

# A dynamic objective function technique for generating multiple solution models in seismic tomography

N. Rawlinson, M. Sambridge and E. Saygin

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

Accepted 2008 April 1. Received 2008 April 1; in original form 2007 December 3

## SUMMARY

A new technique designed for generating multiple solutions to seismic tomography problems using gradient based inversion is presented. The basic principle is to exploit information gained from previous solutions to help drive the search for new models. This is achieved by adding a feedback or evolution term to the objective function that creates a local maximum at each point in parameter space occupied by the previously computed models. The advantage of this approach is that it only needs to produce a relatively small ensemble of solutions, since each model will substantially differ from all others to the extent permitted by the data. Common features present across the ensemble are, therefore, likely to be well constrained. A synthetic test using surface wave traveltimes and a highly irregular distribution of sources and receivers shows that a range of different velocity models are produced by the new technique. These models tend to be similar in regions of good path coverage, but can differ substantially elsewhere. A simple measure of the variation across the solution ensemble, given by one standard deviation of the velocity at each point, accurately reflects the robustness of the average solution model. Comparison with a standard bootstrap inversion method unequivocally shows that the new approach is superior in the presence of inhomogeneous data coverage that gives rise to under or mixed-determined inverse problems. Estimates of posterior covariance from linear theory correlate more closely with the dynamic objective function results, but require accurate knowledge of *a priori* model uncertainty. Application of the new method to traveltimes derived from long-term cross-correlations of ambient noise contained in passive seismic data recorded in the Australian region demonstrates its effectiveness in practice, with results well corroborated by prior information. The dynamic objective function scheme has several drawbacks, including a somewhat arbitrary choice for the shape of the evolution term, and no guarantee of a thorough exploration of parameter space. On the other hand, it is tolerant of non-linearity in the inverse problem, is relatively straightforward to implement, and appears to work well in practice. For many applications, it may be a useful addition to the suite of synthetic resolution tests that are commonly used.

**Key words:** Inverse theory; Seismic tomography; Computational seismology; Australia.

## 1 INTRODUCTION

The under or mixed determined nature of the non-linear inverse problem in seismic tomography usually means that an infinite number of different models satisfy the data. Monte Carlo direct search methods, like the Metropolis algorithm, genetic algorithms and Neighbourhood Algorithm, widely explore parameter space and produce multiple solutions that can be interrogated for consistent information using ensemble inference techniques (e.g. Lomax & Snieder 1994; Sambridge 1998, 1999a,b; Koper *et al.* 1999; Mosegaard & Sambridge 2002). However, the large number of unknowns present in most classes of seismic tomography makes a fully non-linear approach computationally impractical. As a result,

iterative non-linear tomography based on using local gradient information to direct the search remains the method of choice in many applications (e.g. Farra & Madariaga 1988; Zelt & Smith 1992; Zelt & Barton 1998; Graeber & Asch 1999; Bijwaard & Spakman 2000; Rawlinson *et al.* 2001; Gorbato & Kennett 2003; Bai & Greenhalgh 2005; Rawlinson *et al.* 2006).

The basic assumption underlying gradient based minimization is that the inverse problem is locally linearizable, which requires the objective function to be smooth, and the initial model  $\mathbf{m}_0$  to be sufficiently close to the required minimum, for the technique to successfully converge. Backus & Gilbert (1970) refer to a solution  $\mathbf{m}$  obtained in this way as being ‘g-near’ to  $\mathbf{m}_0$ , since the functional  $\mathbf{g}(\mathbf{m})$  (e.g. traveltimes predictions) is computed from

$\mathbf{g}(\mathbf{m}_0)$  using first-order perturbation theory. If second or higher order terms make a significant contribution to  $\mathbf{g}(\mathbf{m}) - \mathbf{g}(\mathbf{m}_0)$ , then  $\mathbf{m}$  is ‘g-far’ from  $\mathbf{m}_0$ . While it is feasible for iterative non-linear inversion to produce ‘g-far’ solutions, both stability and solution robustness will degrade as higher order terms become more dominant. The wealth of new information on Earth structure and composition that has been revealed by seismic tomography over the last several decades is testament to the validity of the ‘g-near’ assumption for many data sets (see Nolet 1987; Iyer & Hirahara 1993; Rawlinson & Sambridge 2003, for many examples). However, since only a single model is usually produced by this technique, the problem of solution non-uniqueness must still be addressed. To date, a number of different approaches have been used to examine the robustness of solutions obtained using gradient based tomography.

One of the most common schemes for assessing the validity of a solution model is the synthetic resolution test, in which an input model is used to generate an artificial data set using the same source and receiver configuration as the data set recorded in the field. The ability of the inversion scheme to recover the input model can then be used to evaluate the robustness of the solution model. The ubiquitous checkerboard resolution test, which uses an input model consisting of an alternating pattern of fast and slow anomalies, is an example of this class of numerical experiment (e.g. Walck 1988; Glahn & Granet 1993; Ritsema *et al.* 1998; Day *et al.* 2001; Graeber *et al.* 2002; Rawlinson & Urvoy 2006). Although useful, synthetic resolution tests have a number of limitations, including that the results can vary according to the input structure used (e.g. L ev eque *et al.* 1993).

Another means of assessing solution robustness comes from linear theory by way of formal estimates of posterior covariance and resolution (Tarantola 1987; Menke 1989). Quantitative information regarding model uncertainty is valuable, which explains the widespread use of these measures in seismic tomography (e.g. Aki *et al.* 1977; White 1989; Benz *et al.* 1992; Wang & Braile 1996; Graeber & Asch 1999). Drawbacks of using posterior covariance and resolution include that their validity decreases as the non-linearity of the inverse problem increases; they require the inversion of a potentially large  $M \times M$  matrix, where  $M$  is the number of unknowns; and errors in model representation are not taken into account (although this problem is hard to overcome with most forms of robustness analysis). Furthermore, the absolute values of the resolution and covariance estimates can become rather meaningless when the *a priori* model covariance and data uncertainties are poorly estimated. In addition, regularized solutions, which by their very nature suppress noise propagation from the data, also suppress formal model uncertainties in the least well-constrained parts of the model. Nevertheless, their relative values can still be useful indicators of the effect the data have in constraining the solution model. The problem of attempting to directly invert large sparse matrices when many model parameters are involved has been mitigated in recent times by modifying iterative approaches such as LSQR (Zhang & McMechan 1995; Yao *et al.* 1999) to approximate the generalized inverse, and using sophisticated packages such as PROPACK, which accurately estimate singular values and vectors for large sparse matrices (Zhang & Thurber 2007).

A number of other techniques for analysing solution non-uniqueness in seismic tomography have been investigated. Debayle & Sambridge (2004) estimate the minimum length scale of resolvable structure as a function of location in the model using Voronoi diagrams; the size and shape of Voronoi cells are constrained by a quality criterion which is based on ray density and azimuthal cover-

age. This approach shares similarities with irregular mesh tomography in which the distribution of grid points or cells which define the model parametrization is adjusted according to some measure of data constraint (Sambridge & Rawlinson 2005). Statistical methods of error estimation based on multiple inversions with different components of the data set, such as bootstrapping and jackknifing, have also been used (Lees & Crosson 1989; Su & Dziewonski 1997; Zelt 1999; Gung & Romanowicz 2004). However, Nolet *et al.* (1999) point out that both bootstrapping and jackknifing rely on overdetermined inverse problems which don’t normally occur in seismic tomography.

Direct search methods which generate multiple solution models, and then apply some form of statistical analysis to decide which features of the ensemble are significant, have much appeal. For instance, they can obviate the need for static regularization that is often used in gradient based tomography to tune the solution, and explore parameter space more thoroughly. The question that is addressed in this paper is whether it is feasible to generate a meaningful ensemble of data satisfying solution models via iterative non-linear inversion. One possibility that has been investigated previously is to use multiple starting models (e.g. Vasco *et al.* 1996). Thus, the gradient-based procedure is initialized from different points in model space, and therefore, may converge to different solutions. Vasco *et al.* (1996) apply cluster analysis to an ensemble of 1075 tomography models generated in this way from cross-hole data. In theory, one could also generate a family of solutions by identifying the model null space vectors (using Singular Value Decomposition, SVD, for example), and then varying a given solution model only in the null space, such that the data fit remains unchanged. The so-called ‘nullspace shuttle’ proposed by Deal & Nolet (1996) essentially follows this principle, as it provides the ability to move from one solution to another without compromising the data fit. The main drawback of this technique is that it is strictly valid for linear inverse problems. The new approach tested here is quite different, in that it modifies the objective function after each model is produced in order to prevent subsequent solutions from residing near the same point in parameter space. The aim is to produce a relatively small number of acceptable models that are as different from one another as possible.

After describing the dynamic objective function scheme, several tests are carried out using a synthetic surface wave traveltime data set that features a strongly heterogeneous path distribution. This initial validation process allows the behaviour of the scheme and its sensitivities with respect to data coverage and model complexity to be examined. Information extracted from the solution ensemble can also be directly compared to the true model. A bootstrap analysis is performed with the same data set to enable comparison with a standard statistical technique. Formal estimates of posterior covariance from linear theory are also computed for the purposes of comparison. Following the synthetic tests, the new scheme is applied to surface wave traveltimes extracted from the long-term cross-correlations of ambient noise data collected in the Australian region.

## 2 METHOD

### 2.1 Formulation of the dynamic objective function

The inverse problem in seismic tomography can be posed in terms of minimizing an objective function consisting of a data term and one or more regularization terms. A common form of objective function

used in gradient based inversion is:

$$S(\mathbf{m}) = \frac{1}{2} \{ [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}]^T \mathbf{C}_d^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}] + \epsilon (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) + \eta \mathbf{m}^T \mathbf{D}^T \mathbf{D} \mathbf{m} \}, \quad (1)$$

where  $\mathbf{m}$  are the model parameters,  $\mathbf{g}(\mathbf{m})$  the data predictions,  $\mathbf{d}_{\text{obs}}$  the data observations,  $\mathbf{C}_d$  the *a priori* data covariance matrix,  $\mathbf{m}_0$  the reference model,  $\mathbf{C}_m$  the *a priori* model covariance matrix and  $\mathbf{D}$  the second derivative smoothing operator (see Rawlinson & Sambridge 2003, for a detailed discussion).  $\epsilon$  and  $\eta$  are known as the damping parameter and smoothing parameter, respectively and govern the trade-off between how well the solution satisfies the data, the proximity of the solution model to the starting model and the smoothness of the solution model.

