SEISMIC TRAVELTIME TOMOGRAPHY OF THE CRUST AND LITHOSPHERE

N. Rawlinson and M. Sambridge
Research School of Earth Sciences, Australian National University,
Canberra ACT 0200, Australia

1 Introduction

1.1 Motivation

Seismic data represent one of the most valuable resources for investigating the internal structure and composition of the earth. One of the first people to deduce earth structure from seismic records was Mohorovičić, a Serbian seismologist who, in 1909, observed two distinct traveltime curves from a regional earthquake. He determined that one curve corresponded to a direct crustal phase and the other to a wave refracted by a discontinuity in elastic properties between crust and upper mantle. This world-wide discontinuity is now known as the Mohorovičić discontinuity or Moho for short. On a larger scale, the method of Herglotz and Wiechart (see, for example, Gubbins, 1992) was first implemented in 1910 to construct a 1-D whole earth model. The method uses the relationship between angular distance and ray parameter to determine velocity as a function of radius within the earth.

Today, an abundance of methods exist for determining earth structure from seismic waves. Different components of the seismic record may be used, including traveltimes, amplitudes, waveform spectra, full waveforms or the entire wavefield. Source-receiver configurations also differ - receiver arrays may be in-line or 3-D, sources may be close or distant to the receiver array, sources may be natural or artificial, and the scale of the study may be from tens of meters to the whole earth. Finally, there are a multitude of ways of translating the data extracted from the seismogram into a representation of seismic structure.

The purpose of this article is to review a particular class of methods for imaging earth structure called seismic traveltime tomography. This is a form of seismic traveltime inversion that is used to constrain 2-D and 3-D models of the Earth represented by a significant number of parameters. The word *tomography* literally means *slice picture* (from the Greek word *tomos* meaning *slice*) and was first used in medical imaging to describe the process of mapping the internal density distribution of the human body using x-rays (Lee & Pereyra, 1993). The term was later appropriated by the seismological community to describe a similar process using seismic waves to map earth structure. Seismologists now routinely use *tomography* to refer to 3-D structural imaging even though, strictly speaking, the word was originally designed to describe the imaging of 2-D slices only.

Inversion of source-receiver traveltimes of seismic waves is undoubtedly the most popular technique for imaging subsurface structure at all scales. However, comprehensive up-to-date reviews of the methodology and their application are rarely found in the literature (some useful reference works include Nolet, 1987b; Iyer & Hirahara, 1993; Kennett, 1998). We hope to at least partially address this problem. Here, we restrict ourselves to traveltime tomography used in studies of the crust and lithosphere. These local scale studies typically involve the deployment of seismometer arrays with a spatial coverage of several hundred km or less in any dimension. The class of data that may be

recorded in such experiments include normal incidence reflection, refraction and wide-angle reflection, teleseismic and local earthquake arrival times. Although the inversion methods used in these studies are often common to a wide range of other seismic tomography applications, the point here is that they are presented in the context of traveltime inversion for local-scale structure.

The widespread use of body wave traveltimes in seismic tomography is undoubtedly related to the relative ease with which they may be extracted from a seismogram and the simple relationship that exists between traveltime and wavespeed. However, much more information is contained in a seismic waveform than simply the arrival time of a particular phase. Surface wave tomography, which utilizes the surface waveform component of an arriving wavetrain to build 3-D images of shear wavespeed, is the most commonly used type of waveform tomography. It is generally carried out at regional or global scales and has been particularly important in mapping beneath oceans (Nolet, 1987a); oceanic upper mantle is rarely probed by body wave tomography since few seismic recorders are placed in an ocean setting. The methodology and application of surface wave tomography falls outside the scope of this review, so we refer the interested reader to the texts of Iyer & Hirahara (1993) and Kennett (2002) and the journal papers of Cara & Lévêque (1987), Nolet (1990), Zielhuis & van der Hilst (1996) and Debayle & Kennett (2000) for more information on the subject.

This paper sets out to review commonly used methods of traveltime inversion for crustal and lithospheric imaging from the mid 1970s, when seismic tomography was first used, until recently, and provide a comprehensive list of references. However, we would like the paper to be more than just a concatenation of method descriptions and a discussion of their relative merits. This is achieved in several ways. First, we impart a tutorial flavor to the paper by being instructive as well as informative; this will be of particular benefit to readers who are not very familiar with seismic tomography. For example, many schematic diagrams and technical drawings are included to try and illustrate basic concepts, or clarify important ideas. Second, recent techniques that have seen little or no application in seismic tomography but show significant potential are also explained (e.g. the fast marching method of traveltime determination and global optimization techniques). Third, to help understand how the various methods are used in real data applications, and how different classes of data influence the formulation of the inverse problem, we present a number of case studies in detail. Generally, these examples are very recent, although we also discuss earlier applications to emphasize how the methodology has evolved. Lastly, we discuss in some detail the future of seismic traveltime tomography as a tool for subsurface imaging, and in particular the frontier areas of research that remain to be explored.

In the remainder of this section, the basic concepts underlying seismic traveltime tomography are introduced, and the four types of data for analyzing crustal and lithospheric structure are described (i.e. coincident reflection, refraction and wide-angle reflection, local earthquake and teleseismic). In Section 2, we review methods of seismic traveltime tomography, and in particular focus on model parameterization, techniques for determining traveltimes, inversion schemes and practical methods for analyzing solution robustness. Application of these methods to each of the four data types are then presented and compared in Section 3, with most examples taken from the existing literature. In Section 4, we conclude with a discussion on possible future developments in seismic traveltime tomography.

1.2 Seismic Traveltime Tomography: Formulation

If we represent some elastic property of the subsurface (e.g. velocity) by a set of model parameters **m**, then a set of data (e.g. traveltimes) **d** can be predicted for a given source-receiver array by line

integration through the model. The relationship between data and model parameters, $\mathbf{d} = \mathbf{g}(\mathbf{m})$, forms the basis of any tomographic method. For an observed dataset $\mathbf{d}_{obs}$ and an initial model $\mathbf{m}_0$, the difference $\mathbf{d}_{obs} - \mathbf{g}(\mathbf{m}_0)$ gives an indication of how well the current model predictions satisfy the data. The inverse problem in tomography is then to manipulate $\mathbf{m}$ in order to minimize the difference between observed and predicted data subject to any regularization that may be imposed. The end result will be a mathematical representation of the true structure whose accuracy will depend on a number of factors including: i) how well the observed data are satisfied by the model predictions, ii) assumptions made in parameterizing the model, iii) errors in the observed data, iv) accuracy of the method for determining model predictions $\mathbf{g}(\mathbf{m})$, and v) the extent to which the data constrain the model parameters. The tomographic method therefore depends implicitly on the general principles of inverse theory (Tarantola, 1987; Menke, 1989; Parker, 1994) .

The steps required to produce a tomographic image from seismic data can thus be defined as follows:

1. **Model parameterization:** The seismic structure of the region being mapped is defined in terms of a set of unknown model parameters. Tomographic methods generally require an initial estimate of model parameter values to be specified.

2. **Forward calculation:** A procedure is defined for the calculation of model data (e.g. traveltimes) given a set of values for the model parameters.

3. **Inversion:** Automated adjustment of the model parameter values with the object of better matching the model data to the observed data subject to any regularization that may be imposed.

4. **Analysis of solution robustness:** May be based on estimates of covariance and resolution from linear theory or on the reconstruction of test models using synthetic datasets.

In seismic traveltime tomography the model data are traveltimes and the model parameters define velocity variations. The traveltime of a ray in a continuous velocity medium $v(\mathbf{x})$ is:

$$t = \int_{L(v)} \frac{1}{v(\mathbf{x})} \mathrm{d}l \tag{1}$$

where L is the ray path and $v(\mathbf{x})$ is the velocity field. Eq. 1 is non-linear since the integration path depends on the velocity. This inherent non-linearity means that the inverse problem can be very difficult to solve. There are three basic approaches that are used, which we define as (i) linear tomography, (ii) iterative non-linear tomography, and (iii) fully non-linear tomography. In linear tomography, the relationship between traveltime residual and velocity perturbation is linearized about a reference model and corrections to the velocity field are made under this assumption. Thus, ray paths are determined only once (through the initial or reference model) and are not re-traced. Iterative non-linear tomography also ignores the path dependence of the velocity correction in the inversion step, but accounts for the non-linearity of the problem by iteratively applying corrections and re-tracing rays (i.e. repeating steps 2 and 3 in the above formulation) until, for example, the data are satisfied, or the rate of data fit improvement per iteration satisfies a given tolerance. Fully non-linear tomography locates a solution without relying on linearization in any way, but is rarely done in practice.

The linearization assumption commonly adopted in traveltime tomography is reasonable provided it can be shown that the source-receiver path is not significantly perturbed by the adjustments

made to the model parameter values in the inverse step. If we now consider a perturbation $\delta v(\mathbf{x})$ to a reference velocity field $v_0(\mathbf{x})$, so that $v(\mathbf{x}) = v_0(\mathbf{x}) + \delta v(\mathbf{x})$, then both the ray path and source-receiver traveltime must also be perturbed in the new velocity field $v(\mathbf{x})$ relative to $v_0(\mathbf{x})$. If the new path is $L(v) = L_0 + \delta L$ where L_0 is the path in $v_0(\mathbf{x})$ and $t = t_0 + \delta t$ where t_0 is the traveltime along L_0 in $v_0(\mathbf{x})$, then the traveltime in $v(\mathbf{x})$ can be written:

$$t = \int_{L_0+\delta L} \frac{1}{v_0 + \delta v} \mathrm{d}l \tag{2}$$

The integrand in Eq. 2 may be expanded using the geometric series:

$$\frac{1}{v_0 + \delta v} = \frac{1/v_0}{1 - (-\delta v/v_0)} = \frac{1}{v_0} - \frac{\delta v}{v_0^2} + \frac{\delta v^2}{v_0^3} - \dots \tag{3}$$

Substitution of this expression into Eq. 2 and ignoring second-order terms yields:

$$t = \int_{L_0+\delta L} \frac{1}{v_0} - \frac{\delta v}{v_0^2} \mathrm{d}l + O(\delta v^2) \tag{4}$$

which to first order may be separated as:

$$t = \int_{L_0} \frac{1}{v_0} - \frac{\delta v}{v_0^2} \mathrm{d}l + \int_{\delta L} \frac{1}{v_0} - \frac{\delta v}{v_0^2} \mathrm{d}l + O(\delta v^2) \tag{5}$$

The second integral on the RHS can be set to zero since Fermat's Principle states that, for fixed endpoints, the traveltime along a ray path is stationary with respect to perturbations in the path ($\partial t/\partial L = 0$). Since $t = t_0 + \delta t$, then the perturbation in traveltime resulting from a perturbation to the velocity field is given by:

$$\delta t = -\int_{L_0} \frac{\delta v}{v_0^2} \mathrm{d}l + O(\delta v^2) \tag{6}$$

The implication of Eq. 6 is that if the velocity along the path is perturbed, then the corresponding traveltime perturbation calculated along the original path will be accurate to first order (see Snieder & Sambridge, 1993, for a discussion of the case in which ray end points are perturbed). It is interesting to note that if the above calculation is performed in terms of a perturbation in slowness $s(\mathbf{x}) = 1/v(\mathbf{x})$ rather than velocity, then the resulting expression for a perturbation in traveltime is given by:

$$\delta t = \int_{L_0} \delta s \mathrm{d}l + O(\delta s^2) \tag{7}$$

which is linearly dependent on δs.

Studies that use linear tomography may deal with tens to hundreds of thousands of model parameters (e.g. Inoue et al., 1990; Spakman, 1991; van der Hilst et al., 1997; Bijwaard et al., 1998), making an iterative non-linear approach computationally expensive, or have data geometries (such as teleseismic) that make a linear assumption more valid (e.g. Achauer 1994). It is worth noting, however, that a number of recent regional and global scale traveltime tomography studies involving hundreds of thousands to millions of ray paths and hundreds of thousands of model parameters have used iterative non-linear inversion schemes (Bijwaard & Spakman, 2000; Gorbatov et al., 2000,

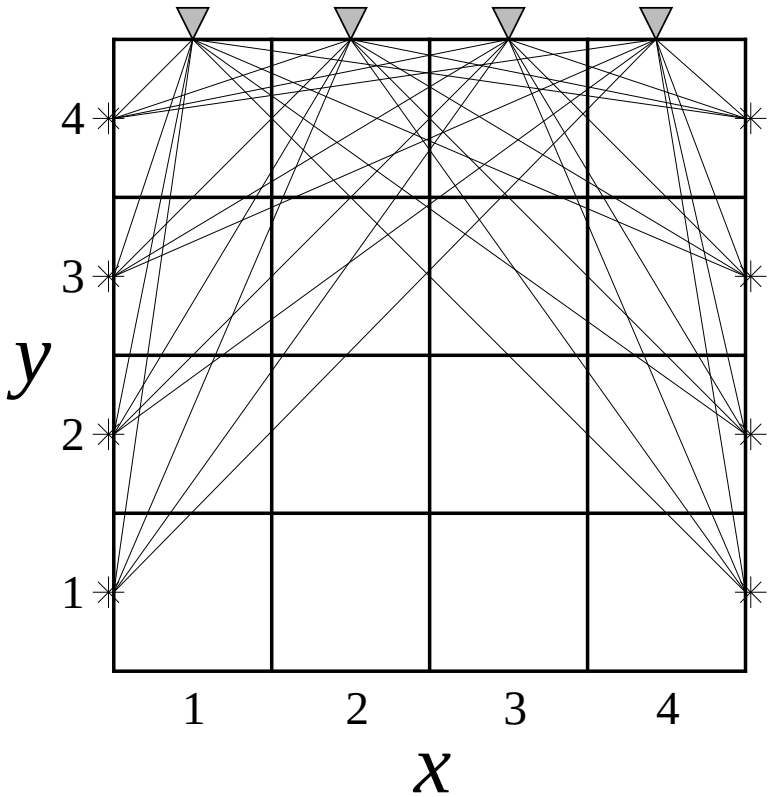

Figure 1: Schematic representation of a source-receiver array (sources denoted by asterisks, receivers by triangles) that bounds a model volume represented by a set of 16 constant velocity blocks labeled using (x, y) coordinates. Ray paths connect sources and receivers and demonstrate why the inverse problem often requires regularization. Blocks (2,1) and (3,1) are not constrained by the data, blocks (1,1), (2,2), (3,2) and (4,1) are relatively poorly constrained by the data while blocks like (2,4) and (3,4) are relatively well constrained by the data.

2001). Local earthquake and wide-angle tomography, for which ray paths depend strongly on velocity structure, generally use an iterative non-linear approach (e.g. Hole, 1992; Graeber & Asch, 1999).

Common model parameterizations include constant velocity (or slowness) blocks, a grid of velocity nodes with a specified interpolant like trilinear or cubic splines, and to a lesser extent, spectral parameterizations like truncated Fourier series. Interface parameterizations use similar schemes except those appropriate to a surface rather than a volume. The forward problem of finding source-receiver ray paths and traveltimes is often solved using ray tracing techniques like shooting and bending, first-arrival wavefront tracking on a grid (e.g. eikonal solvers) and network methods, also known as Shortest Path Raytracing (SPR). The inverse step of adjusting model parameters to satisfy observed data subject to regularization constraints is frequently solved using gradient methods (e.g. Gauss-Newton, subspace inversion) and backprojection methods like ART or SIRT. Global optimization methods like genetic algorithms (Goldberg, 1989) have been used but only rarely. Regularization (i.e. applying other constraints to the model in addition to those supplied by the data) is an important consideration in the inverse step due to the often under-determined or mixed-determined nature of the problem (see Fig. 1). A means of analyzing solution robustness is a critical step in a meaningful interpretation of an inversion result. Estimates of *a posteriori* model resolution and covariance from linear theory have been used, as have synthetic reconstructions (e.g. the checkerboard

test) that use the same source-receiver geometry as the original experiment. All the above methods are discussed in more detail in Section 2.

Due to its origins in radiology and early seismic imaging, the term tomography is generally only applied to methods that invert for a property (e.g. velocity) that is described throughout the model volume. If only interface structure is inverted for (e.g. Hole, 1992; Rawlinson & Houseman, 1998), then the term tomography is generally not applied, even though the method conforms to the procedures described above. However, in this paper we assume that the term tomography may be used for any method that generally follows the four steps listed above and results in 2-D or 3-D representations of subsurface structure.

1.3 Traveltime Data used in Studies of the Crust and Lithosphere

The four classes of data we consider for the tomographic determination of crustal and lithospheric structure are normal incidence reflection, refraction and wide-angle reflection, local earthquake and teleseismic. Normal incidence reflection surveys and refraction and wide-angle reflection surveys use controlled or artificial sources (e.g. airgun shots, explosions, vibroseis) to generate seismic energy. The benefits of controlled sources include precise identification of source location and origin time, control over data coverage, and knowledge of source waveform. However, compared to experiments that use earthquake sources, surveys of this nature tend to be more expensive and cannot be used to probe deep structure (air-gun shots tend not to penetrate very far beyond the base of the crust) unless powerful sources such as nuclear explosions are used. PNEs (Peaceful Nuclear Explosions) can be detected thousands of kilometers away and have occasionally been recorded by long in-line receiver arrays (Priestley et al., 1994; Ryberg et al., 1996; Nielsen et al., 1999). Due to the advent of the Comprehensive Test Ban Treaty, PNEs are unlikely to be used in the future.

Normal incidence reflection seismic surveys use in-line arrays of sources and receivers to image crustal structure on depth scales of tens of meters to tens of kilometers. This seismic method is referred to as *normal incidence* (or *coincident*) because the aim is to map reflections from sub-horizontal interfaces using near-vertically propagating P-waves (see Fig. 2a). The most common way of analyzing this data is to plot traces from adjacent sources or receivers next to each other on a horizontal distance versus two-way traveltime axis to produce a reflection section. Variations in the earth's impulse response can then be linked to variations in earth structure. To produce a usable reflection section, significant processing is required (Yilmaz, 1987; Telford et al., 1990) to remove or reduce effects caused by the source wavetrain, multiples, normal moveout, near-surface complexities, diffraction and data noise. The entire seismic wavefield is used and is mapped directly to the model space. For this reason, reflection seismic sections are unequaled in terms of detail and resolution. The method is most effective when used to image discontinuous changes in seismic structure (i.e. reflectors).

As shown in Fig. 2a, the paths taken by the seismic energy in a seismic reflection experiment are not strictly vertical, as the shot is recorded by an array of receivers at varying offsets (usually small compared to the maximum depth of the recorded energy) from the source. The multiplicity of data permits stacking to boost signal-to-noise ratio and the variation of ray trajectories also makes it possible to explicitly image elastic properties using traveltimes in a process known as reflection tomography (Bishop et al., 1985; Farra & Madariaga, 1988; Williamson, 1990; Blundell, 1993; Carroll & Beresford, 1996; Kosloff et al., 1996). Structure is commonly represented by a series of sub-horizontal layers separated by continuous interfaces. Usually, both interface geometry and layer velocity are varied by the inversion to satisfy the observed traveltimes.

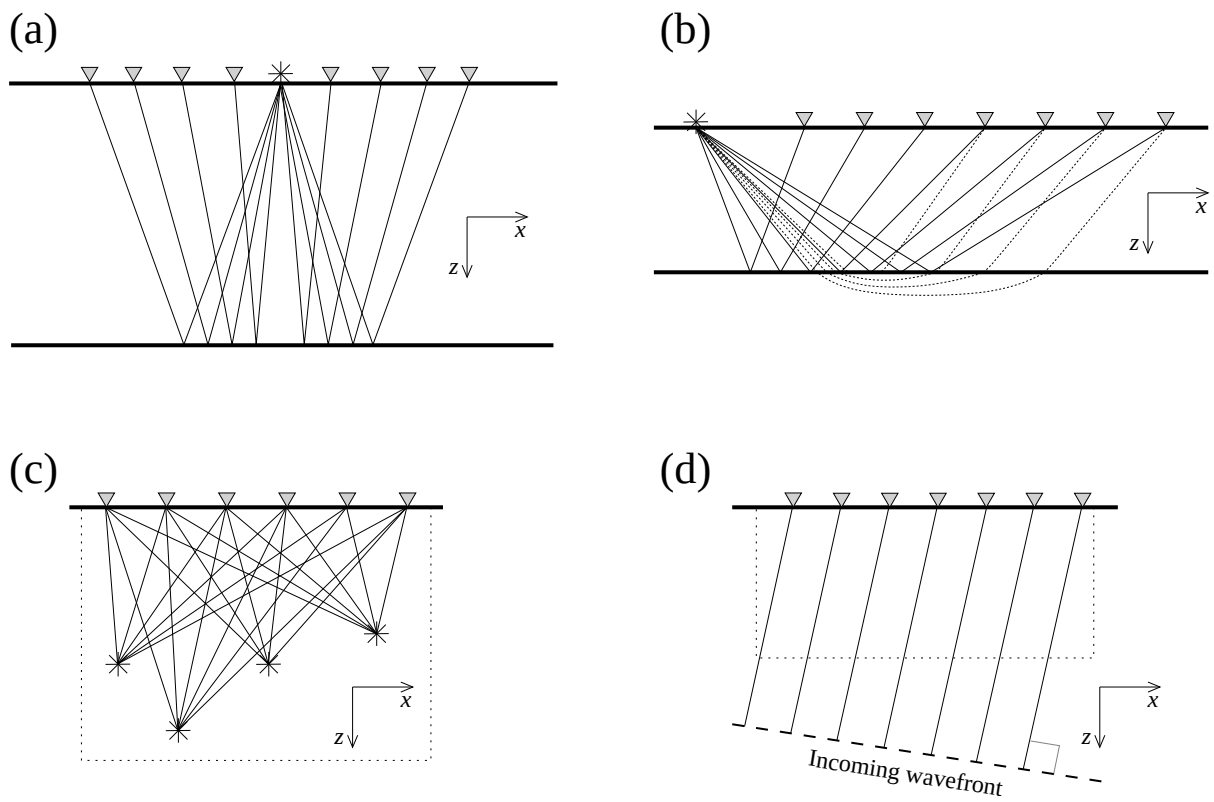

Figure 2: Schematic source-receiver geometries for various seismic surveys (sources are denoted by asterisks, receivers by triangles): (a) Normal incidence reflection array (common source), (b) wide-angle seismic array; thin black lines denote reflected rays, dotted lines denote refracted rays, (c) local earthquake array - sources lie within modeled region (dotted line), and (d) teleseismic array - sources lie outside modeled region (dotted line).

Although it is sometimes possible for coincident reflection data to resolve the trade-off that exists between lateral variation of velocity within a layer and interface depth (Blundell, 1993), the often small offsets between source and receiver make it difficult. Refraction and wide-angle reflection surveys use arrays with large offsets between sources and receivers (see Fig. 2b). In this case, traces from adjacent receivers recorded from a single shot, or from adjacent shots recorded at a single receiver, may be plotted on a time versus offset plot to reveal the presence of coherent traveltime curves. The term *wide-angle data* normally implies the presence of both refraction and reflection information (e.g. Riahi & Lund, 1994; Zelt et al., 1996). Wide-angle tomography shares many similarities with reflection tomography. Model structure is commonly represented by layers with both interface geometry and layer velocity constrained by the inversion procedure. Reflected rays tend to be more sensitive to variations in interface structure, while refracted rays tend to be more sensitive to variations in layer velocity.

Recently, interpretation of wide-angle seismic data has been carried out in 3-D using multiple in-line arrays of sources and large receiver arrays. Inversion methods for 3-D wide-angle data have been presented by a number of authors including Hole (1992), Hole et al. (1992), Riahi et al. (1997), Zelt & Barton (1998), Zelt et al. (1999) and Rawlinson et al. (2001a). Interpretation methods for such datasets have developed rapidly in recent years, although many still cannot deal with the level of structural detail commonly obtained from 2-D datasets (e.g. Kodaira et al., 1998).

One of the first papers to be published on seismic tomography was that of Aki & Lee (1976), who

inverted first-arrival P-wave traveltimes from local earthquakes for velocity structure and hypocenter location in Bear Valley, California. The source-receiver geometry for this type of study is shown schematically in Fig. 2c - the earthquake sources lie beneath the receiver array within the model volume. The hypocenter coordinates, which are not accurately known, must be included in the inversion. Although Fig. 2c shows a 2-D experiment, most local earthquake studies are 3-D. Since the publication of the Aki & Lee (1976) paper, this branch of tomography has come into common usage and is now popularly known as Local Earthquake Tomography or LET (Thurber, 1993).

LET has been used to image the lithosphere and upper asthenosphere to depths of up to 200 km in subduction zone settings (Abers, 1994; Graeber & Asch, 1999). High resolution images of the crust have also been obtained using shallow earthquakes (Thurber, 1983; Chiarabba et al., 1997). In such cases, the results of LET can be usefully compared with wide-angle studies of the same region (Eberhart-Phillips, 1990). Advantages of LET over wide-angle tomography include greater depth of penetration and the added structural information provided by the relocated hypocenters, e.g. the existence of double seismic zones (Hasegawa et al., 1978; Kawakatsu, 1985; Kao & Rau, 1999). On the other hand, the relocation of hypocenters adds to the non-uniqueness of the solution and phases other than first-arrival P and S can be difficult to incorporate. For this reason, LET models rarely include interfaces, although Zhao et al. (1992) included interfaces in their LET model of NE Japan by using observed SP waves converted at the Moho and PS/SP waves converted at the upper boundary of the subducted Pacific Plate.

A significant difference between LET and teleseismic tomography is source location; in a teleseismic study, earthquakes are generally thousands of km away from the receiver array. The target region of the crust and upper mantle lies beneath the receiver array. A key assumption of teleseismic tomography is that only the region beneath the receiver array contains significant lateral variations in velocity. Elsewhere, a 1-D earth model is adequate to predict the geometry and inclination of the wavefront before it strikes the target region. Therefore, it is possible to trace the rays through a 1-D reference model of the earth until they penetrate the teleseismic model. Normally, relative traveltime residuals rather than absolute traveltimes are used in the inversion; this helps to account for errors in hypocenter location and large scale mantle heterogeneities. Fig. 2d shows a schematic diagram of a wavefront from a distant earthquake incident on a teleseismic receiver array. As in LET, teleseismic tomography is usually carried out in 3-D.

A seminal paper by Aki et al. (1977) used teleseismic data recorded at the Norsar array to invert for velocity anomalies to a depth of 126 km. The final solution was produced by linear inversion. The assumption of linearity is more accurate in teleseismic tomography than in LET or wide-angle inversion. This occurs because the ray paths tend to be near-vertical as they transmit through the model volume and hence are less affected by the dominant variations of velocity with depth. Consequently, many teleseismic tomography images, even those published recently, are the result of linear inversions (Humphreys & Clayton, 1990; Glahn & Granet, 1993; Seber et al., 1996; Saltzer & Humphreys, 1997). If the traveltime residuals are suggestive of significant lateral structure, then an iterative non-linear approach may be required (Weiland et al., 1995; McQueen & Lambeck, 1996; Rawlinson & Houseman, 1998; Frederiksen et al., 1998; Steck et al., 1998; Graeber et al., 2002).

The main difficulties in teleseismic tomography arise because of the irregular and unpredictable nature of earthquakes. Earthquakes tend to occur at plate boundaries, so it is common to have a very uneven distribution (in terms of azimuth and inclination) of ray paths through the target volume. Another factor is that relative traveltime residuals only provide good constraints on lateral variations in velocity relative to an *a priori* lateral average. Vertical variations in the velocity field of the

solution model are poorly constrained and therefore must be interpreted with caution (Lévêque & Masson, 1999).

The depth extent of teleseismic investigations may range from crustal (e.g. Lambeck et al., 1988; Rawlinson & Houseman, 1998) to many hundreds of kilometers (e.g. Humphreys & Clayton, 1990). The horizontal extent of the receiver array and the source distribution determines the depth to which features may be resolved. The vertical dimension of the model volume is often chosen on this basis, but it is always possible that structure outside the solution region causes some of the variation between traveltime residuals (e.g. Benz et al., 1992).

2 Methods of Traveltime Inversion

2.1 Representation of Structure

The traveltime of a seismic wave between source and receiver is solely dependent on the velocity structure of the medium through which the wave propagates. Therefore, subsurface structure in a seismic traveltime inversion is represented by variations in P or S wave velocity (or slowness). As mentioned in Section 1.2, velocity variations may be defined by a set of interfaces whose geometry is varied to satisfy the data, a set of constant velocity blocks or nodes with a specified interpolation function, or a combination of velocity and interface parameters. The most appropriate choice will depend on the *a priori* information (e.g. known faults or other interfaces), whether or not the data indicates the presence of interfaces (e.g. reflections, mode conversions), whether data coverage is adequate to resolve the trade-off between interface position and velocity, and finally, the capabilities of the inversion routine.

2.1.1 Velocity Parameterization

Constant velocity blocks (Fig. 3a) are simple to define and result in linear ray paths within each block. On the other hand, they are not a natural choice for representing smooth variations in subsurface structure due to the velocity discontinuities that exist between adjacent blocks. These artificial discontinuities may also cause unwarranted ray shadow zones and triplications. However, if a large number of blocks are used and restrictions are placed on the size of the velocity changes between adjacent blocks, then a reasonable approximation to a continuously varying velocity field is possible. In teleseismic tomography, constant velocity blocks have been used by many authors including Aki et al. (1977), Oncescu et al. (1984), Humphreys & Clayton (1988), Humphreys & Clayton (1990), Benz et al. (1992), Achauer (1994) and Saltzer & Humphreys (1997). In wide-angle traveltime inversions, the use of constant velocity blocks is not as common. Zhu & Ebel (1994) and Hildebrand et al. (1989) use constant velocity blocks in the inversion of 3-D refraction traveltimes while Williamson (1990) and Blundell (1993) use them in an inversion of reflection traveltimes. Similarly, constant velocity blocks are only rarely used in local earthquake tomography (Aki & Lee, 1976; Nakanishi, 1985). This scheme for representing structure is often avoided when strong ray curvature is expected.

