the noise determined from the real data travel-time picks. We then performed an McMC inversion and obtained an inversion result (Fig. 9) which is very similar compared to the input model. In most parts of the inverted model, including regions of very poor ray coverage, the difference between both models stays below 2 per cent. The typical number of cells for the synthetic data set (∼70, see Fig. 4) is of the same order as for the real data set, and thus both models have a comparable complexity.

11 CONCLUSIONS

Increasing computer power makes massive exploration of the model space in tomographic imaging possible. We applied the McMC technique to invert traveltime refraction data sets. Instead of a single final model, we thoroughly exploited the model space and derived a large number of representative models which all fit the observed data set very well. In addition, we inverted for the model dimension (number and position of cells in the mesh) and for the data noise, driven by data only. By modifying the averaging procedure, we obtained a stable and reliable estimate for the reference model and its standard deviation (uncertainty and error map) even in regions of reduced resolution. We employed an FD eikonal solver for fast forward (traveltime) calculation and model evaluation. For refraction style data, triangulation (of the irregular point models) and barycentric interpolation within the triangles proved to be a good approach for model parametrization needed to estimate the traveltimes. We would give preference to the model description (parametrization) by triangles instead of constant velocity Voronoi cells, since the triangulated models are more suitable to describe realistic earth models and have much smaller velocity jumps, which might cause traveltime errors when using the eikonal solver. The forward problem (traveltime calculation) is completely performed for each individual model during the search—this seems to be particularly important in refraction type scenarios (Scenario 'B') in which even subtle velocity variations (of the predominantly vertical gradients) can result in very different ray paths, particularly very different depth penetration. Given the nature of the McMC technique, no starting model, pre-set parametrization values, damping or smoothing parameters were needed. The result of the presented inversion technique was compared to conventional inversion and successfully tested with a synthetic data set, thus indicating a similar performance compared to conventional tomographic methods.

However, the main advantage of the McMC approach is that it frees one from the choice of sometimes subjective and/or arbitrary parameters used in conventional inversion techniques (inversion grid size, damping, smoothing, pre-set data noise, starting model, etc.), while providing important additional model constraints (error maps) and information on data noise. It would be straightforward to extend the algorithm to the 3-D case, apply it to local earthquake data or cross-hole tomographic scenarios, invert for other hyperparameters (i.e. forward grid size) or include other data sets (i.e. *S*-wave traveltimes, traveltimes of reflected phases, etc.) for joint inversions. The method can easily be adapted to different scales and performs well with multiscale problems.

ACKNOWLEDGEMENTS

REFERENCES

Agostinetti, N.P., Giacomuzzi, G. & Malinverno, A., 2015. Local three-dimensional earthquake tomography by transdimensional Monte Carlo sampling, *Geophys. J. Int.,* **201,** 1598–1617.

Bijwaard, H., Spakman, W. & Engdahl, E.R., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.,* **103**(B12), 30 055–30 078.

Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.,* **178**(3), 1411–1436.

Bodin, T., Sambridge, M., Rawlinson, N. & Arroucau, P., 2012a. Transdimensional tomography with unknown data noise, *Geophys. J. Int.,* **189**(3), 1536–1556.

Bodin, T., Sambridge, M., Tkalčić, H., Arroucau, P., Gallagher, K. & Rawlinson, N., 2012b. Transdimensional inversion of receiver functions and surface wave dispersion, *J. geophys. Res.,* **117,** B02301, doi:10.1029/2011JB008560

Bottero, A., Gesret, A., Romary, Th., Noble, M. & Maisons, Ch., 2016. Stochastic seismic tomography by interacting Markov chains, *Geophys. J. Int.,* **207**(1), 374–392.

Burdick, S. & Lekić, V., 2017. Velocity variations and uncertainty from transdimensional P-wave tomography of North America, *Geophys. J. Int.,* **209**(2), 1337–1351.

Debski, W., 2010. Seismic tomography by Monte Carlo sampling, *Pure appl. Geophys.,* **167**(1–2), 131–152.

Debski, W., 2013. Bayesian approach to tomographic imaging of rock-mass velocity heterogeneities, *Acta Geophys.,* **61**(6), 1395–1436.