Damping and smoothing regularization are usually required in gradient based tomography to tame the data term in eq. (1) so that the objective function becomes smooth and well behaved, that is, more locally quadratic so that for some initial model  $\mathbf{m}_0$ , there exists a ‘g-near’ solution  $\mathbf{m}$ . However, they generally do not contribute meaningful *a priori* information to the solution model, which is why they are normally dispensed with when direct search methods are used (e.g. Lomax & Snieder 1994; Sambridge 1998). The appeal of producing multiple solution models, discarding static regularization, yet retaining a gradient based inversion framework, gave rise to the idea of continuously modifying the shape of the misfit surface defined by  $S(\mathbf{m})$  in response to information extracted from the data. The basic principle is that once a point in model space has been sampled, then there is little to be gained from revisiting its neighbourhood in future searches unless the data constraints are particularly compelling. This can be formalized in an objective function by the cumulative addition of an

explicit evolution or feedback term that is a maximum at each point in parameter space occupied by previously generated models (see Fig. 1). Thus, by beginning at some point  $\mathbf{m}_0$  in parameter space, each successive solution model will be distinct from all of its predecessors.

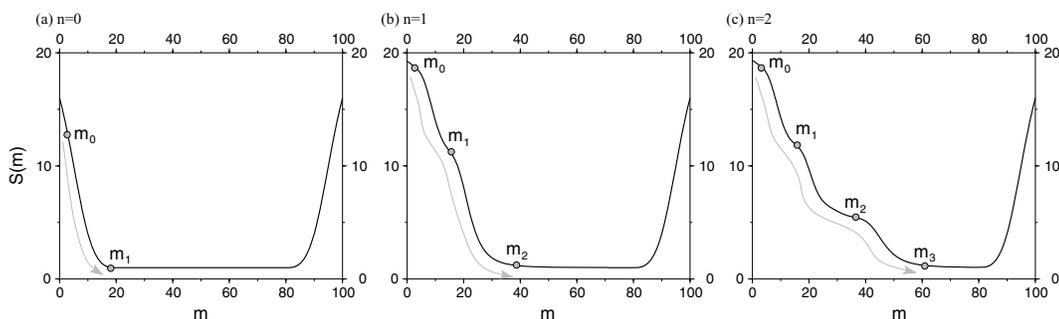
The feedback or evolution term  $\Omega(\mathbf{m})$  that is added to  $S(\mathbf{m})$  could potentially take many forms, but it is important that it does not strongly influence the objective function in regions of model space distant from any of the previously generated solutions. Therefore, simply setting:

$$\Omega_j(\mathbf{m}) = - \sum_{i=1}^j (\mathbf{m} - \mathbf{m}_i)^T (\mathbf{m} - \mathbf{m}_i) \quad j = 1, \dots, N, \quad (2)$$

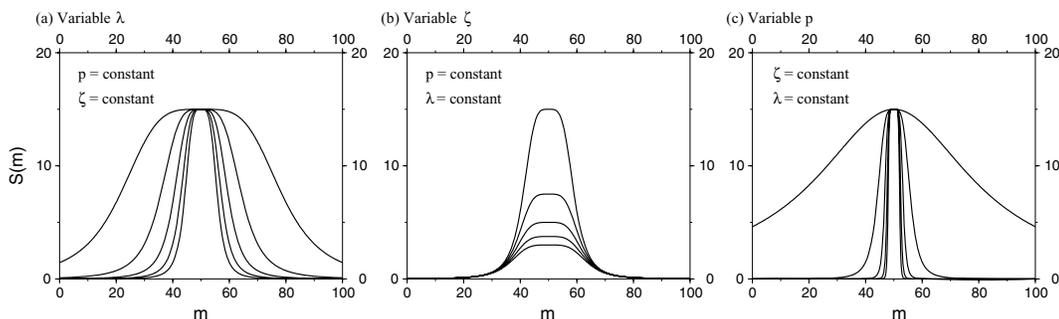
where  $N$  is the total number of models generated, would be unsatisfactory, for while it is a maximum where  $\mathbf{m} = \mathbf{m}_i$ , it decreases quadratically away from this point to  $-\infty$ . A better choice is:

$$\Omega_j(\mathbf{m}) = \sum_{i=1}^j \frac{1}{\lambda [(\mathbf{m} - \mathbf{m}_i)^T (\mathbf{m} - \mathbf{m}_i)]^p + \zeta} \quad j = 1, \dots, N, \quad (3)$$

where  $\lambda$ ,  $p$  and  $\zeta$  are free parameters. In this case,  $\Omega(\mathbf{m})$  is maximum when  $\mathbf{m} = \mathbf{m}_i$ , and decreases to zero away from this point. The factor  $\zeta$  determines the maximum value of  $\Omega(\mathbf{m})$ ,  $\lambda$  the width of  $\Omega(\mathbf{m})$ , and  $p$  the curvature of  $\Omega(\mathbf{m})$  (see Fig. 2). By adjusting the values of these free parameters, the influence of the evolution term on the objective function can be effectively controlled. Ultimately, the values of  $\zeta$ ,  $\lambda$  and  $p$  are chosen by trial and insight, but it turns out that cogent choices can be made based on the character of the data misfit function (see next section).



**Figure 1.** Schematic illustration of the dynamic objective function approach for generating multiple solution models. In this case, the objective function  $S(\mathbf{m})$  is simply a function of a single model parameter  $m$ . (a) No feedback term is present when the first solution is found; (b) when the second solution  $m_2$  is computed, the minimization is influenced by the presence of a feedback term due to the first model  $m_1$ ; (c) the third solution  $m_3$  is deterred from converging to the previous two solutions by the presence of two feedback terms centred at  $m_1$  and  $m_2$ .



**Figure 2.** Demonstration of how the geometry of the evolution term varies as a function of each of the free parameters (a)  $\lambda$ , (b)  $\zeta$  and (c)  $p$ .

In the following treatment, the damping and smoothing terms in eq. (1) are discarded so that the objective function becomes:

$$S_j(\mathbf{m}) = \frac{1}{2} \left\{ [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}]^T \mathbf{C}_d^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}] + \sum_{i=1}^j \frac{1}{\lambda [(\mathbf{m} - \mathbf{m}_i)^T (\mathbf{m} - \mathbf{m}_i)]^p + \zeta} \right\} \quad j = 1, \dots, N. \quad (4)$$

The first solution that is generated has the evolution term set to zero, since no model has yet been produced. Alternatively, one could set  $\mathbf{m}_0 = \mathbf{m}_1$  and use the starting model as the first solution model that is fed back into the objective function, but the former approach is used here. The scheme progresses by feeding each solution model that is produced back into the objective function, which is subsequently minimized to produce a new model. This iterative process continues until  $N$  models are generated (see Fig. 1). The lack of explicit regularization, at least in the first iteration of the scheme, may well destabilize a gradient-based minimization process. However, it turns out that the implicit regularization imposed by the cubic B-spline parametrization and subspace inversion method used in the later examples mitigates this potential hazard.

In comparing eqs (1) and (4), it is worth noting that both techniques rely on regularization to produce a model from a non-unique problem. However, in the case of eq. (1), the regularization is static and a single solution is produced. By contrast, eq. (4) recognizes the presence of multiple solutions, and dynamically adjusts the regularization in order to generate a family of dissimilar data fitting solutions.

Gradient-based inversion methods make use of the derivatives of  $S(\mathbf{m})$  at a specified point in model space, and require  $S(\mathbf{m})$  to be sufficiently smooth to permit a local quadratic approximation:

$$S(\mathbf{m} + \delta\mathbf{m}) \approx S(\mathbf{m}) + \hat{\gamma}\mathbf{m} + \frac{1}{2} \delta\mathbf{m}^T \hat{\mathbf{H}} \delta\mathbf{m}, \quad (5)$$

where  $\delta\mathbf{m}$  is a perturbation to the current model and  $\hat{\gamma} = \partial S / \partial \mathbf{m}$  and  $\hat{\mathbf{H}} = \partial^2 S / \partial \mathbf{m}^2$  are the gradient vector and Hessian matrix, respectively. Evaluating these partial derivatives for the above form of  $S(\mathbf{m})$  yields:

$$\hat{\gamma} = \mathbf{G}^T \mathbf{C}_d^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}] - \sum_{i=1}^j a_i (\mathbf{m} - \mathbf{m}_i) \quad (6)$$

$$\hat{\mathbf{H}} = \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \nabla_{\mathbf{m}} \mathbf{G}^T \mathbf{C}_d^{-1} [\mathbf{g}(\mathbf{m}) - \mathbf{d}_{\text{obs}}] + \sum_{i=1}^j b_i (\mathbf{m} - \mathbf{m}_i)(\mathbf{m} - \mathbf{m}_i)^T, \quad (7)$$

where  $\mathbf{G} = \partial \mathbf{g} / \partial \mathbf{m}$  is the Fréchet matrix of partial derivatives calculated during the solution of the forward problem, and the coefficients  $a_i$  and  $b_i$  are given by:

$$a_i = \frac{\lambda p q_i^{p-1}}{[\lambda q_i^p + \zeta]^2} \quad (8)$$

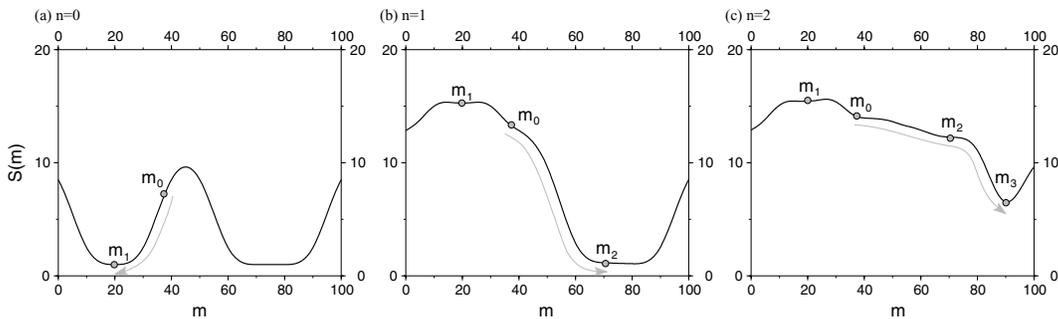
$$b_i = \frac{2\lambda p q_i^{p-2} (\lambda q_i^p + \zeta) [\lambda q_i^p (p+1) - \zeta (p-1)]}{[\lambda q_i^p + \zeta]^4}, \quad (9)$$

where  $q_i = (\mathbf{m} - \mathbf{m}_i)^T (\mathbf{m} - \mathbf{m}_i)$ . Note that  $a_i$  and  $b_i$  depend non-linearly on the model  $\mathbf{m}$  and act as weights in the gradient vector and Hessian matrix.

