

Transdimensional tomography with unknown data noise

Thomas Bodin,^{1,*} Malcolm Sambridge,¹ Nick Rawlinson¹ and Pierre Arroucau²

¹Research School of Earth Sciences, Australian National University, Canberra ACT 0200, Australia. E-mail: thomas.bodin@anu.edu.au

²Center for Research Excellence in Science and Technology, North Carolina Central University, Durham NC 27707, USA

Accepted 2012 February 6. Received 2012 January 25; in original form 2011 September 26

SUMMARY

A meaningful interpretation of seismic measurements requires a rigorous quantification of the uncertainty. In an inverse problem, the data noise determines how accurately observations should be fit, and ultimately the level of detail contained in the recovered model. A common problem in seismic tomography is the difficulty in quantifying data uncertainties, and thus the required level of data fit. Traditionally, the complexity of the solution model (defined by both the number of basis functions and the regularization) is defined arbitrarily by the user prior to inversion with only limited use of data errors. In the context of multiscale problems, dealing with multiple data sets that are characterized by different noise variances and that span the Earth at different scales is a major challenge. Practitioners are usually required to arbitrarily weigh the contribution of each data type into the final solution. Furthermore, the basis functions are usually spatially uniform across the velocity field and regularization procedures are global, which prevents the solution model from accounting for the uneven spatial distribution of information. In this work we propose to address these issues with a Hierarchical Bayesian inversion. The new algorithm represents an extension of the transdimensional tomography to account for uncertainties in data noise. This approach has the advantage of treating the level of noise in each data set, as well as the number of model parameters, as unknowns in the inversion. It provides a parsimonious solution that fully represents the degree of knowledge one has about seismic structure (i.e. constraints, resolution and trade-offs). Rather than being forced to make decisions on parametrization, level of data fit and weights between data types in advance, as is often the case in an optimization framework, these choices are relaxed and instead constrained by the data themselves. The new methodology is presented in a synthetic example where both the data density and the underlying structure contain multiple length scales. Three ambient noise data sets that span the Australian continent at different scales are then simultaneously inverted to infer a multiscale tomographic image of Rayleigh wave group velocity for the Australian continent. The procedure turns out to be particularly useful when dealing with multiple data types with different unknown levels of noise as the algorithm is able to naturally adjust the fit to the different data sets and provide a velocity map with a spatial resolution adapted to the spatially variable information present in the data.

Key words: Inverse theory; Probability distributions; Surface waves and free oscillations; Seismic tomography; Computational seismology; Australia.

1 INTRODUCTION

Seismic ground motion represents one of the most valuable resources for investigating the internal structure, composition and dynamics of the Earth. When combined with petrophysical experiments and geodynamical modelling to aid interpretation, seismic data allow us to examine the scale and nature of heterogeneities in the crust and upper mantle in different tectonic regions. To-

mographic imaging has been a favoured tool of seismologists for interpretation of seismic data for more than 30 yr, and has resulted in numerous models of seismic wave speed, attenuation and anisotropy at different scales. Recent summaries can be found in Rawlinson & Sambridge (2003); Romanowicz (2003); Rawlinson *et al.* (2010).

1.1 A heterogeneous distribution of information

Due to the increasing quantity and quality of seismic data, the number of overlapping tomographic models and their consistency has dramatically increased in recent years (Nolet 2008). However, several problems remain, such as the uneven spatial distribution

*Now at: Berkeley Seismological Laboratory, University of California, Berkeley, California, USA.

of information contained in seismic data, due mainly to irregular source–receiver path coverage. Another issue is the difficulty in accurately quantifying the level of data noise in seismic waveforms, and hence the level of interpretable information present in observed data (Gouveia & Scales 1998; Scales & Snieder 1998).

The unbalanced spatial and azimuthal sampling of the Earth and the presence of data uncertainties results in an inherent non-uniqueness of the tomographic inverse problem, which often leads to ambiguity in the interpretation of seismic models. Furthermore, with increasing computational power, tomographic studies simultaneously treating different data sets that sample the Earth at different scales—and hence with different resolving power—are gaining in popularity. For example, mixing teleseismic and regional seismic information generates a tomographic system that is especially ill-posed given the tremendous variability of data sampling. In such situations, the level of interpretable information in seismic data is spatially variable, difficult to measure and differs between data types and processing schemes. This raises the issue of how much information should be present in the recovered image (i.e. how complex should the model be?), and how this information should be spatially distributed.

The level of model complexity is directly determined by the choice of model parametrization and regularization. Traditionally tomographic studies have largely ignored the spatial variability in the data's resolving power and opted to build models in terms of uniform local or globally supported basis functions, for example, regular cells or spherical harmonics. The limitation of a regular parametrization is that in order to resolve densely sampled areas, small-scale artefacts in regions where the velocity field is much less well constrained may be introduced. This effect is usually avoided by adding non data-driven constraints on the model (i.e. regularization). However, most smoothing and damping procedures are global (responsive to the entire model) and while averaging over large areas, they make resolvable small-scale details difficult to see, or even mask them all together. Moreover, sharp discontinuities present in the model are typically blurred by smoothing processes into gradual transitions. Note however that several groups are currently working on developing alternative regularization procedures, which allow sharp discontinuities, or multiscale feature. See for example in Loris *et al.* (2007, 2010), Hu *et al.* (2009). Another well known limitation of regularization procedures is that they often result in artificially small model uncertainty estimates. This is because regularization stabilizes the model construction process at the cost of biasing the solution in a statistical sense (Aster *et al.* 2005).

1.2 A brief review of multiresolution methods

To maximize the amount of information extracted from the data, several studies have used irregular parametrizations of non-overlapping cells, where the grid size is based on some criteria prior to inversion (Abers & Roecker 1991; Fukao *et al.* 1992; Sambridge *et al.* 1995; Sambridge & Gudmundsson 1998; Spakman & Bijwaard 1998; Bijwaard *et al.* 1998; Boschi *et al.* 2004; Zhang & Thurber 2005; Nolet & Montelli 2005; Li *et al.* 2008; Boschi *et al.* 2009; Schiavardi & Morelli 2009). In most of these studies, the irregular mesh is chosen to match variations in ray path density (quantified by hit count). However, hit count is not always an ideal indicator of resolvability, particularly when considering clusters of seismic rays travelling in a near uniform direction. Schaefer *et al.* (2011) address this issue with an adaptive-voxel parametrization, with voxel size also accounting for azimuthal coverage. Note also that cell boundaries can be deformed to eliminate rank deficiency

in the tomographic system of equations (i.e. the null space energy) (Vesnaver 1996) or to minimize the condition number of this system of equations (Curtis & Snieder 1997).

Instead of building an irregular grid prior to inversion, another possibility is to dynamically adapt the grid during the inversion process, and thus the tomographic system of equations is resolved for each new parametrization. For example, Michelini (1995) proposed a scheme with simultaneous determination of seismic velocities and nodes perturbations. There is a natural tendency for nodes to cluster where velocity is changing more rapidly, but the method does not lead to a denser node distribution where the sampling is denser. In a larger scale problem, Sambridge & Faletic (2003) proposed an iterative method where model nodes are added in places where amplitude changes from the previous step are higher than a predefined criterion. However, the optimal criterion for parameter refinement is uncertain.

Although irregular parametrization schemes dramatically reduce the size of the tomographic system and yield more uniformly resolved parameters, the inverse problem often remains ill-posed and non-unique and smoothing or norm damping are still needed. Determining the optimum regularization operators and corresponding weights remains difficult and often speculative. Another problem is the need of specific forward solvers to compute sensitivity kernels on irregular meshes, which generally introduce additional difficulties at the implementation stage (Sambridge & Rawlinson 2005).

Alternatively, the uneven sampling of the Earth can be addressed by expressing the inverse problem on overlapping regular grids of different scales, for example, through a wavelet parametrization (Chiao & Kuo 2001; Chiao & Liang 2003; Loris *et al.* 2007; Tikhotsky & Achauer 2008; Delost *et al.* 2008; Gung *et al.* 2009; You *et al.* 2010; Hung *et al.* 2010). Wavelet decomposition is particularly convenient in multiscale tomography as it enables one to account for both spatial and spectral resolution while avoiding the choice of an irregular grid. Recently, Simmons *et al.* (2011) proposed a multilevel tessellation inversion where a hierarchy is defined between regular grids of different scales. In their scheme, high resolution solutions can be expressed in terms of perturbations of low resolution solutions. The parameters associated with the coarse grid carry heavier weights than the parameters associated with the base level grid subdivisions. Note that many other multiresolution methods for seismic tomography exist in the literature, and we refer the reader to the review papers by Sambridge & Rawlinson (2005) and Thurber & Ritsema (2007).

Although multiresolution methods enable automatic inference of the relative resolution between different areas of a model, the overall number of parameters (or the size of the finest grid) is still to be defined. The question is when do we stop subdividing cells, or adding more unknowns into the problem? All the aforementioned algorithms are based on an optimization procedure (i.e. least-squares fit), and the level of data fit can always be improved by adding more unknowns. Some methods can be used to find the 'optimal' model complexity (or regularization parameters) such as the chi-square statistical test or the discrepancy principle (Aster *et al.* 2005). However, these schemes operate within a given fixed dimension for the model, and hence several tests with different degrees of freedom need to be carried out separately. Furthermore, they rely on knowing the level of data noise in advance, which is difficult in seismology.

1.3 A transdimensional framework

The difficulties outlined above can be addressed with a transdimensional Bayesian formulation. In Bodin & Sambridge (2009),

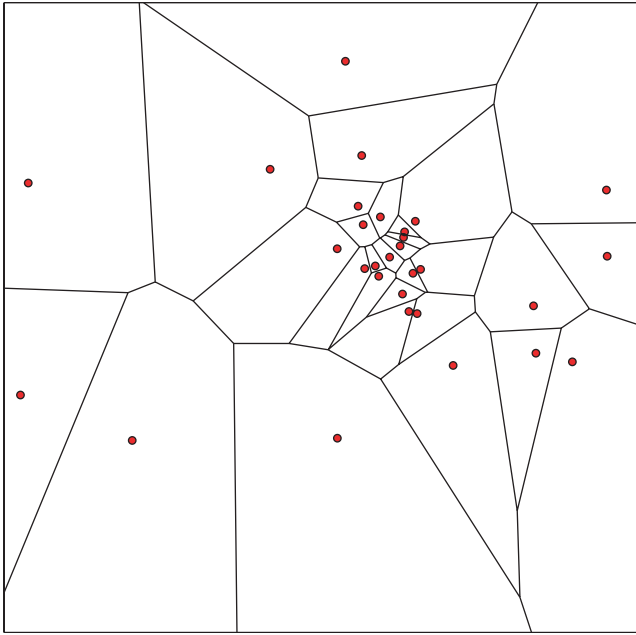


Figure 1. Example of a multiscale Voronoi diagram (black), which forms a set of irregular cells that partition the plane. Any point inside a cell is closer to the node of that cell than any other node, so the shape of the parametrization is entirely defined by the location of nodes (red dots). In transdimensional tomography (Bodin & Sambridge 2009), the model is parametrized with a variable number of nodes and a constant velocity value is assigned to each cell. The solution is a large number of Voronoi models with variable parametrization distributed according to the posterior probability function. A smooth solution model is produced for interpretation by pointwise averaging this large ensemble of spatially partitioned models.

we proposed the transdimensional tomography algorithm which is based on a fully nonlinear stochastic search for model complexity, parameter definitions, and seismic models. Although it is currently too computationally expensive for 3-D global problems, our transdimensional approach to tomography has proven to be feasible for regional scale 2-D problems. As showed in Fig. 1, the velocity (or slowness) field is partitioned by a variable number of non-overlapping cells defined by a set of Voronoi nodes (Voronoi 1908; Okabe *et al.* 1992). The number and the position of the cells, as well as the velocity field itself are unknowns in the inversion.

In a Bayesian formulation of the inverse problem, the knowledge concerning the unknown model parameters is given by the *a posteriori* probability distribution (or posterior distribution) $p(\mathbf{m}|\mathbf{d})$ which is the probability density of the model parameters given the observed data (Tarantola & Valette 1982). While optimization schemes only provide a single model, that is, the best combination of parameters that maximize the fit to observations, the Bayesian solution is a large ensemble of models, whose density is directly proportional to the posterior distribution. This ensemble solution contains much more information than a single model, and allows quantitative assessment of model parameter uncertainties, resolution and non-uniqueness (see Khan *et al.* 2011, for a recent example of Bayesian tomography).

In transdimensional tomography (Bodin *et al.* 2009; Bodin & Sambridge 2009), the number of cells is an unknown, and so the dimension of the model space is itself a variable. In this case the posterior distribution becomes a transdimensional function. A fundamental property of Bayesian inference is the preference for the least complex explanation for an observation, referred to in the lit-

erature as ‘natural parsimony’ (Malinverno 2002). Given a choice between a simple model with fewer unknowns and a more complex model that provides a similar fit to data, the simpler one will be favoured (see MacKay 2003, for a discussion). In the last decade, transdimensional inversion has become popular in Earth sciences and has been applied to a wide range of areas (e.g. Malinverno 2002; Stephenson *et al.* 2004, 2006; Jasra *et al.* 2006; Hopcroft *et al.* 2009; Charvin *et al.* 2009; Gallagher *et al.* 2011; Dettmer *et al.* 2010; Luo 2010; Piana Agostinetti & Malinverno 2010; Minsley 2011; Bodin *et al.* 2012).

In transdimensional tomography, the solution is represented by a large ensemble of Voronoi models with variable parameterisations. At each pixel of the 2-D velocity map, one can compute the mean and standard deviation over all the models in the ensemble solution. When these posterior expectations are computed, models with variable geometries overlap providing continuous smooth solution and uncertainty maps that have a better resolution than any single model. This way, information is extracted from the ensemble as a whole, and the approach provides a parsimonious solution with no need for explicit regularization. For a detailed description of transdimensional tomography, the reader is referred to Bodin & Sambridge (2009).

In this study, we present an improved version of transdimensional tomography, and use it to simultaneously invert different data sets that span the Australian crust at various scales. The Rayleigh wave group traveltimes data set from Saygin & Kennett (2010) is inverted in conjunction with similar cross-correlation traveltimes obtained from the WOMBAT experiment which sample the southeast part of the continent at much higher resolution (Arroucau *et al.* 2010; Young *et al.* 2011). The data sets used here are different both in size (i.e. the number of picked traveltimes) and in the scale of the region they sample (Fig. 2). Although considerable detail might be resolvable for densely sampled areas, only relative large-scale information is available for sparsely sampled regions. The goal is to see whether additional information can be revealed by simultaneously inverting data sets that would not be easy to combine using a standard approach.

1.4 Uncertain data noise and Hierarchical Bayes

An important feature of transdimensional schemes is that the level of data uncertainty estimated by the user prior to inversion (i.e. the covariance matrix of data errors) directly affects the complexity of the solution, that is, the number of model parameters. Note that data uncertainties contain two terms: the errors done on the measurements but also modelling errors, that is, the inability of our forward model to explain the data (Gouveia & Scales 1998). Therefore in this work we follow the definition of Scales & Snieder (1998) and define data noise as ‘that part of the data that we do not expect the model to explain’, and hence the residuals (difference between estimated data and observations) are viewed as a realization of a random data noise.

