



Seismic Tomography and the Assessment of Uncertainty

N. Rawlinson^{*,1}, A. Fichtner[§], M. Sambridge[¶] and M.K. Young^{||}

^{*}School of Geosciences, University of Aberdeen, Aberdeen, Scotland, UK

[§]Institute of Geophysics, Department of Earth Sciences, ETH Zurich, Zurich, Switzerland

[¶]Research School of Earth Sciences, Australian National University, Canberra, ACT, Australia

^{||}DownUnder GeoSolutions Pty Ltd, West Perth, WA, Australia

¹Corresponding author: E-mail: nrawlinson@abdn.ac.uk

Contents

1. Introduction	2
1.1 Motivation	2
1.2 Historical Perspective	5
1.3 Uncertainty in the Age of Big data	12
2. Nonuniqueness in Seismic Tomography	14
2.1 Data Coverage	15
2.2 Data Noise	17
2.3 The Parameterization Problem	21
2.4 The Data Prediction Problem	23
2.5 The Inverse Problem	25
3. Practical Assessment Methods	28
3.1 Covariance and Resolution	28
3.2 Jackknife and Bootstrap	36
3.3 Synthetic Reconstruction Tests	38
3.4 Linear and Iterative Nonlinear Sampling	39
3.5 Fully Nonlinear Sampling	44
4. Case Studies	46
4.1 Synthetic Reconstruction Test: Teleseismic Tomography Example	46
4.2 Iterative Nonlinear Sampling: Surface Wave Tomography Example	52
4.3 Transdimensional Inversion: Surface Wave Tomography Example	55
4.4 Full Waveform Inversion: Resolution Analysis Based on Second-Order Adjoints	60
5. Concluding Remarks	64
References	66



1. INTRODUCTION

1.1 Motivation

For over 40 years seismic tomography has been the primary tool for revealing the heterogeneous nature of Earth's internal structure across a large range of scales. From its origins in early active source (Bois, La Porte, Lavergne, & Thomas, 1971) and passive source (Aki, Christofferson, & Husebye, 1977; Aki & Lee, 1976; Dziewonski, Hager, & O'Connell, 1977) travel time studies, seismic tomography has become increasingly sophisticated and powerful in response to advances in methodology, rapid improvements in computing power, and growth in the availability of high-quality digital data. Today, we have reached the point where massive inverse problems involving millions of unknowns and tens of millions of data values can be tackled (e.g., Burdick et al., 2014); where the entire waveform can be inverted rather than a derivative component such as travel time (e.g., Chen, Zhao, & Jordan, 2007; Fichtner, 2011; Tape, Liu, Maggi, & Tromp, 2010); where multiscale structures can be recovered in a single inversion (Bodin, Sambridge, Tkalcic, et al., 2012; Burdick et al., 2008; Fichtner et al., in press); where multiple data sets can be jointly inverted (Obrebski, Allen, Pollitz, & Hung, 2011; Rawlinson & Urvoy, 2006; West, Gao, & Grand, 2004), including data sets of different classes such as surface wave dispersion, gravity, and heat flow (Afonso, Fullea, Yang, Connolly, & Jones, in press); and where various seismic properties, including P- and S-wave velocity and attenuation can be recovered, as well as, in some cases, other physical and material properties such as temperature and composition (Khan, Boschi, & Connolly, 2011; Afonso, Fullea, Griffin, et al., in press, Afonso, Fullea, et al., in press). As a consequence, our knowledge of the Earth's internal structure, composition, and dynamics is rapidly improving.

Despite the growing power of seismic tomography as a tool to image the Earth's interior, there remains one crucial facet of the technique that has only seen limited improvement in recent times. This is the issue of solution robustness, which arises from the ill-posed nature of the tomographic inverse problem. According to the original definition of Hadamard (1902, pp. 49–52), a well-posed problem in mathematics is characterized by having a solution that exists, is unique, and changes continuously with respect to initial conditions. In most practical seismic tomography applications, the inverse problem is under- or mixed-determined, so multiple data-satisfying

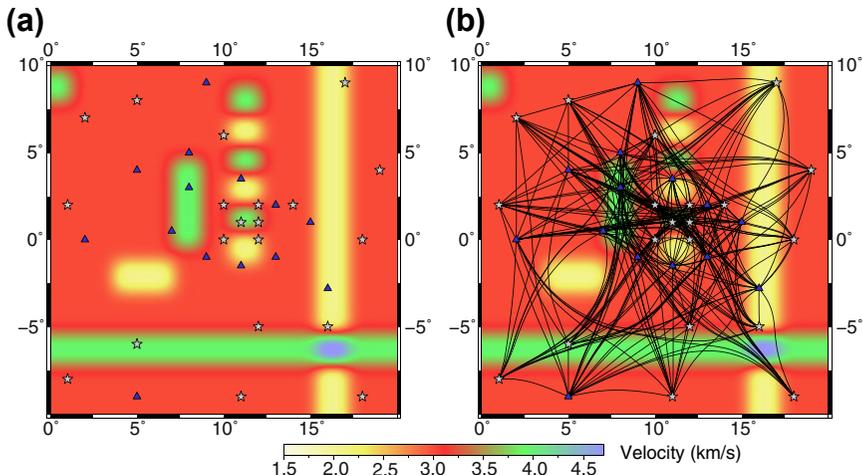


Figure 1 Simple ray tracing example that demonstrates why seismic tomography problems tend to be ill-posed. (a) Velocity perturbations overlain by a set of sources (white stars) and receivers (blue triangles); (b) two-point rays traced between all sources and receivers. (For interpretation of the references to color in this figure legend, the reader is referred to the online version of this book.)

solutions exist, and solutions (e.g., maximum likelihood in a linearized least squares formulation) tend to be unstable with respect to small perturbations in prior information and data noise in the absence of regularization.

Although relatively simple, [Figure 1](#) provides useful insight into the ill-posed nature of seismic tomography. The synthetic model consists of high- and low-wave speed perturbations relative to a background value of 3.0 km/s and is of variable scale length ([Figure 1\(a\)](#)). The source of data is provided by an irregular distribution of sources and receivers. If we assume that geometric ray theory is valid and that only first arrivals are identifiable, then the data coverage ([Figure 1\(b\)](#)) is uneven, not just because of the source–receiver configuration, but also because the ray paths bend in response to variations in wave speed. For first arrivals, the tendency is to avoid low-velocity regions and preferentially sample high-velocity regions. In this case, the tomographic inverse problem can be formulated as one of finding a pattern of wave speeds that satisfy the two-point travel time data. From the path coverage in [Figure 1\(b\)](#), it seems obvious that the solution to such a problem would be nonunique, for instance, the high velocity anomaly in the top left corner could assume any value without influencing the data. Similarly, a number of low-velocity regions within the array are

poorly sampled, and could likely assume a range of data-satisfying values. Another consideration is that all seismic data contain noise, and as this noise increases, so does the range of models that fit the data equally well.

In applications involving data recorded in the field, uncertainty arises not only in the manner described above with regard to [Figure 1](#), but also as a result of simplifying assumptions in the physics of the forward problem, limitations on the range of possible structures imposed by the choice of parameterization, and assumptions about the distribution and magnitude of the noise. All of these influences on the solution are extremely difficult to quantify. To further compound the problem, when it actually comes to interpreting the seismic results (e.g., P-wave velocity image) in terms of temperature, composition, and other physical properties (e.g., grain size, presence of melt) that provide direct insight into subsurface structure and processes, there is an additional layer of nonuniqueness; for instance, a decrease in P-wave velocity could be due to an increase in temperature, an increase in melt, or a compositional change. As a consequence, even if features appear to stand out clearly in a tomographic image, their meaning is very often open to debate. For instance, while some authors cite images of low wave speeds extending throughout much of the mantle as evidence in support of mantle plume theory (e.g., [Montelli, Nolet, Dahlen, & Masters, 2006](#); [Montelli et al., 2004](#); [Wolfe et al., 2009](#)), others have suggested quite different convective mantle regimes (e.g., [Foulger et al., 2013](#)). Thus, uncertainty in seismic tomography affects a wide range of Earth scientists (e.g., geodynamicists, mantle geochemists) who utilize the singular insights into deep Earth structure provided by this branch of geophysics, and is not something that should be regarded as of interest to specialists only.

The pioneers of seismic tomography (e.g., [Aki et al., 1977](#)), and indeed geophysical inverse problems ([Backus & Gilbert, 1967, 1968, 1969, 1970](#)), were well aware of the issues surrounding solution robustness, and that simply producing a model that satisfied the data was not meaningful, unless associated error bounds could be determined. Yet even today, with vast computing and intellectual resources at our fingertips, it is all too common to find examples of seismic tomography in the literature where solution robustness is either ignored or given minimal treatment. It therefore seems timely to provide an overview of the various methods that have been devised to assess model uncertainty and consider their strengths and weaknesses. We here restrict ourselves to methods implemented in practical seismic tomography, noting that inverse theory is a vast field with many

applications throughout the physical sciences of which complete coverage is well beyond the scope of this paper. That said, there are other geophysical inverse problems that share similar challenges, for example, magnetic and gravity data inversion (Li & Oldenburg, 1996, 1998), which can also involve thousands or millions of unknowns. As such, much of what is covered here is also applicable to other fields.

In the following sections, a descriptive approach is favored over one involving detailed derivations in order to appeal to nonspecialists and students who may have limited background in this area. For those interested in the more mathematical aspects of the theory, sufficient references are included throughout the text. After providing a brief historical perspective, we discuss the causes of solution nonuniqueness in seismic tomography and then go on to describe a range of methods used to assess model robustness. A series of case studies are then presented to showcase a number of methods, ranging from the more traditional to the cutting edge. Note that some prior knowledge of seismic tomography methodology is assumed. For further information in this regard, interested readers are referred to several books and review articles on the subject (Iyer & Hirahara, 1993; Liu & Gu, 2012; Nolet, 1987; Nolet, 2008; Rawlinson & Sambridge, 2003; Romanowicz, 2003; Rawlinson, Pozgay, & Fishwick, 2010; Trampert & Fichtner, 2013a).

1.2 Historical Perspective

Although much of the fundamental framework for geophysical data inversion was laid down by Backus and Gilbert (1967, 1968, 1969, 1970) and Wiggins (1972), the first examples of seismic tomography were published half a decade later (Aki & Lee, 1976; Aki et al., 1977; Dziewonski et al., 1977). In the seminal work of Aki et al. (1977), on what is now known as teleseismic tomography, P-wave arrival time residuals from distant earthquakes are inverted for 3-D slowness variations beneath a seismic array. It is assumed that the wave impinging on the model region from below is planar, and that variations in wave speed can be described by a regular grid of constant slowness blocks. A further assumption is that the geometry of ray paths that penetrate the 3-D model region are only influenced by depth variations in wave speed. As a result of these assumptions, the inverse problem, although ill-posed, is linear. The authors use a damped least squares approach to solve the linear inverse problem, and also produce formal estimates of model resolution and covariance. Despite being published a year earlier, the subsequent study of Aki and Lee (1976) represents the first

example of local earthquake tomography, in which hypocenter parameters as well as slowness structure are simultaneously inverted for using arrival times. In this case, posterior covariance and resolution estimates are made for slowness structure, source location, and source origin time.

In the study of [Dziewonski et al. \(1977\)](#), 700,000 teleseismic P-wave travel time residuals are inverted for the 3-D velocity structure of the mantle described in terms of spherical harmonics. The authors use a similar approach to [Aki et al. \(1977\)](#) to solve the linearized inverse problem and also produce formal estimates of resolution. The earliest published example of seismic tomography in an active source (cross-hole) context ([Bois et al., 1971](#)) is 2-D but accounts for the path dependence on velocity structure by using a shooting method of ray tracing, in which the trajectory of rays are iteratively adjusted until source–receiver paths are obtained. A damped least squares approach similar to that used by [Aki et al. \(1977\)](#) is applied in an iterative manner to solve the inverse problem. Although no estimates of model uncertainty are provided in this case, the authors clearly recognize the issue of solution nonuniqueness and perform several inversions using different input parameters (such as cell size) to examine their sensitivity on the solution.

Of all methods for assessing robustness in seismic tomography, the synthetic reconstruction test is by far the most ubiquitous in the published literature; even today, most seismic tomography results are accompanied by a test of this nature (e.g., [Rawlinson, Pozgay, et al., 2010](#)). Although there are many variants, the basic commonality is that there is some contrived, synthetic, or known structure through which the forward problem is solved, using identical sources, receivers, and phase types as the observational data set. This creates a synthetic data set, which is as accurate as permitted by any approximations made in the forward solution. The next step is to carry out an inversion of the synthetic data set in an attempt to recover the known structure. Differences between the reconstruction and the known structure provide insight into the resolution limits of the data set. As discussed later, this approach to analyzing solution nonuniqueness has its drawbacks, but its relative ease of implementation, even with very large data sets, and the apparent simplicity of interpreting the output has made it extremely popular.

The first use of synthetic reconstruction tests in seismic tomography was actually made by [Aki and Lee \(1976\)](#) in their simultaneous inversion of local earthquake travel times for 3-D P-wave velocity structure and hypocenter

location. In this study, they examine three synthetic models: the first is a simple constant velocity half-space; the second a layered medium; and the third a simple laterally heterogeneous model that simulates the presence of a transform fault cross-cutting two media characterized by different velocities. The aim of these tests was to examine trade-offs between hypocenter location and velocity variation. To simulate the effects of observational error, random noise was also added to the synthetic data sets. Future synthetic reconstruction tests gradually introduced larger data sets, more sophisticated forward solvers, and more complex synthetic models, but the underlying approach used is essentially the same.

Today, the most commonly used model for synthetic tests is the so-called checkerboard model, which consists of a regular alternating pattern of positive and negative anomalies (e.g., positive and negative velocity perturbations relative to some reference model) along each spatial dimension of the model. This is an extension of the spike test (Walck & Clayton, 1987) in which the synthetic model contains one or more short-wavelength anomalies; inverting the associated synthetic data provides insight into smearing. The checkerboard test was first introduced by Spakman and Nolet (1988) and rapidly became very popular (Day, Peirce, & Sinha, 2001; Glahn & Granet, 1993; Graeber, Houseman, & Greenhalgh, 2002; Granet & Trampert, 1989; Rawlinson & Kennett, 2008; Rawlinson, Salmon, & Kennett, 2013; Ritsema, Nyblade, Owens, Langston, & VanDecar, 1998) due largely to its relative ease of interpretation. However, as discussed in more detail later, the insight into solution nonuniqueness provided by a checkerboard test is relatively limited (e.g., L ev eque, Rivera, & Wittlinger, 1993).

The drawbacks of synthetic testing and the difficulty of computing formal estimates of resolution and covariance for large inverse problems motivated researchers to look elsewhere for estimates of solution uncertainty. Statistics provides a number of standard tests for measuring accuracy that can be readily applied to potentially large inverse problems. These include bootstrapping and jackknifing, both of which are based on carrying out repeat inversions using different subsets of the data and then making an assessment of uncertainty from the ensemble of solutions that are produced. Both bootstrapping and jackknifing have been used in seismic tomography (Gung & Romanowicz, 2004; Lees & Crosson, 1989; Su & Dziewonski, 1997; Zelt, 1999), but examples in the published literature are few and far between.

As computing power increased during the 1990s and new methods were developed to tackle very large linear inverse problems, the issue of trying to estimate covariance and resolution in the presence of many thousands of unknowns was revisited. For example, [Zhang and McMechan \(1995\)](#) use an extension of LSQR, a variant of the conjugate gradient method developed by [Paige and Saunders \(1982\)](#), to approximate resolution and covariance matrices for problems involving thousands of unknowns and tens of thousands of observations. [Yao, Roberts, and Tryggvason \(1999\)](#) provide an alternative approach to estimating resolution and covariance using LSQR and [Zhang and Thurber \(2007\)](#) apply a method that also relies on Lanczos bidiagonalization but yields the full resolution matrix and sidesteps the issue of whether subspace methods, like LSQR, can produce useful estimates of uncertainty given that they are restricted to exploring a small subspace of the full model space at each iteration ([Nolet, Montelli, & Virieux, 1999](#)).

A variety of other more peripheral techniques—in the sense that they have not gained common usage—have been suggested in the last few decades for assessing model robustness in the context of linear and iterative nonlinear inversion methods. Several of these fall into the category of producing multiple data-satisfying models from which summary information is produced. For example, [Vasco, Peterson, and Majer \(1996\)](#) use multiple starting models to generate a set of solutions to which cluster analysis is applied to retrieve the more robust features. [Deal and Nolet \(1996\)](#), within a strictly linear framework, identify model null-space vectors along which the solution can change but the data fit is essentially invariant. This “null-space shuttle” enables one to produce an ensemble of data fitting solutions with high computational efficiency, as demonstrated by the recent paper of [de Wit, Trampert, and van der Hilst \(2012\)](#), in which the uncertainty of detailed global P-wave models is assessed. [Rawlinson, Sambridge, and Saygin \(2008\)](#) develop a dynamic objective function approach to generating multiple solution models in which the objective function is modified in response to the generation of each new model so that future models are deterred from visiting previously sampled regions of model space. The inverse problem therefore only needs to be solved a limited number of times before the full range of features allowed by the data is revealed. For all of the above sampling techniques, taking the average model and the standard deviation of the ensemble as summary information is one way of interpreting the results.

Traditionally, seismic tomography has relied on regular parameterizations to represent structure. Due to the well-known trade-off between

resolution and variance (Backus & Gilbert, 1968), most data sets yield models in which the uncertainty can vary significantly as a function of location while spatial resolution is held constant. The other end-member approach is to attempt to keep model variance constant and vary the spatial resolution of the recovered model. Although this presents certain computational challenges, it has the potential advantage that solution robustness is relatively uniform across the model. Early work by Chou and Booker (1979) and Tarantola and Nercessian (1984), in which “blockless” strategies are developed, pioneered this approach, and were eventually followed by variable mesh methods, which are becoming increasingly common (Abers & Roecker, 1991; Bijwaard, Spakman, & Engdahl, 1998; Burdick et al., 2014; Curtis & Snieder, 1997; Fukao, Obayashi, Inoue, & Nebai, 1992; Michelini, 1995; Montelli et al., 2004; Nolet & Montelli, 2005; Sambridge & Gudmundsson, 1998; Sambridge, Braun, & McQueen, 1995; Spakman & Bijwaard, 2001). However, the challenge of working out how to spatially vary the resolution of recovery in response to the information content of the data is nontrivial, and to date there is no method that can guarantee that model variance is constant throughout the model, let alone what the value of the variance might be. Recent advances in wavelet decomposition methods for representing structure (Chiao & Kuo, 2001; Loris, Nolet, Dautbechies, & Dahlen, 2007; Simons et al., 2011; Tikhotsky & Achauer, 2008) may help alleviate this limitation.

In the last decade, there has been an increased focus on nonlinear sampling methods that produce an ensemble of data fitting models that can be subsequently interrogated for robust features. In many cases, these methods do not rely on the assumption of local linearization, which makes them attractive for highly nonlinear problems. The down side is, of course, the requirement for huge computational resources, but with rapid improvements in computing power, such problems are gradually becoming more tractable. Early attempts at fully nonlinear tomography, which were cast in the form of global optimization problems, include Pullammanappallil and Louie (1993) and Boschetti, Dentith, and List (1996) for 2-D reflection and refraction travel time tomography, and in 3-D, Asad, Pullammanappallil, Anooshehpour, and Louie (1999) for local earthquake tomography. Apart from the limited computing power available at the time, which necessitated the use of relatively small data sets, these pioneering efforts also relied on regular static parameterizations which did not account for spatial variations in the constraining power of the data.

Monte Carlo methods form the basis of most fully nonlinear inversion techniques developed for seismic tomography. [Sambridge and Mosegaard \(2002\)](#) define Monte Carlo methods as “experiments making use of random numbers to solve problems that are either probabilistic or deterministic in nature.” The origin of Monte Carlo methods can be traced back to the beginning of the nineteenth century, if not before ([Sambridge & Mosegaard, 2002](#)), but much of the pioneering work on modern techniques that are still used today originated in the 1960s ([Hammersley & Handscomb, 1964](#); [Press, 1968](#)). The first paper to introduce Monte Carlo inversion methods into geophysics was by [Keilis-Borok and Yanovskaya \(1967\)](#), which is based on earlier work in the Union of Soviet Socialist Republics where much of the initial development took place. Simulated annealing, a nonuniform Monte Carlo method for global optimization, was introduced into geophysics in the work of [Rothman \(1985, 1986\)](#). Genetic algorithms were first used in geophysics in the early 1990s ([Sambridge & Drijkoningen, 1992](#); [Stoffa & Sen, 1991](#)) for global optimization problems, and proved popular for solving highly nonlinear inverse problems involving a relatively small number of unknowns. The works cited in the previous paragraph by [Pullammanappallil and Louie \(1993\)](#) and [Boschetti et al. \(1996\)](#) used inversion methods based on simulated annealing and genetic algorithms, respectively.

Instead of using Monte Carlo techniques to directly solve global optimization problems, which produces a best fitting model, an alternative is to exploit the sampling they produce to assess uncertainty and trade-off issues, which are inherent to most geophysical inverse problems ([Sambridge, 1999](#)). It is in this context that Monte Carlo methods are seeing a resurgence in seismic tomography today, with the development of several techniques that promise to reshape the traditional linear optimization framework that is still favored in most studies.

Of the various nonlinear sampling techniques available, it is the advent of so-called Bayesian transdimensional tomography that has perhaps shown the most promise for improving the way we do seismic imaging ([Bodin & Sambridge, 2009](#)). A key feature of the approach is that the number and distribution of model unknowns, in addition to their values (e.g., velocity), are determined by the inversion. The advantage is that the level of detail recovered is strongly data driven, and potential increases in compute time caused by these additional degrees of freedom are offset by the exclusion of redundant parameters. The term “Bayesian” refers to the formal statistical framework for combining a priori model information (i.e., information

about model unknowns that are independent of the data) and data to produce a result (cast in terms of a posterior model distribution) that is more tightly constrained than the a priori model distribution (Scales & Snieder, 1997). Monte Carlo search techniques do not require a Bayesian setting of the inverse problem, and not all Bayesian inverse problems are solved using Monte Carlo methods, but the two are often linked. This is probably because Monte Carlo methods generally avoid implementing ad hoc regularization—common with optimization methods—that is at odds with the underlying philosophy of Bayes’ theorem.

The transdimensional inversion scheme of Bodin and Sambridge (2009) is driven by a reversible jump Markov chain Monte Carlo (rj-McMC) scheme, which produces a posterior probability density distribution of Earth models. This ensemble of models can be interrogated for summary information such as the average model and the standard deviation, which provides a measure of uncertainty. Bodin et al. (2012b) apply the scheme to multiscale ambient seismic noise data from the Australian region to produce group velocity maps. Young, Rawlinson, and Bodin (2013) extend the method to include inversion for shear wave speed and produce high-resolution 3-D images of the crust in southeast Australia using ambient noise data from a large transportable array. In these applications, ray path trajectory is not updated for every model generated by the rj-McMC scheme due to computational cost, which means that the inversion is not fully nonlinear. However, in practice, the frequency of update can be chosen to optimize the trade-off between compute time and invariance of the posterior probability density distribution. Galetti, Curtis, Baptie, and Meles (2014) use the scheme of Bodin et al. (2012b) with ray trajectory updates for each new model.

With modern computing power, Bayesian transdimensional tomography is becoming tractable even for relatively large, fully 3-D tomography problems. For example, Piana Agostinetti, Giacomuzzi, and Malinverno (submitted for publication) have developed a scheme that can be applied to 3-D local earthquake tomography, which involves inverting not only for V_p (P-wave velocity) and V_p/V_s (ratio of P-wave and S-wave velocity), but also for hypocenter location. Computational requirements for a realistic problem involving over 800 events, nearly 60 stations, and of the order of 5500–6500 unknowns is of the order of a week on a cluster of 250 central processing units (CPUs). As computing power grows, this class of Bayesian approach will no doubt become increasingly popular.

In the last decade, full waveform tomography has emerged as a viable tool for imaging the subsurface across a range of scales (Fichtner, 2011;

Fichtner, Kennett, Igel, & Bunge, 2009; Fichtner et al., in press; Operto, Virieux, Dessa, & Pascal, 2006; Ravaut et al., 2004; Smithyman, Pratt, Hayles, & Wittebolle, 2009; Tape, Liu, Maggi, & Tromp, 2009). Numerical solution of the elastic wave equation in three dimensions means that the full recorded wave train generated by a seismic event can be exploited, which has the potential to yield more information than more traditional approaches, like travel time tomography, which rely on picking the onset of identifiable phases. The main drawbacks are the computational cost of solving the wave equation and the nonlinear nature of the inverse problem, which until recently, have limited application to relatively small data sets. A consequence of these challenges is that meaningful estimates of solution robustness are difficult to make; for example, formal estimates of covariance and resolution under the assumption of local linearity are difficult to recover without full realization of the sensitivity matrix, which is generally not done in order to limit the computational burden. As a result, resolution analysis in full waveform tomography has been limited to synthetic recovery tests (e.g., Chen and Jordan, 2007) and estimates of composite volumetric sensitivity (Tape et al., 2010). However, in a recent paper, Fichtner and Trampert (2011b) demonstrate that under the assumption of a quadratic approximation to the misfit function, it is possible to produce quantitative estimates of resolution in full waveform tomography with a computational burden that is less than a synthetic reconstruction test.

1.3 Uncertainty in the Age of Big data

In the last few years, the term “Big data” has become popular to describe our rapidly growing ability to generate vast quantities of digital data (Dobbs, Manyika, Roxburgh, & Lund, 2011; Lohr, 2012; Marx, 2013). While there is no precise definition for this term, it generally refers to data sets that are too large to store, manage, or effectively analyze (Dobbs et al., 2011). Gantz and Reinsel (2010) estimate that the global rate of data collection is increasing at a rate of 58% per year, which in 2010 alone amounted to 1250 billion gigabytes, more bytes than the estimated number of stars in the universe. Moreover, since 2007, we have been generating more bits of data per year than can be stored in all of the world’s storage devices (Gantz & Reinsel, 2010).