Fig. 1 illustrates how the dynamic objective function scheme can explore parameter space when the data term is characterized by a broad minimum, as might occur in a very underdetermined inverse problem. If the functional  $\mathbf{g}(\mathbf{m})$  is sufficiently non-linear, it is also possible that multiple minima exist, in which case there may be more than one distinct solution. Direct search techniques such as the niching genetic algorithm (Koper *et al.* 1999) and Neighbourhood algorithm (Sambridge 1999a) are able to target and explore these favourable regions of model space. In contrast, iterative non-linear schemes will converge to a single solution with the aid of suitable regularization, as they are not designed to explore multimodal objective functions. Using a dynamic objective function approach, it is possible to explore multimodal landscapes, as illustrated in Fig. 3, but it would be difficult to ensure that a thorough search is performed when  $\mathbf{g}(\mathbf{m})$  is highly non-linear.

## 2.2 Minimization of the dynamic objective function

In the examples that follow, surface wave traveltime tomography will be used to evaluate the new technique for generating multiple models and assessing solution robustness. The forward problem of computing source–receiver traveltimes in order to evaluate the data prediction vector  $\mathbf{g}(\mathbf{m})$  is solved using a grid based eikonal solver known as the fast marching method or FMM (Sethian 1996; Popovici & Sethian 2002; Rawlinson & Sambridge 2004a,b; de Kool *et al.* 2006). The scheme works by implicitly tracking the first-arrival wave front along an evolving narrow band of grid points, the traveltimes of which are updated by solving the eikonal equation using upwind entropy satisfying operators. Here, the finite difference operators are defined for 2-D spherical shell coordinates (constant radius, but variable latitude and longitude). The Fréchet matrix of partial derivatives  $\mathbf{G}$  is computed by integration along each ray path, which can be located by following the traveltime gradient from the receiver, through the traveltime field, back to the source. Source



**Figure 3.** A simple example showing how the dynamic objective function scheme could work in a multimodal landscape. (a) First iteration with no feedback term; (b) second iteration in which the feedback term produced by the first solution enables the remaining minima to be explored and (c) third iteration.

grid refinement (Rawlinson & Sambridge 2004b) is implemented to account for the undersampling of the traveltime field in the source neighbourhood; ultimately, this leads to significant improvements in accuracy with little additional computational cost.

An iterative non-linear procedure is used to minimize the objective function, which successively applies the FMM scheme to solve the forward problem, and a subspace inversion method to solve the linearized inverse problem. Thus, for some current model  $\mathbf{m}^k$ , a new model  $\mathbf{m}^{k+1} = \mathbf{m}^k + \delta\mathbf{m}^k$  is produced where the model perturbation  $\delta\mathbf{m}^k$  is supplied by the subspace inversion method. The scheme progresses until the observed data are satisfied or when the change in  $S(\mathbf{m})$  with iteration gets sufficiently small.

The subspace inversion method works by projecting the quadratic approximation of  $S(\mathbf{m})$  onto an  $n$ -dimensional subspace of the model space (Kennett *et al.* 1988; Sambridge 1990; Williamson 1990). For the dynamic objective function defined above (eq. 4), the perturbation  $\delta\mathbf{m}$  is given by

$$\delta\mathbf{m} = -\mathbf{A} \left\{ \mathbf{A}^T \left[ \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \sum_{i=1}^j b_i (\mathbf{m} - \mathbf{m}_i)(\mathbf{m} - \mathbf{m}_i)^T \right] \mathbf{A} \right\}^{-1} \mathbf{A}^T \hat{\boldsymbol{\gamma}}, \quad (10)$$

where  $\mathbf{A} = [\mathbf{a}^l]$  is the  $M \times n$  projection matrix (for  $M$  unknowns),  $\mathbf{G}$  is the matrix of Fréchet derivatives and  $\hat{\boldsymbol{\gamma}}$  is the gradient vector ( $\hat{\boldsymbol{\gamma}} = \partial S / \partial \mathbf{m}$ ). The basis vectors that span the  $n$ -dimensional subspace are in our case based on the gradient vector in model space  $\boldsymbol{\gamma} = \mathbf{C}_m \hat{\boldsymbol{\gamma}}$  and the model space Hessian  $\mathbf{H} = \mathbf{C}_m \hat{\mathbf{H}}$ , where  $\hat{\mathbf{H}} = \partial^2 S / \partial \mathbf{m}^2$  (note that for eq. 4,  $\boldsymbol{\gamma} = \hat{\boldsymbol{\gamma}}$  and  $\mathbf{H} = \hat{\mathbf{H}}$ ). The first search direction is given by  $\mathbf{a}^1 = \boldsymbol{\gamma}$ , the direction of steepest ascent. All subsequent search directions are given by  $\mathbf{a}^{l+1} = \mathbf{H}\mathbf{a}^l$  (for  $l = 2, \dots$ ), which means that  $\mathbf{a}^{l+1}$  is based on the rate of change of  $\mathbf{a}^l$ . In order to avoid linear dependence between different  $\mathbf{a}^l$ , SVD is used to produce an orthonormal basis. For large  $n$ , the set of  $\mathbf{a}^l$  may not completely span all  $n$  dimensions. In such cases, SVD can identify the unnecessary basis vectors so that they can be removed from  $\mathbf{A}$ . In all subsequent calculations, we use a subspace dimension of 10, but allow our SVD algorithm to dynamically remove unnecessary basis vectors at each iteration. Therefore, in practice, the subspace dimension tends to vary between 7 and 10 from iteration to iteration.

Once the iterative non-linear process converges, or produces a model which adequately satisfies the data, it is then fed back into the objective function (eq. 4) and the process is repeated starting from the same initial model. If  $N$  solutions are required, then the iterative non-linear scheme is applied  $N$  times. Rather than initiate each new search for a data satisfying model from the same initial model, another possibility is to begin at the point in parameter space occupied by the most recent solution.

At this point, it is worth noting the regularization assumptions that are inherent to the subspace inversion scheme. The subspace in which the function minimization takes place is spanned by a family of orthogonal vectors that are defined in terms of the gradient vector in model space  $\boldsymbol{\gamma}$  and its rate of change. This means that any model parameter  $m_L$  for which  $\partial \mathbf{g} / \partial m_L = \mathbf{0}$  (i.e. all traveltimes are unaffected by changes in  $m_L$ ) will not vary from the initial model to the first solution model (i.e. only those regions of a model with adequate ray path coverage will be perturbed during the inversion process). Subsequent inversions with the evolution term in place will also not vary these parameters as its contribution to the corresponding elements of the gradient vector (see eq. 6) is zero, that is, the objective function defines a local maximum in these dimensions of parameter space. In order to generate a family of distinct solutions, it is generally undesirable to have any parameters of the starting model located near a local maximum of the evolution term. Here,

the problem is overcome by randomly perturbing the gradient of the evolution term if the difference between model parameters from successive generations falls below a given tolerance. In effect, this gives a ‘kick start’ to those parameters stuck near a local maximum of  $S(\mathbf{m})$ . Although in one sense this added step is undesirable in that it introduces additional free parameters to the problem, it turns out that a wide range of values for the model tolerance and standard deviation of the random perturbation can be chosen without the technique breaking down.

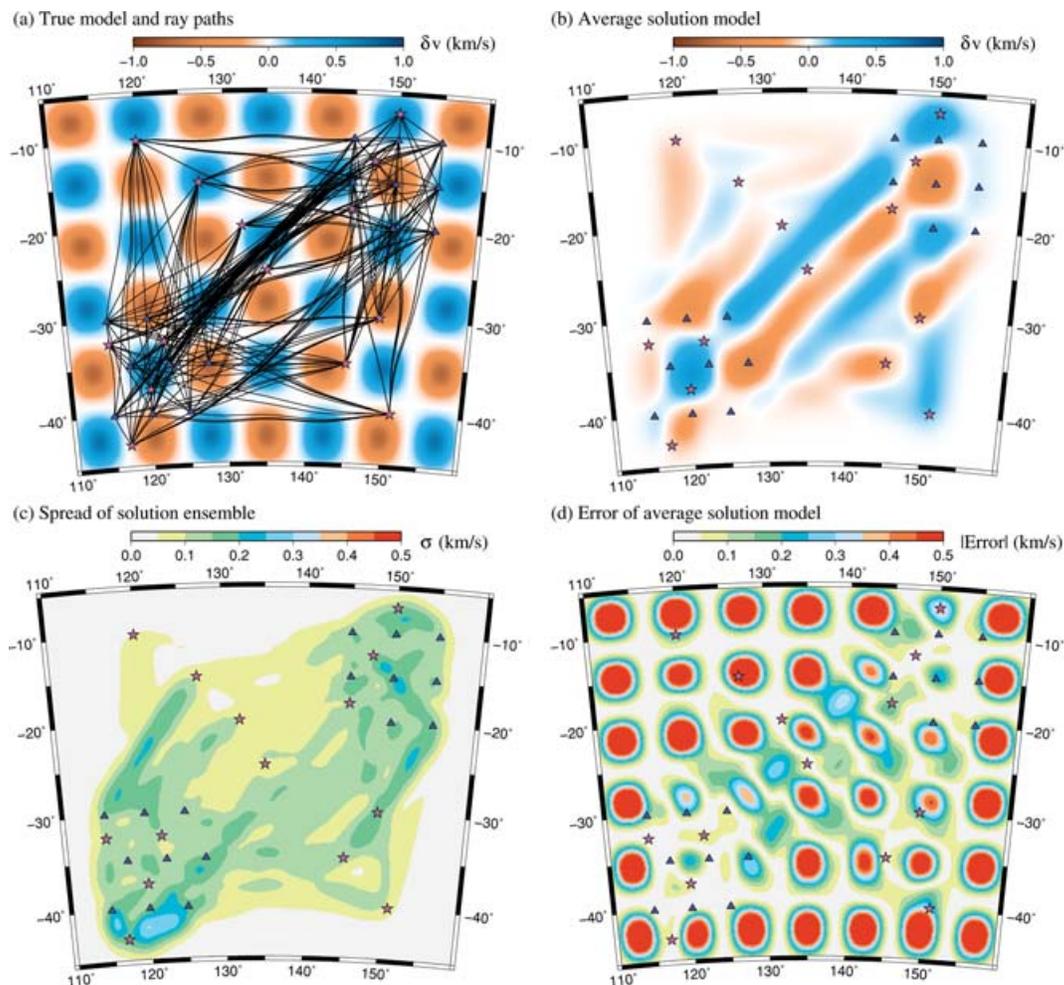
### 3 RESULTS

Two tomography examples are presented below. Synthetic tests that feature a highly variable path distribution are first performed to examine the validity of the new technique. This is followed by an application that exploits high frequency Rayleigh wave empirical Green’s functions extracted from ambient noise data in order to image the Australian crust. In both cases, structure is represented by smooth variations in wave speed on a spherical shell. A mosaic of cubic B-spline functions (see Rawlinson & Sambridge 2003, for more details) defined by a grid of control nodes is used to describe the velocity continuum; the values of these control nodes constitute the unknowns in the inversion. The use of cubic B-splines in the velocity parametrization imposes implicit smoothing regularization on the inverse problem, as it helps to suppress sharp variations in the shape of the misfit surface.