The level of data noise determines the required level of data fit, and high dimensional (many cell) models that fit the data more than this level are given a low probability (see Dettmer *et al.* 2010; Piana Agostinetti & Malinverno 2010, for recent examples). A key issue of transdimensional inversion is therefore the quantification of data noise, which is often difficult. This is especially true in the case of joint inversion of multiple data sets, where noise levels effectively control the influence of different data types.

Using ambient-noise surface wave tomography, seismologists have constructed maps of seismic wave velocity (e.g. Shapiro *et al.*

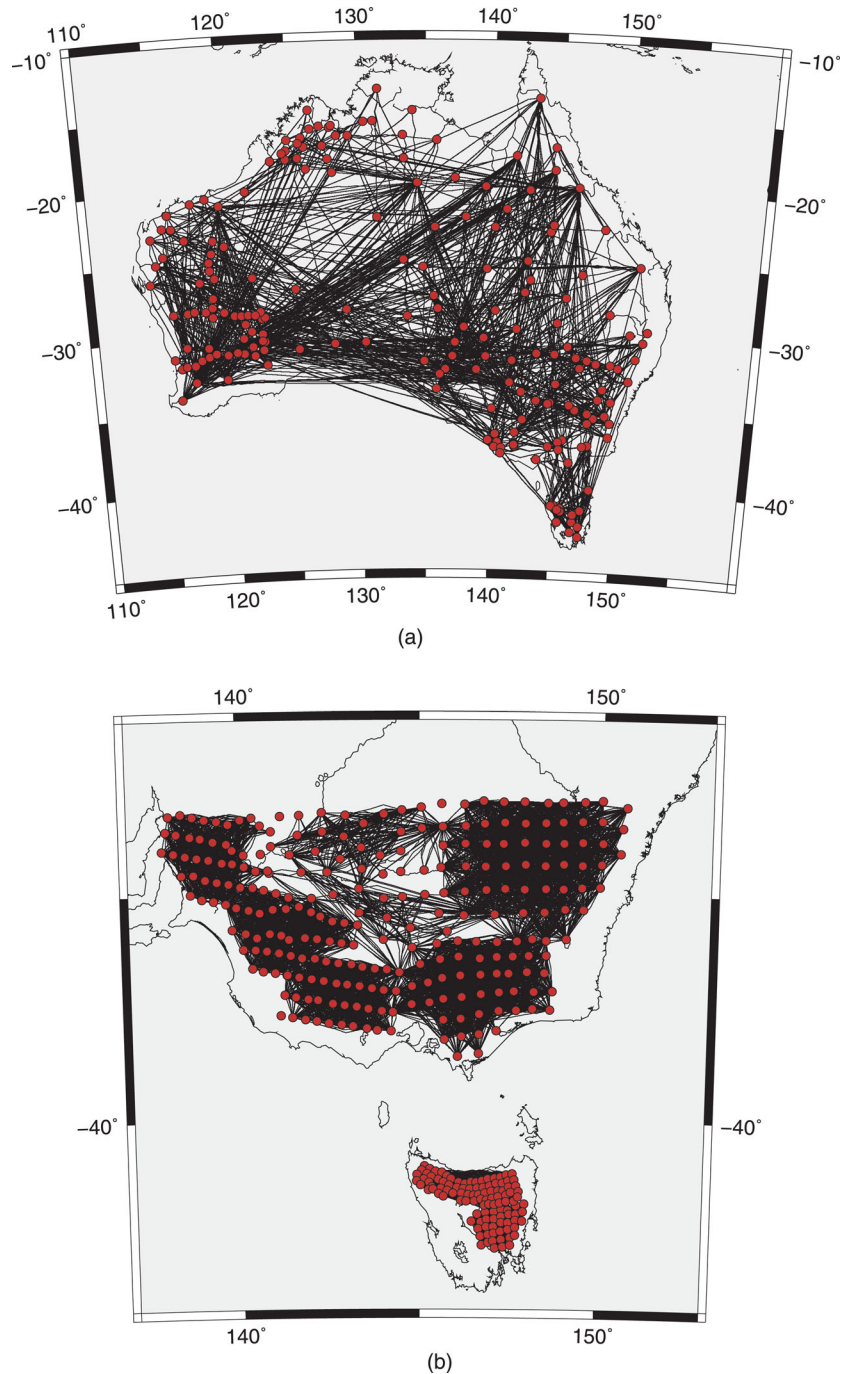


Figure 2. Ray coverage for the three ambient noise data sets which sample the Australian crust at different scales. Black lines indicate rays between simultaneously recording station pairs (red circles). Top panel: large scale data set. Bottom panel: Wombat Experiment that can be divided into two subarrays (mainland southeast Australia and Tasmania).

2005; Sabra *et al.* 2005; Yang *et al.* 2006; Villasenor *et al.* 2007; Brenguier *et al.* 2008; Stehly *et al.* 2009; Yao & Van der Hilst 2009) in many different regions of the world and at a variety of scales. Although such techniques are now well established and widely used, the uncertainties associated with ambient noise measurements are still poorly understood. While recent studies propose to quantitatively measure the bias due to different effects (Larose *et al.* 2008; Weaver *et al.* 2009, 2011; Yao & Van der Hilst 2009), this is ongoing work and there is presently no consensus on the way to quantify errors.

In this multiscale study, interstation distances, ambient noise recording period, azimuthal distribution of noise and processing techniques differ with each data set. Therefore the uncertainty on the computed traveltimes may vary considerably between data sets as well as within a data set. Without any reliable estimation of data noise, we shall show that it appears difficult to directly use the conventional tomography as described in Bodin & Sambridge (2009).

Hence, we here propose an improved version of the algorithm that can address the issue of noise estimation. The Bayesian formulation

is extended to hierarchical models which allow data error to be an unknown parameter (Malinverno & Briggs 2004; Malinverno & Parker 2006). Overall, a joint posterior probability distribution is defined both for data noise parameters and Earth parameters. The result is that the arbitrary choice of a weighting factor to balance different data sets in the misfit function is avoided, and instead this aspect is driven by the data itself, thereby allowing a probabilistic velocity model to be obtained.

The purpose of this paper is twofold. First, we show that the improved version of transdimensional tomography can be used to jointly invert different data sets in a multiple scale tomography problem. Secondly, we show that using a broad prior distribution for the data uncertainties and letting the data infer its own level of noise is a new and powerful feature of the extended algorithm. We first present the three ambient noise data sets, and then establish the necessity of the Hierarchical Bayes formulation. A synthetic experiment is then carried out prior to simultaneously invert the three observational data sets to construct a detailed map of Rayleigh wave group velocity at 5 s for Australia.

2 THE DATA

We use the 5 s Rayleigh wave group traveltime data set computed from ambient noise by Saygin & Kennett (2010) (this data set will be referred to as the large scale data set) in conjunction with similar cross-correlation traveltimes obtained from the WOMBAT experiment by Arroucau *et al.* (2010) and Young *et al.* (2011) (Fig. 2).

Saygin & Kennett (2010) compiled all the seismic broadband data from temporary and permanent stations across the Australian continent from 1992 to 2006. They used a large volume of recorded noise, which comes from diffuse sources of seismicity such as oceanic and atmospheric disturbances, and computed the cross-correlation on the vertical component for all simultaneously recording station pairs (see Bensen *et al.* 2007, for details about the data processing scheme). The extracted waveforms exhibit a dispersed wave-train which can be interpreted as the Rayleigh wave component of the Green's function (Campillo & Paul 2003; Shapiro & Campillo 2004). For different frequencies, the Rayleigh wave arrival time is picked on the envelope of the bandpass filtered seismogram. We note here that there is no information available on the data uncertainty for the large scale data set shown in Fig. 2(a).

WOMBAT is an extensive program of temporary seismic array deployments throughout southeast mainland Australia and Tasmania. Each array consists of between 30 to 60 short period instruments that continuously record for between 5 to 10 months. Over the last decade, a total of over 550 sites have been occupied resulting in a very large passive seismic data set that has been used for several studies (e.g. Graeber *et al.* 2002; Rawlinson & Urvoy 2006; Rawlinson *et al.* 2006; Clifford *et al.* 2007; Rawlinson & Kennett 2008; Rawlinson *et al.* 2010, 2011; Tkalčić *et al.* 2011; Young *et al.* 2011). Recently, Arroucau *et al.* (2010) computed cross-correlograms using a subset of WOMBAT array data in a similar manner as Saygin & Kennett (2010). However, the technique used to determine Rayleigh wave group traveltimes from the obtained cross-correlograms was slightly different from the one used for the large scale data set by Saygin & Kennett (2010). In a first stage, preliminary dispersion curves for periods ranging from 1 to 20 s were constructed and averaged to build a phase-matched filter (Levshin & Ritzwoller 2001) which was subsequently applied to the seismograms prior to a second round of traveltime picking. This allowed additional dispersion measurements to be included that were discarded during the first

stage due to low signal-to-noise ratio. Relative uncertainties for the picked traveltimes were obtained following the procedure presented in Cotte & Laske (2002) and Harmon *et al.* (2007).

Arrays were not deployed at the same time in Tasmania and in mainland southeast Australia. As a consequence, there are no apparent traveltimes available between Tasmanian and mainland stations. Furthermore, the average interstation distance for the mainland arrays is about 50 km, while it is only 15 km in Tasmania. Hence the WOMBAT data can be divided into two subsets corresponding to two separate regions with different scales: mainland southeast Australia and Tasmania (see Fig. 2a).

The traveltimes obtained using WOMBAT arrays were used to generate tomographic maps for each frequency. Due to the difference in ray densities, data from mainland arrays and Tasmania were inverted separately with two different grid sizes. The internode distance for mainland regions (20 km) was four times larger than that for Tasmania (5 km). Furthermore, the regularization parameters (smoothing and damping) were different for the two inversions.

In this study we carry out a joint inversion of these three data sets which have different scales and have been processed separately. Each data set samples a particular region (whole Australia, mainland southeast Australia and Tasmania) and is characterized with a scale length defined by its average ray length (respectively 200 km, 50 km, 15 km). We note that, while the two WOMBAT data sets independently sample two separate regions (Fig. 2a), the largest scale data set spans the whole continent and includes areas sampled by the smaller arrays (Fig. 2a).

3 WHY HIERARCHICAL BAYES?

In Bodin & Sambridge (2009), the largest scale data set was inverted with the conventional transdimensional tomography algorithm. The transdimensional posterior probability distribution was defined with Bayes' theorem (Bayes 1958) which combines prior information on the model \mathbf{m} with the observed data \mathbf{d} :

$$p(\mathbf{m}|\mathbf{d}) \propto p(\mathbf{d}|\mathbf{m}) \times p(\mathbf{m}), \quad (1)$$

where $a|b$ means a given, or conditional on, b . Hence, $p(a|b)$ is the probability of having a when b is fixed. $p(\mathbf{m})$ is the prior probability distribution (what we know about the model before measuring the data) which is set uniform with relatively wide bounds to assume unobstructive prior knowledge.

The likelihood function $p(\mathbf{d}|\mathbf{m})$ determines how well a given model with a particular set of parameter values can reproduce the observed data. Here it is based on a Gaussian distribution which quantifies the agreement between simulated and observed data (Sambridge *et al.* 2006):

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\sqrt{(2\pi)^N |\mathbf{C}_e|}} \times \exp \left\{ -\frac{1}{2} \left([\mathbf{g}(\mathbf{m}) - \mathbf{d}]^T \mathbf{C}_e^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}] \right) \right\}, \quad (2)$$

where $\mathbf{g}(\mathbf{m})$ is the vector (of length N) of traveltimes estimated for the seismic model \mathbf{m} , and \mathbf{C}_e is the covariance matrix of data noise, which represents data uncertainty and the covariability between data errors ($|\mathbf{C}_e|$ is the determinant of this matrix). The model vector \mathbf{m} has a variable number of components, and the number of Voronoi cells needed to explain the data is directly determined by the estimated data noise \mathbf{C}_e (see Figs 3 and 4). Because the level of detail in the solution is automatically adjusted as a function

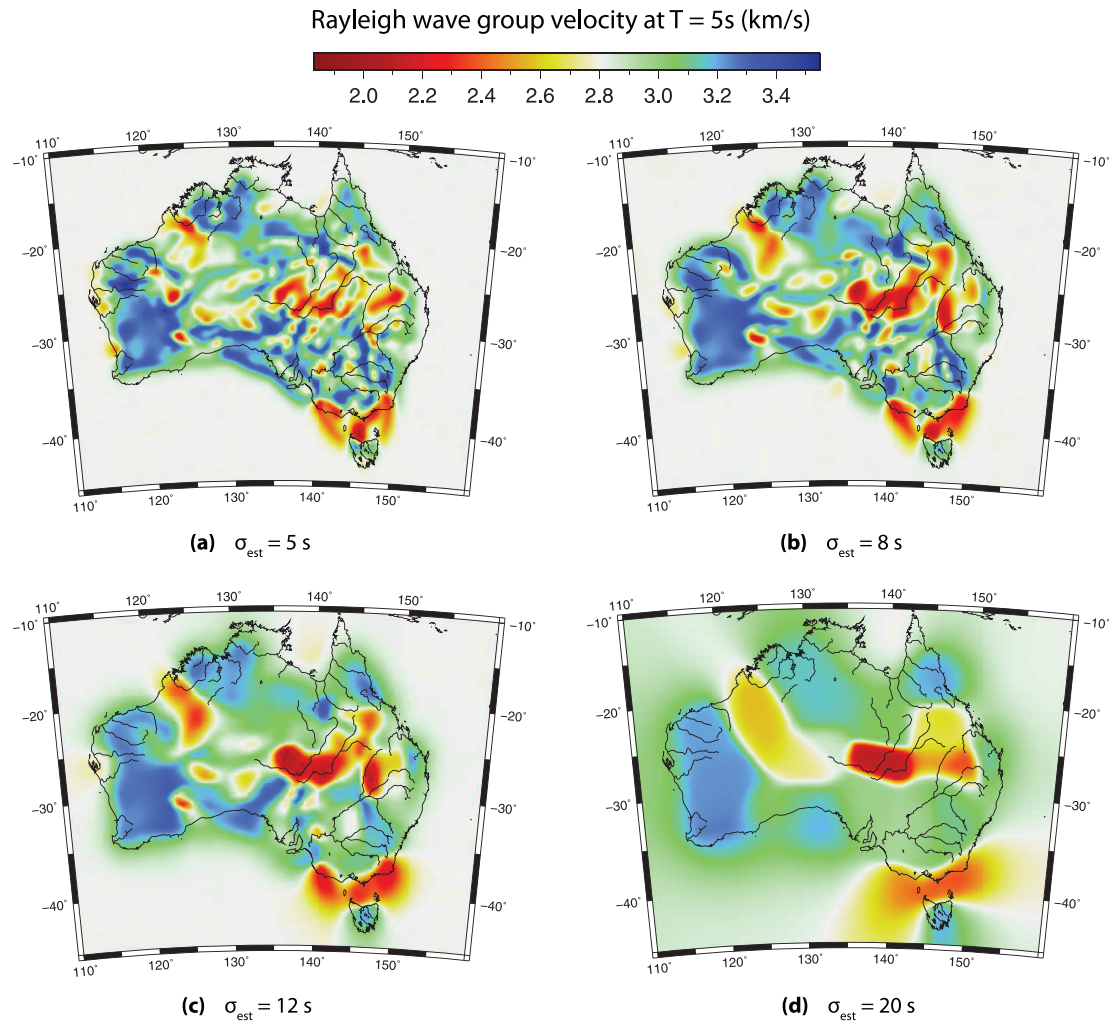


Figure 3. Conventional transdimensional tomography (Bodin & Sambridge 2009) with ambient noise cross-correlation for the Australian continent (large scale data set). Average solution model (km s^{-1}) obtained with four different (but realistic) values for σ_{est} .