Seismology is not immune from this data explosion, with increasing amounts of high-quality data being recorded, stored, and made available over the Internet. A good example of the rapid growth in seismic data comes from the Incorporated Research Institutions in Seismology, Data

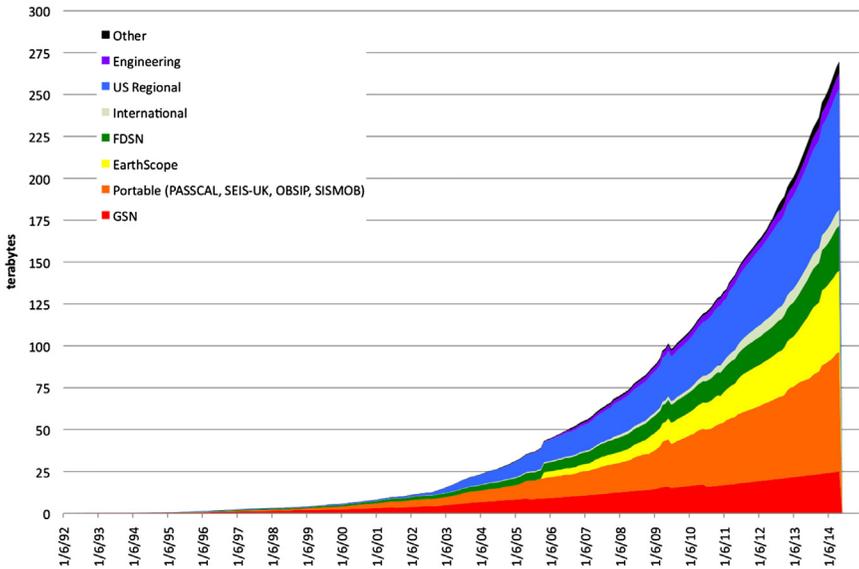


Figure 2 Cumulative volume of the IRIS DMC seismic data archive as on May 1, 2014. Source: <http://www.iris.edu>.

Management Centre (IRIS DMC), which since the early 1990s has been archiving local, regional, and global data sets. [Figure 2](#) plots the cumulative size of the archive since 1992, which suggests an exponential rate of growth. In the field of seismic tomography, the challenge will be to try and make use of as much of these data as possible. At regional (e.g., continent-wide) and global scales, there will be additional pressure to update models more regularly to keep pace with the deluge of new data that potentially may result in significant improvements to the imaging results. As it is, current global models already utilize millions of data measurements and push the boundaries of current computational resources (e.g., [Burdick et al., 2014](#)).

As the size of the inverse problem increases due to the addition of more data, the challenge and importance of assessing model uncertainty becomes arguably greater. For example, if one were to compute formal estimates of covariance and resolution using linear theory, then the compute time of the matrix inversion is of the order of $O(n^2 - n^3)$ for an $n \times n$ matrix [it could potentially be less if matrix sparsity is exploited, but the relationship is still nonlinear]; clearly, then, as the number of unknowns is increased, computational requirements will grow rapidly. Similarly, if a sampling approach such as the aforementioned rj-McMC is used to generate a set

of data-satisfying models from which summary information such as model standard deviation is extracted, compute time will not be linear. Yet, it is crucial that we have quantitative information on uncertainty as models become increasingly detailed and complex, and more inferences can potentially be made about the physical state of the region that is imaged. New methods will need to be brought to bear to properly deal with model uncertainty in the age of Big Data.



2. NONUNIQUENESS IN SEISMIC TOMOGRAPHY

Nonuniqueness in seismic tomography refers to when more than one model satisfies the observations, and is a consequence of the ill-posed nature of the problem. The reason that this arises is succinctly explained by [Snieder \(1991\)](#): “The inverse problem where one wants to estimate a continuous model with infinitely many degrees of freedom from a finite data set is necessarily ill-posed.” Although this appears to be indisputable, it is nonetheless at odds with a statement made by [Aki and Lee \(1976\)](#) in one of the first papers on seismic tomography when they evaluate the results of their inversion of local earthquake data: “Thus, we confirm our earlier contention that we can obtain a unique solution of our problem when sufficient data are available.” Ostensibly, this might seem contradictory, but in reality it is merely a case of viewing the problem at different stages of the solution process. From the outset, all seismic data sets are finite, so it follows that any number of models, with no restrictions on their flexibility, could be conceived that satisfy the observations to the same level. However, if we impose a limit on the minimum scale length of the model, for example, based on the dominant wavelength of the data that is being exploited (with the argument that the observables are insensitive to variations of smaller scale length), then the range of data-satisfying models will be dramatically reduced. Taking it a step further, if we now define an objective function to which the inverse problem is tied, then a unique solution may be possible, particularly if the assumption of linearization is imposed. The above statement made by [Aki and Lee \(1976\)](#) is essentially tied to an inverse problem that has been reduced to this state. However, in most seismic tomography problems, the presence of (often poorly constrained) data noise means that solution uniqueness is difficult to achieve, even if a variety of limiting assumptions are imposed on the permissible variations in structure.

Below, a brief description is provided of the various factors that play a role in constraining the solution of an inverse problem in seismic tomography.

2.1 Data Coverage

Increasing the volume of available data by adding contributions from additional sources or receivers will in many cases produce a better outcome in seismic tomography. However, it is well known that adding more data does not necessarily result in a better constrained inverse problem. This is illustrated by the simple case where a ray path traverses two blocks with different velocities. When a second ray passes through the same blocks with the same ratio of path lengths, then the two linear equations relating travel time and slowness are linearly dependent and so the new ray adds no new information to the inverse problem. Although there are a variety of tomography problems where this issue arises, it is particularly notable when earthquake sources are used (Fishwick & Rawlinson, 2012; Rawlinson, Kennett, Vanacore, Glen, & Fishwick, 2011). In such cases, earthquakes tend to cluster around seismogenic regions (e.g., subduction zones, active faults), and after a period of time most subsequent earthquakes occur within the neighborhood of previous earthquakes, such that they contribute little new structural information in the recorded seismogram.

Figure 3 shows a simple synthetic example, based on Figure 1, which demonstrates this concept. Figure 3(a) is a reconstruction of Figure 1(a), based on a constant velocity starting model, which uses the source–receiver travel times of the paths shown in Figure 1(b) (see Rawlinson et al., 2008; for an explanation of the iterative nonlinear inversion scheme). Figure 3(b) is a repeat of this experiment but with a travel time data set that is twice the size; this is simply accomplished by repeating each source location with a 0.3° perturbation in latitude and longitude. Despite the significant increase in the number of data, the recovered model is virtually identical. Given that the new source locations are perturbed by at least 2% of the model width, one might have expected the reconstruction in Figure 3(b) to be slightly better. However, given the minimum scale length of the anomalies, which is around 5% of the model width, and the nonlinearity of the problem, the lack of improvement is hardly surprising. In the latter case, since first-arrival rays are attracted to higher velocity regions, rays from nearby sources tend to bunch together, and do little to help constrain structure. In applications involving real observations, the presence of noise, provided that it is uncorrelated, should mean that adding more data, even if it samples identical

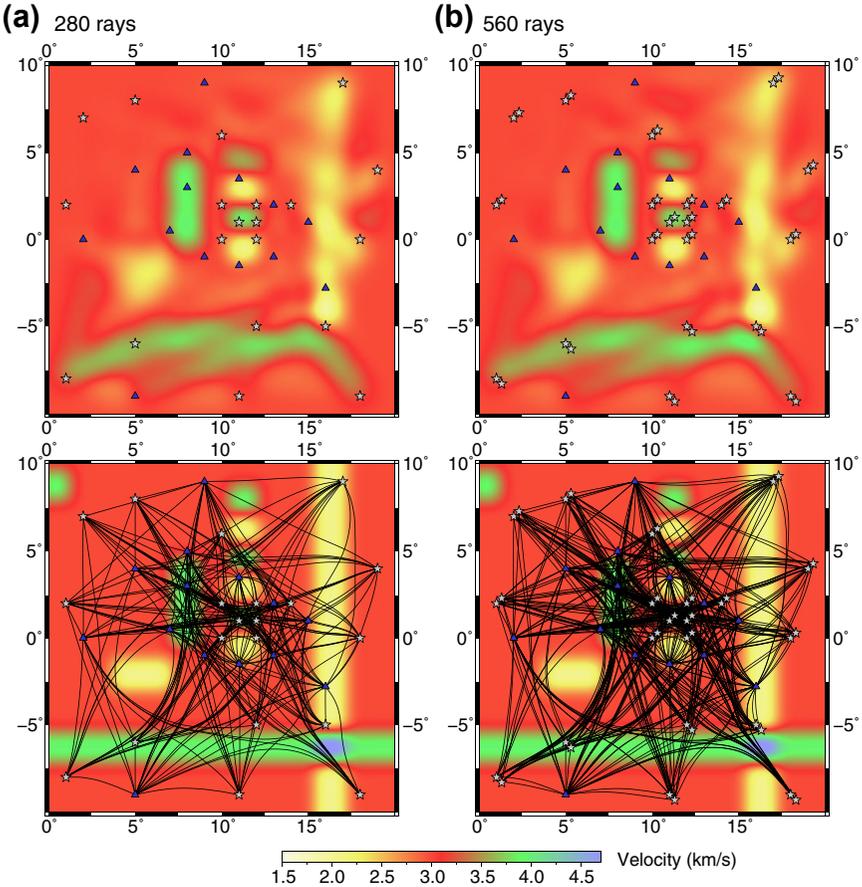


Figure 3 Two synthetic reconstruction tests based on the structure and path coverage shown in Figure 1. The initial model has a uniform velocity of 3.0 km/s. (a) Data set consists of 280 travel times between 20 sources and 14 receivers; (b) Data set consists of 560 travel times between 40 sources and 14 receivers. For both (a) and (b), the top plot shows the reconstructed model and the bottom plot shows the ray paths superimposed on the input model.

along-path structure, will improve the result due to an “averaging out” of the noise. This is the same philosophy behind data binning, which is often done prior to teleseismic, regional, or global tomography (e.g., Rawlinson & Fishwick, 2012).

Ray coverage or density maps are often used in seismic tomography to provide insight into the resolving power of seismic data and the quality of a reconstruction (Bijwaard & Spakman, 2000; Nielsen, Thybo, & Solodilov, 1999; Ramachandran, Hyndman, & Brocher, 2006; Walck & Clayton,

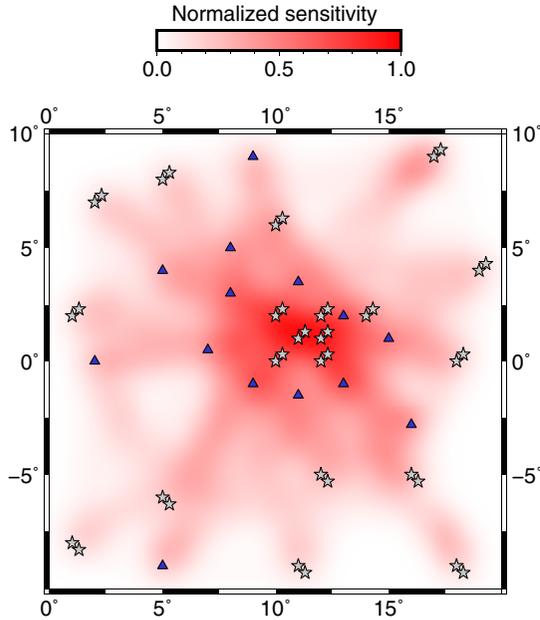


Figure 4 Normalized cumulative sensitivity of the travel time data set with respect to the model parameterization for the synthetic travel time data set shown in [Figure 1](#).

1987). However, at best they are an indirect tool, and have the potential to be misleading. A variant of the ray coverage or density map is to instead plot some measure of the sensitivity of the observables with respect to the model parameters ([Chen and Jordan, 2007](#); [Tape et al., 2010](#)). [Figure 4](#) shows the normalized cumulative sensitivity (obtained by summing the Fréchet derivatives at each control node and dividing by the largest value) of the travel time data set in [Figure 1](#); here the underlying grid is interpolated using a smooth mosaic of cubic B-spline functions, which is why the sensitivity plot is smooth. However, just because the data are sensitive to a change in the value of a parameter does not automatically mean that the parameter is well resolved. For example, if a unidirectional bundle of rays traverses a pair of cells, each ray travel time will vary if the velocity of either of the cells is changed, but the data cannot discriminate between a change made to the velocity of one or the other of the cells.

2.2 Data Noise

Noise is ubiquitous to all seismic data, and is often very difficult to accurately quantify. For example, with manual picking of phases, it is common

for even experienced analysts to disagree on onset times (Leonard, 2000), let alone some measure of picking uncertainty. Automated picking algorithms (Allen, 1982; Di Stefano et al., 2006; Vassallo, Satriano, & Lomax, 2012; Wang & Houseman, 1997) have the potential to offer more rigorous and consistent estimates of uncertainty; for example, Di Stefano et al. (2006) automatically compute arrival time uncertainty using a quality-weighting algorithm that takes into account waveform sampling rate, spectral density analysis, and signal to noise ratio. However, these estimates are calibrated using a series of reference picks and error estimates provided by the user. Picking methods that use some measure of waveform coherence (Chevrot, 2002; Rawlinson & Kennett, 2004; VanDecar & Crosson, 1990) have the potential to produce accurate estimates of relative onset times, and can yield estimates of uncertainty. In the case of Chevrot (2002), picking error is determined by computing the correlation coefficient between each seismic waveform and the optimal waveform (determined by simulated annealing), and comparing the result to the auto-correlation of the optimal waveform; the point where the two correlation functions intersect gives the time delay error. While this may produce good relative estimates of picking uncertainty, it is unclear whether the absolute values are very meaningful. In full waveform tomography, it is the seismogram itself that represents the data, so no explicit picking of phases is usually required. However, particular care is required as to how waveform misfit is defined if imaging artifacts caused by the presence of data noise are to be minimized (Bozdağ, Trampert, & Tromp, 2011). Since noise-induced measurement uncertainties are almost impossible to assess quantitatively for complete waveforms, full waveform inversion mostly operates with data characterized by high signal to noise ratios (e.g., Chen et al., 2007; Fichtner et al., 2009; Tape et al., 2010).

It is clear that the presence of data noise is unavoidable in seismic tomography, so it remains to be seen how it influences the analysis of uncertainty in seismic tomography. In general, as the level of data noise increases, the range of data-fitting models increases. Thus, in practice, even with an ostensibly overdetermined problem (more independent data than unknowns) any hope of solution uniqueness cannot be realized because of the noise. Within a linearized least squares inversion framework (Menke, 1989; Tarantola, 1987), where the data noise is assumed to have a Gaussian distribution, large levels of noise can be handled, provided some prior knowledge of the standard deviation of the noise is known. Figure 5 shows the result of applying the regularized (damped and smoothed) least squares inversion method of

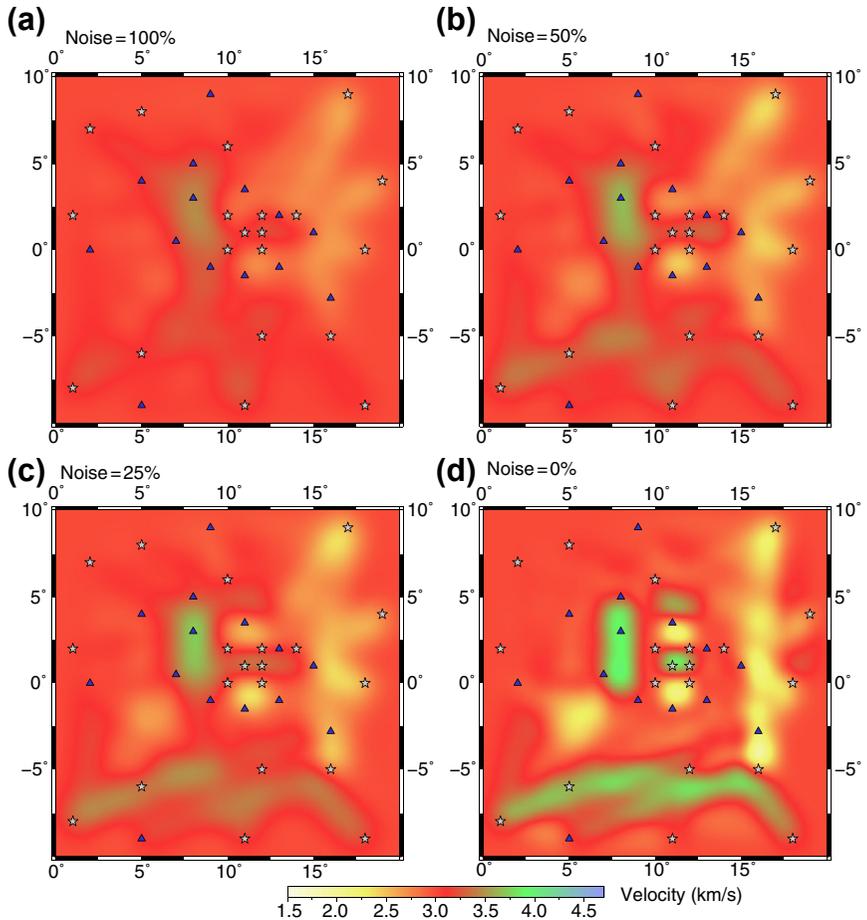


Figure 5 The effect of varying levels of data noise on the inversion of the travel time data set shown in [Figure 1](#). Gaussian noise with a standard deviation of (a) 100%, (b) 50%, (c) 25%, and (d) 0% are applied to the data, where 100% represents the data misfit of the initial (constant velocity) model (21 s).

[Rawlinson et al. \(2008\)](#) to the synthetic data set illustrated in [Figure 1](#) with varying amounts of Gaussian noise imposed, ranging up to a standard deviation that equals the standard deviation of the data misfit of the constant velocity initial model. In each case, the regularization is tuned so that inversion converges to a point where the standard deviation of the data misfit matches the standard deviation of the imposed data noise. As the noise level increases, the solution model essentially grades toward a uniform velocity model, with recovered model heterogeneity resembling the true structure,

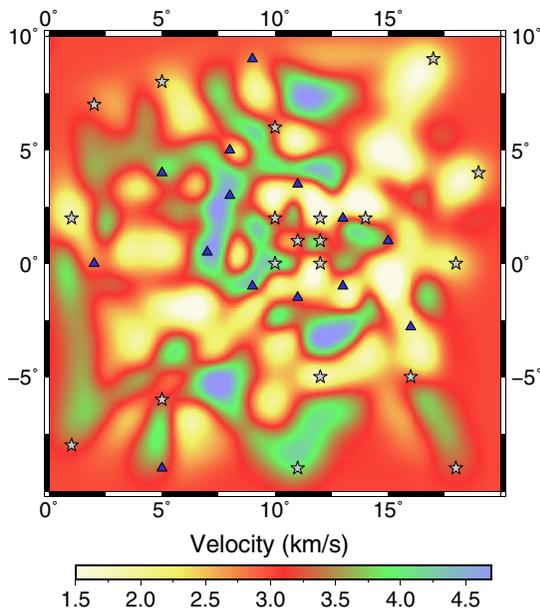


Figure 6 The same as [Figure 5\(a\)](#) (i.e., 100% noise), but this time the damping and smoothing are set such that the standard deviation of the final model data misfit is 75% of the standard deviation of the imposed data noise (i.e., the data are overfit).

albeit with increasingly lower amplitude and fewer short wavelength variations.

The example shown in [Figure 5](#) demonstrates the type of solution model behavior that is desirable, with structure only being recovered where required by the signal contained in the data. If one were to compute posterior covariance for these models, they would show that the posterior uncertainty would approach the prior uncertainty as the noise increases. However, if the standard deviation of the noise is poorly known, then it is relatively easy for spurious structure to be introduced. For example, [Figure 6](#) shows what happens when the standard deviation of data misfit is reduced to 75% of the standard deviation of the imposed data noise. Clearly, the results bear little resemblance to the truth, and any computation of the posterior covariance would likewise be extremely misleading. The rather disturbing result of [Figure 6](#) is in part the reason why the use of least squares misfit measures in seismic tomography is not universally adopted ([Djikpesse & Tarantola, 1999](#); [Pulliam, Vasco, & Johnson, 1993](#); [Scales, Gersztenkorn, & Treitel, 1988](#)). Another drawback of least squares misfit measures is that they are not robust to outliers, which could easily be introduced in a number

of ways including phase misidentification and Global Positioning System (GPS) timing issues. In the latter case, with data repositories such as IRIS DMC storing large data sets collected by different groups from various parts of the world, it is not unusual for GPS timing failures to be improperly flagged.

A novel approach to overcoming the issue of unknown or poorly understood levels of data noise is to treat the standard deviation of the noise as an unknown in the inversion (Malinverno & Briggs, 2004; Malinverno & Parker, 2006); this has recently been implemented via a Hierarchical Bayesian inversion scheme (Bodin, Sambridge, Tkalčić, et al., 2012), which exhibits natural parsimony, and has shown to be effective in a number of applications (Young, Cayley, et al., 2013; Young et al., 2013). Although a notable advance in the field of seismic tomography, it still requires the noise distribution to be assumed in advance.

2.3 The Parameterization Problem

One of the biggest, and in many cases the least, justifiable assumptions that is made in seismic tomography is with regard to the permissible range of seismic structure that can be recovered. In the Earth, seismic properties can vary smoothly or sharply in three dimensions over a great range of scales. Yet it is common to use regular basis functions in the spatial or wave number domain to represent structure. One of the simplest representations uses regular blocks with constant seismic properties (e.g., Achauer, 1994; Aki et al., 1977; Hildebrand, Dorman, Hammer, Schreiner, & Cornuelle, 1989; Nakanishi, 1985; Oncescu, Burlacu, Anghel, & Smalbergher, 1984; Vasco & Johnson, 1998), which has certain advantages such as simple initial value ray tracing. More sophisticated parameterizations use a grid of control nodes tied to a function that produces a continuum, such as trilinear (Eberhart-Phillips, 1986; Graeber & Asch, 1999; Zhao et al., 1994) or cubic splines (Farra & Madariaga, 1988; McCaughey & Singh, 1997; Rawlinson, Reading, & Kennett, 2006). In the spectral domain, truncated Fourier series (Hammer, Dorman, Hildebrand, & Cornuelle, 1994; Hildebrand et al., 1989; Wang & Houseman, 1997) have been used at local and regional scales, and spherical harmonics have been used at the global scale (Dziewonski et al., 1977; Dziewonski & Woodhouse, 1987; Romanowicz & Gung, 2002; Trampert & Woodhouse, 1995).

The drawback of all these parameterizations is that they impose severe limits on the types of structures that can be recovered, and can potentially result in the appearance of artifacts if the observed data are due to structure

that cannot be represented by the chosen parameterization. The imposition of a parameterization dramatically reduces the range of data-satisfying models, but the choice is often driven by convenience or computational tractability rather than the underlying physics of the problem. The use of a regular parameterization also does not take into account spatial variability in the information content of the data, which commonly occurs in seismic tomography due to irregular station and/or source distributions. Thus, some parts of a model may be well resolved by the data and other parts poorly resolved, but the spatial resolution of the model is unchanged. This means that due to the trade-off between resolution and covariance, poorly constrained parts of the model have a high error if they are oversampled by the parameterization, or well-constrained parts of the model have a low error if they are undersampled by the parameterization. In either case, information recovery is unlikely to be optimal.

From the point of view of solution nonuniqueness, the inverse problem can be well constrained or poorly constrained, depending on the spacing of the parameterization that is chosen. For instance, if the model was represented by a single parameter, then the inverse problem would be overdetermined and the posterior error associated with the solution would be small. Of course, it is unlikely that such a simple model would satisfy the data, but the aim of this book-end example is to show how subjective the assessment of model uncertainty can become as a result of its inextricable link to the choice of model parameterization.

To overcome the limitations of a fixed regular parameterization, a variety of studies have attempted to use either static or adaptive irregular parameterizations that are based on some measure of the information content of the data set (Abers & Roecker, 1991; Bijwaard & Spakman, 2000; Bijwaard et al., 1998; Burdick et al., 2008, 2014; Chou & Booker, 1979; Curtis & Snieder, 1997; Fukao et al., 1992; Michelini, 1995; Montagner & Nataf, 1986; Montelli et al., 2004; Sambridge et al., 1995; Tarantola & Necessian, 1984; Vesnaver, Böhm, Madrussani, Rossi, & Granser, 2000; Zhang, Rector, & Hoversten, 2005). A common static approach is to match the parameterization to the path density prior to inversion (Abers & Roecker, 1991). Adaptive approaches often use some kind of bottom-up splitting strategy in which new parameters are added in regions where data constraints appear greater (Sambridge & Faletic, 2003). The principle of trying to use the data itself to drive the spatial variability of recovered information appears sensible, but the additional degrees of freedom that this requires can quickly make the inverse problem intractable. Certainly one might imagine

that with the ability to vary the length scale of recovered structure, the goal would be to end up with a model in which the uncertainty associated with each parameter is identical; if not, then making inferences from the results becomes even more difficult, because one would need to account for variations in both uncertainty and scale length. Yet, no study published to date has produced a model with this property.

In the last decade or so, a number of advances have been made in the development of data-driven parameterizations. These include wavelet decomposition (Chiao & Kuo, 2001; Loris et al., 2007; Simons et al., 2011; Tikhotsky & Achauer, 2008) and partition modeling (Bodin & Sambridge, 2009; Bodin, Sambridge, Rawlinson, & Arroucau, 2012; Young et al., 2013). In the latter case, the number of unknowns, the spatial distribution of basis functions and the values of their coefficients are all unknowns in the inversion, which makes it an extremely data-driven process. Within a Bayesian framework, whereby the data are combined with prior model information to produce a posterior distribution of data-fitting models, the partition approach can recover structure over a large range of scale lengths and yield meaningful estimates of solution uncertainty.

2.4 The Data Prediction Problem

The way in which the forward problem is solved in seismic tomography can play a role in the recovery of structure, and therefore should not be ignored when assessing the robustness of the result. There are three basic ways in which the forward solution method can have an influence: (1) accuracy of the forward solver, (2) simplifying assumptions about the physics of wave propagation, and (3) completeness of the solution.