An alternative to a block parameterization is to define velocities at the vertices of a rectangular grid (see Fig. 3b) together with a specified interpolation function. One of the first examples of this approach was by Thurber (1983) in the context of local earthquake tomography. To describe the velocity at any point (x, y, z) within a rectangular grid of nodes, he used a trilinear interpolation

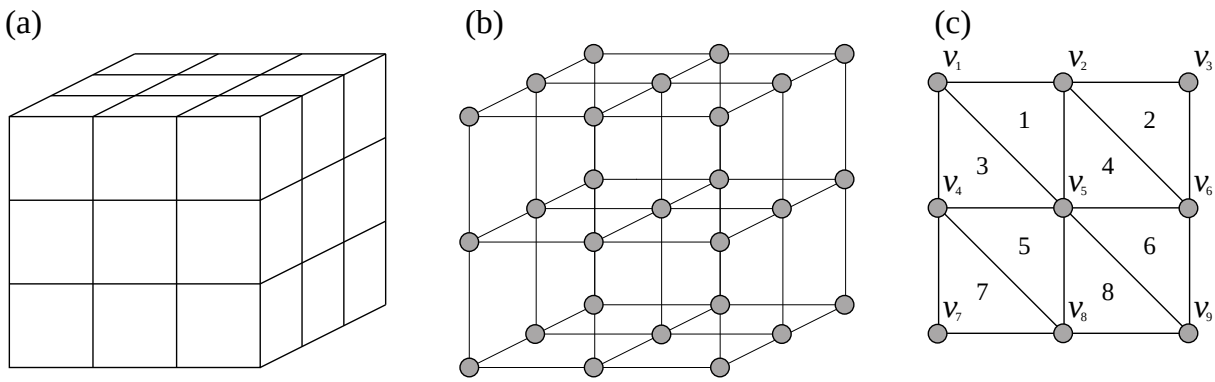

Figure 3: Different types of velocity parameterization: (a) constant velocity blocks, (b) a grid of velocity nodes, and (c) triangulated velocity grid designed for constant velocity gradient cells (after White, 1989).

function:

$$v(x, y, z) = \sum_{i=1}^{2}\sum_{j=1}^{2}\sum_{k=1}^{2} V(x_i, y_j, z_k)\left(1 - \left|\frac{x - x_i}{x_2 - x_1}\right|\right)\left(1 - \left|\frac{y - y_j}{y_2 - y_1}\right|\right)\left(1 - \left|\frac{z - z_k}{z_2 - z_1}\right|\right) \quad (8)$$

where $V(x_i, y_j, z_k)$ are the velocity values at the eight grid points surrounding (x, y, z). The use of Eq. 8 ensures that the velocity field will be continuous throughout the model volume, although the velocity gradient will be discontinuous from cell to cell. This model parameterization is now commonly used in local earthquake tomography (Eberhart-Phillips, 1986, 1990; Zhao et al., 1992; Eberhart-Phillips & Michael, 1993; Scott et al., 1994; Haslinger et al., 1999), presumably because most of these inversions are based on the SIMULPS code devised by Thurber (1983). Zhao et al. (1994) and Steck et al. (1998) have used this parameterization in teleseismic tomography, although Zhao et al. (1994) use a modified form of Eq. 8 for spherical coordinates.

Higher order interpolation functions must be used if the velocity field is to have continuous first and second derivatives, which are required for some ray tracing methods (Thomson & Gubbins, 1982). Cubic spline interpolation results in continuous first and second derivatives and has been used by a number of authors. Thomson & Gubbins (1982), in a NORSAR teleseismic study, use the following cubic spline function to describe the slowness field within a regular 3-D spherical grid of nodes:

$$s(r, \theta, \phi) = \sum_{i=1}^{4}\sum_{j=1}^{4}\sum_{k=1}^{4} S_{ijk} C_i(R) C_j(\Theta) C_k(\Phi) \quad (9)$$

where S_{ijk} are the slowness values at the nodes of the $4 \times 4 \times 4$ grid surrounding the point (r, θ, ϕ). $C_i(R)$, $C_j(\Theta)$ and $C_k(\Phi)$ are known as the cardinal splines (modified by Thomson & Gubbins, 1982, for local support) and R, Θ and Φ are the local coordinates of r, θ and ϕ. Nodal values do not necessarily equal the spline values at the node points. Sambridge (1990) uses a similar parameterization in Cartesian coordinates to describe a 3-D model constrained by traveltimes from local earthquakes and explosions.

Cubic B-splines are similar to the cardinal splines described above in that they are locally supported and do not necessarily pass through the node values. Conventional cubic spline interpolation forces the spline to pass through node values and is not locally supported. Undesirable effects of non-local support include poorly resolved portions of the model having a global influence and unrealistic velocity fluctuations between nodes (Shalev, 1993). In 2-D wide-angle traveltime tomography, cubic

spline interpolation has been employed by Lutter et al. (1990), while Farra & Madariaga (1988) and McCaughey & Singh (1997) have used cubic B-spline bases.

An interpolation method which features inherent flexibility is splines under tension (Smith & Wessel, 1990). Here, a tension factor is used to control the mode of interpolation, which, in the case of Neele et al. (1993), can vary between near trilinear interpolation and cubic spline interpolation. The scheme features continuous first and second derivatives. Usually, one will choose a tension factor that results in a smooth model but minimizes unrealistic oscillations and maximizes local control. Neele et al. (1993), VanDecar et al. (1995) and Ritsema et al. (1998) all use this approach in the inversion of teleseismic traveltimes.

A method of parameterization that goes some way towards bridging the gap between a block approach and a grid approach is one which uses cells with a constant velocity gradient. White (1989) describes a method of 2-D refraction tomography in which a rectangular grid of nodes is used to define triangular regions of constant velocity gradient (see Fig. 3c). The velocity within each cell is given by:

$$v(x, y) = v_o + (x - x_o)\nabla_x v + (z - z_o)\nabla_z v \tag{10}$$

where v_o, $\nabla_x v$ and $\nabla_z v$ are determined using the velocities at the vertices of the triangle (e.g. $v_o = v_1$, $\nabla_x v = (v_2 - v_1)/\Delta x$, $\nabla_z v = (v_5 - v_2)/\Delta z$ in cell 1 of Fig. 3c). The attributes of this method are that velocity varies continuously throughout the medium and rays can be traced analytically within each cell. However, the velocity gradient is discontinuous at each cell boundary which, in conjunction with their triangular shape, can result in difficulties in finding the source-receiver ray path. A similar method was used by Chapman & Drummond (1982) for refracted rays. Another interpolation function that can be used with the triangular cell geometry which also allows analytic ray tracing is the constant gradient of quadratic slowness $1/v^2$ (Červený, 1987). The extension of these procedures to 3-D involves the use of tetrahedral cells, with the linear interpolation functions described in terms of the velocity nodes at the four vertices of the tetrahedron. It is important to note that these methods of parameterization are used primarily because they facilitate analytic ray tracing and only secondly because they provide an adequate approximation to actual subsurface velocity distribution, which is inevitably more complex.

Rather than use a block or grid-based parameterization, one could use a scheme in which velocity is discretized in the wavenumber domain rather than the spatial domain. Spectral parameterizations that use some form of truncated Fourier series fall into this category. In their inversion of reflection amplitudes and traveltimes, Wang & Pratt (1997) describe a 2-D slowness distribution using the Fourier series:

$$\begin{aligned} s(\mathbf{r}) &= a_{00} + \sum_{m=1}^{N}[a_{m0}\cos(\mathbf{k}\cdot\mathbf{r}) + b_{m0}\sin(\mathbf{k}\cdot\mathbf{r})] \\ &\quad + \sum_{m=-N}^{N}\sum_{n=1}^{N}[a_{mn}\cos(\mathbf{k}\cdot\mathbf{r}) + b_{mn}\sin(\mathbf{k}\cdot\mathbf{r})] \end{aligned} \tag{11}$$

where $\mathbf{r} = x\mathbf{i} + z\mathbf{j}$ and $\mathbf{k} = m\pi k_0\mathbf{i} + n\pi k_0\mathbf{j}$ are the position and wavenumber vector respectively, and a_{mn} and b_{mn} are the amplitude coefficients of the $(m, n)^{th}$ harmonic term, which become the unknowns in the inversion step. The advantage of this type of parameterization is that it defines a velocity field which is infinitely differentiable and whose smoothness can be controlled through the choice of the number of harmonic terms N that are used. Eq. 11 defines a non-local parameterization, however, so poorly resolved portions of a solution model may have a detrimental effect on other

regions of the model. Spectral parameterizations have been used in wide-angle traveltime inversion by Hildebrand et al. (1989), Hammer et al. (1994) and Wiggins et al. (1996) to study crustal structure beneath deep oceans.

2.1.2 Including Interfaces

Velocity discontinuities are most commonly included in velocity models when the subsurface is represented by sub horizontal layers (see Fig. 4). In 2-D and 3-D traveltime inversion, the use of layered parameterizations has almost exclusively been the domain of reflection and refraction tomography. Reflection sections only image reflectors and refraction sections usually contain obvious later-arriving traveltime curves associated with velocity discontinuities. The issues related to choosing an appropriate interface parameterization are similar to those for choosing an appropriate velocity parameterization - representation of the geological structure and suitability for use in the forward and inverse solution steps.

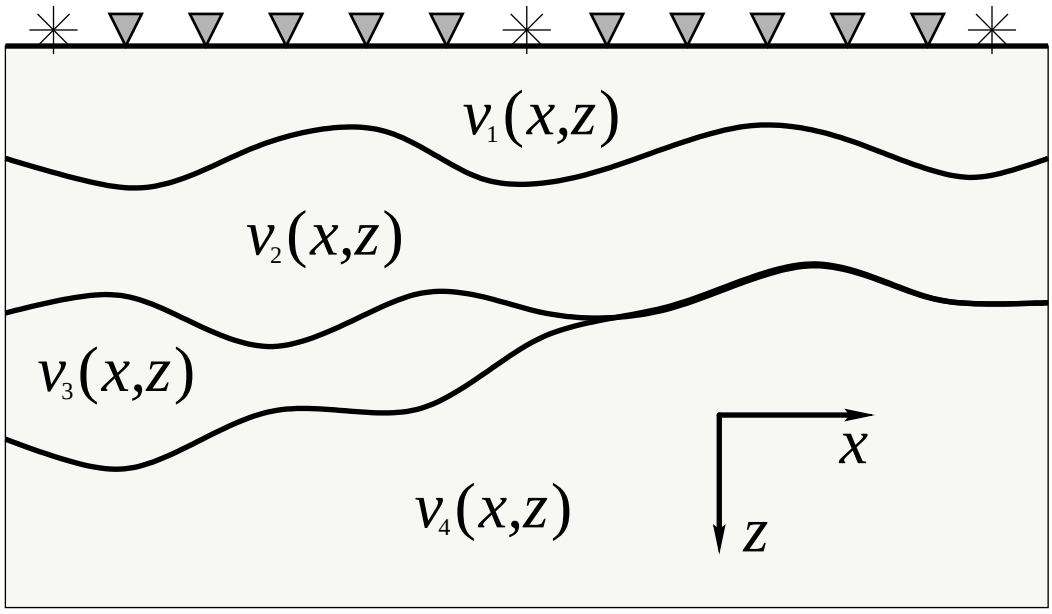

Figure 4: Schematic representation of the kind of layered velocity structure that can be imaged in reflection and wide-angle traveltime inversion. The velocity functions $v_i(x, z)$ describe the velocity variation within a layer.

In 2-D, piecewise linear segments (see Fig. 5a) are probably the simplest means of representing interface geometry. The wide-angle method of Zelt & Smith (1992) and the reflection method of Williamson (1990) employ this type of interface parameterization. One obvious problem with using piecewise linear segments is that the gradient of the interface is discontinuous at the joins between segments. Such discontinuity may not be geologically realistic and will create artificial shadow zones because incident rays with very similar paths may depart from the interface along very different paths if they intersect the interface on either side of a point of gradient discontinuity. Zelt & Smith (1992) avoid this problem by using an averaging filter to smooth the interface so that the departing ray has a trajectory consistent with the smooth interface, but the point of projection is still given by the intersection point of the incident ray with the piecewise linear interface.

The logical extension of piecewise linear segments to a 3-D model with interface surfaces is

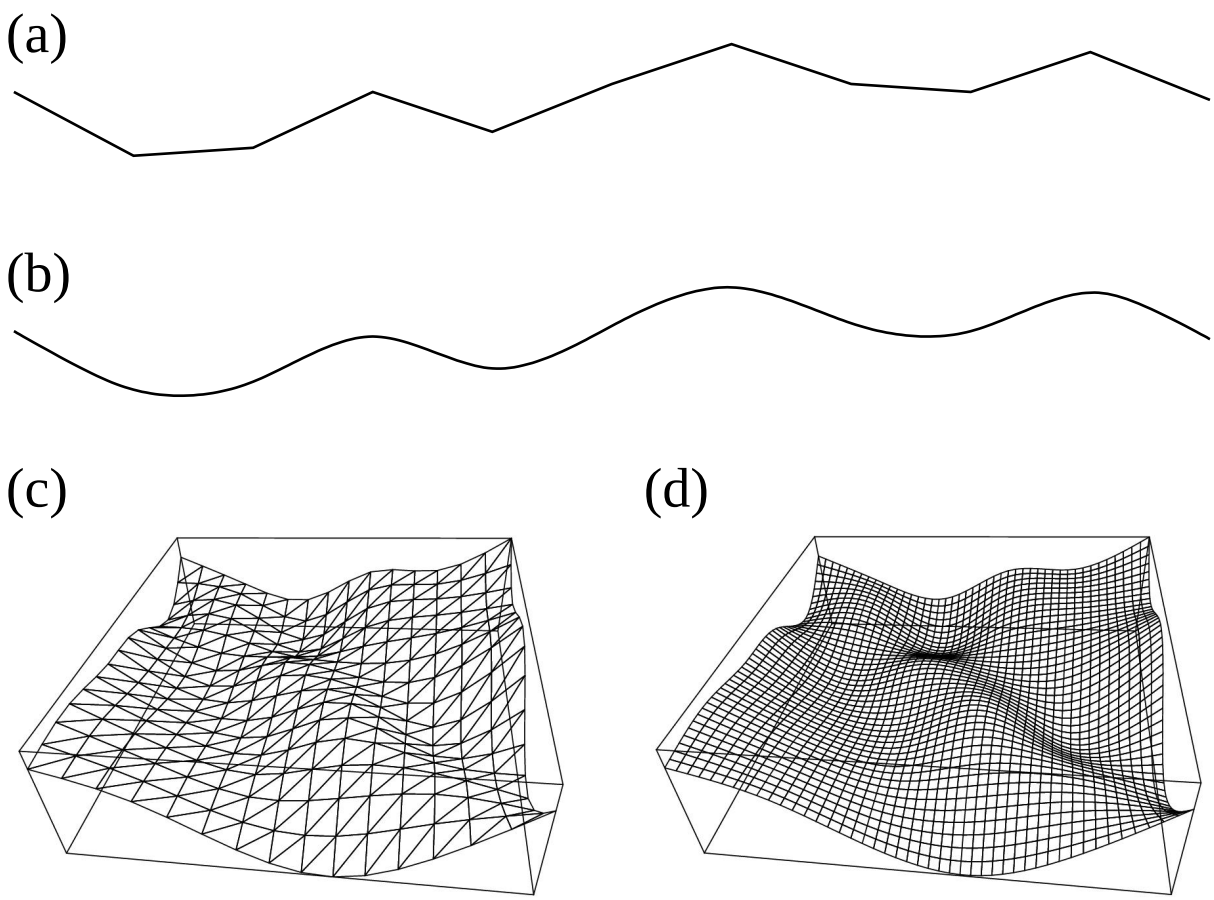

Figure 5: Types of interface parameterization for 2-D (a-b) and 3-D (c-d) models. (a) Piecewise linear segments, (b) piecewise cubic B-spline interpolation, (c) surface defined by mosaic of triangular patches, (d) surface defined by mosaic of bicubic B-spline patches - note that the surface is visualized here by orthogonal sets of lines.

to use piecewise triangular area segments (see Fig. 5c). Sambridge (1990) has used this approach in the inversion of local earthquake and quarry blast traveltimes. Guiziou et al. (1996) also used a triangulated interface structure in the tomographic inversion of reflection traveltimes in order to work with geological models created in GOCAD (a computer aided design tool for modeling geological objects). One advantage of triangulation is that multi-valued surfaces are easily described.

Analogous to defining velocity on a grid of nodes, an interface may be described by a grid of depth nodes with a specified interpolation function (piecewise linear segments can be viewed as a special case of this). Piecewise cubic spline functions with C_2 continuity (see Fig. 5b) are commonly used in wide-angle inversions. Conventional cubic splines have been used in a number of 2-D schemes including those by White (1989), Lutter & Nowack (1990) and Rawlinson & Houseman (1998). Červený et al. (1984) parameterize a layered model with splines under tension for both interfaces and layer velocity fields. Cubic B-splines, which feature local control of interface geometry, have been used by Farra & Madariaga (1988) and McCaughey & Singh (1997). In 3-D, the use of smooth interfaces (see Fig. 5d) is less common, mainly because methods for the inversion of 3-D layered structures are less wide spread. The ray tracing method of Gjøystdal et al. (1984) parameterizes interfaces in terms of bicubic splines, and the reflection tomography method of Chiu et al. (1986) describes interfaces using n^{th} order polynomials (in practice, they use $n \leq 3$). Like con-

tinuous velocity variations, interfaces are also amenable to spectral parameterization. For example, Wang & Houseman (1994) describe interfaces using a truncated Fourier series where the number of harmonic terms controls the allowable flexibility of the interface. The problems associated with the use of a global parameterization outlined above for velocity also apply to interface structure.

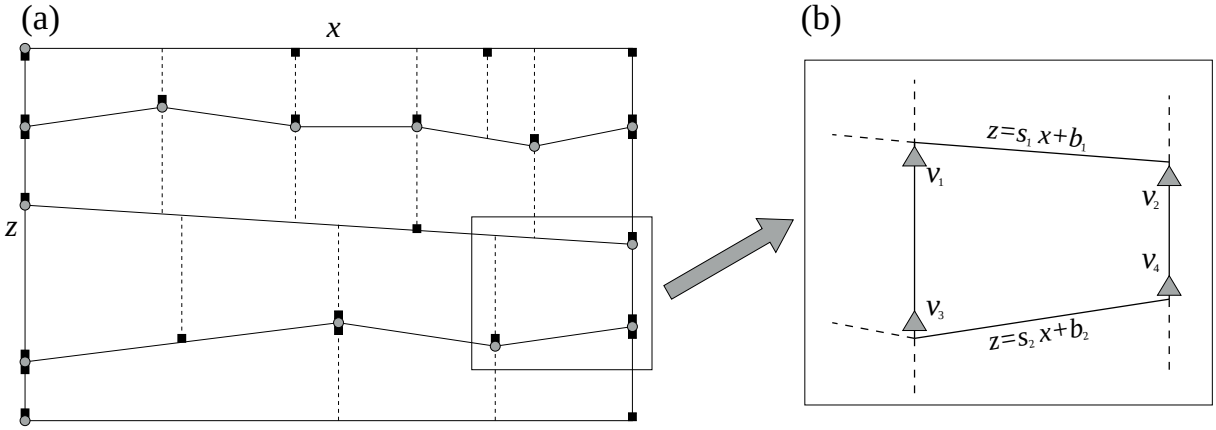

Figure 6: Example of a layered structure parameterized using the method of Zelt & Smith (1992). (a) Structure composed of four layers with layer velocity defined by 25 velocity nodes (black squares) and interface structure defined by 14 boundary nodes (grey circles). There is no velocity discontinuity between the second and third layers and dashed lines indicate the lateral boundaries of trapezoids. (b) The velocity field within a trapezoid is defined by interpolating between four corner values (grey triangles). If no velocity node lies at a corner, then the required value is obtained by linear interpolation from adjacent nodes.

Irregular parameterizations are not very commonly used in traveltime tomography. However, irregular shaped elements can be adapted to suit variations in subsurface data coverage. The frequently used 2-D wide-angle inversion method of Zelt & Smith (1992) uses such a method. Layer boundaries are described by a set of one or more arbitrarily spaced nodes interpolated linearly. Within each layer, velocity nodes are specified on the upper and lower boundaries, the number and spacing of which may vary (Fig. 6a). To facilitate velocity interpolation, each layer is divided laterally into trapezoidal blocks separated by vertical boundaries, which occur at each upper and lower boundary node and velocity node. The velocity within each trapezoid is interpolated using the velocity values at the four corners (obtained by linear interpolation if a velocity node does not occupy a corner) such that the velocity within each layer varies linearly between the upper and lower boundaries in the vertical direction, and linearly along the upper and lower boundaries between nodes. Layer boundaries may or may not represent velocity discontinuities. Fig. 6b shows the design of a trapezoid in more detail. The velocity within a trapezoid is given by (Zelt & Smith, 1992):

$$v(x, z) = \frac{c_1 x + c_2 x^2 + c_3 z + c_4 xz + c_5}{c_6 x + c_7} \tag{12}$$

where $\{c_i\}$ are linear combinations of the corner velocities. The inherent flexibility of this technique means that a velocity structure with or without layering can be represented. If layers are present, then it is possible for the velocity within the layers to vary arbitrarily.

Another example of irregular parameterization is given by Rawlinson et al. (2001a) in their method of wide-angle traveltime inversion for 3-D layered crustal structure. They make use of bicubic B-splines in parametric form to describe interface geometry. For a set of control vertices

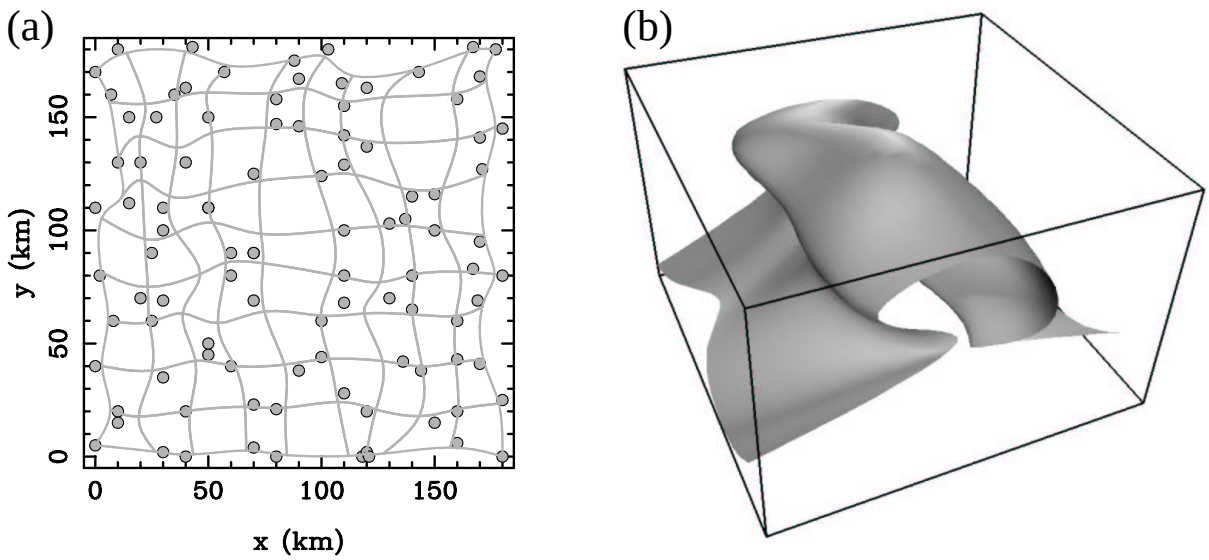

Figure 7: Flexibility of a cubic B-spline parameterization in parametric form. (a) Irregular grid of nodes describing a surface. Grey lines indicate surface patch boundaries. (b) Multivalued surface described by cubic B-splines.

$\mathbf{p}_{i,j} = (x_{i,j}, y_{i,j}, z_{i,j})$ where $i = 1, \ldots, m$ and $j = 1, \ldots, n$, the B-spline for the ij^{th} surface patch is:

$$\boldsymbol{B}_{i,j}(u, v) = \sum_{k=-1}^{2} \sum_{l=-1}^{2} b_k(u) b_l(v) \mathbf{p}_{i+k,j+l} \tag{13}$$

so that any point on a surface patch is a function of two independent parameters u ($0 \leq u \leq 1$) and v ($0 \leq v \leq 1$). The weighting factors $\{b_i\}$ are the uniform cubic B-spline basis functions (Bartels et al., 1987). Properties of a surface constructed using a mosaic of B-spline patches defined by Eq. 13 include C_2 continuity and local control. Fig. 5d shows a bicubic B-spline surface described by a regular grid of 16×16 vertices. In addition to inherent smoothness and local control, vertices may lie on an irregular grid (Fig. 7a) and the surface may also be multivalued (Fig. 7b), although Rawlinson et al. (2001a) did not make use of the latter property in their inversions. Control vertices may be widely spaced in regions of poor data coverage and closely spaced in regions of good data coverage.

Fault surfaces are another common feature of earth structure that may need to be represented in a model. Explicit representation of faults is not common in seismic methods. Faults are often near-vertical and cause discontinuities in the interfaces and velocity fields of sub-horizontal layers offset by the fault. Ray-tracing through such a medium may be difficult and the inverse problem is likely to be highly non-linear. Lambeck et al. (1988) and Lambeck & Burgess (1992) computed teleseismic traveltime residuals for a 2-D model in which constant velocity layers bounded by piecewise linear interfaces were not required to be laterally continuous across the model, thus allowing faults to be defined. Inversion of traveltime residuals was not attempted in either study, however. Wide-angle inversion methods that allow for complex lateral structure (e.g. Zelt & Smith, 1992) are usually not designed to represent faults. Laterally discontinuous structures like those found at subduction zones (Zelt, 1999) can be represented because interfaces tend to have a shallow dip and layer dislocation is not usually needed. Wang & Braile (1996), using a method based on that of Zelt & Smith (1992), represent faults implicitly by sharp near-vertical jumps in sub-horizontal interface structure in adja-

cent interfaces. However, these structures were constrained manually during the inversion process.

When interfaces are used in conjunction with layer velocities specified by blocks or a grid of nodes, then it is usually necessary to extrapolate the velocity field of each layer beyond the surrounding interfaces. These velocity parameters are redundant unless changes in interface geometry made by the inversion step cause the relevant nodes to lie within the layer. The velocity within each layer is usually defined to be independent of velocities in other layers, so any spatial overlap of velocity nodes from adjacent layers is of no consequence. In a 3-D wide-angle inversion study, Zelt et al. (1999) describe velocity structure using a continuous velocity parameterization but include "floating" reflectors. These floating reflectors allow reflections to be used to constrain interface structure and velocity but simplify traveltime determination by associating the reflectors with sharp gradients in velocity rather than with velocity discontinuities.

As mentioned earlier, the use of interfaces in teleseismic traveltime tomography is rare. Zhao et al. (1994) employ fixed interfaces described either by a power series or piecewise linear interpolation in their simultaneous inversion of local, regional and teleseismic traveltimes for velocity variation. Davis (1991) uses backprojection to invert teleseismic traveltime residuals for the structure of the lithosphere-asthenosphere boundary in East Africa, defining interface geometry in terms of a polynomial expansion. Kohler & Davis (1997) use a similar procedure to determine 2-D crustal thickness variations in California from teleseismic traveltime residuals.

2.2 Traveltime Determination

The calculation of ray traveltimes between known end points through a given velocity structure is often called the forward problem. When more than one ray path exists between a given source and receiver, the path with minimum traveltime is the one usually required because first-arrivals are always easier to identify on a seismogram. Often, other quantities such as Fréchet derivatives are calculated together with traveltime, but these parameters and their methods of computation are described in the next section in the context of the relevant inversion method.

The traveltime of a seismic wave between source S and receiver R is given by the integral:

$$t = \int_S^R \frac{1}{v(\mathbf{x})} \mathrm{d}l \tag{14}$$

where $\mathrm{d}l$ is differential path length, $\mathbf{x}$ is the position vector and v is velocity. The difficulty in performing this integration, as pointed out in Section 1.2, is that the path taken by the seismic energy depends on the velocity structure, and the path needs to be known in order to evaluate the integral. For an elastic medium, the propagation of seismic wavefronts can be described by the eikonal equation:

$$(\nabla_{\mathbf{x}} T)^2 = \frac{1}{[v(\mathbf{x})]^2} \tag{15}$$

where T is the traveltime of the wavefront. This description of wave propagation is subject to the so-called high frequency assumption: the wavelength of a seismic wave should be much less than the length scale of the velocity variations of the medium through which it passes. If traveltime is described by the equation $T = T(\mathbf{x})$ and time is held constant, then $T_A = T(\mathbf{x})$ is an implicit equation for the wavefront at time T_A (Aki & Richards, 1980). If T_A is increased to, say, T_B, then the equation $T_B = T(\mathbf{x})$ will describe the new geometry and position of the wavefront at a time $T_B - T_A$ later. If instead a point of constant phase on the wave is described as $\mathbf{x} = \mathbf{x}(T)$ then,

rather than implicitly describing a wavefront (i.e. a surface), we explicitly describe a ray path (i.e. a curve). Ray paths are by definition everywhere normal to wavefronts. The equation that governs the geometry of ray paths can be derived from the eikonal equation by considering how a small change in time dt effects a point $\mathbf{x}$ on a wavefront (see Aki & Richards, 1980). The resultant *ray equation*:

$$\frac{\mathrm{d}}{\mathrm{d}l}\left(\frac{1}{v(\mathbf{x})}\frac{\mathrm{d}\mathbf{x}}{\mathrm{d}l}\right) = \nabla\left(\frac{1}{v(\mathbf{x})}\right) \tag{16}$$

can be used to describe ray path geometry for any given velocity field $v(\mathbf{x})$. A consequence of Eq. 16 is Fermat's principle - that of all the paths that join two points A and B in a velocity medium, the true ray path(s) will be stationary in time. In other words, the path along which the integral in Eq. 14 is performed is one which extremizes t. This property was used to derive Eq. 6.