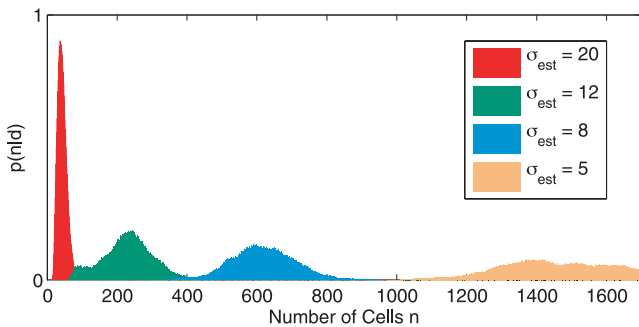


Figure 4. Posterior distribution on the number of Voronoi cells corresponding to the four solutions shown in Fig. 3 obtained using conventional transdimensional tomography. These distributions are simply obtained by plotting the histogram of the number of cells of sampled models across the ensemble solution. Here it is clear that as the estimated data noise given by the user decreases, model complexity increases.

of the required data fit, this has been seen as an advantage over optimization based inversions where the level of data noise is not accounted for and the complexity of the solution (i.e. the number of model parameter or level of smoothing) is manually adjusted by the user.

Indeed it is worth noting that the general least squares solution used in linear inverse problems (where $\mathbf{d} = \mathbf{G}\mathbf{m}$) is given by

$$\mathbf{m}_{L_2} = (\mathbf{G}^T \mathbf{C}_e^{-1} \mathbf{G})^{-1} \mathbf{G}^T \mathbf{C}_e^{-1} \mathbf{d}. \quad (3)$$

Here \mathbf{m}_{L_2} does not depend on the absolute value of the data noise. That is, \mathbf{m}_{L_2} does not change when \mathbf{C}_e is multiplied by a constant factor. In fact, only the estimated error on the model (i.e. the posterior model covariance matrix) depends on the data noise. In most traveltimes tomography studies, the linear system of equation \mathbf{G} is regularized and the damped least square solution can be written as:

$$\mathbf{m}_{L_2} = (\mathbf{G}^T \mathbf{C}_e^{-1} \mathbf{G} + \mu \mathbf{C}_m^{-1})^{-1} (\mathbf{G}^T \mathbf{C}_e^{-1} \mathbf{d} + \mu \mathbf{C}_m^{-1} \mathbf{m}_0), \quad (4)$$

where \mathbf{m}_0 is a reference model, \mathbf{C}_m is the *a priori* model covariance matrix and μ a regularization parameter that has to be manually chosen by the user (Aster *et al.* 2005). In this case, a scaling factor in \mathbf{C}_e simply absorbs into the regularization parameter. That is, if we multiply \mathbf{C}_e and divide μ by the same factor, the solution \mathbf{m}_{L_2} is unchanged. Hence, most tomographic models obtained with optimization schemes do not directly depend on the magnitude of data noise but only on relative uncertainties. For example, Steck *et al.* (1998) used relative weights (1, 2, 3 and 4) to describe the diagonal elements of \mathbf{C}_e .

Here the large scale data set has been inverted with the conventional transdimensional algorithm (Bodin & Sambridge 2009),

where data errors are assumed to be independent and normally distributed with zero mean and a fixed standard deviation σ_{est} , that is, \mathbf{C}_e is diagonal with all diagonal elements equal to σ_{est}^2 . The solution maps obtained for four different values of σ_{est} are shown in Fig. 3. The posterior distribution on the number of cells (i.e. the histogram on model dimension across the ensemble solution) for each map is shown in Fig. 4. There is clearly a strong correlation between σ_{est} and the number of cells of the sampled models. As the assigned data error decreases, more cells are added in the model and the solution maps show more complexity. Here the model complexity is adapted automatically to fit the data up to the given level of noise. In the absence of information about the data noise, it is impossible to give a preference to any of these four solutions. The details appearing in 3(a) could be unrealistic and noise induced as a result of the data being ‘over-fitted’. Conversely, in Fig. 3(d), data might have been considered too noisy and uninformative and consequently the solution model may be missing finer details. This shows that by choosing σ_{est} , we also choose the model complexity.

4 HIERARCHICAL BAYES ALGORITHM

By adding the Hierarchical Bayes formulation to the algorithm and by treating \mathbf{C}_e as unknown, we let the data infer its own degree of uncertainty without imposing any fixed value for the required data fit. In Fig. 3 the model complexity is spatially variable but the algorithm operates within a fixed data noise level. By freeing up this constraint and also treating the data variance as an unknown, we allow the overall model complexity level to be driven by the data.

If there are N data in the problem, \mathbf{C}_e is a symmetric $N \times N$ matrix defined with $(N^2 + N)/2$ independent values which are obviously impossible to estimate separately from only N data. Hence some assumptions need to be made in order to ‘parameterize’ the matrix of data errors. Following current statistical terminology, the data noise covariance matrix is expressed with a number of ‘hyper-parameters’ $\mathbf{C}_e = f(h_1, h_2, \dots)$ (Gelman *et al.* 2004). For example, assuming an independent (i.e. not correlated) and invariant Gaussian random noise, the noise covariance matrix can be considered proportional to the identity matrix \mathbf{I}_N and can be parameterized with a single hyperparameter $h_1 = \sigma$ representing the standard deviation of data errors ($\mathbf{C}_e = \sigma^2 \mathbf{I}_N$). In principle any scale factor removed from \mathbf{C}_e can be treated in a similar manner.

The model to be inverted for is defined by the combined set $\mathbf{m} = [\mathbf{c}, \mathbf{v}, n, \mathbf{h}]$, where \mathbf{c} is the vector containing the nodes of Voronoi cells (Fig. 1), and \mathbf{v} contains the constant velocity values assigned to each cell. The vectors \mathbf{c} and \mathbf{v} are of size n (i.e. the number of Voronoi cells) which is itself a variable parameter [see Bodin & Sambridge (2009) for a complete description of the model parameterization]. In this study we include as unknowns the vector of hyper-parameters defining data errors $\mathbf{h} = (h_1, h_2, \dots)$. Each component of \mathbf{m} is given a wide uniform prior distribution. In this way, the final models will be dominated by the data rather than by prior information. In the following sections we show how we define noise parameters \mathbf{h} in \mathbf{C}_e .

At a coarse level, the algorithm can be summarized by Fig. 5. At each iteration, ray paths are determined in the current continuous reference velocity model using the fast marching method (Sethian & Popovici 1999; Rawlinson & Sambridge 2004) (prior to the first iteration this is a laterally homogeneous reference model). At each successive iteration the Hierarchical Bayes transdimensional algorithm produces a large number of Voronoi models $\mathbf{m} = [\mathbf{c}, \mathbf{v}, n, \mathbf{h}]$

with variable dimensions from which Bayesian statistics can be derived (e.g. expected number of cells, expected level of data noise, etc. . . .). Then, the reference model is updated by spatially averaging the entire ensemble of Voronoi models. Each Voronoi model has discontinuities throughout the velocity field, but the ensemble average tends to be spatially smooth with continuously varying gradients.

Given a fixed geometry of rays, the Hierarchical Bayes transdimensional algorithm is a method for obtaining a sequence of random samples \mathbf{m} , where each is a perturbation of the last, and which are distributed according to the transdimensional posterior distribution $p(\mathbf{m}|\mathbf{d})$. The approach is based on a generalization of the well known Metropolis-Hastings algorithm (Metropolis *et al.* 1953; Hastings 1970) termed the reversible-jump Markov chain Monte Carlo (rj-McMC) sampler (Geyer & Møller 1994; Green 1995, 2003). The Hierarchical Bayes algorithm is implemented in the same manner as the conventional transdimensional tomography, the only difference being that here we add an extra type of model perturbation in the Markov chain, that is, a change in the hyperparameter vector \mathbf{h} . Having randomly initialized the model parameters $\mathbf{m} = [\mathbf{c}, \mathbf{v}, n, \mathbf{h}]$ by drawing values from the prior distribution of each parameters, the algorithm proceeds iteratively. Each step of the Markov chain is divided into three stages:

(1) Propose a new model by drawing from a probability distribution $q(\mathbf{m}'|\mathbf{m})$ such that the proposed model \mathbf{m}' is conditional only on the current model \mathbf{m} . This involves one randomly selected type of change, with probability 1/5, out of five possible:

(i) **Change a velocity value:** randomly pick one Voronoi cell (from a uniform distribution) and change the velocity value assigned to this cell according to a Gaussian probability distribution $q(v'_i|v_i)$ centred at the current value v_i .

(ii) **Move a Voronoi node:** randomly pick on Voronoi cell, and randomly perturb the position of its node according to a 2-D Gaussian proposal probability density $q(\mathbf{c}'_i|\mathbf{c}_i)$ centred at the current position \mathbf{c}_i .

(iii) **Birth:** create a new Voronoi cell by randomly drawing a point in the 2-D map. A velocity value needs to be assigned to the new cell. This is drawn from a Gaussian proposal probability density centred on the current velocity value where the birth takes place.

(iv) **Death:** delete a Voronoi node chosen randomly from the current set of n cells.

(v) **Change one noise parameter:** randomly pick one component of the vector \mathbf{h} , and randomly perturb its value according to a Gaussian proposal probability density $q(h'_i|h_i)$ centred at the current value h_i .

(2) Compute traveltimes estimated for the proposed Voronoi model. The new estimated traveltimes $g(\mathbf{m}')$ are compared with the observations \mathbf{d} to build the likelihood (2) and the posterior value of the proposed model $p(\mathbf{m}'|\mathbf{d})$.

(3) Randomly accept or reject the proposed model (in terms of replacing the current model), according to the acceptance criterion ratio given in eq. (18) of Bodin & Sambridge (2009). When the proposed model is rejected, the current model is retained for the next step and also added again to output ensemble.

The first part of the chain (called the burn-in period) is discarded, after which the random walk is assumed to be stationary and starts to produce a type of ‘importance sampling’ of the model space (Green 1995, 2003). This means that models generated by the chain are asymptotically distributed according to the posterior probability distribution $p(\mathbf{m}|\mathbf{d})$ (see Gallagher *et al.* 2009, for a description of the Metropolis sampler).

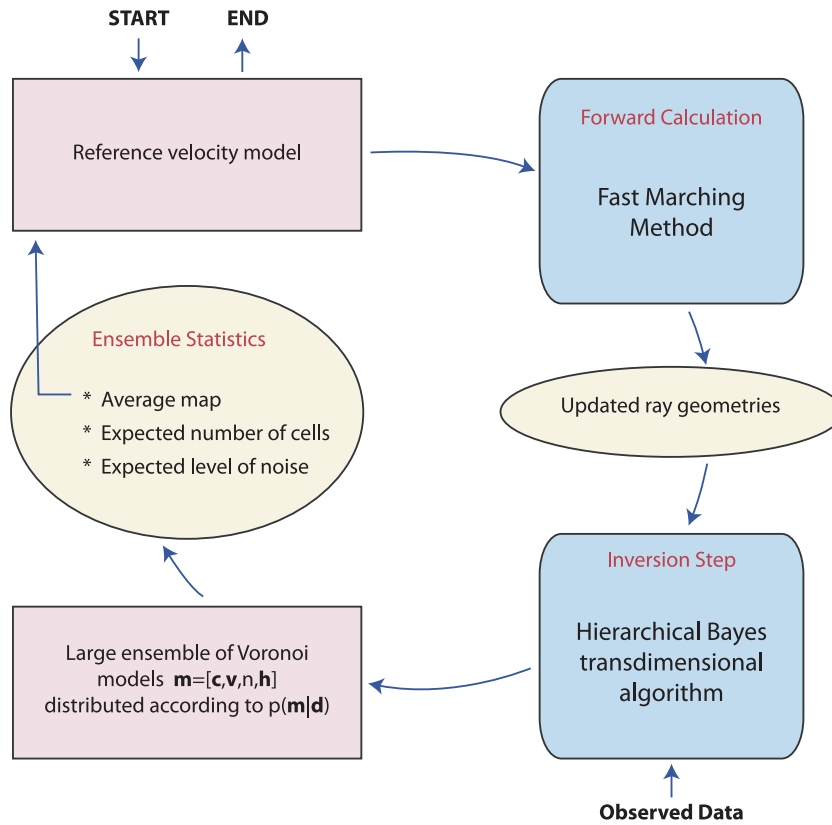


Figure 5. The Hierarchical Bayes transdimensional algorithm is used in conjunction with Fast Marching eikonal solver to build an iterative tomography that takes into account ray bending. At each iteration, the Hierarchical Bayes algorithm produces an ensemble of models with variable dimensions from which posterior expectations can be extracted. The ensemble solution of Voronoi models are spatially averaged to produce a reference model for the next iteration.

When treating the data noise as a variable, one might intuitively expect the algorithm to choose high values for the variance of data noise (i.e. the diagonal elements of C_e in (2)) as this would minimize the misfit and increase the likelihood. However, note that the Gaussian likelihood function in (2) is normalized by the determinant of the data covariance matrix, and hence large data errors would also tend to reduce the likelihood. Hence in a sampling scheme the values taken by the magnitude of data noise will be a balance between these two competing effects.

In our multiscale problem, data noise is likely to vary between data sets due to variations of spectral and azimuthal characteristics of the noise field on different regions and at different spatial scales. For example, the azimuthal distribution of sources is much more even in Tasmania than for continental stations due to the close proximity of oceanic microseisms on all sides of the island (Young *et al.* 2011). The Hierarchical Bayes formulation can account for this by independently treating the uncertainty on each data set. For example, the levels of noise for the large scale data set and for WOMBAT arrays can be separately represented by two scale factors to be inverted for [i.e. $\mathbf{h} = (\sigma_1, \sigma_2)$]. In this way, the inversion procedure can consistently combine data sets with differing levels of uncertainty.

5 A SYNTHETIC TEST

5.1 Experimental setup

A synthetic multiscale checkboard velocity model (Fig. 6) is constructed in which the square size is approximately proportional to

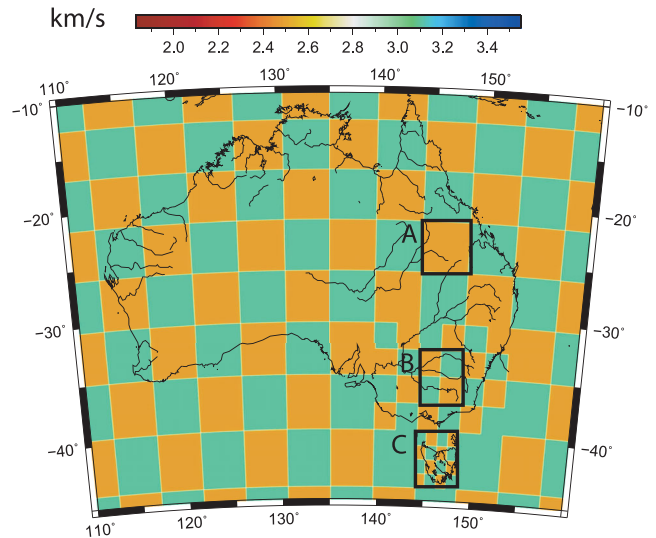


Figure 6. Synthetic velocity model (km s^{-1}) with a multiscale resolution. The three regions A, B and C have equal area but different spatial scales of velocity structure.

the spatial sampling of the data shown in Fig. 2. The three regions of equal area A, B and C are examples of regions with different unit square. The areas in orange have a velocity of 2.5 km s^{-1} and the green have a velocity of 3.1 km s^{-1} which makes the tomographic problem fairly non-linear. Three synthetic traveltime data sets corresponding to the same configurations as shown before are

constructed by using the Fast Marching Method (Sethian & Popovici 1999; Rawlinson & Sambridge 2004).