The limited accuracy of forward problem solvers can be a significant source of error because they numerically solve equations for which no analytical solutions are available. Numerical approximations are made in the interests of computational efficiency and if these approximations are poor, the resulting error will impact on the solution and any quantitative assessment of uncertainty. To illustrate the effect of forward problem errors on solution accuracy, Figure 7 shows the result of inverting the Figure 1 data set with an inaccurate forward solution. In this case, the eikonal equation, a nonlinear partial differential equation, is solved using a grid-based finite difference scheme. The use of a very coarse grid means that the finite difference approximations are poor and the travel time predictions inaccurate, resulting in a poorer solution. Compared to other sources of uncertainty in the tomography problem, inaccuracies in the forward solver, such as illustrated

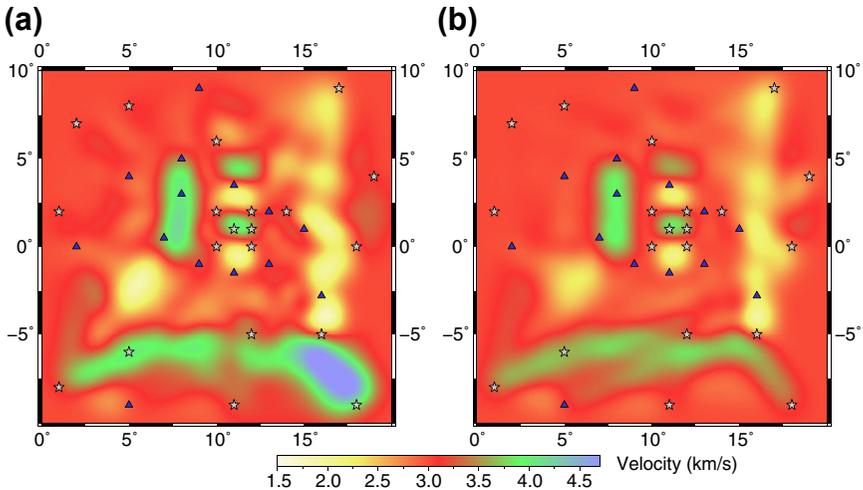


Figure 7 Inversion of the [Figure 1](#) data set using (a) inaccurate estimates of travel time and (b) accurate estimates of travel time. In (a) grid spacing for numerical solution of the eikonal equation is 0.8° , while in (b) it is 0.1° .

in [Figure 7](#), are relatively straightforward to overcome. However, there are more insidious sources of error that can be difficult to diagnose. For instance, most methods of two-point ray tracing, which can be viewed as a potentially highly nonlinear inverse problem, are nonrobust in that they use locally linear approximations to achieve convergence ([Cassell, 1982](#); [Farrar & Madariaga, 1988](#); [Julian & Gubbins, 1977](#); [Koketsu & Sekine, 1998](#); [Pereyra, Lee, & Keller, 1980](#); [Rawlinson, Houseman, & Collins, 2001](#); [Sambridge, 1990](#); [Um & Thurber, 1987](#); [VanDecar, James, & Assumpção, 1995](#); [Zhao, Hasegawa, & Horiuchi, 1992](#)). As a result, rays may not be found that exist, or the ray that is found is not the one that matches the phase identified on the seismogram due to multipathing, a phenomenon that increases as a function of velocity heterogeneity. This will invariably lead to unquantifiable uncertainties in the final model.

Errors caused by simplifications of wave propagation physics are an active area of research, especially in the context of so-called finite-frequency tomography. The majority of seismic tomography undertaken today is based on geometric ray theory, in which the underlying assumption is that the wavelength of the seismic wave is much smaller than the minimum scale length of heterogeneity. This high-frequency assumption ignores a variety of seismic wave behaviors including diffraction, scattering, and wave front healing. As a result, the sensitivity of the observable, such as travel time, is

dependent on off-path effects, and ignoring this relationship means that the accuracy of the recovery may be diminished. Studies that apply this approach are in general not able to provide a quantitative measure of how such simplifying assumptions impact on the uncertainty of the tomography result. Finite-frequency tomography (Chevrot & Zhao, 2007; Marquering, Dahlen, & Nolet, 1999; Montelli et al., 2004; Yang et al., 2009) attempts to overcome some of these limitations by using first-order perturbation theory to account for the presence of single scatterers. A simplification of wave propagation physics frequently made in local- to regional-scale full waveform inversion is the acoustic approximation (e.g., Bleibinhaus, Hole, & Ryberg, 2007; Kamei, Pratt, & Tsuji, 2013; Pratt & Shipp, 1999). The Earth is assumed to act as a fluid where wave propagation is governed by the acoustic wave equation that can be solved with much less computational resources than the complete elastic wave equation. While the acoustic approximation produces kinematically correct first arrivals (the travel times of the direct P and S waves are correct), later parts of the seismogram may not be accurately represented, thus introducing forward modeling errors that are difficult to quantify.

Both ray theory and the acoustic approximation illustrate that simplifications in the physics of seismic wave propagation can go hand in hand with incomplete solutions of the forward problem in the sense that specific types of waves cannot be modeled. Consequently, only specific aspects of the seismic wave field, e.g., early-arriving waveforms in the case of the acoustic approximation, can be exploited for tomography. This limitation, in turn, contributes to the nonuniqueness of the solution. Ultimately, errors in the forward problem can only be minimized by the robust solution of the full elastic wave equation, using, for instance, finite-difference (e.g., Moczo, Kristek, Vavrycuk, Archuleta, & Halada, 2002), spectral-element (e.g., Komatitsch & Vilotte, 1998), or other numerical techniques. The bandwidth of the solutions is, however, still very much limited by the available computational resources.

2.5 The Inverse Problem

The tomographic inverse problem involves adjusting model parameters in order to satisfy the data to an acceptable level and any a priori constraints that may be available. It is often formulated as a minimization problem in which an objective function or penalty function is defined and a search algorithm is applied to find regions of model space with a high level of fit. The way in which the objective function is defined can potentially have a major

influence on the inversion result and its associated uncertainty. Gradient-based inversion methods often use an objective function of the form (Rawlinson et al., 2010a)

$$S(\mathbf{m}) = (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}) + \varepsilon (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) + \eta \mathbf{m}^T \mathbf{D}^T \mathbf{D} \mathbf{m} \quad (1)$$

where $\mathbf{g}(\mathbf{m})$ are the predicted data, \mathbf{d}_{obs} are the observed data, \mathbf{C}_d is the a priori data covariance matrix, \mathbf{m}_0 is the reference model, \mathbf{C}_m is the a priori model covariance matrix, and \mathbf{D} is a second derivative smoothing operator. Ideally, \mathbf{C}_d represents the total covariance due to all sources of noise, including those due to observation and assumptions (e.g., in the parameterization and forward problem). However, the reality is that \mathbf{C}_d often only contains some ad hoc estimate of picking uncertainty and therefore would be better described as a weighting matrix rather than a true prior covariance matrix. Yet \mathbf{C}_d is crucial in controlling both the output model and its associated uncertainties, so if it is poorly representative of prior uncertainty, then the reliability of the solution will likewise be diminished. In recognition of this issue, there have been several recent studies that attempt to recover \mathbf{C}_d , or some component of it, during the inversion. For example, Duputel, Agram, Simons, Minson, and Beck (2014) attempt to recover the “prediction error” component of \mathbf{C}_d during earthquake source inversion and Bodin et al. (2012a) invert for the standard deviation of the diagonal elements of \mathbf{C}_d in surface wave tomography (a case study is provided in Section 4.3).

The prefactors ε and η in Eqn (1) are referred to as the damping and smoothing parameters, respectively, and control the trade-off between data fit, model perturbation relative to a reference model, and model smoothness. These regularization terms have different origins; in the case of damping, if we set $\varepsilon = 1$ and \mathbf{C}_d and \mathbf{C}_m truly represent the prior data and model covariance, respectively, then we have a Bayesian style inversion in which prior information on a model is combined with data constraints to produce a posterior distribution. Smoothing, on the other hand, appeals to Occam’s razor, in which parsimony is favored over complexity. The inclusion of both damping and smoothing in which ε and η are real positive variables results in the definition of a somewhat ad hoc objective function for which meaningful estimates of covariance and resolution, even for a linear inverse problem, are difficult to obtain.

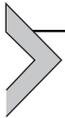
Another potential drawback of Eqn (1) is that it assumes data noise has a Gaussian distribution. However, there is no guarantee that this is the case, and outliers (e.g., from picking the wrong phase, GPS timing errors) may have an unjustifiably large influence on the inversion result. Removal of outliers on the basis of some assumption about the spread of acceptable values (e.g., greater than N standard deviations from the mean, where $N \geq 1$) is one approach for reducing their influence on the final result, but as pointed out by [Jeffreys \(1932\)](#), “a process that completely rejects certain observations, while retaining with full weight others with comparable deviations, possibly in the opposite direction, is unsatisfactory in principle.” [Jeffreys \(1932\)](#) developed a scheme, known as uniform reduction, which reduces the influence of outliers without directly needing to identify the anomalous data. The effect of uniform reduction is to assign outliers small weights so that they do not have a strong effect on the solution. The implementation and potential benefits of this approach in the context of 3-D local earthquake tomography are demonstrated in the study of [Sambridge \(1990\)](#).

In seismic tomography, the use of an L_2 measure of misfit such as Eqn (1) is almost universal, but in many cases there is little evidence for errors having a Gaussian distribution (e.g., phase arrival times in the International Seismological Centre (ISC) bulletin [Buland, 1986](#); [Pulliam et al., 1993](#)). As a result, alternative misfit measures have been considered, most notably the L_1 measure of misfit, which is known to be robust in the presence of outliers ([Claerbout & Muir, 1973](#); [Jeong, Pyun, Son, & Min, 2013](#); [Pulliam et al., 1993](#); [Scales et al., 1988](#)). [Claerbout and Muir \(1973\)](#) advocate the use of absolute error criteria, and find from studying numerous examples that it rarely exceeds two to four times the computing requirements of its least squares equivalent, and in many cases produces much better results. It is interesting to note that despite these early efforts, it is still very common to find studies that use an L_2 norm and simply cull “outliers” that are defined in a relatively arbitrary manner.

Deterministic inversion methods that produce a single solution, such as linearized least squares, generally require some kind of regularization to stabilize the inversion and produce a plausible result. When sampling methods are used, it is possible to dispense with explicit regularization altogether. For example, in the Bayesian transdimensional scheme of [Bodin and Sambridge \(2009\)](#), the objective function that is used is Eqn (1) without any damping or smoothing term; in other words, a simple least squares data misfit function.

Once an objective function or measure of misfit has been defined, there are various ways in which the inverse problem can be solved. Most

techniques rely on linearization of the inverse problem, which ultimately results in solution of a large system of linear equations. Back projection techniques like algebraic reconstruction technique and simultaneous iterative reconstruction technique avoid direct solution of these equations but tend to suffer from poor convergence properties (Blundell, 1993). Gradient-based methods such as damped least squares and its many variants (Aki et al., 1977; Graeber et al., 2002; Rawlinson et al., 2006; Thurber, 1983; Zhao et al., 1992) require the solution of a large and often sparse linear system of equations. There are various direct and approximate ways of solving such systems including Lower Upper (LU) decomposition, Cholesky decomposition, singular value decomposition, conjugate gradient and its LSQR variants, and more general subspace schemes (Hestenes & Stiefel, 1952; Kennett, Sambridge, & Williamson, 1988; Nolet, 1985; Scales, 1987). Ultimately, the aim is to move from one point in model space (the initial model) to another point (the final model) that lies within the bounds of all data-satisfying models. Assessing the uncertainty of this single solution involves trying to quantify the limits of the data-satisfying region of model space. However, the linearization assumption means that the method will only be effective if the objective function exhibits a single minimum and its surrounding architecture is approximately quadratic. The application of regularization essentially helps to conform the objective function to this shape. As such, any estimate of model uncertainty depends not only on a good knowledge of data noise, but also on the imposed regularization being consistent with prior information.



3. PRACTICAL ASSESSMENT METHODS

Below, a summary is given of the strengths and weaknesses of a variety of methods that have been devised for assessing model uncertainty in seismic tomography.

3.1 Covariance and Resolution

For inverse problems that are linear or linearizable, the calculation of formal estimates of posterior covariance and resolution is computationally tractable, although for larger problems, it is usually only a subset of the full information that is extracted (Nolet et al., 1999; Yao et al., 1999; Zhang & McMechan, 1995; Zhang & Thurber, 2007). The pioneering work of Backus and Gilbert (1968, 1970) and Wiggins (1972) established the foundation of general linear inverse theory for solving ill-posed inverse problems, which includes a

quantitative assessment of solution reliability. If the inverse problem is linearized, it is common to use an objective function of the form (Rawlinson, Pozgay, et al., 2010)

$$S(\mathbf{m}) = (\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d})^T \mathbf{C}_d^{-1} (\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d}) + \varepsilon \delta\mathbf{m}^T \mathbf{C}_m^{-1} \delta\mathbf{m} + \eta \delta\mathbf{m}^T \mathbf{D}^T \mathbf{D} \delta\mathbf{m} \quad (2)$$

where the last term on the right-hand side of the equation perturbs the prior model. The local minimum of this function occurs where $\partial S(\mathbf{m})/\partial\mathbf{m} = 0$, which results in a solution of the form

$$\delta\mathbf{m} = [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \varepsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1} \delta\mathbf{d} \quad (3)$$

The term $[\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \varepsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1}$ is often referred to as the generalized inverse \mathbf{G}^{-g} , the exact form of which is dependent on the choice of objective function. The resolution matrix can then be written as $\mathbf{R} = \mathbf{G}^{-g} \mathbf{G}$, where $\delta\mathbf{m} = \mathbf{R} \delta\mathbf{m}_{\text{true}}$, and estimates the averaging of the true model $\delta\mathbf{m}_{\text{true}}$ in its representation by $\delta\mathbf{m}$. For Eqn (3), the resolution matrix can be written:

$$\mathbf{R} = [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \varepsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} \quad (4)$$

The posterior covariance matrix is defined by $\mathbf{C}_M = \mathbf{G}^{-g} [\mathbf{G}^{-g}]^T$ (e.g., Yao et al., 1999) and measures the degree to which two model unknowns, m_i and m_j , vary together (or covary) i.e., $\text{cov}(m_i, m_j) = E[(m_i - \mu_i)(m_j - \mu_j)]$, where $\mu_i = E(m_i)$ and $\mu_j = E(m_j)$. \mathbf{C}_M can be related to the resolution matrix by $\mathbf{R} = \mathbf{I} - \mathbf{C}_M \mathbf{C}_m^{-1}$ (see Tarantola, 1987 for more details). In most applications, it is the diagonal elements of \mathbf{C}_M that indicate the posterior uncertainty associated with each parameter that is interpreted. Likewise, it is generally the diagonal elements of the resolution matrix that are considered, which have a value approaching unity for well-resolved parameters.

Another metric that can be useful in analyzing solution robustness is the correlation matrix, which can be defined (Tarantola, 1987):

$$\rho^{ij} = \frac{C_M^{ij}}{(C_M^{ii})^{\frac{1}{2}} (C_M^{jj})^{\frac{1}{2}}} \quad (5)$$

where $-1 \leq \rho^{ij} \leq 1$ and $i, j = 1, \dots, M$. A strong correlation between parameters indicates that they have not been independently resolved by the data. Covariance and resolution are commonly used to assess solution quality in seismic tomography (Aki et al., 1977; Riahi & Juhlin, 1994; Steck et al.,

1998; White, 1989; Yao et al., 1999; Zelt & Smith, 1992; Zhang & Thurber, 2007), although correlation is less frequently used (McCaughy & Singh, 1997; Zhang & Toksöz, 1998). The chief drawbacks of these measures of uncertainty are that (1) their validity decreases as the nonlinearity of the inverse problem increases, (2) the inversion of a large matrix is required, (3) implicit regularization imposed by the ad hoc choice of model parameterization is not accounted for, and (4) a priori model covariance and data errors are usually poorly known, which at the very least make absolute values of posterior uncertainty rather meaningless.

A major obstacle in the computation of resolution and covariance is, as previously mentioned, the need to explicitly store and invert potentially very large matrices. This difficulty sparked the development of matrix probing techniques where information about specific properties of a matrix, e.g., its largest eigenvalues or its trace, can be estimated through the application of the matrix to random vectors. While very general matrix probing techniques have been developed in applied mathematics (see Halko, Martinsson, and Tropp (2011) for a comprehensive review), more specialized methods have been developed recently in order to estimate resolution proxies, such as the trace or the diagonal of the resolution matrix (e.g., An, 2012; MacCarthy, Brochers, & Aster, 2011; Trampert & Fichtner, 2013b).

Figure 8 shows the result of computing posterior covariance for an iterative nonlinear damped least squares inversion based on the data set shown in Figure 1. In this case, no explicit smoothing is applied, so the posterior covariance is defined by

$$\mathbf{C}_M = \mu [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \varepsilon \mathbf{C}_m^{-1}]^{-1} \quad (6)$$

where $\mu = \varepsilon$ when $\varepsilon \geq 1$ and $\mu = 1$ when $\varepsilon < 1$. Strictly speaking, when the covariance matrix is estimated by $\mathbf{C}_M = \mathbf{G}^{-g} [\mathbf{G}^{-g}]^T$, $\mu = 1$ and ε is absorbed into the definition of \mathbf{C}_m . However, if ε is interpreted as a prefactor that allows one to tune prior uncertainty, as $\varepsilon \rightarrow \infty$, $\mathbf{C}_M \rightarrow 0$, which can be misleading. By using the alternative approach suggested in Eqn (6), as $\varepsilon \rightarrow \infty$, $\mathbf{C}_M \rightarrow \mathbf{C}_m$ and as $\varepsilon \rightarrow 0$, $\mathbf{C}_M \rightarrow [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}]^{-1}$ (covariance completely controlled by data), which is more desirable.

The initial model chosen for the Figure 8 example has a constant velocity of 3.0 km/s and the standard deviation of the uncertainty associated with the initial model is set at 0.5 km/s. Figure 8(a) shows the result for $\varepsilon = 1$ and Figure 8(b) shows the result for $\varepsilon = 2000$. The effect of increasing the damping is to decrease the amplitude of the recovered model and increase the

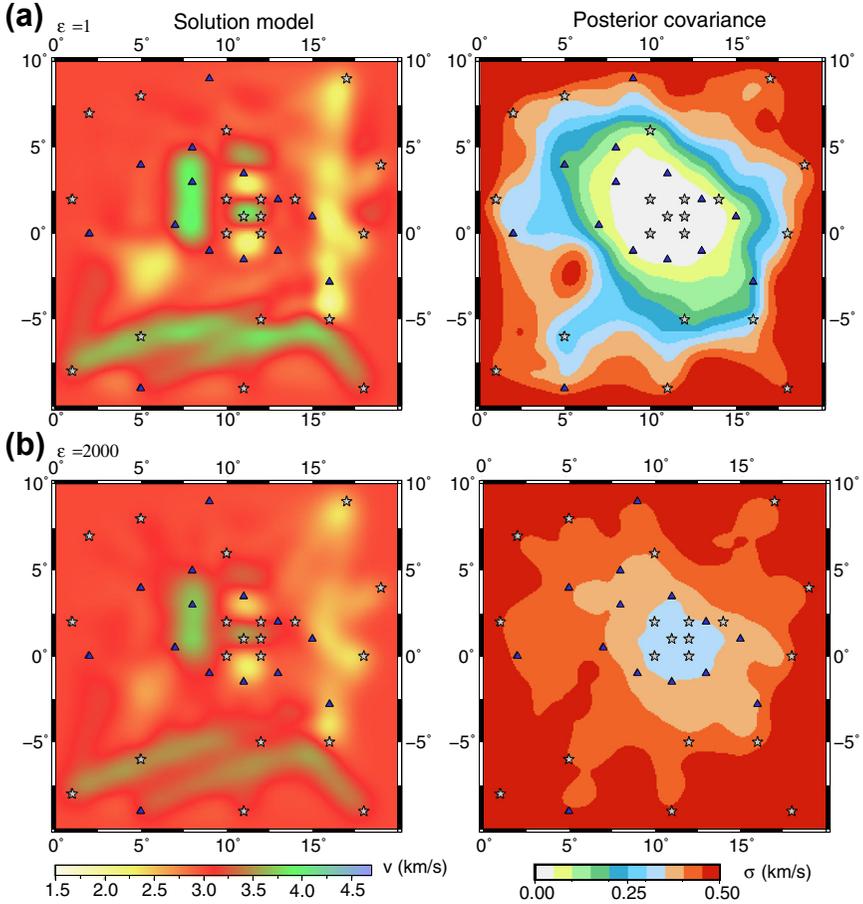


Figure 8 Estimate of posterior covariance for the iterative nonlinear inversion of the data set shown in Figure 1. (a) Damping factor of $\varepsilon = 1$ chosen; (b) damping factor of $\varepsilon = 2000$ chosen. Left-hand column shows the solution model and right-hand column shows the associated estimate of posterior covariance.

posterior covariance estimates (plotted as the standard deviation σ , which is the square root of the diagonal elements of \mathbf{C}_M). The importance of \mathbf{C}_m in the result is clear, as regions of little or no path coverage have a $\sigma \rightarrow 0.5$. The estimates of uncertainty provided by the posterior covariance matrix appear reasonable, at least in a relative sense; for instance, uncertainty is low near the center of the model where path density is high, and higher toward the margins where path coverage drops off. Also, at about 5° east and 2.5° south in Figure 8(a) (right), the local zone of high uncertainty

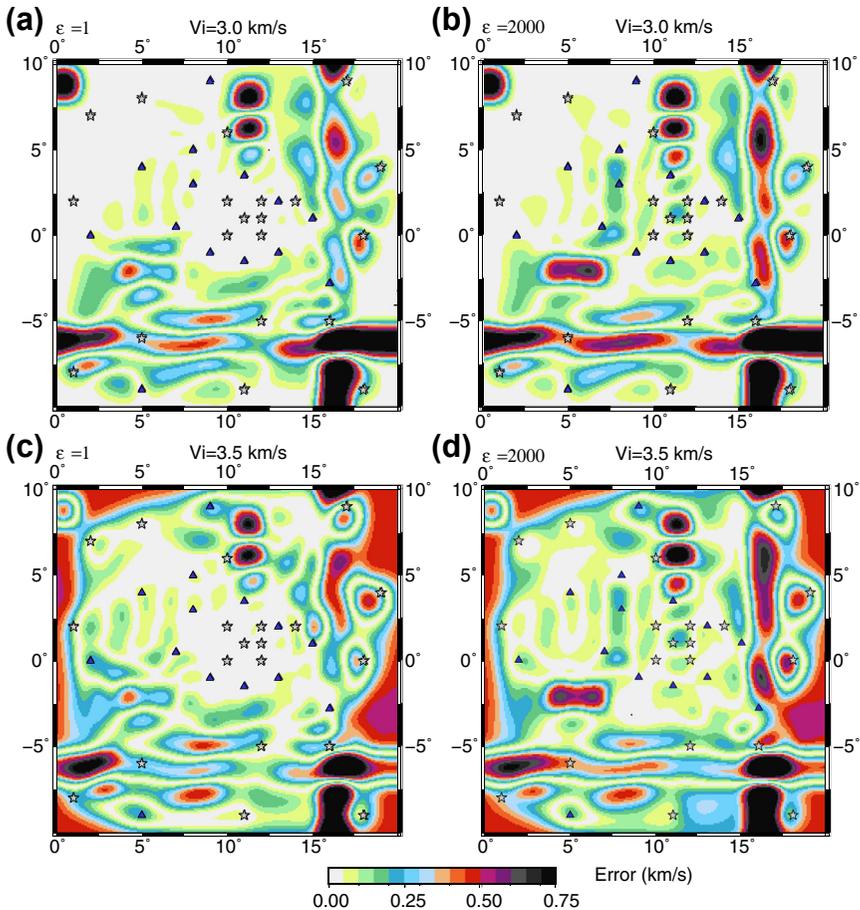


Figure 9 (a) Actual error associated with the solution model shown in Figure 8(a) ($\epsilon = 1$); (b) actual error associated with the solution model shown in Figure 8(b) ($\epsilon = 2000$); (c) same as (a) except using a starting model with a constant velocity of $V_i = 3.5$ km/s; (d) same as (b) except using a starting model with a constant velocity of $V_i = 3.5$ km/s.

corresponds to a low-velocity zone in which path coverage is poor (see Figure 1(b)).

However, if we illustrate the actual errors (Figure 9(a) and (b)) associated with the two inversion results shown in Figure 8, it becomes clear that posterior covariance estimates, certainly for a nonlinear inverse problem, are not very meaningful. Although uncertainty estimates and actual error are not expected to be correlated in general, one would at least hope that where errors are significant, the uncertainty estimate is able to accommodate the

difference, which is often not the case. It should be noted that in this example, the initial model has a velocity (3.0 km/s) that is equal to the background velocity of the true model, which illustrates why the posterior uncertainty need not match the actual error, given that it is reasonable to have higher uncertainty where there is no path coverage. Figure 9(c) and (d) shows the error when an initial model with a velocity of 3.5 km/s is used instead.

Another instructive synthetic example is illustrated in Figures 10–12. In this case, we have a 2-D wide-angle experiment in which refracted and reflected phases are generated and recorded at the surface, and sample a three-layer model in which velocity varies linearly with depth in each layer, and layer boundaries have variable geometry (Figure 10). Interfaces are described by cubic B-spline functions and layer velocities by the linear equation $v = v_0 + kz$, where v_0 is the velocity at the surface and k is the velocity gradient. The inverse problem is to reconstruct the Moho geometry and layer velocities using the synthetic travel times of both the refracted and reflected phases. Gaussian noise with a standard deviation of 70 ms is added to the synthetic travel time data set to simulate picking uncertainties. A shooting method of ray tracing is used to compute the two-point travel times. An iterative nonlinear damped least squares inversion scheme is used that does not include smoothing. Figure 11(a) shows the inversion result, which uses a laterally invariant starting model and velocities with around 10% perturbation from the true model. In general, the reconstruction is quite accurate, except for the concave-up zones of the interfaces, which are not well sampled by first-arrival reflection phases. The covariance and resolution (Figure 11(b) and (c)) plots appear to reflect these uncertainties quite well, with the largest σ values tending to occur in the concave-up regions of the interface (Figure 11(b)). Part of the reason for the results appearing to be more reliable in comparison to the Figures 8 and 9 example is that there are many more data than unknowns (overdetermined inverse problem) in the wide-angle example, whereas the surface wave example is much more mixed determined.