Evans, J.R. & Achauer, U., 1993. Teleseismic velocity tomography using the ACH method: theory and applications to continental-scale studies, in *Seismic Tomography: Theory and Practice*, pp. 319–360, eds Iyer, H.M. & Hirahara, K., Chapman and Hall.

Evans, J.R., Eberhart-Phillips, D. & Thurber, C.H., 1994. User's manual for SIMULps for imaging Vp and Vp/Vs: a derivative of the "Thurber" tomographic inversion SIMUL3 for local earthquakes and explosions. Open-File Rep., U.S. Geol. Surv. USGS-OFR-94-431.

Galetti, E., Curtis, A., Baptie, B., Jenkins, D. & Nicolson, H., 2017. Transdimensional Love-wave tomography of the British Isles and shear-velocity structure of the East Irish Sea Basin from ambient-noise interferometry, *Geophys. J. Int.,* **208**(1), 36–58.

Gallagher, K, Charvin, K, Nielsen, S., Sambridge, M. & Stephenson, J., 2009. Markov chain Monte Carlo (MCMC) sampling methods to determine optimal models, model resolution and model choice for Earth Science problems, *Mar. Petrol. Geol.,* **26**(4), 525–535.

Green, P., 1995. Reversible jump Markov chain Monte Carlo computation and Bayesian model determination, *Biometrika,* **82**(4), 711–732.

Hawkins, R. & Sambridge, M., 2015. Geophysical imaging using transdimensional trees, *Geophys. J. Int.,* **203**(2), 972–1000.

Lawson, C.L., 1977. Software for C1 Surface Interpolation, in *Mathematical Software III*, pp. 161–194, ed. Rice, J.R., Academic Press.

Lee, D.T. & Schachter, B.J., 1980. Two algorithms for constructing a Delaunay triangulation, *Int. J. Comput. Inform. Sci.,* **9**(3), 219–242.

Liu, Q. & Gu, Y.J., 2012. Seismic imaging: from classical to adjoint tomography, *Tectonophysics,* **566,** 31–66.

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

Metropolis, N., Rosenbluth, M.N., Rosenbluth, A.W., Teller, A.H. & Teller, E., 1953. Equation of state calculations by fast computing machines, *J. Chem. Phys.,* **21**(6), 1087–1092.

Mira, A., 2001. On Metropolis-Hastings algorithm with delayed rejection, *Metron,* **LIX**(3–4), 231–241.

Möbius, A.F., 1827. *Der Barycentrische Calcul,* Johann Ambrosius Barth Verlag.

Mosegaard, K. & Sambridge, M., 2002 Monte Carlo analysis of inverse problems, *Inverse Probl.,* **18**(3), R29–R54.

Mosegaard, K. & Tarantola, A., 1995. Monte Carlo sampling of solutions to inverse problems, *J. geophys. Res.,* **100**(B7), 12 431–12 447.

Nolet, G. & Montelli, R., 2005. Optimal parametrization of tomographic models, *Geophys. J. Int.,* **161**(2), 365–372.

Pachhai, S., Tkalčić, H. & Dettmer, J., 2014. Bayesian inference for ultralow velocity zones in the Earth's lowermost mantle: complex ULVZ beneath the east of the Philippines, *J. geophys. Res.,* **119,** 8346–8365.

Podvin, P. & Lecomte, I., 1991. Finite difference computation of traveltimes in very contrasted velocity models: a massively parallel approach and its associated tools, *Geophys. J. Int.,* **105**(1), 271–284.

Prodehl, C. & Mooney, W.D., 2012. Exploring the Earth's crust: history and results of controlled-source seismology, *Geol. Soc. Am. Mem.,* **208,** 764, doi:10.1130/2012.2208(002).

Pullammanappallil, S.K. & Louie, J., 1994. A generalized simulated-annealing optimization for inversion of first-arrival times, *Bull. seism. Soc. Am.,* **84**(5), 1397–1409.

Rawlinson, N. & Sambridge, M., 2003. Seismic traveltime tomography of the crust and lithosphere, *Adv. Geophys.,* **46,** 81–198.

Rawlinson, N., Pozgay, S. & Fishwick, S., 2010. Seismic tomography: a window into deep Earth, *Phys. Earth planet. Inter.,* **178**(3–4), 101–135.

Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years, *Annu. Rev. Earth Planet. Sci.,* **31**(1), 303–328.