As explained later, we think that it is realistic to have smaller uncertainties at shorter interstation distances. Consequently, some random Gaussian noise has been added to the synthetic traveltimes with a standard deviation of 4 s for the large scale data set and of 1 s for the WOMBAT arrays. The hierarchical formulation will attempt to recover these noise values along with the model and its complexity.

5.2 Data noise hyperparameters

We use the Hierarchical Bayes formulation and invert for two noise hyperparameters: σ_1 for the large scale data set and σ_2 for the WOMBAT arrays (southeast Australia + Tasmania). Hence, the likelihood function takes the form:

$$p(\mathbf{d}|\mathbf{m}) = \frac{1}{\prod_{i=1}^N (\sqrt{2\pi}\sigma'_i)} \times \exp \left\{ \sum_{i=1}^N \frac{-(g(\mathbf{m})_i - \mathbf{d}_i)^2}{2(\sigma'_i)^2} \right\}, \quad (5)$$

where

$$\begin{cases} \sigma'_i = \sigma_1 & \text{for data belonging to the largest scale set.} \\ \sigma'_i = \sigma_2 & \text{for WOMBAT data.} \end{cases} \quad (6)$$

The model parameters (\mathbf{c} , \mathbf{v} , n , σ_1 , σ_2), where n is the number of Voronoi cells, are successively perturbed along the Markov chain to collect an ensemble of models that samples the posterior probability distribution.

5.3 Results

Posterior inference was made using an ensemble of around about 4×10^5 models. A total of 96 Markov chains were run in parallel. Each chain was run for 2×10^6 steps. The first million were discarded as burn-in steps, only after which the sampling algorithm was judged to have converged. Then, every 250th model visited was taken in the ensemble. Three passes were made around the general tomography algorithm with an update of the ray geometry for each pass (see Fig. 5).

The spatial average of the post burn-in samples collected during the last iteration is shown in Fig. 7 and clearly recovers features of the true velocity field at different scales. By retrieving the different sizes of the checkerboard squares, the parametrization has been able to adapt to the underlying structure of the model as well as to the spatial distribution of rays. The solution model varies smoothly in regions of limited coverage like region A without sacrificing any sharp or small-scale features in well-sampled regions like region C.

The uniform prior distributions on both hyperparameters σ_1 and σ_2 were defined over the range [0.5 7] s. Their posterior distribution after each pass are shown in Fig. 8. For the last iteration, the posterior maxima are close to the true value of data noise (relative to the wide range of the prior). The posterior maxima for σ_1 is about four times as large as σ_2 which is the ratio between the two true values of data noise. Note that the posterior maxima for both hyperparameters would be closer to the true value if a Jeffreys prior (i.e. $1/\sigma$) was used instead of a uniform prior. This is because the standard deviation of the noise is a scale parameter, and a Jeffreys prior would be the proper non-informative prior to use (the same is true for the number of model parameters). However in this study, for simplicity, we always use uniform priors for hyperparameters.

With scant information on the data noise prior to the inversion, the Hierarchical Bayes procedure recovers the standard deviations

of the true data noise (4 s and 1 s). At the same time it provides a parsimonious solution model with a complexity and resolution that varies spatially and locally conforms to the level of information provided by the data. Indeed, the low gradients in regions sampled by the largest data set (e.g. region A in Fig. 7) indicate that observations are less well fit there than under the WOMBAT arrays (e.g. regions B and C) where the discontinuities are better recovered. This can be seen quantitatively in Fig. 8. Note that inferring the level of data noise for each data set is tantamount to inverting for the weighting factor between data sets in a joint inversion. This is because the estimated noise given to a particular data set directly weights the contribution of this data set to the total data fit.

5.4 Hyperparameters and theory errors

In an inverse problem the data noise is usually defined as the difference between the observed and the predicted measurements ($\mathbf{d} - g(\mathbf{m})$). In practice, ‘noise’ is whatever component of the measured data that $g(\mathbf{m})$ cannot account for. See Gouveia & Scales (1998) and Scales & Snieder (1998) for a discussion.

In traveltome tomography, incorrect estimates of ray geometries implies an incorrect forward model, that is, an incorrect function g in eq. (5). This is because by keeping the ray paths fixed, the tomographic problem is linearized in slowness around a reference model. This linear approximation on the forward model may also contribute to the misfit, and should be accounted for as data noise. For example, let us imagine a tomographic problem where traveltome measurements are perfect, but where the ray geometries used in the inversion are completely wrong. In this case it would not be sensible to perfectly fit the data (although measurements are perfect), and hence the covariance matrix of data errors needed in the inversion would have to account for modelling uncertainties.

Fig. 8 illustrates how the hyperparameters σ_1 and σ_2 take into account these modelling errors. The Hierarchical Bayes tomography algorithm was run over iterations (see Fig. 5). At the first iteration, traveltimes for all the sampled models were computed assuming straight rays between station pairs. Then the average of all collected models was used to update ray geometries which were used in the next pass. We show in Fig. 8 the posterior distribution on the two hyperparameters obtained after each of these three passes, that is for three different approximations of the forward model. The values taken by the hyperparameters during the random walk are clearly higher than the true data noise and seem to decrease as the iterations progress. By retracing the rays and iterating the process, the true rays geometries can be better approximated and the expected posterior value of the noise parameters decrease and converge towards the noise value that has been added to the ‘true’ synthetic data. As ray paths are better approximated, the error present in the forward model g decreases and contributes less to the data misfit. Hence the hyperparameters σ_1 and σ_2 effectively quantify the ability of the model to fit the data and take into account both observational and theoretical errors contributions.

5.5 Comparison with subspace inversion

To compare our result to a fixed grid optimization based inversion, we jointly inverted the three data sets with the subspace method (Kennett *et al.* 1988; Rawlinson *et al.* 2006, 2008). Here, the velocity field is defined by a uniform grid of nodes with bicubic B-spline interpolation. The nodes are evenly distributed and do not move during the inversion process. This way of parametrizing the velocity field is common in surface wave tomographic studies Here, the

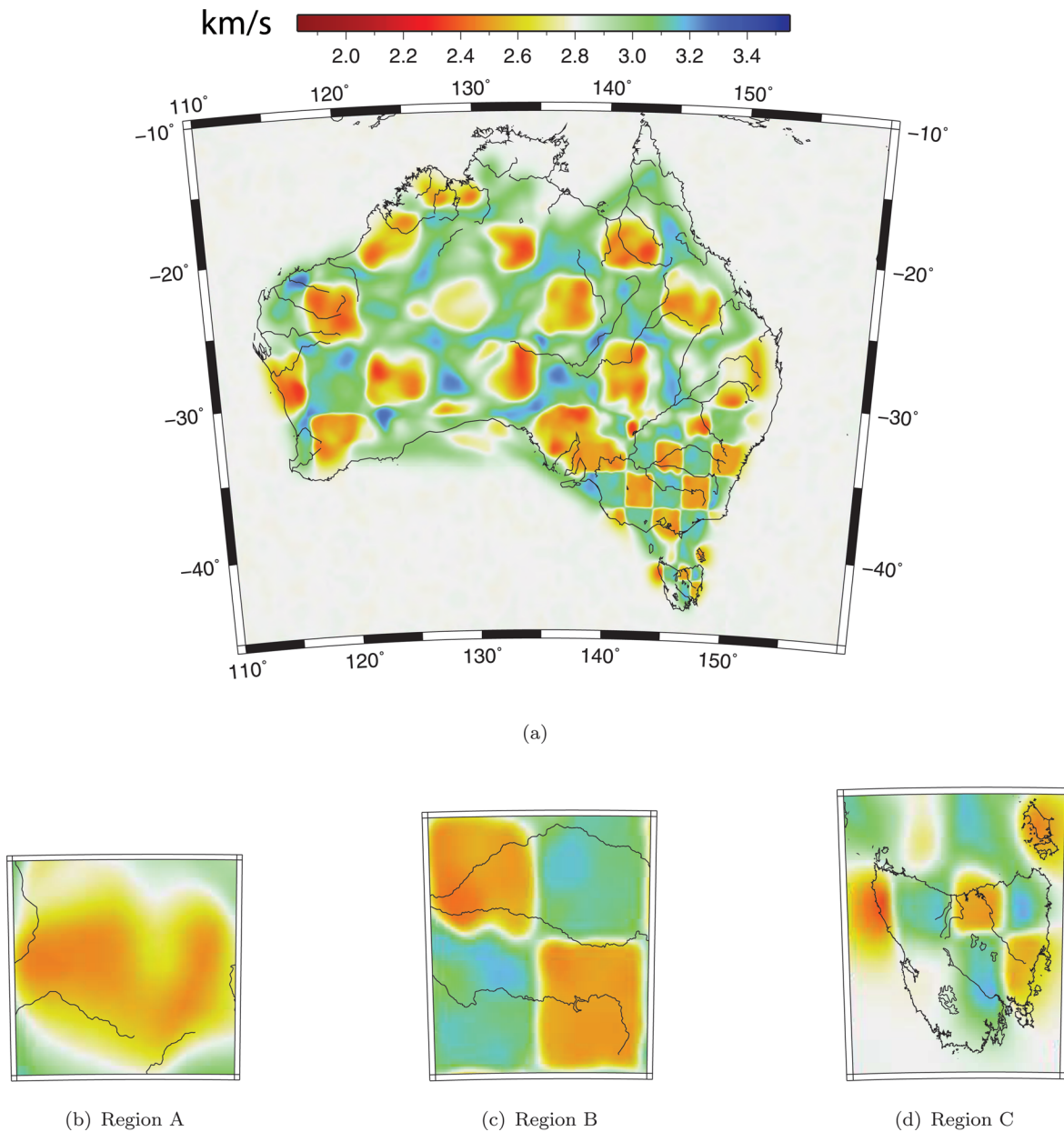


Figure 7. Average solution model obtained with the Hierarchical Bayes transdimensional algorithm by jointly inverting the three data sets. The three lower panels show details for Regions A, B and C.

unknown which is sought for during the inversion step is a perturbation of the reference field. The problem is then locally linearized around the reference model. The method is a subspace inversion because at each iteration, it projects the full linearized inverse problem onto a smaller m -dimensional model space to reduce computational effort. Details of the subspace method are given in Kennett *et al.* (1988), Rawlinson *et al.* (2006), and Rawlinson *et al.* (2008). Note here that the level of data noise is not accounted for as the problem is regularized using *ad hoc* damping and smoothing. Several grid sizes were tried, and for each an iterative L-curve method (Aster *et al.* 2005) that successively ‘tuned’ both damping and smoothing parameters was carried out as in Rawlinson *et al.* (2006).

Fig. 9 shows two solutions obtained with two grid sizes: a coarse parametrization with 56×39 nodes, which equates to a B-spline node separation of about 1° ; and a denser parametrization with a finer grid of 112×78 nodes, that is, an internode separation of

0.45° . The two solutions obtained with the two grid sizes are shown in Fig. 10. The coarse grid solution shown in Fig. 10(a) recovers the amplitudes of the true model relatively well, but misses the sharp discontinuities and the small scale features under WOMBAT arrays. By using a finer grid, in Fig. 10(b) the velocity field in southeast Australia is better recovered but small artefacts are introduced elsewhere and overall the amplitudes are worse. In both cases, information appears to be lost compared to the transdimensional solution. The coarse grid appears to be adequate for the large scale data set whereas the finer grid is better for WOMBAT arrays.

The data misfit for both solutions is relatively similar with an ‘rms’ value of 3.3 s for the coarse grid solution and 3.1 s for the finer grid solution. In the absence of information on the measurement error, there is no way to objectively quantify which of these two solutions better describes the true velocity model.

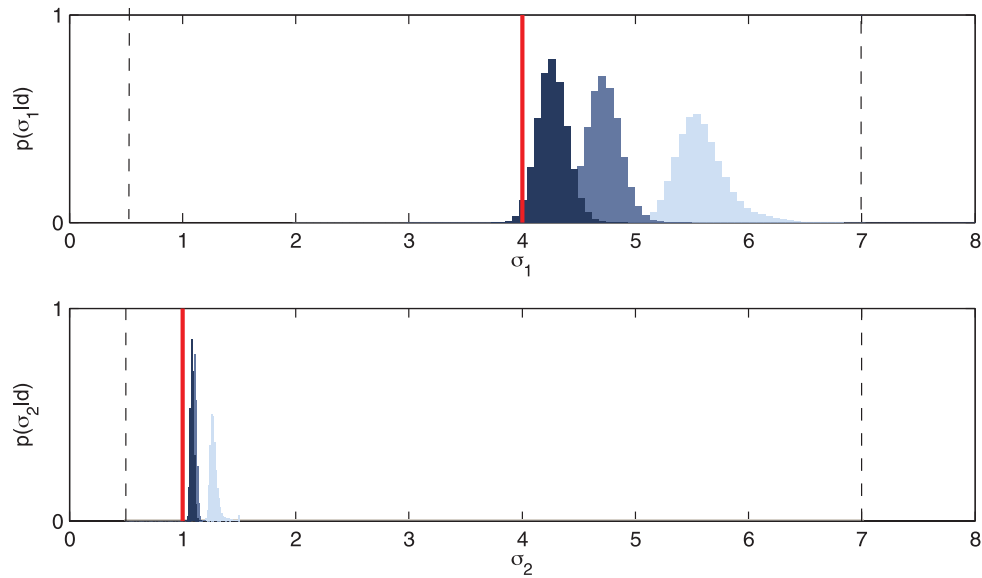


Figure 8. Posterior distribution for hyperparameters σ_1 (top panel) and σ_2 (bottom panel) for three iterations of the algorithm, with an update of ray geometries at each iteration (see Fig. 5). The posterior distribution after the first iteration is shown in light blue, second iteration in blue, and third iteration in dark blue. Red lines indicate the true values of noise added to the data. The bounds of the uniform prior distribution on the noise parameters are shown with dashed lines.