Figure 12 shows the correlation between three different interface nodes and the remaining unknowns for the Figure 11(a) solution. Again, this plot shows that the interface nodes tend to be well resolved, although it is interesting to observe oscillatory behavior between the reference node and surrounding interface nodes. This may be due to the use of cubic B-splines, which use a weighted combination of neighboring nodes to define a value at a single point. Increasing or decreasing the depth of a single node can be

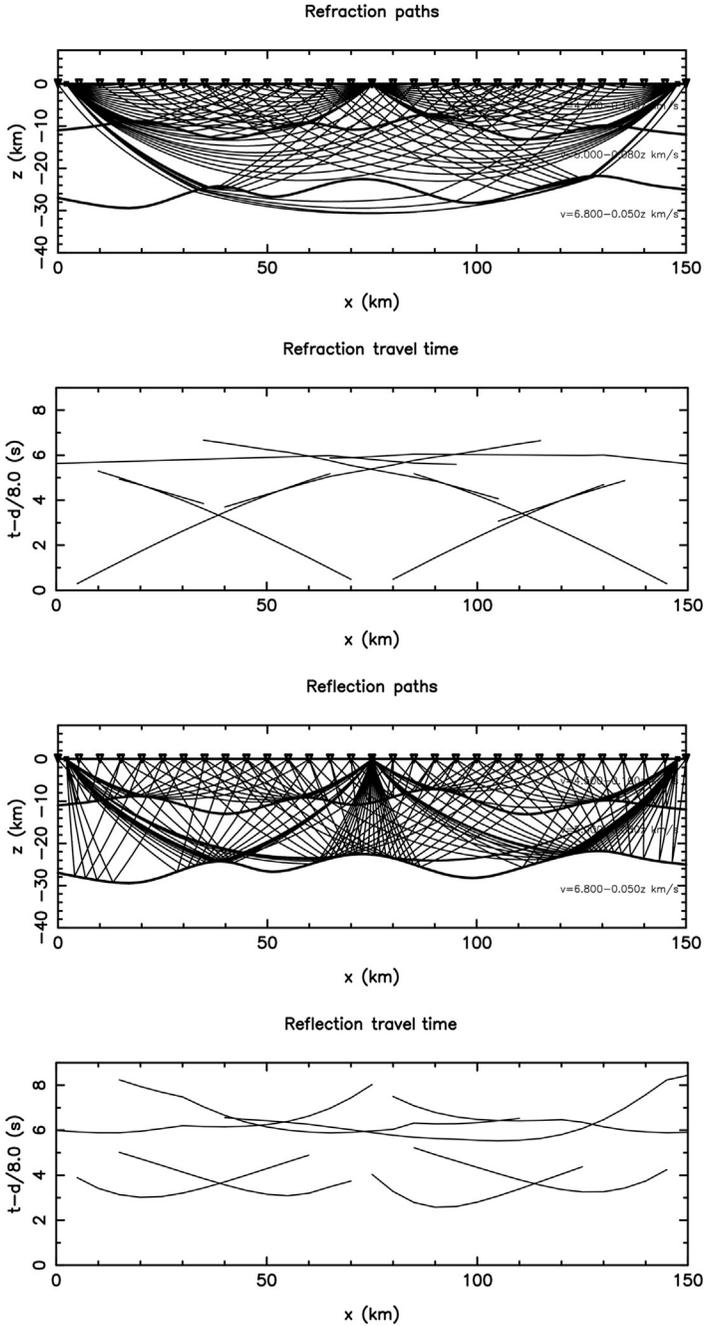
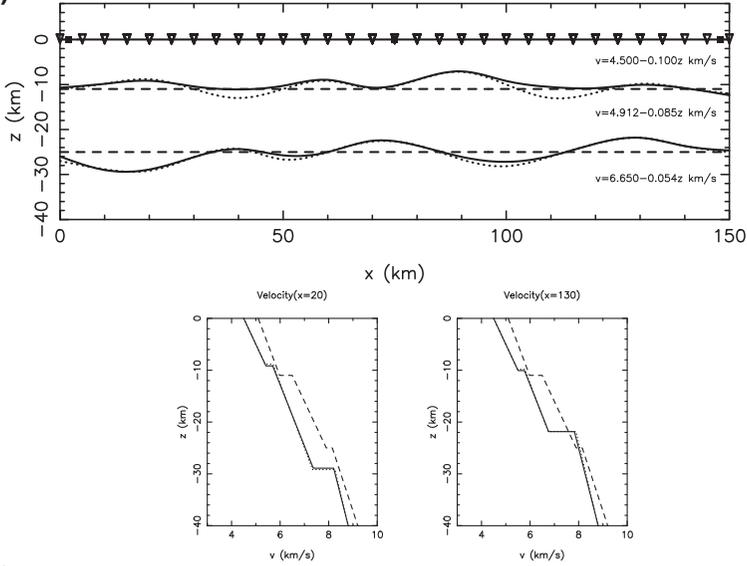
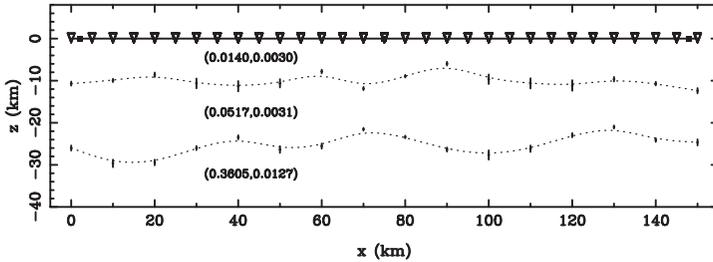


Figure 10 Synthetic 2-D wide-angle data set consisting of refraction and reflection phases. The associated travel time curves are shown beneath each phase type. Top: refraction arrivals; bottom: reflection arrivals.

(a) Damped least squares solution



(b) Covariance



(c) Resolution

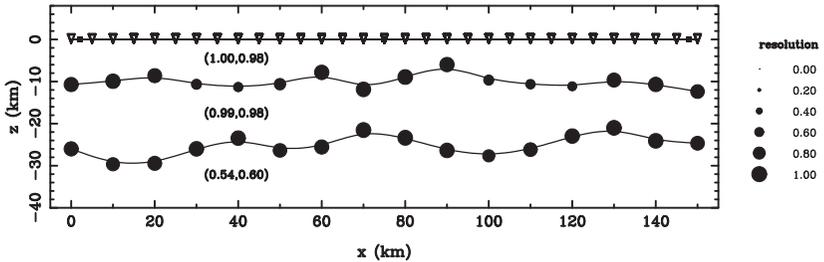


Figure 11 (a) Damped least squares inversion of data shown in Figure 10, using a laterally invariant starting model. Dashed lines show initial interfaces, solid lines show recovered interfaces, and dotted line shows true interfaces; (b) posterior covariance for the model shown in (a); (c) diagonal element of resolution matrix for the model shown in (a). For both (b) and (c), the two numbers in parentheses within each layer represents the error and resolution of the layer velocity parameters, respectively.

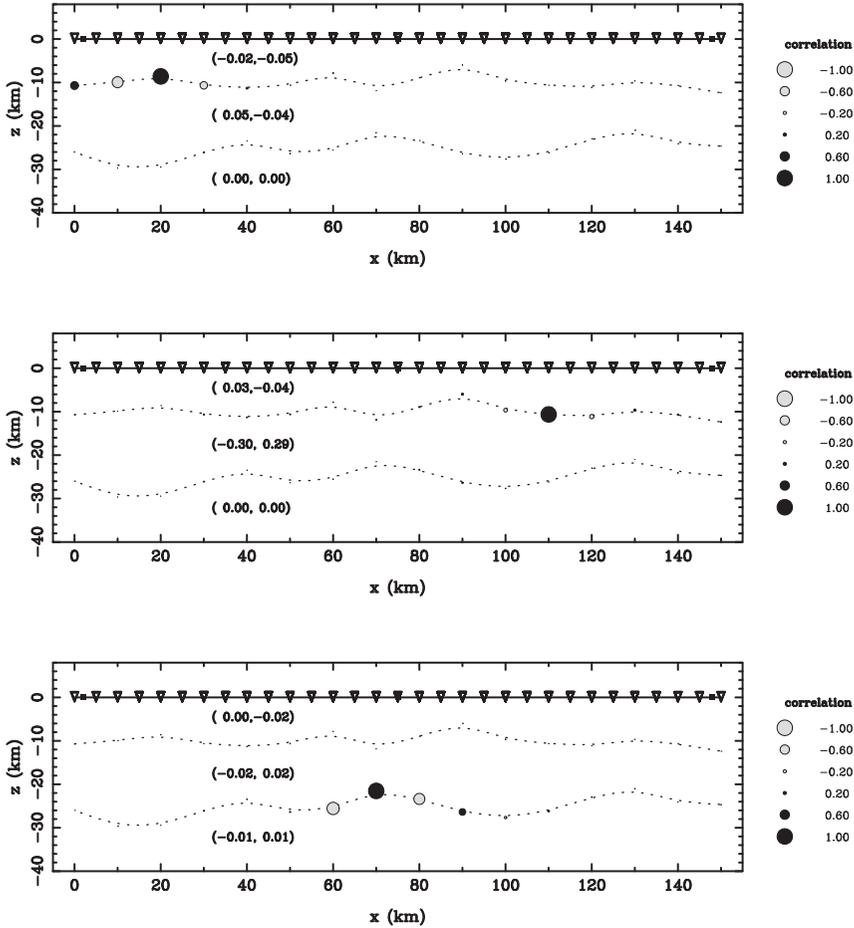


Figure 12 Examples of correlations computed between three separate interface nodes and all other model parameters for the solution shown in [Figure 11\(a\)](#). The two numbers in parentheses within each layer represent the correlation values of the two velocity parameters with respect to the reference interface node.

traded off to some extent by decreasing or increasing, respectively, the depth of an adjacent node.

3.2 Jackknife and Bootstrap

Both the jackknife and bootstrap tests are standard statistical methods of error assessment. The bootstrap test involves performing repeat inversions with a resampled data set (i.e., a new data set formed by taking samples from an original data set) and examining the characteristics of the model ensemble

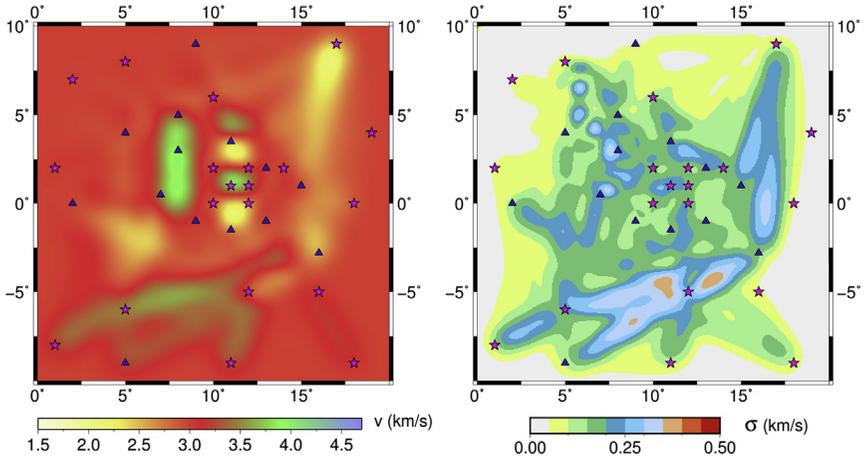


Figure 13 Bootstrap test applied to the [Figure 1](#) data set. The plot on the left shows the average model, while the plot on the right shows the standard deviation of the model ensemble; 50 models were generated for this test. Comparison with [Figure 8](#) shows that where data constraints are absent, the bootstrap uncertainty tends to zero, whereas the covariance tends to the prior estimate of model uncertainty. The former effect is purely due to the implicit regularization imposed by the subspace inversion scheme, which does not alter the value of a parameter unless it is influenced by data.

that is produced. The resampling of the data set can be performed randomly with replacement; thus, a single piece of data can be used more than once, and the size of the new data set is set equal to that of the original ([Efron & Tibshirani, 1993](#)). The jackknife test is similar, but instead of random sampling of a data set, each separate inversion is carried out by omitting a different set of observations. In the case where a single observation is omitted per iteration, for N observations, N inversions are carried out and the resulting ensemble of solutions can be interrogated to produce summary information. A number of tomography studies have used jackknifing to assess solution robustness ([Gung & Romanowicz, 2004](#); [Lees & Crosson, 1989, 1990](#); [Su & Dziewonski, 1997](#); [Zelt, 1999](#)); however, as pointed out by [Nolet et al. \(1999\)](#), both bootstrapping and jackknifing rely on overdetermined inverse problems, and these do not often arise in seismic tomography.

[Figure 13](#) shows an example of the bootstrap test applied to the [Figure 1](#) data set. As before, the initial model is defined by a constant velocity of 3.0 km/s. Damping and smoothing are turned off in the inversion, but implicit regularization is still in place via the choice of a cubic B-spline parameterization with a finite separation of grid nodes. The solution model is defined in this case by the average model, while the uncertainty is

represented by the standard deviation of an ensemble of 50 models. Where there is no path coverage, the uncertainty drops to zero despite the absence of explicit regularization; this occurs because a subspace inversion technique is used (Kennett et al., 1988), which will not adjust parameters that have a zero Fréchet derivative. Consequently, the uncertainty estimate only has meaning in regions of good path coverage, where the pattern of model variability bears some resemblance to path density. The amplitude of σ is significantly underestimated, however, which is in part due to the need to regularize (implicitly in this case) mixed and underdetermined inverse problems. As such, it appears that bootstrapping is not very useful for seismic tomography, particularly when heterogeneous path coverage is present.

3.3 Synthetic Reconstruction Tests

The synthetic reconstruction test is the most common, and perhaps the most criticized, method for assessing solution robustness in seismic tomography. All it essentially requires is for a synthetic or test model to be defined and an artificial data set to be generated in the presence of this model using an identical source–receiver geometry and phase types as the observational data set. The inversion method is then applied in an attempt to recover the synthetic structure. Differences between the true model and the reconstruction form a basis for assessing the reliability of the solution. A particular variant of this approach known as the checkerboard test, in which the synthetic model is defined by an alternating pattern of positive and negative anomalies in each dimension, has been one of the mainstays of seismic tomography studies for the last quarter of a century (e.g., Achauer, 1994; Aloisi, Cocina, Neri, Orecchio, & Privitera, 2002; Chen and Jordan, 2007; Fishwick, Kennett, & Reading, 2005; Glahn & Granet, 1993; Gorbатов, Widiyantoro, Fukao, & Gordeev, 2000; Pilia, Rawlinson, Direen, Cummins, & Balfour, 2013; Rawlinson, Tkalčić, & Reading, 2010; Spakman & Nolet, 1988; Zelt & Barton, 1998). Other types of synthetic structures have also been used, including discrete spikes, volumes of various geometric shapes, and structures designed to mimic particular features such as subduction zones (Eberhart–Phillips & Reyners, 1997; Graeber & Asch, 1999; Hole, 1992; Rawlinson et al., 2006; Walck & Clayton, 1987; Wolfe, Solomon, Silver, VanDecar, & Russo, 2002).

The synthetic reconstruction test has a number of weaknesses, including (1) accounting for data noise, which is often poorly constrained, is difficult. Simply adding Gaussian noise with a particular standard deviation to the synthetic data may poorly represent the actual noise content of the

observational data set; (2) using identical parameterization for the synthetic model and reconstructed model will yield a result biased in favor of a good reconstruction; (3) similarly, using the same forward problem solver for the computation and the inversion of the artificial data yields overly optimistic results because errors in the forward problem solution are disregarded; (4) results can vary according to the input structure used, particularly when the inverse problem is nonlinear. In the latter case, [Lévêque et al. \(1993\)](#) demonstrated with a simple test example that even for linear inverse problems the checkerboard test can be misleading, and in some circumstances can reproduce small-scale structure more accurately than large-scale structure. Partly as a result of such caveats, it is common to find examples of checkerboard tests carried out across a range of scales (e.g., [Fishwick et al., 2005](#)), coupled with some other style of reconstruction test (e.g., [Rawlinson et al., 2006](#)) or carried out together with some other measure of uncertainty such as covariance and resolution (e.g., [Graeber & Asch, 1999](#)).

[Figure 14](#) shows an example of the synthetic checkerboard test applied to the data set in [Figure 1](#). The original heterogeneous model (input model) is shown in [Figure 8\(a\)](#). Although the reconstruction indicates that the basic pattern of the checkerboard is recovered, this is not really the case when one inspects [Figure 8](#) or [9](#), and so the checkerboard could be construed as somewhat misleading in this regard. In general, the amplitudes are underestimated, which is typical of a damped least squares solution. The amplitudes are, unsurprisingly, most accurate in the region of dense path coverage near the center of the model. Part of the reason for the relatively poor performance of the checkerboard in this case as a proxy for the uncertainties in the [Figure 8\(a\)](#) reconstruction can be attributed to the very different path coverage between [Figure 14\(b\)](#) and [\(d\)](#), which is a function of the significant wave speed anomalies that are present. This is particularly noticeable in regions of low ray density when paths have some distance to travel, such as in the southern region of the model. The general tendency with synthetic reconstruction tests is to appraise them qualitatively, which in this case may result in misleading inferences about the robustness of the actual model recovery.

3.4 Linear and Iterative Nonlinear Sampling

As noted earlier, the underdetermined nature of the seismic tomography problem means that a potentially wide range of models may satisfy the data and a priori constraints. Yet most solution strategies end up yielding a single data satisfying model from which inferences are made. Published

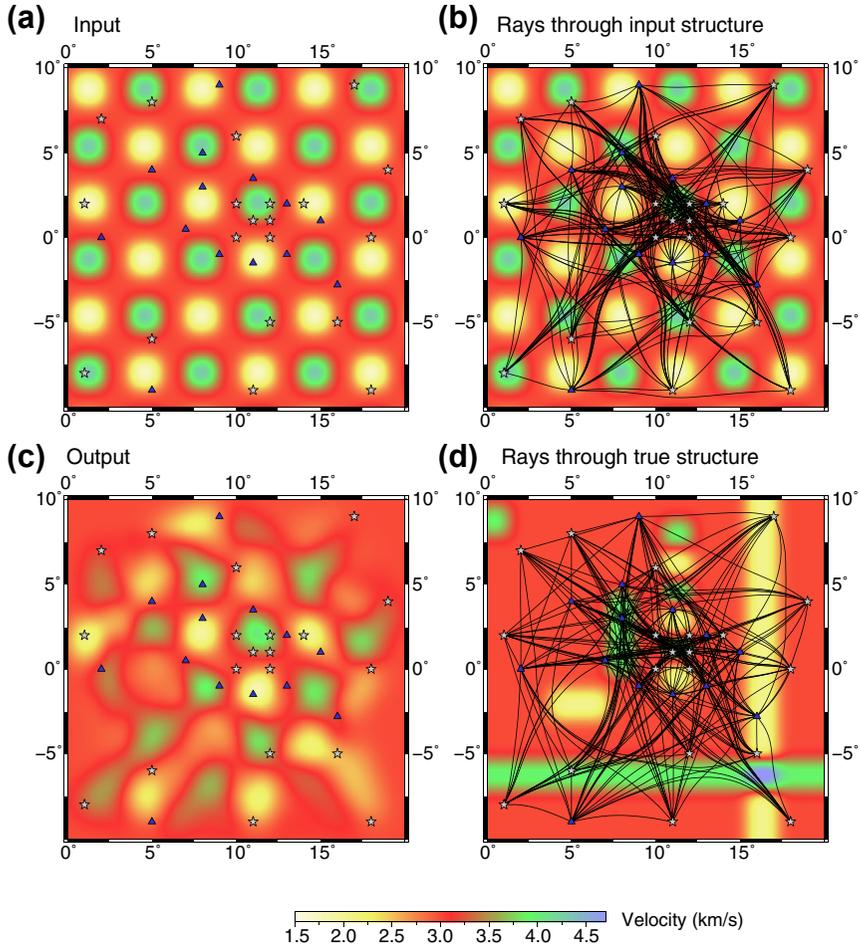


Figure 14 Example of a checkerboard reconstruction test for the [Figure 1](#) data set and a iterative nonlinear damped least squares solution. (a) Synthetic checkerboard model; (b) ray path geometry through synthetic checkerboard model; (c) reconstructed model using inversion method; (d) actual path coverage through true model.

studies tend to implement some kind of qualitative or quantitative assessment of solution robustness but, as has been pointed out, these are often of limited value. Inversion strategies that aim to produce an ensemble of data-fitting models are not tied to a particular configuration of features in the solution; instead, a range of potentially plausible structures are recovered from which summary information can be extracted that highlight those features most required by the data.

Within a linear framework, [Deal and Nolet \(1996\)](#) develop the so-called null-space shuttle, which exploits the underdetermined nature of the linear system of equations that define the inverse problem in order to yield more than one data-satisfying solution. The null-space shuttle is the operator that allows movement from one solution to another without corrupting data fit. It does this by filtering a solution model a posteriori, where the filter is restricted to operate only on components of the solution that do not affect the data fit. The filter that is used can assume a variety of forms depending on the a priori information; examples include a smoothing filter, or one designed to emphasize sharp boundaries. Once the filter operates on the solution model, the difference between the new model and the filtered model is projected onto the null-space, which has the effect of removing any changes that degrade the fit to data. In their study, [Deal and Nolet \(1996\)](#) apply the technique to synthetic data to demonstrate that, where a filter based on good a priori information is available, a more accurate model can be obtained by applying the null-space shuttle method to the minimum norm solution. In a subsequent application of the method ([Deal, Nolet, & van der Hilst, 1999](#)) to image the Tonga subduction zone, the travel time tomography model obtained from the inversion of teleseismic and local P-wave travel times is enhanced by biasing it toward a theoretical slab temperature model based on the diffusion equation. Projecting the difference between the seismic tomography model and the temperature model (assuming velocity is a function of temperature) onto the null-space of the inversion removes components of the slab temperature model that violates the travel time data fit.

[de Wit et al. \(2012\)](#) generalize the null-space shuttle method proposed by [Deal and Nolet \(1996\)](#) in order to estimate quantitative bounds on the tomographic model with the goal of producing a range of different but acceptable models. The new technique is applied to a very large global body wave travel time data set. They found that accurate estimates of data uncertainty are crucial for obtaining a reliable ensemble of models. Furthermore, the solution range also depends on the choice of regularization that is required by the inversion of the underdetermined system of equations; in particular, the range of acceptable models becomes larger as the regularization is decreased.

The scheme proposed by [de Wit et al. \(2012\)](#) is similar to the so-called regularized extremal bounds analysis (REBA) of [Meju \(2009\)](#), which finds a range of data-fitting models given a set tolerance on the objective function. Although it is designed for nonlinear geophysical inverse problems, REBA is

based on iterative updates using a local quadratic approximation to the objective function and regularized system of linear equations, and therefore is dependent on a number of assumptions, unrelated to the data, in order to estimate the range of permissible models. Vasco (2007) uses an alternative approach to exploiting the null-space of the tomographic inverse problem by applying Lie group methods that do not require linearization about a reference model. As such, it can be viewed as a generalization of the null-space shuttle method of Deal and Nolet (1996) for nonlinear problems.

An alternative approach for searching model space for data fitting models within an iterative nonlinear framework is the so-called dynamic objective function scheme of Rawlinson et al. (2008). The basic principle behind the method is to exploit information gained from previous solutions to help drive the search for new models. Rather than attempt to minimize a fixed objective function, a feedback or evolution term is included that modifies the misfit landscape in accordance with the location of previous solutions. The form of the objective function used in Rawlinson and Kennett (2008) is

$$S_j(\mathbf{m}) = \frac{1}{2} \left[(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}) + \sum_{i=1}^j \frac{1}{\lambda \left[(\mathbf{m} - \mathbf{m}_i)^T (\mathbf{m} - \mathbf{m}_i) \right]^p + \zeta} \right] \quad j = 1, \dots, N \quad (7)$$

which discards the usual damping and smoothing terms and instead introduces a function that creates a local peak in the objective function at values of \mathbf{m} corresponding to all previous solutions $j = 1, \dots, n$ (where $n < N$) that have been located. The aim is to penalize new solutions from converging on previous solutions unless the data are sufficiently persuasive. The terms p , λ , and ψ control the shape and amplitude of the local maximum as defined in Figure 15. With appropriate choices of these terms, it is possible to produce a relatively small ensemble of models that together contain the most robust features that can be inferred from the data. The main weakness of the method is that, although damping and smoothing regularization is discarded, appropriate choices of p , λ , and ψ need to be found and are problem dependent. An example of the dynamic objective function technique applied to observational data is given in the next section.

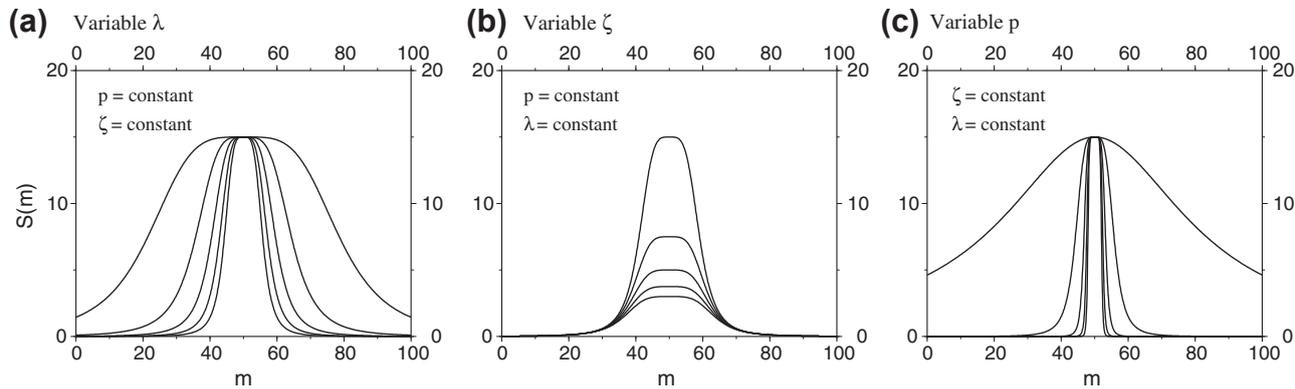


Figure 15 Graphical representation of how the variables p , λ , and ζ in Eqn (7) influence the shape of the evolution term. From [Rawlinson et al. \(2008\)](#). Copyright 2008 Royal Astronomical Society. Reproduced by permission of Oxford University Press.

