Geophysical Journal International

Geophys. J. Int. (2010) **180,** 847–857 doi: 10.1111/j.1365-246X.2009.04452.x

Reducing errors in seismic tomography: combined inversion for sources and structure

Andrew P. Valentine and John H. Woodhouse

Department of Earth Sciences, University of Oxford, Oxford OX1 3PR, *UK. E-mail: andrew.valentine@earth.ox.ac.uk*

Accepted 2009 November 10. Received 2009 November 5; in original form 2009 September 14

SUMMARY

To perform seismic tomography, accurate determinations of event locations and focal mechanisms are required. These are usually obtained prior to tomographic inversion; often, little attention is paid to the earth model used for their determination. We show that an imprint of this model is found in the model recovered after tomographic inversion. To reduce this problem, it is important to recalculate earthquake source parameters as the inversion proceeds; synthetic tests suggest that this yields a better correspondence between recovered and true models. However, alternate source and structure inversions lead to a slow rate of convergence, and significant errors remain. We therefore propose combining source and structure inversion into a single inverse problem, and present an efficient algorithm to do this. We demonstrate that this reduces the number of iterations required to achieve a given accuracy; in our experiments, we observe a fourfold improvement on alternate inversions. We focus on full-waveform inversions for both source and structure, although the methods presented are general and should be applicable to other techniques.

Key words: Inverse theory; Earthquake source observations; Seismic tomography.

1 INTRODUCTION

Seismic tomography is a well-established tool for imaging the Earth's interior—a variety of methods and algorithms have been developed, for both regional and global studies (recent reviews include Romanowicz 2003; Thurber & Ritsema 2007). These all share the same basic approach: parameters describing an earth model are adjusted to obtain a better fit between recorded and synthetic seismic data. However, calculating synthetic data requires an estimate of the seismic source parameters, and obtaining this usually involves calculations within a particular earth model. By definition, this model cannot be correct—otherwise the tomographic inversion would be unnecessary—so the resulting source parameters must be assumed to be inaccurate. We are therefore updating our model to satisfy criteria that are known to be erroneous; effects arising from this appear to be rarely considered, and to date have often been assumed to be negligible.

We will demonstrate that systematic errors in source parameters can lead to significant errors in the model recovered after tomographic inversion, with a bias towards the model used for source determination. This situation can be improved by re-evaluating the source parameters during structure inversion, using the best available model at each step (e.g. Dziewonski 1984); however, this is far from optimal. For best results, it is necessary to invert for all source parameters and the earth model simultaneously. Implemented naïvely, this requires the solution of an extremely large system of equations, growing linearly with the number of events used. However, by exploiting the structure of the matrices involved, it is possible to obtain an algorithm where each iteration has a similar computational cost to that incurred in performing alternate source and structure inversions. As we shall see, fewer iterations are required to achieve a given accuracy when the two are combined, leading to an overall reduction in inversion time.

Some work has been done on related problems. Pavlis & Booker (1980) and Morelli & Dziewonski (1991) consider the problem of simultaneously relocating events while performing traveltime tomography, and present algorithms related to those used in this paper; it was found that the relocation was beneficial to the recovered model. Kennett & Sambridge (1998) also address problems with more than one parameter class through an iterative technique.

The results and methods we describe in this paper may be applied to many tomographic problems, and choices of modelling strategy. However, we shall concentrate on global, full-waveform tomography. Our approach to source determination is based on that of Dziewonski *et al.* (1981), and our structure inversion strategy builds on that of Woodhouse & Dziewonski (1984). In both cases, we treat the seismogram as being approximately linearly dependent on model parameters, and use least-squares minimization of the waveform misfit to evaluate the optimal updates to the parameters. For clarity, we will begin by setting out the standard approach to solving a general linearized inverse problem; we then consider applying this to the problem of source determination, and illustrate the errors that may occur if an inaccurate earth

model is used. By using a synthetic data set calculated using a known earth model, we can then investigate the effect these errors have on structure determination.

2 THE STANDARD LINEARIZED INVERSE PROBLEM

If we have a model, $\mathbf{m}$, then the assertion that data are linearly dependent on the model allows us to calculate synthetic data via the relationship

$$\mathbf{s} = \mathbf{A}\mathbf{m}\,, \tag{1}$$

where the linear operator $\mathbf{A}$ is known. We now seek the model, $\mathbf{m}$, such that the difference between synthetic and recorded data is minimized. Our approach will be based on the least-squares criterion; we therefore minimize the misfit function

$$m^2 = \frac{(\mathbf{d}-\mathbf{s})^{\mathbf{T}}(\mathbf{d}-\mathbf{s})}{\mathbf{d}^{\mathbf{T}}\mathbf{d}} = \frac{(\mathbf{d}-\mathbf{A}\mathbf{m})^{\mathbf{T}}(\mathbf{d}-\mathbf{A}\mathbf{m})}{\mathbf{d}^{\mathbf{T}}\mathbf{d}}\,, \tag{2}$$

with respect to the model parameters, and obtain the standard result (see, e.g. Menke 1989)

$$\mathbf{m} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{d}\,. \tag{3}$$

If our problem is exactly linear, eq. (3) returns the best-fitting model parameters in a single step. However, we can accommodate weak non-linearity by an iterative approach: at each step, we seek the update to the current estimate of the model that most improves the misfit. This is equivalent to finding the update that best explains the difference between data and synthetic, leading to the iterative scheme

$$\mathbf{m_1} = \mathbf{m_0} + \left(\mathbf{A}^{\mathbf{T}}\mathbf{A}\right)^{-1}\mathbf{A}^{\mathbf{T}}(\mathbf{d}-\mathbf{s_0})\,, \tag{4}$$

where $\mathbf{s_0}$ is the synthetic data calculated using the model $\mathbf{m_0}$.

If we have additional linear constraints on our model parameters, we can express these in the form

$$\mathbf{B}\mathbf{m} - \mathbf{c} = \mathbf{0}\,. \tag{5}$$

We can incorporate these into our determination via the method of Lagrange multipliers. This modifies eq. (4), so that the update to model parameters, $\mathbf{\Delta m}$, at each iteration can be written

$$\begin{pmatrix} \mathbf{\Delta m} \\ \lambda \end{pmatrix} = \begin{pmatrix} \mathbf{A}^{\mathbf{T}}\mathbf{A} & \mathbf{B}^{\mathbf{T}} \\ \mathbf{B} & \mathbf{0} \end{pmatrix}^{-1} \begin{pmatrix} \mathbf{A}^{\mathbf{T}}(\mathbf{d}-\mathbf{s_0}) \\ \mathbf{c} \end{pmatrix}. \tag{6}$$

The values of the Lagrange parameters, λ, are discarded, as usual.

It often turns out that the matrix $\mathbf{A}^{\mathbf{T}}\mathbf{A}$ is ill-conditioned, so that the inversion is numerically unstable. This arises where the available data are insufficient to fully constrain all model parameters, and this must be taken into account in the formulation of the problem. A number of different treatments are possible (e.g. Boschi & Dziewonski 1999); we follow that set out by Tarantola & Valette (1982). A statistical approach to inverse theory implies that a given set of data defines a probability density function in model-space; provided measurement errors are Gaussian, the least-squares formalism corresponds to obtaining the most probable model. However, a full treatment requires us to consider the covariances between data and model parameters, and recognize that we usually have *a priori* information about acceptable classes of model. It can be shown that this leads to an iterative scheme

$$\mathbf{m_1} = \mathbf{m_0} + \left(\mathbf{A}^{\mathbf{T}}\mathbf{C_d^{-1}}\mathbf{A} + \mathbf{C_m^{-1}}\right)^{-1}\left(\mathbf{A}^{\mathbf{T}}\mathbf{C_d^{-1}}(\mathbf{d}-\mathbf{s_0}) - \mathbf{C_m^{-1}}(\mathbf{m_0}-\mathbf{m_p})\right)\,, \tag{7}$$

where $\mathbf{C_d}$ is the data covariance matrix, $\mathbf{C_m}$ the model parameter covariance matrix and $\mathbf{m_p}$ our *a priori* model (Tarantola & Valette 1982, eq. (25), adapted to our current notation). In broad terms, the data covariance matrix represents our assessment of the measurement error in our data set; the model parameter covariance matrix represents the size of error that we are willing to accept on each model parameter.