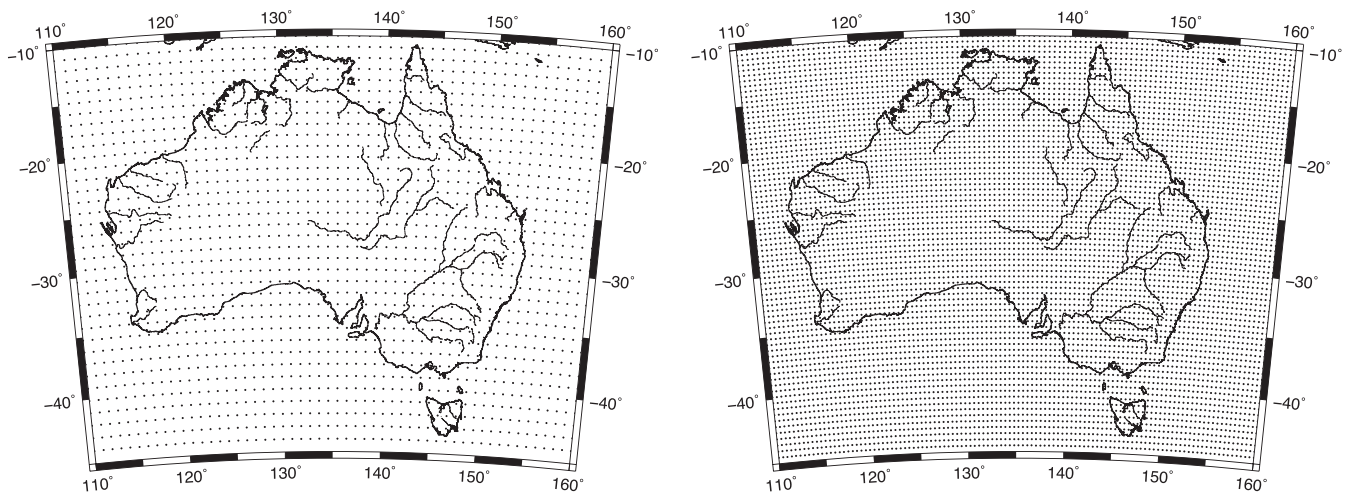


Figure 9. Fixed B-spline node parametrization used for the subspace inversion.

Fig. 11 shows the L-curve and its curvature obtained for the finer grid in 10(b) for a fixed damping value ($\epsilon = 3$) and with a varying smoothing parameter η . The curve shows the model roughness against the data fit for several values of η . The sharpness of the corner is not very well defined, resulting from the fact that different data sets with different properties (error and scale) are inverted together. The subspace inversion only uses a single global smoothing value although the optimal regularization parameter may be different according to data sets. This L-curve solution gives a mean residual of 3.1 s (which is a weighted average between the two true values of data noise 1 and 4), and there is no way to discriminate between different data types.

Smoothing constraints are applied equally to all parts of the model regardless of the actual resolution capability which depends on the ray coverage. This smoothing prevents unconstrained artefacts from appearing in areas of poor sampling, but also suppresses model details in the well-sampled areas.

6 FIELD DATA APPLICATION

We applied the transdimensional Hierarchical Bayes tomography algorithm to the three real data sets in Fig. 2 and constructed a multiscale tomographic image of Rayleigh wave group velocities at a period of 5 s.

6.1 Data noise parametrization

As previously, errors are assumed to be independent and normally distributed with zero mean and standard deviation σ_i . As a consequence, the data covariance matrix is diagonal and σ_i represent its diagonal elements.

In the synthetic example shown earlier, the data noise was parametrized with only two hyperparameters σ_1 (for the large scale data set) and σ_2 (for WOMBAT arrays). Here, different arrays are also modelled separately but the nature of the data may require

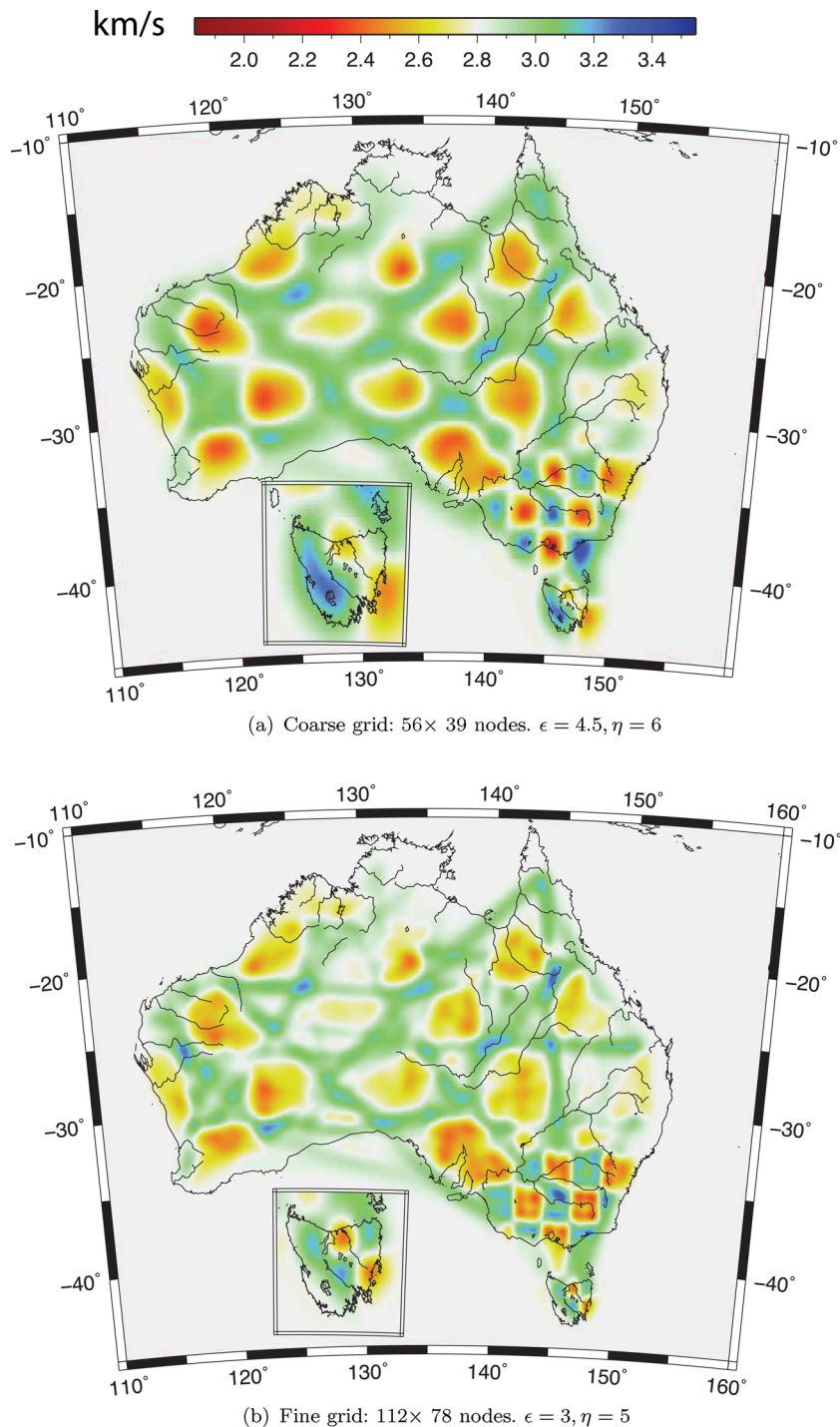


Figure 10. Solution models obtained with the subspace inversion for the two different grid sizes shown in Fig. 9. The smoothing and damping regularization parameters have been chosen by successively finding the point of maximum curvature of the L-curve.

the noise parametrization to be slightly more complex rather than a single constant number for each array type. This subsection presents the noise parametrization chosen for each array and explains how it is based on physical arguments.

6.1.1 WOMBAT arrays

Arroucau *et al.* (2010) produced traveltime picking uncertainties for WOMBAT arrays following the procedure presented by Cotte &

Laske (2002) and Harmon *et al.* (2007). Uncertainties were defined as the half-width of the time interval during which the amplitude of the envelope was 50 per cent of its peak amplitude. The aforementioned authors state that this choice for error bars sometimes results in a large scatter that may not reflect the actual precision of the measurements. Thus their error estimates need to be empirically rescaled, that is, to be multiplied by a constant factor.

These traveltime error estimates for WOMBAT arrays are shown as a function of the interstation distance in Fig. 12, and appear

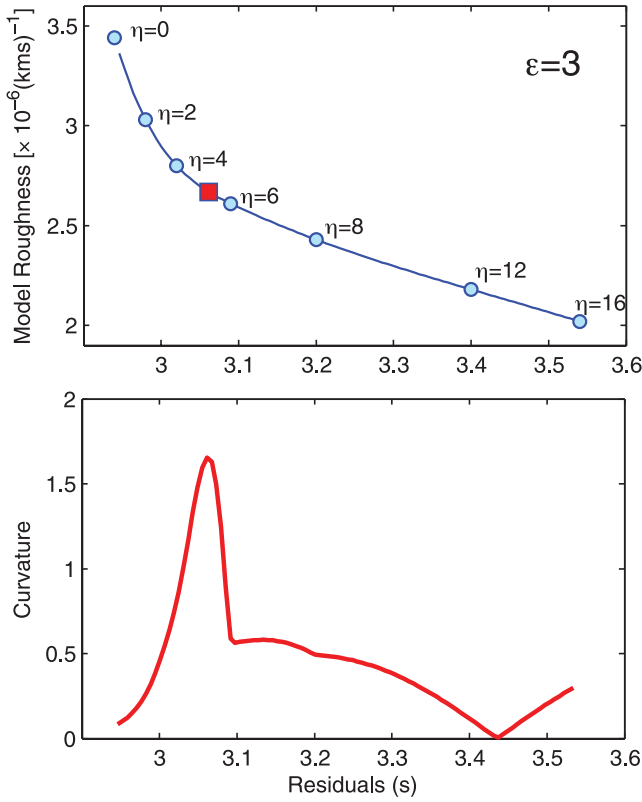


Figure 11. Upper panel: L-curve for the 112×78 nodes grid, ϵ is kept constant and η is changed. Lower panel: curvature of the L-curve. The maximum curvature gives the corner of the L-curve and provides the optimum η .

positively correlated with interstation distance. Short traveltimes are well constrained with an estimated uncertainty of around 1.5 s while those for stations further apart are less precise. For an angular distance of three degrees, the estimated noise is around 3 s.

In an optimization based inversion, this estimate of relative errors is sufficient to weight the information in the data. However, the Bayesian sampling approach depends on absolute errors, and so we invert for a scaling factor applied to the data covariance matrix. In particular, the σ'_i in (5) take the form

$$\sigma'_i = \lambda \times \sigma_{\text{rel}}^i \quad (7)$$

where σ_{rel}^i corresponds to the relative uncertainty, and λ is the unknown model hyperparameter. In this way, the algorithm uses the information available on relative errors in conjunction with posterior inference on a scaling factor provided by the Hierarchical Bayes procedure.

6.1.2 Large scale data set

For the large scale data set, no information is available on data uncertainties. Instead of parametrizing the data noise with a single hyperparameter, as in the synthetic example, we treat the noise as a linear function of the interstation distance and use two noise parameters, and write:

$$\sigma'_i = a \times d_i + b, \quad (8)$$

where d_i is the interstation distance, and a and b are hyperparameters to be inverted for.

The linear trend described earlier is a common observation in real data sets and is called a proportional effect (Aster *et al.* 2005).

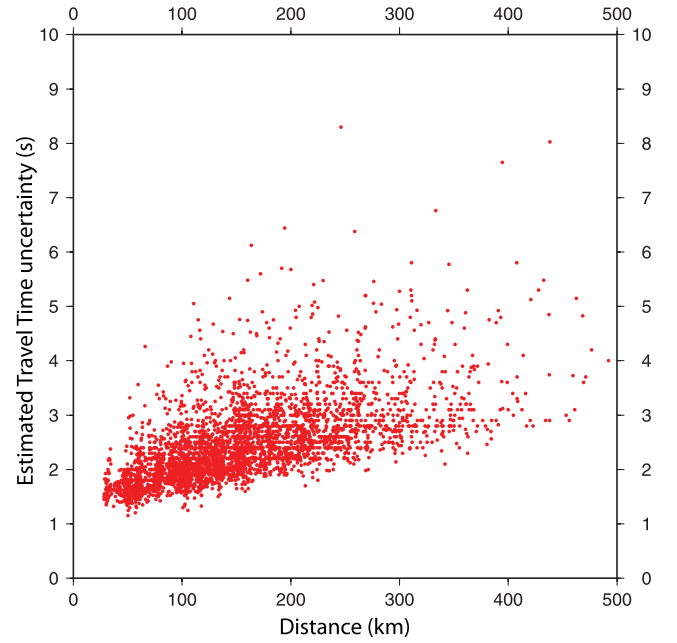


Figure 12. Estimated traveltime picking error as a function of interstation distance for WOMBAT arrays. According to Cotte & Laske (2002) and Harmon *et al.* (2007), this method of estimating uncertainties only provides relative errors between data points.

It occurs when the size of measurement errors are proportional to the measurement magnitude, and is clearly observable in Fig. 12. There are some theoretical arguments to explain the proportional effect in ambient noise data. At longer distances, the coherent part of the noise between two stations is more attenuated, and hence more recording time may be needed to construct the Green's function (Bensen *et al.* 2007). There is also a decrease in signal-to-noise ratio with distance due to the smaller range of azimuths of propagating surface waves that contribute constructively to the cross-correlation and to the scattering and multipathing along the great circle path between stations (Harmon *et al.* 2007; Weaver *et al.* 2009).

6.2 Results

The Hierarchical Bayes algorithm was run using two different choices for the likelihood. Data sets were first inverted with an L_2 misfit measure as in (5), and then with an L_1 norm, that is

$$\phi_{L_1}(\mathbf{m}) = \sum_{i=1}^N \frac{|d_i - g(\mathbf{m})_i|}{\sigma'_i}. \quad (9)$$

The likelihood function associated with the L_1 norm is a double sided exponential distribution, which is commonly named a Laplacian probability distribution

$$p(\mathbf{d}_{\text{obs}}|\mathbf{m}) = \prod_{i=1}^N \left[(2\sigma'_i)^{-1} \right] \times \exp\{-\phi_{L_1}(\mathbf{m})\}. \quad (10)$$

The advantage of using a Laplacian likelihood distribution is that the average solution will be more outlier resistant, or robust, than the expected Earth model obtained with a Gaussian likelihood (Aster *et al.* 2005).