3.5 Fully Nonlinear Sampling

Inversion methods that avoid the assumption of local linearization and provide a thorough interrogation of model space in order to produce an ensemble of data-satisfying models are the most attractive for addressing the nonlinear relationship between observables and model parameters. For some tomography problems, such as global or teleseismic travel time tomography, where ray paths do not strongly deviate from global reference model predictions, linear and iterative nonlinear schemes can be relatively robust. However, when wave speed heterogeneity is significant and prior information is limited (e.g., crustal or near-surface studies), nonlinear sampling methods are potentially of the greatest benefit. Full waveform tomography also becomes increasingly nonlinear at higher frequencies due to cycle skipping issues. However, nonlinear sampling methods are much more computationally expensive than methods based on linearization, and consequently have only enjoyed limited exposure in realistic seismic tomography problems.

Common nonlinear inversion methods used in the physical sciences, including genetic algorithms, which use an analog to biological evolution to drive the search for new models, and simulated annealing, which is based on an analog with physical annealing in thermodynamic systems, have been used to solve geophysical inverse problems (Mosegaard & Sambridge, 2002; Sambridge & Mosegaard, 2002); however, application to 2-D and 3-D tomography problems is limited due to the large number of unknowns involved (Asad et al., 1999; Boschetti et al., 1996; Pullammanappallil & Louie, 1993).

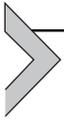
Surface wave tomography is more amenable to nonlinear inversion methods than body wave tomography, because the problem can be posed as a composite 2-D and 1-D inverse problem rather than a fully 3-D inverse problem. For instance, one can invert group or phase dispersion for 2-D period-dependent group or phase velocity maps, and then carry out a point-by-point inversion for 1-D shear wave velocity in order to build a 3-D model. Meier, Curtis, and Trampert (2007a, 2007b) use neural network inversion to invert fundamental mode Love and Rayleigh phase and group velocity maps for a global model of the crust and uppermost mantle. As well as addressing the nonlinearity of the inverse problem, this approach has the benefit of providing the posterior probability distribution of model parameters, thus allowing a quantitative assessment of uncertainty. However, the number of unknowns in the 1-D inverse problem is limited (29 in this

case), and the reliance on 2-D group and phase velocity maps derived from a linear inversion means that the full nonlinearity of the complete problem is not addressed. [Shapiro and Ritzwoller \(2002\)](#) carry out a similar study using a large data set of fundamental mode Rayleigh and Love wave group and phase velocities, but instead use a Markov chain Monte Carlo (MCMC) method to yield an ensemble of data-satisfying models.

The MCMC approach to solving the nonlinear inverse problem is gradually growing in popularity in seismic tomography. [Bodin and Sambridge \(2009\)](#) implement the reversible jump variant to solve a transdimensional inverse problem in which the number and spatial distribution of model unknowns vary in addition to their values. The inverse problem is solved within a Bayesian framework, which means that information is represented by probability density functions. The goal of Bayesian inference, within a linear or nonlinear setting, is to quantify the posterior probability distribution given a prior distribution and constraints provided by the data. The posterior probability distribution is defined by an ensemble of data satisfying models generated by the Markov chain following an initial burn-in phase. Information such as the mean and standard deviation can be extracted from the ensemble. [Bodin and Sambridge \(2009\)](#) apply the scheme to a 2-D surface wave test problem in which rays are only updated after every N th model is generated (where N is large) in order to minimize computational resources. In this sense, the technique is ultimately iterative nonlinear rather than fully nonlinear. However, since N is a variable and linearization is not inherent to the inversion scheme, it is possible to make the scheme fully nonlinear by setting $N = 1$. This is done in the transdimensional tomography study of [Galetti et al. \(submitted for publication\)](#), where surface wave group dispersion is inverted for period-dependent group velocity maps.

Stochastic sampling methods provide a robust way of extracting meaningful information from sparse data sets, but they still require an accurate knowledge of data noise; in the absence of such information, the range of data-fitting models becomes an unknown variable. In the context of seismic tomography, [Bodin, Sambridge, Tkalcic, et al. \(2012\)](#) introduce the so-called hierarchical Bayesian inversion scheme, an extension of the Bayesian transdimensional scheme, which in addition to the number, value, and distribution of model parameters, allows the level of noise (e.g., represented by the standard deviation) to be an unknown in the inversion. This is particularly useful, as the absolute level of noise (including picking error and

modeling error, the latter being the inability of the forward model to explain the data) is usually poorly known. [Bodin, Sambridge, Rawlinson and Arroucau \(2012\)](#) apply the new scheme to surface wave dispersion data from Australia that comprises three separate experiments carried out at very different scales. The standard deviation of the data noise is treated as a linear function of interstation distance in order to account for the large range of interstation path lengths.



4. CASE STUDIES

Four different case studies are presented below, which use different means of assessing model robustness. The first example showcases the synthetic reconstruction test that is commonly used in seismic tomography. The remaining three examples apply more recently developed techniques for assessing model robustness, including iterative nonlinear sampling, trans-dimensional tomography, and resolution analysis for full waveform tomography.

4.1 Synthetic Reconstruction Test: Teleseismic Tomography Example

Following the early work of [Aki et al. \(1977\)](#), teleseismic tomography has become very popular for imaging the structure of the crust and lithosphere in 3-D ([Glahn & Granet, 1993](#); [Graeber et al., 2002](#); [Humphreys & Clayton, 1990](#); [Oncescu et al., 1984](#); [Rawlinson & Fishwick, 2012](#); [Rawlinson & Kennett, 2008](#); [Saltzer & Humphreys, 1997](#)) despite its well-known drawbacks. These include ignoring lateral variations in structure outside the model region that may contribute to the measured arrival time residual and the subvertical incidence of the seismic energy at the receiver. The latter effect results in relatively poor resolution in the vertical direction, while the mapping of arrival time residuals as wave speed variations within a limited model region beneath the array may introduce unwanted artifacts.

Here we present an example of teleseismic tomography applied to Tasmania, southeast Australia, with the main goal of assessing the results of an associated synthetic reconstruction test. Full details of the methods, results, and interpretation can be found in [Rawlinson et al. \(2006\)](#). Data for the study comes from an array of 72 recorders deployed across northern Tasmania in 2001 and 2002 (see [Figure 16\(a\)](#)). A total of 6520 relative

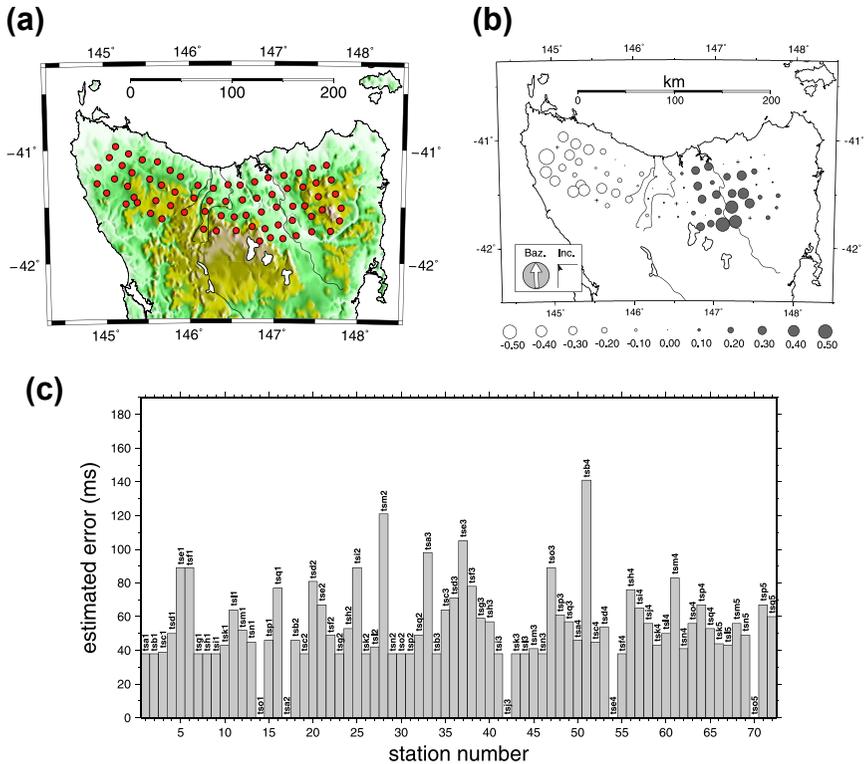


Figure 16 (a) A 72-station Tigger array (deployed between 2001 and 2002) with an average station separation of 15 km; (b) plot of teleseismic arrival time residuals for an event from the Marianas; (c) estimate of uncertainty associated with the extraction of arrival time residuals for the Marianas event.

P-wave arrival time residuals from 101 teleseismic sources are extracted from the seismic records using the adaptive stacking technique of [Rawlinson and Kennett \(2004\)](#). [Figure 16\(b\)](#) shows a map of the P-wave residuals for an event from the Mariana Islands. The adaptive stacking technique also produces an estimate of picking uncertainty ([Figure 16\(c\)](#)), which is used to weight the contribution of residuals in the tomography. A minimum uncertainty threshold of 37.5 ms (75% of the sample interval) is imposed in recognition of noise and waveform incoherence across the array. An iterative nonlinear inversion scheme is applied to map the arrival time residuals as velocity variations; the objective function includes damping and smoothing regularization to control the amplitude and wavelength of retrieved structure. Trade-off curves are used to decide the appropriate damping and

smoothing. The forward problem of travel time prediction is solved using a grid-based eikonal scheme that robustly finds first arrivals (Rawlinson & Sambridge, 2004). The inverse problem is solved iteratively using a subspace inversion scheme (Kennett et al., 1988), with arrival times recomputed after each model update.

Figure 17 shows a depth slice and an east–west slice through the Tasmania solution model, obtained after six iterations using a 10-D subspace scheme. The data variance is reduced by 74%, which corresponds to an Root mean square (RMS) reduction from 193.7 ms to 98.7 ms. From the adaptive stacking results, the estimated data noise is 77 ms, which indicates that there is likely to be a component of “modeling noise” due to implicit (from the parameterization) and explicit (from the damping and smoothing) regularization and forward modeling assumptions. Most of the recovered structures look plausible, although the edge of the vertical slices appear to contain unrealistic streaking effects.

In order to investigate the robustness of the solution, a synthetic checkerboard test is carried out using three different checkerboard sizes ranging between approximately 25 and 50 km (see Figure 18). Gaussian noise with a standard deviation of 77 ms is added to all three synthetic data sets in order to simulate the effects of picking noise. Arguably, one could use noise with a standard deviation of 98.7 ms in order to reproduce the same fit to data experienced by the real model. In addition, rather than use identical parameterizations for the synthetic and recovered model, it would be more realistic to use difference parameterizations. However, this approach is the convention in seismic tomography, and the checkerboard recovery that is achieved can be regarded as being on the optimistic side of the truth. Figure 19 shows the output model, which in general shows a good recovery of the pattern of anomalies across all three scales. The region of good recovery is most extensive for the large checkerboard, and most restricted for the small checkerboard. This is an expected result given the known trade-off between resolution and covariance. On the vertical sections (Figure 19(b) and (c)), significant streaking can be observed toward the edge of the model where crossing path coverage diminishes.

One of the limitations of a checkerboard test such as that illustrated in Figures 18 and 19 is that the extent of near-vertical distortion of structure is difficult to fully appreciate due to the structure of the checkerboard, where the diagonal elements in the vertical plane are closely aligned with dominant ray directions. To address this issue, spike tests in which discrete anomalies placed some distance apart represent a more robust test. Figure 20 shows

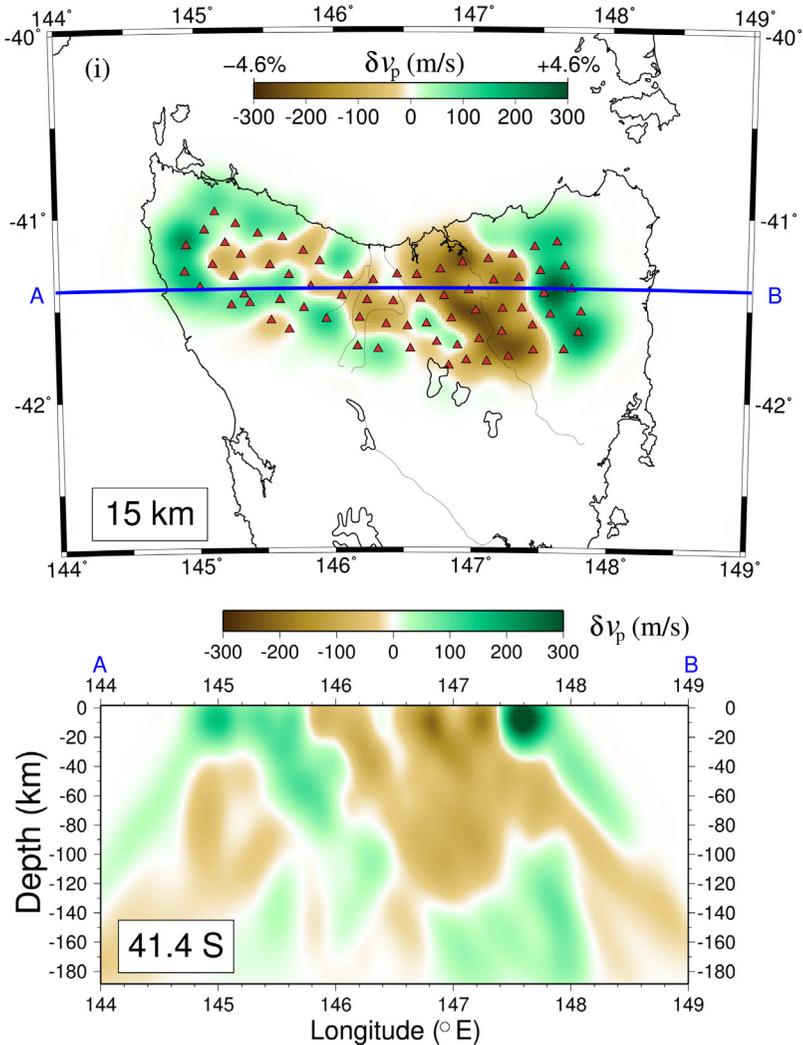


Figure 17 Horizontal and vertical slice through the Tasmania solution model obtained via iterative nonlinear inversion of teleseismic arrival time residuals. *Modified from Rawlinson et al. (2006). Copyright 2006 American Geophysical Union. Reproduced by permission of American Geophysical Union.*

the result of a synthetic spike recovery test, which, apart from the structure, uses the same settings as the previous checkerboard test. The output shows that even in regions with good path coverage, vertical smearing of structure takes place. This characteristic of the recovery must be accounted for in the interpretation of the results.

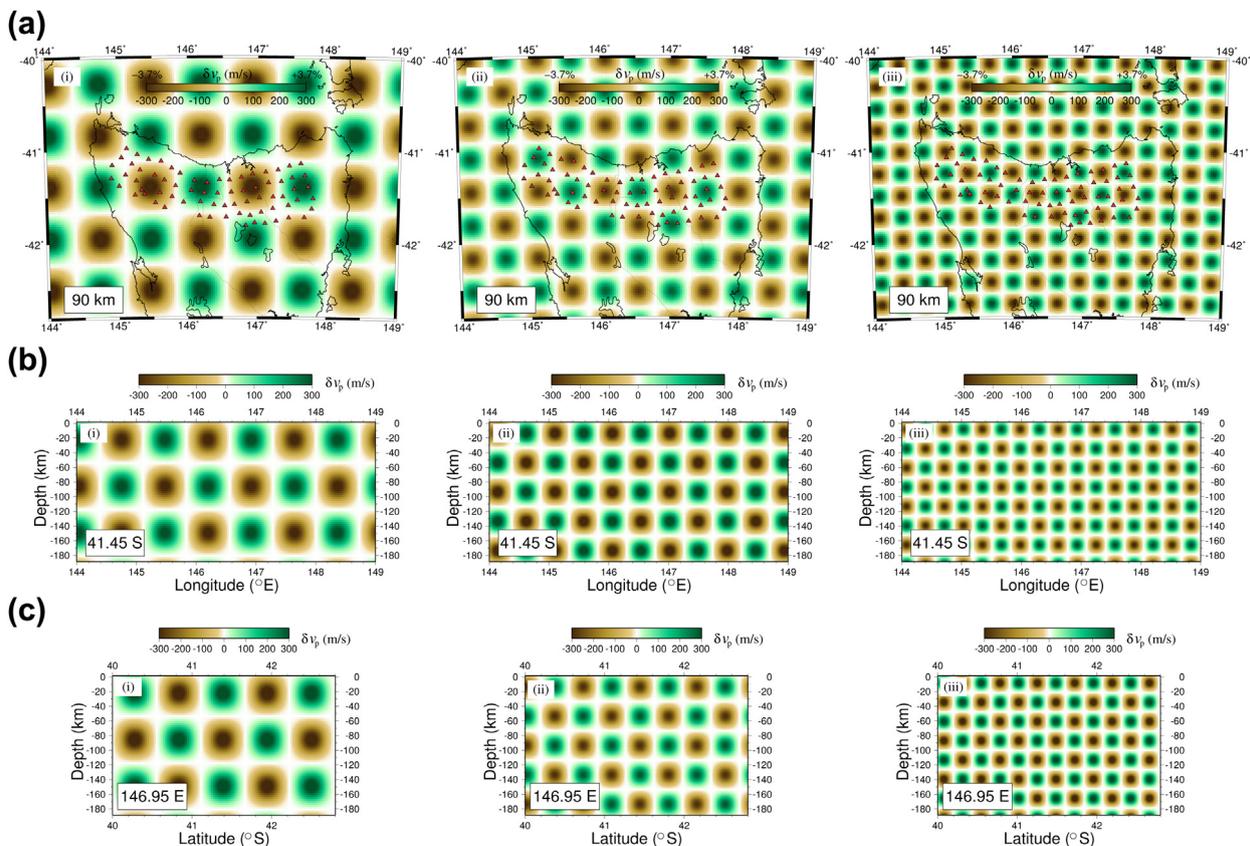


Figure 18 Input checkerboard model for the synthetic recovery test. Three different checkerboard sizes are used. (a) Depth slice at 90 km; (b) E-W slice at 41.45° S; (c) N-S slice at 146.95° E. From [Rawlinson et al. \(2006\)](#). Copyright 2006 American Geophysical Union. Reproduced by permission of American Geophysical Union.

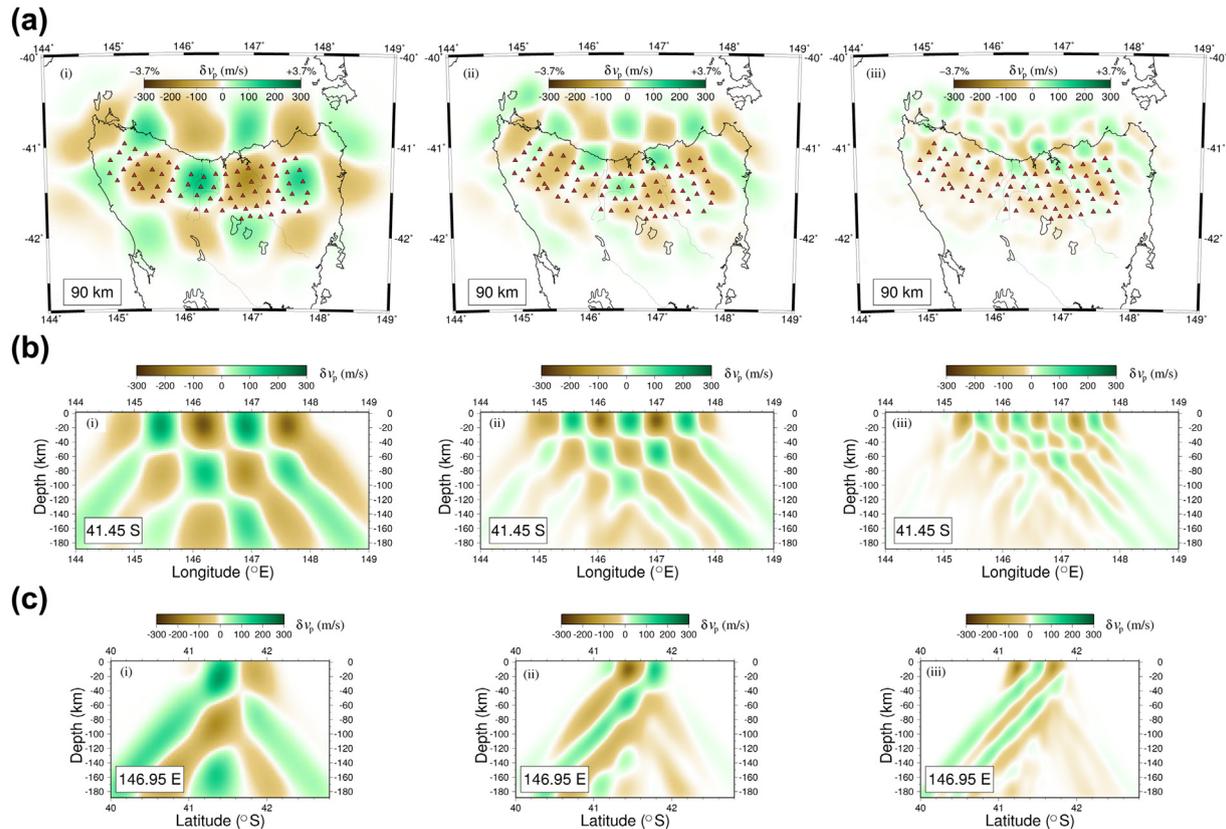


Figure 19 Output of the checkerboard recovery test. Figure 18 shows the input checkerboard pattern. From Rawlinson et al. (2006). Copyright 2006 American Geophysical Union. Reproduced by permission of American Geophysical Union.

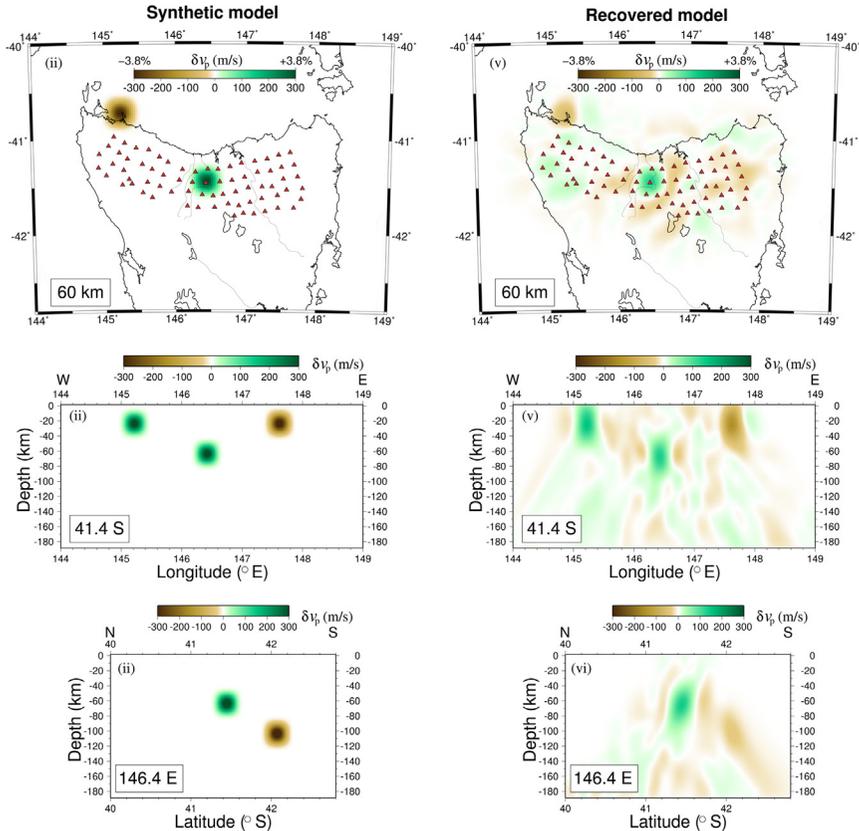


Figure 20 Synthetic test results that involve the recovery of discrete spikes. Apart from the smearing of the spike, there is also a small amount of background “noise” that can be attributed to the inclusion of Gaussian noise in the synthetic data set. From [Rawlinson et al. \(2006\)](#). Copyright 2006 American Geophysical Union. Reproduced by permission of American Geophysical Union.

4.2 Iterative Nonlinear Sampling: Surface Wave Tomography Example

In this case study, the dynamic objective function technique of [Rawlinson et al. \(2008\)](#) is used to invert a regional ambient noise surface wave dispersion data set for the seismic structure of the Australian crust. Since the emergence of virtual source seismology over a decade ago ([Campillo & Paul, 2003](#); [Shapiro & Campillo, 2004](#)), a new class of seismic tomography that exploits long-term cross-correlations of ambient noise data between station pairs has gained significant traction ([Saygin & Kennett, 2009](#); [Shapiro,](#)

Campillo, Stehly, & Ritzwoller, 2005; Yang, Ritzwoller, Levshin, & Shapiro, 2007; Young et al., 2013; Young, Rawlinson, Arroucau, Reading, & Tkalčić, 2011). There are a number of reasons for its popularity, including that ambient noise information can be exploited from legacy data sets; the data coverage is largely determined by the station spacing; and the region of good depth resolution often complements results obtained from other methods (e.g., teleseismic tomography). In most cases, it is the surface wave component of the so-called empirical Green's function that emerges from the cross-correlation, which means that standard surface wave dispersion analysis can be used to constrain structure.