By assuming each datum is independent, and by weighting the data set appropriately, we can ensure that the *a priori* errors on each datum are uniform. In this case, the *a priori* data covariance matrix is an identity matrix. If we now define $\mathbf{D} = \mathbf{C_m^{-1}}$, we obtain

$$\mathbf{m_1} = \mathbf{m_0} + \left(\mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}\right)^{-1}\left(\mathbf{A}^{\mathbf{T}}(\mathbf{d}-\mathbf{s_0}) - \mathbf{D}\left(\mathbf{m_0}-\mathbf{m_p}\right)\right)\,. \tag{8}$$

This algorithm is similar to that obtained by making the transformation $\mathbf{A}^{\mathbf{T}}\mathbf{A} \rightarrow \mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}$ in eq. (4), as is done when performing, for example, Tikhonov regularization. However, eq. (8) contains an additional term, $\left(\mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}\right)^{-1}\mathbf{D}\left(\mathbf{m_0}-\mathbf{m_p}\right)$ (as in Tarantola & Valette 1982, eq. (25)), which in practice is essential to achieve convergence in a non-linear problem. When the term is present, convergence will be reached when

$$\mathbf{A}^{\mathbf{T}}(\mathbf{d}-\mathbf{s_0}) = \mathbf{D}\left(\mathbf{m_0}-\mathbf{m_p}\right)\,, \tag{9}$$

with the difference between data and synthetic being explained to within the accepted errors in model parameters.

A somewhat more intuitive explanation for the incorporation of this term can be seen on defining the resolution matrix,

$$\mathbf{R} = \left(\mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}\right)^{-1}\mathbf{A}^{\mathbf{T}}\mathbf{A}\,. \tag{10}$$

It is then straightforward to show that eq. (8) can be written

$$\mathbf{m_1} - \mathbf{m_p} = \mathbf{R}\left(\mathbf{m_0}-\mathbf{m_p}\right) + \left(\mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}\right)^{-1}\mathbf{A}^{\mathbf{T}}(\mathbf{d}-\mathbf{s_0})\,, \tag{11}$$

showing that an update is made with respect to the model as resolved by the data, which need not be equivalent to the raw model parameters. The form of eq. (11) may be more suitable for computational purposes than eq. (8), especially when the regularization is performed in such a way that the matrix **D** is never explicitly obtained.

The resolution matrix represents the effect of regularization on the model, and defines how the model is recovered by an inversion. For a perfect recovery of the input model, the resolution matrix would be an identity matrix; in reality, this is not attainable. Instead, for most reasonable choices of regularization, we will observe a smoothing effect: short-wavelength features in the earth model will be lost or 'smeared out', and the resolution matrix will encapsulate this. In this case, not all components of our model parameterization may be determined: some may have no impact on the measurements we can make. To characterize our model, we can calculate the trace of the resolution matrix, representing the effective number of linearly independent parameters determined in the inversion.

3 SOURCE INVERSION

A seismic event can be described in terms of a 10-component source vector,

$$\mathbf{f} = \left(M_{rr}, M_{\theta\theta}, M_{\phi\phi}, M_{r\theta}, M_{r\phi}, M_{\theta\phi}, \theta_c, \phi_c, z_c, t_c\right)^{\mathbf{T}} , \tag{12}$$

where the M_{ij} represent the six independent components of the seismic moment tensor, and $(\theta_c, \phi_c, z_c, t_c)$ defines the spatio-temporal location of the event centroid. We can then express the synthetic seismogram as recorded at the point **r** as (e.g. Gilbert & Dziewonski 1975)

$$\mathbf{s}(t, \mathbf{r}, \mathbf{f}, \oplus) = \sum_{i=1}^{6} \left[\mathbf{G}_i(t, \mathbf{r}, \theta_c, \phi_c, z_c, t_c, \oplus) * \tau(t, t_c)\right] f_i , \tag{13}$$

where $\oplus$ represents a given earth model, and τ is a function describing the temporal evolution of the source. The $\mathbf{G}_i$ represent the Green's functions for the source, and can be calculated for a given earth model and source–receiver pair. We therefore see that the seismogram is exactly linear in the six moment tensor components, allowing the direct application of eq. (3) provided that the centroid location is known. Because we typically have a reasonable estimate of this from, for example, measurements of *P* and *S* wave traveltimes, we are able to obtain an initial estimate of the source vector. In practice, we apply the constraint that the source should have vanishing isotropic part, so that $M_{rr} + M_{\theta\theta} + M_{\phi\phi} = 0$, enforced via Lagrange multipliers as described earlier.

Having obtained this estimate, we can refine it allowing both the moment tensor and the centroid location to vary. For small changes in the source parameters, the first-order perturbation to the seismogram can be expressed

$$\mathbf{\Delta s} = \frac{\partial \mathbf{s}}{\partial f_1}\Delta f_1 + \cdots + \frac{\partial \mathbf{s}}{\partial f_6}\Delta f_6 + \frac{\partial \mathbf{s}}{\partial \theta_c}\Delta\theta_c + \frac{\partial \mathbf{s}}{\partial \phi_c}\Delta\phi_c + \frac{\partial \mathbf{s}}{\partial z_c}\Delta z_c + \frac{\partial \mathbf{s}}{\partial t_c}\Delta t_c . \tag{14}$$

A variety of methods exist to allow the calculation of the partial derivatives occurring in this expression, within a given earth model. Consequently, we can implement an algorithm of the form of eq. (4) to iteratively update all 10 components of the source vector to improve the agreement between data and synthetic. Defining **B** as the matrix of partial derivatives,

$$\mathbf{B} = \left(\frac{\partial \mathbf{s}}{\partial f_1} \cdots \frac{\partial \mathbf{s}}{\partial f_6} \quad \frac{\partial \mathbf{s}}{\partial \theta_c} \quad \frac{\partial \mathbf{s}}{\partial \phi_c} \quad \frac{\partial \mathbf{s}}{\partial z_c} \quad \frac{\partial \mathbf{s}}{\partial t_c} \right) , \tag{15}$$

we have

$$\mathbf{f}_{i+1} = \mathbf{f}_i + \left(\mathbf{B}_i^{\mathbf{T}}\mathbf{B}_i\right)^{-1}\mathbf{B}_i^{\mathbf{T}}\left(\mathbf{d} - \mathbf{s}_i\right) . \tag{16}$$

The subscript indices on **B** and **s** are included to emphasise that these quantities are recalculated on each iteration, using the source information obtained at the previous step.

This method allows us to find the source vector which best describes an event within a particular earth model—it need not correspond to the 'true' source vector if the earth model is inaccurate. To make meaningful use of the source information, we must develop an appreciation of how errors in the earth model are reflected in the recovered source parameters. The relationship is obviously complex; we shall seek only to illustrate the scale of uncertainty in the source parameters. To do this, we perform a simple experiment: synthetic seismograms are generated for known sources in a known earth model, then used as the data for source inversions in a different earth model. A comparison of the recovered and true source vectors may then be performed.