In an optimization framework, finding the L_1 norm solution is complicated because the misfit function is then a non differentiable function of \mathbf{m} at any point where one of the residuals $d_i - g(\mathbf{m})_i$ is zero. However, Monte Carlo schemes do not use derivatives and

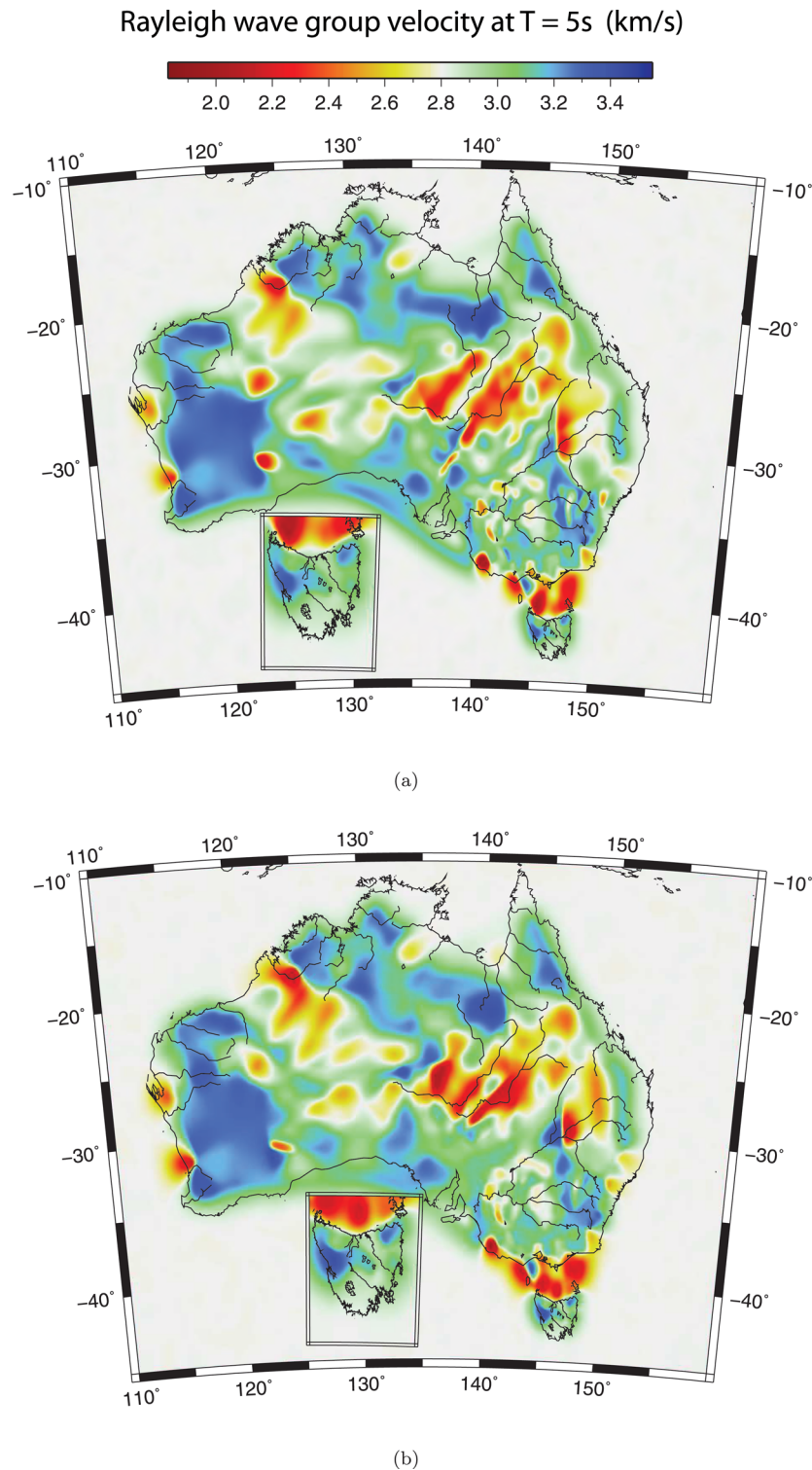


Figure 13. Hierarchical Bayes average solution (km s^{-1}). Top panel: data misfit defined with an L_2 norm, that is, the likelihood is a Gaussian distribution. Bottom panel: data misfit defined with an L_1 norm, that is, the likelihood is Laplacian distribution.

sampling a Laplacian distribution is no more difficult as sampling a Gaussian distribution. Fig. 13 shows the average solution maps for both L_1 and L_2 misfit definitions. Both were obtained after four updates of the ray paths. For each iteration, a total of 96 independent Markov chains were run independently on separate processors and posterior inference was made using an ensemble of 57 600 models. Each chain was run for 3×10^6 steps, the first half of which were

discarded as burn-in. Then, every 250th model visited was taken in the ensemble. In our case, it turns out that the solution model obtained with an L_1 norm is relatively similar to the one obtained with an L_2 norm.

The frequencies of ambient noise used here are largely sensitive to structure in the first 3–8 km of the crust, so one would expect the solution models in Fig. 13 to most readily discriminate between

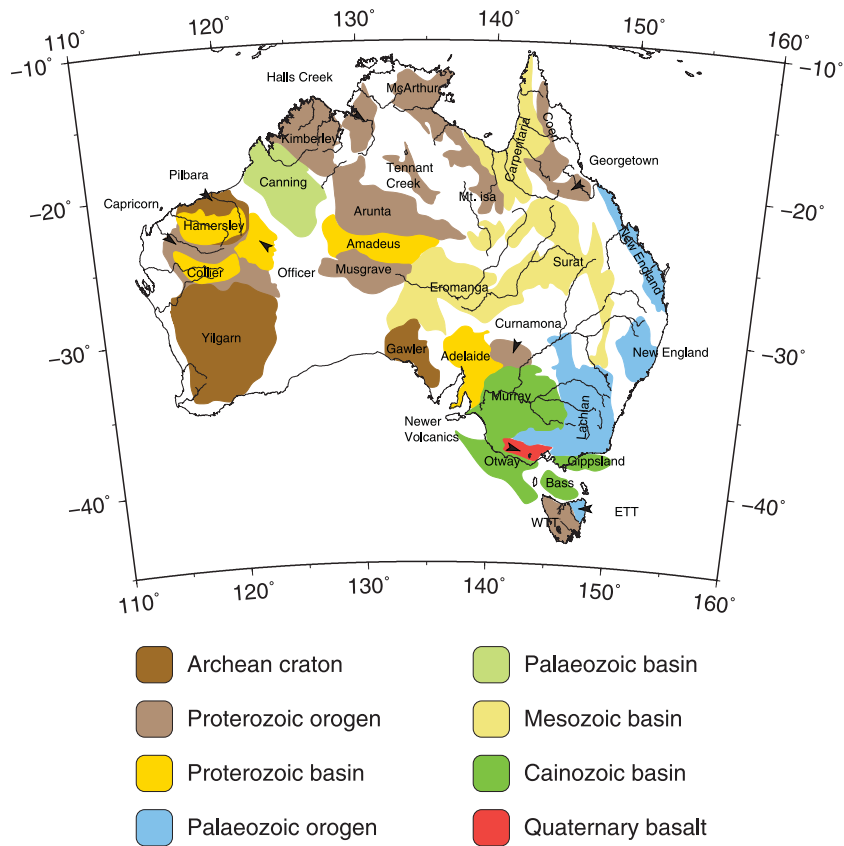


Figure 14. Simplified geological map of the Australian continent which shows regions containing significant sedimentary basins and exposed orogens classified according to age. WTT, West Tasmania Terrane; ETT, East Tasmania Terrane.

sedimentary and hard rock regions in the shallow crust. Fig. 14 shows a simplified broad-scale geological map of the Australian continent with terranes marked according to age and dominant exposure type. The correlation between higher velocities and the presence of exposed crystalline basement, and lower velocities and the presence of sedimentary basins of substantial thickness, is quite distinct. The Archean terranes of central and western Australia, comprising the Gawler, Yilgarn and Pilbara cratons (Fig. 14), are clearly revealed as regions of elevated velocity, although the full extent of the Pilbara craton is somewhat masked by the presence of overlying sediments. Although sampling much deeper in the lithosphere, the shear wave speed images obtained from regional surface wave tomography (Fishwick *et al.* 2005) of the Australian continent show a very similar pattern of high velocities beneath the Archean cratons.

Although less distinct, there is also a relationship between elevated velocities and the presence of Proterozoic and Palaeozoic basement. For example, the Kimberly Craton and Lachlan Orogen appear to correspond to regions of higher velocity; in the latter case, this region includes much of the southern Great Dividing Range, which contains extensive exposure of igneous rocks and metamorphosed sediments (Foster & Gray 2000). Sedimentary basins, which cover vast tracts of the Australian continent, are generally responsible for most of the low velocity features present in Fig. 13. For example, the Great Artesian Basin in central eastern Australia, which encompasses the Eromanga, Surat and Carpentaria basins (Fig. 14) contains large regions in which the sediment thickness exceeds 2 km (Laske & Masters 1997). The presence of such thick sediments would be responsible for the distinct low velocity

zone observed in central and central eastern Australia (Fig. 13). The Canning Basin in northwest Australia also hosts extremely thick sedimentary sequences—up to 6 km in some places (Laske & Masters 1997); this would explain the presence of the low velocity zone dividing the Kimberley and Pilbara Cratons.

In southeast Australia, the vast intracratonic Murray Basin is not very clearly defined in the velocity images (Fig. 13), but this is probably due to the sediment layer being relatively thin (Knight *et al.* 1995). However, the small scale length low velocity lineations that can be observed, due to the multiscale nature of the tomographic technique used to recover structure, probably represent the presence of pre-Tertiary infra-basins that underlie the Murray Basin (Arroucau *et al.* 2010). One remarkable feature of the tomography results is that all three Cainozoic basins in Bass Strait—the Bass, Otago and Gippsland basins (Fig. 14) appear as distinct low velocity zones, even though they are largely resolved by the continent-wide data set alone. In Tasmania, the divide between the Proterozoic West Tasmania Terrane and Phanerozoic East Tasmania Terrane is represented by a lower velocity transition zone, which is consistent with the recent results of (Young *et al.* 2011).

The posterior information on hyperparameters (n , λ , a , b) is shown in Fig. 15. Note that the collected velocity models in the ensemble solution have an average of 1200 cells. Each cell is defined by three parameters (2-D location of nodes + velocity) which makes the dimension of the model space around 3600. The Monte Carlo integration is feasible because it was implemented on parallel computing architecture. To give the reader an idea of the computational cost of such an inversion, each ‘outer-loop’ iteration of the algorithm (as shown in Fig. 5) requires approximately five days,

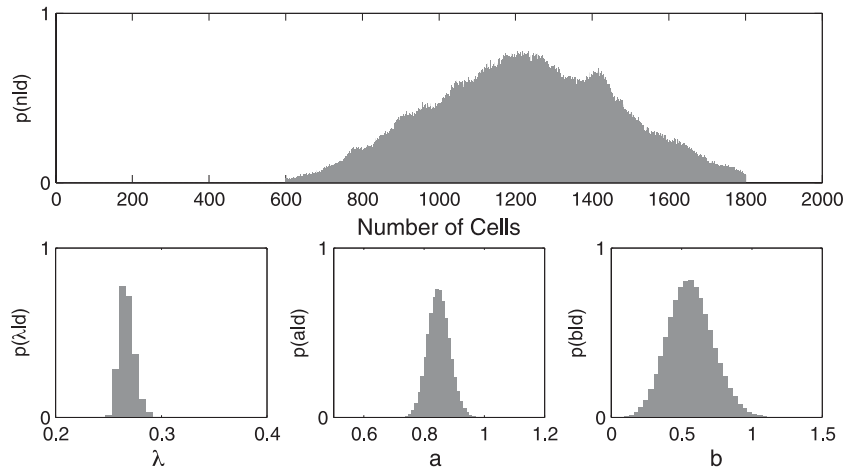


Figure 15. Posterior probability distribution on hyperparameters for the L_1 misfit solution shown in Fig. 13(b). The three inferred noise parameters λ , a and b define the level of data noise in data sets as given by (7) and (8).

so each panel of Fig. 13 represents about 15 days of computation time. The inferred information on the level of noise in traveltimes indicates that the uncertainties provided for the WOMBAT arrays have been rescaled to around $\lambda = 0.28$. The posterior value on hyperparameter a indicates that the data noise for the large scale is expected to increase 0.85 s each time the interstation distance increases by 1 degree with an expected data error of 0.6 s at 0° degrees (see lower panels in Fig. 15).

Fig. 16(a) shows the error map for the L_1 solution obtained with transdimensional tomography. This is constructed by taking the standard deviation of the ensemble of sampled Voronoi models at each point of the velocity field. This locally shows how well the solution model in 13(b) is constrained. As expected, well sampled areas in Western Australia, South East Australia and Tasmania show a lower velocity uncertainty.

Because the underlying parametrization is mobile, it is also interesting to look at the spatial density of Voronoi nodes across the ensemble of models collected. To do this we discretized the region into cells of $0.5 \times 0.5^\circ$, and calculated the average number of Voronoi nodes per cell over the posterior ensemble (see Fig. 16b). This map displays the average size of Voronoi cells at each point of the velocity field. A large number of small Voronoi cells are concentrated within WOMBAT arrays with larger cells elsewhere, thereby demonstrating the adaptive character of the transdimensional parametrization.

It is interesting to see that the estimated error on the model is not necessarily correlated with the density of cells. The Archaean cratons in Western Australia are well constrained without need of small cells (This area shows low values for model uncertainty in 16(a) with the lowest density of cells in 16b). There is good ray coverage in Western Australia, and one might expect the algorithm to introduce a lot of small cells to provide a high level of detail. However, the velocity field seems to be quite homogeneous and there is no need to introduce high levels of complexity in this region. This example shows the parsimonious nature of the algorithm and indicates that the transdimensional parametrization not only adapts to the density of rays but also to the character of the velocity structure itself.

6.3 Comparison with the Subspace Inversion

Finally, we compare our results with maps obtained with a standard fixed grid optimization approach. As described in Section 5.5, the

three data sets are simultaneously inverted with the subspace method (Kennett *et al.* 1988; Rawlinson *et al.* 2006, 2008). Here the inverse problem is regularized using ‘ad-hoc’ damping and smoothing, and the level of data noise is not accounted for. Fig. 17 shows two solutions obtained with the two grid sizes showed in Fig. 9. For each grid size, regularization parameters were successively ‘tuned’ with an iterative L-curve method as in Rawlinson *et al.* (2006).

As with the synthetic experiments, the coarse grid solution in Fig. 17(a) misses the details and small scale features under WOMBAT arrays that have been recovered with transdimensional tomography. Although the node spacing seems appropriate for the large scale data set, it is too coarse and cannot map information present under Southeast Australia and Tasmania. By using a finer grid, in Fig. 17(b) the velocity field in Southeast Australia presents more details but the structure under Tasmania is still missing. In this case, small artefacts resulting from data noise are introduced such as under Yilgarn craton, and overall the amplitudes are worse due to damping.

By visual inspection of the two images on Fig. 17, one can observe a striking feature of fixed grid inversion approaches: the scale of recovered velocity heterogeneities is spatially constant over the velocity field. Here the various scales of structural heterogeneities in the Earth, as well as inhomogeneities in data coverage, are not accounted for.

Although results are generally similar to transdimensional tomography for both choices of grid size, information appears to be lost in the well sampled areas compared to the transdimensional solution. Furthermore, here there is no way to measure which of these two solutions best describes the underlying seismic structure, which can result in misinterpretation. These are best fitting models and there is no information available on the level of uncertainty on the recovered velocity field.