#### 3.1 Synthetic tests

A synthetic data set is constructed by using FMM to compute traveltimes for seismic energy propagating across a spherical shell between 14 sources and 17 receivers in the presence of velocity heterogeneity. This scenario, therefore, mimics the propagation of surface waves between earthquake sources and receivers in the Earth. The model was generated by specifying a grid of  $26 \times 30$  control points in latitude and longitude, respectively (resulting in a total of 780 unknowns), setting them to a constant value of  $5.0 \text{ km s}^{-1}$ , and then superimposing a checkerboard pattern with a maximum amplitude of  $0.8 \text{ km s}^{-1}$ . The resultant smooth velocity model, sources, receivers and associated ray paths are shown in Fig. 4(a). Clearly, the chosen configuration of sources and receivers produces a highly heterogeneous data coverage, and ray paths bend significantly in response to wave speed variations. Overall, this scenario represents a good test for the new scheme, as the resolving power of the data set varies considerably across the model, and is not always correlated with path density. For instance, in the region between the two clusters of sources and receivers (longitude  $130^\circ$  to  $140^\circ$ , latitude  $-20^\circ$  to  $-30^\circ$ ), there is dense path coverage, but little change in azimuth between the adjacent rays. To simulate realistic levels of noise in the data, Gaussian noise with a standard deviation of 2.0 s is added to the synthetic traveltimes.

Prior to testing the new scheme, a reference inversion is carried out using a standard bootstrap analysis (e.g. Efron & Tibshirani 1993). The bootstrap is a method of statistical inference that exploits the power of modern computers to produce estimates of probabilistic phenomena (e.g. variance, bias, coverage) from large and complex data sets without the need for explicit formulae (unlike formal measures of covariance and resolution). It works by performing repeat inversions with a re-sampled data set and examining the characteristics of the model ensemble that is produced. Here, the data set used to generate each new model is obtained by randomly sampling the



**Figure 4.** (a) Velocity model, sources (red stars), receivers (blue triangles) and ray paths associated with the synthetic data set; (b) average model produced from 50 iterations of the bootstrap technique; (c) spread of the solution ensemble as defined by one standard deviation of the distribution and (d) absolute value of the error associated with the average solution model.

pool of synthetic traveltimes with replacement. A single traveltimes can be used more than once, and the size of the new data set equals that of the original. In seismic tomography, the mean of the solution ensemble and its standard deviation are the most commonly used pieces of summary information (e.g. Calvert *et al.* 2000; Gung & Romanowicz 2004). Although it was pointed out earlier that the appropriateness of bootstrapping for under or mix-determined inverse problems is questionable, it is one of the few techniques available that allows ensemble analysis to be carried out with large seismic data sets.

Figs 4(b)–(d) show the result of a 50 iteration bootstrap scheme applied to the synthetic data set shown in Fig. 4(a). The starting model used to produce each solution comprises a simple medium with a constant velocity of  $5.0 \text{ km s}^{-1}$  (i.e. the background model). The pattern of recovered anomalies exhibited by the average solution model (Fig. 4b) closely matches that of the true model (Fig. 4a) in regions of good angular path coverage. In the corridor between the two receiver clusters, significant smearing is produced by the large number of paths with similar azimuths. The standard deviation of the solution ensemble (Fig. 4c) shows that much of the model variability occurs in regions with the highest concentrations of sources and receivers, which is opposite to the actual error trend seen in Fig. 4(d). It is important to note that the solution spread (Fig. 4c)

is not expected to emulate the actual error (Fig. 4d); the checkerboard pattern evident in the latter case is due largely to the choice of initial model, which exactly matches the true model along the zero lines of the checkerboard. However, the solution spread, which is principally a function of data coverage and uncertainty, should provide an indication of the maximum variability of each parameter as permitted by the data constraints.

The potentially misleading result of the bootstrap method stems from the underdetermined nature of the inverse problem, which requires regularization in order to generate a solution. Here, regularization is imposed implicitly using the subspace scheme, which causes the solution model to remain largely unperturbed from its original state in regions of poor data coverage. Gradient based inversion techniques such as the subspace method will not adjust parameters that have zero Fréchet derivatives. Therefore, Fig. 4(c) only has meaning in regions of good path coverage, which can be found principally in the top right and bottom left sectors of the image. Although underestimated, the pattern of model variability in these locations appears reasonable—for instance, the largest values of  $\sigma$ , which occur at  $120^\circ\text{E}$ ,  $42^\circ\text{S}$ , coincide with a small region of poor azimuthal ray coverage. On the other hand, the zone between the two receiver clusters, where many rays bundle together (Fig. 4a), does not cause significant variability in the solution

ensemble. The explanation for this phenomenon reveals a weakness in the bootstrap method in the presence of inhomogeneous data coverage. The number of rays that get bundled together define approximately half the total number of traveltimes in the complete data set, so random selections from the pool of traveltimes will invariably include a significant number of these paths. Consequently, the structure in this region will always be the same smeared out anomaly.

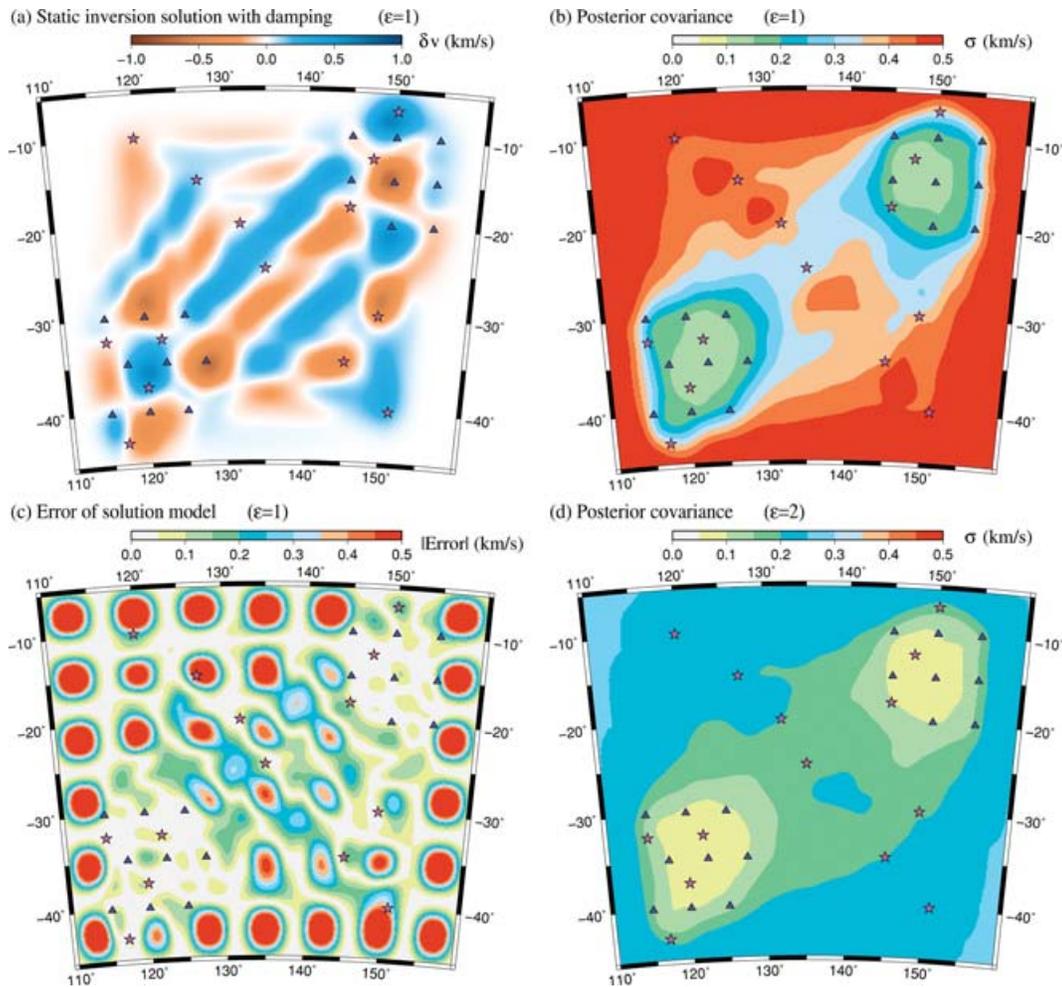
Although the bootstrap comparison is instructive, many studies that deal with large tomographic problems opt for either synthetic reconstruction tests or formal estimates of posterior covariance and resolution in order to assess solution non-uniqueness. In the latter case, local linearity is usually assumed to make the problem tractable. For eq. (1) with  $\eta = 0$ , the posterior model covariance matrix can be expressed (Menke 1989):

$$\mathbf{C}_M = (\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1})^{-1}. \quad (11)$$

As pointed out by Rawlinson & Sambridge (2003), posterior covariance and resolution are really only meaningful in a Bayesian context, where the error statistics associated with the *a priori* model are well known, thus allowing the data and prior constraints to be combined to yield a more accurate posterior model distribution. In such circumstances  $\epsilon = 1$ , and  $\mathbf{C}_m$  represents the true *a priori* model covariance. Fig. 5(a) shows the solution model produced using a static inversion

(i.e. no model feedback) with  $\epsilon = 1$  and the diagonal elements of  $\mathbf{C}_m$  set to  $(0.5)^2$ ; thus, uncertainties in the initial model are assumed to be uncorrelated and have a uniform value of  $0.5 \text{ km s}^{-1}$ . Six iterations of a 10-D subspace inversion scheme are applied to obtain the solution. Fig. 5(b) is the associated posterior model covariance obtained by solving eq. (6) using LU decomposition with  $\mathbf{G}$  computed at the solution point. Regions of low uncertainty generally match areas of good pattern recovery, where the error in the solution model is low (Fig. 5c). High path density does not necessarily yield low posterior covariance, as shown in the corridor between the two receiver arrays in Fig. 5(b) (cf. Fig. 4a). Overall, this result is promising, and clearly superior to the bootstrap test results, which is why posterior covariance estimates remain popular in seismic tomography (Rawlinson & Sambridge 2003).