To generate our data, we use a simple degree-8 model containing heterogeneity of a scale similar to that seen in the Earth (model M84C, Woodhouse & Dziewonski 1984). The low angular order has been chosen to facilitate rapid experimentation; it is sufficient for illustrative purposes. This model is parameterized in terms of spherical harmonics for the angular part, with polynomial radial basis functions. We calculate synthetic seismograms using normal mode summation, and incorporate lateral heterogeneity via the path average approximation (Woodhouse & Dziewonski 1984). Seismograms are processed to yield two frequency–time windows per trace. The first is designed to enhance sensitivity to body waves, and contains frequencies below 0.016 Hz; the second contains frequencies below 0.007 Hz and is intended to enhance mantle waves. These methods are used for all seismograms discussed in this paper.

A common starting point for a global inversion is the laterally homogeneous model PREM (Dziewonski & Anderson 1981)—even if a 3-D model is used, this is likely to itself be derived from PREM. As such, we can assume that it is the 'least correct' earth model that we are likely to use: models obtained on successive iterations are intended to be increasingly representative of the true Earth. We therefore invert our synthetic data in PREM; the errors observed give an order-of-magnitude estimate of those expected in any reasonable tomographic inversion.

Table 1. Source parameter recovery in an incorrect earth model.

	Location			Errors			Focal mechanism	
Event	Lat	Lon	Depth (km)	Lateral (km)	Depth (km)	Time (s)	Correct	Recovered
012800B	43.05	146.84	61.10	54.56	6.44	0.99		
020600C	−5.84	150.88	33.00	27.16	1.65	0.30		
022600B	13.80	144.78	132.20	32.83	1.49	0.22		
042300B	−28.31	−62.99	608.50	56.49	2.09	0.81		
050600B	−11.30	165.43	12.00	42.55	1.46	2.26		

Note: Synthetic data for five typical events are generated in a degree-8 model, then inverted within PREM. Recovery of focal mechanism and spatio-temporal event locations after five iterations is shown.

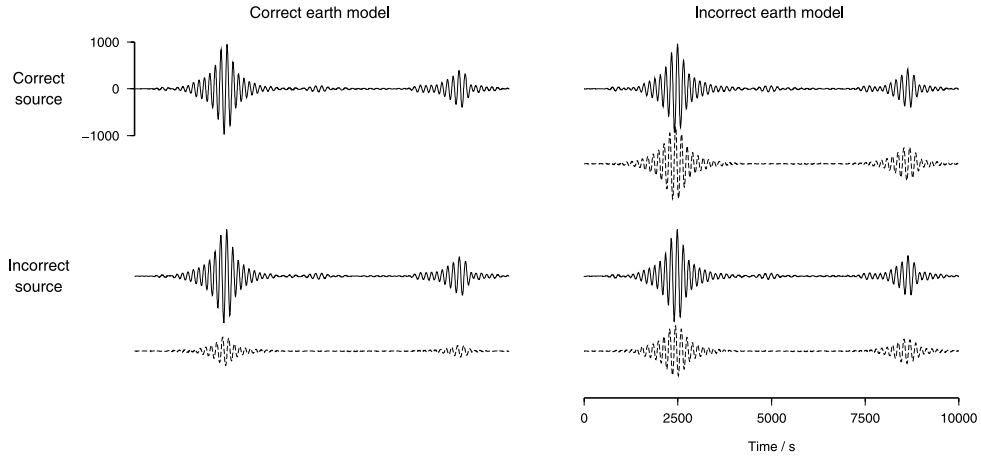

Figure 1. Synthetic vertical-component seismograms calculated at station ESK (Eskdalemuir, Scotland), for event 012800B. Left-hand column shows calculations performed in a degree-8, laterally heterogeneous model; right-hand column shows seismograms in PREM. The upper row was produced using the Harvard CMT catalogue source parameters for this event; the lower row uses source parameters recovered by a source inversion of the degree-8 data in PREM. Dotted lines denote difference between the corresponding seismogram, and the 'correct' trace (top left panel). All traces are plotted on the same scale.

These errors are presented in Table 1, for five characteristic events. The results represent average values across a number of experiments, with randomly chosen initial estimates for the centroid location within 50 km of the correct value. The recovered focal mechanism is seen to be good in all cases; however, errors in spatial location are large, at around 50 km. This is in agreement with the experience of Dziewonski *et al.* (1981), when first presenting the source inversion algorithm. We observe that this error is mainly due to lateral mislocation—depth resolution has not been strongly affected by the erroneous earth model. Others have reported similar results: for example, Antolik *et al.* (2001) experimented with source relocations using *P*-wave observations in a variety of models, and found average lateral mislocations of around 20 km regardless of the model used; they identify the size of the data set as a critical factor in improving accuracy.

The fact that there are differences between recovered and true source parameters are unsurprising: our source inversion algorithm attempts to compensate for errors in the fixed parameters—here, the earth model—by altering the free parameters. However, as we see from Fig. 1, the recovered model parameters no longer generate a zero misfit in the correct model. As we shall see, this causes significant difficulties when a naïve approach to structure inversion is taken.

4 STRUCTURE INVERSION

The general concept underlying structure inversion mirrors that for source inversions: to first order, the seismograms depend linearly on small changes in the structure parameters, $p_{1\ldots N}$, so that

$$\Delta \mathbf{s} = \frac{\partial \mathbf{s}}{\partial p_1}\Delta p_1 + \frac{\partial \mathbf{s}}{\partial p_2}\Delta p_2 + \cdots + \frac{\partial \mathbf{s}}{\partial p_N}\Delta p_N \,. \tag{17}$$

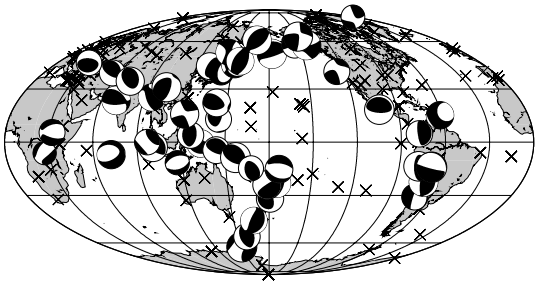

Figure 2. Events and stations in data set used for experiments presented in this paper. Crosses mark locations of stations in the data set; only those within the epicentral distance range $40^\circ < \theta < 140^\circ$ are used for any event.

Again, the partial derivatives can readily be computed, by a variety of methods, and an algorithm to iteratively refine the model parameters may be implemented. To obtain a stable solution, it is necessary to introduce regularization, so our inversion process implements eq. (8)

$$\mathbf{p}_{i+1} = \mathbf{p}_i + \left(\mathbf{A}_i^{\mathbf{T}}\mathbf{A}_i + \mathbf{D}\right)^{-1}\left(\mathbf{A}_i^{\mathbf{T}}\left(\mathbf{d} - \mathbf{s}_i\right) - \mathbf{D}\left(\mathbf{p}_i - \mathbf{p}_\mathbf{p}\right)\right), \tag{18}$$

with $\mathbf{A}$ representing the matrix of partial derivatives,

$$\mathbf{A} = \left(\frac{\partial \mathbf{s}}{\partial p_1} \cdots \frac{\partial \mathbf{s}}{\partial p_N}\right). \tag{19}$$

Subscript indices are again included to highlight the need to recalculate quantities using the best available earth model. In this paper, we take the regularization matrix to be a small, constant multiple of the identity matrix

$$\mathbf{D} = \epsilon\mathbf{I}. \tag{20}$$

Our prior model, $\mathbf{p}_\mathbf{p}$, will here be taken to be PREM.