In traveltime tomography, the traditional means of determining source-receiver traveltimes is ray tracing (Červený, 1987, 2001). More recently, wavefront tracking schemes such as finite difference solutions of the eikonal equation have been employed (Vidale, 1988, 1990; Qin et al., 1992). Another method that has seen recent application is network/graph theory, which makes direct use of Fermat's principle (Nakanishi & Yamaguchi, 1986; Moser, 1991). Each of these methods of traveltime determination is described below.

2.2.1 Ray Tracing

The problem of finding a ray path between a source and receiver is a two point boundary value problem, for which there are two basic methods of solution: shooting and bending.

2.2.1.1 Shooting method Shooting methods of ray tracing rely on formulating the ray equation (Eq. 16) as an initial value problem, where a complete ray path can be determined provided the source coordinates and initial ray direction are known. The boundary value problem is then solved by shooting rays through the medium from the source and using information from the computed paths to update the initial ray trajectories so that they more accurately target the receivers (see Fig. 8). Rays may also be shot from receiver to source as the principle of reciprocity applies.

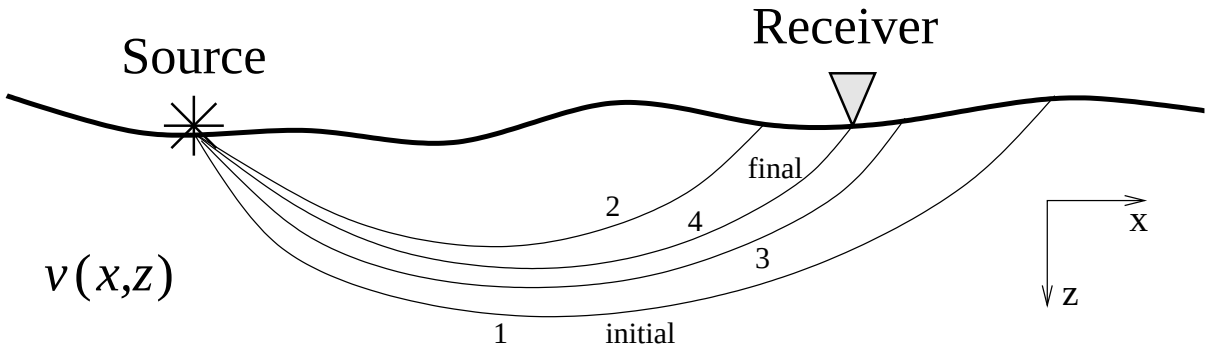

Figure 8: Principle of the shooting method. The initial projection angle of ray 1 is iteratively adjusted until the final ray (4) passes sufficiently close to the receiver.

The ease with which the initial value problem can be solved depends on how the velocity distribution is parameterized. For constant velocity (or slowness) block models, the path within a block

is simply a straight line with traveltime varying linearly with distance. At cell boundaries, new trajectories are calculated using Snell's Law:

$$\frac{\sin\theta_i}{v_i} = \frac{\sin\theta_r}{v_r} \tag{17}$$

where θ_i and θ_r are the angles of incident and refracted rays relative to the normal vector to the interface, and v_i and v_r are the velocities of the media containing the incident and refracted rays respectively.

Analytic ray tracing is also possible for media with a constant velocity gradient. Telford et al. (1976) derive parametric equations for ray path geometry and traveltime when velocity varies linearly with depth z:

$$\left.\begin{aligned} x &= \frac{1}{pk}(\cos i_o - \cos i) \\ z &= \frac{1}{pk}(\sin i - \sin i_o) \\ t &= \frac{1}{k}\ln\left(\frac{\tan\frac{i}{2}}{\tan\frac{i_o}{2}}\right) \end{aligned}\right\} \tag{18}$$

where i is the inclination of the ray (from the downgoing vertical), i_o is the initial inclination, k is the velocity gradient and p is the ray parameter. When the velocity gradient is arbitrarily oriented, such as when a grid of constant velocity gradient triangles or tetrahedra are used (see Section 2.1), then these expressions are modified by rotation of the coordinate system (White, 1989). Simple analytic solutions are also possible when there is a constant gradient of quadratic slowness (see Červený, 1987).

Only a few simple velocity functions allow for analytic solutions of the initial value problem. For the case of an arbitrary differentiable velocity function $v(\mathbf{x})$, numerical solution of an initial value formulation of the ray equation is required. For example, Zelt & Smith (1992), as part of their 2-D wide-angle traveltime inversion method, solve two pairs of first-order differential equations:

$$\frac{\mathrm{d}z}{\mathrm{d}x} = \cot\theta \qquad \frac{\mathrm{d}\theta}{\mathrm{d}x} = \frac{v_z - v_x\cot\theta}{v} \tag{19}$$

or

$$\frac{\mathrm{d}x}{\mathrm{d}z} = \tan\theta \qquad \frac{\mathrm{d}\theta}{\mathrm{d}z} = \frac{v_z\tan\theta - v_x}{v} \tag{20}$$

with initial conditions provided by source location (x_o, z_o) and ray take-off angle θ_o. θ is the angle of incidence (relative to the z-axis), $v_z = \partial v/\partial z$ and $v_x = \partial v/\partial x$. Eq. 19 is used for near horizontal rays and Eq. 20 is used for near vertical rays. Both systems of equations are solved using a Runge-Kutta method with error control. Traveltime is determined by numerical integration of Eq. 14 using the trapezoidal rule. Sambridge & Kennett (1990) use the following set of equations to solve the

initial value problem in 3-D:

$$\left.\begin{aligned}
\frac{\partial x}{\partial t} &= v \sin i \cos j \\
\frac{\partial y}{\partial t} &= v \sin i \sin j \\
\frac{\partial z}{\partial t} &= v \cos i \\
\frac{\partial i}{\partial t} &= -\cos i \left(\frac{\partial v}{\partial x} \cos j + \frac{\partial v}{\partial y} \sin j \right) + \frac{\partial v}{\partial z} \sin i \\
\frac{\partial j}{\partial t} &= \frac{1}{\sin i} \left(\frac{\partial v}{\partial x} \sin j - \frac{\partial v}{\partial y} \cos j \right)
\end{aligned}\right\} \tag{21}$$

where i and j represent the incidence angle and azimuth respectively of the ray. They also use a Runge-Kutta method to solve the system with the ray traveltime t as the integration variable. The accuracy of the ray path and associated traveltime determined via numerical ray tracing depends on the accuracy of the solution scheme (4th order accurate in this case) and the length of the integration step.

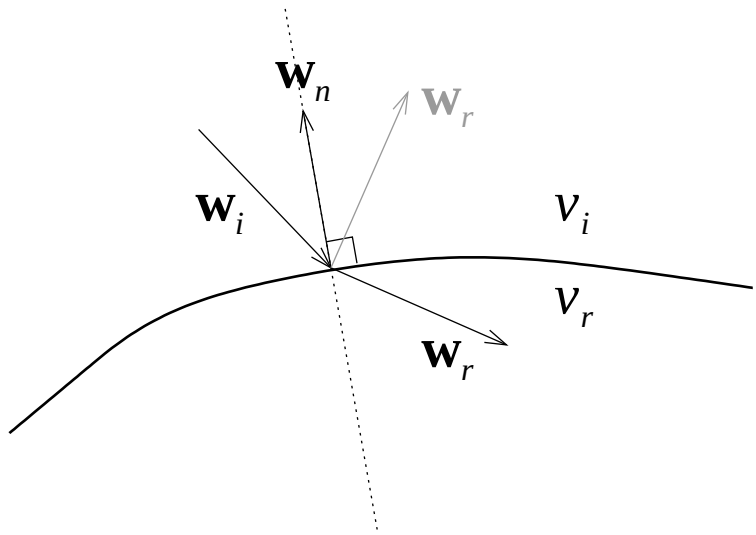

Figure 9: At an interface, rays may refract and/or reflect. $\mathbf{w}_i$ is tangent to the incident path, $\mathbf{w}_r$ is tangent to the refracted (or reflected) path and $\mathbf{w}_n$ is normal to the interface.

When interfaces are included in a discontinuous velocity model, the reflection and transmission laws for a ray path at an interface can be described in terms of Snell's Law (Eq. 17). The geometrical consequence of Snell's Law is that the angle of reflection equals the angle of incidence, and that the reflected/transmitted ray lies in the same plane as the incident ray and the vector normal to the interface at the intersection point. Formulating these constraints into a procedure for determining the new ray direction is relatively straightforward once the ray-interface intersection point is located. For example, if $\mathbf{w}_i$ is a vector tangent to the incident ray path at the intersection point, and $\mathbf{w}_r$ is a vector tangent to the refracted ray path at the intersection point (see Fig. 9), then the relation between these

two vectors is given by (see Červený, 1987):

$$\mathbf{w}_r = \mathbf{w}_i + \left\{ \kappa \left[\frac{1}{v_r^2} - \frac{1}{v_i^2} + (\mathbf{w}_i \cdot \mathbf{w}_n)^2 \right]^{1/2} - \mathbf{w}_i \cdot \mathbf{w}_n \right\} \mathbf{w}_n \tag{22}$$

where $\mathbf{w}_n$ is a normal vector to the interface at the intersection point and v_i and v_r are the velocities of the incident and refracted rays respectively at the intersection point. $\kappa = \text{sign}(\mathbf{w}_i \cdot \mathbf{w}_n)$ and equals $+1$ if $\mathbf{w}_i$ makes an acute angle with $\mathbf{w}_n$ and -1 otherwise. When reflected rays are required, then $v_i = v_r$ and Eq. 22 reduces to:

$$\mathbf{w}_r = \mathbf{w}_i - 2(\mathbf{w}_i \cdot \mathbf{w}_n)\mathbf{w}_n \tag{23}$$

where $\mathbf{w}_r$ now points in the direction of the reflected ray. When analytic ray tracing is used, the ray-interface intersection point can often be found by solving a system of equations which equate a point on the ray with a point on the surface (e.g. Rawlinson et al., 2001a). In numerical ray tracing, the step length of the integration may be iteratively updated in order to obtain a point on the ray path sufficiently close to the interface (e.g. Sambridge & Kennett, 1990).

Solution of the initial value problem is the first step in finding a ray path from source to receiver. The next and generally more difficult step is to solve the two-point boundary value problem. Julian & Gubbins (1977) suggest two iterative methods of solution. The first of these is Newton's method which can be written for the 3-D problem as:

$$\begin{bmatrix} \dfrac{\partial h}{\partial i_o} & \dfrac{\partial h}{\partial j_o} \\ \dfrac{\partial g}{\partial i_o} & \dfrac{\partial g}{\partial j_o} \end{bmatrix}_n \begin{bmatrix} i_o^{n+1} - i_o^n \\ j_o^{n+1} - j_o^n \end{bmatrix} = \begin{bmatrix} H - h(i_o^n, j_o^n) \\ G - g(i_o^n, j_o^n) \end{bmatrix} \tag{24}$$

where (h, g) are the calculated horizontal coordinates of the ray endpoint, (H, G) are the desired coordinates and (i_o, j_o) are respectively the inclination and azimuth of the ray at the source. Solution of this system gives the updated projection coordinates (i_o^{n+1}, j_o^{n+1}), and the process is iterated until an appropriate tolerance criterion is met. The difficulty with this scheme is the accurate determination of the partial derivative matrix. The second method is that of false position, which involves fitting a plane to the $h(i_o, j_o)$ and $g(i_o, j_o)$ of three known rays. The improved estimate (i_o^{n+1}, j_o^{n+1}) corresponds to where (H, G) lies on the plane. The method of false position is quicker at each iteration than Newton's method but converges more slowly. Sambridge & Kennett (1990) use Eq. 24 to solve the two point problem and determine accurate values for the partial derivatives by solving, in conjunction with the initial value problem, two systems of first-order differential equations that describe the geometrical spreading of the wavefront. A perturbation is applied to the initial ray projection angle if the ray gets trapped in a local minimum. In an application of the method (Sambridge, 1990), the initial trajectory of the first-guess ray path is provided by solving the two point problem for a laterally averaged version of the model, as suggested by Thurber & Ellsworth (1980).

The method used by Sambridge (1990) to calculate the partial derivatives in Eq. 24 is a specific application of paraxial ray approximation (Červený & Pšenčik, 1983; Červený et al., 1984; Červený, 1987; Farra & Madariaga, 1988; Červený, 2001), a method that is commonly used to solve the two-point problem in reflection and refraction ray tracing. The method is based on using a ray-centered coordinate system, where a particular ray Ω is taken to define one of the three coordinate axes. The wavefield in the vicinity of the central ray can be determined from quantities that are integrated along

the central ray using dynamic ray tracing. Geometric spreading and wavefront curvature parameters along the initial ray Ω can be used to rapidly locate the two-point ray path from an initial ray that is not too far from the target (see Červený et al., 1984).

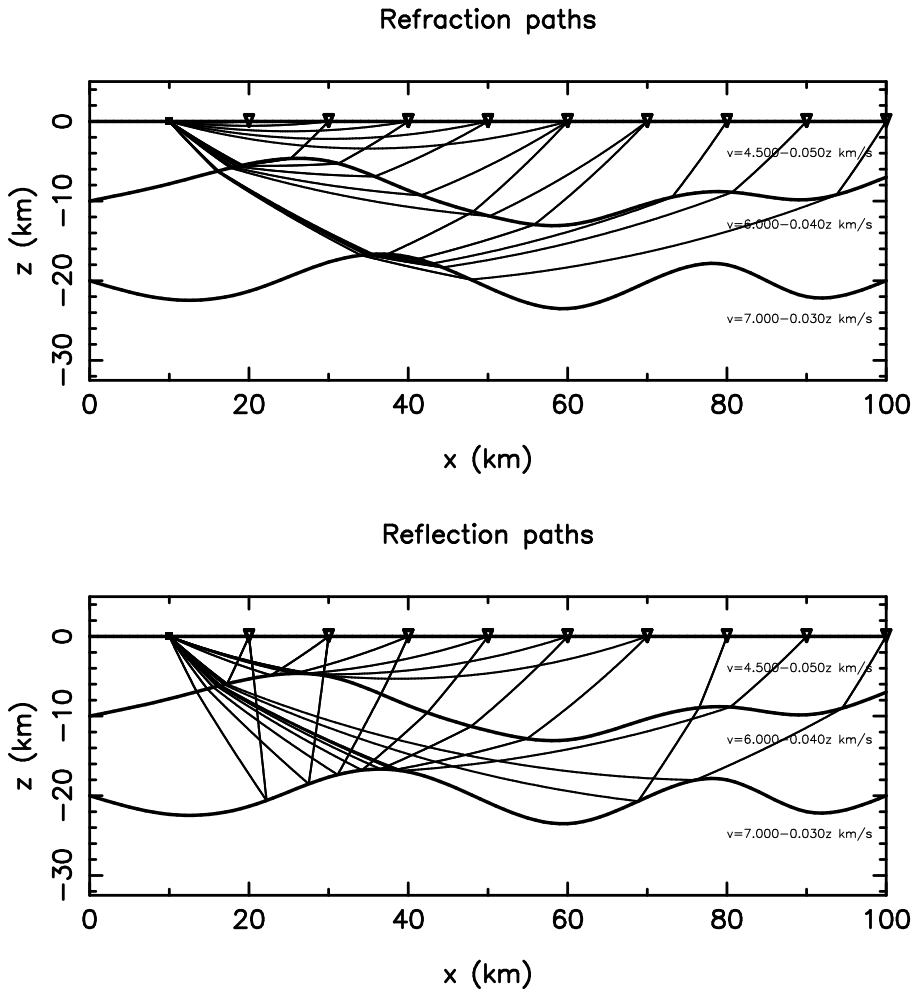

Figure 10: Shooting method of Rawlinson et al. (2001a) used to find source-receiver refraction and reflection ray paths.

In 2-D problems, a shooting approach is often used because the source-receiver array lies on a single vertical plane, making the shooting of a single fan of rays an effective way of obtaining nearby rays to all targets. Zelt & Smith (1992) use a bisection method to find rays that bound each required phase (e.g. a set of rays that all reflect back to the surface from a particular interface). The boundary value problem is approximately solved by shooting a fan of rays into each defined region and linearly interpolating the required quantities between the two closest rays that bracket a receiver. Blundell (1993) uses a similar approach of shooting a fan of rays from the source to find 2-D reflection arrivals. The two-point problem is then solved using a secant algorithm or a bisection algorithm with a pair of rays that bracket the receiver. Similar methods were used by Cassell (1982) and Langan et al. (1985). Other applications of shooting methods in 2-D reflection and/or refraction traveltime inversion include those by Farra & Madariaga (1988), White (1989), Lutter et al. (1990), Williamson (1990), Zelt & Smith (1992) and McCaughey & Singh (1997). Examples of its use in 3-D reflection

and/or refraction traveltime inversions are harder to find although several 3-D tomographic studies that combined refraction and local earthquake data (Benz & Smith, 1984; Ankeny et al., 1986; Sambridge, 1990) and some teleseismic tomography studies (Neele et al., 1993; VanDecar et al., 1995) have used shooting methods of ray tracing. Recently, Rawlinson et al. (2001a) developed a shooting method for finding refraction and reflection arrivals in 3-D layered media. Layer velocity varies linearly with depth in their model, so they were able to analytically trace rays within layers using equations similar to Eq. 18. The boundary value problem was solved using the Newton scheme of Eq. 24. Fig. 10 shows two-point paths through a 2-D layered model found using this method.

2.2.1.2 Bending method The bending method of ray tracing operates by adjusting the geometry of an initial arbitrary path that joins the source and receiver (Fig. 11) until it becomes a true ray path (i.e. it satisfies Fermat's principle). The bending method proposed by Julian & Gubbins (1977) is designed for a continuous 3-D velocity medium and locates a two-point ray path by solving a system of first-order differential equations. If the ray path is described parametrically as $\mathbf{x} = \mathbf{x}(q)$ where the

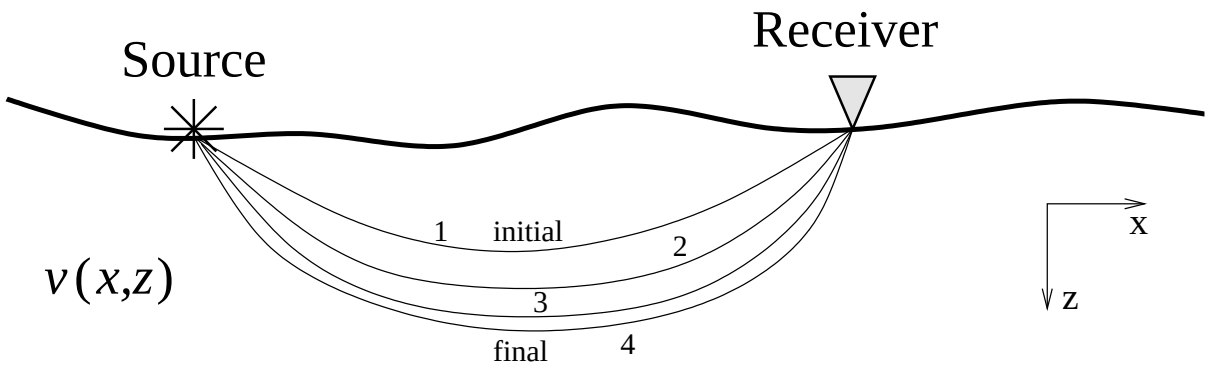

Figure 11: Principle of the bending method. The geometry of the initial path (ray 1) is adjusted until it satisfies Fermat's principle (ray 4).

choice for q can be made later, then Eq. 14 can be written:

$$t = \int_{q_S}^{q_R} sF\,\mathrm{d}q \tag{25}$$

where s is slowness and:

$$F = \frac{\mathrm{d}l}{\mathrm{d}q} = \sqrt{\dot{x}^2 + \dot{y}^2 + \dot{z}^2} \tag{26}$$

with the differentials $\dot{x}$, $\dot{y}$ and $\dot{z}$ being taken with respect to q. The calculus of variations can be employed to describe the path which extremizes t. The Euler-Lagrange equations are (Julian & Gubbins, 1977):

$$\left.\begin{aligned} \frac{\mathrm{d}}{\mathrm{d}q}\frac{\partial}{\partial \dot{x}}(sF) &= \frac{\partial}{\partial x}(sF) \\ \frac{\mathrm{d}}{\mathrm{d}q}\frac{\partial}{\partial \dot{y}}(sF) &= \frac{\partial}{\partial y}(sF) \\ \frac{\partial F}{\partial q} &= 0 \end{aligned}\right\} \tag{27}$$

where $q = l/L$; L is the total length of the source-receiver ray path and $0 \leq l \leq L$. This choice for q results in a single-valued representation of the ray. The boundary conditions are then $\mathbf{x}(0) = \mathbf{x}_S$ and

$\mathbf{x}(1) = \mathbf{x}_R$ where $\mathbf{x}_S$ and $\mathbf{x}_R$ are the source and receiver coordinates respectively. These equations are non-linear and cannot be solved directly. If some initial path $\mathbf{x}^0(q)$ is chosen that passes through S and R, then an improved estimate may be given by:

$$\mathbf{x}^1(q) = \mathbf{x}^0(q) + \boldsymbol{\xi}^0(q) \tag{28}$$

where $\boldsymbol{\xi}^0(q)$ represents a perturbation to the initial path. If Eq. 28 is substituted into Eq. 27, then the resulting equations for $\boldsymbol{\xi}^0$ can be linearized and solved (see Julian & Gubbins, 1977), thus giving the improved estimate $\mathbf{x}^1$. This process can be repeated until the solutions converge.

Pereyra et al. (1980) use a similar approach to locate two-point paths in arbitrary continuous media. They also extend their method to allow for the presence of interfaces. For a medium with an arbitrary number of interfaces that separate regions of smooth velocity variation, the bending problem can be treated by considering a separate system of non-linear differential equations in each smooth region. It is then possible to use the known discontinuity condition at each interface that is traversed by the ray to couple the separate systems. The disadvantage here is that the order in which the interfaces are traversed needs to be known in advance.

Um & Thurber (1987) develop a pseudo-bending technique for solving the two-point problem in continuous 3-D media. Their method is based on a perturbation scheme in which the integration step size is progressively halved. The initial guess path is defined by three points which are linearly interpolated. The center point is then iteratively perturbed using a geometric interpretation of the ray equation until the traveltime extremum converges within a specified limit, at which point the ray equation will be approximately satisfied. The number of path segments is then doubled and the three-point perturbation scheme is repeated working from both endpoints to the middle (a total of three times for this step). The number of segments is doubled again and the procedure is repeated iteratively (see Fig. 12), until the change in traveltime between successive iterations satisfies some convergence criterion.

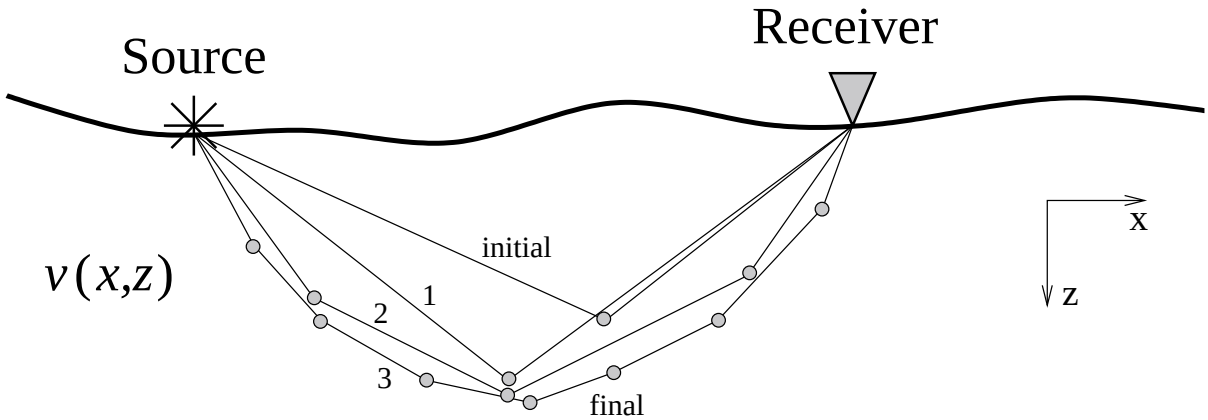

Figure 12: Principle of the pseudo-bending method of Um & Thurber (1987). An initial guess ray defined by three points is provided. The center point is perturbed to best satisfy the ray equation. Then the number of segments is doubled and the process is repeated. This figure schematically represents three such iterations.

Compared to earlier bending methods, the pseudo-bending technique is much faster (Um & Thurber, 1987). Zhao et al. (1992) modify this technique to cope with interfaces as follows. Consider two points A and B close to but on either side of an interface. Straight lines connect A and B separately to a point C on the interface. The correct ray-interface intersection point is obtained by adjusting the point C using a bisection method until Snell's Law is satisfied.

Prothero et al. (1988) develop a 3-D bending method based on the simplex method of function minimization. The first stage of the method is to locate the minimum-time circular path between source and receiver using an exhaustive search method. Perturbations to this path, described by a sum of sine wave harmonics, are then made using the simplex method which searches for the amplitude coefficients that produce the path of least time. The method is more robust than the pseudo-bending method of Um & Thurber (1987) but is significantly slower (Prothero et al., 1988).

Bending methods of ray tracing have been used by a number of authors in studies that invert teleseismic data (Thomson & Gubbins, 1982; Zhao et al., 1994, 1996; Steck et al., 1998), but not by many in the inversion of reflection or refraction data. Chiu et al. (1986) use a bending method in the inversion of 3-D reflection traveltimes and Zhao et al. (1997) use the pseudo-bending method in the inversion of refraction traveltimes for 2-D crustal structure. In local earthquake tomography, bending methods are probably most commonly used to find source-receiver paths and traveltimes (Eberhart-Phillips, 1990; Zhao et al., 1992; Scott et al., 1994; Eberhart-Phillips & Reyners, 1997; Graeber & Asch, 1999). In comparing their bending and shooting methods, Julian & Gubbins (1977) found that bending is computationally faster than shooting by a factor of 10 or more in media with continuous velocity variations. When discontinuities are present, however, the formulation of the bending problem becomes much more complex. In general, for smooth velocity structures that do not cause complex ray geometries, bending methods are more efficient, but when interfaces or strong velocity gradients are present, shooting methods tend to be more robust and therefore preferable (Červený, 1987; Sambridge & Kennett, 1990).

The only other type of ray tracing scheme that is mentioned here is approximate ray tracing (Thurber & Ellsworth, 1980). Here, the velocity in a region local to the source and receiver is laterally averaged, and a 1-D ray tracer is used to find the minimum time-path through this laterally invariant structure. The resultant traveltime and path approximate the true first-arrival traveltime and path through the 3-D model. If more accuracy is required, the ray path estimate can be used as a starting path in a bending routine (Thurber & Ellsworth, 1980). A variant of this technique was introduced by Thurber (1983), in which a large number of circular arcs with differing curvature and dip are joined between source and receiver. The traveltime along each arc is then computed using the 3-D velocity model. An approximation to the first-arrival ray is then selected by choosing the arc with minimum traveltime. Thurber (1983) and Eberhart-Phillips (1986) have used this style of approximate ray tracing in local earthquake tomography.

2.2.2 Wavefront Tracking

Rather than tracing rays from point to point through a medium to determine source-receiver traveltimes, an alternative is to track the propagation path of the entire wavefront. The traveltime from the source to all points in the medium is found using this approach. The most common means of wavefront tracking employs finite-difference solutions of the eikonal equation on a regular grid to calculate the first-arrival traveltime field.

2.2.2.1 Finite difference schemes Vidale (1988) proposed a finite difference scheme that involves progressively integrating the traveltimes along an expanding square in 2-D. Strictly speaking, this method doesn't track wavefronts to determine the traveltime field, but it represents a precursor

to the class of schemes that do, and is still widely used. The eikonal equation (Eq. 15) in 2-D is:

$$\left(\frac{\partial T}{\partial x}\right)^2 + \left(\frac{\partial T}{\partial z}\right)^2 = [s(x,z)]^2 \tag{29}$$

where $s(x,z)$ is the slowness field and $T(x,z)$ is the traveltime of a propagating wave. Vidale's method is formulated for a structure defined by a square grid of velocity nodes. Consider the grid points surrounding some local source point A in Fig. 13. If the traveltime to point A is T_0 then the traveltime to the points B_i are given by:

$$T_{B_i} = T_0 + \frac{h}{2}(s_{B_i} + s_A) \tag{30}$$

where h is the node separation and s_{B_i} and s_A are the slowness at the nodes B_i and A respectively. The next step is to find the traveltime to the corner points T_{C_i}. If the top right hand group of nodes in Fig. 13 with known traveltimes to $A(T_0)$, $B_1(T_1)$ and $B_2(T_2)$ are considered, then the traveltime to the point $C_1(T_3)$ can be determined from the eikonal equation. The two differential terms in Eq. 29 can be approximated with finite differences:

$$\left.\begin{aligned} \frac{\partial T}{\partial x} &= \frac{1}{2h}(T_1 + T_3 - T_0 - T_2) \\ \frac{\partial T}{\partial z} &= \frac{1}{2h}(T_2 + T_3 - T_0 - T_1) \end{aligned}\right\} \tag{31}$$

which, when substituted into Eq. 29, gives:

$$T_3 = T_0 + \sqrt{2(h\bar{s})^2 - (T_2 - T_1)^2} \tag{32}$$

where $\bar{s}$ is the average slowness of all four points under consideration.

The resulting scheme can be used to calculate the traveltimes to all the C_i. The traveltimes to the next set of grid points can then be determined as the scheme progresses by solving along squares of increasing size around the source point (see Fig. 14). Solving for the traveltime to node points along a new square cannot be done in an arbitrary order; a scheme (e.g. Vidale, 1988) is required to determine the order of solution that will result in the least traveltime to each new node. Only these times will be valid seismic traveltimes. Vidale (1988) also gives another formulation that assumes locally circular wavefronts. The locally circular wavefront approximation is most accurate for strongly curved wavefronts and the locally plane wavefront approximation is most accurate for wavefronts with low curvature. Vidale (1990) extends the method to 3-D.