The need for accurate *a priori* information in the computation of  $\mathbf{C}_M$  is starkly revealed in Fig. 5(d), which is the same as Fig. 5(b), except that  $\epsilon = 2$ , which is equivalent to setting the *a priori* uncertainty to  $0.25 \text{ km s}^{-1}$ . Since the posterior uncertainty can never exceed the prior uncertainty,  $\sigma$  in Fig. 5(d) is approximately half that of  $\sigma$  in Fig. 5(b). In cases where good *a priori* information is unavailable (which is often the case), the absolute values of  $\mathbf{C}_M$ , therefore, have little meaning. If  $\epsilon = 0$ , which corresponds to no *a priori* constraints, then  $\mathbf{C}_M$  cannot be computed for underdetermined problems since  $\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}$  will be a singular matrix.

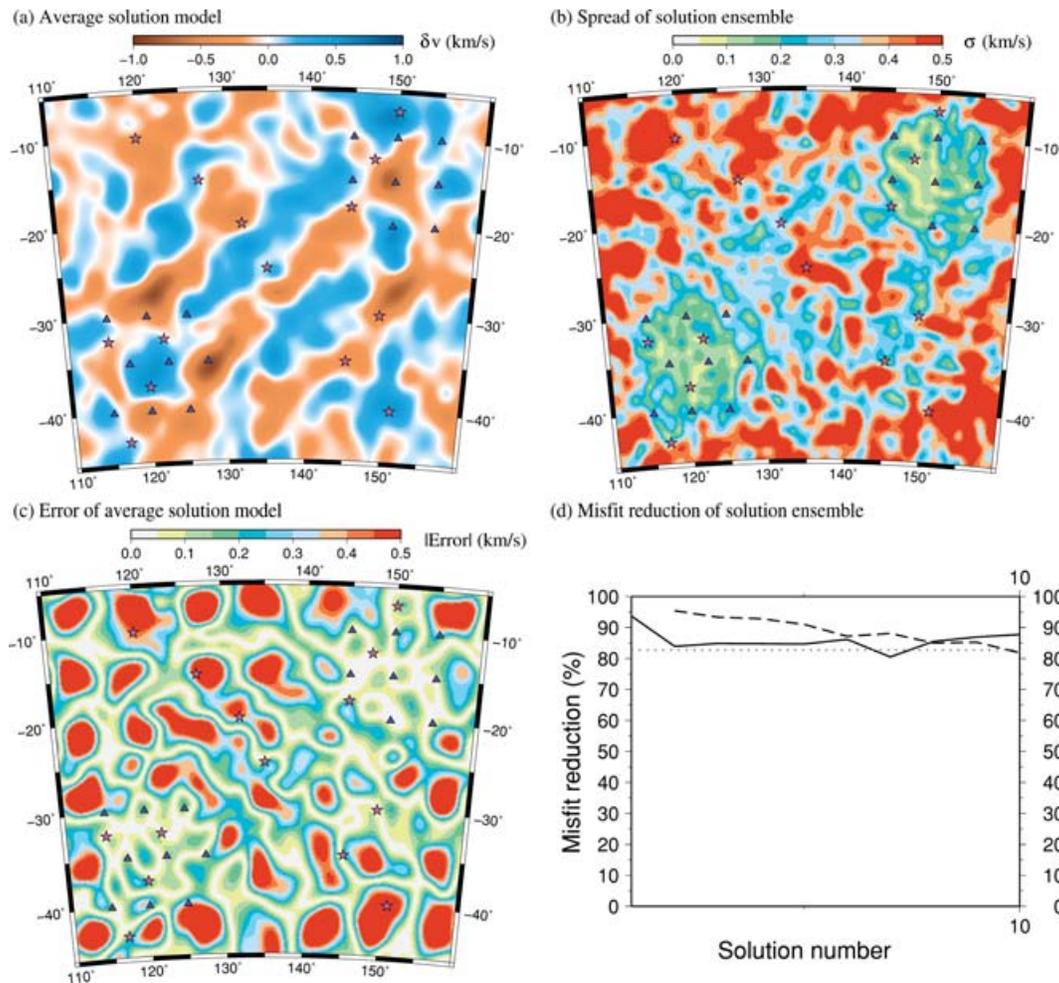


**Figure 5.** (a) Velocity model obtained using a static inversion with a damping value of  $\epsilon = 1$ ; (b) estimate of posterior covariance with  $\epsilon = 1$ ; (c) absolute value of the error associated with the solution model shown in (a) and (d) estimate of posterior covariance with  $\epsilon = 2$ .

Following on from the bootstrap and posterior covariance results, we generate an ensemble of 10 solution models using the dynamic objective function minimization procedure. The free parameters  $\lambda$ ,  $p$  and  $\zeta$  are set to  $7 \times 10^{-5}$ , 1 and  $1 \times 10^{-4}$ , respectively. When each new model is generated, the starting model used on each occasion is the  $5.0 \text{ km s}^{-1}$  constant velocity background model. In this case,  $\zeta$  has little effect on the evolution term, so it is just set to a small number. Tests show that the optimum choice for  $\lambda$  produces solution models that satisfy the data, but are as different from one another as possible. As  $\lambda$  decreases from this threshold value (evolution term increases in magnitude), the solution models become more dissimilar, but no longer satisfy the data; as  $\lambda$  increases (evolution term decreases in magnitude), the solution models still satisfy the data, but become more similar. In practice, it turns out that relatively few trial and error iterations are usually required to find the optimum  $\lambda$  value. The parameter  $p$  tends to trade off with  $\lambda$  to some extent, and it was found that little appears to be gained from using  $p$  values other than unity.

The average model of the 10 member ensemble is shown in Fig. 6(a); the corresponding standard deviation is shown in Fig. 6(b); and the error between the average model (Fig. 6a) and the true model

(Fig. 4a) is shown in Fig. 6(c). In regions of good path coverage, the average solution model recovers the pattern of anomalies quite accurately (*cf.* Figs 6a and c), although there does appear to be some ‘noisiness’ in the final image. The spread of the solution model, as given by one standard deviation of the ensemble (Fig. 6b), clearly distinguishes between well constrained and poorly constrained regions of the model. In contrast to the unrealistic bootstrap error map (Fig. 4c), the error measure from the dynamic approach shows a consistent and more believable pattern—in particular, maximum variability occurs in areas of low or absent path coverage. Notably, the zone between the two receiver clusters in which rays bunch together (Fig. 4a) has a high standard deviation, correctly indicating that these paths poorly resolve structure. Regions in which path coverage is totally absent (e.g. near the edge of the model) have a finite standard deviation (Fig. 6b), the value of which is arbitrarily controlled by the choice of  $\lambda$ , and to a lesser extent  $p$  and  $\zeta$ . Strictly speaking, the uncertainty in these regions should be large, but in the absence of prior information, cannot be quantified. Therefore, the spread of the solution model in areas of no ray coverage should be viewed as a lower bounds on the possible range of error.



**Figure 6.** (a) Average solution model from 10 generations of the dynamic objective function scheme; (b) spread of solution ensemble as defined by one standard deviation of the distribution; (c) absolute value of the error associated with the average solution model and (d) two measures of misfit reduction for each solution in the ensemble relative to the starting model. The solid black line denotes the rms traveltimes misfit reduction and the dashed line denotes the misfit reduction of the evolution term in eq. (4). The horizontal dotted line corresponds to the rms traveltimes misfit reduction required to satisfy the data to the level of the imposed noise (i.e. one standard deviation of the Gaussian noise).

Fig. 6(d) plots the behaviour of the rms traveltimes residual and evolution term (see eq. 4) as a function of model number in the ensemble. There is little point in keeping track of the value of the objective function from one solution to the next due to presence of the evolution term, which changes after each new model is produced. All 10 solution models approximately satisfy the data to the level of the noise, and the behaviour of the evolution term indicates that each new solution model is much more distant in parameter space from all preceding generations than the initial model, which is a desirable result. Note that the misfit reduction of the evolution term is not defined for the first model produced. The almost monotonic decrease of the evolution term misfit reduction with solution generation is due to the increasing number of data satisfying solutions that populate parameter space, that is, it becomes harder to find points in parameter space that are distant from other solutions.

A final test is performed to investigate whether the properties of the solution ensemble change significantly if a larger number of dynamic iterations is used. Exactly the same scenario as Fig. 6 is adopted, but now the dynamic objective function procedure is allowed to run for 50 iterations. The results are summarized in Fig. 7. Compared to Fig. 6(a), the average solution model (Fig. 7a) differs mainly in that the amplitude of anomalies is much smaller in regions of poor data coverage; this is reflected in the checkerboard error pattern which has become more coherent (cf. Figs 6c and 5c). Another desirable feature is that the spread of solution models has

become more bimodal (cf. Figs 7b and 6b): in the neighbourhood of the two source–receiver clusters,  $\sigma$  is small, and elsewhere,  $\sigma$  is much more uniformly large. This suggests that the 10 models used to generate Fig. 6 was arguably an insufficient population from which to infer reliable statistics. The progressive decrease of the evolution term misfit reduction with iteration (Fig. 7d) is expected due to the increasing population of distinct solution models. In most cases, the rms traveltimes misfit of the solution models approximately satisfies the data, which is a desirable outcome.

### 3.2 Application to ambient noise data from Australia

Over the past 15 years, the passive broad-band station coverage of Australia has gradually improved, with a cumulative total of over 200 instruments to date (see Kennett 2006, for a recent overview). The availability of these data permit a wide variety of studies to be undertaken, including surface wave tomography (Debayle & Kennett 2000; Fishwick *et al.* 2005), body wave tomography (Kennett 2003), receiver function analysis (Clitheroe *et al.* 2000; Reading & Kennett 2003) and shear wave splitting (Heintz & Kennett 2005). The emergence of virtual source seismology over the last few years (Campillo & Paul 2003; Shapiro & Campillo 2004) has given rise to a new class of seismic tomography that exploits the long-term cross-correlations of ambient noise data between station pairs (Shapiro *et al.* 2005; Kang & Shin 2006; Yang *et al.* 2007).