The simplest approach to structure inversion assumes that the initial source determinations are accurate; repeated application of eq. (18) will then yield an improved earth model. To illustrate the errors that this approach can introduce, we generate a larger synthetic data set in our degree-8 model. This contains 46 events, as shown in Fig. 2, chosen by dividing the Earth's surface into 15° regions and choosing the most recent event in each of magnitude 6.5+, where any exist. Station distribution is based on the reporting stations for the IRIS/IDA global network. We process the resulting data as described earlier, and determine the best-fitting sources within PREM. We find that, on average, the sources are mislocated by 43.5 km, in broad agreement with our earlier experiments. We then determine the best-fitting 3-D structure, performing five iterations of our structure inversion.

We choose a damping parameter $\epsilon = 0.01$, yielding a model with 140 effective free parameters. This is a conservative choice, and does not represent the best resolution obtainable with our data; however, it ensures that differences between models can be clearly seen. Note that the value of the damping parameter is given in arbitrary units, and depends on the various scaling choices made when assembling the inverse problem: it is stated only to facilitate comparison between the experiments presented here. An interpretation of the impact of any particular damping parameter should be based upon the effective number of model parameters.

The reduction in waveform misfit over the five structure iterations is shown as the first row of Table 2; we see that a modest improvement is obtained. However, the recovered model contains large errors, as can be seen in figures 3 (after the first iteration) and 4 (after five iterations). It appears that successive iterations generally act to increase the errors in the model, despite reductions in misfit. This is due to the errors in the initial source determination—these induce waveform changes that the structure inversion attempts to fit.

For comparison purposes, we repeat the experiment, but determine the initial source vectors using a degree-8 chequerboard earth model. We then perform structure inversions as before, with PREM as our starting model. Fig. 5 shows the errors in the recovered model after five iterations, and the difference between this and the model recovered when source determinations are performed in PREM. We clearly see an imprint from the degree-8 chequerboard—despite this not being used during structure inversion.

One obvious approach to reducing these effects is to recompute the source parameters at each structure iteration, using the best available earth model. This should allow a gradual improvement in source estimates on successive iterations, with a resulting reduction in structure errors. To test this, we repeat our structure determination, using initial sources as determined in PREM. Iteration-by-iteration misfits are as presented as experiment 2 in Table 2; we see that these show a considerable improvement on those obtained without source parameter refinement. The recovered model after the first iteration is identical to that found earlier; however, after five iterations the errors are considerably reduced—see Fig. 4. In addition, Table 3 shows that the average error in source locations reduces significantly with successive iterations.

Performing alternate source and structure inversions is straightforward to implement, and has clear benefits. However, this method relies on the assumption that at any point in the inversion, all waveform errors must be resolved by *only* source parameter corrections, or *only* earth model updates. As we have seen, this is not realistic, and at best can be regarded as sub-optimal. We therefore require a more complete treatment, that removes this problem.

Table 2. Iteration-by-iteration evolution of overall waveform misfit, calculated as in eq. (2), during 16 tomographic experiments.

							Overall waveform misfit					
	Method	ϵ	μ	N_{str}	N_{src}	N_{tot}	PREM	I1	I2	I3	I4	I5
1	SNG	0.01	–	140	414	564	0.134	0.084	0.080	0.078	0.077	0.077
2	ALT	0.01	–	140	414	564	0.134	0.063	0.043	0.035	0.032	0.030
3	COM	0.03	0.00	140	414	564	0.134	0.054	0.034	0.029	0.027	0.027
4	COM	0.0185	0.02	140	392	532	0.134	0.052	0.034	0.028	0.026	0.026
5	COM	0.01	0.00	181	414	595	0.134	0.050	0.029	0.023	0.022	0.022
6	COM	0.01	0.01	170	392	562	0.134	0.049	0.030	0.024	0.023	0.022
7	COM	0.01	0.02	162	384	546	0.134	0.049	0.031	0.025	0.024	0.023
8	COM	0.01	0.1	152	373	525	0.134	0.050	0.032	0.026	0.025	0.024
9	COM	0.001	0.00	284	414	698	0.134	0.050	0.027	0.021	0.020	0.020
10	COM	0.001	0.01	270	392	662	0.134	0.049	0.026	0.021	0.019	0.019
11	COM	0.001	0.02	265	384	649	0.134	0.049	0.026	0.020	0.019	0.019
12	COM	0.001	0.1	257	373	630	0.134	0.050	0.027	0.021	0.019	0.019
13	COM	0.0001	0.00	318	414	732	0.134	0.052	0.028	0.021	0.020	0.020
14	COM	0.0001	0.01	315	392	707	0.134	0.051	0.027	0.021	0.019	0.019
15	COM	0.0001	0.02	313	384	697	0.134	0.051	0.027	0.020	0.019	0.019
16	COM	0.0001	0.1	311	373	684	0.134	0.051	0.027	0.020	0.019	0.019

Note: The same data set is used throughout; the inversion method and regularization parameters vary. 'SNG' denotes inversion using a single determination of source parameters, 'ALT' signifies that alternate source and structure inversions were performed (Section 4). 'COM' refers to inversion using the combined source and structure algorithm (Section 5). ϵ and μ represent damping parameters for structure and source inversions, respectively, as defined in eqs (20), (33) and (34). N_{str} denotes the effective number of free structural parameters in the inversion, and N_{src} the effective number of free source parameters. N_{tot} is the sum of these two quantities.

Table 3. Mean spatial error in event locations on an iteration-by-iteration basis during 16 tomographic experiments.

				Average spatial error (km)					
	Method	ϵ	μ	PREM	I1	I2	I3	I4	I5
1	SNG	0.01	–	43.5 (15.4)	–	–	–	–	–
2	ALT	0.01	–	43.5 (15.4)	27.5 (9.3)	19.0 (6.9)	13.7 (5.8)	10.6 (5.0)	8.7 (4.5)
3	COM	0.03	0.00	43.5 (15.4)	14.8 (7.8)	6.7 (3.6)	6.0 (3.1)	5.7 (2.8)	5.6 (2.7)
4	COM	0.0185	0.02	43.5 (15.4)	14.1 (7.3)	6.3 (3.4)	5.5 (2.8)	5.2 (2.5)	5.1 (2.4)
5	COM	0.01	0.00	43.5 (15.4)	13.1 (7.1)	5.2 (2.7)	3.9 (2.1)	3.5 (1.8)	3.4 (1.7)
6	COM	0.01	0.01	43.5 (15.4)	12.9 (6.8)	5.2 (2.8)	4.1 (2.1)	3.8 (1.9)	3.7 (1.8)
7	COM	0.01	0.02	43.5 (15.4)	12.9 (6.7)	5.3 (2.9)	4.3 (2.2)	4.0 (2.0)	3.9 (1.9)
8	COM	0.01	0.1	43.5 (15.4)	13.2 (6.7)	5.7 (3.0)	4.8 (2.3)	4.5 (2.1)	4.5 (2.0)
9	COM	0.001	0.00	43.5 (15.4)	12.4 (6.9)	4.6 (2.4)	3.0 (1.7)	2.4 (1.5)	2.3 (1.4)
10	COM	0.001	0.01	43.5 (15.4)	11.8 (6.3)	4.3 (2.3)	2.7 (1.5)	2.3 (1.4)	2.1 (1.3)
11	COM	0.001	0.02	43.5 (15.4)	11.6 (6.2)	4.3 (2.3)	2.7 (1.5)	2.2 (1.4)	2.1 (1.3)
12	COM	0.001	0.1	43.5 (15.4)	11.6 (6.1)	4.3 (2.3)	2.8 (1.4)	2.4 (1.3)	2.3 (1.2)
13	COM	0.0001	0.00	43.5 (15.4)	12.4 (6.9)	4.6 (2.4)	2.9 (1.6)	2.3 (1.4)	2.2 (1.3)
14	COM	0.0001	0.01	43.5 (15.4)	11.7 (6.3)	4.2 (2.2)	2.5 (1.5)	2.1 (1.3)	2.0 (1.3)
15	COM	0.0001	0.02	43.5 (15.4)	11.5 (6.2)	4.2 (2.2)	2.5 (1.4)	2.1 (1.3)	1.9 (1.3)
16	COM	0.0001	0.1	43.5 (15.4)	11.5 (6.0)	4.2 (2.2)	2.6 (1.3)	2.2 (1.2)	2.1 (1.1)

Note: Quantities in parentheses represent the standard deviation on each average. See Table 2 for corresponding misfit evolution and definitions of the inversion methods, ϵ and μ.