The problem with using an expanding square to progressively determine the traveltime field is that its geometry does not, in general, resemble the shape of the first-arrival wavefront. Consequently, the computed traveltimes may not represent first-arrivals, especially if the structure contains large velocity contrasts. Fig. 15 shows a schematic example in which the traveltime from A to B is determined for path 1 by the expanding square method, but path 2 has the least traveltime by virtue of the high velocity zone. Qin et al. (1992) propose a scheme that calculates the traveltime field using an expanding wavefront method. They use the same propagator equations (Equations 30 and 32) and start by calculating the traveltimes to the eight grid points (in 2-D) about the source. Thereafter, the point of global minimum traveltime along the perimeter of the points processed so far is used as the next source to locally expand the solution region. Using this approach, the traveltime field

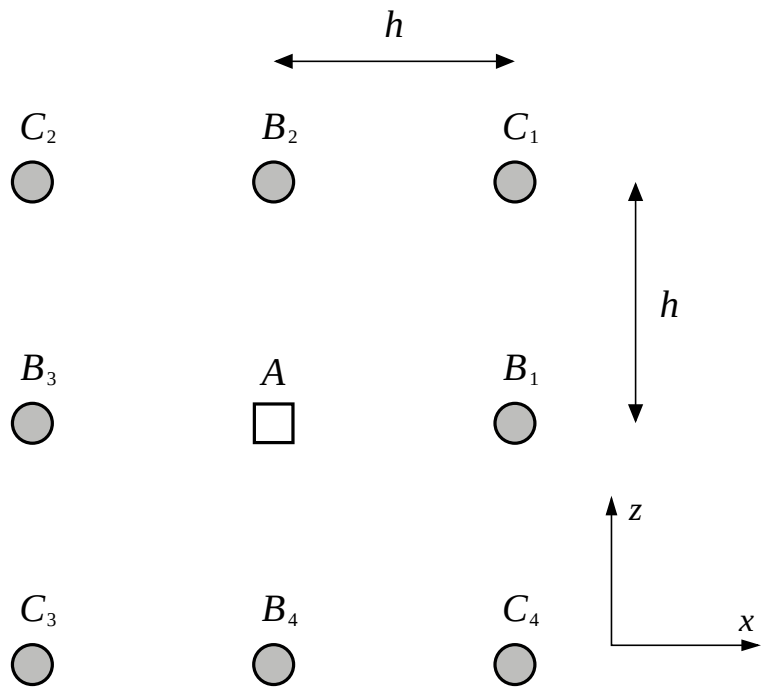

Figure 13: Method used by Vidale (1988) to find the first-arrival traveltime field for a continuous velocity medium. See text for details.

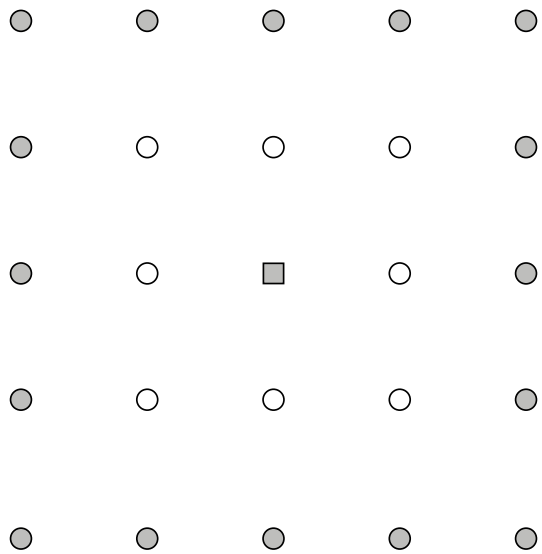

Figure 14: The expanding square method for determining the traveltime field. Traveltimes to the filled circles are determined from the open circles. The filled square is the source.

is determined using an expanding geometry that closely resembles the true shape of the wavefront and the possibility of computing arrivals other than first-arrivals is minimized. Where steep velocity gradients or discontinuities are encountered, however, problems can still occur as only outward propagating rays are considered.

Cao & Greenhalgh (1994) also solve the eikonal equation using a finite difference scheme and a solution region defined by an expanding wavefront. They consider two different model discretization schemes; one in which each node is placed at the center of a cell (i.e. same as the Vidale scheme),

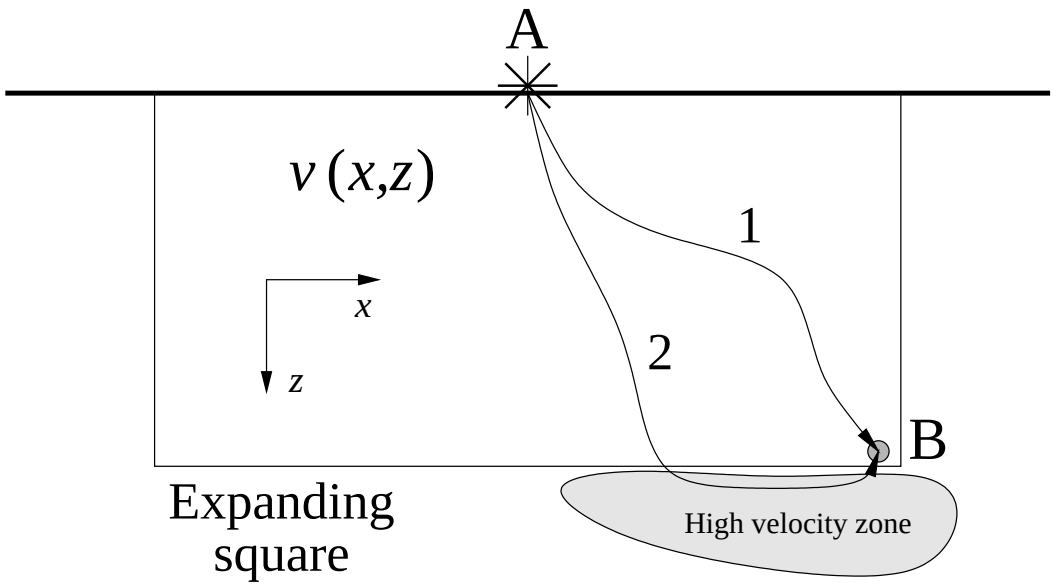

Figure 15: Schematic illustration showing how the expanding square method can fail. The traveltime along path 1 is determined by the expanding square but path 2 has a shorter traveltime due to the high velocity zone.

and one in which the nodes are placed at the corner of a cell with uniform slowness. They found that the corner node discretization provided superior solutions.

The presence of discontinuities such as interfaces between layers can be simulated by assigning the appropriate velocity values to the grid points that lie on either side of the interface. This means that accurate first-arrival times can be determined without separately parameterized interfaces forming a part of the model for the forward step. However, the expanding square formalism of Vidale (1988) may fail for head-waves traveling along an interface with a large velocity contrast as causality may be violated in a similar way as suggested in Fig. 15. Hole & Zelt (1995) and Afnimar & Koketsu (2000) address this problem by introducing special head-wave operators.

If reflection traveltimes are sought, the finite difference method needs to be modified. Riahi & Juhlin (1994) and Hole & Zelt (1995) both develop schemes for calculating reflection traveltimes by modifying the Vidale (1990) method. Riahi & Juhlin (1994) solve the eikonal equation starting from both source and receiver and tracking first-arrivals through the grid to the interface. The correct reflection point will then be the one which minimizes source-receiver traveltime. The drawback of this approach is the need to track traveltimes through the grid from all sources and all receivers. Hole & Zelt (1995) overcome this problem by assuming that the incoming wavefront from the source and the interface are sufficiently smooth to validate a local planar approximation. Thus, the reflected traveltimes to nodes that are adjacent to the interface can be determined using only the depth to the reflector, the normal vector to the reflector, and the direction vectors of the incident ray and reflected ray (from Snell's law). If the wavefront incident on the reflector contains gradient discontinuities or strong curvature, the accuracy of this scheme will be reduced.

Unlike ray tracing methods of traveltime determination, wavefront tracking approaches do not explicitly find ray paths. If they are used as part of a tomographic-style inversion scheme, then some way of locating ray paths is required. One way of doing this is to start at the receiver and follow the traveltime gradient ∇T back through the computed traveltime field to the source. ∇T will always be oriented perpendicular to the first-arrival wavefront and will therefore trace out the first-arrival

ray path. In practice, this could be done on a cell by cell basis using the average traveltime gradient within each cell to orientate a local line segment approximation to the path. Thus, the complete ray path will be described in terms of piecewise linear segments. For example, if we consider a 2-D cell surrounded by grid points $T_{i,j}$, $T_{i+1,j}$, $T_{i,j-1}$ and $T_{i+1,j-1}$, then the average traveltime gradient is approximately:

$$\nabla T = \left(\frac{T_{i+1,j} - T_{i,j} - T_{i,j-1} + T_{i+1,j-1}}{2\delta x}, \frac{T_{i,j} - T_{i,j-1} - T_{i+1,j-1} + T_{i+1,j}}{2\delta z} \right) \tag{33}$$

A number of authors have developed schemes that draw on the basic idea of Vidale (1988); van Trier & Symes (1991) determine traveltimes on a regular grid using an upwind finite-difference method which solves a hyperbolic conservation law that describes changes in the gradient components of the traveltime field. Podvin & Lecomte (1991) employ a method that uses Huygen's principle in the finite difference approximation; Faria & Stoffa (1994) propose a scheme that explicitly uses Fermat's Principle to determine first-arrivals on a gridded traveltime field. These three methods are more suited to solution by massively parallel or vector computation than the methods of Vidale (1988) and Qin et al. (1992).

2.2.2.2 Fast marching method A problem with many of the eikonal grid-based methods (e.g. Vidale, 1988, 1990; Qin et al., 1992) is that they have numerical difficulties when the true wavefront is not differentiable. In other words, the first-arriving wavefront may contain kinks (discontinuities in gradient); this is particularly the case in complex velocity media where multi-pathing (the wavefront crosses itself) can occur. One way of addressing this problem is to search for "weak solutions" of Eq. 15. A weak solution to a differential equation is an entropy satisfying approximate solution that is not differentiable everywhere but satisfies an integral formulation of the equation. The key advantage of such a formulation is that more general solutions are permitted, in particular ones that don't necessarily satisfy the differentiability demands of the original equation.

The Fast Marching Method (FMM) of Sethian & Popovici (1999) for solving the eikonal equation on a 3-D grid uses this approach. In its simplest form, the FMM uses the first-order upwind difference scheme:

$$\begin{bmatrix} \max(D_{ijk}^{-x}T, -D_{ijk}^{+x}T, 0)^2 + \\ \max(D_{ijk}^{-y}T, -D_{ijk}^{+y}T, 0)^2 + \\ \max(D_{ijk}^{-z}T, -D_{ijk}^{+z}T, 0)^2 \end{bmatrix}^{\frac{1}{2}} = s_{ijk} \tag{34}$$

where the following finite difference operator notation is used:

$$D^{+x}T = \frac{T(x+\delta x) - T(x)}{\delta x} \tag{35}$$

$$D^{-x}T = \frac{T(x) - T(x-\delta x)}{\delta x} \tag{36}$$

and s_{ijk} is slowness at the grid point (i, j, k). Eq. 34 is a non-linear equation (quadratic) for the traveltime T_{ijk}. Of the two possible solutions, the larger value is always the correct value. The FMM of Sethian & Popovici (1999) systematically constructs traveltimes T in a downwind fashion from known values upwind using a narrow band method. The narrow band basically represents the propagating wavefront, and grid points are tagged as either *alive*, *close* or *far*, depending on whether

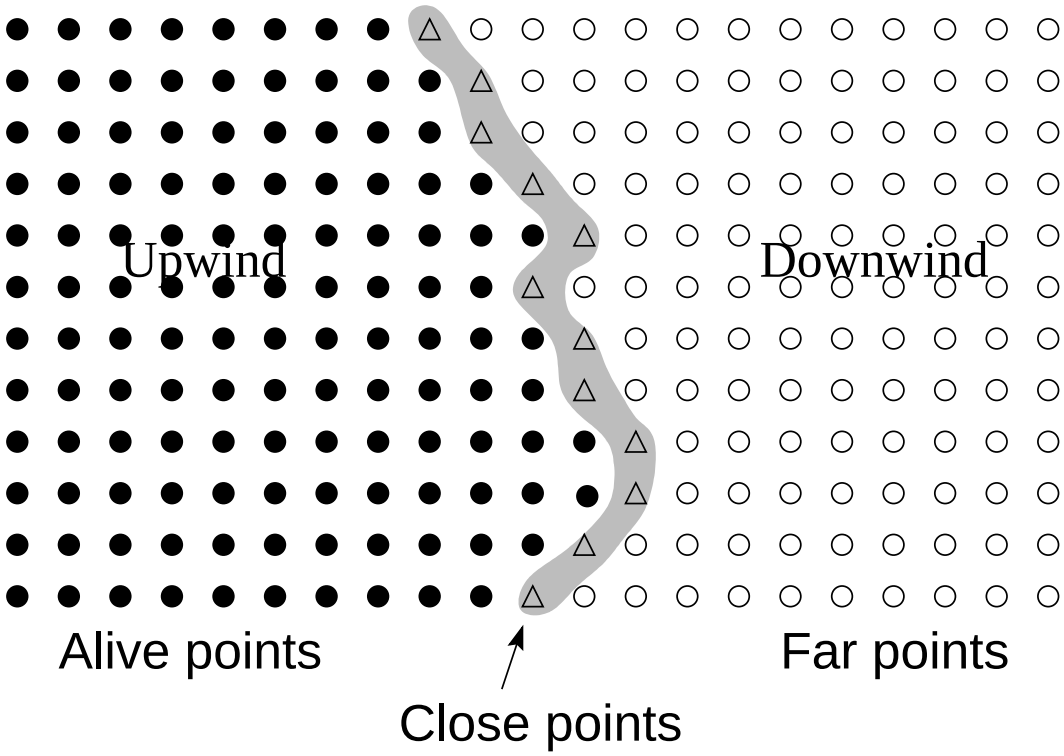

Figure 16: Principle of the narrow band method. *Alive* points have their traveltimes correctly calculated. *Close* points form a band about the *alive* points and have trial values. *Far* points have no values calculated. *Alive* points lie upwind of the narrow band while *far* points lie downwind.

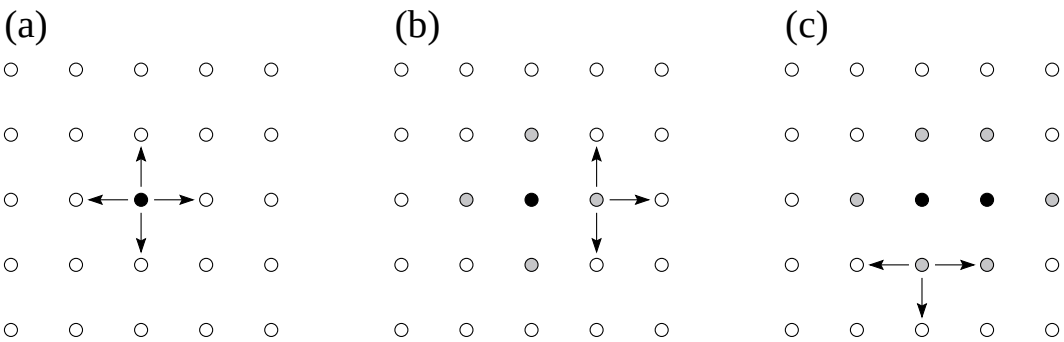

Figure 17: Principal of the FMM in 2-D. (a) Starting from the source point (black dot) in the center of a grid, traveltimes to the four neighboring grid points are determined using Eq. 34. (b) The smallest of these four values (grey dots) must be correct, so all *close* neighbors to this point that are not *alive* (white dots) have their values computed, and added to the narrow band defined by the grey dots. (c) The smallest of these six *close* points again must be correct, and all neighboring points have their values computed (or recomputed).

they have had the band pass through them, they are inside the band, or are yet to be touched by the band, respectively (see Fig. 16).

The FMM begins from a source point (or wavefront), and calculates the traveltimes at neighboring grid points using Eq. 34 (for example) to form the first stage of the narrow band. The point with minimum traveltime is then accepted as *alive* (i.e. it is a true first-arrival traveltime), and all neighboring points to this *alive* point are updated (if *close*) or calculated for the first time (if *far*), in

which case they become *close* and the narrow band progresses downwind (see Fig. 17). Choosing the *close* point with minimum traveltime means that causality is satisfied. The key to an efficient implementation of the narrow band scheme is to be able to rapidly locate the *close* point with minimum traveltime. Heap sorting of traveltimes using a binary tree (Sethian & Popovici, 1999) means that the FMM will have an efficiency of $O(N \log N)$ where N is the number of grid points. Fig. 18 shows a wavefront calculated using the FMM propagating through a complex 2-D velocity model. Source-receiver ray paths, calculated by following the steepest descent direction (using Eq. 33) through the computed traveltime field from each receiver, are also shown. The velocity variations have been made extremely large to illustrate the robustness of the scheme, which remains stable despite the propagating wavefront exhibiting significant discontinuities in gradient.

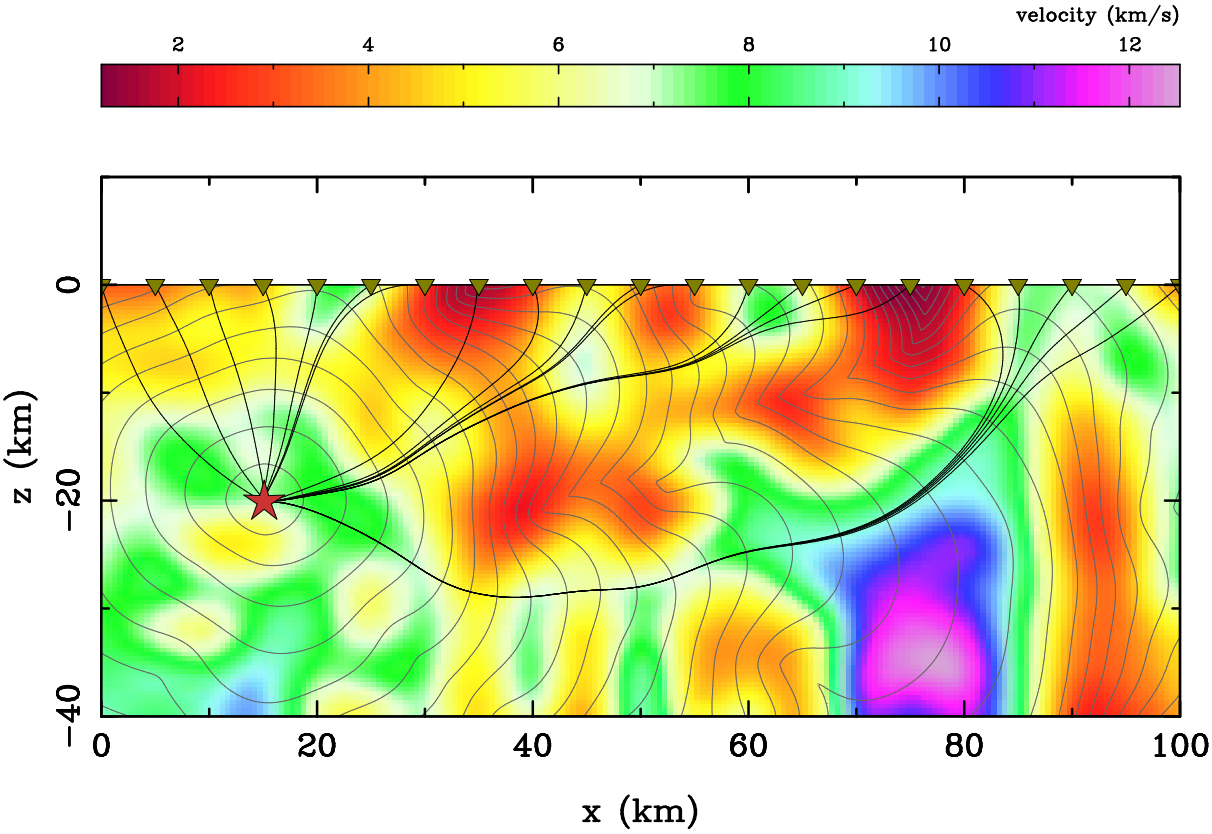

Figure 18: Example of the FMM in a complex 2-D velocity medium. The grid on which the eikonal equation is solved has a spacing of 200 m in both x and z; the total number of points is 100,701. Wavefronts are shown by thin grey lines and are visualized at 0.5 s intervals; rays are denoted by black lines and are always perpendicular to the wavefronts.

To increase the accuracy of the FMM scheme without increasing the number of grid points, higher order differences may be used to approximate the traveltime gradient (Popovici & Sethian, 2002). However, since causality must be respected, they cannot simply replace the first-order differences in all cases (Sethian, 1999). Kim (2001) advocates using the average normal slowness $\hat{s}_{ijk}$ in place of s_{ijk} to increase accuracy since velocities specified at grid points don't account for the true variation across a cell. For point sources, Alkhalifah & Fomel (2001) suggest using a spherical rather than Cartesian grid since the latter tends to under-sample the wavefront near the source where curvature is high, and hence introduces traveltime errors. This source of error occurs for any Carte-

sian grid-based method that determines the traveltime field from a point source. Although the FMM scheme is yet to be applied to tomographic problems, we have described it in some detail because it is the first unconditionally stable eikonal scheme and will probably come into more common use in the near future.

Finite difference grid methods that solve the eikonal equation are generally much quicker (Vidale, 1988; Sethian & Popovici, 1999) than the ray tracing methods described in Section 2.2.1, especially for problems involving relatively few receivers and many sources or vice versa. They are designed to calculate first-arrival traveltimes, they find diffraction paths in shadow zones, and they can often work in regions of complex velocity (e.g. Sethian & Popovici, 1999). Ray methods do not necessarily find the first-arrival path and often fail in shadow zones. The grid-based methods can also determine the arrivals of head-waves and diffractions that cannot be found by conventional geometric ray tracing.

Finite-difference schemes do have their disadvantages, one of the principal being algorithm stability, although the FMM scheme recently introduced by Sethian & Popovici (1999) overcomes this problem. They also only locate first-arrivals, unless explicit conditions, such as reflections from an interface, are specified. While this is desirable in many cases, later arrivals can be of interest, such as in the generation of synthetic seismograms. Another disadvantage of grid methods is that the medium needs to be densely sampled by velocity nodes in order to achieve accurate traveltimes, with consequent demands on processing power and memory. For a 3-D problem, computation time will increase at least in proportion to M^3 where M is the number of nodes in one dimension. The increase in computation time for ray-tracing methods is generally not as dramatic. Regional and global studies, which often use a large number of model parameters, do not usually employ grid-based methods to find traveltimes and ray paths. Apart from computation time, specific phases (e.g. PcP, PP, PcS) which are not first-arrivals are sometimes used, which would also complicate the implementation of a grid method. Grid methods are also not common in teleseismic tomography, possibly because of the need to find ray paths from the source point to the model region through a global 1-D velocity model, a task easily accomplished by ray tracing. Similarly, examples of applications in local earthquake tomography are hard to find, although there doesn't seem to be any particular reason why grid methods are unsuited to this class of problem.

In contrast, finite difference solutions of the eikonal equation have been used frequently in the forward step of wide-angle traveltime inversions, especially in 3-D. Hole (1992) presents a method for the inversion of first-arrival traveltimes for 3-D velocity variation using a finite-difference approach. Hole et al. (1992) use a similar forward scheme in the inversion for interface structure using broadside refractions from the Queen Charlotte Basin, Canada. Also using finite-difference techniques, Riahi et al. (1997) invert wide-angle reflections for Moho structure beneath the Gulf of Bothnia and Zelt et al. (1996) invert both reflection and refraction traveltimes for velocity and interface structure beneath the southwestern Canadian Cordillera. Other studies to use finite difference solutions of the eikonal equation in 3-D wide-angle traveltime inversions include Zelt & Barton (1998), Zelt et al. (1999, 2001) and Day et al. (2001). Parsons et al. (1996) use a finite difference approach in the inversion of wide-angle traveltimes for the 2-D crustal structure of the Colorado Plateau.

2.2.3 Shortest Path Ray Tracing (SPR)

The shortest path or network method uses Fermat's principle directly to find the path of the first-arrival ray between source and receiver. To do this, a grid of nodes is specified within the velocity

medium and a network or graph is formed by connecting neighboring nodes with traveltime path segments. The first-arrival ray path between source and receiver will then correspond to the path through the network which has the least traveltime.

In a seminal paper by Nakanishi & Yamaguchi (1986), the velocity field is defined by a set of constant velocity blocks with network nodes placed on the interface between the blocks. Connection paths between adjacent nodes do not cross any cell boundaries (see Fig. 19a), so the traveltime t between two nodes is simply $t = ds$ where d is the distance between the two nodes and s is cell slowness. A similar approach is used by Fischer & Lees (1993). Moser (1991) uses a rectangular grid with the network nodes coinciding with the velocity nodes (see Fig. 19b). The traveltime between two connected nodes is estimated by $t = d(s_1 + s_2)/2$ where s_1 and s_2 are the slowness at the two nodes.

Once the network structure and method of traveltime determination between two nodes has been chosen, the next step is to use a shortest path algorithm to locate the ray path. Essentially, the problem is to locate the path of minimum traveltime from all the possible paths between source and receiver through the given network. An algorithm that is often used in network theory is that of Dijkstra (1959) for which computation time is proportional to the number of nodes squared.

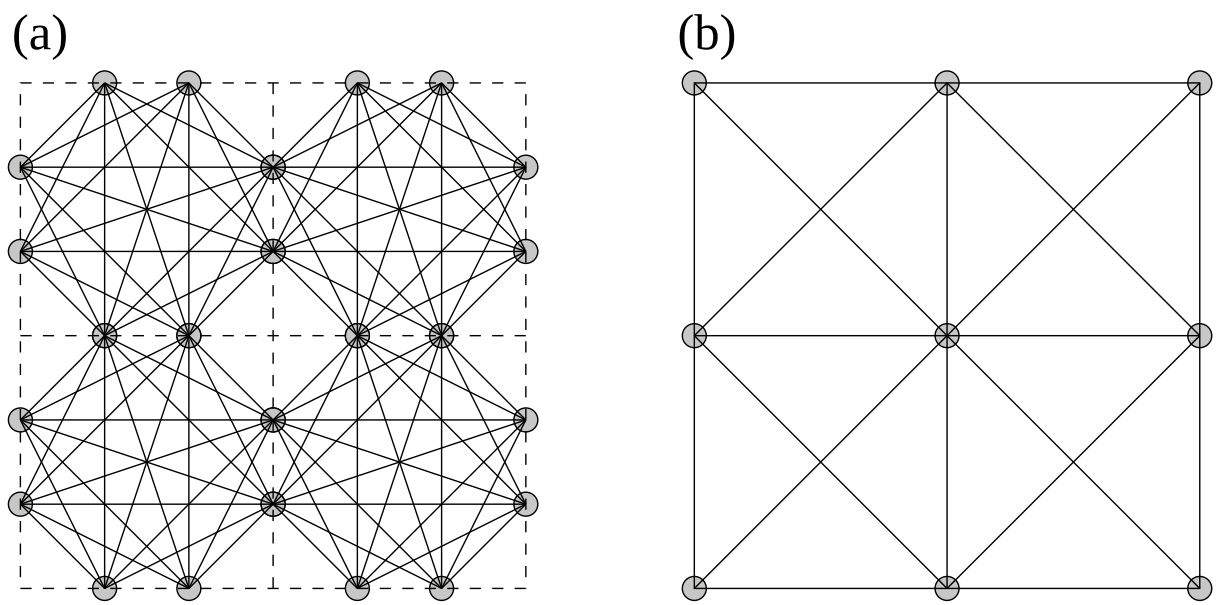

Figure 19: Two types of node arrangements for an SPR network (after Moser, 1991). Connectors are indicated by solid lines in both cases. (a) Network nodes along constant slowness cell boundaries (dashed). (b) Network nodes that coincide with velocity nodes.

Errors in SPR are due to the finite node spacing and angular distribution of node connectors (Moser, 1991). A coarse grid of nodes may poorly approximate the velocity variations while a limited range of angles between adjacent connectors may result in a poor approximation to the true path. Obviously, increasing the number of nodes and connectors will result in superior solutions but may come at a significant computational cost. Much work has been done to increase the computational speed of the shortest path algorithm, with particular attention given to the use of efficient sorting algorithms (Moser, 1991; Klimeš & Kvasnička, 1994; Cheng & House, 1996; Zhang & Toksöz, 1998).

SPR will by definition find the first-arrival traveltime between any given source and receiver.

However, it is possible to impose constraints on the path so that some other arrivals such as reflections or multiples can be determined. Moser (1991) demonstrated a method for reflections which requires the shortest path to visit a specified set of nodes that lie on the interface.

In their original implementation of SPR, Nakanishi & Yamaguchi (1986) inverted traveltimes from local earthquakes, while Zhang & Toksöz (1998) used it in the inversion of refraction traveltimes. Toomey et al. (1994) inverted first-arrival refraction traveltimes for 3-D crustal velocity structure using a scheme similar to that of Moser (1991) to solve the forward problem. Apart from these examples, the use of SPR in tomographic inversions is not common. SPR shows similar advantages to finite differences (see Section 2.2.2) relative to conventional ray tracing methods. It can correctly locate diffraction paths and head waves and always finds first-arrivals. The main advantage SPR has over most finite difference methods is robustness; it is capable of working in highly complex media. Cheng & House (1996) claim it to be the most robust numerical scheme for traveltime calculations, although the introduction of the FMM challenges this claim. SPR methods also tend to be slower than eikonal methods.

2.3 Solving the Inverse Step

The inversion step, which involves the adjustment of the model parameters $\mathbf{m}$ to better satisfy the observed data $\mathbf{d}_{obs}$ through the known relationship $\mathbf{d} = \mathbf{g}(\mathbf{m})$, can be performed in a number of ways. In traveltime tomography, the functional $\mathbf{g}$ is non-linear because the ray path depends on the velocity structure. Ideally, an inversion scheme should account for this non-linearity. The three approaches to solving the inversion step that will be considered below are backprojection, gradient methods and global optimization techniques.