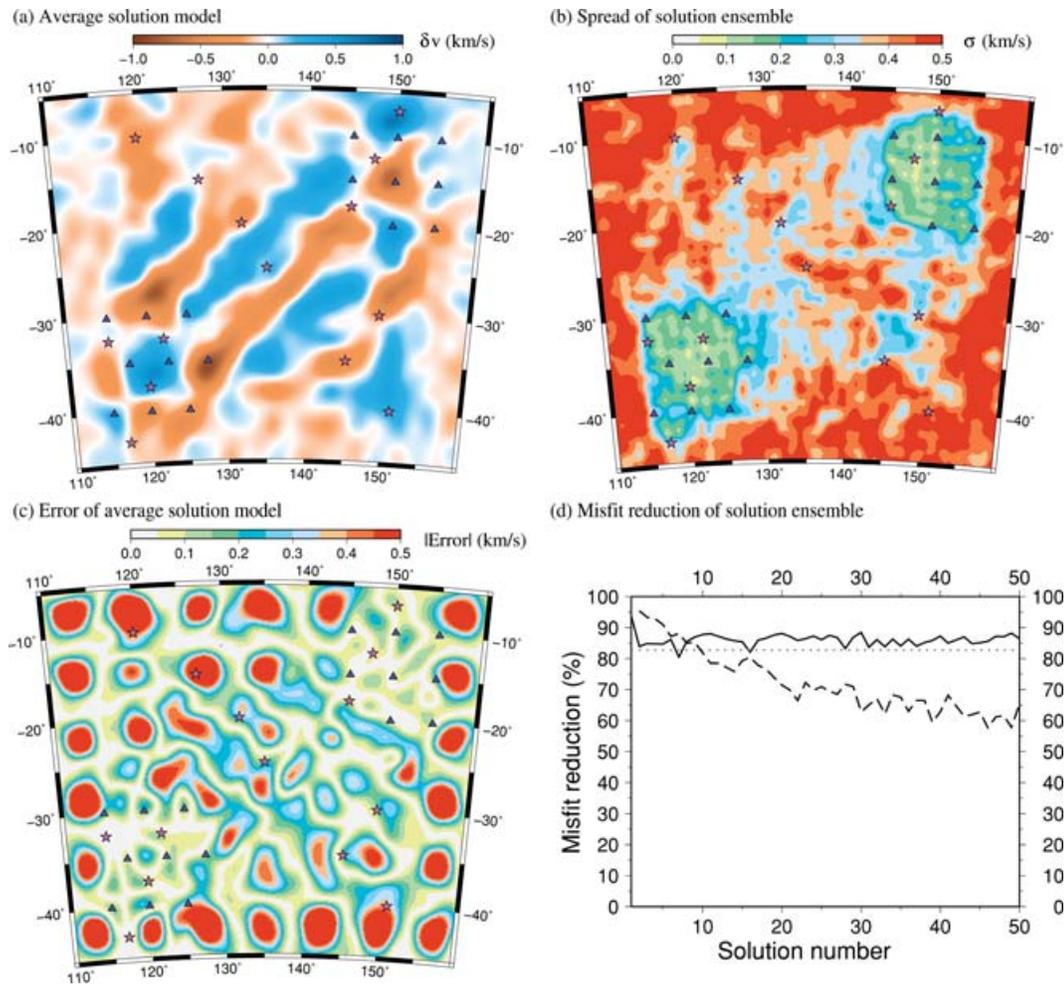
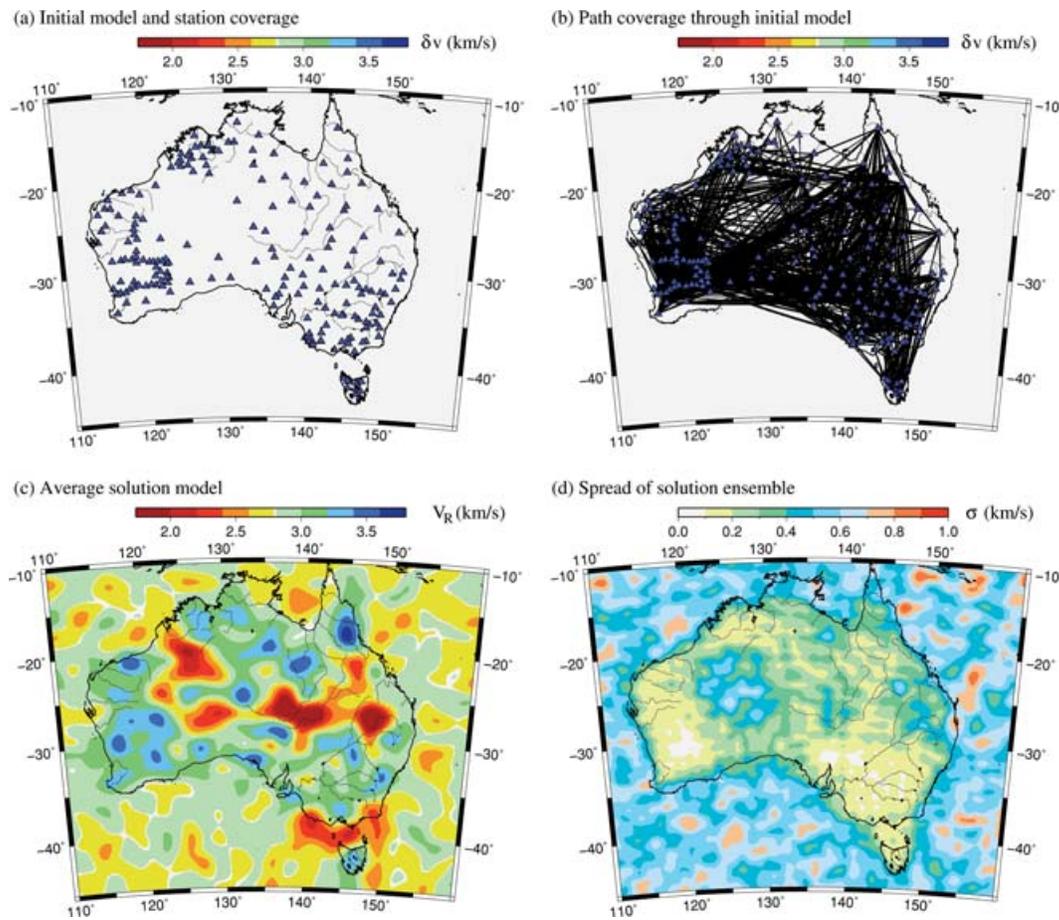


Figure 7. Same as Fig. 6, but now using 50 iterations of the dynamic objective function procedure.



**Figure 8.** (a) Stations used in the cross-correlation of ambient noise data; (b) path coverage through the initial model; (c) average solution model computed from an ensemble of 25.  $V_R$  denotes Rayleigh wave group velocity and (d) variation of the model ensemble as represented by one standard deviation of the distribution.

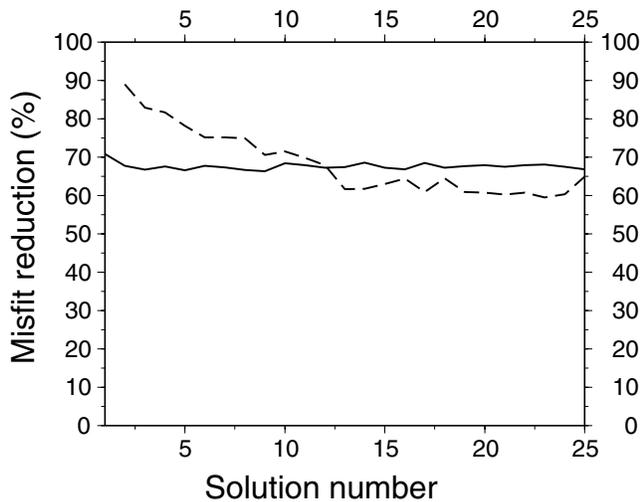
In a recent study (Saygin 2007), high frequency Rayleigh wave empirical Green's functions were extracted from the records of 208 seismometers covering the Australian region using a cross-correlation procedure. The traveltimes of these coherent energy packets were subsequently inverted to produce tomographic images of the mid-shallow crust beneath Australia. In this example, traveltimes associated with 6.7 s period Rayleigh waves are used to investigate the potential benefits of the new dynamic objective function scheme when applied in an observational context.

The location of the 208 broad-band recorders is shown in Fig. 8(a); note that traveltimes do not exist for every station pair, as sequential array deployments were used to achieve the cumulative coverage. The total number of traveltimes used to constrain structure is, therefore, 2332. The initial model is described by a grid of  $30 \times 30$  control points in latitude and longitude, with the velocity at each node set to  $2.8 \text{ km s}^{-1}$ . Fig. 8(b) shows the path coverage through the initial model; clearly some regions will be better constrained by the data than others. The free parameters  $\lambda$ ,  $p$  and  $\zeta$  used in the dynamic objective function to construct the solution ensemble are set to  $5 \times 10^{-6}$ , 1 and  $5 \times 10^{-6}$ , respectively. Here, a total of 25 solutions are generated—increasing this number has only a negligible effect on the average solution model that is produced (Fig. 8c). The variation of the solution ensemble as a function of geographical location (Fig. 8d) suggests that the average solution model is generally well constrained within the Australian mainland, which is consistent with the path coverage (Fig. 8b). However, there

are regions, particularly in western Australia, where the data constraints appear to be relatively poor. Although difficult to discern in Fig. 8(b), these areas tend to have lower path density and/or poorer azimuthal path coverage.

The behaviour of the data and evolution terms during the model generation process is summarized in Fig. 9. On average, the traveltime misfit reduction exhibited in this example is less than the synthetic test examples, but this is probably largely due to the greater level of noise present in the observations. Unfortunately, no quantitative estimates of the data uncertainty are available for this data set, although the final rms misfit of around 10 s for all of the solution models implies that it may be in the vicinity of this value (assuming data noise is the principle contributor to the misfit). As in the synthetic test examples, the evolution term misfit reduction tends to decrease monotonically as more solution models are produced. This is an almost inevitable consequence, as it only takes two solutions in the same proximity to increase the size of the evolution term substantially.

In order to help validate the results produced in this test, a conventional checkerboard resolution analysis is performed. Traveltimes are computed in the presence of an alternating pattern of fast and slow anomalies using the same source and receiver configuration as the observational data set. The synthetic traveltimes are then inverted in an attempt to recover the pattern. Gaussian noise with a standard deviation of 5 s is added to the synthetic traveltimes to simulate the noise content of the observations. Fig. 10(a) shows the



**Figure 9.** Two measures of misfit reduction for each solution in the ensemble (see Fig. 8) relative to the starting model. The solid black line denotes the rms traveltime misfit reduction and the dashed line denotes the misfit reduction of the evolution term in eq. (4).

input checkerboard pattern used to create the data set, and Fig. 10(b) shows the corresponding output pattern generated by the inversion.