5 COMBINED SOURCE–STRUCTURE INVERSION

We can combine both structure and all source determinations into a single inverse problem. Denoting our earth model parameters by the vector $\mathbf{p}$, and N source vectors by $\mathbf{f}_{1\ldots N}$, and treating each seismogram as depending linearly on both earth model and the relevant source vector, we can write

$$\begin{pmatrix} \mathbf{e}_1 \\ \mathbf{e}_2 \\ \vdots \\ \mathbf{e}_N \end{pmatrix} = \begin{pmatrix} \mathbf{A}_1 & \mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{A}_2 & \mathbf{0} & \mathbf{B}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{A}_N & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_N \end{pmatrix} \begin{pmatrix} \mathbf{p} \\ \mathbf{f}_1 \\ \mathbf{f}_2 \\ \vdots \\ \mathbf{f}_N \end{pmatrix}, \tag{21}$$

where for clarity we have defined $\mathbf{e}_i = \mathbf{d}_i - \mathbf{s}_i$, the error in the synthetic seismograms for the ith event. $\mathbf{A}_i$ is defined as in eq. (19), and represents the sensitivity of our data to the earth model parameters; $\mathbf{B}_i$ represents sensitivity to source as in eq. (15). Subscript indices are now used

to distinguish quantities arising from different events. From eq. (4), we see that the optimal update to the model parameters requires the solution of

$$\begin{pmatrix} \mathbf{A}^{\mathbf{T}}\mathbf{A} & \mathbf{A}_1^{\mathbf{T}}\mathbf{B}_1 & \mathbf{A}_2^{\mathbf{T}}\mathbf{B}_2 & \cdots & \mathbf{A}_N^{\mathbf{T}}\mathbf{B}_N \\ \mathbf{B}_1^{\mathbf{T}}\mathbf{A}_1 & \mathbf{B}_1^{\mathbf{T}}\mathbf{B}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{B}_2^{\mathbf{T}}\mathbf{A}_2 & \mathbf{0} & \mathbf{B}_2^{\mathbf{T}}\mathbf{B}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \vdots & \ddots & \vdots \\ \mathbf{B}_N^{\mathbf{T}}\mathbf{A}_N & \mathbf{0} & \mathbf{0} & \cdots & \mathbf{B}_N^{\mathbf{T}}\mathbf{B}_N \end{pmatrix} \begin{pmatrix} \Delta\mathbf{p} \\ \Delta\mathbf{f}_1 \\ \Delta\mathbf{f}_2 \\ \vdots \\ \Delta\mathbf{f}_N \end{pmatrix} = \begin{pmatrix} \mathbf{A}^{\mathbf{T}}\mathbf{e} \\ \mathbf{B}_1^{\mathbf{T}}\mathbf{e}_1 \\ \mathbf{B}_2^{\mathbf{T}}\mathbf{e}_2 \\ \vdots \\ \mathbf{B}_N^{\mathbf{T}}\mathbf{e}_N \end{pmatrix}, \tag{22}$$

where we have defined

$$\mathbf{A}^{\mathbf{T}}\mathbf{A} = \sum_{i=1}^{N} \mathbf{A}_i^{\mathbf{T}}\mathbf{A}_i \,, \tag{23}$$

$$\mathbf{A}^{\mathbf{T}}\mathbf{e} = \sum_{i=1}^{N} \mathbf{A}_i^{\mathbf{T}}\mathbf{e}_i \,. \tag{24}$$

Direct solution of eq. (22) is likely to be unfeasible for realistic data sets and model paramaterizations—the size of the matrix to be inverted scales with the number of events used, and the computational cost scales as the cube of this (e.g. Golub & van Loan 1996). However, we can exploit the fact that most blocks in the matrix contain only zeroes, and use a process similar to Gaussian elimination to obtain a solution (see also Morelli & Dziewonski 1991). Incorporating a symmetric regularization term in analogy with eq. (8), and grouping similar terms under the submatrices $\mathbf{A}^{\mathbf{T}}\mathbf{B}$, $\mathbf{B}^{\mathbf{T}}\mathbf{B}$, $\mathbf{B}^{\mathbf{T}}\mathbf{e}$ and $\mathbf{\Delta F}$ for brevity, eq. (22) can be written

$$\begin{pmatrix} \mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_1 & \mathbf{A}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_2 \\ \mathbf{B}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_2^{\mathbf{T}} & \mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3 \end{pmatrix} \begin{pmatrix} \mathbf{\Delta p} \\ \mathbf{\Delta F} \end{pmatrix} = \begin{pmatrix} \mathbf{A}^{\mathbf{T}}\mathbf{e} - \mathbf{D}_1\mathbf{p}_0 - \mathbf{D}_2\mathbf{F}_0 \\ \mathbf{B}^{\mathbf{T}}\mathbf{e} - \mathbf{D}_2^{\mathbf{T}}\mathbf{p}_0 - \mathbf{D}_3\mathbf{F}_0 \end{pmatrix}. \tag{25}$$

We therefore have a pair of simultaneous equations relating the source and structure corrections to the difference between data and synthetic:

$$(\mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_1)\mathbf{\Delta p} + (\mathbf{A}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_2)\mathbf{\Delta F} = \mathbf{A}^{\mathbf{T}}\mathbf{e} - \mathbf{D}_1\mathbf{p}_0 - \mathbf{D}_2\mathbf{F}_0 \,, \tag{26}$$

$$(\mathbf{B}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_2^{\mathbf{T}})\mathbf{\Delta p} + (\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3)\mathbf{\Delta F} = \mathbf{B}^{\mathbf{T}}\mathbf{e} - \mathbf{D}_2^{\mathbf{T}}\mathbf{p}_0 - \mathbf{D}_3\mathbf{F}_0 \,. \tag{27}$$

Solving eq. (27) for $\mathbf{\Delta F}$, we find

$$\mathbf{\Delta}\boldsymbol{F} = \left(\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3\right)^{-1}\left(\mathbf{B}^{\mathbf{T}}\mathbf{e} - (\mathbf{B}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_2^{\mathbf{T}})\mathbf{\Delta}\boldsymbol{p} - \mathbf{D}_2^{\mathbf{T}}\mathbf{p}_0 - \mathbf{D}_3\mathbf{F}_0\right). \tag{28}$$

This can be substituted back into eq. (26), and solved to yield an expression for the optimal model update, $\mathbf{\Delta p}$, free from any explicit dependence on the source corrections

$$\begin{aligned} \mathbf{\Delta}\boldsymbol{p} = &\left[\mathbf{A}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_1 - (\mathbf{A}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_2)(\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3)^{-1}(\mathbf{B}^{\mathbf{T}}\mathbf{A} + \mathbf{D}_2^{\mathbf{T}})\right]^{-1}\left[\mathbf{A}^{\mathbf{T}}\mathbf{e} - (\mathbf{A}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_2)(\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3)^{-1}\mathbf{B}^{\mathbf{T}}\mathbf{e}\right. \\ &\left. -(\mathbf{D}_1 - (\mathbf{A}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_2)(\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3)^{-1}\mathbf{D}_2^{\mathbf{T}})\mathbf{p}_0 - (\mathbf{D}_2 - (\mathbf{A}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_2)(\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3)^{-1}\mathbf{D}_3)\mathbf{F}_0\right]. \end{aligned} \tag{29}$$

We can use this to calculate an update to our earth model, which in turn may be used in conjunction with eq. (28) to extract the implied source corrections. These two equations therefore represent an exact solution to eq. (25), and define an algorithm for iteratively refining both source and structure parameters, simultaneously.