2.3.1 Backprojection

In Section 1.2, we showed that the perturbation of a ray path only has a second order effect on traveltime. In terms of slowness, this was written (see Eq. 7):

$$\delta t = \int_{L_0} \delta s(\mathbf{x}) \mathrm{d}l + O(\delta s(\mathbf{x})^2) \tag{37}$$

If a continuum is described by M constant slowness blocks, then the discrete form of Eq. 37 for N rays can be written:

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{38}$$

where $\mathbf{d}$ are the traveltime residuals, $\mathbf{m}$ the slowness perturbations and $\mathbf{G}$ an $N \times M$ matrix of ray lengths l_{ij} corresponding to the distance traversed by each ray in each block. Note that for the general case $\mathbf{m}$ (e.g. velocity nodes, interface depths etc.) in Eq. 38, $\mathbf{G} = \partial\mathbf{g}/\partial\mathbf{m}$ where $\mathbf{g}(\mathbf{m})$ is the model prediction. Many of the elements of $\mathbf{G}$ will be zero since each ray path will usually only traverse a small subset of the M blocks. Backprojection methods can be used to solve Eq. 38 for the slowness perturbations $\mathbf{m}$ by iteratively mapping traveltime anomalies into slowness perturbations along the ray paths until the data are satisfied. Backprojection methods generally use constant slowness (or velocity) blocks. Two well known backprojection techniques for solving Eq. 38 are the Algebraic Reconstruction Technique (ART) and the Simultaneous Iterative Reconstruction Technique (SIRT), both of which originate from medical imaging.

In ART, the model is updated on a ray by ray basis. The residual d_n for the nth ray path is distributed along the path by adjusting each component of $\mathbf{m}$ in proportion to the length l_{nj} of the ray segment in the jth cell:

$$m_j^{k+1} = m_j^k + \frac{t_n^{k+1} l_{nj}}{\sum_{m=1}^{M} l_{nm}^2} \tag{39}$$

where $t_n^{k+1} = d_n - t_n^k$ is the difference between the residuals at the 0th and k^{th} iteration, m_j^k is the approximation to the jth model parameter at the kth iteration, $m_j^1 = 0$ and $t_n^1 = 0$. The residual along the $(n+1)^{th}$ ray is then determined for the updated velocity field $\mathbf{m}$ using the original path and is backprojected in the same manner. A single iteration of the method consists of performing this backprojection for the N ray paths. Rays are then retraced and the backprojection repeated until the data are satisfied to within tolerance, or the solution converges. The main problem with ART is that it suffers from poor convergence properties (Blundell, 1993). It has been used by McMechan (1983) in cross-hole tomography and Nakanishi & Yamaguchi (1986) in local earthquake tomography.

SIRT addresses some of the convergence problems associated with ART by averaging the perturbations applied to each parameter from all the rays that are influenced by the parameter. Thus, the SIRT algorithm may be written (Blundell, 1993):

$$m_j^{k+1} = m_j^k + \frac{1}{R_j^k} \sum_{n=1}^{R_j^k} \left[\frac{t_n^{k+1} l_{nj}}{\sum_{m=1}^{M} l_{nm}^2} \right] \tag{40}$$

where R_j^k is the number of rays that the jth model parameter influences for the kth iteration. The SIRT method has been used in the inversion of teleseismic traveltime residuals by Dueker et al. (1993), Granet & Trampert (1989) and McQueen & Lambeck (1996). Blundell (1993) used SIRT (as well as other methods) in the inversion of reflection traveltimes for both velocity structure and interface depth.

Authors who have used variants of these backprojection schemes include Humphreys & Clayton (1990), Hole (1992) and Zelt & Barton (1998). Humphreys & Clayton (1990) used block subbinning, filtering and spatial averaging in the backprojected inversion of teleseismic traveltimes. Block subbinning reduces the weight of rays that come from dominant directions, and thus reduces blurring of the image along these paths. Filtering is done using point spread functions to reduce the natural tendency of backprojection to blur an image if resolution is not perfect, and spatial averaging is used to smooth the solution. Hole (1992) also uses smoothing in the backprojected inversion of wide-angle traveltimes. Zelt & Barton (1998), in their 3-D wide-angle seismic inversion method, implement several other modifications aimed at improving the convergence and accuracy of backprojection.

Inversion using backprojection tends to be computationally more rapid at each iteration compared to other techniques, but often converges more slowly and with less stability. This is at least partly due to the use of more *ad hoc* regularization (like spatial averaging) compared to, for example, the formal inclusion of such constraints in the inversion permitted by gradient methods.

2.3.2 Gradient Methods

The inverse problem in seismic tomography can be formulated as one of minimizing an objective function consisting of a data residual term and one or more regularization terms. As before, let $\mathbf{d}$ denote a data vector of length N which is dependent on a model vector $\mathbf{m}$ of length M as $\mathbf{d} = \mathbf{g}(\mathbf{m})$. For an initial estimate $\mathbf{m}_0$ of the model parameters, comparing $\mathbf{d} = \mathbf{g}(\mathbf{m}_0)$ with the observed traveltimes $\mathbf{d}_{obs}$ gives an indication of the accuracy of the model. The misfit can be quantified by constructing an objective function $S(\mathbf{m})$, consisting of a weighted sum of data misfit and regularization terms, that is to be minimized.

An essential component of the objective function is a term $\Psi(\mathbf{m})$ which measures the difference between the observed and predicted data. If it is assumed that the error in the relationship $\mathbf{d}_{obs} \approx \mathbf{g}(\mathbf{m}_{true})$ is Gaussian, then a least squares or L_2 measure of this difference is suitable:

$$\Psi(\mathbf{m}) = \|\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}\|^2 \tag{41}$$

If uncertainty estimates have been made for the observed data (usually based on picking error), then more accurate data are given a greater weight in the objective function by writing $\Psi(\mathbf{m})$ as:

$$\Psi(\mathbf{m}) = (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs})^T \mathbf{C}_d^{-1} (\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}) \tag{42}$$

where $\mathbf{C}_d$ is a data covariance matrix. If the errors are assumed to be uncorrelated, then $\mathbf{C}_d = [\delta_{ij}(\sigma_d^j)^2]$ where σ_d^j is the uncertainty of the jth traveltime. Strictly speaking, $\mathbf{C}_d$ is best referred to as a data weighting matrix rather than a data covariance matrix unless it truly reflects the uncertainty associated with the data. A major weakness in this definition of data misfit is that the L_2-norm is sensitive to outliers. This means that if only a few data have spurious values (e.g. from incorrect phase identification), then they will have a significant influence on the size of $\Psi(\mathbf{m})$ since each residual is squared; the L_2-norm is a non-robust measure. A solution produced by the minimization of $\Psi(\mathbf{m})$ is then likely to be less reliable than one produced using, for example, an L_1 norm (Menke, 1989). Despite this weakness, most inversion methods appeal to Gaussian statistics and adopt an L_2 norm. See Claerbout & Muir (1973) for a discussion of robustness in error distribution.

A common problem with tomographic inversion is that not all model parameters will be well constrained by the data alone (i.e. the problem may be under-determined or mixed-determined). A regularization term $\Phi(\mathbf{m})$ is often included in the objective function to provide additional constraints on the model parameters, thereby reducing the non-uniqueness of the solution. The regularization term is typically defined as:

$$\Phi(\mathbf{m}) = (\mathbf{m} - \mathbf{m}_0)^T \mathbf{C}_m^{-1} (\mathbf{m} - \mathbf{m}_0) \tag{43}$$

where $\mathbf{C}_m$ is an *a priori* model covariance matrix. If uncertainties in the initial model are assumed to be uncorrelated, then $\mathbf{C}_m = [\delta_{ij}(\sigma_m^j)^2]$ where σ_m^j is the uncertainty associated with the jth model parameter of the initial model. Again, this should really be referred to as a model weighting matrix unless its entries reflect the true statistical uncertainties of the initial model. The effect of $\Phi(\mathbf{m})$ is to encourage solution models $\mathbf{m}$ that are near a reference model $\mathbf{m}_0$. The values used in $\mathbf{C}_m$ are usually based on prior information.

Another approach to regularization is the minimum structure solution (Constable et al., 1987) which attempts to find an acceptable trade-off between satisfying the data and finding a model with the minimum amount of structural variation. One way of including this requirement in the objective function is to use the term (Sambridge, 1990):

$$\Omega(\mathbf{m}) = \mathbf{m}^T \mathbf{D}^T \mathbf{D} \mathbf{m} \tag{44}$$

where **Dm** is a finite difference estimate of a specified spatial derivative. For example, if $m_1, m_2, \ldots, m_M$ represent contiguous depth nodes of an interface in 2-D space, then the gradient of the interface could be regulated using:

$$\mathbf{Dm} = \begin{bmatrix} -1 & 1 & 0 & \ldots & \ldots & 0 \\ 0 & -1 & 1 & 0 & \ldots & 0 \\ \ldots & 0 & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & 0 & -1 & 1 \\ \ldots & \ldots & \ldots & \ldots & 0 & 0 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \ldots \\ m_{M-1} \\ m_M \end{bmatrix} \tag{45}$$

Alternatively, the curvature of the interface could be regulated using:

$$\mathbf{Dm} = \begin{bmatrix} 0 & 0 & 0 & 0 & \ldots & 0 \\ 1 & -2 & 1 & 0 & \ldots & 0 \\ 0 & 1 & -2 & 1 & \ldots & \ldots \\ \ldots & \ldots & \ldots & \ldots & \ldots & \ldots \\ \ldots & \ldots & \ldots & 1 & -2 & 1 \\ \ldots & \ldots & \ldots & \ldots & 0 & 0 \end{bmatrix} \begin{bmatrix} m_1 \\ m_2 \\ m_3 \\ \ldots \\ m_{M-1} \\ m_M \end{bmatrix} \tag{46}$$

An explicit smoothing term such as Eq. 46 in the objective function may be necessary if crude parameterizations such as constant velocity blocks are used to simulate a continuously varying velocity field. However, if an implicitly smooth parameterization like cubic splines is used, then an explicit smoothing term may be unnecessary. In other words, the same smooth result could be achieved by reducing the number of parameters and hence the size of the inverse problem. The appropriate number of parameters required to represent the model for a given dataset could be chosen on a statistical basis, such as employing the F test (Menke, 1989).

Using the L_2 terms described above in Equations 42 - 44, the objective function $S(\mathbf{m})$ can be written in full as:

$$S(\mathbf{m}) = \frac{1}{2}\left[\Psi(\mathbf{m}) + \epsilon\Phi(\mathbf{m}) + \eta\Omega(\mathbf{m})\right] \tag{47}$$

where ϵ is referred to as the *damping factor* and η as the *smoothing factor* (when **D** is the second derivative operator, which is usually the case). Multiplication of all terms by $1/2$ is done simply to prevent the expressions for the first and second derivatives of S having all elements multiplied by 2 (see Eq. 49 and Eq. 50). ϵ and η govern the trade-off between how well the solution $\mathbf{m}_{est}$ will satisfy the data, how closely $\mathbf{m}_{est}$ is to $\mathbf{m}_0$, and the smoothness of $\mathbf{m}_{est}$. There are several means for choosing appropriate values for ϵ and η. One way is to use the largest values of ϵ and η for which the data are still satisfied, but there will be a trade-off between ϵ and η. When only damping or smoothing is invoked, then this approach is simpler to implement. However, it may be that the relationship between the data fit and the model perturbation or model roughness is highly non-linear, in which case this criterion may not be robust. A better approach is to inspect the trade-off curves between data fit and model roughness (or perturbation) for different values of η (or ϵ), as shown schematically in Fig. 20. If both η and ϵ are non-zero, then a contour plot which shows traveltime fit contoured on a plot of model roughness vs. model perturbation (Fig. 21) could be used, but the main drawback here is the considerable computational effort required. An alternative approach is to perform a synthetic reconstruction using the same source-receiver geometry as the real experiment. The appropriate amount of damping and smoothing would then be given by the values of η and ϵ which resulted in the most accurate reconstruction of the synthetic model.

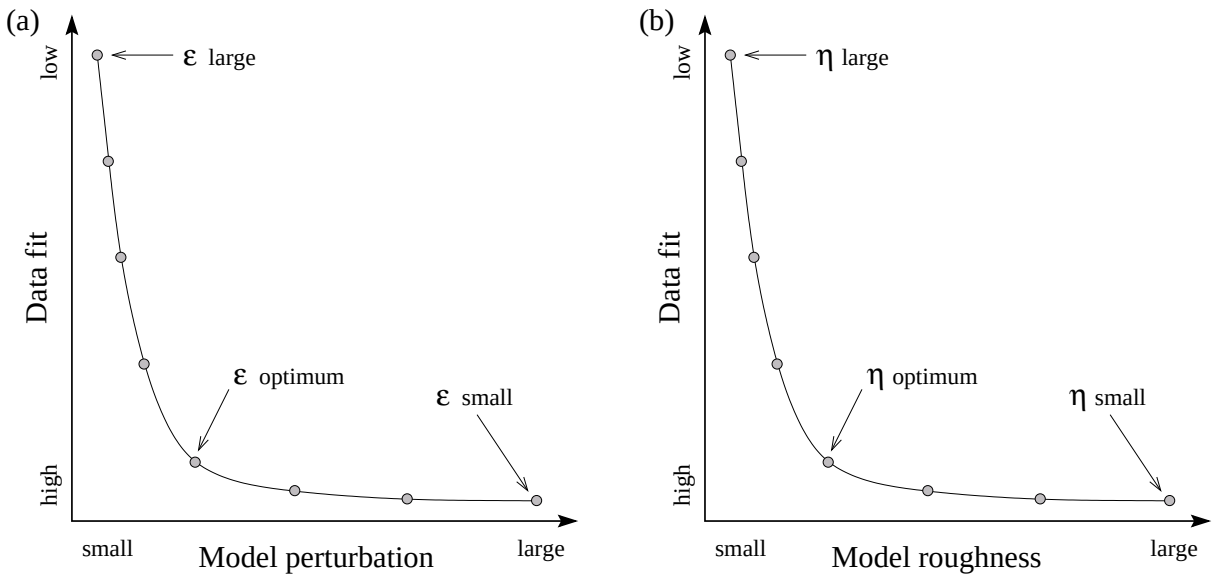

Figure 20: Schematic illustration of trade-off curves that could be used to choose appropriate damping or smoothing parameters for an inversion. A number of separate (eight in these examples) inversions with different values of ϵ or η are required in order to construct these curves. (a) Data fit vs. model perturbation for different values of ϵ. (b) Data fit vs. model roughness for different values of η.

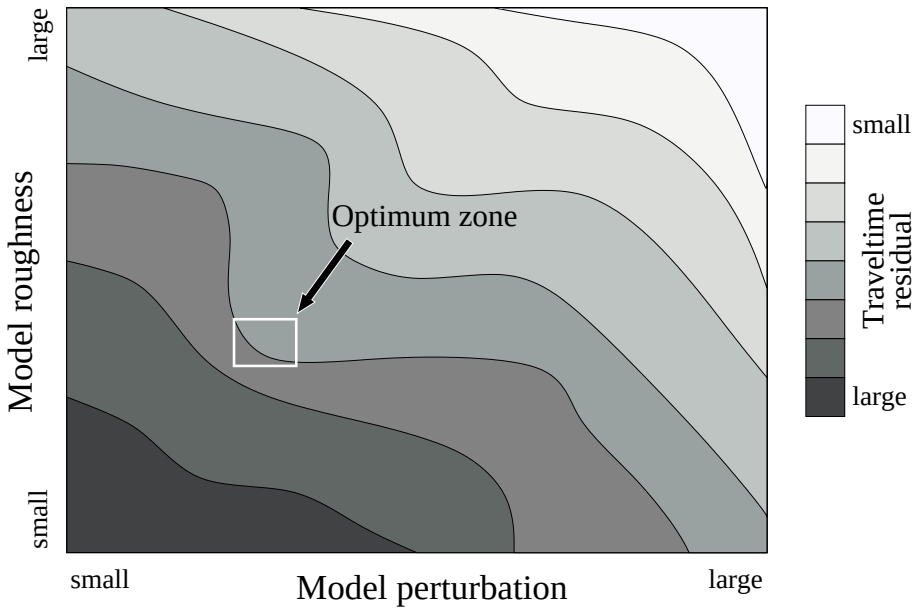

Figure 21: Possible scheme for choosing ϵ and η simultaneously. Contours of traveltime misfit for a set of solution models are plotted on a graph of model perturbation vs. model roughness. The optimum region has low roughness and model perturbation but adequately satisfies the data. Computational effort is the main drawback of this approach.

Many studies involving the tomographic inversion of real data use a semi-quantitative approach to choosing η and/or ϵ (or other trade-off parameters) like those outlined above (e.g. Oncescu et al., 1984; White, 1989; Neele et al., 1993; Steck et al., 1998; Graeber & Asch, 1999; Rawlinson et al., 2001a). At this stage, it is important to point out that the objective function expressed in Eq. 47 in a sense juxtaposes two different regularization frameworks: Bayesian and Occam's. In a Bayesian-

style inversion, knowledge of *a priori* information is of paramount importance. To properly honor this approach, smoothing would be ignored (i.e. set η=0), the data covariance matrix would reflect the known statistical properties of the data and the initial model estimate would be based on *a priori* model information. In addition, we would set $\epsilon = 1$ since the *a priori* model covariance matrix would reflect the uncertainties associated with the *a priori* information. Minimization of the objective function then results in a solution that assimilates the information contained in the data with the *a priori* information, resulting in an *a posteriori* model distribution. The posterior model uncertainties should then be less than the prior uncertainties, which is the desired outcome (how much less depends on how good the data are). The principal difficulty with this approach is that meaningful *a priori* information concerning model and data errors is difficult to obtain in practice. Nevertheless, this Bayesian approach has been adopted by a number of authors (e.g. Aki et al., 1977; Lutter & Nowack, 1990). Scales & Snieder (1997) discuss the merits and difficulties of adopting a Bayesian framework for inversion.

The other regularization framework is suggested by Occam's principle of seeking a solution with the least structure necessary to fit the data (Constable et al., 1987). In this case $\epsilon = 0$ since we don't want the initial model, whose accuracy is poorly known, to unduly influence the solution model. This type of scheme is also popular (e.g. Sambridge, 1990; Zelt & Barton, 1998; Day et al., 2001). Often, however, a mixture of both frameworks is used, in which case one seeks a physically reasonable model that contains no unnecessary structure, is not highly perturbed from the initial model and satisfies the data. This explains to some extent why the choice of ϵ and η is often subjective.

Gradient-based inversion methods make use of the derivatives of $S(\mathbf{m})$ at a specified point in model space. A basic assumption that is shared by all practical gradient methods is that $S(\mathbf{m})$ is sufficiently smooth to allow a local quadratic approximation about some current model:

$$S(\mathbf{m} + \delta\mathbf{m}) \approx S(\mathbf{m}) + \hat{\boldsymbol{\gamma}}\delta\mathbf{m} + \tfrac{1}{2}\delta\mathbf{m}^T\hat{\mathbf{H}}\delta\mathbf{m} \tag{48}$$

where $\delta\mathbf{m}$ is a perturbation to the current model and $\hat{\boldsymbol{\gamma}} = \partial S/\partial\mathbf{m}$ and $\hat{\mathbf{H}} = \partial^2 S/\partial\mathbf{m}^2$ are the gradient vector and Hessian matrix respectively. Evaluating these partial derivatives for Eq. 47 gives:

$$\hat{\boldsymbol{\gamma}} = \mathbf{G}^T\mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}] + \epsilon\mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0) + \eta\mathbf{D}^T\mathbf{D}\mathbf{m} \tag{49}$$

$$\hat{\mathbf{H}} = \mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \nabla_{\mathbf{m}}\mathbf{G}^T\mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}] + \epsilon\mathbf{C}_m^{-1} + \eta\mathbf{D}^T\mathbf{D} \tag{50}$$

where $\mathbf{G} = \partial\mathbf{g}/\partial\mathbf{m}$ is the Fréchet matrix of partial derivatives calculated during the solution of the forward problem. As mentioned earlier, for the case of constant slowness blocks, $\mathbf{G} = [l_{ij}]$ where l_{ij} is the ray segment length of the ith ray in the jth block. Usually, the second derivative term in $\hat{\mathbf{H}}$ is neglected since it is time consuming to evaluate, and its effect is small if $\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}$ is small, or if the forward problem is quasi-linear ($\nabla_{\mathbf{m}}\mathbf{G} \approx 0$). Both $\hat{\boldsymbol{\gamma}}$ and $\hat{\mathbf{H}}$ do not lie in model space, but in the dual of model space (Tarantola, 1987). If $\boldsymbol{\gamma}$ is the steepest ascent vector in model space then $\boldsymbol{\gamma} = \mathbf{C}_m\hat{\boldsymbol{\gamma}}$ and $\mathbf{H}$, the curvature operator in model space, is $\mathbf{H} = \mathbf{C}_m\hat{\mathbf{H}}$.

Since $\mathbf{g}$ is generally non-linear, the minimization of Eq. 47 requires an iterative approach:

$$\mathbf{m}_{n+1} = \mathbf{m}_n + \delta\mathbf{m}_n \tag{51}$$

where $\mathbf{m_0}$ is the initial model. The objective function is minimized for the current ray path estimate at each step to produce $\mathbf{m}_{n+1}$, after which new ray paths are computed for the next iteration. The iterations cease either when the observed traveltimes are satisfied or when the change in $S(\mathbf{m})$ with

iteration gets sufficiently small. A useful measure of data fit is provided by the normalized χ^2 misfit function defined by:

$$\chi^2 = \frac{1}{N}\sum_{i=1}^{N}\left(\frac{d_m^i - d_{obs}^i}{\sigma_d^i}\right)^2 \tag{52}$$

where $\{d_m^i\} = \mathbf{g}(\mathbf{m})$, $\{d_{obs}^i\}$ is the set of observed data and $\{\sigma_d^i\}$ are the traveltime uncertainties (or weights). An inversion solution fits the data to the level of the noise when $\chi^2 = 1$. Once this value is achieved, there is little point in continuing with the iterative inversion process. However, Eq. 52 measures the data fit in an average sense, and it is still possible to have $\chi^2 < 1$ while one or more model traveltime residuals are larger than their respective error estimates. If the data misfit does not fall below $\chi^2 = 1$, then statistical hypothesis testing (e.g. Kreyszig, 1993) can be used to stop the iterative process. For example, if the data misfit has a normal distribution, we can test whether the data variance at iteration n is significantly different from the variance at iteration $n+1$ (usually referred to as an F-test). The F-test has been used by numerous authors (e.g. Thurber, 1983; Eberhart-Phillips, 1986; Steck et al., 1998; Graeber & Asch, 1999).

The following gradient-based inversion methods can be used to determine $\delta\mathbf{m}_n$ in Eq. 51.

2.3.2.1 Gauss-Newton method and damped-least squares The Gauss-Newton method locates the updated point $\mathbf{m}_{n+1}$ by finding the minimum of the tangent paraboloid to $S(\mathbf{m})$ at $\mathbf{m}_n$. At the minimum of S, the gradient will vanish, so $\mathbf{m}$ is required such that:

$$\mathbf{F}(\mathbf{m}) = \mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}) + \epsilon\mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0) + \eta\mathbf{D}^T\mathbf{D}\mathbf{m} = 0 \tag{53}$$

where $\mathbf{F}(\mathbf{m}) = \hat{\boldsymbol{\gamma}}$. If we are at some point $\mathbf{m}_n$, then a more accurate estimate $\mathbf{m}_{n+1}$ can be obtained using a Taylor series expansion of Eq. 53 and ignoring second order terms:

$$F_i(m_{n+1}^1, \ldots, m_{n+1}^M) = F_i(m_n^1, \ldots, m_n^M) + \sum_{j=1}^{M}(m_{n+1}^j - m_n^j)\left.\frac{\partial F_i}{\partial m^j}\right|_{\mathbf{m}_n} = 0 \tag{54}$$

which may be rewritten as:

$$\mathbf{m}_{n+1} = \mathbf{m}_n - \left[\frac{\partial\mathbf{F}}{\partial\mathbf{m}}\right]_n^{-1}[\mathbf{F}_n] = \mathbf{m}_n - \left[\frac{\partial^2 S}{\partial\mathbf{m}^2}\right]_n^{-1}\left[\frac{\partial S}{\partial\mathbf{m}}\right]_n \tag{55}$$

where $(\partial S/\partial\mathbf{m})_n$ is the gradient vector and $(\partial^2 S/\partial\mathbf{m}^2)_n$ is the Hessian matrix. Substituting Eq. 49 and Eq. 50 into Eq. 55 gives the Gauss-Newton solution:

$$\begin{aligned}\delta\mathbf{m}_n \;=\; & -[\mathbf{G}_n^T\mathbf{C}_d^{-1}\mathbf{G}_n + \nabla_{\mathbf{m}}\mathbf{G}_n^T\mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}) + \epsilon\mathbf{C}_m^{-1} + \eta\mathbf{D}^T\mathbf{D}]^{-1} \\ & \times[\mathbf{G}_n^T\mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}] + \epsilon\mathbf{C}_m^{-1}(\mathbf{m}_n - \mathbf{m}_0) + \eta\mathbf{D}^T\mathbf{D}\mathbf{m}_n]\end{aligned} \tag{56}$$

As mentioned earlier, the second derivative term in the Hessian matrix is usually ignored, which gives the quasi-Newton solution:

$$\begin{aligned}\delta\mathbf{m}_n \;=\; & -[\mathbf{G}_n^T\mathbf{C}_d^{-1}\mathbf{G}_n + \epsilon\mathbf{C}_m^{-1} + \eta\mathbf{D}^T\mathbf{D}]^{-1}[\mathbf{G}_n^T\mathbf{C}_d^{-1}[\mathbf{g}(\mathbf{m}_n) - \mathbf{d}_{obs}] \\ & +\epsilon\mathbf{C}_m^{-1}(\mathbf{m}_n - \mathbf{m}_0) + \eta\mathbf{D}^T\mathbf{D}\mathbf{m}_n]\end{aligned} \tag{57}$$

Straightforward implementation of this method requires an $M \times M$ matrix equation to be solved. If the number of model parameters is large, the solution will be computationally expensive, and if data coverage is poor, the problem may well be ill-conditioned.

If instead we assume that the relationship $\mathbf{d} = \mathbf{g}(\mathbf{m})$ is linearizable then (c.f. Eq. 38):

$$\mathbf{d}_{obs} \approx \mathbf{g}(\mathbf{m}_0) + \mathbf{G}(\mathbf{m} - \mathbf{m}_0) \tag{58}$$

or $\delta\mathbf{d} = \mathbf{G}\delta\mathbf{m}$ with $\delta\mathbf{d} = \mathbf{d}_{obs} - \mathbf{g}(\mathbf{m}_0)$ and $\delta\mathbf{m} = \mathbf{m} - \mathbf{m}_0$ (i.e. Eq. 38 with $\mathbf{G} = \partial\mathbf{g}/\partial\mathbf{m}$). If Eq. 58 is exactly linear, then the Newton and quasi-Newton solutions are the same because the second derivative terms in the Hessian matrix are zero. Because a one-step solution is possible in the linear case, the objective function is sometimes written:

$$S(\mathbf{m}) = \frac{1}{2}\left[(\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d})^T\mathbf{C}_d^{-1}(\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d}) + \epsilon\delta\mathbf{m}^T\mathbf{C}_m^{-1}\delta\mathbf{m} + \eta\delta\mathbf{m}^T\mathbf{D}^T\mathbf{D}\delta\mathbf{m}\right] \tag{59}$$

where last term on the RHS smooths the perturbations to the prior model. The functional in this case is:

$$\mathbf{F}(\mathbf{m}) = \mathbf{G}^T\mathbf{C}_d^{-1}(\mathbf{G}\delta\mathbf{m} - \delta\mathbf{d}) + \epsilon\mathbf{C}_m^{-1}\delta\mathbf{m} + \eta\mathbf{D}^T\mathbf{D}\delta\mathbf{m} = 0 \tag{60}$$

and the solution can be written as:

$$\delta\mathbf{m} = [\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \epsilon\mathbf{C}_m^{-1} + \eta\mathbf{D}^T\mathbf{D}]^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}\delta\mathbf{d} \tag{61}$$

When no smoothing is used ($\eta = 0$) and the matrices $\mathbf{C}_d$ and $\mathbf{C}_m$ represent the known *a priori* error statistics, then Eq. 61 becomes:

$$\delta\mathbf{m} = [\mathbf{G}^T\mathbf{C}_d^{-1}\mathbf{G} + \mathbf{C}_m^{-1}]^{-1}\mathbf{G}^T\mathbf{C}_d^{-1}\delta\mathbf{d} \tag{62}$$

which is the maximum likelihood solution to the inverse problem or the *stochastic inverse* (Aki et al., 1977). The expressions for $\delta\mathbf{m}$ in Eq. 57, Eq. 61 or Eq. 62 are often referred to as Damped Least Squares (DLS) solutions to the inverse problem (particularly when $\eta = 0$).

It is interesting to note the differences between the solution given by Eq. 57 and Eq. 61. First, the smoothing term in Eq. 61 smoothes the perturbations to the model, not the model itself. Second, if Eq. 61 is applied iteratively, then the damping regularization is not necessarily the same as that imposed by the iterative implementation of Eq. 57. In the latter, the term $\mathbf{m}_0$ usually represents the initial or starting model and the effect of the regularization is to favor a solution near the initial model (how "near" will depend on the value of ϵ). If, however, $\mathbf{m}_0 = \mathbf{m}_{prior}$, where $\mathbf{m}_{prior}$ is the solution at the previous iteration, then the damping in Eq. 57 will be the same as that for Eq. 61, in which the misfit of the current model compared to the previous model is regulated. In damping, the use of $\mathbf{m}_0$ is often referred to as "jumping" while using $\mathbf{m}_{prior}$ is often referred to as "creeping" (Shaw & Orcutt, 1985). Similarly, smoothing the perturbation to the model is "creeping" while smoothing the model is "jumping". The DLS-type solution scheme is the technique most commonly used to solve the inverse step in seismic tomography. Many other variants to those discussed above have also been used - Spakman (1993) describes and compares several of them.