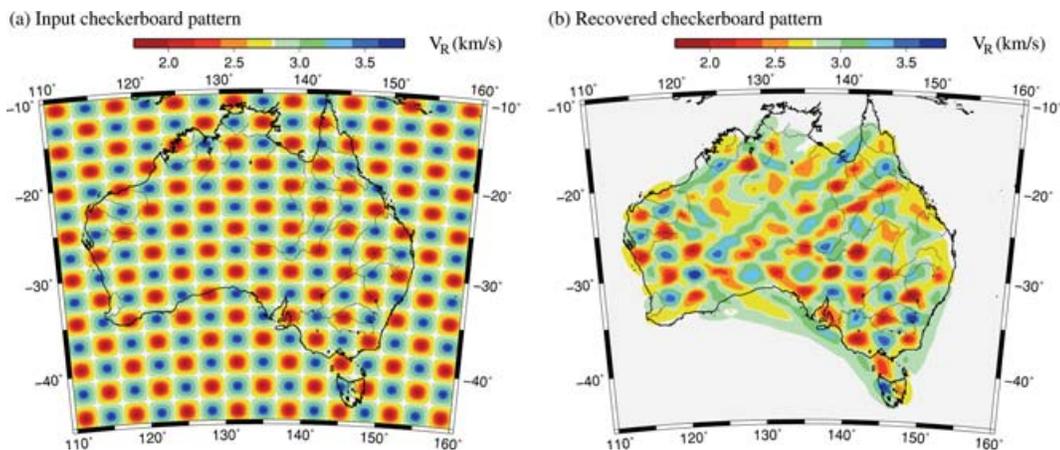
Although frequently used to analyse the robustness of tomographic images, checkerboard resolution tests are not a universally acclaimed technique; for example, Lévêque *et al.* (1993) point out their propensity to retrieve small scale patterns of alternating wave speed, even though larger scale structure may not be well resolved by the data. Also, if the non-linearity of the inverse problem is accounted for, checkerboard tests have an added model dependence, which could be significant in the presence of strong wave speed heterogeneity such as that exhibited in Fig. 8(c). Despite these recognized limitations, the regions of good pattern recovery (Fig. 10b) seem to correspond to regions of low model ensemble variability (Fig. 8d). Likewise, where smearing of the anomalies is evident, such as in eastern Queensland and Western Australia, the models in the ensemble become less similar. Overall, there does not appear to be any major discrepancy between the results of the two techniques, which is reassuring.

#### 4 DISCUSSION AND CONCLUSIONS

In this study, a dynamic objective function approach to generating multiple solution models in seismic tomography is formulated and applied to both synthetic and field data. The aim of the technique is to modify the topography of the objective function surface based on information provided by previously computed solutions. This is accomplished via the introduction of an evolution or feedback term in the objective function, which creates a local maximum in parameter space at all points occupied by previous generations of solution model. Consequently, any new model that is generated is penalized if it becomes too close to these regions of parameter space. Provided the objective function remains reasonably well behaved, gradient based minimization procedures can be used to generate an ensemble of solutions. Clearly, this is a potential advantage, as it means that large inverse problems can be addressed.

A series of synthetic tests (Figs 4–7) using a data set featuring highly variable path coverage shows that the new method is capable of producing an ensemble of different data-satisfying models that yield meaningful information on both seismic structure and the degree to which it is constrained by the data. Simply taking the average of the ensemble produces a model which generally replicates the true wave speed patterns in regions of good angular path coverage. Mapping the variation of the ensemble, as measured by the standard deviation of all solutions, correctly shows which regions are reliably recovered. Bootstrapping, a standard numerical technique for statistical inference, is used to produce a family of models by repeat random sampling and inversion of the original data set with replacement. The reliance of this technique on over-determined inverse problems is clearly exposed in this case as the standard deviation of the model ensemble poorly reflects the actual error associated with the average model. In particular, low model variability in regions of sparse data coverage and ray bundling (high ray density but narrow angular distribution), stands in stark contrast to the dynamic objective function results.

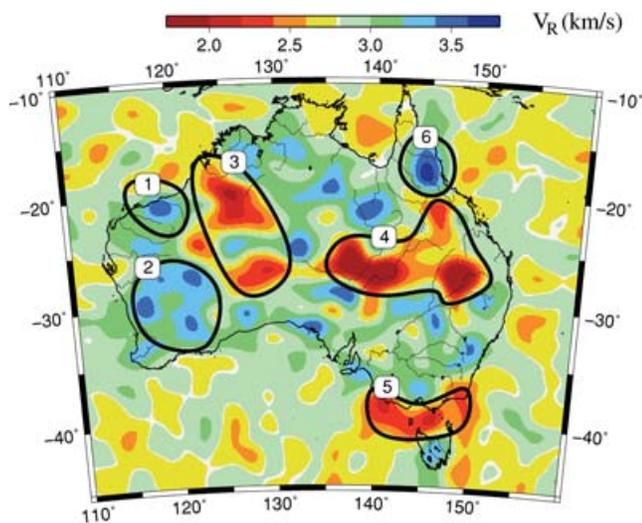
Bootstrapping and dynamic regularization share some similarities, namely that they can be applied to large non-linear inverse problems, produce an ensemble of models that can be interrogated for common features, and use stochastic processes to help drive the search for new models. In the case of bootstrapping, each solution is produced by inverting a data set obtained by random selection from a pool of values. The dynamic objective function technique



**Figure 10.** Checkerboard resolution test results for the Australian ambient noise data set. The associated path coverage is shown in Fig. 7(b). Note that the implicit regularization imposed by the subspace inversion scheme means that the solution model does not deviate from the starting model ( $2.8 \text{ km s}^{-1}$ ) in regions of no path coverage.

randomly perturbs the gradient of the evolution term in order to prevent model parameters from becoming stuck near local maxima of the misfit surface from one solution generation to the next. This helps to ensure that poorly constrained regions exhibit large ensemble variability. An additional benefit of the new technique compared to bootstrapping is reduced computation time. This occurs due to the need for fewer solution models in order to obtain a stable estimate of model uncertainty. For very large data sets, the need to solve the full inverse problem for each model produced (and at least 10 are required) may mean that dynamic regularization is computationally prohibitive, but for most forms of tomography, this is unlikely to be the case.

Application of the dynamic objective function technique to surface wave traveltimes derived from long-term cross-correlations of the ambient noise data collected in Australia reveals an intriguing pattern of wave speed variation (see Fig. 8c) that can be correlated with the surface geology. Fig. 11 shows the average solution model plotted as in Fig. 8(c), but with several of the outstanding features highlighted. The zones of elevated wave speed in western Australia correspond very closely with the surface expressions of the Pilbara (region 1) and Yilgarn (region 2) cratons (Betts *et al.* 2002), which represent fragments of ancient Archean lithosphere. The low velocity anomalies that characterize regions 3, 4 and 5 have a remarkable correspondence to regions of thick sedimentary cover and/or elevated heat flow (Chopra & Holgate 2005). In the case of region 3, the north to south pattern of slow-fast-slow anomalies correlates closely with the presence of the Officer Basin, Musgrave Block (preserved Proterozoic orogen) and Amadeus Basin. However, it should be noted that this sector of the model is not very well constrained by the data set (Fig. 8d). Region 4 approximately overlies the Cooper-Eromanga Basin, and Region 5 overlies the Gippsland, Bass and Otway Basins. Finally, the zone of elevated wave speed in northern Queensland (region 6) coincides with the location of the Proterozoic Georgetown inlier, which contrasts with the younger Phanerozoic terrane that characterizes much of the Tasmanides further south (Betts *et al.* 2002). Although our analysis of the geological implications of Fig. 11 is rather limited, the aim here is to argue that the average solution model produced by the new technique stands up to scrutiny against *a priori* information. For further details on the interpretation, refer to Saygin (2007).



**Figure 11.** Same as Fig. 8(c), but with several features of interest highlighted.

The variation between models in the ambient noise solution ensemble, as summarized in Fig. 8(d), appears to indicate that the average solution model is optimally constrained in southeast Australia and the western half of Western Australia. A synthetic checkerboard resolution test (Fig. 10) using the same source and receiver geometry appears to support this result. Although it is difficult to directly compare robustness estimates produced by these two methods, there does not appear to be any major discrepancies between the maps in Figs 8(d) and 10(b). This is a reassuring result, particularly considering the amplitude of the velocity heterogeneity that is recovered ( $\pm 30$  per cent).

The dynamic objective function scheme proposed in this paper is not free from ‘tuning’ parameters. The most obvious are  $\lambda$ ,  $\rho$  and  $\zeta$  (see eq. 4), which control the shape of the local maximum inserted into the misfit landscape. Indeed, the function used to define the evolution term itself was chosen arbitrarily; the only requirement was that it is a maximum at a given point in model space, and decays to zero away from that point. Many other penalty functions could be used including quadratic, exponential and logcosh (Pierro 1995), although in the case of quadratic functions, the global influence on the misfit surface is undesirable (see Section 2.1). However, any function that exhibits a central maximum and decays to zero will contain free parameters that control its shape. Ideally, if one could use the data to help constrain these parameters in the inversion rather than choose them via trial and error, the scheme would be much more appealing. Ultimately, we have only just started to explore the possibilities of this dynamic regularization technique, and it may be that other ways of modifying the misfit landscape in response to the generation of successive solutions will turn out to be superior. On the other hand, we have clearly shown that our approach can work effectively, and sensible choices for the free parameters can be made by taking into account the behaviour of the data misfit and model evolution term. While it will never be able to explore parameter space as extensively or thoroughly as Monte Carlo direct search methods, the clear advantage of the gradient-based technique is that it can deal with very large tomographic problems, and does not need to produce a highly populated solution ensemble in order to make meaningful inferences.

In conclusion, the results of this study suggest that the new gradient based technique is capable of revealing robustly constrained information contained in noisy and inhomogeneous data sets, and offers a viable alternative to the various classes of synthetic resolution tests and linear error analyses that are currently in use. The modifications required to implement the method with pre-existing gradient based codes should be minimal, which is an added incentive for its use. Although the scheme advocated in this paper has shown itself to be practicable, it may well be improved by further work into the choice of evolution term and starting model used to generate each new solution in the ensemble.

## ACKNOWLEDGMENTS

This work was supported by Australian Research Council Discovery Project DP0556282. Luis Rivera and an anonymous reviewer are thanked for their constructive comments, which greatly improved the paper.

## REFERENCES

- Aki, K., Christofferson, A. & Husebye, E.S., 1977. Determination of the three-dimensional seismic structure of the lithosphere, *J. geophys. Res.*, **82**, 277–296.