7 CONCLUSION AND FUTURE DIRECTIONS

We have shown here that transdimensional tomography is particularly suited for inversion of multiple data sets that sample the Earth at different scales. Synthetic and real data examples have illustrated the adaptive character of the parametrization which enables us to image small scale features in well sampled areas without introducing spurious artefacts elsewhere. The level of smoothing is spatially

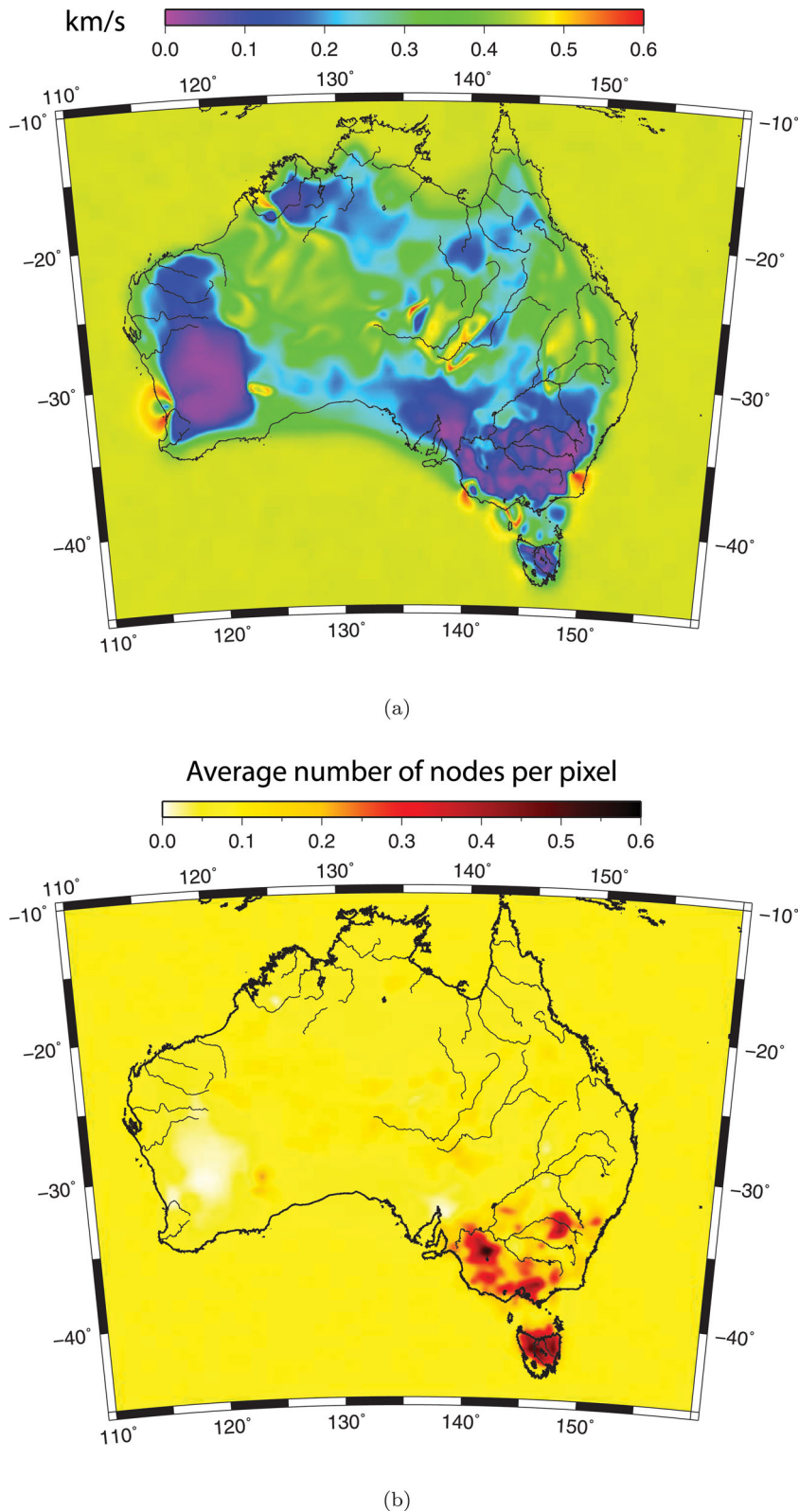


Figure 16. Top panel: error map (km s^{-1}) associated with the L1 norm solution in 13(b). Bottom panel: density of Voronoi nuclei across the ensemble of sampled models. The colour scale represents the expected number of nuclei per pixel.

variable and is naturally determined by the data. Contrary to other multiscale tomography methods, recovered structure is not only regulated by the density of rays, but also by the inferred data noise and by the structure of the underlying velocity field.

As the complexity of the model is variable, the estimated level of data noise takes an important role in the inversion as it directly determines the number of model parameters needed to fit the data to the required level. We have shown that an extended Bayesian

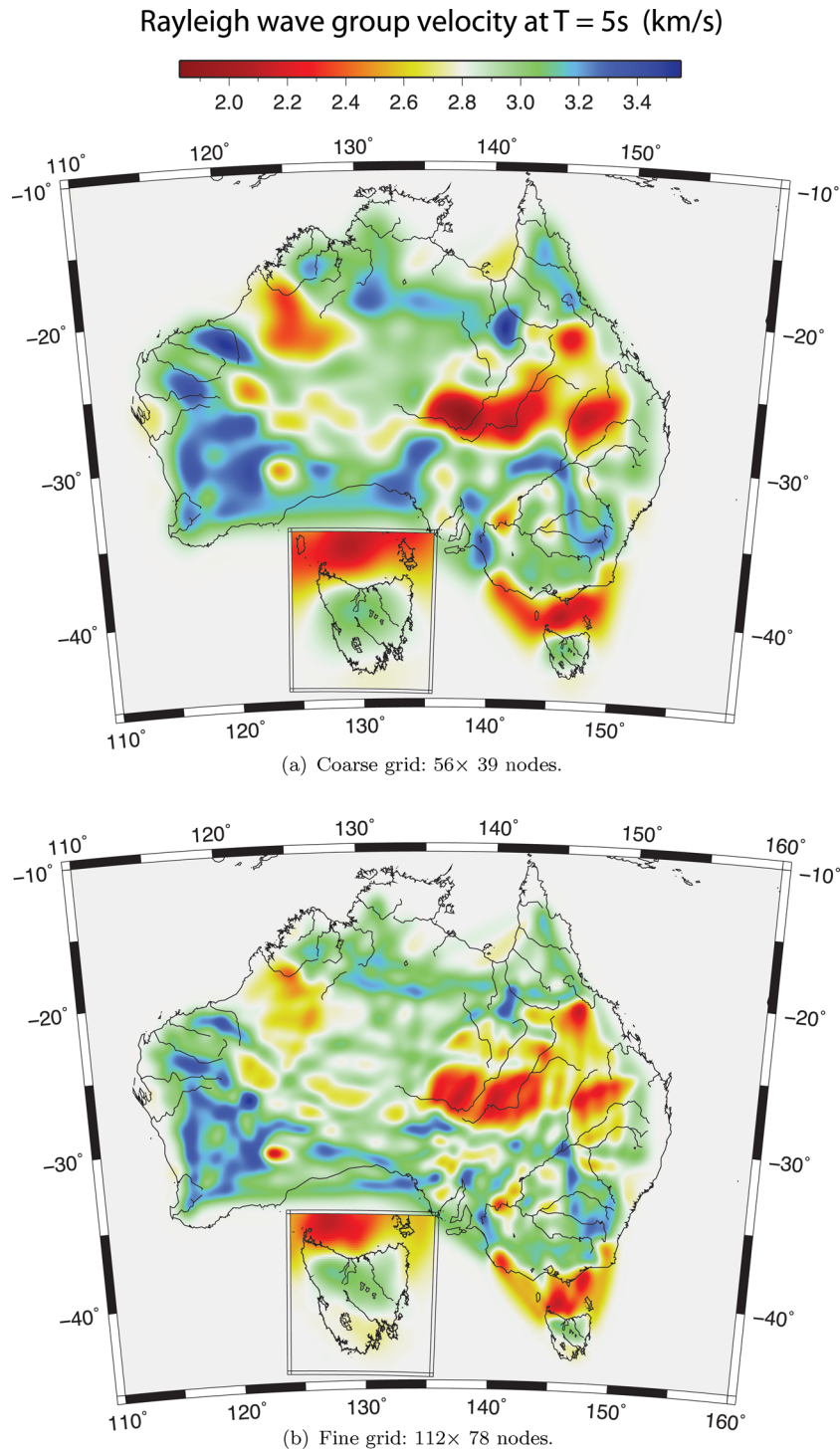


Figure 17. Solution models obtained with the subspace inversion for the two different grid sizes shown in Fig. 9. The smoothing and damping regularization parameters have been chosen by successively finding the point of maximum curvature of the L-curve.

formulation called Hierarchical Bayes can take into account the lack of knowledge on data uncertainty. When assessment of measurement errors is difficult to achieve *a priori* (as in ambient noise tomography), this procedure treats the standard deviation of data noise as an unknown and makes a joint posterior inference on both model complexity and data uncertainty. The Hierarchical Bayes procedure turns out to be particularly useful when dealing with

multiple data types having different unknown levels of noise. With scant prior knowledge on data errors, the algorithm is able to infer the level of information brought by each data type and to naturally adjust the fit to different data sets.

In our ambient noise tomography application, the data noise from WOMBAT arrays was naturally rescaled while the noise for the large scale data set was parametrized as a linear function of the interstation

distance. The inversion resulted in a parsimonious velocity map with a spatial resolution adapted to the quantity of information present in the data.

Uncertainty assessment on apparent traveltimes from ambient noise cross-correlation is an active area of research and Hierarchical Bayes could be used as a tool in the future to quantify the behaviour of noise with different parameters like interstation distance, azimuthal source distribution, or recording time.

It will be soon possible to incorporate new data from on-going deployments at different scales in Australia. Saygin & Kennett (2012) recently processed additional traveltimes that sample the crust at the continental scale. There are also supplementary ambient noise data for Tasmania available (Young *et al.* 2011). Furthermore, 67 short-period seismometers have recently been positioned across the Gawler and Curnamona Cratons in South Australia (Salmon & Arroucau 2010). Station spacing was approximately 60 km and covers the area from the Streaky Bay in the west to the New South Wales border in the east. Stations recorded continuous three component data for a period of 6–8 months and ambient noise traveltimes are currently being processed.

Another possibility is the inclusion of azimuthal anisotropy in the inversion. It is possible to observe an azimuthal dependence on the path-averaged velocities extracted from WOMBAT arrays. Therefore, instead of inverting for a single velocity value within each cell, one could invert for three anisotropic parameters per cell (a maximum velocity, a minimum velocity and a direction).

Other possible extensions include combining ambient noise recordings with receiver functions, earthquake surface wave dispersion measurements, regional and teleseismic traveltimes or SKS splitting measurements. The Hierarchical Bayes procedure may prove to be practically useful for joint inversion, because it is able to naturally weight the contribution of different data types in the misfit function, thus removing the user driven selection of weighting factors.

ACKNOWLEDGMENTS

We thank Kerry Gallagher for discussions on Hierarchical Bayes, Alberto Malinverno, Andrew Curtis and Stan Dosso for useful feedback, and Erdinc Saygin for assisting with seismic data. This project was supported by French-Australian Science and Technology travel grant (FR090051) under the International Science Linkages program from the Department of Innovation, Industry, Science and Research, and Australian Research Council Discovery project DP110102098. Calculations were performed on the TerraWulf II, a computational facility supported through the AuScope inversion laboratory. AuScope Ltd is funded under the National Collaborative Research Infrastructure Strategy (NCRIS) and the Education Infrastructure Fund (EIF), both Australian Commonwealth Government Programmes.

REFERENCES

- Abers, G. & Roecker, S., 1991. Deep structure of an arc-continent collision: earthquake relocation and inversion for upper mantle P and S wave velocities beneath Papua New Guinea, *J. geophys. Res.*, **96**(B4), 6379–6401.
- Arroucau, P., Rawlinson, N. & Sambridge, M., 2010. New insight into Cainozoic sedimentary basins and Palaeozoic suture zones in southeast Australia from ambient noise surface wave tomography, *Geophys. Res. Lett.*, **37**(7), L07303, doi:10.1029/2009GL041974.
- Aster, R., Borchers, B. & Thurber, C., 2005. *Parameter Estimation and Inverse Problems*, International Geophysics Series Vol. 90, Elsevier, Amsterdam.
- Bayes, T., 1958. An essay towards solving a problem in the doctrine of chances, *Biometrika*, **45**, 295–315 [reprint of the original article which appeared in *Philos Trans Roy Soc London*, 1763, **53**, 370–418].
- Bensen, G., Ritzwoller, M., Barmin, M., Levshin, A., Lin, F., Moschetti, M., Shapiro, N. & Yang, Y., 2007. Processing seismic ambient noise data to obtain reliable broad-band surface wave dispersion measurements, *Geophys. J. Int.*, **169**(3), 1239–1260.
- Bijwaard, H., Spakman, W. & Engdahl, E., 1998. Closing the gap between regional and global travel time tomography, *J. geophys. Res.*, **103**(B12), 30055–30078.
- Bodin, T. & Sambridge, M., 2009. Seismic tomography with the reversible jump algorithm, *Geophys. J. Int.*, **178**(3), 1411–1436.
- Bodin, T., Sambridge, M. & Gallagher, K., 2009. A self-parametrizing partition model approach to tomographic inverse problems, *Inverse Problems*, **25**, 055009, doi:10.1088/0266-5611/25/5/055009.
- Bodin, T., Sambridge, M., Tkalčić, H., P. A. & Gallagher, K., 2012. Transdimensional inversion of receiver functions and surface wave dispersion with a Hierarchical Bayes Algorithm, *J. geophys. Res.*, **117**, B02301, doi:10.1029/2011JB008560.
- Boschi, L., Ekström, G. & Kustowski, B., 2004. Multiple resolution surface wave tomography: the mediterranean basin, *Geophys. J. Int.*, **157**, 293–304.
- Boschi, L., Fry, B., Ekström, G. & Giardini, D., 2009. The european upper mantle as seen by surface waves, *Surv. Geophys.*, **30**(4), 463–501.
- Brenguier, F., Shapiro, N., Campillo, M., Ferrazzini, V., Duputel, Z., Coutant, O. & Necessian, A., 2008. Towards forecasting volcanic eruptions using seismic noise, *Nat. Geosci.*, **1**(2), 126–130.
- Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science*, **299**(5606), 547–549.
- Charvin, K., Hampson, G., Gallagher, K. & Labourdette, R., 2009. A Bayesian approach to inverse modelling of stratigraphy, part 2: Validation tests, *Basin Res.*, **21**(1), 27–45.
- Chiao, L. & Kuo, B., 2001. Multiscale seismic tomography, *Geophys. J. Int.*, **145**(2), 517–527.
- Chiao, L. & Liang, W., 2003. Multiresolution parameterization for geophysical inverse problems, *Geophysics*, **68**(1), 199–209, doi:10.1190/1.1543207.
- Clifford, P., Greenhalgh, S., Houseman, G. & Graeber, F., 2007. 3-D seismic tomography of the Adelaide fold belt, *Geophys. J. Int.*, **172**(1), 167–186.
- Cotte, N. & Laske, G., 2002. Testing group velocity maps for Eurasia, *Geophys. J. Int.*, **150**(3), 639–650.
- Curtis, A. & Snieder, R., 1997. Reconditioning inverse problems using the genetic algorithm and revised parameterization, *Geophysics*, **62**(5), 1524–1532.
- Delost, M., Virieux, J. & Operto, S., 2008. First-arrival traveltimes tomography using second generation wavelets, *Geophys. Prospect.*, **56**(4), 505–526.
- Detmer, J., Dosso, S. & Holland, C., 2010. Trans-dimensional geoaoustic inversion, *J. acoust. Soc. Am.*, **128**, 3393–3405, doi:10.1121/1.3500674.
- Fishwick, S., Kennett, B. & Reading, A., 2005. Contrasts in lithospheric structure within the Australian craton. Insights from surface wave tomography, *Earth planet. Sci. Lett.*, **231**(3–4), 163–176.
- Foster, D. & Gray, D., 2000. Evolution and structure of the Lachlan Fold Belt (Orogen) of eastern Australia, *Annu. Rev. Earth planet. Sci.*, **28**(1), 47–80.
- Fukao, Y., Obayashi, M., Inoue, H. & Nembai, M., 1992. Subducting slabs stagnant in the mantle transition zone, *J. geophys. Res.*, **97**(B4), 4809–4822.
- 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.
- Gallagher, K., Bodin, T., Sambridge, M., Weiss, D., Kylander, M. & Large, D., 2011. Inference of abrupt changes in noisy geochemical records using Bayesian Transdimensional changepoint models, *Earth planet. Sci. Lett.*, **311**, 182–194, doi:10.1016/j.epsl.2011.09.015.
- Gelman, A., Carlin, J., Stern, H. & Rubin, D., 2004. *Bayesian Data Analysis. Texts in Statistical Sciences*, 2nd ed., Chapman and Hall, Boca Raton, FL.