Studies that have used a DLS-type solution to the inverse problem in teleseismic tomography include those by Aki et al. (1977), Zhao et al. (1994), Weiland et al. (1995), Wiggins et al. (1996) and Steck et al. (1998). Authors who have used DLS solutions to invert wide-angle traveltimes include Kanasewich & Chiu (1985), Chiu et al. (1986), Farra & Madariaga (1988), White (1989), Lutter &

Nowack (1990), Lutter et al. (1990), Zelt & Smith (1992), Lutter et al. (1994), Kosloff et al. (1996), Wang & Braile (1996), McCaughey & Singh (1997) and Zelt & Barton (1998). Similarly, DLS is also popular in local earthquake tomography (Aki & Lee, 1976; Thurber, 1983; Eberhart-Phillips, 1986; Zhao et al., 1992; Graeber & Asch, 1999).

The most computationally expensive part of the DLS solution is to solve a matrix equation of dimension M. If the number of parameters is not large, then the solution may be found using methods like LU decomposition (Press et al., 1992) or Cholesky decomposition (Tarantola, 1987). A particularly useful solution technique for small to mid-size problems is Singular Value Decomposition (Press et al., 1992) or SVD. It is capable of robustly dealing with matrices that are singular or nearly singular, which is often the case in tomographic problems. In addition, SVD can be used to diagnose problems with the system of equations, such as the presence of equations that do not help constrain the solution. Another attractive feature of SVD is that covariance and resolution estimates associated with the solution model may be obtained at virtually no extra cost (e.g. White, 1989).

For models defined by large numbers of parameters, direct solution methods are cumbersome and iterative techniques are more practical. One such method is the conjugate gradient method of Hestenes & Stiefel (1952), which is able to take advantage of the sparse nature of linear systems commonly associated with seismic tomography problems (Scales, 1987). Conjugate gradients and LSQR, a variant of the conjugate gradient algorithm, are probably the most commonly used methods for solving linear systems of the form of Eq. 61 with a large number (e.g. 1000's - 100,000's) of unknowns (Nolet, 1985; Scales, 1987; VanDecar & Snieder, 1994).

The DLS-type solution to the inverse problem need not be formulated as a set of normal equations, such as Eq. 57 or Eq. 62. For example, Eq. 59 may be minimized by finding the least squares solution of the system:

$$\begin{bmatrix} \mathbf{C}_d^{-1/2}\mathbf{G} \\ \sqrt{\epsilon}\mathbf{C}_m^{-1/2} \\ \sqrt{\eta}\mathbf{D} \end{bmatrix} \delta\mathbf{m} = \begin{bmatrix} \mathbf{C}_d^{-1/2}\delta\mathbf{d} \\ 0 \\ 0 \end{bmatrix} \tag{63}$$

which is equivalent to solving Eq. 61. SVD or iterative solvers like LSQR can be used to solve Eq. 63 since they can equally well be applied to non-square systems and will solve the equations in the least-squares sense.

2.3.2.2 Steepest descent The method of steepest descent is probably the simplest gradient-based method for iterative minimization of an objective function. It is based on the idea that the objective function S may be minimized by successive searches along local directions of steepest descent. If $\boldsymbol{\gamma}_n$ is the direction of steepest ascent in model space at a point $\mathbf{m}_n$, the model correction is given by:

$$\delta\mathbf{m}_n = -\mu_n\boldsymbol{\gamma}_n \tag{64}$$

where the scalar μ_n is a positive real number whose value locates the minimum in the direction specified by $\boldsymbol{\gamma}_n$. The parameter μ_n can be found by line minimization as follows. By assuming that $\mathbf{g}(\mathbf{m})$ is locally linear, so that $\mathbf{g}(\mathbf{m}_{n+1}) = \mathbf{g}(\mathbf{m}_n) + \mathbf{G}_n(\mathbf{m}_{n+1} - \mathbf{m}_n)$, then $S(\mathbf{m})$ is quadratic so $\partial S(\mathbf{m})/\partial\mu_n = 0$ will give the value of μ_n. Substituting Eq. 64 into Eq. 48 gives:

$$S(\mathbf{m}_n - \mu_n\boldsymbol{\gamma}_n) = S(\mathbf{m}_n) - \mu_n\hat{\boldsymbol{\gamma}}_n^T\boldsymbol{\gamma}_n + \tfrac{1}{2}\mu_n^2\boldsymbol{\gamma}_n^T\hat{\mathbf{H}}_n\boldsymbol{\gamma}_n \tag{65}$$

The partial derivative of Eq. 65 is:

$$\frac{\partial S(\mathbf{m}_n - \mu_n\boldsymbol{\gamma}_n)}{\partial\mu_n} = -\hat{\boldsymbol{\gamma}}_n^T\boldsymbol{\gamma}_n + \mu_n\boldsymbol{\gamma}_n^T\hat{\mathbf{H}}_n\boldsymbol{\gamma}_n = 0 \tag{66}$$

which can be rearranged to give:

$$\mu_n = \frac{\hat{\boldsymbol{\gamma}}_n^T \boldsymbol{\gamma}_n}{\boldsymbol{\gamma}_n^T \hat{\mathbf{H}}_n \boldsymbol{\gamma}_n} \tag{67}$$

An iterative line search could also be implemented to determine μ_n, but this may require numerous calculations of the forward step. The efficiency of the steepest descent method depends on the character of $S(\mathbf{m})$, but usually it is slow because the local direction of steepest descent may differ greatly from the direction in which the minimum is to be found. Fig. 22 schematically illustrates how the steepest descent method may be inefficient. Blundell (1993) investigates the properties of the steepest descent method in the context of inverting reflection traveltimes for velocity and interface depth but the method has not been widely used in seismic tomography.

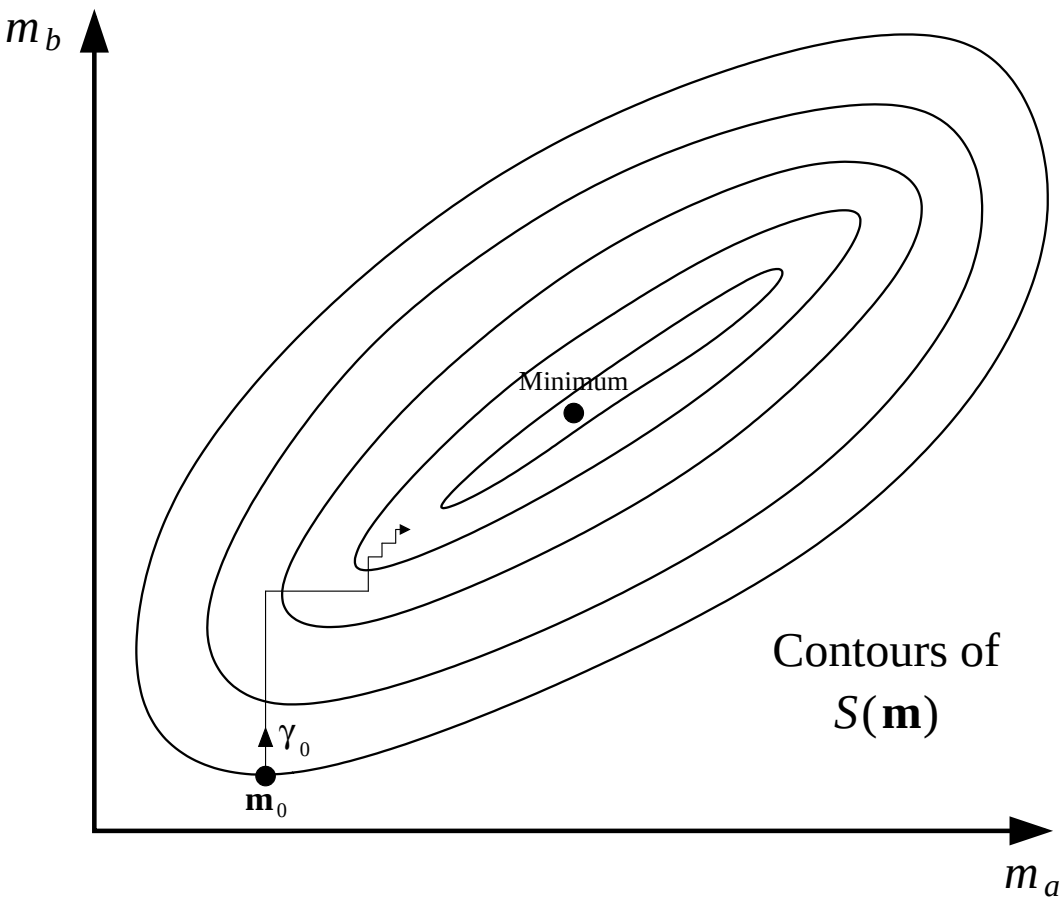

Figure 22: Illustration of the method of steepest descent in 2-D model space. The method can be inefficient in locating the solution if the local direction of steepest descent is oblique to the direction in which the minimum is to be found.

2.3.2.3 Conjugate gradients Hestenes & Stiefel (1952) first suggested the idea of conjugate gradient methods in regard to solving systems of linear equations. Fletcher & Reeves (1964) subsequently applied it to unconstrained optimization. At each iteration, the conjugate gradient method uses a new search direction that is conjugate to all the previous ones. In other words, the nth iteration of the conjugate gradient method locates the minimum in an n-dimensional subspace spanned by the

current search direction and the $n-1$ previous search directions. The algorithm may be defined by (see Tarantola, 1987):

$$\boldsymbol{\phi}_n = \boldsymbol{\gamma}_n + \alpha_n \boldsymbol{\phi}_{n-1} \tag{68}$$

$$\delta \mathbf{m}_n = -\mu_n \boldsymbol{\phi}_n \tag{69}$$

where $\{\boldsymbol{\phi}_n\}$ is a set of conjugate directions and $\boldsymbol{\phi}_0 = \boldsymbol{\gamma}_0$. The scalar μ_n can be determined by Eq. 67 (using $\boldsymbol{\phi}_n$ in place of $\boldsymbol{\gamma}_n$) and α_n is defined so that the new search direction is conjugate to the previous search directions (see Press et al., 1992):

$$\alpha_n = \frac{\boldsymbol{\gamma}_n^T \hat{\boldsymbol{\gamma}}_n}{\boldsymbol{\gamma}_{n-1}^T \hat{\boldsymbol{\gamma}}_{n-1}} \tag{70}$$

or

$$\alpha_n = \frac{\boldsymbol{\gamma}_n^T \hat{\boldsymbol{\gamma}}_n - \boldsymbol{\gamma}_n^T \hat{\boldsymbol{\gamma}}_{n-1}}{\boldsymbol{\gamma}_{n-1}^T \hat{\boldsymbol{\gamma}}_{n-1}} \tag{71}$$

Eq. 70 is the Fletcher-Reeves variant and Eq. 71 is the Polak-Ribiere variant. The latter formulation sometimes gives superior results when $S(\mathbf{m})$ is non-quadratic (Tarantola, 1987). In the context of seismic traveltime inversion, the conjugate gradient method has been used infrequently for direct minimization of an objective function. Rawlinson & Houseman (1998) have used it to invert teleseismic traveltimes and Blundell (1993) has used it in the inversion of reflection traveltimes.

The principal advantage that both the steepest descent and conjugate gradient method have over the Newton and DLS methods is that a large system of linear equations does not need to be solved. Consequently, these methods are much more rapid at the inverse step. However, since they only minimize in one dimension at each iteration, they tend to converge more slowly. If the procedure for solving the forward step is fast compared to solving the $M \times M$ matrix equation, then steepest descent or conjugate gradients may be preferable. However, if the forward step is slow compared to solving the matrix inverse, then a Newton or DLS approach may be quicker.

2.3.2.4 Subspace method Both the steepest descent method and conjugate gradient method are examples of 1-D subspace methods in that they perform a line minimization at each iteration. In general, however, subspace methods may be constructed in which the minimization is carried out simultaneously along several search directions that together span a subspace of the model space. The basic theory for the general subspace inversion method is presented here; more details can be found in Kennett et al. (1988), Sambridge (1990) and Williamson (1990).

At each iteration, the subspace method restricts the minimization of the quadratic approximation of $S(\mathbf{m})$ to a p-dimensional subspace of model space, so that the perturbation $\delta\mathbf{m}$ (ignoring the iteration subscript n in $\delta\mathbf{m}_n$ for convenience) occurs in the space spanned by a set of p M-dimensional basis vectors $\{\mathbf{a}^j\}$:

$$\delta \mathbf{m} = \sum_{j=1}^{p} \mu_j \mathbf{a}^j = \mathbf{A}\boldsymbol{\mu} \tag{72}$$

where $\mathbf{A} = [\mathbf{a}^j]$ is the $M \times p$ projection matrix. The component μ_j determines the length of the corresponding vector $\mathbf{a}^j$ that minimizes the quadratic form of $S(\mathbf{m})$ in the space spanned by $\mathbf{a}^j$. Hence, $\boldsymbol{\mu}$ is found by substituting Eq. 72 into Eq. 48, which gives in summation form:

$$S(\mathbf{m} + \delta\mathbf{m}) = S(\mathbf{m}) + \sum_{j=1}^{p} \mu_j \hat{\boldsymbol{\gamma}}^T \mathbf{a}^j + \tfrac{1}{2} \sum_{j=1}^{p} \sum_{k=1}^{p} \mu_j \mu_k [\mathbf{a}^k]^T \hat{\mathbf{H}} [\mathbf{a}^j] \tag{73}$$

and locating the minimum of S with respect to $\boldsymbol{\mu}$:

$$\frac{\partial S(\mathbf{m})}{\partial \mu_q} = \hat{\boldsymbol{\gamma}}^T \mathbf{a}^q + \sum_{k=1}^{p} \mu_k [\mathbf{a}^k]^T \hat{\mathbf{H}} [\mathbf{a}^j] = 0 \tag{74}$$

for $q = 1, \ldots, p$. Rearranging Eq. 74 for $\boldsymbol{\mu}$ gives:

$$\boldsymbol{\mu} = -[\mathbf{A}^T \hat{\mathbf{H}} \mathbf{A}]^{-1} \mathbf{A}^T \hat{\boldsymbol{\gamma}} \tag{75}$$

and since $\delta\mathbf{m} = \mathbf{A}\boldsymbol{\mu}$, the solution, taking $\hat{\mathbf{H}} = \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}$, is:

$$\delta\mathbf{m} = -\mathbf{A}[\mathbf{A}^T (\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1} + \eta \mathbf{D}^T \mathbf{D}) \mathbf{A}]^{-1} \mathbf{A}^T \hat{\boldsymbol{\gamma}} \tag{76}$$

which can be used iteratively in the manner specified by Eq. 51. The quantities $\mathbf{A}$, $\hat{\boldsymbol{\gamma}}$ and $\mathbf{G}$ are re-evaluated between successive iterations. Most implementations of the subspace method construct the basis vectors $\{\mathbf{a}^j\}$ in terms of the steepest ascent vector in model space $\boldsymbol{\gamma}$ and its rates of change (e.g. Kennett et al., 1988; Sambridge, 1990; Williamson, 1990; Blundell, 1993).

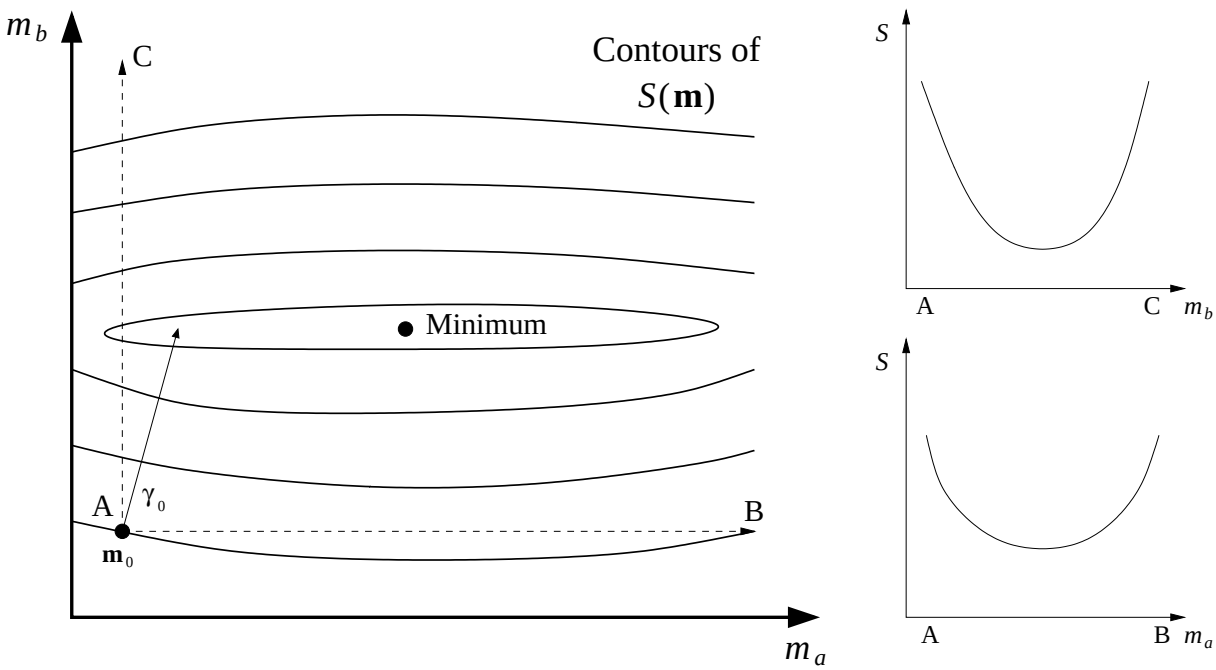

Figure 23: A contour plot of $S(\mathbf{m})$ which is a function of two parameters of different physical dimensions. $S(\mathbf{m})$ is much more sensitive to m_b than m_a, and a gradient method like steepest descents will converge slowly. Searching in directions specified by basis vectors (dotted lines) that only lie in space defined by a single parameter class eliminates these problems, since $S(\mathbf{m})$ is a function of only one parameter class in each of these directions.

The subspace method has several desirable characteristics. First, the determination of $\delta\mathbf{m}$ only requires the solution of a relatively small $p \times p$ system of linear equations - Williamson (1990) uses $p = 6$ and Blundell (1993) uses $p \leq 8$. Second, it offers a natural way of dealing with multiple parameter classes, such as velocity parameters and interface depth parameters, that are to be inverted for simultaneously. If the basis vectors $\{\mathbf{a}^j\}$ are chosen such that each vector only lies in the space spanned by a particular parameter class, then the minimization will account for the

different sensitivities of $S(\mathbf{m})$ with respect to the different parameter classes in a balanced way (see Fig. 23). Other gradient methods in which parameters of different physical dimensions are mixed can exhibit slow convergence and a strong dependence on relative scaling of the different parameter types (Kennett et al., 1988).

As an example of how one might choose a set of basis vectors, consider an inverse problem that has three different parameter classes that are to be constrained by the data. Three separate search directions can simply be obtained by partitioning the gradient vector in model space $\boldsymbol{\gamma} = \mathbf{C}_m\hat{\boldsymbol{\gamma}}$ on the basis of parameter class:

$$\boldsymbol{\gamma} = \mathbf{a}^1 + \mathbf{a}^2 + \mathbf{a}^3 = \begin{bmatrix} \boldsymbol{\gamma}^1 \\ 0 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ \boldsymbol{\gamma}^2 \\ 0 \end{bmatrix} + \begin{bmatrix} 0 \\ 0 \\ \boldsymbol{\gamma}^3 \end{bmatrix} \tag{77}$$

where $\mathbf{a}^1$, $\mathbf{a}^2$ and $\mathbf{a}^3$ represent ascent vectors that lie in the space of only one parameter type. A further nine basis vectors can be obtained by pre-multiplying $\mathbf{a}^1$, $\mathbf{a}^2$ and $\mathbf{a}^3$ by the model space Hessian $\mathbf{H} = \mathbf{C}_m\hat{\mathbf{H}}$ and partitioning the three vectors that result in the same way as Eq. 77. Additional basis vectors can be produced by repeating the process of pre-multiplication of the latest set of vectors by the partitioned model space Hessian. Once the required number of basis vectors is obtained, Gram-Schmidt orthogonalization should be applied to avoid interdependence:

$$\mathbf{a}^i_{orth} = \mathbf{a}^i - \sum_{j=1}^{i-1} \frac{\mathbf{a}^i \cdot \mathbf{a}^j}{\mathbf{a}^j \cdot \mathbf{a}^j}\mathbf{a}^j \tag{78}$$

for $i = 1, \ldots, p$. The basis vectors $\{\mathbf{a}^i_{orth}\}$ are normalized and used in Eq. 76 via the projection matrix $\mathbf{A}$. If the subspace is large, then SVD may be more effective in finding the orthonormal set of subspace vectors due to the build-up of round-off error associated with the numerical implementation of Eq. 78 (Press et al., 1992). Choosing an appropriate number of basis vectors requires finding an acceptable balance between computational effort and rate of convergence.

Subspace methods have been applied to the inversion of reflection traveltimes for velocity and interface structure (Williamson, 1990; Blundell, 1993), to the inversion of local earthquake and artificial source traveltimes for velocity, interface structure and hypocenter location (Sambridge, 1990) and to the inversion of reflection amplitude data for interface structure (Wang & Houseman, 1994) and velocity (Wang & Houseman, 1995). Rawlinson et al. (2001a,b) use it in the inversion of refraction and wide-angle reflection traveltimes for the determination of layered crustal structure.

2.3.2.5 Fréchet matrix All gradient methods require the calculation of $\mathbf{G} = \partial\mathbf{g}/\partial\mathbf{m}$, which describes the rate of change of traveltimes with respect to model parameters. For such discrete models, $\mathbf{G}$ defines the "Gateaux" derivative of $\mathbf{g}$; the term "Fréchet" derivative should really only be used to describe the derivative of $\mathbf{g}$ for a continuous model (Shaw & Orcutt, 1985). However, since $\mathbf{G}$ is widely referred to as the Fréchet derivative/matrix even if the model is discrete, we will continue to do so here. The two basic parameter types that are normally encountered in traveltime inversion, especially of wide-angle data, are velocity (or slowness) and interface depth. First-order accurate expressions for the derivatives can be derived quite simply for both cases. They are usually calculated as part of the forward step of the tomographic problem. In local earthquake tomography, the source location is also an unknown, so derivatives of traveltime with respect to these parameters are required.

The linearized relationship between traveltime residual and velocity perturbation was given by Eq. 6:

$$\delta t = -\int_{L(v_0)} \frac{\delta v}{v_0^2} \mathrm{d}l \tag{79}$$

where δv is the velocity perturbation and v_0 is the reference velocity field. If the velocity field is defined by a grid of velocity nodes, then to first-order the Fréchet derivatives are given by:

$$\frac{\partial t}{\partial v_n} = -\int_{L(v_0)} v_0^{-2} \frac{\partial v}{\partial v_n} \mathrm{d}l \tag{80}$$

where v_n is the velocity of a particular node and $\partial v / \partial v_n$ is the change of velocity along the ray with respect to a change in v_n. This expression is usually straightforward to calculate if the velocity interpolation function $v = f(\mathbf{v}_{nodes})$ has a simple form (e.g. cubic B-splines). Fréchet derivatives have been calculated using Eq. 80 (or its equivalent for slowness) by most authors who use gradient methods and models parameterized by a grid of velocity nodes (e.g. White, 1989; Lutter et al., 1990; Sambridge, 1990).

First-order accurate analytic expressions can also be obtained for the Fréchet derivatives when the model parameters describe interface depths. The basic approach is to partition the problem:

$$\frac{\partial t}{\partial z_n} = \frac{\partial t}{\partial h_{int}} \frac{\partial h_{int}}{\partial z_{int}} \frac{\partial z_{int}}{\partial z_n} \tag{81}$$

where z_n is the depth coordinate of the interface node, h_{int} is displacement normal to the interface at the point of intersection by the ray and z_{int} is the depth coordinate of the intersection point. The first two derivatives on the RHS of Eq. 81 can be worked out analytically to first-order accuracy by assuming a locally linear wavefront and interface (e.g. Bishop et al., 1985; Nowack & Lyslo, 1989; White, 1989; Sambridge, 1990; Zelt & Smith, 1992; Blundell, 1993; Riahi & Juhlin, 1994). Consider Fig. 24 which shows a plane wave impinging on a planar interface that is perturbed by a distance Δh. Rays A and B show the path taken by the wave before and after the perturbation. It is easy to see that the difference in traveltime Δt between rays A and B from position 1 to position 2 is:

$$\Delta t = \frac{a}{v_j} - \frac{b}{v_{j+1}} \tag{82}$$

and since $a = -\mathbf{w}_j \cdot \mathbf{w}_n \Delta h$ and $b = -\mathbf{w}_{j+1} \cdot \mathbf{w}_n \Delta h$ (all vectors are unit vectors), substitution into Eq. 82 gives:

$$\Delta t = \left[\frac{\mathbf{w}_{j+1} \cdot \mathbf{w}_n}{v_{j+1}} - \frac{\mathbf{w}_j \cdot \mathbf{w}_n}{v_j} \right] \Delta h \tag{83}$$

and the approximation to the derivative is:

$$\frac{\partial t}{\partial h_{int}} \approx \frac{\mathbf{w}_{j+1} \cdot \mathbf{w}_n}{v_{j+1}} - \frac{\mathbf{w}_j \cdot \mathbf{w}_n}{v_j} \tag{84}$$

The second term in Eq. 81 can be derived from the fact that $\mathbf{w}_n \cdot \mathbf{w}_z = -\Delta h / \Delta z$ so the derivative is:

$$\frac{\partial h_{int}}{\partial z_{int}} \approx -\mathbf{w}_n \cdot \mathbf{w}_z \tag{85}$$

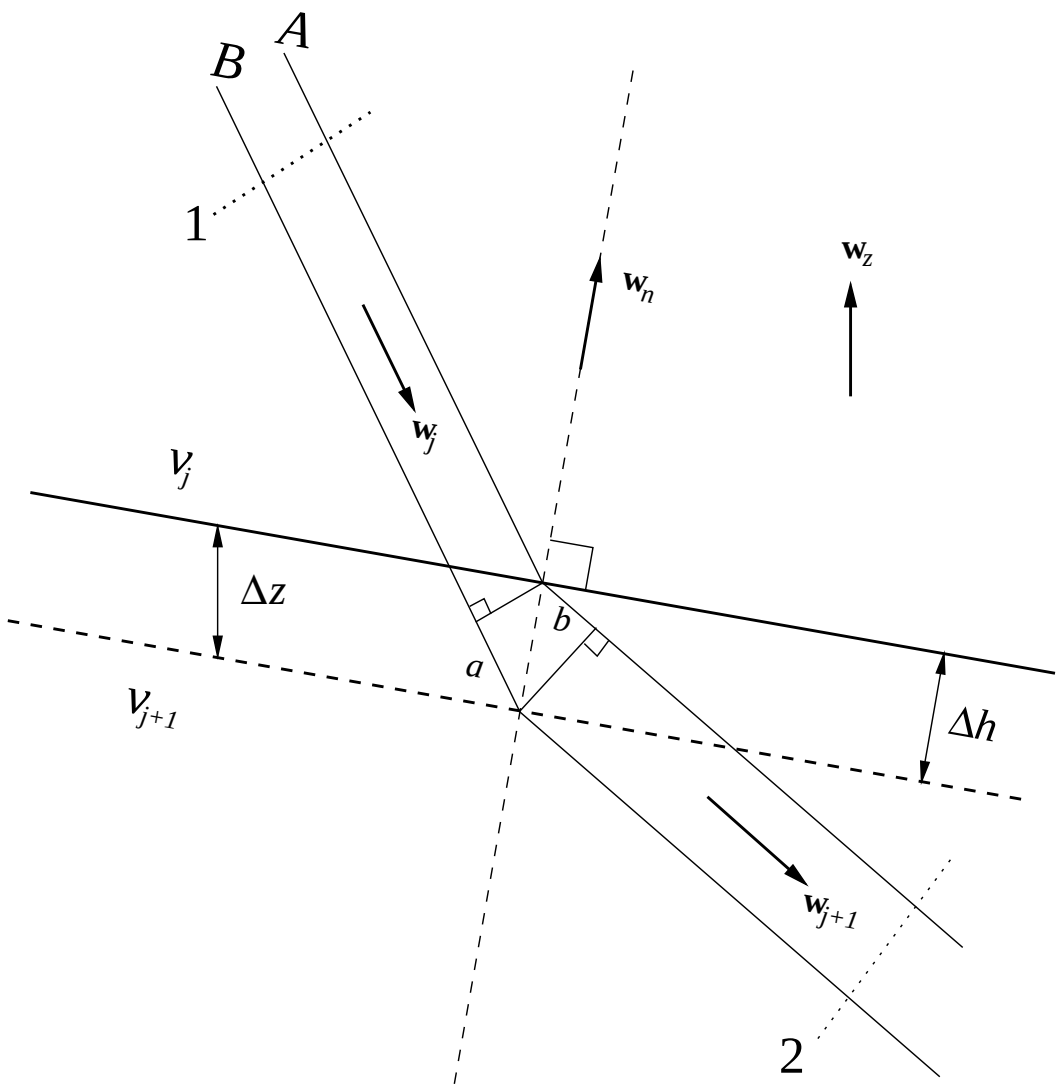

Figure 24: Plane wave incident on a perturbed planar interface for first-order approximation of interface Fréchet derivatives. $\mathbf{w}_j$ is a unit vector parallel to rays A and B in layer j and $\mathbf{w}_{j+1}$ is a unit vector parallel to rays A and B in layer $j+1$. $\mathbf{w}_n$ is the unit normal vector to the interface at the intersection point and $\mathbf{w}_z = [0, 0, 1]$.