- Backus, G. & Gilbert, F., 1970. Uniqueness in the inversion of inaccurate gross Earth data, *Philos. Trans. R. Soc.*, **266**, 123–192.
- Bai, C.-Y. & Greenhalgh, S., 2005. 3-D non-linear travel-time tomography: imaging high contrast velocity anomalies, *Pure appl. Geophys.*, **162**, 2029–2049.
- Benz, H.M., Zandt, G. & Oppenheimer, D.H., 1992. Lithospheric structure of northern California from teleseismic images of the upper mantle, *J. geophys. Res.*, **97**, 4791–4807.
- Betts, P.G., Giles, D., Lister, G.S. & Frick, L., 2002. Evolution of the Australian lithosphere, *Aust. J. Earth Sci.*, **49**, 661–695.
- Bijwaard, H. & Spakman, W., 2000. Non-linear global *P*-wave tomography by iterated linearized inversion, *Geophys. J. Int.*, **141**, 71–82.
- Calvert, A. *et al.*, 2000. Geodynamic evolution of the lithosphere and upper mantle beneath the Alboran region of the western Mediterranean: constraints from travel time tomography, *J. geophys. Res.*, **105**, 10 871–10 898.
- Campillo, M. & Paul, A., 2003. Long-range correlations in the diffuse seismic coda, *Science*, **299**, 547–549.
- Chopra, P. & Holgate, F., 2005. A GIS analysis of temperature in the Australian crust, in *Proceedings of the World Geothermal Congress, 2005*, Antalya, Turkey.
- Clitheroe, G., Gudmundsson, O. & Kennett, B.L.N., 2000. The crustal thickness of Australia, *J. geophys. Res.*, **105**, 13 697–13 713.
- 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, *Geophys. J. Int.*, **146**, 31–52.
- de Kool, M., Rawlinson, N. & Sambridge, M., 2006. A practical grid based method for tracking multiple refraction and reflection phases in 3D heterogeneous media, *Geophys. J. Int.*, **167**, 253–270.
- Deal, M.M. & Nolet, G., 1996. Nullspace shuttles, *Geophys. J. Int.*, **124**, 372–380.
- Debayle, E. & Kennett, B.L.N., 2000. The Australian continental upper mantle: structure and deformation inferred from surface waves, *J. geophys. Res.*, **105**, 25 423–25 450.
- Debayle, E. & Sambridge, M., 2004. Inversion of massive surface wave data sets: model construction and resolution assessment, *J. geophys. Res.*, **109**, B02316, doi:10.1029/2003JB002652.
- Efron, B. & Tibshirani, 1993. *An Introduction to the Bootstrap*, Chapman and Hall/CRC, New York.
- Farra, V. & Madariaga, R., 1988. Non-linear reflection tomography, *Geophys. J.*, **95**, 135–147.
- Fishwick, S., Kennett, B.L.N. & Reading, A.M., 2005. Contrasts in lithospheric structure within the Australian craton—insights from surface wave tomography, *Earth planet. Sci. Lett.*, **231**, 163–176.
- Glahn, A. & Granet, M., 1993. Southern Rhine Graben: small-wavelength tomographic study and implications for the dynamic evolution of the graben, *Geophys. J. Int.*, **113**, 399–418.
- Gorbatov, A. & Kennett, B.L.N., 2003. Joint bulk-sound and shear tomography for Western Pacific subduction zones, *Earth planet. Sci. Lett.*, **210**, 527–543.
- 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, *J. geophys. Res.*, **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, *Geophys. J. Int.*, **149**, 249–266.
- Gung, Y. & Romanowicz, B., 2004. *Q* tomography of the upper mantle using three-component long-period waveforms, *Geophys. J. Int.*, **157**, 813–830.
- Heintz, M. & Kennett, B.L.N., 2005. Continental scale shear-wave splitting analysis: investigation of seismic anisotropy underneath the Australian continent, *Earth planet. Sci. Lett.*, **236**, 106–119.
- Iyer, H. & Hirahara, K., 1993. *Seismic Tomography: Theory and Practice*, Chapman & Hall, London.
- Kang, T.-S. & Shin, J.S., 2006. Surface-wave tomography from ambient seismic noise of accelerograph networks in southern Korea, *Geophys. Res. Lett.*, **33**, L17303, doi:10.1029/2006GL027044.
- Kennett, B.L.N., 2003. Seismic structure in the mantle beneath Australia, in *Evolution and Dynamics of the Australian Plate*, Vol. 22: Special Publication, pp. 7–23, eds Hillis, R.R. & Müller, R.D., Geological Society of Australia.
- Kennett, B.L.N., 2006. Developments in passive seismic techniques through the ANSIR National Research Facility, *Explor. Geophys.*, **37**, 278–285.
- Kennett, B.L.N., Sambridge, M.S. & Williamson, P.R., 1988. Subspace methods for large scale inverse problems involving multiple parameter classes, *Geophys. J.*, **94**, 237–247.
- Koper, K.D., Wysession, M.E. & Wiens, D.A., 1999. Multimodal function optimization with a niching genetic algorithm: a seismological example., *Bull. seism. Soc. Am.*, **89**, 978–988.
- Lees, J.M. & Crosson, R.S., 1989. Tomographic inversion for three-dimensional velocity structure at Mount St. Helens using earthquake data, *J. geophys. Res.*, **94**, 5716–5728.
- Lévesque, J.J., Rivera, L. & Wittlinger, G., 1993. On the use of the checkerboard test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**, 313–318.
- Lomax, A. & Snieder, R., 1994. Finding sets of acceptable solutions with a genetic algorithm with application to surface wave group dispersion in Europe, *Geophys. Res. Lett.*, **21**, 2617–2620.
- Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, New York.
- Mosegaard, K. & Sambridge, M., 2002. Monte Carlo analysis of inverse problems, *Inverse Problems*, **18**, R29–R54.
- Nolet, G., 1987. Waveform tomography, in *seismic Tomography: with Applications in Global Seismology and Exploration Geophysics*, pp. 301–322, ed. Nolet, G., D. Reidel, Dordrecht.
- Nolet, G., Montelli, R. & Virieux, J., 1999. Explicit, approximate expressions for the resolution and a posteriori covariance of massive tomographic systems, *Geophys. J. Int.*, **138**, 36–44.
- Pierro, A.R.D., 1995. A modified expectation maximization algorithm for penalized likelihood estimation in emission tomography, *IEEE Trans. Med. Imag.*, **14**, 132–137.
- Popovici, A.M. & Sethian, J.A., 2002. 3-D imaging using higher order fast marching traveltimes, *Geophysics*, **67**, 604–609.
- Rawlinson, N. & Sambridge, M., 2003. Seismic traveltime tomography of the crust and lithosphere, *Adv. Geophys.*, **46**, 81–198.
- Rawlinson, N. & Sambridge, M., 2004a. Wavefront evolution in strongly heterogeneous layered media using the fast marching method, *Geophys. J. Int.*, **156**, 631–647.
- Rawlinson, N. & Sambridge, M., 2004. Multiple reflection and transmission phases in complex layered media using a multistage fast marching method, *Geophysics*, **69**, 1338–1350.
- Rawlinson, N. & Urvoy, M., 2006. Simultaneous inversion of active and passive source datasets for 3-D seismic structure with application to Tasmania, *Geophys. Res. Lett.*, **33**, L24313, doi:10.1029/2006GL028105.
- 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, *Geophys. J. Int.*, **145**, 381–401.
- Rawlinson, N., Reading, A.M. & Kennett, B.L.N., 2006. Lithospheric structure of Tasmania from a novel form of teleseismic tomography, *J. geophys. Res.*, **111**, B02301, doi:10.1029/2005JB003803.
- Reading, A.M. & Kennett, B.L.N., 2003. Lithospheric structure of the Pilbara Craton, Capricorn Orogen and northern Yilgarn craton, Western Australia, from teleseismic receiver functions, *Aust. J. Earth Sci.*, **50**, 439–445.
- 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, *J. geophys. Res.*, **103**, 21 201–21 213.
- Sambridge, M., 1998. Exploring multidimensional landscapes without a map, *Inverse Problems*, **14**, 427–440.
- Sambridge, M., 1999a. Geophysical inversion with a Neighbourhood algorithm—I. Searching parameter space, *Geophys. J. Int.*, **138**, 479–494.
- Sambridge, M., 1999b. Geophysical inversion with a Neighbourhood algorithm—II. Appraising the ensemble, *Geophys. J. Int.*, **138**, 727–746.

- Sambridge, M. & Rawlinson, N., 2005. Seismic tomography with irregular meshes, in *Seismic Earth: Array Analysis of Broadband Seismograms*, Vol. 157, pp. 49–65, eds Levander, A. & Nolet, G., American Geophysical Union.
- Sambridge, M.S., 1990. Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D, *Geophys. J. Int.*, **102**, 653–677.
- Saygin, E., 2007. Seismic receiver and noise correlation based studies in Australia, *PhD thesis*, The Australian National University.
- Sethian, J.A., 1996. A fast marching level set method for monotonically advancing fronts, *Proc. Natl. Acad. Sci.*, **93**, 1591–1595.
- Shapiro, N.M. & Campillo, M., 2004. Emergence of broadband Rayleigh waves from correlations of the ambient seismic noise, *Geophys. Res. Lett.*, **31**, L07614, doi: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.
- Su, W.-J. & Dziewonski, A.M., 1997. Simultaneous inversion for 3-D variations in shear and bulk velocity in the mantle, *Phys. Earth planet. Inter.*, **100**, 135–156.
- Tarantola, A., 1987. *Inverse Problem Theory*, Elsevier, Amsterdam.
- Vasco, D.W., Peterson, J.E. & Majer, E.L., 1996. Nonuniqueness in traveltime tomography: ensemble inference and cluster analysis, *Geophysics*, **61**, 1209–1227.
- Walck, M.C., 1988. Three-dimensional  $V_p/V_s$  variations for the Coso region, California, *J. geophys. Res.*, **93**, 2047–2052.
- Wang, B. & Braile, L.W., 1996. Simultaneous inversion of reflection and refraction seismic data and application to field data from the northern Rio Grande rift, *Geophys. J. Int.*, **125**, 443–458.
- White, D.J., 1989. Two-dimensional seismic refraction tomography, *Geophys. J.*, **97**, 223–245.
- Williamson, P.R., 1990. Tomographic inversion in reflection seismology, *Geophys. J. Int.*, **100**, 255–274.
- Yang, Y., Ritzwoller, M.H., Levshin, A.L. & Shapiro, N.M., 2007. Ambient noise Rayleigh wave tomography across Europe, *Geophys. J. Int.*, **168**, 259–274.
- Yao, Z.S., Roberts, R.G. & Tryggvason, A., 1999. Calculating resolution and covariance matrices for seismic tomography with the LSQR method, *Geophys. J. Int.*, **138**, 886–894.
- Zelt, C.A., 1999. Modelling strategies and model assessment for wide-angle seismic traveltimes data, *Geophys. J. Int.*, **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, *J. geophys. Res.*, **103**, 7187–7210.
- Zelt, C.A. & Smith, R.B., 1992. Seismic traveltime inversion for 2-D crustal velocity structure, *Geophys. J. Int.*, **108**, 16–34.
- Zhang, H. & Thurber, C.H., 2007. Estimating the model resolution matrix for large seismic tomography problems based on Lanczos bidiagonalization with partial reorthogonalization, *Geophys. J. Int.*, **170**, 337–345.
- Zhang, J. & McMechan, G.A., 1995. Estimation of resolution and covariance for large matrix inversions, *Geophys. J. Int.*, **121**, 409–426.