Ryberg, T., Haberland, C., Haberlau, T., Weber, M., Bauer, K., Behrmann, J.H. & Jokat, W., 2015. Crustal structure of northwest Namibia: evidence for plume-rift-continent interaction, *Geology,* **43**(8), 739–742.

Sambridge, M. & Guṁundsson, Ó., 1998. Tomographic systems of equations with irregular cells, *J. geophys. Res.,* **103**(B1), 773–781.

Sambridge, M. & Faletic, R., 2003. Adaptive whole Earth tomography, *Geochem. Geophys. Geosyst.,* **4**(3), 1022, doi:10.1029/2001GC000213.

Sambridge, M. & Mosegaard, K., 2002. Monte Carlo methods in geophysical inverse problems, *Rev. Geophys.,* **40**(3), 1009, doi:10.1029/2000RG000089.

Sambridge, M. & Rawlinson, N., 2005. Seismic tomography with irregular meshes, in *Seismic Earth: Array Analysis of Broadband Seismograms*, pp. 49–65, eds Levander, A. & Nolet, G. American Geophysical Union.

Sambridge, M., Braun, J. & McQueen, H., 1995. Geophysical parametrization and interpolation of irregular data using natural neighbours, *Geophys. J. Int.,* **122**(3), 837–857.

Sambridge, M., Gallagher, K., Jackson, A. & Rickwood, P., 2006. Transdimensional inverse problems, model comparison and the evidence, *Geophys. J. Int.,* **167**(2), 528–542.

Shapiro, N.M. & Ritzwoller, M.H., 2002. Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle, *Geophys. J. Int.,* **151**(1), 88–105.

Shen, W., Ritzwoller, M.H., Schulte-Pelkum, V. & Lin, F.C., 2013. Joint inversion of surface wave dispersion and receiver functions: a Bayesian Monte-Carlo approach, *Geophys. J. Int.,* **192**(2), 807–836.

Shewchuk, J.R., 1996. Triangle: engineering a 2D quality mesh generator and delaunay triangulator, in *Applied Computational Geometry: Towards Geometric Engineering'*, pp. 203–222, eds Lin, M.C. & Manocha, D., Lecture Notes in Computer Science, Vol. 1148, Springer-Verlag.

Shewchuk, J.R., 2014. Reprint of: Delaunay refinement algorithms for triangular mesh generation, *Comput. Geom.,* **47**(7), 741–778.

Thurber, C.H., 1993. Local earthquake tomography: velocities and VP/VS-theory, in *Seismic Tomography*, pp. 563–583, eds Iyer, H.M. & Hirahara, K., Chapman and Hall.

Thurber, C.H. & Aki, K., 1987. Three-dimensional seismic imaging, *Annu. Rev. Earth planet. Sci.,* **15**(1), 115–139.

Thurber, C.H. & Eberhart-Phillips, D., 1999. Local earthquake tomography with flexible gridding, *Comput. Geosci.,* **25**(7), 809–818.

Tierney, L. & Mira, A., 1999. Some adaptive Monte Carlo methods for Bayesian inference, *Stat. Med.,* **8**(17–18), 2507–2515.

Weber, Z., 2000. Seismic traveltime tomography: a simulated annealing approach, *Phys. Earth planet. Inter.,* **119**(1–2), 149–159.

Wessel, P. & Smith, W., 1995. New version of the generic mapping tools, *EOS, Trans. Am. geophys. Un.,* **76**(33), 329–329.

Wessel, P. & Smith, W., 1998. New, improved version of generic mapping tools released, *EOS, Trans. Am. geophys. Un.,* **79**(47), 579–579.

Young, M.K., Rawlinson, N. & Bodin, T., 2013a. Transdimensional inversion of ambient seismic noise for 3D shear velocity structure of the Tasmanian crust, *Geophysics,* **78**(3), WB49–WB62.

Young, M.K., Cayley, R.A., McLean, M.A., Rawlinson, N., Arroucau, P. & Salmon, M., 2013b. Crustal Structure of the east Gondwana margin in southeast Australia revealed by transdimensional ambient seismic noise tomography, *Geophys. Res. Lett.,* **40**(16), 4266–4271.

Zelt, C. & Barton, P.J., 1998. Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe Basin, *J. geophys. Res.,* **103,** 7187–7210.