Substitution of both partial derivatives into Eq. 81 gives:

$$\frac{\partial t}{\partial z_n} \approx \left[\frac{\mathbf{w}_j \cdot \mathbf{w}_n}{v_j} - \frac{\mathbf{w}_{j+1} \cdot \mathbf{w}_n}{v_{j+1}}\right]\left[\mathbf{w}_n \cdot \mathbf{w}_z\right]\frac{\partial z_{int}}{\partial z_n} \tag{86}$$

Note that Eq. 86 will work for any ray direction provided $\mathbf{w}_j$ always points towards the interface and $\mathbf{w}_{j+1}$ always points away from the interface, irrespective of whether the ray is upgoing or downgoing. For reflections, $\mathbf{w}_{j+1} \cdot \mathbf{w}_n = -\mathbf{w}_j \cdot \mathbf{w}_n$ and $v_{j+1} = v_j$. The term $\partial z_{int}/\partial z_n$ depends on the form of the interface depth interpolation function $z = f(\mathbf{z}_{nodes})$.

In 3-D space, four parameters describe the location of an earthquake hypocenter. The partial derivative of the traveltime t with respect to origin time t_o is simply:

$$\frac{\partial t}{\partial t_o} = -1 \tag{87}$$

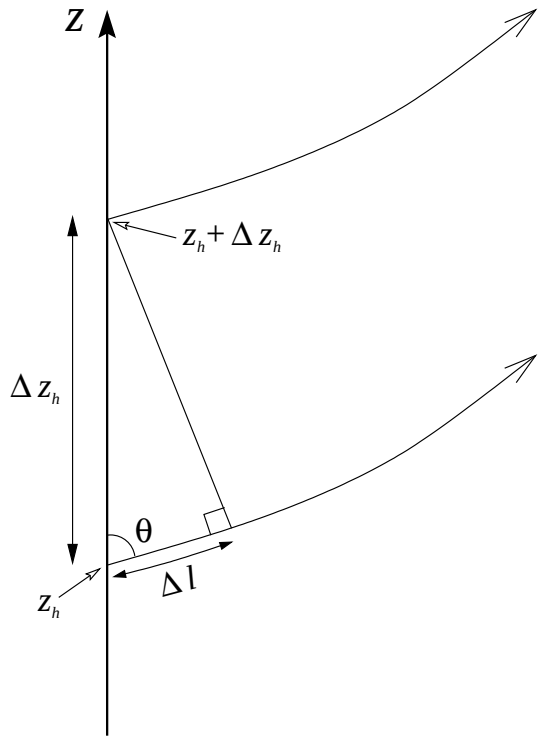

Figure 25: A change in depth Δz_h of the hypocenter results in a change in the path length Δl of the ray.

because $t = t_a - t_o$ where t_a is the arrival time. In Cartesian coordinates, the partial derivatives of traveltime with respect to the three spatial coordinates are straightforward to derive. For example, consider a perturbation in depth Δz_h of the hypocenter (see Fig. 25). The corresponding change in path length Δl is given by:

$$\Delta l = -\Delta z_h \cos\theta_h \tag{88}$$

where θ_h is the angle between the ray and the z-axis at the hypocenter. Note that the change in path length is negative since the hypocenter is perturbed to a shallower depth. The change in traveltime is given by:

$$\Delta t = \frac{\Delta l}{v_h} = -\frac{\Delta z_h \cos\theta_h}{v_h} \tag{89}$$

where v_h is the velocity at the hypocenter. The first-order accurate expression for the partial derivative is thus given by:

$$\frac{\partial t}{\partial z_h} = -\frac{\cos\theta_h}{v_h} \tag{90}$$

Similarly, the partial derivatives for the remaining two parameters are:

$$\frac{\partial t}{\partial x_h} = -\frac{\cos\phi_h}{v_h} \tag{91}$$

$$\frac{\partial t}{\partial y_h} = -\frac{\cos\psi_h}{v_h} \tag{92}$$

where ϕ_h and ψ_h subtends the horizontal projection of the ray path and the x and y axes respectively at the hypocenter. Most local earthquake tomography schemes use this kind of hypocenter partial derivative (e.g. Thurber, 1983; Eberhart-Phillips, 1986; Sambridge, 1990).

2.3.3 Global Optimization

The inversion methods described in Section 2.3.1 and Section 2.3.2 are local in that they exploit information in regions of model space near an initial model estimate and thus avoid an extensive search of model space. Consequently, they cannot guarantee convergence to a global minimum solution. Local methods are prone to entrapment in local minima, especially if the subsurface velocity structure is complex and the starting model is not close to the true model. Fig. 26 illustrates these problems.

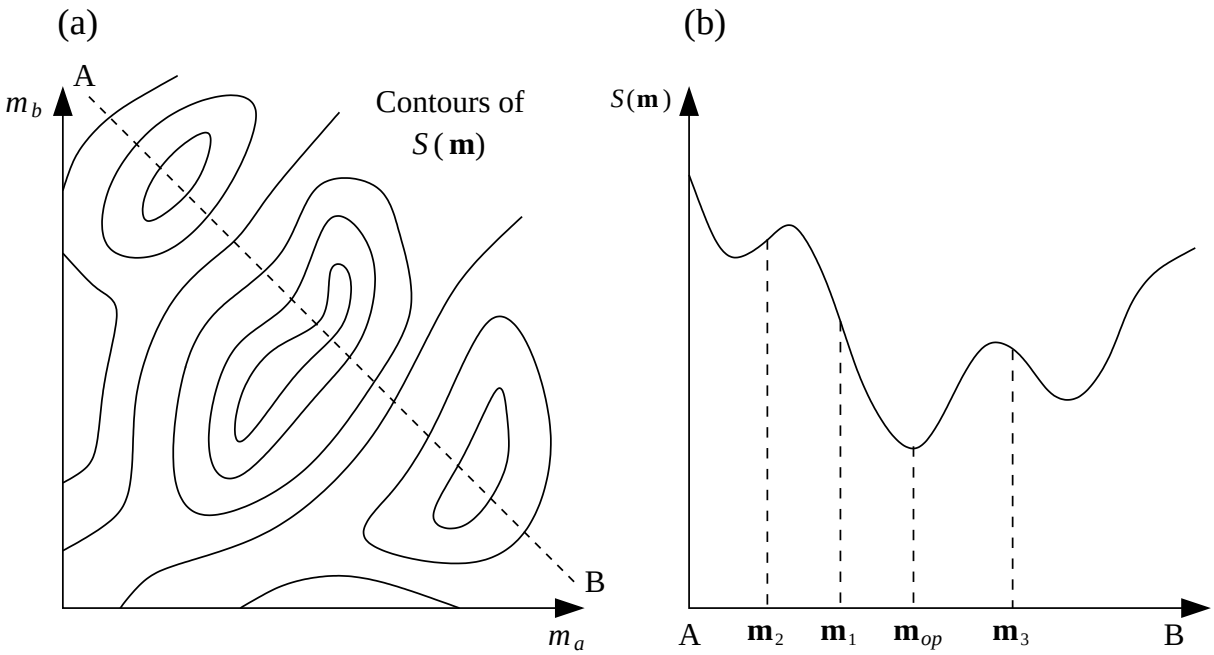

Figure 26: (a) Contours of $S(\mathbf{m})$ for a two parameter model. (b) Cross section through (a). Local methods are likely to locate the global optimum solution $\mathbf{m}_{op}$ if initial model $\mathbf{m}_1$ is used, but are more likely to find local minima if initial models $\mathbf{m}_2$ or $\mathbf{m}_3$ are used.

In many realistic applications of traveltime tomography, particularly at regional and global scales, the need for global optimization techniques is hard to justify, because the *a priori* model information is relatively accurate and lateral heterogeneities are not very large (e.g. Widiyantoro & van der Hilst, 1997; Gorbatov et al., 2000). Therefore, the local minimum of the objective function in the vicinity of the initial model is also likely to be the global minimum of the objective function. However, the crust and lithosphere are generally less well constrained by *a priori* information and are also much more heterogeneous. This means that the initial model is likely to be more distant from the global minimum solution, and entrapment in a local minimum becomes more of a concern. A second motivation for using global methods is that they often produce an ensemble of solutions that satisfy the data to a similar level. This enables one to choose the model deemed most likely to represent the geology of the region (Pullammanappallil & Louie, 1993), and estimate posterior model uncertainties.

The computational burden of exploring large regions of model space is immense, especially for the large number of unknowns typically encountered in traveltime tomography. Recent interest in the use of global optimization techniques for solving geophysical inverse problems has been generated to a large degree by rapid advancements in computing power. Further advancements will continue to

make these techniques more practical and hence popular. Several types of global optimization that have been applied to seismic inversion use random processes to search model space and find better models. These Monte Carlo (MC) methods include genetic algorithms and simulated annealing. For recent reviews on the use of MC techniques for geophysical inverse problems, see Sambridge & Mosegaard (2001) and Mosegaard & Sambridge (2002).

The earliest Monte Carlo approach, defined simply as a uniform random search of model space, is the simplest of these global methods. The misfit function is evaluated at a set of points in model space that are randomly chosen between pairs of upper and lower bounds (chosen *a priori*). The selection of new points has no dependence on previous points. Each model generated by this process is tested for fitness and then accepted or rejected. The final set of accepted models can then be used for interpretation. If model space is not very large, say $M = 80$, and each m_i is discretized to assume only 10 possible values, the number of different models is 10^{80}. Of course, the uniform random search approach will not test every model, but it is likely to spend significant amounts of computation time exploring unfavorable regions of model space. Computation requirements will therefore become prohibitive if large numbers of model parameters are involved (Sambridge & Drijkoningen, 1992). Consequently, a basic Monte Carlo approach is not suited to most seismic traveltime inversion problems.

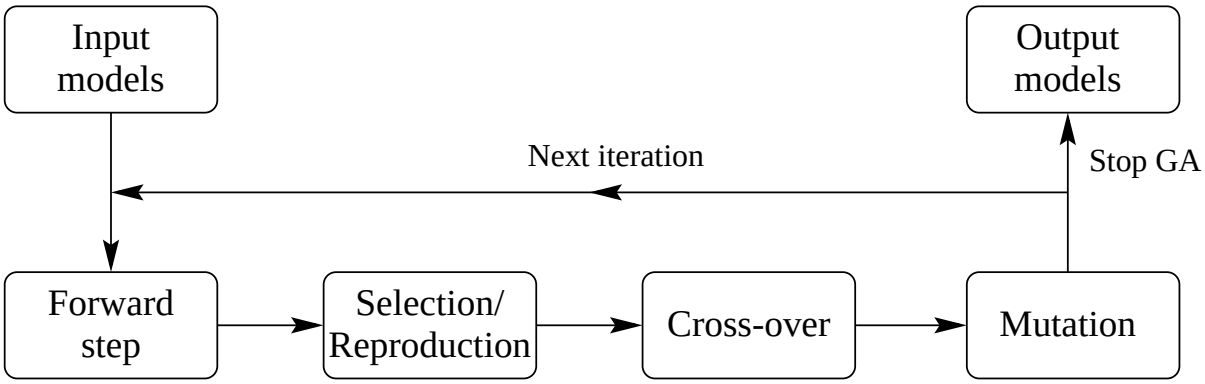

Figure 27: Flow chart for genetic algorithm solution of the inversion problem. Each step is carried out on a population of models and the process is terminated when the maximum fitness parameter of the population exceeds a given threshold.

Genetic algorithms use an analogue to biological evolution to develop new models from an initial pool of randomly picked models. The process of producing a new set of models from a pre-existing set involves four basic steps, summarized in the flowchart of Fig. 27. The first step is to solve the forward problem (i.e. determine traveltimes) for the set of input models. The next step, selection and reproduction, assigns a measure of fitness to each model in the pool based on the magnitude of the misfit function at the corresponding point in model space. Then a rule for selecting which models are to be used to create the next generation is applied. Two common choices for the selection operator are linear normalization selection and parent selection (Boschetti et al., 1996). Linear normalization selection ranks each model according to its fitness and then allows each model to generate a number of offspring proportional to its rank position. Parent selection causes pairs of models to be mated randomly so that each couple creates two offspring in the cross-over stage. The cross-over step creates a pool of offspring, each of which is a “mix” of its two parents. Sambridge & Drijkoningen (1992) represent models by binary strings and perform the cross-over by simply

cutting and transposing two segments at a randomly chosen point along the two strings.

Mutation involves randomly changing some parameter values (or bits if a binary string representation is used) in selected models. This ensures that some "freshness" remains in the model pool, although the mutation rate should be kept low (Boschetti et al., 1996, use a mutation rate of 1%) so that "good" models are not corrupted. The basic principle behind genetic algorithms is that models with a high fitness index will pass their characteristics to subsequent generations, while those aspects of the cycle that introduce random changes to the population permit new parts of the model space to be tested. For more detailed information on genetic algorithms, refer to Goldberg (1989) and Whitley (1994).

Simulated annealing is based on an analogy with physical annealing in thermodynamic systems to guide variations to the model parameters. The process of annealing in metallurgy involves the slow cooling of a metal, which allows the atoms to order themselves into stable, structurally strong, low energy configurations. In the analogue process, some starting model is represented by a collection of atoms (parameters) in equilibrium at a given temperature T. At each iteration, an atom is displaced (a parameter is varied) and the resultant change in the energy of the system ΔE (change in the objective function) is computed. If $\Delta E \leq 0$ the displacement is accepted and the new model is used as the starting point of the next iteration. If $\Delta E > 0$, then acceptance is probabilistic. Kirkpatrick et al. (1983) use the probability $P(\Delta E) = \exp(-\Delta E/K_B T)$ where T is analogous to temperature in controlling the probability of a randomly selected move and the constant K_B is analogous to Boltzmann's constant. By repeating this step many times, the model evolves with the variation of parameters simulating the thermal motion of atoms held at a temperature T. The simulated annealing process consists of "melting" the system at a high T, then progressively lowering T until the system reaches an equilibrium state (it "freezes"). At each T, the simulation is iterated until a steady state is reached before moving to the next temperature level. More detailed explanations of this method can be found in Kirkpatrick et al. (1983) and Aarts & Korst (1989).

Global optimization using stochastic methods is a rapidly developing field of science. However, current applications to seismic traveltime inversion problems have been limited due to computational expense. Genetic algorithms have been used in the 1-D inversion of marine refraction waveforms (Sambridge & Drijkoningen, 1992; Drijkoningen & White, 1995) and the 2-D inversion of refraction traveltimes (Boschetti et al., 1996). The latter paper inverts for 45 model parameters. While this number of parameters is insufficient to adequately parameterize many seismic datasets, it is possible to use such a coarse model as a starting point for subsequent refinement using local optimization techniques. The idea behind this two-stage inversion is that the globally optimized coarse model will be near enough the global minimum to allow the local method to locate the global solution. Boschetti et al. (1996) use genetic algorithms in this context. Simulated annealing has been used by Pullammanappallil & Louie (1993) in the inversion of reflection traveltimes for 2-D velocity structure and interface geometry. Asad et al. (1999) use simulated annealing to produce a coarse 3-D model from local earthquake traveltimes before refining it with the gradient method of Thurber (1983), in a scheme similar in principle to that of Boschetti et al. (1996). Global optimization techniques can find global minimum solutions to highly non-linear inverse problems, but the computational expense when large numbers of parameters are involved currently limits their use in seismic data inversion.

2.4 Analysis of Solution Quality

The process of producing a solution to an inverse problem using the above methods is not complete until some estimate of solution robustness or quality is made. Simply producing a single solution

that minimizes an objective function (i.e. best satisfies the data and *a priori* constraints) without knowledge of resolution or non-uniqueness is inadequate. Two approaches are commonly used to assess solution robustness in traveltime tomography. The first approach assumes local linearity to estimate model covariance and resolution; the second tests resolution by reconstructing a synthetic model using the same source-receiver geometry as the real experiment.

2.4.1 Resolution and Covariance Matrices

To derive expressions for posterior model covariance and resolution, we will assume that the objective function is of the form of Eq. 47 with $\eta = 0$. The reason this is done is because it is more straightforward and common to consider covariance and resolution matrices in a Bayesian framework. In other words, the error statistics associated with the *a priori* model information and data are well known, thus allowing the two sets of information to be objectively combined to produce a more accurate posterior model distribution. In such circumstances, $\epsilon = 1$ and $\mathbf{C}_m$ represents the true *a priori* model covariance. In the following derivation, however, we will retain ϵ as a variable to examine its effect on covariance and resolution if it is used as a tuning parameter in the inversion. For an objective function of the form Eq. 47 with $\eta = 0$, the maximum likelihood solution is given by the $\mathbf{m}$ that satisfies Eq. 53 with $\eta = 0$, which may be written as:

$$\epsilon \mathbf{C}_m^{-1}(\mathbf{m} - \mathbf{m}_0) = -\mathbf{G}^T \mathbf{C}_d^{-1}(\mathbf{g}(\mathbf{m}) - \mathbf{d}_{obs}) \tag{93}$$

Adding $\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}(\mathbf{m} - \mathbf{m}_0)$ to both sides gives:

$$\mathbf{m} - \mathbf{m}_o = [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1} [\mathbf{d}_{obs} - \mathbf{g}(\mathbf{m}) + \mathbf{G}(\mathbf{m} - \mathbf{m}_o)] \tag{94}$$

which is an implicit equation for $\mathbf{m}$. Following Tarantola (1987), let $\mathbf{m}_{true}$ represent the true model, which is unknown. The observed data are related to $\mathbf{m}_{true}$ by:

$$\mathbf{d}_{obs} = \mathbf{g}(\mathbf{m}_{true}) + \boldsymbol{\zeta} \tag{95}$$

where $\boldsymbol{\zeta}$ represents observational and model representation errors. The resolution operator $\mathbf{r}$ defines the relationship between the calculated solution $\mathbf{m}$ and the true solution:

$$\mathbf{m} = \mathbf{r}(\mathbf{m}_{true}) \tag{96}$$

If $\mathbf{r}$ is linear, then $\mathbf{m} = \mathbf{r}(\mathbf{m}_0) + \mathbf{R}(\mathbf{m}_{true} - \mathbf{m}_0)$, where $\mathbf{R} = \partial \mathbf{r} / \partial \mathbf{m}$ and $\mathbf{m}_0 = \mathbf{r}(\mathbf{m}_0)$ so that:

$$\mathbf{m} - \mathbf{m}_0 = \mathbf{R}(\mathbf{m}_{true} - \mathbf{m}_0) \tag{97}$$

If we assume that $\boldsymbol{\zeta} = \{0\}$ in Eq. 95 and let $\mathbf{m} = \mathbf{m}_{true}$ on the RHS of Eq. 94, then Eq. 97 can be written in the form of Eq. 94 with the resolution matrix $\mathbf{R}$ given by:

$$\mathbf{R} = [\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1}]^{-1} \mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} \tag{98}$$

The diagonal elements of R range between zero and one. If $\mathbf{R} = \mathbf{I}$, then, according to Eq. 97, $\mathbf{m} = \mathbf{m}_{true}$ and the solution model is perfectly resolved. If $\mathbf{R} \neq \mathbf{I}$, then the model parameter estimates represent weighted averages of the true model parameters.

The matrix $\mathbf{C}_m$ describes *a priori* model covariance, with the square root of the diagonal entries indicating the uncertainty associated with the initial model parameter values. The constraints supplied by the data will result in changes to these uncertainties. The *a posteriori* covariance matrix $\mathbf{C}_M$

describes the error in the solution parameters and is related to the resolution matrix by (Tarantola, 1987):

$$\mathbf{R} = \mathbf{I} - \mathbf{C}_M \mathbf{C}_m^{-1} \tag{99}$$

Substituting Eq. 98 for $\mathbf{R}$ and solving for $\mathbf{C}_M$ gives:

$$\mathbf{C}_M = \epsilon[\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1}]^{-1} \tag{100}$$

The diagonal elements of $\mathbf{C}_M$ indicate the posterior uncertainty associated with each model parameter. The Fréchet matrix $\mathbf{G}$ in Equations 98 and 100 is calculated at the solution point. Off-diagonal elements of the posterior covariance matrix are more conveniently interpreted in terms of correlations (Tarantola, 1987):

$$\rho^{ij} = \frac{C_M^{ij}}{(C_M^{ii})^{\frac{1}{2}} (C_M^{jj})^{\frac{1}{2}}} \tag{101}$$

where $-1 \leq \rho^{ij} \leq 1$ and $i, j = 1, \ldots, M$. A strong correlation between uncertainties means that the two parameters have not been independently resolved by the dataset.

If ϵ is treated as a damping factor in the inversion (i.e. its value is varied to tune the solution), then $\mathbf{C}_m$ no longer truly represents the *a priori* model covariance. In this case, the resolution and posterior covariance will have a dependence on the value chosen for ϵ. From the above definition of $\mathbf{R}$ (Eq. 98), as $\epsilon \to 0$, then $\mathbf{R} \to \mathbf{I}$ and the solution approaches perfect resolution. As $\epsilon \to \infty$, then $\mathbf{R} \to \mathbf{0}$ and the model is not resolved by the data at all. If we rearrange Eq. 100 as:

$$\mathbf{C}_M^{-1} = \frac{\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G}}{\epsilon} + \mathbf{C}_m^{-1} \tag{102}$$

we see that as $\epsilon \to 0$, the *a priori* covariance becomes increasingly irrelevant to the value of the posterior covariance, whereas when $\epsilon \to \infty$, $\mathbf{C}_M \to \mathbf{C}_m$. From a Bayesian viewpoint, having $\epsilon \to 0$ means that there is no *a priori* model information and the information contained in the data is totally responsible for the state of the posterior model information. In contrast, $\epsilon \to \infty$ means that there are no errors associated with the *a prior* model information, in which case the data are irrelevant. These two end member states are not possible in practice, and reflect the fact that the inclusion of a damping parameter is not consistent with a Bayesian paradigm. In short, the use of ϵ as a damping parameter to tune the solution makes the absolute values of resolution and posterior covariance rather meaningless. However, their relative values will still be useful indicators of the effect the data have had in constraining the solution model. For example, it is reasonable to interpret parameters associated with diagonal elements of the resolution matrix that are large as meaning that the data have been more effective in constraining them than those parameters with smaller resolution values.

The principal difficulties with $\mathbf{R}$ and $\mathbf{C}_M$ are that (i) they are derived from linear theory and are less meaningful as the non-linearity of the problem increases (Snieder, 1998, has considered perturbation theory as a means of extending these concepts to a non-linear regime), (ii) errors in model representation are not taken into account, and (iii) they require the inversion of an $M \times M$ matrix, $\mathbf{G}^T \mathbf{C}_d^{-1} \mathbf{G} + \epsilon \mathbf{C}_m^{-1}$, which may be impractical for large numbers of parameters. Nevertheless, they have been used in many teleseismic traveltime inversions (e.g. Aki et al., 1977; Benz et al., 1992; Steck et al., 1998), wide-angle traveltime inversions (e.g. White, 1989; Hole, 1992; Zelt & Smith, 1992; Riahi & Lund, 1994; Wang & Braile, 1996) and local earthquake tomography studies

(e.g. Eberhart-Phillips, 1990; Eberhart-Phillips & Michael, 1993; Abers, 1994; Protti et al., 1996; Graeber & Asch, 1999). McCaughey & Singh (1997) and Zhang & Toksöz (1998) also consider correlations (Eq. 101) as part of their interpretation of solution quality.

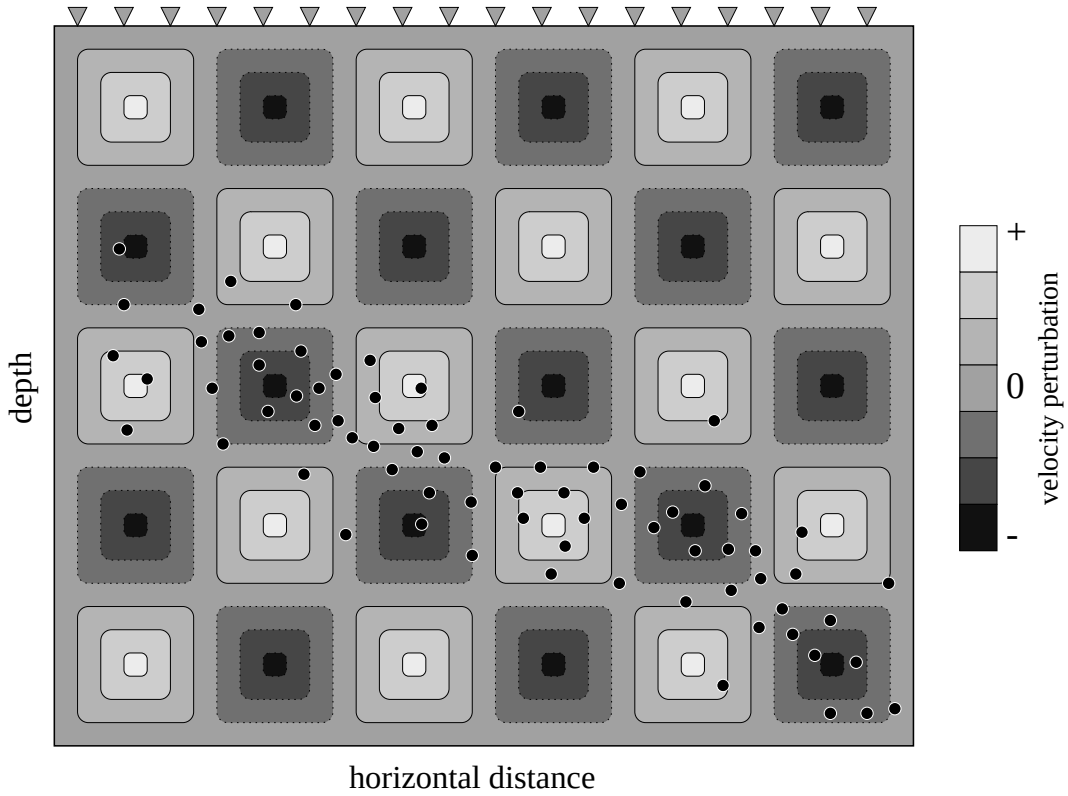

Figure 28: Schematic diagram showing a checkerboard test model for a local earthquake tomography scenario. Sources (black dots with white edges) and receivers (grey-filled triangles) are positioned identically to the real experiment and rays are traced through the synthetic structure. The synthetic traveltimes are then inverted, starting from some given initial model, in an attempt to recover the checkerboard pattern.

2.4.2 Synthetic Tests

Parameterizations that describe continuous velocity fields often opt for resolution tests that attempt to reconstruct a synthetic model using the same source-receiver geometry as the real experiment. The rationale behind this approach is that if a known structure with similar length scales to the solution model can be recovered using the same (for linearized solutions) or similar (for iterative non-linear solutions) ray paths, then the solution model should be reliable. The quality criterion is the similarity between the recovered model and the synthetic model. The so-called "checkerboard test" (Fig. 28), in which the synthetic model is divided into alternating regions of high and low velocity with a length scale equal (or greater) to the smallest wavelength structure recovered in the solution model, is a common test model. The initial model used for the test is the same as that used for the real inversion. Regions in which the checkerboard pattern is recovered clearly are those regions in which structure in the solution model can be considered to be well resolved. Lévêque et al. (1993) demonstrate that such an approach is not necessarily as reliable as it might seem. It is possible for the small scale structure of the checkerboard test to be well retrieved while larger-scale structure

is poorly retrieved. Furthermore, if the solution takes into account the non-linearity of the inverse problem, then the ray-path coverage will have a dependence on the velocity distribution. Thus, while a checkerboard reconstruction can account for the non-linearity of the traveltime dependence on the checkerboard structure, it cannot account for the non-linearity of the traveltime dependence on the true structure. Thus, rather than iteratively invert for the checkerboard structure, it may be better to simply use the ray-path geometry from the model produced by the inversion of the real data in a linear inversion. Alternatively, one could test a number of different synthetic models to investigate the effect of different structures (and hence ray geometries) on the resolution; this would be advisable especially if the solution model is complex.

Teleseismic traveltime tomography studies that have used checkerboard resolution tests include those by Glahn & Granet (1993), Achauer (1994) and Seber et al. (1996). Ritsema et al. (1998) used several different synthetic models rather than checkerboards and analyzed the accuracy of their recovery. This kind of analysis is not commonly used in wide-angle studies, although recently, Zelt (1998), Zelt et al. (1999, 2001) and Day et al. (2001) implemented checkerboard tests for analyzing the resolution of velocity structure (Zelt et al., 1999, also use it for interface structure) derived from wide-angle traveltime inversion. In local earthquake tomography, checkerboard tests have been implemented by Chiarabba et al. (1997) and Graeber & Asch (1999); Walck & Clayton (1987) and Walck (1988) used synthetic reconstructions with anomalies positioned in key localities (i.e. regions of high geological interest).

3 Applications to Observed Data

In this section, we review some applications of traveltime inversion methods to real data. Our aim is to give an idea of how the methods described above can be put together to solve a 2-D or 3-D inverse problem in practice. With this is mind, we tend to focus our attention on several selected case studies, rather than just briefly describe numerous examples. The choice of which methods to use is usually influenced by the class of data (reflection, wide-angle, local earthquake, teleseismic) that is available (see Section 1.3). For example, normal incidence reflection data, as its name suggests, predominantly contains reflected phases, so interfaces must be included in the parameterization. Teleseismic data, on the other hand, do not contain reflected phases, so structure in this case is generally represented by a continuous velocity variation. Thus, the different data types often resolve different aspects of structure, a feature that will also be discussed in some detail below.

3.1 Reflection Tomography

A number of 2-D schemes for the tomographic interpretation of reflection data have been presented over the years but not many of them have been extensively applied to real data (at least, there are not many examples in the literature). Bishop et al. (1985) presented one of the first methods for the simultaneous determination of velocity and depth in laterally varying media. In their method, the subsurface is represented by sub-horizontal layers separated by interfaces with a cubic spline parameterization. Within each layer, velocity is permitted to vary laterally and vertically by means of a grid of constant velocity gradient boxes. Within each box, a ray path segment will have the geometry of a circular arc, and therefore has an analytic expression. A shooting method of ray tracing determines source-receiver ray paths and traveltimes, and a Gauss-Newton method is used to iteratively minimize an objective function consisting of a data residual term and a model length

term (i.e. damping but no smoothing). The system of linear equations is solved using a Gauss-Seidel algorithm with successive over-relaxation. The method was applied to Common Depth Point (CDP) data from a pair of intersecting profiles in a region with permafrost. Significant shallow velocity variations due to the presence of the permafrost were imaged and the depth estimates to the interfaces at the intersection of the profiles agreed significantly better that those determined by conventional seismic processing. The maximum depth to which structure was imaged was ∼5.5 km.