- Geyer, C. & Møller, J., 1994. Simulation procedures and likelihood inference for spatial point processes, *Scand. J. Stat.*, **21**(4), 359–373.
- Gouveia, W. & Scales, J., 1998. Bayesian seismic waveform inversion: parameter estimation and uncertainty analysis, *J. geophys. Res.*, **103**(B2), 2759–2780.
- Graeber, F., Houseman, G. & Greenhalgh, S., 2002. Regional teleseismic tomography of the western Lachlan Orogen and the Newer Volcanic Province, southeast Australia, *Geophys. J. Int.*, **149**(2), 249–266.
- Green, P., 1995. Reversible jump MCMC computation and Bayesian model selection, *Biometrika*, **82**, 711–732.
- Green, P., 2003. Trans-dimensional Markov chain Monte Carlo, *Highly Struct. Stoch. Syst.*, **27**, 179–198.
- Gung, Y., Hsu, Y., Chiao, L. & Obayashi, M., 2009. Multiscale waveform tomography with two-step model parameterization, *J. geophys. Res.*, **114**(B11), B11301, doi:10.1029/2008JB006275.
- Harmon, N., Forsyth, D. & Webb, S., 2007. Using ambient seismic noise to determine short-period phase velocities and shallow shear velocities in young oceanic lithosphere, *Bull. seism. Soc. Am.*, **97**(6), 2009–2023, doi:10.1785/0120070050.
- Hastings, W., 1970. Monte Carlo simulation methods using Markov chains and their applications, *Biometrika*, **57**, 97–109.
- Hopcroft, P., Gallagher, K. & Pain, C., 2009. A Bayesian partition modelling approach to resolve spatial variability in climate records from borehole temperature inversion, *Geophys. J. Int.*, **178**(2), 651–666.
- Hu, W., Abubakar, A. & Habashy, T., 2009. Simultaneous multifrequency inversion of full-waveform seismic data, *Geophysics*, **74**(2), R1–R14, doi:10.1190/1.3073002.
- Hung, S., Chen, W., Chiao, L. & Tseng, T., 2010. First multi-scale, finite-frequency tomography illuminates 3-D anatomy of the Tibetan Plateau, *Geophys. Res. Lett.*, **37**(6), L06304, doi:10.1029/2009GL041875.
- Jasra, A., Stephens, D., Gallagher, K. & Holmes, C., 2006. Bayesian mixture modelling in geochronology via Markov chain Monte Carlo, *Math. Geol.*, **38**(3), 269–300.
- Kennett, B., Sambridge, M. & Williamson, P., 1988. Subspace methods for large inverse problems with multiple parameter classes, *Geophys. J. Int.*, **94**(2), 237–247.
- Khan, A., Zunino, A. & Deschamps, F., 2011. The thermo-chemical and physical structure beneath the North American Continent from Bayesian Inversion of surface-wave phase velocities, *J. geophys. Res.*, **116**, B09304, doi:10.1029/2011JB008380.
- Knight, L.A., McDonald, P.A., Frankel, E. & Moore, D.H., 1995. A preliminary appraisal of the pre-tertiary infrabasins beneath the Murray Basin, Northwestern Victoria, Department of agriculture, energy and minerals, Victorian Initiative for Minerals and Petroleum Report.
- Larose, E., Roux, P., Campillo, M. & Derode, A., 2008. Fluctuations of correlations and green's function reconstruction: role of scattering, *J. appl. Phys.*, **103**(11), 114907–114907.
- Laske, G. & Masters, G., 1997. A global digital map of sediment thickness, *EOS, Trans. Am. geophys. Un.*, **78**, F483.
- Levshin, A. & Ritzwoller, M., 2001. Automated detection, extraction, and measurement of regional surface waves, *Pure appl. Geophys.*, **158**(8), 1531–1545.
- Li, C., van der Hilst, R., Engdahl, E. & Burdick, S., 2008. A new global model for P wave speed variations in Earth's mantle, *Geochem. Geophys. Geosyst.*, **9**(5), Q05018, doi:10.1029/2007GC001806.
- Loris, I., Nolet, G., Daubechies, I. & Dahlen, F., 2007. Tomographic inversion using 1-norm regularization of wavelet coefficients, *Geophys. J. Int.*, **170**(1), 359–370, doi:10.1111/j.1365-246X.2007.03409.x.
- Loris, I., Douma, H., Nolet, G., Daubechies, I. & Regone, C., 2010. Nonlinear regularization techniques for seismic tomography, *J. Comput. Phys.*, **229**(3), 890–905.
- Luo, X., 2010. Constraining the shape of a gravity anomalous body using reversible jump Markov chain Monte Carlo, *Geophys. J. Int.*, **180**(3), 1067–1079.
- MacKay, D., 2003. *Information Theory, Inference, and Learning Algorithms*, Cambridge University Press, Cambridge.
- Malinverno, A., 2002. Parsimonious Bayesian Markov chain Monte Carlo inversion in a nonlinear geophysical problem, *Geophys. J. Int.*, **151**(3), 675–688.
- Malinverno, A. & Briggs, V., 2004. Expanded uncertainty quantification in inverse problems: Hierarchical Bayes and empirical Bayes, *Geophysics*, **69**, 1005–1016, doi:10.1190/1.1778243.
- Malinverno, A. & Parker, R., 2006. Two ways to quantify uncertainty in geophysical inverse problems, *Geophysics*, **71**, W15–W27, doi:10.1190/1.2194516.
- Metropolis, N. et al., 1953. Equations of state calculations by fast computational machine, *J. Chem. Phys.*, **21**(6), 1087–1091.
- Micheline, A., 1995. An adaptive-grid formalism for traveltimes tomography, *Geophys. J. Int.*, **121**(2), 489–510.
- Minsley, B., 2011. A trans-dimensional Bayesian Markov chain Monte Carlo algorithm for model assessment using frequency-domain electromagnetic data, *Geophys. J. Int.*, **187**, 252–272, doi:10.1111/j.1365-246X.2011.05165.x.
- Nolet, G., 2008. *A Breviary of Seismic Tomography: Imaging the Interior of the Earth and Sun*, Cambridge University Press, Cambridge.
- Nolet, G. & Montelli, R., 2005. Optimal parametrization of tomographic models, *Geophys. J. Int.*, **161**(2), 365–372.
- Okabe, A., Boots, B. & Sugihara, K., 1992. *Spatial Tessellations: Concepts and Applications of Voronoi Diagrams*, John Wiley & Sons Inc., New York, NY.
- Piana Agostinetti, N. & Malinverno, A., 2010. Receiver function inversion by trans-dimensional Monte Carlo sampling, *Geophys. J. Int.*, **181**(2), 858–872.
- Rawlinson, N. & Kennett, B., 2008. Teleseismic tomography of the upper mantle beneath the southern Lachlan Orogen, Australia, *Phys. Earth planet. Inter.*, **167**(1–2), 84–97.
- Rawlinson, N. & Sambridge, M., 2003. Seismic traveltimes tomography of the crust and lithosphere, *Adv. Geophys.*, **46**, 81–198.
- Rawlinson, N. & Sambridge, M., 2004. Wave front evolution in strongly heterogeneous layered media using the fast marching method, *Geophys. J. Int.*, **156**(3), 631–647.
- Rawlinson, N. & Urvoy, M., 2006. Simultaneous inversion of active and passive source datasets for 3-D seismic structure with application to Tasmania, *Geophys. Res. Lett.*, **33**, doi:10.1029/2006GL028105.
- Rawlinson, N., Reading, A. & Kennett, B., 2006. Lithospheric structure of Tasmania from a novel form of teleseismic tomography, *J. geophys. Res.*, **111**(B2), B02301, doi:10.1029/2005JB003803.
- Rawlinson, N., Sambridge, M. & Saygin, E., 2008. A dynamic objective function technique for generating multiple solution models in seismic tomography, *Geophys. J. Int.*, **174**(1), 295–308.
- Rawlinson, N., Tkalcic, H. & Reading, A., 2010. Structure of the Tasmanian lithosphere from 3D seismic tomography, *Aust. J. Earth Sci.*, **57**(4), 381–394.
- Rawlinson, N., Kennett, B., Vanacore, E., Glen, R. & Fishwick, S., 2011. The structure of the upper mantle beneath the Delamerian and Lachlan orogens from simultaneous inversion of multiple teleseismic datasets, *Gondwana Res.*, **19**, 788–799.
- Romanowicz, B., 2003. Global mantle tomography: progress status in the past 10 years, *Annu. Rev. Earth planet. Sci.*, **31**(1), 303–328.
- Sabra, K., Gerstoft, P., Roux, P., Kuperman, W. & Fehler, M., 2005. Surface wave tomography from microseisms in Southern California, *Geophys. Res. Lett.*, **32**, L14311, doi:10.1029/2005GL023155.
- Salmon, M. & Arroucau, P., 2010. New Results from the South Australian Seismic Arrays, in *Proceedings of the AESC*, Abstracts, Canberra.
- Sambridge, M. & Faletic, R., 2003. Adaptive whole Earth tomography, *Geochem. Geophys. Geosyst.*, **4**(3), 1022, doi:10.1029/2001GC000213.
- Sambridge, M. & Gudmundsson, O., 1998. Tomographic systems of equations with irregular cells, *J. geophys. Res.*, **103**(B1), 773–782.
- Sambridge, M. & Rawlinson, N., 2005. Seismic tomography with irregular meshes, *Geophys. Monogr.*, **157**, 49–65.
- 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.
- Saygin, E. & Kennett, B., 2010. Ambient seismic noise tomography of australian continent, *Tectonophysics*, **481**(1–4), 116–125.
- Saygin, E. & Kennett, B., 2012. Crustal structure of Australia from ambient seismic noise tomography, *J. geophys. Res.*, **117**, B01304, doi:10.1029/2011JB008403.
- Scales, J. & Snieder, R., 1998. What is noise, *Geophysics*, **63**(4), 1122–1124.
- Schaefer, J., Boschi, L. & E., K., 2011. Adatively parameterized surface wave tomography: methodology and a new model of the european upper mantle, *Geophys. J. Int.*, **186**(3), 1431–1453.
- Schivardi, R. & Morelli, A., 2009. Surface wave tomography in the European and Mediterranean region, *Geophys. J. Int.*, **177**(3), 1050–1066.
- Sethian, J. & Popovici, A., 1999. 3-D travelttime computation using the fast marching method, *Geophysics*, **64**(2), 516–523.
- Shapiro, N. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**(7), L07614, doi:10.1029/2004GL019491.
- Shapiro, N., Campillo, M., Stehly, L. & Ritzwoller, M., 2005. High-resolution surface-wave tomography from ambient seismic noise, *Science*, **307**, 1615–1618, doi:10.1126/science.1108339.
- Simmons, N., Myers, S. & Johannesson, G., 2011. Global-scale P wave tomography optimized for prediction of teleseismic and regional travel times for Middle East events: 2. Tomographic inversion, *J. geophys. Res.*, **116**(B4), B04305, doi:10.1029/2010JB007969.
- Spakman, W. & Bijwaard, H., 1998. Irregular cell parameterization of tomographic problems, *Ann. Geophys.*, **16**, 18.
- Steck, L., Thurber, C., Fehler, M., Lutter, W., Roberts, P., Baldrige, W., Stafford, D. & Sessions, R., 1998. Crust and upper mantle P wave velocity structure beneath Valles caldera, New Mexico: results from the Jemez teleseismic tomography experiment, *J. geophys. Res.*, **103**(B10), doi:10.1029/98JB00750.
- Stehly, L., Fry, B., Campillo, M., Shapiro, N., Guilbert, J., Boschi, L. & Giardini, D., 2009. Tomography of the alpine region from observations of seismic ambient noise, *Geophys. J. Int.*, **178**(1), 338–350.
- Stephenson, J., Gallagher, K. & Holmes, C., 2004. Beyond kriging: dealing with discontinuous spatial data fields using adaptive prior information and Bayesian partition modelling, *Geol. Soc. Lond., Spec. Publ.*, **239**(1), 195, doi:10.1144/GSL.SP.2004.239.01.13.
- Stephenson, J., Gallagher, K. & Holmes, C., 2006. Low temperature thermochronology and strategies for multiple samples 2: Partition modelling for 2D/3D distributions with discontinuities, *Earth planet. Sci. Lett.*, **241**(3–4), 557–570.
- Tarantola, A. & Valette, B., 1982. Inverse problems: quest for information, *J. Geophys.*, **50**, 159–170.
- Thurber, C. & Ritsema, J., 2007. Theory and observations: seismic tomography and inverse methods, *Treatise on Geophysics*, Vol. 1, pp. 323–360, Elsevier, Amsterdam.
- Tikhotsky, S. & Achauer, U., 2008. Inversion of controlled-source seismic tomography and gravity data with the self-adaptive wavelet parametrization of velocities and interfaces, *Geophys. J. Int.*, **172**(2), 619–630.
- Tkalčić, H., Rawlinson, N., Arroucau, N., Amarjeet, K. & Kennett, B., 2011. Multi-step modeling of receiver-based seismic and ambient noise data from WOMBAT Array: crustal structure beneath Southeast Australia, *J. geophys. Res.*, in press.
- Vesnaver, A., 1996. Irregular grids in seismic tomography and minimum time ray tracing, *Geophys. J. Int.*, **125**, 147–165.
- Villasenor, A., Yang, Y., Ritzwoller, M. & Gallart, J., 2007. Ambient noise surface wave tomography of the Iberian Peninsula: implications for shallow seismic structure, *Geophys. Res. Lett.*, **34**, L11304, doi:10.1029/2007GL030164.
- Voronoi, G., 1908. Nouvelles applications des parametres continus a la theorie des formes quadratiques, *J. Reine Angew. Math.*, **134**, 198–287.
- Weaver, R., Froment, B. & Campillo, M., 2009. Correlation of non-isotropically distributed ballistic scalar diffuse waves in two dimensions, *J. acoust. Soc. Am.*, **125**, 2536.
- Weaver, R., Hadziioannou, C., Larose, E. & Campillo, M., 2011. On the precision of noise correlation interferometry, *Geophys. J. Int.*, **185**, 1384–1392.
- Yang, Y., Ritzwoller, M., Levshin, A. & Shapiro, N., 2006. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.*, **168**(1), 259–274, doi:10.1111/j.1365-246X.2006.03203.x.
- Yao, H. & Van der Hilst, R., 2009. Analysis of ambient noise energy distribution and phase velocity bias in ambient noise tomography, with application to SE Tibet, *Geophys. J. Int.*, **179**(2), 1113–1132.
- You, S., Gung, Y., Chiao, L., Chen, Y., Lin, C., Liang, W. & Chen, Y., 2010. Multiscale ambient noise tomography of short-period rayleigh waves across northern taiwan, *Bull. seism. Soc. Am.*, **100**(6), 3165, doi:10.1785/0120090394.
- Young, M., Rawlinson, N., Arroucau, P., Reading, A. & Tkalčić, H., 2011. High-frequency ambient noise tomography of southeast australia: new constraints on tasmania's tectonic past, *Geophys. Res. Lett.*, **38**(13), L13313, doi:10.1029/2011GL047971.
- Zhang, H. & Thurber, C., 2005. Adaptive mesh seismic tomography based on tetrahedral and Voronoi diagrams: application to Parkfield, California, *J. geophys. Res.*, **110**, doi:10.1029/2004JB003186.