Over the last two decades, the coverage of broadband stations across Australia has greatly improved, with a cumulative total of over 200 to date (see Figure 21(a)). Many of these have been part of temporary arrays, which means that not every station pair has recorded simultaneously, a condition required for the extraction of ambient noise information. However, there are sufficient station pairs available to produce good path coverage throughout much of the Australian land mass (see Figure 21(b)). Here, we invert 6.7 s Rayleigh wave group travel times taken from Saygin and Kennett (2009) for a map of group velocity, which is largely a function of shear wave structure in the upper crust. A total of 2332 travel times are used from 208 stations. Picking uncertainty is set to 5.0 s for all data, which reflects the difficulties associated with estimating noise in dispersion data. A 2-D version of the iterative nonlinear scheme described in Rawlinson et al. (2006)—see previous example—is used as a basis for the reconstruction, although with the objective function modified as in Eqn (7). The initial model is described by a grid of 30×30 cubic B-spline control points in latitude and longitude, with a constant velocity of 2.8 km/s. The parameters λ , p , and ζ (see Eqn (7)) are set to 5×10^{-6} , 1, and 5×10^{-6} , respectively. An ensemble of 25 solutions is generated by repeated minimization of Eqn (7) using the subspace inversion scheme; the eikonal equation is re-solved between each subspace iteration in order to address the nonlinearity of the problem. Increasing the size of the ensemble does little to change the average solution model that is produced (Figure 21(c)). Figure 21(d) shows the standard deviation of the ensemble, which to some extent correlates with path density (Figure 21(b)). However, unlike path coverage or density, the dynamic objective function approach provides a quantitative estimate of uncertainty, and can attach high uncertainty to regions of high ray density if, for example, the angular coverage is poor. In the offshore regions of Australia, where there is no data coverage at all, both the recovered model and the standard

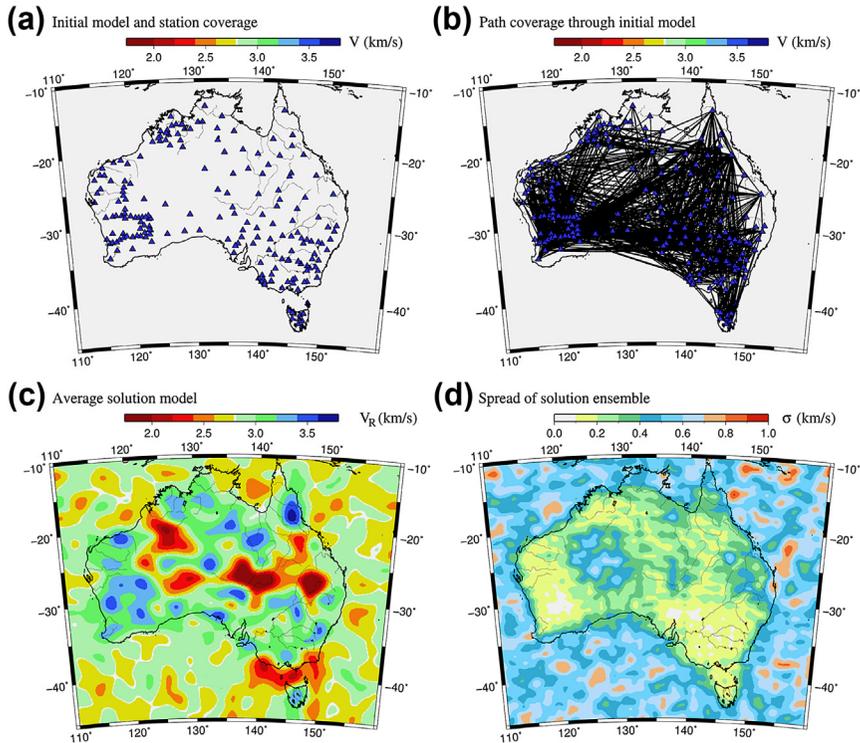


Figure 21 Application of the dynamic objective function technique to an ambient noise data set from Australia. (a) Cumulative broadband station coverage to 2009; (b) interstation surface wave path coverage for ambient noise data set; (c) average solution model taken from an ensemble of 25; (d) standard deviation of model ensemble. From [Rawlinson et al. \(2008\)](#). Copyright 2008 Royal Astronomical Society. Reproduced by permission of Oxford University Press.

deviation contain high-frequency “noise” that reflects the fact that the method does not behave in a truly random fashion in the absence of information. However, the main features of [Figure 21\(c\)](#) appear to correlate well with the broadscale geology of the Australian crust ([Rawlinson et al., 2008](#); [Saygin & Kennett, 2009](#)), and the spread of the ensemble appears reasonable.

In order to help validate the spread of the solution ensemble, which is shown in [Figure 21\(d\)](#), a checkerboard test is carried out ([Figure 22](#)). Gaussian noise with a standard deviation of 5 s is added to the synthetic travel times computed from the input checkerboard ([Figure 22\(a\)](#)). The regions of good pattern recovery in the output model ([Figure 22\(b\)](#)) appear to correlate well with the regions of low model ensemble variability ([Figure 21\(d\)](#)). Where smearing of the anomalies is evident, such as in eastern Queensland

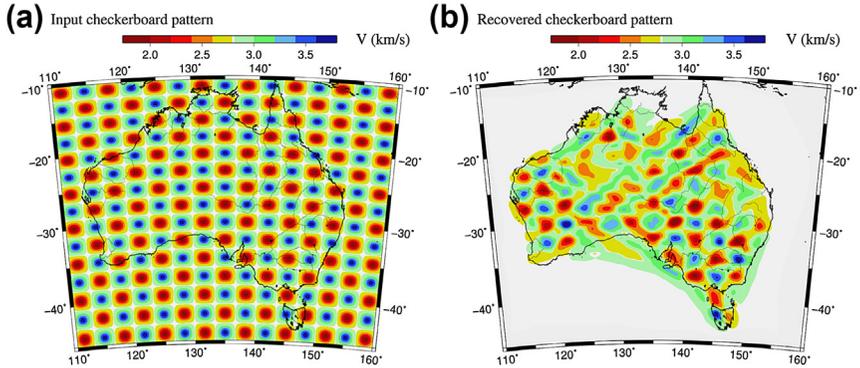


Figure 22 Checkerboard resolution test for the Australian ambient noise data set. (a) Synthetic or input model; (b) recovered or output model. From *Rawlinson et al. (2008)*. Copyright 2008 Royal Astronomical Society. Reproduced by permission of Oxford University Press.

and Western Australia, the model ensembles become less similar. It is reassuring that there are no major discrepancies between the two techniques, but it could be argued that a quantitative measure of uncertainty is more useful.

4.3 Transdimensional Inversion: Surface Wave Tomography Example

In this example, ambient noise data from the WOMBAT transportable array in southeast Australia is inverted for the 3-D shear wave velocity structure of the crust using the transdimensional scheme outlined in the previous section. WOMBAT is the largest transportable array of its kind in the southern hemisphere, and comprises over 700 station locations with a spacing of 50 km on the mainland and 15 km in Tasmania (*Figure 23*). Here, data from a subset of the mainland stations (using arrays enclosed by the red line in *Figure 23*) are used in the transdimensional inversion. Full results can be seen in the paper of *Young et al. (2013a)*. Rayleigh wave phase velocity dispersion measurements from the cross-correlation of data from 434 stations over a period range of 1–20 s is used in the inversion. The hierarchical form of the transdimensional tomography scheme is used to invert the interstation phase dispersion for period-dependent phase velocity maps. This means that the standard deviation of the data noise is treated as an unknown in the inversion, where the noise includes whatever the model cannot explain, including measurement errors, approximations in the forward solver, and assumptions implicit to the parameterization. In the latter

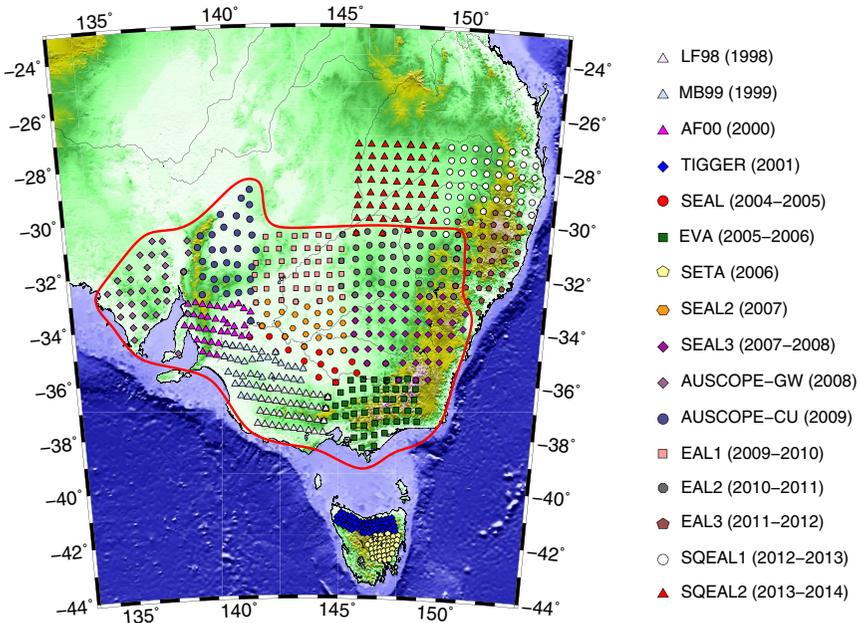


Figure 23 Station locations of the WOMBAT transportable seismic array in southeast Australia. The names of the component arrays and their date of installation are shown in the key on the right. The thick red line encloses those arrays from which data are extracted for the transdimensional inversion. (For interpretation of the references to color in this figure legend, the reader is referred to the online version of this book.)

case, constant velocity Voronoi cells of variable size, number, and distribution are used, which allows for great flexibility in the representation of structure, and imposes a form of implicit regularization.

The compute time for a single-phase velocity map is of the order of 1000 CPU hours, which is roughly four orders of magnitude greater than that required for a more traditional iterative nonlinear approach. However, the improvement in quality, plus the robust estimate of model uncertainty, makes this extra cost acceptable. The prior bound on the velocity variations is set at 2.4 km/s, and is centered on the average phase velocity observed for the period in question. Increasing the bounds has little effect on the resultant solution. [Figure 24](#) shows an example of the output at 1 and 5 s. The average velocity model and its associated standard deviation is obtained from the posterior distribution. Due to the transdimensional nature of the inversion, acceptable models can be represented by different numbers and distributions of the Voronoi cells. As [Figure 24\(a\)](#) shows, there is a tendency for the

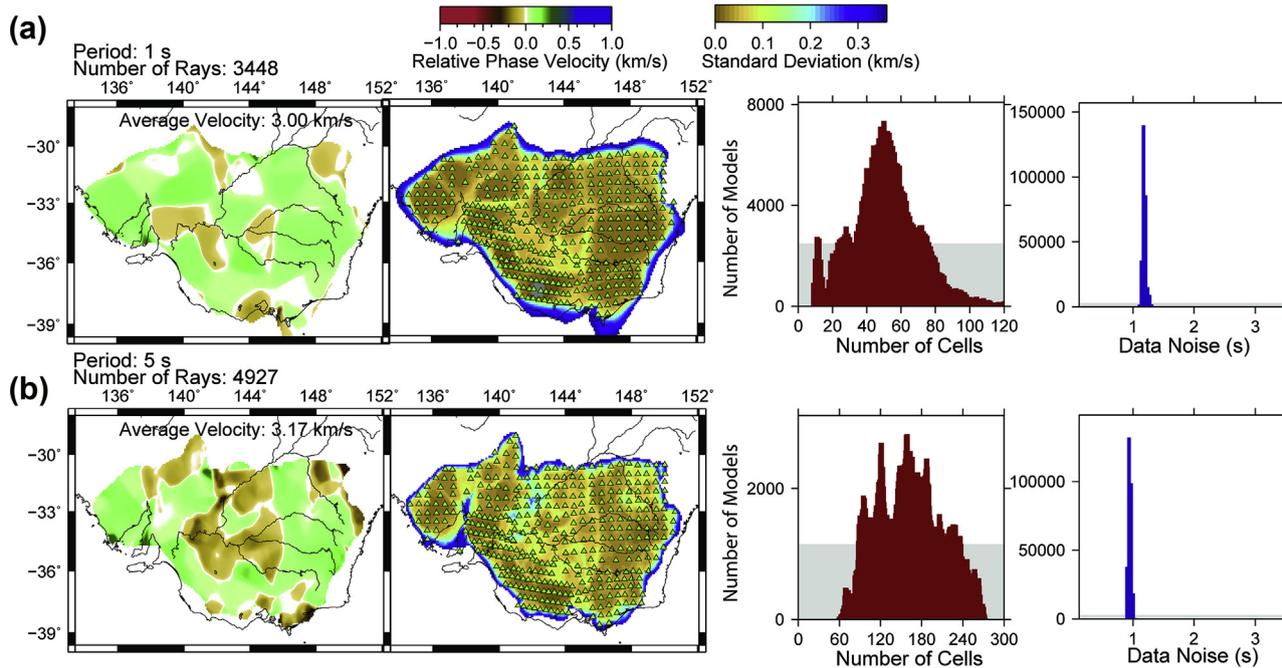


Figure 24 Average phase velocity maps for periods of (a) 1.0 s and (b) 5.0 s. For each phase velocity map, there is an accompanying map of the standard deviation, the posterior probability distribution histogram for the number of cells and the data noise. In the latter two cases, the prior distributions are indicated by shaded gray rectangles. From *Young, Cayley, et al. (2013)*. Copyright 2013 American Geophysical Union. Reproduced by permission of American Geophysical Union.

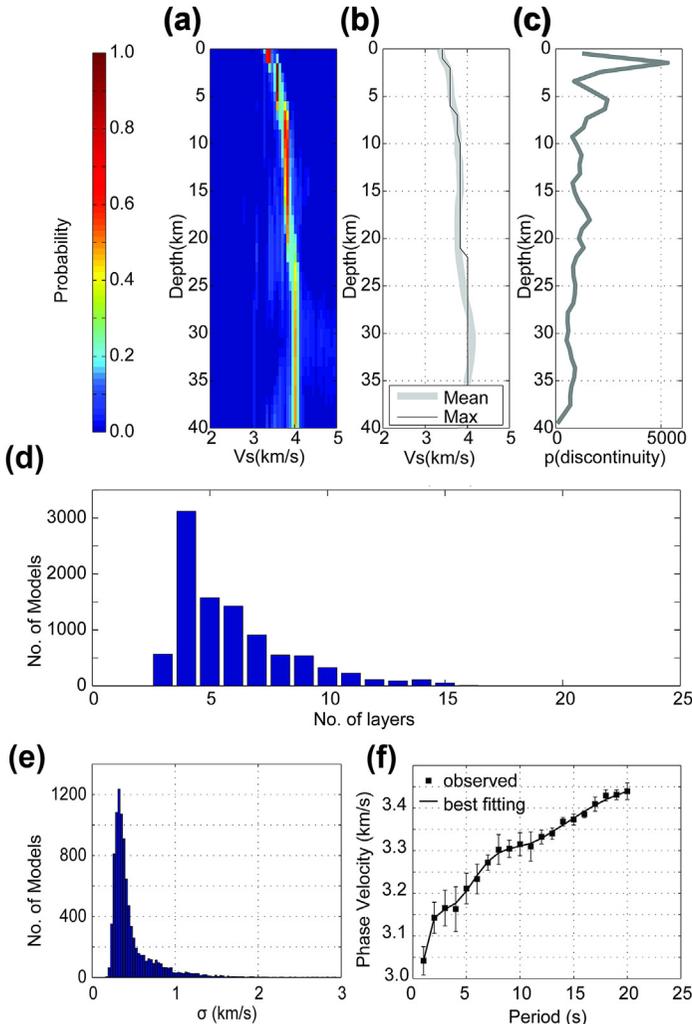


Figure 25 Inversion result for 1-D shear wave velocity at 30.64° S and 140.13° E. (a) Posterior probability distribution for crustal shear wave velocity; (b) mean and maximum likelihood of the shear wave velocity as a function of depth; (c) probability of a discontinuity; (d) posterior probability distribution on the number of layers in the model; (e) posterior probability distribution on the data noise; (f) phase dispersion for best fitting shear velocity model (black curve) superimposed on the actual phase velocity curve (shown by error bars). From *Young, Cayley, et al. (2013)*. Copyright 2013 American Geophysical Union. Reproduced by permission of American Geophysical Union.

posterior probability distribution to peak at a preferred number of cells, which is indicative of the level of detail present in the average model. In this case, the 5-s period map has more detail than the 1-s map. Although path coverage plays a role in this result, the level of noise in the data set is also a factor. In [Figure 24\(a\)](#), the posterior probability distribution for the data noise indicates a higher level of uncertainty than the corresponding plot in [Figure 24\(b\)](#). This is partly because the phase velocity at the 1-s period is difficult to pick, and is likely contaminated by unmodeled near-surface effects like scattering and the presence of topography. That said, the standard deviation of the 1-s map is lower than that of the 5-s map because it has a lower resolution (fewer cells) and hence the variance is improved.

The next step of the inversion process is to sample the phase velocity maps at regular intervals in latitude and longitude (approximately 20 km) to produce a set of 1372 dispersion curves. Standard deviations of the ensemble are also extracted and used as uncertainty inputs into the inversions for 1-D shear wave velocity structure. The Bayesian inversion method used for the 1-D inversion is essentially the same as before, except that the uncertainty of each phase velocity measurement is assumed to be relative, so a constant of proportionality is inverted for. Furthermore, since the problem is now 1-D, Voronoi cells are not required and the subsurface is instead defined by a variable number of constant velocity layers, where the number, position, and velocities of the layers are unknowns in the inversion. The forward method used for calculating phase velocity dispersion curves uses the method described by [Saito \(1988\)](#). In this case, each 1-D shear model takes about 30 h of CPU time to produce. [Figure 25](#) shows an example of the 1-D inversion for a particular grid node. In most cases, between three and seven layers are required to explain the data ([Figure 25\(d\)](#)). As well as yielding an estimate of uncertainty for the 1-D shear model ([Figure 25\(a\)](#)), other useful information such as the probability of a discontinuity can be extracted ([Figure 25\(c\)](#)).

A 3-D view of the average shear wave velocity model, obtained by joining together all 1-D shear wave models, is shown in [Figure 26](#). A variety of features have been inferred from this model, including crustal scale rifts, major faults, and the presence of Archean cratonic material in the lower crust ([Young, Cayley, et al., 2013](#)). Furthermore, the near-surface low-velocity region shown in red in the vertical slice corresponds very well with the known sediment thickness in the region. Horizontal slices through the average model and the associated standard deviation are shown for slices

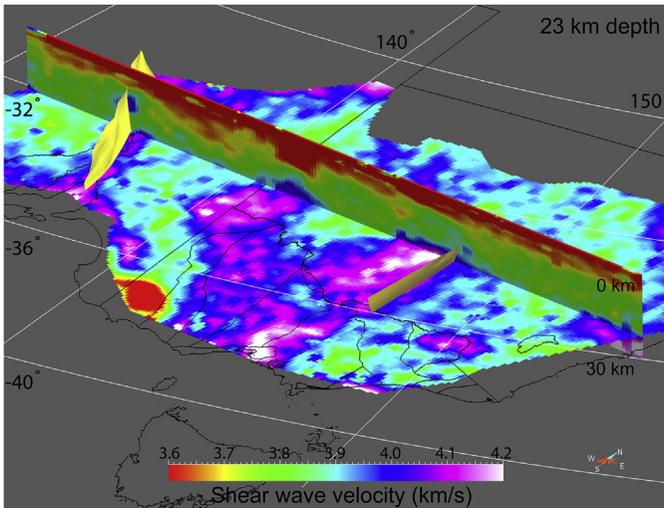


Figure 26 3-D perspective view of the final (average) 3-D shear wave velocity model, which is created by interpolating the regular grid of 1-D models. Yellow surfaces show the locations of the inferred Bootheragandra Fault (east) and the Torrens Hinge Zone (west). From *Young, Cayley, et al. (2013)* Copyright 2013 American Geophysical Union. Reproduced by permission of American Geophysical Union.

between 15 and 30 km depth in [Figure 27](#) (note that the color scale used in [Figure 27](#) is different from that used in [Figure 26](#)). The standard deviation tends to increase with depth due to a loss of structural sensitivity caused by the maximum period of the cross-correlated waveform being limited to 20 s (WOMBAT largely comprises three-component sensors with a natural period of 1 s). The high velocities seen at depths of 30 km may partly be caused by the sensitivity of the longer period dispersion measurements to mantle velocities; the jump from crust to mantle velocities at the Moho is not resolved by this data set. One of the main outcomes of this example is that the transdimensional scheme yields a vast amount of quantitative information that can be used to assess the robustness of any of the observed features that emerge from the inversion.

4.4 Full Waveform Inversion: Resolution Analysis Based on Second-Order Adjoints

During the past decade, steadily growing computational resources have led to an increased popularity of fully numerical methods for the solution of the seismic wave equation. Finite difference, spectral element, discontinuous Galerkin, and many other methods allow us to simulate 3-D seismic wave

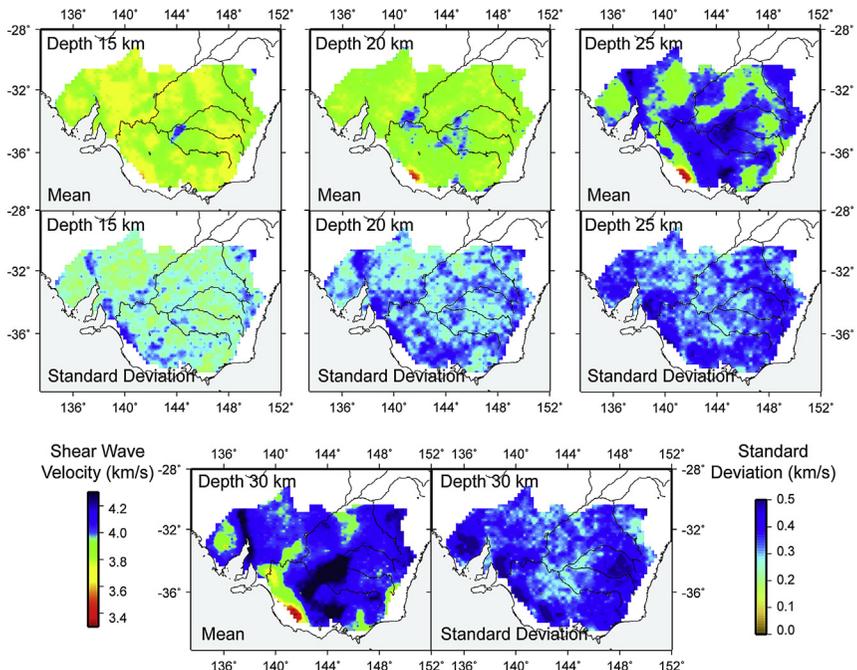


Figure 27 Horizontal sections between 15 and 30 km depth through the composite 3-D model, along with the associated standard deviations. *From Young, Cayley, et al. (2013). Copyright 2013 American Geophysical Union. Reproduced by permission of American Geophysical Union.*

propagation through complex Earth models at useful frequencies (e.g., Dumbser, Käser, & de la Puente, 2007; Komatitsch & Vilotte, 1998; Moczo et al., 2002). The ability to model the complete wave field with high accuracy sparked the development of full waveform inversion techniques that exploit complete seismograms for the benefit of improved tomographic resolution (e.g., Chen et al., 2007; Fichtner et al., 2009; Tape et al., 2009).

Full waveform inversion is mostly formulated as an optimization problem where the minimum of an objective functional $S(\mathbf{m})$, similar to the one in Eqn (1), is sought. The minimization proceeds iteratively with the help of gradient methods such as steepest descent or conjugate gradients. The gradient of the objective functional S with respect to the model parameters \mathbf{m} is computed efficiently using adjoint techniques (e.g., Tarantola, 1988; Tromp, Tape, & Liu, 2005; Fichtner et al., 2006a, 2006b).

Figure 28 presents a full waveform inversion model of Europe and Western Asia, taken from the work of Fichtner et al. (in press). While the

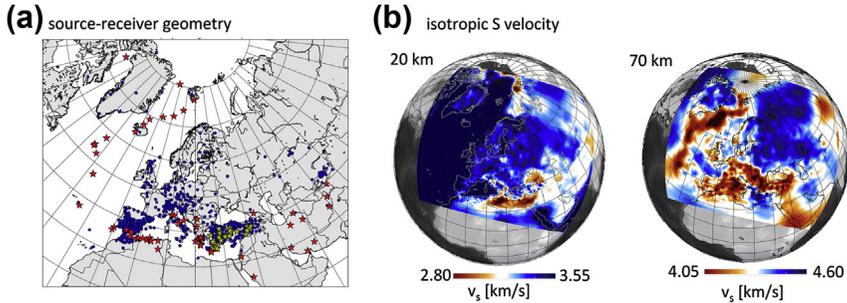


Figure 28 Full waveform inversion for upper mantle structure beneath Europe and Western Asia. (a) Source–receiver geometry. Earthquake epicenters are marked by stars and station locations by blue circles. Yellow stars indicate local events with magnitudes below $M_w = 4.5$. (b) Horizontal slices through the distribution of the isotropic S velocity at 20 and 70 km depth. (For interpretation of the references to color in this figure legend, the reader is referred to the online version of this book.) *Modified from Fichtner et al. (in press). Copyright 2013 Royal Astronomical Society. Reproduced by permission of Oxford University Press.*

exploitation of complete seismograms should intuitively improve the quality of the tomographic model, the quantitative assessment of resolution in full waveform inversion is a major challenge for the following reasons: (1) The computation of covariance and resolution is not possible because the full sensitivity matrix \mathbf{G} needed for their calculation is too expensive to compute explicitly. (2) Synthetic reconstruction tests, using, for instance, checkerboard input patterns, are also computationally expensive because each of these tests requires the same resources as the real data inversion itself. (3) Also for reasons of computational costs, the Monte Carlo sampling of S is not possible unless the dimension of the model space is drastically reduced, using, for instance, tectonic regionalization (Käuff, Fichtner, & Igel, *in press*).