It may not be immediately clear that we have gained anything by this process. However, provided that $\mathbf{D}_3$ has the same block-diagonal structure as $\mathbf{B}^{\mathbf{T}}\mathbf{B}$—a reasonable requirement, because there is no reason to suppose that source parameters for separate events should be coupled—the inverse of $(\mathbf{B}^{\mathbf{T}}\mathbf{B} + \mathbf{D}_3)$ is straightforward to compute. Under these circumstances, we can use the fact that

$$\begin{pmatrix} \mathbf{M}_1 & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2 & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_N \end{pmatrix}^{-1} = \begin{pmatrix} \mathbf{M}_1^{-1} & \mathbf{0} & \cdots & \mathbf{0} \\ \mathbf{0} & \mathbf{M}_2^{-1} & \cdots & \mathbf{0} \\ \vdots & \vdots & \ddots & \vdots \\ \mathbf{0} & \mathbf{0} & \cdots & \mathbf{M}_N^{-1} \end{pmatrix}, \tag{30}$$

for general square matrices $\mathbf{M}_i$, so that the cost of this inversion scales only linearly with the number of events, and a parallelized implementation is straightforward. The second matrix that must be inverted to solve eq. (29) has a dimension equivalent to the number of model parameters, and this is the main computational bottleneck in any implementation of this algorithm. However, a similar matrix must be inverted when performing alternate structure and source iterations: combining the two does not require increased computational resources. In fact, we find

that simultaneous inversion has modest performance advantages, as the total overhead costs can be reduced. However, as we shall see, the main improvement in computational performance is associated with a reduction in the number of iterations required to achieve a given accuracy.

To progress further, we must define the regularization matrices $\mathbf{D}_{1\ldots3}$. As before, there is no single correct choice; however, one reasonable approach is to take $\mathbf{D}_1 = \epsilon\mathbf{I}$, $\mathbf{D}_3 = \mu\mathbf{I}$, $\mathbf{D}_2 = \mathbf{0}$. This corresponds to the assertion that the *a priori* errors on source parameters are characterized by a Gaussian distribution of width $\sqrt{\mu}$, and those on structure parameters by a distribution of width $\sqrt{\epsilon}$. Using these values in eqs 28 and 29 yields

$$\mathbf{\Delta F} = (\mathbf{B^T B} + \mu\mathbf{I})^{-1}(\mathbf{B^T e} - \mu\mathbf{F}_0 - \mathbf{B^T A \Delta p}), \tag{31}$$

$$\mathbf{\Delta p} = \left[\mathbf{A^T A} + \epsilon\mathbf{I} - \mathbf{A^T B}(\mathbf{B^T B} + \mu\mathbf{I})^{-1}\mathbf{B^T A}\right]^{-1}\left[\mathbf{A^T e} - \epsilon\mathbf{p}_0 - \mathbf{A^T B}(\mathbf{B^T B} + \mu\mathbf{I})^{-1}(\mathbf{B^T e} - \mu\mathbf{F}_0)\right]. \tag{32}$$

This can be broken down into quantities computed for individual events:

$$\mathbf{\Delta f}_i = \left(\mathbf{B}_i^{\mathrm{T}}\mathbf{B}_i + \mu\mathbf{I}\right)^{-1}\left(\mathbf{B}_i^{\mathrm{T}}\mathbf{e}_i - \mu\mathbf{f}_{i0} - \mathbf{B}_i^{\mathrm{T}}\mathbf{A}_i\,\mathbf{\Delta p}\right), \tag{33}$$

$$\Delta\mathbf{p} = \left[\sum_{i=1}^{N}\left(\mathbf{A}_i^{\mathrm{T}}\mathbf{A}_i - \mathbf{A}_i^{\mathrm{T}}\mathbf{B}_i\left(\mathbf{B}_i^{\mathrm{T}}\mathbf{B}_i + \mu\mathbf{I}\right)\mathbf{B}_i^{\mathrm{T}}\mathbf{A}_i\right) + \epsilon\mathbf{I}\right]^{-1}\left[\sum_{i=1}^{N}\left(\mathbf{A}_i^{\mathrm{T}}\mathbf{e}_i - \mathbf{A}_i^{\mathrm{T}}\mathbf{B}_i\left(\mathbf{B}_i^{\mathrm{T}}\mathbf{B}_i + \mu\mathbf{I}\right)\left(\mathbf{B}_i^{\mathrm{T}}\mathbf{e}_i - \mu\mathbf{f}_{i0}\right)\right) - \epsilon\mathbf{p}_0\right]. \tag{34}$$

As a result, the computation is relatively straightforward, and code developed to implement alternate structure and source inversions can be adapted to perform simultaneous inversion.

To allow a direct comparison with the results from alternate inversions, we first choose $\epsilon = 0.03$, $\mu = 0.00$. This again gives a model containing 140 effective free parameters; the waveform misfits over five iterations are tabulated as the third example in Table 2, and the source location errors in Table 3. We see that misfits are comparable with those obtained using alternate source and structure inversions, but source locations are significantly better, with one combined iteration yielding a similar improvement to three pairs of alternate iterations. The errors in recovered earth models after one and five iterations are shown in Figs 3 and 4 respectively, with counterparts obtained by the methods already discussed. We see that the overall spatial form of these errors remains largely unchanged—this is principally governed by the spatial distribution of sources and receivers. However, the amplitude of the errors is significantly reduced when the new algorithm is used.

If we instead choose $\epsilon = 0.0185$, $\mu = 0.02$, we still obtain a model with 140 effective free structure parameters, but the regularization on the source determination reduces the effective number of source parameters sought. Again, misfits and source location errors are presented in Tables 2 and 3, as experiment four; we see that there is a small decrease in these. The recovered earth model changes only slightly from that seen with $\mu = 0.00$.

To explore the effect of the damping parameters further, Fig. 6 shows the errors at a depth of 200 km in the recovered earth model for different combinations of ϵ and μ. As we expect, we see that reducing ϵ increases the effective number of structure parameters recovered, and thus improves the fit to the model. However, if we reduce it too much, we lose the benefits of regularization, and errors begin to increase.