While none of the classical resolution analysis tools seems directly applicable to full waveform inversion, the following variant of a spike test is computationally feasible: We perturb the optimal model \mathbf{m}_{opt} , e.g., the model shown in Figure 28, with a (nearly) point-localized heterogeneity $\delta\mathbf{m}_{\text{in}}$ and use $\mathbf{m}_{\text{opt}} + \delta\mathbf{m}_{\text{in}}$ as input for a synthetic recovery test. Instead of performing a complete iterative full waveform inversion, we use the fact that the output model $\delta\mathbf{m}_{\text{out}}$ is equal to the Hessian matrix \mathbf{H} applied to $\delta\mathbf{m}_{\text{in}}$, that is, $\delta\mathbf{m}_{\text{out}} = \mathbf{H}\delta\mathbf{m}_{\text{in}}$ (Fichtner & Trampert, 2011a). The Hessian \mathbf{H} contains the second derivatives of the objective functional S with respect to the model parameters \mathbf{m} . The matrix–vector product

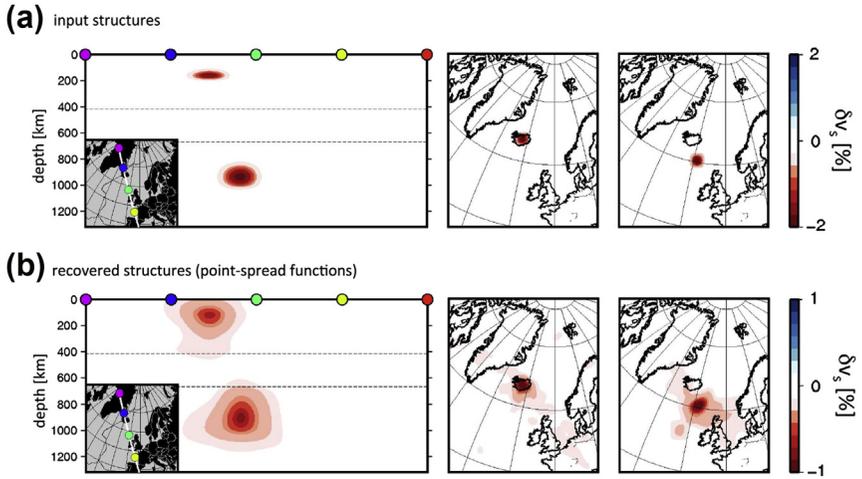


Figure 29 Point-spread functions used to assess resolution in the North Atlantic region. (a) Localized input S velocity heterogeneities $\delta\mathbf{m}_{\text{in}}$ with a peak amplitude of 2%. (b) The output structures $\delta\mathbf{m}_{\text{out}}$ are computed by applying the Hessian \mathbf{H} to $\delta\mathbf{m}_{\text{in}}$. For (nearly) point-localized input structures, $\mathbf{H}\delta\mathbf{m}_{\text{in}}$ is referred to as the point-spread function. Modified from *Fichtner et al. (in press)*. Copyright 2013 Royal Astronomical Society. Reproduced by permission of Oxford University Press.

$\mathbf{H}\delta\mathbf{m}_{\text{in}}$ can be computed efficiently with the help of so-called second-order adjoints without requiring \mathbf{H} to be determined explicitly (Fichtner & Trampert, 2011a; Santosa & Symes, 1988).

Figure 29 shows a synthetic recovery test for the North Atlantic part of the model from Figure 28. The input structures $\delta\mathbf{m}_{\text{in}}$ are spheroidal S velocity heterogeneities centered at positions in the model where information on resolution is particularly critical for interpretation, i.e., beneath the Iceland hot spot (Rickers, Fichtner, & Trampert, in press). The recovered structures $\delta\mathbf{m}_{\text{out}}$, referred to as point-spread functions, are smeared versions of $\delta\mathbf{m}_{\text{in}}$. As in the classical spike test, the point-spread function contains information on spatial resolution and amplitude recovery. Specifically, the width of the point-spread function determines the resolution length at the position where the input spike is located.

For a comprehensive assessment of resolution, we would ideally compute point-spread functions for every position inside the model volume, a formidable task in a 3-D problem with potentially millions of model parameters. To lighten the computational burden, we can take advantage of the approximate Gaussian shape of the point-spread function, as seen in Figure 29. Parameterizing the point-spread function in terms of a position-dependent

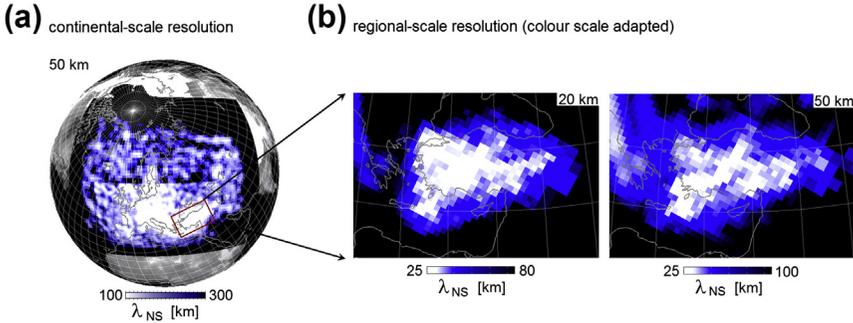


Figure 30 Resolution length in north–south direction, λ_{NS} , at 50 km depth. (a) Continental-scale view. (b) Zoom into the Anatolian region where resolution is particularly high thanks to the incorporation of regional earthquakes. The color scale is shifted toward lower values for better visibility. (For interpretation of the references to color in this figure legend, the reader is referred to the online version of this book.) Modified from *Fichtner et al. (in press)*. Copyright 2013 Royal Astronomical Society. Reproduced by permission of Oxford University Press.

Gaussian reduces the problem to the determination of the parameters of the Gaussian that control its amplitude and width. These parameters can be computed efficiently from application of the Hessian \mathbf{H} to a set of seven linearly independent model perturbations $\delta\mathbf{m}$ (Fichtner & Trampert, 2011c). Once the parameters of the Gaussian approximation to the point-spread function are found, we can extract various position-dependent resolution proxies, including the direction-dependent resolution length.

Resolution length in north–south direction, λ_{NS} , for the tomographic model from Figure 28 is shown in Figure 30. Throughout most of central Europe, λ_{NS} is below 100 km, meaning that structures with length scales of more than 100 km in the north–south direction are resolved. Resolution is particularly good in the Anatolian region where regional data place strong constraints on smaller scale structures (Fichtner et al., in press).

5. CONCLUDING REMARKS

Comprehensive assessment of solution robustness is the most challenging task in modern seismic tomography. In this era of “Big data” where it is not unusual to tackle massive inverse problems involving millions of data and a similar number of unknowns (e.g., Burdick et al., 2014), robust quantitative estimates of model uncertainty are often beyond the capability of available computing power. Those methods that can be applied to large

problems, such as synthetic reconstruction tests and estimates of covariance and resolution from linear theory, suffer from a range of drawbacks that limit their usefulness. More robust methods are available, but for the most part they have been developed with much smaller problems in mind. Stochastic nonlinear sampling techniques, such as the family of Bayesian transdimensional tomography schemes that are based on the rj-McMC method, appear to show a lot of promise in terms of producing models that maximize the extraction of reliable information from the data and quantifying uncertainty. Although most applications to date have been to relatively modest 2-D problems, extension to 3-D is feasible, as has been demonstrated by [Piana Agostinetti et al. \(submitted for publication\)](#) for the problem of local earthquake tomography. Recent advances in inversion methodology also promise to extend transdimensional inversion to larger problems. For example, [Sambridge \(2014\)](#) develops a parallel tempering algorithm for probabilistic sampling and multimodal optimization, and demonstrates that the convergence of the McMC process is significantly accelerated by its inclusion. In one example involving receiver functions ([Sambridge, 2014](#)), it is shown that convergence of the Markov chain is improved by an order of magnitude.

It is clear that there is a real need for further development of robust methods for assessing model uncertainty in seismic tomography, as the current generation is in many cases not up to the task. With the advent of full waveform tomography, which is much more computationally demanding than traditional approaches such as travel time tomography, and arguably more nonlinear, the need is even greater. Most methods for quantifying uncertainty are tied to the approach used to solve the inverse problem; thus, formal estimates of covariance and resolution from linear theory are normally calculated for inverse problems posed within a linear or iterative-nonlinear framework. In full waveform tomography, current methods rely on local linearization due to the huge computing costs associated with the forward solver. As such, it is reasonable to use a consistent approach from linear theory to assess uncertainty ([Fichtner & Trampert, 2011b](#)). However, in regions of significant heterogeneity, lack of prior knowledge, and the availability of high frequencies, the assumption of local linearity can become hard to justify.

ACKNOWLEDGMENTS

Aspects of this work were supported by ARC Discovery Projects DP1101020985 and DP120103673. Most figures were created using the freely distributed software packages

GMT (Wessel, Smith, Scharroo, Luis, & Wobbe, 2013) and xfig. Two anonymous reviewers are thanked for their constructive comments on the original version of this manuscript.

REFERENCES

- Abers, G. A., & Roecker, S. W. (1991). Deep structure of an arc-continent collision; earthquake relocation and inversion for upper mantle P and S wave velocities beneath Papua New Guinea. *Journal of Geophysical Research*, *96*, 6379–6401.
- Achauer, U. (1994). New ideas on the Kenya rift based on the inversion of the combined dataset of the 1985 and 1989/90 seismic tomography experiments. *Tectonophysics*, *236*, 305–329.
- Afonso, J. C., Fullea, J., Griffin, W. L., Yang, Y., Jones, A. G., Connolly, J. A. D., et al. (2013). 3D multi-observable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. I: a priori petrological information and geophysical observables. *Journal of Geophysical Research*, *118*(5), 2586–2617. <http://dx.doi.org/10.1002/jgrb.50124>.
- Afonso, J. C., Fullea, J., Yang, Y., Connolly, J. A. D., & Jones, A. G. (2013). 3D multi-observable probabilistic inversion for the compositional and thermal structure of the lithosphere and upper mantle. II: general methodology and resolution analysis. *Journal of Geophysical Research*, *118*(4), 1650–1676. <http://dx.doi.org/10.1002/jgrb.50123>.
- Aki, K., Christofferson, A., & Husebye, E. S. (1977). Determination of the three-dimensional seismic structure of the lithosphere. *Journal of Geophysical Research*, *82*, 277–296.
- Aki, K., & Lee, W. H. K. (1976). Determination of the three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes 1. A homogeneous initial model. *Journal of Geophysical Research*, *81*, 4381–4399.
- Allen, R. (1982). Automatic phase pickers: their present use and future prospects. *Bulletin of Seismological Society of America*, *72*, 225–242.
- Aloisi, M., Cocina, O., Neri, G., Orecchio, B., & Privitera, E. (2002). Seismic tomography of the crust underneath the Etna volcano, Sicily. *Physics of the Earth and Planetary Interiors*, *134*, 139–155.
- An, M. (2012). A simple method for determining the spatial resolution of a general inverse problem. *Geophysical Journal International*, *191*, 849–864.
- Asad, A. M., Pullammanappallil, S. K., Anoooshehpoo, A., & Louie, J. N. (1999). Inversion of travel-time data for earthquake locations and three-dimensional velocity structure in the Eureka Valley area, eastern California. *Bulletin of Seismological Society of America*, *89*, 796–810.
- Backus, G. E., & Gilbert, J. F. (1967). Numerical applications of a formalism for geophysical inverse problems. *Geophysical Journal of the Royal Astronomical Society*, *13*, 247–276.
- Backus, G. E., & Gilbert, J. F. (1968). The resolving power of gross earth data. *Geophysical Journal of the Royal Astronomical Society*, *16*, 169–205.
- Backus, G., & Gilbert, F. (1969). Constructing P-velocity models to fit restricted sets of travel-time data. *Bulletin of Seismological Society of America*, *59*, 1407–1414.
- Backus, G., & Gilbert, F. (1970). Uniqueness in the inversion of inaccurate gross Earth data. *Philosophical Transactions of the Royal Society A*, *266*, 123–192.
- Bijwaard, H., & Spakman, W. (2000). Non-linear global P-wave tomography by iterated linearized inversion. *Geophysical Journal International*, *141*, 71–82.
- Bijwaard, H., Spakman, W., & Engdahl, E. R. (1998). Closing the gap between regional and global travel time tomography. *Journal of Geophysical Research*, *103*, 30,055–30,078.
- Bleibinhaus, F., Hole, J. A., & Ryberg, T. (2007). Structure of the California Coast Ranges and San Andreas Fault at SAFOD from seismic waveform inversion and reflection imaging. *Journal of Geophysical Research*, *112*. <http://dx.doi.org/10.1029/2006JB004611>.

- Blundell, C. A., (1993). Resolution analysis of seismic p-wave velocity estimates using reflection tomographic inversion. (Ph.D. thesis), Monash University.
- Bodin, T., & Sambridge, M. (2009). Seismic tomography with the reversible jump algorithm. *Geophysical Journal International*, *178*, 1411–1436.
- Bodin, T., Sambridge, M., Rawlinson, N., & Arroucau, P. (2012). Transdimensional tomography with unknown data noise. *Geophysical Journal International*, *189*, 1536–1556.
- Bodin, T., Sambridge, M., Tkalcic, H., Arroucau, P., Gallagher, K., & Rawlinson, N. (2012). Transdimensional inversion of receiver functions and surface wave dispersion. *Journal of Geophysical Research*, *117*, B02301. <http://dx.doi.org/10.1029/2011JB008560>.
- Bois, P., La Porte, M., Lavergne, M., & Thomas, G. (1971). Essai de détermination automatique des vitesses sismiques par mesures entre puits. *Geophysical Prospecting*, *19*, 42–83.
- Boschetti, F., Dentith, M. K., & List, R. D. (1996). Inversion of seismic refraction data using genetic algorithms. *Geophysics*, *61*, 1715–1727.
- Bozdağ, E., Trampert, J., & Tromp, J. (2011). Misfit functions for full waveform inversion based on instantaneous phase and envelope measurements. *Geophysical Journal International*, *185*, 845–870.
- Buland, R. (1986). Uniform reduction error analysis. *Bulletin of Seismological Society of America*, *76*, 217–230.
- Burdick, S., Li, C., Martynov, V., Cox, T., Eakins, J., Astiz, L., et al. (2008). Upper mantle heterogeneity beneath North America from travel time tomography with global and USArray transportable array data. *Seismological Research Letters*, *79*, 384–392.
- Burdick, S., Van der Hilst, R. D., Vernon, F. L., Martynov, V., Cox, T., Eakins, J., et al. (2014). Model Update January 2013: upper mantle heterogeneity beneath North America from travel time tomography with global and USArray transportable array data. *Seismological Research Letters*, *85*, 77–81.
- Campillo, M., & Paul, A. (2003). Long-range correlations in the diffuse seismic coda. *Science*, *299*, 547–549.
- Cassell, B. R. (1982). A method for calculating synthetic seismograms in laterally varying media. *Geophysical Journal of the Royal Astronomical Society*, *69*, 339–354.
- Chen, P., Zhao, L., & Jordan, T. H. (2007). Full 3D tomography for the crustal structure of the Los Angeles region. *Bulletin of Seismological Society of America*, *97*, 1094–1120.
- Chevrot, S. (2002). Optimal measurements of relative and absolute delay times by simulated annealing. *Geophysical Journal International*, *151*, 164–171.
- Chevrot, S., & Zhao, L. (2007). Multiscale finite-frequency Rayleigh wave tomography of the Kaapvaal craton. *Geophysical Journal International*, *169*, 201–215.
- Chiao, L.-Y., & Kuo, B.-Y. (2001). Multiscale seismic tomography. *Geophysical Journal International*, *145*, 517–527.
- Chou, C. W., & Booker, J. R. (1979). A Backus-Gilbert approach to inversion of travel time data for three-dimensional velocity structure. *Geophysical Journal of the Royal Astronomical Society*, *59*, 325–344.
- Claerbout, J. F., & Muir, F. (1973). Robust modelling with erratic data. *Geophysics*, *38*, 826–844.
- Curtis, A., & Snieder, R. (1997). Reconditioning inverse problems using the genetic algorithm and revised parameterization. *Geophysics*, *62*, 1524–1532.
- Day, A. J., Peirce, C., & Sinha, M. C. (2001). Three-dimensional crustal structure and magma chamber geometry at the intermediate-spreading, back-arc Valu Fa Ridge, Lau Basin – results of a wide-angle seismic tomographic inversion. *Geophysical Journal International*, *146*, 31–52.
- de Wit, R. W. L., Trampert, J., & van der Hilst, R. D. (2012). Toward quantifying uncertainty in travel time tomography using the null-space shuttle. *Journal of Geophysical Research*, *117*, B03301.

- Deal, M. M., & Nolet, G. (1996). Null-space shuttles. *Geophysical Journal International*, *124*, 372–380.
- Deal, M. M., Nolet, G., & van der Hilst, R. D. (1999). Slab temperature and thickness from seismic tomography 1. method and application to tonga. *Journal of Geophysical Research*, *104*, 28,789–28,802.
- Di Stefano, R., Aldersons, F., Kissling, E., Baccheschi, P., Chiarabba, C., & Giardini, D. (2006). Automatic seismic phase picking and consistent observation error assessment: application to the italian seismicity. *Geophysical Journal International*, *165*, 121–134.
- Djikpesse, H. A., & Tarantola, A. (1999). Multi-parameter L1 norm waveform fitting: Interpretation of Gulf of Mexico reflection seismograms. *Geophysics*, *64*, 1023–1035.
- Dobbs, R., Manyika, J., Roxburgh, C., & Lund, S. (2011). *Big data: The next frontier for innovation, competition and productivity*. Tech. Rep. McKinsey Global Institute
- Dumbser, M., Käser, M., & de la Puente, J. (2007). Arbitrary high order finite volume schemes for seismic wave propagation on unstructured meshes in 2D and 3D. *Geophysical Journal International*, *171*, 665–694.
- Duputel, Z., Agram, P. S., Simons, M., Minson, S. E., & Beck, J. L. (2014). Accounting for prediction uncertainty when inferring subsurface fault slip. *Geophysical Journal International*, *197*, 464–482.
- Dziewonski, A. M., Hager, B. H., & O’Connell, R. J. (1977). Large-scale heterogeneities in the lower mantle. *Journal of Geophysical Research*, *82*, 239–255.
- Dziewonski, A. M., & Woodhouse, J. H. (1987). Global images of the earth’s interior. *Science*, *236*, 37–48.
- Eberhart-Phillips, D. (1986). Three-dimensional velocity structure in northern California coast ranges from inversion of local earthquake arrival times. *Bulletin of Seismological Society of America*, *76*, 1025–1052.
- Eberhart-Phillips, D., & Reyners, M. (1997). Continental subduction and three-dimensional crustal structure: the northern South Island, New Zealand. *Journal of Geophysical Research*, *102*, 11,848–11,861.
- Efron, B., & Tibshirani. (1993). *An introduction to the bootstrap*. New York: Chapman and Hall/CRC.
- Farra, V., & Madariaga, R. (1988). Non-linear reflection tomography. *Geophysical Journal*, *95*, 135–147.
- Fichtner, A. (2011). *Full seismic waveform modelling and inversion*. Heidelberg: Springer-Verlag.
- Fichtner, A., Bunge, H.-P., & Igel, H. (2006a). The adjoint method in seismology – I. Theory. *Physics of the Earth and Planetary Interiors*, *157*, 86–104.
- Fichtner, A., Bunge, H.-P., & Igel, H. (2006b). The adjoint method in seismology – II. Applications: traveltimes and sensitivity functionals. *Physics of the Earth and Planetary Interiors*, *157*, 105–123.
- Fichtner, A., Kennett, B. L. N., Igel, H., & Bunge, H.-P. (2009). Full waveform tomography for upper-mantle structure in the australasian region using adjoint methods. *Geophysical Journal International*, *179*, 1703–1725.
- Fichtner, A., & Trampert, J. (2011a). Hessian kernels of seismic data functionals based upon adjoint techniques. *Geophysical Journal International*, *185*, 775–798.
- Fichtner, A., & Trampert, J. (2011b). Resolution analysis in full waveform inversion. *Geophysical Journal International*, *187*, 1604–1624.
- Fichtner, A., & Trampert, J. (2011c). Resolution analysis in full waveform inversion. *Geophysical Journal International*, *187*, 1604–1624.
- Fichtner, A., Trampert, J., Cupillard, P., Saygin, E., Taymaz, T., Capdeville, Y., et al. (2013). Multi-scale full waveform inversion. *Geophysical Journal International*, *194*(21), 534–556. <http://dx.doi.org/10.1093/gji/ggt118>.

- Fishwick, S., Kennett, B. L. N., & Reading, A. M. (2005). Contrasts in lithospheric structure within the Australian craton – insights from surface wave tomography. *Earth and Planetary Science Letters*, *231*, 163–176.
- Fishwick, S., & Rawlinson, N. (2012). 3-D structure of the Australian lithosphere from evolving seismic datasets. *Australian Journal of Earth Sciences*, *59*, 809–826.
- Foulger, G., Panza, G. F., Artemieva, I. M., Bastow, I. D., Cammarano, F., Evans, J. R., et al. (2013). Caveats on tomographic images. *Terra Nova*, *25*, 259–281.
- Fukao, Y., Obayashi, M., Inoue, H., & Nebai, M. (1992). Subducting slabs stagnant in the mantle. *Journal of Geophysical Research*, *97*, 4809–4822.
- Galetti, E., Curtis, A., Baptie, B., & Meles, G. (2014). Uncertainty loops in seismic tomography. *Physical Review*, submitted for publication, *194*(21), 534–556. <http://dx.doi.org/doi:10.1093/gji/ggt118>.
- Gantz, J. F., & Reinsel, D. (2010). *The digital universe decade – Are you ready?* IDC white paper May 2010. Framingham, USA: Published by IDC.
- Glahn, A., & Granet, M. (1993). Southern Rhine Graben: small-wavelength tomographic study and implications for the dynamic evolution of the graben. *Geophysical Journal International*, *113*, 399–418.
- Gorbatov, A., Widiyantoro, S., Fukao, Y., & Gordeev, E. (2000). Signature of remnant slabs in the North Pacific from *P*-wave tomography. *Geophysical Journal International*, *142*, 27–36.
- Graeber, F. M., & Asch, G. (1999). Three-dimensional models of *P* wave velocity and *P*-to-*S* velocity ratio in the southern central Andes by simultaneous inversion of local earthquake data. *Journal of Geophysical Research*, *104*, 20,237–20,256.
- Graeber, F. M., Houseman, G. A., & Greenhalgh, S. A. (2002). Regional teleseismic tomography of the western Lachlan Orogen and the Newer Volcanic Province, southeast Australia. *Geophysical Journal International*, *149*, 249–266.
- Granet, M., & Trampert, J. (1989). Large-scale *P*-velocity structures in the Euro-Mediterranean area. *Geophysical Journal International*, *99*, 583–594.
- Gung, Y., & Romanowicz, B. (2004). *Q* tomography of the upper mantle using three-component long-period waveforms. *Geophysical Journal International*, *157*, 813–830.
- Hadamard, J. (1902). *Sur les problèmes aux dérivées partielles et leur signification physique*. Princeton University Bulletin.
- Halko, N., Martinsson, P.-G., & Tropp, J. A. (2011). Finding structure with randomness: probabilistic algorithms for constructing approximate matrix decompositions. *SIAM Rev*, *53*, 217–288.
- Hammer, P. T. C., Dorman, L. M., Hildebrand, J. A., & Cornuelle, B. D. (1994). Jasper Seamount structure: Seafloor seismic refraction tomography. *Journal of Geophysical Research*, *99*, 6731–6752.
- Hammersley, J. M., & Handscomb, D. C. (1964). *Monte Carlo methods*. New York: Chapman & Hall.
- Hestenes, M., & Stiefel, E. (1952). Methods of conjugate gradients for solving linear systems. *Journal of Research of the National Bureau of Standards*, *49*, 409–436.
- Hildebrand, J. A., Dorman, L. M., Hammer, P. T. C., Schreiner, A. E., & Cornuelle, B. D. (1989). Seismic tomography of Jasper Seamount. *Geophysical Research Letters*, *16*, 1355–1358.
- Hole, J. A. (1992). Nonlinear high-resolution three-dimensional travel-time tomography. *Journal of Geophysical Research*, *97*, 6553–6562.
- Humphreys, E. D., & Clayton, R. W. (1990). Tomographic image of the Southern California Mantle. *Journal of Geophysical Research*, *95*, 19,725–19,746.
- Iyer, H., & Hirahara, K. (1993). *Seismic tomography: Theory and practice*. London: Chapman & Hall.

- Jeffreys, H. (1932). An alternative to the rejection of observations. *Proceedings of the Royal Society of London Series A*, 137, 78–87.
- Jeong, W., Pyun, S., Son, W., & Min, D.-J. (2013). A numerical study of simultaneous-source full waveform inversion with l_1 -norm. *Geophysical Journal International*, 194, 1727–1737.
- Julian, B. R., & Gubbins, D. (1977). Three-dimensional seismic ray tracing. *Journal of Geophysics*, 43, 95–113.
- Kamei, R., Pratt, R. G., & Tsuji, T. (2013). On acoustic waveform tomography of wide-angle OBS data – strategies for pre-conditioning and inversion. *Geophysical Journal International*, 194, 1250–1280.
- Käufel, P., Fichtner, A., & Igel, H. (2013). *Probabilistic full waveform inversion based on tectonic regionalisation – development and application to the Australian upper mantle*. *Geophysical Journal International*, 193(1), 437–451. <http://dx.doi.org/10.1093/gji/ggs131>.
- Keilis-Borok, V. I., & Yanovskaya, T. B. (1967). Inverse problems of seismology. *Geophysical Journal International*, 13, 223–234.
- Kennett, B. L. N., Sambridge, M. S., & Williamson, P. R. (1988). Subspace methods for large scale inverse problems involving multiple parameter classes. *Geophysical Journal International*, 94, 237–247.
- Khan, A., Boschi, L., & Connolly, J. A. D. (2011). Mapping the Earth's thermochemical and anisotropic structure using global surface wave data. *Journal of Geophysical Research*, 116. <http://dx.doi.org/10.1029/2010JB007828>.
- Koketsu, K., & Sekine, S. (1998). Pseudo-bending method for three-dimensional seismic ray tracing in a spherical earth with discontinuities. *Geophysical Journal International*, 132, 339–346.
- Komatitsch, D., & Vilotte, J. P. (1998). The spectral element method: an effective tool to simulate the seismic response of 2D and 3D geological structures. *Bulletin of Seismological Society of America*, 88, 368–392.
- Lees, J. M., & Crosson, R. S. (1989). Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data. *Journal of Geophysical Research*, 94, 5716–5728.
- Lees, J. M., & Crosson, R. S. (1990). Tomographic imaging of local earthquake delay times for three-dimensional velocity variation in western Washington. *Journal of Geophysical Research*, 95, 4763–4776.
- Leonard, M. (2000). Comparison of manual and automatic onset time picking. *Bulletin of Seismological Society of America*, 90, 1384–1390.
- Lévêque, J. J., Rivera, L., & Wittlinger, G. (1993). On the use of the checker-board test to assess the resolution of tomographic inversions. *Geophysical Journal International*, 115, 313–318.
- Li, Y., & Oldenburg, D. W. (1996). 3-D inversion of magnetic data. *Geophysics*, 61, 394–408.
- Li, Y., & Oldenburg, D. W. (1998). 3-D inversion of gravity data. *Geophysics*, 63, 109–119.
- Liu, Q., & Gu, Y. J. (2012). Seismic imaging: from classical to adjoint tomography. *Tectonophysics*, 566, 31–66.
- Lohr, S. (February 11, 2012). The age of big data. *The New York Times*.
- Loris, I., Nolet, G., Daubechies, I., & Dahlen, F. (2007). Tomographic inversion using l_1 -norm regularization of wavelet coefficients. *Geophysical Journal International*, 170, 359–370.
- MacCarthy, J. K., Brochers, B., & Aster, R. C. (2011). Efficient stochastic estimation of the model resolution matrix diagonal and generalised cross-validation for large geophysical inverse problems. *Journal of Geophysical Research*, 116. <http://dx.doi.org/10.1029/2011JB008234>.
- Malinverno, A., & Briggs, V. (2004). Expanded uncertainty quantification in inverse problems: hierarchical Bayes and empirical Bayes. *Geophysics*, 69, 1005–1016.

- Malinverno, A., & Parker, R. (2006). Two ways to quantify uncertainty in geophysical inverse problems. *Geophysics*, *71*, W15–W27.
- Marquering, H., Dahlen, F. A., & Nolet, G. (1999). Three-dimensional sensitivity kernels for finite-frequency travel times: the banana-doughnut paradox. *Geophysical Journal International*, *137*, 805–815.
- Marx, V. (2013). The big challenges of big data. *Nature*, *498*, 255–260.
- McCaughey, M., & Singh, S. C. (1997). Simultaneous velocity and interface tomography of normal-incidence and wide-aperture seismic traveltime data. *Geophysical Journal International*, *131*, 87–99.
- Meier, U., Curtis, A., & Trampert, J. (2007a). Fully nonlinear inversion of fundamental mode surface waves for a global crustal model. *Geophysical Research Letters*, *34*, L16304.
- Meier, U., Curtis, A., & Trampert, J. (2007b). Global crustal thickness from neural network inversion of surface wave data. *Geophysical Journal International*, *169*, 706–722.
- Meju, M. A. (2009). Regularized extremal bounds analysis (REBA): an approach to quantifying uncertainty in nonlinear geophysical inverse problems. *Geophysical Research Letters*, *36*, L03304.
- Menke, W. (1989). *Geophysical data analysis: Discrete inverse theory*. New York: Academic Press.
- Michelini, A. (1995). An adaptive-grid formalism for traveltime tomography. *Geophysical Journal International*, *121*, 489–510.
- Moczko, P., Kristek, J., Vavrycuk, V., Archuleta, R., & Halada, L. (2002). 3D heterogeneous staggered-grid finite-difference modeling of seismic motion with volume harmonic and arithmetic averaging of elastic moduli. *Bulletin of the Seismological Society of America*, *92*, 3042–3066.
- Montagner, J.-P., & Nataf, H.-C. (1986). A simple method for inverting the azimuthal anisotropy of surface waves. *Journal of Geophysical Research*, *91*, 511–520.
- Montelli, R., Nolet, G., Dahlen, F. A., & Masters, G. (2006). A catalogue of deep mantle plumes: new results from finite-frequency tomography. *Geochemistry, Geophysics, Geosystems*, *7*. <http://dx.doi.org/10.1029/2006GC001248>.
- Montelli, R., Nolet, G., Dahlen, F. A., Masters, G., Engdahl, E. R., & Hung, S.-H. (2004). Finite-frequency tomography reveals a variety of plumes in the mantle. *Science*, *303*, 338–343.
- Mosegaard, K., & Sambridge, M. (2002). Monte Carlo analysis of inverse problems. *Inverse Problems*, *18*, R29–R54.
- Nakanishi, I. (1985). Three-dimensional structure beneath the Hokkaido-Tohoku region as derived from a tomographic inversion of *P*-arrival times. *Journal of Physics of the Earth*, *33*, 241–256.
- Nielsen, L., Thybo, H., & Solodilov, L. (1999). Seismic tomographic inversion of Russian PNE data along profile Kraton. *Geophysical Research Letters*, *26*, 3413–3416.
- Nolet, G. (1985). Solving or resolving inadequate and noisy tomographic systems. *Journal of Computational Physics*, *61*, 463–482.
- Nolet, G. (1987). *Seismic tomography: With applications in global seismology and exploration geophysics*. Dordrecht: D. Reidel.
- Nolet, G. (2008). *A Breviary of Seismic Tomography: Imaging the interior of the Earth and Sun*. Cambridge: Cambridge University Press.
- Nolet, G., & Montelli, R. (2005). Optimal parameterization of tomographic models. *Geophysical Journal International*, *161*, 1–8.
- Nolet, G., Montelli, R., & Virieux, J. (1999). Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems. *Geophysical Journal International*, *138*, 36–44.
- Obrebski, M., Allen, R. M., Pollitz, F., & Hung, S.-H. (2011). Lithosphere-asthenosphere interaction beneath the western United States from the joint inversion of body-wave

- traveltimes and surface-wave phase velocities. *Geophysical Journal International*, *185*, 1003–1021.
- Oncescu, M. C., Burlacu, V., Anghel, M., & Smalbergher, V. (1984). Three-dimensional P-wave velocity image under the Carpathian Arc. *Tectonophysics*, *106*, 305–319.
- Operto, S., Virieux, J., Dessa, J. X., & Pascal, G. (2006). Crustal seismic imaging from multifold ocean bottom seismometer data by frequency domain full waveform tomography: application to the eastern nankai trough. *Journal of Geophysical Research*, *111*, B09306.
- Paige, C. C., & Saunders, M. A. (1982). LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Transactions on Mathematical Software*, *8*, 43–71.
- Pereyra, V., Lee, W. H. K., & Keller, H. B. (1980). Solving two-point seismic-ray tracing problems in a heterogeneous medium. *Bulletin of Seismological Society of America*, *70*, 79–99.
- Piana Agostinetti, N., Giacomuzzi, G., Malinverno, A. Local 3D earthquake tomography by trans-dimensional Monte Carlo sampling. *Geophysical Journal International*, submitted for publication.
- Pilia, S., Rawlinson, N., Direen, N. G., Cummins, P. R., & Balfour, N. (2013). Structural controls on localized intraplate deformation and seismicity in Southern Australia: Insights from local earthquake tomography of the Flinders Ranges. *Journal of Geophysical Research*, *118*. <http://dx.doi.org/10.1002/jgrb.50168>.
- Pratt, R. G., & Shipp, R. M. (1999). Seismic waveform inversion in the frequency domain, Part 2: fault delineation in sediments using crosshole data. *Geophysics*, *64*, 902–914.
- Press, F. (1968). Earth models obtained by monte carlo inversion. *Journal of Geophysical Research*, *73*, 5223–5234.
- Pullammanappallil, S. K., & Louie, J. N. (1993). Inversion of seismic reflection traveltimes using a nonlinear optimization scheme. *Geophysics*, *58*, 1607–1620.
- Pulliam, R. J., Vasco, D. W., & Johnson, L. R. (1993). Tomographic inversions for mantle P-wave velocity structure based on the minimization of L2 and L1 norms of International Seismological Centre travel time Residuals. *Journal of Geophysical Research*, *98*, 699–734.
- Ramachandran, K., Hyndman, R. D., & Brocher, T. M. (2006). Regional P wave velocity structure of the Northern Cascadia Subduction Zone. *Journal of Geophysical Research*, *111*, B12301.
- Ravaut, C., Operto, S., Improta, L., Virieux, J., Herrero, A., & Dell'Aversana, P. (2004). Multiscale imaging of complex structures from multifold wide-aperture seismic data by frequency-domain full-waveform tomography: application to a thrust belt. *Geophysical Journal International*, *159*, 1032–1056.
- Rawlinson, N., & Fishwick, S. (2012). Seismic structure of the southeast Australian lithosphere from surface and body wave tomography. *Tectonophysics*, *572*, 111–122.
- Rawlinson, N., Houseman, G. A., & Collins, C. D. N. (2001). Inversion of seismic refraction and wide-angle reflection traveltimes for 3-D layered crustal structure. *Geophysical Journal International*, *145*, 381–401.
- Rawlinson, N., & Kennett, B. L. N. (2004). Rapid estimation of relative and absolute delay times across a network by adaptive stacking. *Geophysical Journal International*, *157*, 332–340.
- Rawlinson, N., & Kennett, B. L. N. (2008). Teleseismic tomography of the upper mantle beneath the southern Lachan Orogen, Australia. *Physics of the Earth and Planetary Interiors*, *167*, 84–97.
- 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 Research*, *19*, 788–799.
- Rawlinson, N., Pozgay, S., & Fishwick, S. (2010). Seismic tomography: a window into deep Earth. *Physics of the Earth and Planetary Interiors*, *178*, 101–135.

- Rawlinson, N., Reading, A. M., & Kennett, B. L. N. (2006). Lithospheric structure of Tasmania from a novel form of teleseismic tomography. *Journal of Geophysical Research*, *111*. <http://dx.doi.org/10.1029/2005JB003803>.
- Rawlinson, N., Salmon, N., & Kennett, B. L. N. (2013). Transportable seismic array tomography in southeast Australia: Illuminating the transition from Proterozoic to Phanerozoic lithosphere. *Lithos*, *189*, 65–76.
- Rawlinson, N., & Sambridge, M. (2003). Irregular interface parameterization in 3-D wide-angle seismic traveltime tomography. *Geophysical Journal International*, *155*, 79–92.
- Rawlinson, N., & Sambridge, M. (2004). Wavefront evolution in strongly heterogeneous layered media using the fast marching method. *Geophysical Journal International*, *156*, 631–647.
- Rawlinson, N., Sambridge, M., & Saygin, E. (2008). A dynamic objective function technique for generating multiple solution models in seismic tomography. *Geophysical Journal International*, *174*, 295–308.
- Rawlinson, N., Tkalčić, H., & Reading, A. M. (2010). Structure of the Tasmanian lithosphere from 3-D seismic tomography. *Australian Journal of Earth Sciences*, *57*, 381–394.
- Rawlinson, N., & Urvoy, M. (2006). Simultaneous inversion of active and passive source datasets for 3-D seismic structure with application to Tasmania. *Geophysical Research Letters*, *33*, L24313.
- Riahi, M. A., & Juhlin, C. (1994). 3-D interpretation of reflected arrival times by finite-difference techniques. *Geophysics*, *59*, 844–849.
- Rickers, F., Fichtner, A., & Trampert, J. (2013). The Iceland – Jan Mayen plume system and its impact on mantle dynamics in the North Atlantic region: evidence from full-waveform inversion. *Earth and Planetary Science Letters*, *367*, 39–51.
- Ritsema, J., Nyblade, A. A., Owens, T. J., Langston, C. A., & VanDecar, J. C. (1998). Upper mantle seismic velocity structure beneath Tanzania, east Africa: Implications for the stability of cratonic lithosphere. *Journal of Geophysical Research*, *103*, 21,201–21,213.
- Romanowicz, B. (2003). Global mantle tomography: progress status in the past 10 years. *Annual Review of Earth and Planetary Sciences*, *31*, 303–328.
- Romanowicz, B., & Gung, Y. (2002). Superplumes from the core–mantle boundary to the Lithosphere: Implications for heat flux. *Science*, *296*, 513–516.
- Rothman, D. H. (1985). Nonlinear inversion statistical mechanics, and residuals corrections. *Geophysics*, *50*, 2784–2796.
- Rothman, D. H. (1986). Automatic estimation of large residual statics corrections. *Geophysics*, *51*, 332–346.
- Saito, M. (1988). Disper80: a subroutine package for the calculation of seismic normal mode solutions. In D. Doornbos (Ed.), *Seismological algorithms: Computational methods and computer programs* (pp. 293–319). Academic Press.
- Saltzer, R. L., & Humphreys, E. D. (1997). Upper mantle *P* wave velocity structure of the eastern Snake River Plain and its relationship to geodynamic models of the region. *Journal of Geophysical Research*, *102*, 11,829–11,841.
- Sambridge, M. S. (1990). Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D. *Geophysical Journal International*, *102*, 653–677.
- Sambridge, M. (1999). Geophysical inversion with a Neighbourhood algorithm – I. Searching parameter space. *Geophysical Journal International*, *138*, 479–494.
- Sambridge, M. (2014). A Parallel Tempering algorithm for probabilistic sampling and multi-modal optimization. *Geophysical Journal International*, *196*, 357–374.
- Sambridge, M. S., Braun, J., & McQueen, H. (1995). Geophysical parametrization and interpolation of irregular data using natural neighbours. *Geophysical Journal International*, *122*, 837–857.
- Sambridge, M. S., & Drijkoningen, G. (1992). Genetic algorithms in seismic waveform inversion. *Geophysical Journal International*, *109*, 323–342.

- Sambridge, M., & Faletic, R. (2003). Adaptive whole Earth tomography. *Geochemistry, Geophysics, Geosystems*, 4. <http://dx.doi.org/10.1029/2001GC000213>.
- Sambridge, M., & Gudmundsson, O. (1998). Tomographic systems of equations with irregular cells. *Journal of Geophysical Research*, 103, 773–781.
- Sambridge, M., & Mosegaard, K. (2002). Monte Carlo methods in geophysical inverse problems. *Reviews of Geophysics*, 40. <http://dx.doi.org/10.1029/2000RG000089>.
- Santosa, F., & Symes, W. W. (1988). Computation of the Hessian for least-squares solutions of inverse problems of reflection seismology. *Inverse Problems*, 4, 211–233.
- Saygin, E., & Kennett, B. (2009). Ambient seismic noise tomography of Australian continent. *Tectonophysics*. <http://dx.doi.org/10.1016/j.tecto.2008.11.013>.
- Scales, J. A. (1987). Tomographic inversion via the conjugate gradient method. *Geophysics*, 52, 179–185.
- Scales, J. A., Gersztenkorn, A., & Treitel, S. (1988). Fast lp solution of large, sparse linear systems: application to seismic traveltime tomography. *Journal of Computational Physics*, 75, 314–333.
- Scales, J. A., & Snieder, R. (1997). To Bayes or not to Bayes. *Geophysics*, 62, 1045–1046.
- Shapiro, N. M., & Campillo, M. (2004). Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise. *Geophysical Research Letters*, 31. <http://dx.doi.org/10.1029/2004GL019491>.
- Shapiro, N. M., Campillo, M., Stehly, L., & Ritzwoller, M. H. (2005). High-resolution surface wave tomography from ambient seismic noise. *Science*, 307, 1615–1618.
- Shapiro, N. M., & Ritzwoller, M. H. (2002). Monte-Carlo inversion for a global shear-velocity model of the crust and upper mantle. *Geophysical Journal International*, 151, 88–105.
- Simons, F. J., Loris, I., Nolet, G., Daubechies, I. C., Voronin, S., Judd, J. S., et al. (2011). Solving or resolving global tomographic models with spherical wavelets, and the scale and sparsity of seismic heterogeneity. *Geophysical Journal International*, 187, 969–988.
- Smithyman, B., Pratt, R. G., Hayles, J., & Wittebolle, R. (2009). Detecting near-surface objects with seismic waveform tomography. *Geophysics*, 74, WCC119–WCC127.
- Snieder, R. (1991). An extension of backus-gilbert theory to nonlinear inverse problems. *Inverse Problems*, 7, 409–433.
- Spakman, W., & Bijwaard, H. (2001). Optimization of cell parameterizations for tomographic inverse problems. *Pure and Applied Geophysics*, 158, 1401–1423.
- Spakman, W., & Nolet, G. (1988). Imaging algorithms, accuracy and resolution in delay time tomography. In N. J. Vlaar, G. Nolet, M. J. R. Wortel, & S. A. P. Cloetingh (Eds.), *Mathematical geophysics: A survey of recent developments in seismology and geodynamics* (pp. 155–187). Netherlands: Springer.
- Steck, L. K., Thurber, C. H., Fehler, M., Lutter, W. J., Roberts, P. M., Baldrige, W. S., et al. (1998). Crust and upper mantle P wave velocity structure beneath Valles caldera, New Mexico: results from the Jemez teleseismic tomography experiment. *Journal of Geophysical Research*, 103, 24,301–24,320.
- Stoffa, P. L., & Sen, M. K. (1991). Nonlinear multiparameter optimization using genetic algorithms: Inversion of plane wave seismograms. *Geophysics*, 56, 1794–1810.
- Su, W.-J., & Dziewonski, A. M. (1997). Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle. *Physics of the Earth and Planetary Interiors*, 100, 135–156.
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2009). Adjoint tomography of the Southern California crust. *Science*, 325, 988–992.
- Tape, C., Liu, Q., Maggi, A., & Tromp, J. (2010). Seismic tomography of the Southern California crust based on spectral-element and adjoint methods. *Geophysical Journal International*, 180, 433–462.
- Tarantola, A. (1987). *Inverse problem theory*. Amsterdam: Elsevier.

- Tarantola, A. (1988). Theoretical background for the inversion of seismic waveforms, including elasticity and attenuation. *Pure and Applied Geophysics*, 128, 365–399.
- Tarantola, A., & Nercessian, A. (1984). Three-dimensional inversion without blocks. *Geophysical Journal of the Royal Astronomical Society*, 76, 299–306.
- Thurber, C. H. (1983). Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California. *Journal of Geophysical Research*, 88, 8226–8236.
- Tikhotsky, S., & Achauer, U. (2008). Inversion of controlled-source seismic tomography and gravity data with self-adaptive wavelet parameterization of velocities and interfaces. *Geophysical Journal International*, 172, 619–630.
- Trampert, J., & Fichtner, A. (2013a). Global imaging of the Earth's deep interior: seismic constraints on (an)isotropy, density and attenuation. In S. Karato (Ed.), *Physics and chemistry of the deep Earth* (pp. 324–350). Wiley-Blackwell.
- Trampert, J., & Fichtner, A. (2013b). Resolution tests revisited: the power of random numbers. *Geophysical Journal International*, 192, 676–680.
- Trampert, J., & Woodhouse, J. H. (1995). Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds. *Geophysical Journal International*, 122, 675–690.
- Tromp, J., Tape, C., & Liu, Q. (2005). Seismic tomography, adjoint methods, time reversal and banana-doughnut kernels. *Geophysical Journal International*, 160, 195–216.
- Um, J., & Thurber, C. (1987). A fast algorithm for two-point seismic ray tracing. *Bulletin of Seismological Society of America*, 77, 972–986.
- VanDecar, J. C., & Crosson, R. S. (1990). Determination of teleseismic relative phase arrival times using multi-channel cross-correlation and least squares. *Bulletin of Seismological Society of America*, 80, 150–169.
- VanDecar, J. C., James, D. E., & Assumpção, M. (1995). Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces. *Nature*, 378, 25–31.
- Vasco, D. W. (2007). Invariance, groups, and non-uniqueness: the discrete case. *Geophysical Journal International*, 168, 473–490.
- Vasco, D. W., & Johnson, L. R. (1998). Whole earth structure estimated from seismic arrival times. *Journal of Geophysical Research*, 103, 2633–2671.
- Vasco, D. W., Peterson, J. E., & Majer, E. L. (1996). Nonuniqueness in traveltimes tomography: ensemble inference and cluster analysis. *Geophysics*, 61, 1209–1227.
- Vassallo, M., Satriano, C., & Lomax, A. (2012). Automatic picker developments and optimization: a strategy for improving the performances of automatic phase pickers. *Seismological Research Letters*, 83, 541–554.
- Vesnaver, A., Böhm, G., Madrussani, G., Rossi, G., & Granser, H. (2000). Depth imaging and velocity calibration by 3 D adaptive tomography. *First Break*, 18, 303–312.
- Walck, M. C., & Clayton, R. W. (1987). *P* wave velocity variations in the Coso region, California, derived from local earthquake travel times. *Journal of Geophysical Research*, 92, 393–405.
- Wang, Y., & Houseman, G. A. (1997). Point source $\tau - p$ transform: a review and comparison of computational methods. *Geophysics*, 62, 325–334.
- Wessel, P., Smith, W. H. F., Scharroo, R., Luis, J. F., & Wobbe, F. (2013). Generic mapping tools: improved version released. *Eos, Transactions American Geophysical Union*, 94, 409–410.
- West, M., Gao, W., & Grand, S. (2004). A simple approach to the joint inversion of seismic body and surface waves applied to the southwest U.S. *Geophysical Research Letters*, 31.
- White, D. J. (1989). Two-dimensional seismic refraction tomography. *Geophysical Journal International*, 97, 223–245.
- Wiggins, R. A. (1972). The general linear inverse problem: Implication of surface waves and free oscillations for earth structure. *Journal of Geophysical Research*, 10, 251–285.

- Wolfe, C. J., Solomon, S. C., Laske, G., Collins, J. A., Detrick, R. S., Orcutt, J. A., et al. (2009). Mantle shear-wave velocity structure beneath the Hawaiian Hot Spot. *Science*, *326*, 1388–1390.
- Wolfe, C. J., Solomon, S. C., Silver, P. G., VanDecar, J. C., & Russo, R. M. (2002). Inversion of body-wave delay times for mantle structure beneath the Hawaiian islands: results from the PELENET experiment. *Earth and Planetary Science Letters*, *198*, 129–145.
- Yang, T., Grand, S. P., Wilson, D., Guzman-Speziale, M., Gomez-Gonzalez, J., Dominguez-Teyes, T., et al. (2009). Seismic structure beneath the Rivera subduction zone from finite-frequency seismic tomography. *Journal of Geophysical Research*, *114*, B01302. <http://dx.doi.org/10.1029/2008JB005830>.
- Yang, Y., Ritzwoller, M. H., Levshin, A. L., & Shapiro, N. M. (2007). Ambient noise Rayleigh wave tomography across Europe. *Geophysical Journal International*, *168*, 259–274.
- Yao, Z. S., Roberts, R. G., & Tryggvason, A. (1999). Calculating resolution and covariance matrices for seismic tomography with the LSQR method. *Geophysical Journal International*, *138*, 886–894.
- Young, M. K., Cayley, R. A., Mclean, M. A., Rawlinson, N., Arroucau, P., & Salmon, M. (2013). Crustal Structure of the east Gondwana margin in southeast Australia revealed by transdimensional ambient seismic noise tomography. *Geophysical Research Letters*, *40*, 4266–4271.
- Young, M. K., Rawlinson, N., Arroucau, P., Reading, A. M., & Tkalčić, H. (2011). High-frequency ambient noise tomography of southeast Australia: new constraints on Tasmania's tectonic past. *Geophysical Research Letters*, *38*, L13313. <http://dx.doi.org/10.1029/2011GL047971>.
- Young, M. K., Rawlinson, N., & Bodin, T. (2013). Transdimensional inversion of ambient seismic noise for 3D shear velocity structure of the tasmanian crust. *Geophysics*, *78*, WB49–WB62.
- Zelt, C. A. (1999). Modelling strategies and model assessment for wide-angle seismic travel-time data. *Geophysical Journal International*, *139*, 183–204.
- Zelt, C. A., & Barton, P. J. (1998). Three-dimensional seismic refraction tomography: a comparison of two methods applied to data from the Faeroe Basin. *Journal of Geophysical Research*, *103*, 7187–7210.
- Zelt, C. A., & Smith, R. B. (1992). Seismic traveltime inversion for 2-D crustal velocity structure. *Geophysical Journal International*, *108*, 16–34.
- Zhang, J., & McMechan, G. A. (1995). Estimation of resolution and covariance for large matrix inversions. *Geophysical Journal International*, *121*, 409–426.
- Zhang, J., Rector, J. W., III, III, & Hoversten, G. (2005). Eikonal solver in the celerity domain. *Geophysical Journal International*, *162*, 1–8.
- Zhang, H., & Thurber, C. H. (2007). Estimating the model resolution matrix for large seismic tomography problems based on Lanczos bidiagonalization with partial reorthogonalization. *Geophysical Journal International*, *170*, 337–345.
- Zhang, J., & Toksöz, M. N. (1998). Nonlinear refraction traveltime tomography. *Geophysics*, *63*, 1726–1737.
- Zhao, D., Hasegawa, A., & Horiuchi, S. (1992). Tomographic imaging of *P* and *S* wave velocity structure beneath Northeastern Japan. *Journal of Geophysical Research*, *97*, 19,909–19,928.
- Zhao, D., Hasegawa, A., & Kanamori, H. (1994). Deep structure of Japan subduction zone as derived from local, regional, and teleseismic events. *Journal of Geophysical Research*, *99*, 22,313–22,329.