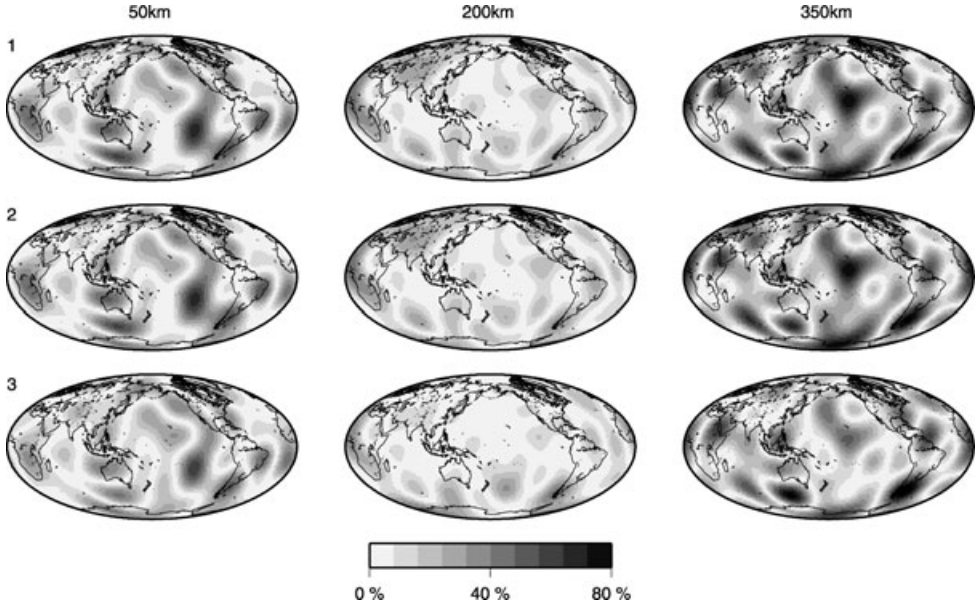

Figure 3. Errors in recovered models after one iteration; numbers refer to the experiments presented in Tables 2 and 3. In all three, a model with 140 effective free structural parameters is sought. Errors are plotted as a percentage of the maximum perturbation in the true model at that depth. Note that experiments 1 and 2 are identical at the first iteration; experiment 3—combined source and structure inversion—shows some improvement on these.

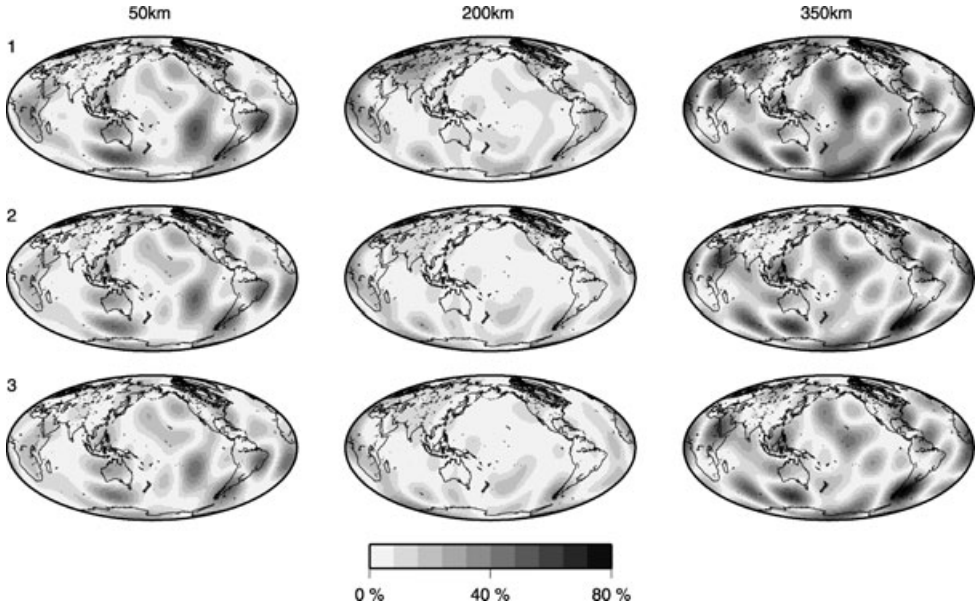

Figure 4. Errors in recovered models after five iterations; compare with Fig. 3. We see that alternate source and structure inversion (experiment 2) is a significant improvement on inversion without re-determination of source parameters; combining source and structure inversion leads to a modest further improvement, with a decrease in the amplitude of the errors.

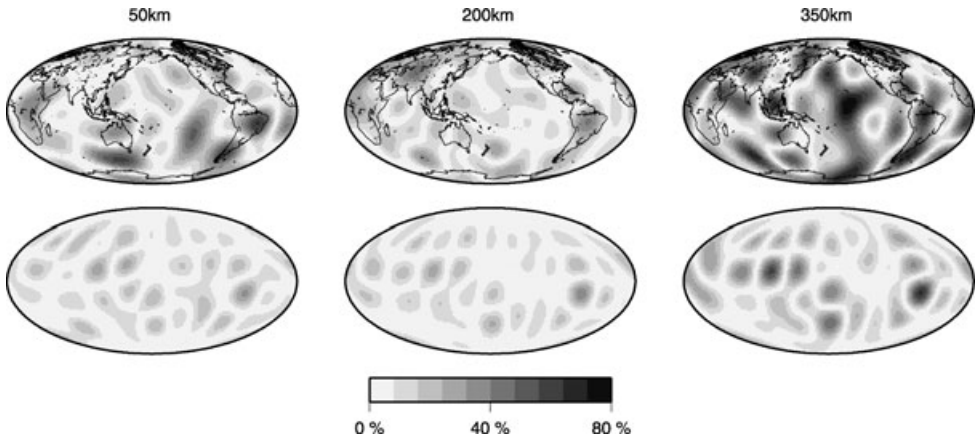

Figure 5. Evidence that the model used for source determination influences the model recovered after tomographic inversion. Upper row shows errors in model recovered after five structure inversions taking PREM as a starting point; source parameters were found in a degree-8 chequerboard model. Lower row shows difference between recovered model and that obtained when source parameters are determined in PREM (experiment 1, Fig. 4). The only difference between these cases is the model used to perform initial source determination; an imprint of the degree-8 chequerboard is clearly present.

Corresponding misfits and source location errors are presented as experiments 5–16 in Tables 2 and 3; notice that the models with over 300 effective structural parameters do not show an improvement in waveform fit compared to those with 250 effective parameters. This suggests that the additional fifty parameters are not providing useful information, and should not be used.

We see that a non-zero value for μ improves matters; however, too high a value results in the loss of useful information, and a deterioration in the earth model. It would appear that a weak regularization on the source parameters acts to improve stability, and therefore allows a more rapid convergence. From the source location errors, we make an interesting observation: the best source determinations are not necessarily made in the best earth model. The minimum average source error in our experiments is obtained for $\epsilon = 0.0001$, $\mu = 0.02$, yet we see that a larger value of ϵ would yield a better estimate of structure.

This effect may seem counter-intuitive: after all, the 'true' model should minimize both source and structure errors. We acknowledge that we have tested only a few values of ϵ and μ—undoubtedly, it will be possible to find values that improve upon the results we have presented, and it may be that values that minimize both source and structure errors can be found. However, it is more likely that this effect arises from a deficiency in our regularization approach: our assessment of the *a priori* errors in the various model parameters is incomplete. Of course, a more thorough treatment of these is possible—but the complexity of the resulting inverse problem is likely to increase markedly as we alter

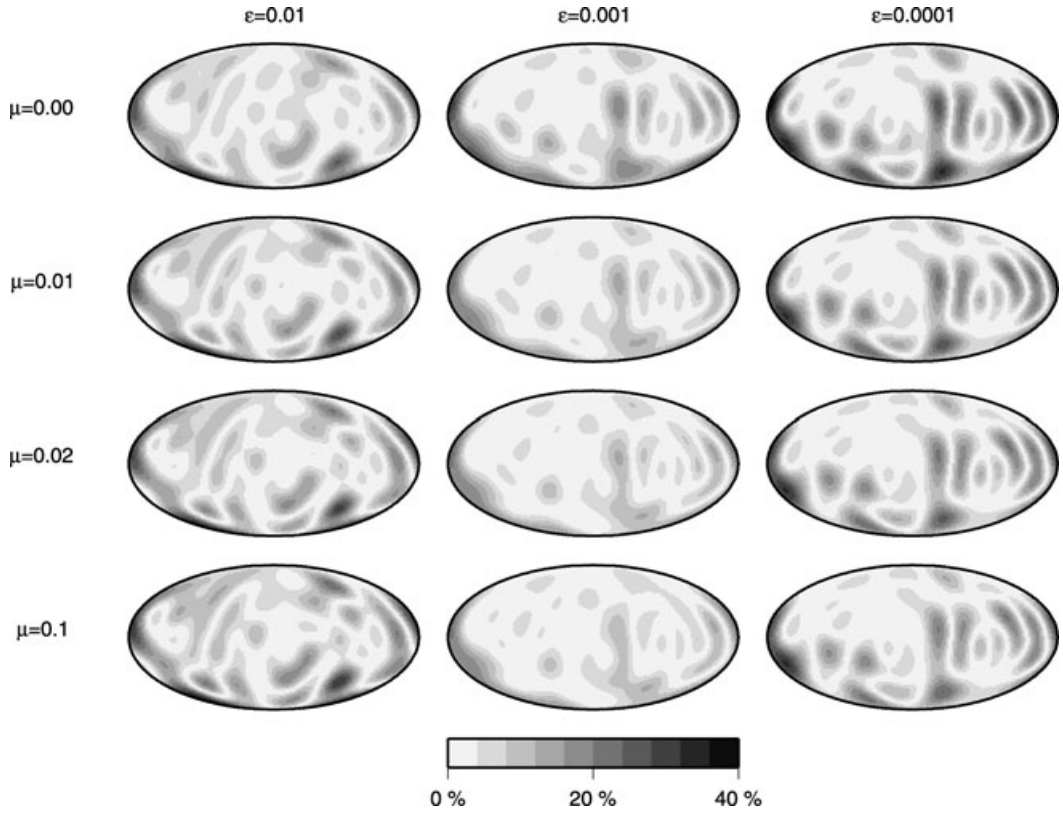

Figure 6. The effects of regularization: errors in model recovered by combined inversion for sources and structure, using a range of different values of ϵ and μ. All slices are at a depth of 200 km. Misfit evolution and source parameter errors are presented as experiments 4–16 in Tables 2 and 3. For the example here, least error in recovered model is obtained when $\epsilon = 0.001$ and with non-zero μ; least error in recovered source parameters occurs when $\epsilon = 0.0001$, $\mu = 0.02$.

the various regularization matrices. It is likely that eqs (33) and (34) represent the best solution to the problem of simultaneous source and structure inversion that can reasonably be implemented; the determination of optimal values for μ and ϵ depends on the particular problem being solved, and on the intended application of the results.

6 CONCLUSIONS

We have demonstrated that source locations determined by methods based on that of Dziewonski *et al.* (1981) are strongly sensitive to the earth model used. This may be important when interpreting results, or when using sources to calculate synthetic seismograms; care should be taken to ensure that any models used are compatible with the one in which sources were determined, otherwise significant errors can arise.

Because of this, tomographic inversions cannot rely on a single initial determination of all source information, or on source vectors obtained from a catalogue. Doing so will yield a model with significant errors, and containing the imprint of the model used to deduce the sources. Instead, some attempt must be made to recalculate sources within the framework of the tomographic algorithm, making use of the information that has been gained.

One solution is to perform alternate source and structure inversions, and this leads to significant reductions in all errors. However, it is slow and inefficient. Instead, we have developed a method that enables simultaneous inversion for all source parameters, and the earth model. By exploiting the structure of the equations, we are able to solve the simultaneous problem for the same computational cost as alternate inversions. We have demonstrated that this yields a remarkable improvement in performance, with a fourfold reduction in the number of iterations required to attain a given accuracy. As is usual in global tomography, regularization is required to allow a stable solution to be obtained; we have shown that applying this to both source and structure terms gives best results.

Our focus in this paper has been on global, full-waveform inversion, and we have not investigated the extent to which our results apply to other tomographic problems. However, we would expect that nearly all will be affected by the interdependence of source and structure determinations, and it is clear that attention must be given to the possible errors, and ways to reduce them. An adaptation of the algorithm presented in this paper may prove beneficial.

Even within our specific application, it is apparent that the precise details of the regularization required, and the errors that may remain after inversion, will vary greatly for different data sets and inversion strategies. It is therefore important that anyone seeking to perform

tomographic inversion performs synthetic experiments to assess the behaviour and reliability of the results they may obtain—and this must in turn inform any interpretation of results.

ACKNOWLEDGMENTS

The authors thank Ana Ferreira and David Al-Attar for helpful discussions, and Lapo Boschi and an anonymous reviewer for their comments and suggestions. APV is supported by the UK Natural Environment Research Council under the grant NE/C510916/1 and by Worcester College, Oxford. Computational and infrastructural support has also been provided through NE/B505997/1.

REFERENCES

Antolik, M., Ekström, G. & Dziewonski, A., 2001. Global event location with full and sparse data sets using three-dimensional models of mantle P-wave velocity, *Pure appl. Geophys.,* **158,** 291–317.

Boschi, L. & Dziewonski, A., 1999. High- and low-resolution images of the Earth's mantle: implications of different approaches to tomographic modeling, *J. geophys. Res.,* **104,** 25 567–25 594.

Dziewonski, A., 1984. Mapping the lower mantle: determination of lateral heterogeneity in P velocity up to degree and order 6, *J. geophys. Res.,* **89,** 5929–5952.

Dziewonski, A. & Anderson, D., 1981. Preliminary reference earth model, *Phys. Earth planet. Inter.,* **25,** 297–356.

Dziewonski, A., Chou, T.-A. & Woodhouse, J., 1981. Determination of earthquake source parameters from waveform data for studies of global and regional seismicity, *J. geophys. Res.,* **86,** 2825–2852.

Gilbert, F. & Dziewonski, A., 1975. An application of normal mode theory to the retrieval of structural parameters and source mechanisms from seismic spectra, *Phil. Trans. R. Soc. Lond. Ser. A, Math. Phys. Sci.,* **278,** 187–269.

Golub, G. & van Loan, C., 1996. *Matrix Computations*, John Hopkins University Press, Baltimore.

Kennett, B. & Sambridge, M., 1998. Inversion for multiple parameter classes, *Geophys. J. Int.,* **135,** 304–306.

Menke, W., 1989. *Geophysical Data Analysis: Discrete Inverse Theory*, Academic Press, New York.

Morelli, A. & Dziewonski, A., 1991. Joint determination of lateral heterogeneity and earthquake location, in *Glacial Isostasy, Sea-Level and Mantle Rheology*, Kluwer Academic Publishers, Dordrecht.

Pavlis, G. & Booker, J., 1980. The mixed discrete-continuous inverse problem: application to the simultaneous determination of earthquake hypocenters and velocity structure, *J. geophys. Res.*, **85**, 4801–4810.

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

Tarantola, A. & Valette, B., 1982. Generalized nonlinear inverse problems solved using the least squares criterion, *Rev. Geophys. Space Phys.*, **20**, 219–232.

Thurber, C. & Ritsema, J., 2007. Theory and observations – seismic tomography and inverse methods, in *Treatise on Geophysics*,Vol. 1, chap. 10, pp. 323–360, Elsevier, Amsterdam.

Woodhouse, J. & Dziewonski, A., 1984. Mapping the upper mantle: three-dimensional modelling of earth structure by inversion of seismic waveforms., *J. geophys. Res.,* **89,** 5953–5986.