It has been recognized by a number of authors that reflection traveltimes poorly resolve vertical variations of velocity within a layer (Farra & Madariaga, 1988; Kosloff et al., 1996). In the method of Farra & Madariaga (1988), a layered parameterization is adopted in which interfaces and lateral velocity variations within layers are described by cubic B-splines. Velocity within a layer is vertically invariant. Source-receiver ray paths and traveltimes are found with a shooting method of ray tracing that uses elements of paraxial ray theory to iteratively correct the ray take-off angle. The objective function they minimize consists of a data residual term and a penalty function which restricts the model behavior in accordance with the available *a priori* model information. A DLS approach is applied to iteratively minimize the objective function, and SVD is used to solve the system of linear equations at each step. They also make use of a layer-stripping approach. In this scheme, reflection traveltimes from the top interface are initially used to constrain the top layer only. Once these data are satisfied, traveltimes for the second interface are also introduced and together they are used to constrain the first two layers. The scheme continues in this manner through each successive layer. Farra & Madariaga (1988) applied their scheme to a synthetic model that included a layer pinchout, and showed that it can be reconstructed using a model with 1-D structure as a starting model. Coupled oscillations of interfaces and velocity variations were found to occur in regions where the data were unable to resolve the trade-off between interface depth variation and lateral velocity variation within a layer. They then applied the scheme to data from the Paris Basin. The maximum offset between source and receiver was 1.68 km, and structure was imaged to a depth of 2 km. A total of 99 parameters (57 velocity, 42 interface) were used to describe the model. The initial model was described by four flat constant velocity layers. In the solution model, the interfaces remained nearly horizontal, but the layer velocities had significant lateral variations. The recovered interface geometries are consistent with the known structure of the region.

Imaging of 3-D structure by reflection tomography is not very common. Chiu et al. (1986) developed and applied a scheme to vibroseis data collected on Vancouver Island. Their method assumes that subsurface structure can be adequately represented by constant velocity layers that separate interfaces described by n^{th}-order polynomial surfaces. A ray bending scheme is used to solve the forward problem, and an iterative damped least squares approach is used to solve the inverse problem. As in the above methods the linear system of equations are solved with SVD. However, in their application to crooked line vibroseis data collected as part of PROJECT LITHOPROBE, only 110 traveltime picks were available to constrain the structure of a two-interface model, consisting of a decollement zone and an under-thrusting oceanic crust. They found that planar interfaces were adequate to satisfy the data, making this a rather limited example of 3-D tomography.

In general, coincident reflection traveltime data alone do not seem to be sufficient to resolve both interface depth variations and arbitrary velocity variations within a layer, despite the relatively dense ray coverage associated with most reflection experiments. Williamson (1990) confirms this limitation by using a multi-stage inversion scheme in which progressively shorter length scales are permitted in both velocity and interface geometry as the iterative process proceeds. Despite satisfying the data and finding that longer wavelength components may be adequately recovered, Williamson

(1990) also found that shorter wavelength velocity-depth trade-offs cannot be resolved without further information. This may take the form of *a priori* information to help constrain the problem, and/or simplifying assumptions made about the structure (e.g. constant velocity layers, layers in which velocity has no vertical variation). While this may be sufficient at shallow depths, a basic drawback of reflection tomography is that the resolving power of the data decreases with depth due to the geometry of the experiment. Thus, it is usually limited to near-surface applications.

3.2 Wide-Angle Tomography

Unlike coincident reflection tomography, inversion of wide-angle traveltimes for 2-D or 3-D crustal structure has been applied to many datasets from around the world. For example, numerous studies have been carried out in Canada (e,g, Hole et al., 1992; Kanasewich et al., 1994; Clowes et al., 1995; Zelt & White, 1995; Morozov et al., 1998; Zelt et al., 2001), the U.S. (e.g. Lutter & Nowack, 1990; Jarchow et al., 1994; Zhu & Ebel, 1994; Parsons et al., 1996; Wang & Braile, 1996; Lizarralde & Holbrook, 1997), Europe (e.g. Riahi & Lund, 1994; Staples et al., 1997; Darbyshire et al., 1998; Louden & Fan, 1998; Mjelde et al., 1998; Korenaga et al., 2000; Morgan et al., 2000) and in oceanic settings (e.g. Hildebrand et al., 1989; Wiggins et al., 1996; Kodaira et al., 1998; Recq et al., 1998; Day et al., 2001; Grevemeyer et al., 2001). Experiments involving marine air-gun sources and land based receivers or OBSs (Ocean Bottom Seismometers) are quite common since they can be coupled with coincident reflection studies. As discussed in Section 1.3, wide-angle data contain both refraction and wide-angle reflection phases. The traveltimes of rays that refract tend to constrain velocity variations better than wide-angle reflections, which in turn are better at constraining interface depth. Thus, with adequate data coverage, the simultaneous inversion of refraction and reflection traveltimes can result in a well constrained solution model that includes both variations in interface depth and layer velocity.

Wide-angle experiments are usually performed in 2-D (i.e. recorded by an in-line array of sources and receivers), although recent studies (Zelt & Barton, 1998; Zelt et al., 1999; Day et al., 2001; Rawlinson et al., 2001b; Zelt et al., 2001) indicate that 3-D wide-angle surveys are becoming more frequent. In 2-D experiments, data coverage is often quite dense, and tomographic-style interpretation techniques are usually designed to allow both interface structure and layer velocities to be constrained by the data. One of the most frequently used methods for the tomographic style inversion of 2-D wide-angle traveltimes was developed by Zelt & Smith (1992). In their method, the model is parameterized in terms of a layered network of irregular blocks (see Section 2.1), which allows velocity to vary both laterally and vertically within a layer. Layer pinchouts and isolated bodies can also be included by reducing layer thickness to zero. The advantage of this approach to structural representation is that velocity and interface node distribution can be adapted to suit the resolving power of the dataset. Refractions, reflections and head waves are traced through the model by numerically solving the initial value problem formulated in terms of a pair of first order ordinary differential equations (Eq. 19 and Eq. 20). Linear interpolation between rays that bracket a given receiver is used to estimate the corresponding traveltime and Fréchet derivatives. The inverse problem is solved using an iterative DLS method (damping but no smoothing), with rays re-traced after each iteration. LU decomposition is used to solve the system of linear equations.

A number of authors have adopted the approach of Zelt & Smith (1992) to invert wide-angle data, including Kanasewich et al. (1994), Riahi & Juhlin (1994), Staples et al. (1997), Ye et al. (1997), Darbyshire et al. (1998), Kodaira et al. (1998), Morozov et al. (1998) and Navin et al. (1998). Darbyshire et al. (1998) applied it to data from the ICEMELT refraction line to image crustal

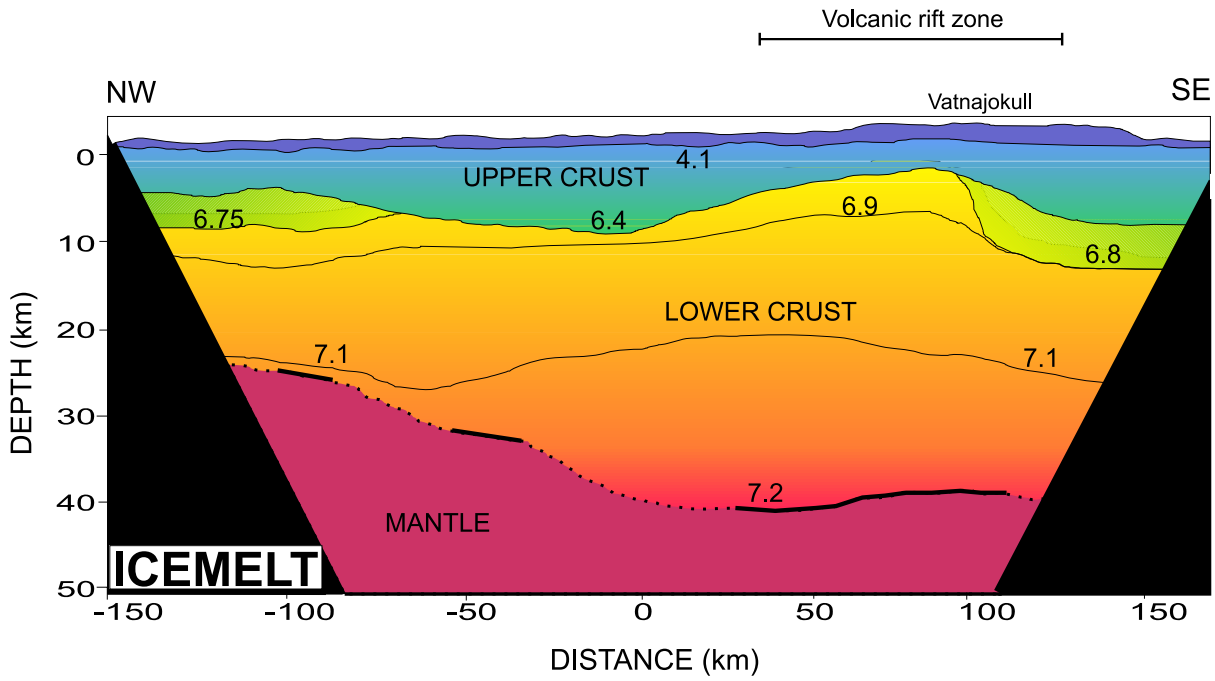

Figure 29: Crustal model obtained from the ICEMELT refraction line in Iceland using the wide-angle traveltime inversion method of Zelt & Smith (1992). Solid lines on the Moho represent those parts of the interface constrained by the data. [From Darbyshire et al. (1998). Copyright 1998 Royal Astronomical Society. Reproduced by permission of Blackwell Science Ltd.]

structure above the Iceland mantle plume. In their study, up to 60 land-based recorders were used to record seismic energy from six explosive shots along a 310 km line traversing Iceland from north to south. A total of 181 traveltime picks were used to constrain the crustal model, which included both interfaces and velocity variations within a layer. The inversion solution produced a normalized χ^2 misfit value of 1.31. Fig. 29 shows the ICEMELT crustal model produced by the inversion of refraction and wide-angle reflection traveltimes. The upper crust is characterized by high vertical velocity gradients (>0.2 s^{-1}) and considerable lateral heterogeneity. By contrast, the lower crust is less complex and features much smaller vertical velocity gradients (<0.03 s^{-1}). The crust thickens to nearly 40 km depth above the mantle plume, a feature interpreted to be caused by increased melt generation due to elevated mantle temperatures and active convection in the plume core. Two of the methods used by Darbyshire et al. (1998) to analyze the robustness of their solution model are ray coverage and estimates of resolution from linear theory. Fig. 30a shows all the ray paths used to constrain the ICEMELT model, and Fig. 30b shows a plot of the diagonal elements of the resolution matrix; values ≥ 0.5 are considered well resolved. Several regions in the upper and lower crust are poorly resolved while the mid crust is generally well resolved.

Darbyshire et al. (1998) did not use the inversion routine of Zelt & Smith (1992) in a completely automated fashion. Manual intervention (i.e. the user empirically tunes one or more parameter values instead of the inversion routine) was used to refine some of the upper crustal structure. In 2-D wide-angle modeling, it is sometimes necessary to intervene during the inversion process to prevent unrealistic model perturbations caused, for example, by insufficient data constraints or extreme non-linearity. In some studies (e.g. Iwasaki et al., 1990; Gürbüz & Evans, 1991; Grad et al., 1997; Iwasaki et al., 1998; Recq et al., 1998), the inversion is performed completely manually in a process called "forward modeling" (i.e. one tries to satisfy the data by adjusting model parameter values using trial

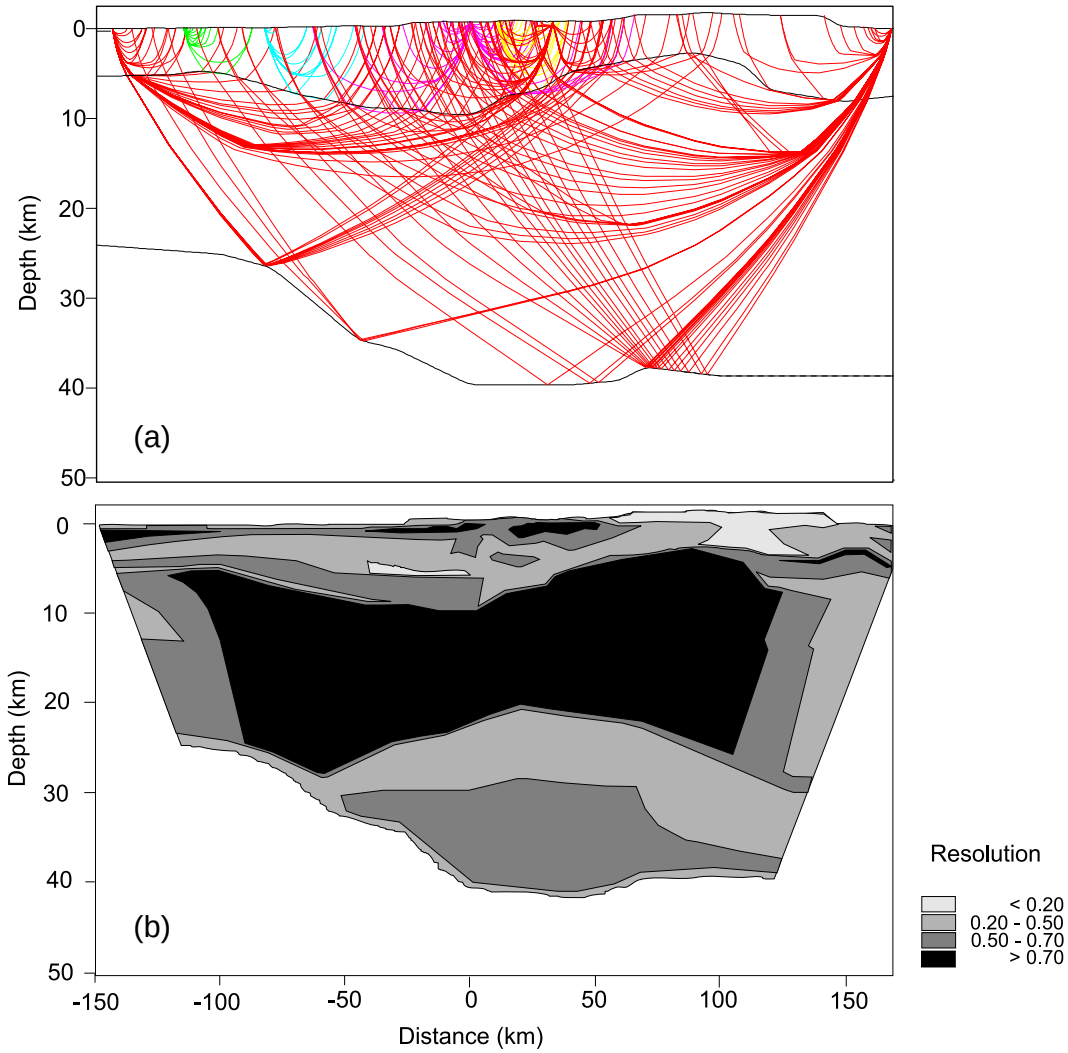

Figure 30: (a) All ray paths used to constrain the model in Fig. 29 and (b) diagonal elements of the resolution matrix obtained from linear theory. Regions that are considered well constrained have a value $\geq$0.5. [From Darbyshire et al. (1998). Copyright 1998 Royal Astronomical Society. Reproduced by permission of Blackwell Science Ltd.]

and error). The advantage of this approach is that the user has complete control over the state of the model, and that hypothetical structures may be tested against the data. On the other hand, if there are a large number of model parameters, repeated forward modeling becomes very time consuming.

Wide-angle seismic studies that attempt to constrain structure in 3-D usually do not consider both interface structure and layer velocity variations. For example, Hole (1992), Hammer et al. (1994), Toomey et al. (1994), Zelt & Barton (1998), Day et al. (2001) and Zelt et al. (2001) inverted first arrivals for continuously varying velocity, and Hole et al. (1992) and Riahi & Juhlin (1994) inverted refraction and reflection traveltimes respectively for interface structure. Studies that have attempted to invert for both parameter types include Zelt et al. (1996), who use a layer stripping approach, and Zelt (1999), who use floating reflectors coupled with sharp velocity gradients to simulate discontinuities. Most of the above 3-D wide-angle studies use methods to solve the forward problem of traveltime and ray path determination based on the finite difference approach of Vidale (1990).

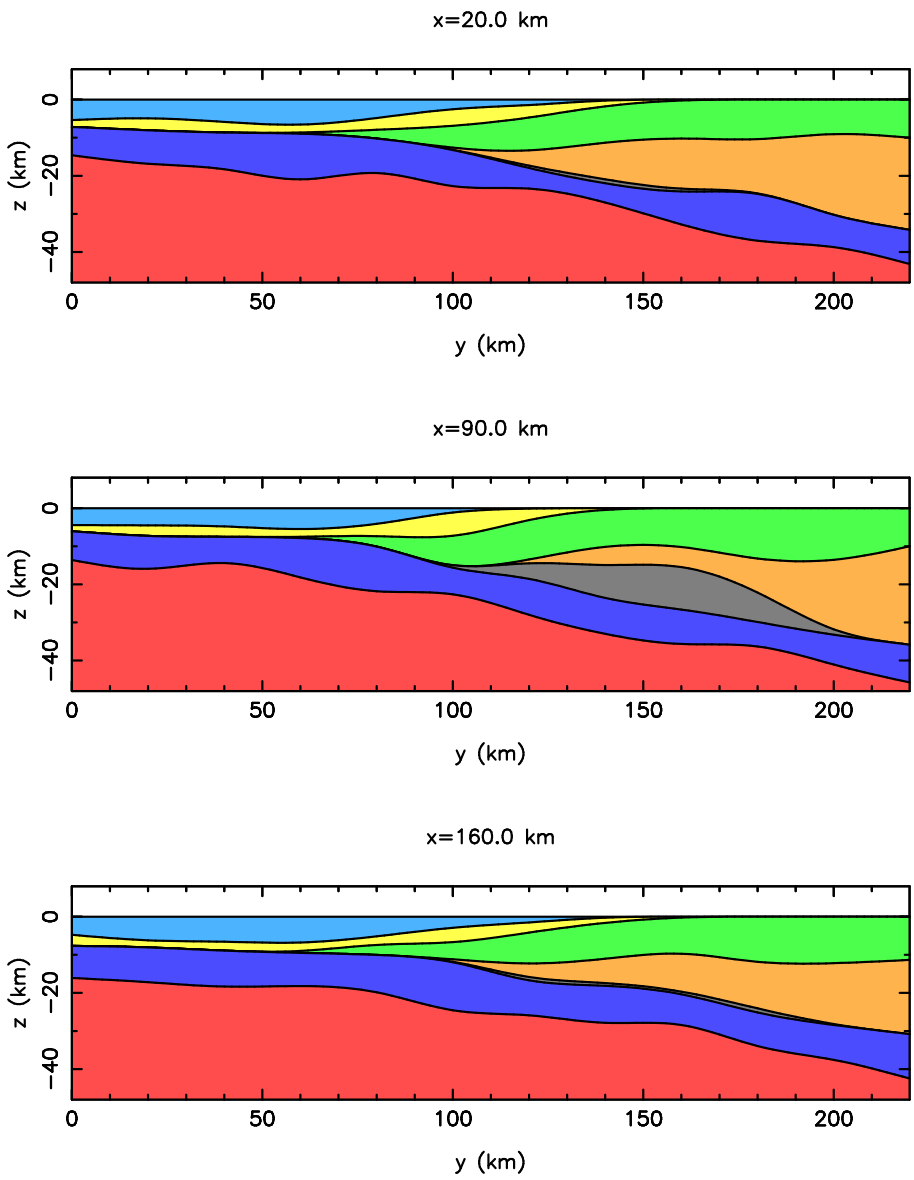

Figure 31: Three cross sections taken approximately perpendicular to the strike of the subducting slab of a synthetic subduction zone. Rawlinson (2000) showed that it was possible to reconstruct the underplated low velocity lens (grey) from an initial model with only 2-D structure using wide-angle refraction and reflection traveltimes. Land based receivers and marine sources were used in the synthetic test, which demonstrated the flexibility of the method developed by Rawlinson et al. (2001a). From top to bottom, the ocean layer is shown in light blue, the sediment layer in yellow, the upper crustal layer in green, the lower crustal layer in amber, the subducting ocean crust in dark blue, and the lithospheric mantle in red.

In fact, it is probably true to say that this is the only branch of seismic tomography in which this method predominates.

A method for the inversion of 3-D wide-angle seismic traveltimes that uses conventional ray tracing, rather than grid-based eikonal methods, was presented by Rawlinson et al. (2001a). In this scheme, crustal structure is represented by layers separated by smoothly varying interfaces with a cubic B-spline parameterization. Within each layer, the velocity varies linearly with depth. This restriction on the layer velocity field means that rays can be traced analytically within a layer, thus allowing rapid solution of the initial value problem. The two point problem is solved using a shooting

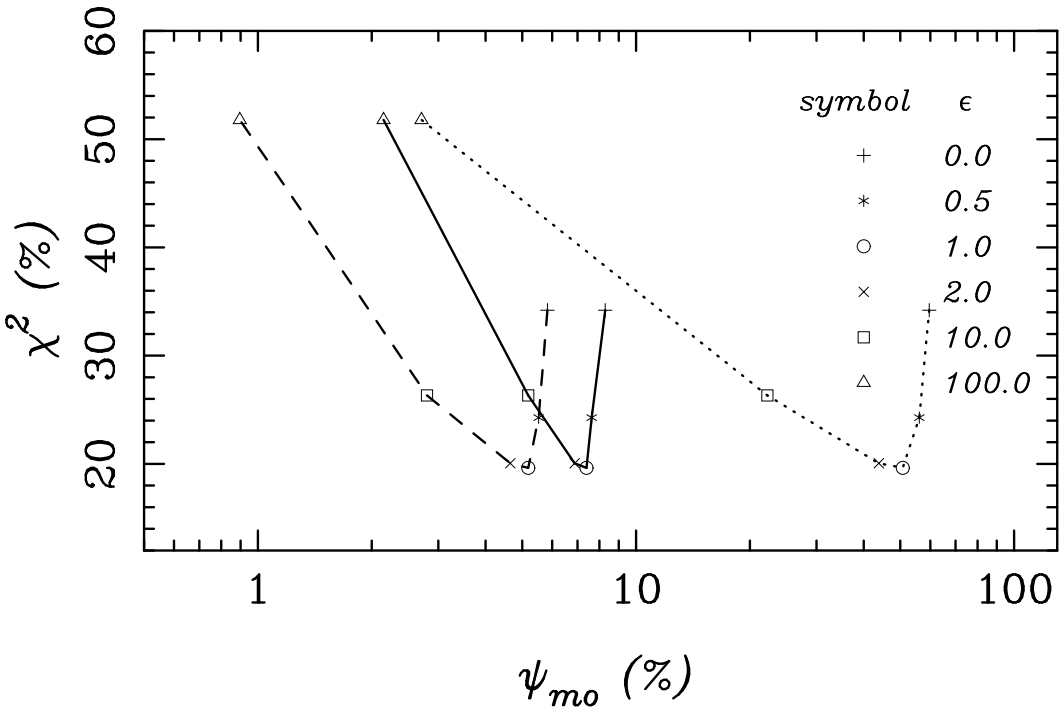

Figure 32: Traveltime inversion trade-off curves for various values of the damping factor ϵ after four iterations of the inversion performed by Rawlinson et al. (2001b) for the crustal structure of Tasmania. The ordinate axis plots the percentage improvement in χ^2 from the starting model, while the abscissa plots the percentage perturbation of the solution model from the initial model; this is plotted for each parameter type: interface node depth (solid line), layer velocity (dashed line) and layer velocity gradient (dotted line).

method which uses a Newton-type method to accurately locate the required two-point paths. A subspace inversion method is used to iteratively solve the inverse problem of minimizing an objective function consisting of a data residual term and a damping term. A smoothing term was not included because the smooth nature of the cubic B-spline parameterization facilitated implicit smoothing by varying the node density as required. An obvious drawback of the method is that the assumption of a crustal layer with laterally invariant velocity may not be applicable in some circumstances. However, if such a model is able to satisfy a given dataset, then it may be difficult to argue for a more complex model. Novel features of the method are that layers may pinch out and the horizontal distribution of the nodes may be varied to suite the resolving power of the dataset. Quite complex structure may therefore be represented. An example is provided by Rawlinson (2000) where a relatively complex synthetic subduction zone (see Fig. 31), which includes an underplated low velocity lens, was successfully reconstructed from a starting model with only 2-D structure.

Rawlinson et al. (2001b) applied the method of Rawlinson et al. (2001a) to wide-angle data collected as part of the TASGO project. In this study, an array of 44 land-based recorders was distributed throughout the island state of Tasmania, SE Australia, to record seismic energy from an encircling array of marine normal incidence reflection shot lines. The size of the dataset was reduced from its maximum potential by a number of recorder failures, low signal to noise ratios at several sites and the limited range of the airguns (usually <200 km). A total of 2590 P_n and P_mP traveltimes from 13 shot lines to 21 receivers were used to constrain a two layer model of Tasmania. This consisted of the crust and the upper mantle separated by a Moho of variable depth. Lateral variations of velocity within the crust were not included because P_n and P_mP arrivals poorly

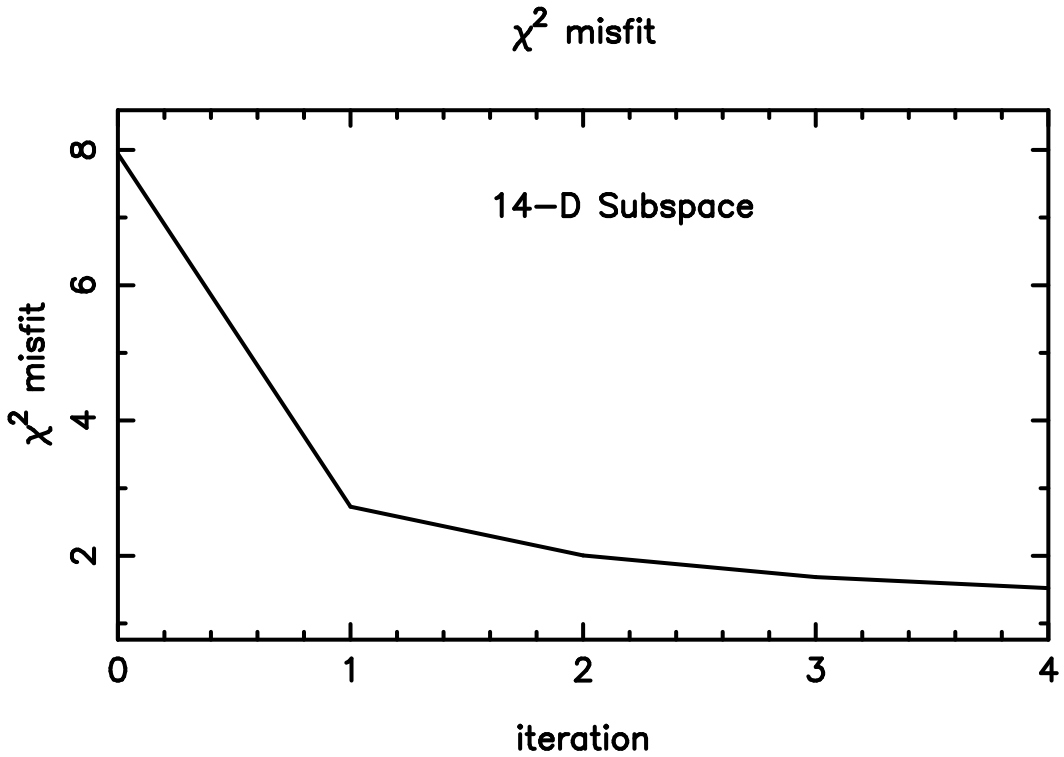

Figure 33: χ^2 misfit versus iteration number for the inversion (using $\epsilon = 1.0$) of P_n and $P_m P$ traveltimes done by Rawlinson et al. (2001b) for the Moho structure of Tasmania.

constrain such features, and the coverage of crustal arrivals was rather poor.

The Moho is represented by a rectangular grid of 600 nodes spaced 20 km apart. The total number of unknowns in the inverse problem is therefore 604 - only two parameters are required to describe the linear velocity gradients in each of the crust and mantle. A simple 1-D model with the Moho at 30 km depth was chosen as a starting model. The damping factor ϵ, that weights the relative importance of the data and the *a prior* model information, was selected by inspection of a trade-off curve (Fig. 32). Since there are three parameter classes in the inverse problem (interface node depth, layer velocity and layer velocity gradient), a trade-off curve is shown for each. From these diagrams, a value of $\epsilon = 1$ was taken. Note that $\epsilon = 2$ generates a near-identical model. A 14-D subspace inversion scheme was applied to minimize the objective function; after four iterations, the RMS traveltime fit was reduced to 176 ms from an initial value of 371 ms. The corresponding normalized χ^2 misfit versus iteration is shown in Fig. 33.

The Moho structure of the solution model is shown in Fig. 34 along with the diagonal elements of the resolution matrix. Outside the horizontal bounds of the source-receiver array, much of the Moho has not been perturbed from its initial value of 30 km due to a lack of ray path coverage. This is supported by the resolution map, which also indicates poor data coverage in central and NE Tasmania (Fig. 34b). A striking feature of the solution model is a major change in crustal thickness beneath the Arthur Lineament metamorphic belt in NW Tasmania (Fig. 34a). This has been interpreted to represent the NW limit of deformation in Tasmania during the Mid-Late Cambrian Tyennan Orogeny.

3-D wide-angle tomographic problems often involve large datasets and a large number of unknowns. For example, in their inversion for continuously varying velocity structure, Zelt & Barton (1998) use 53,479 first-arrival picks and separately invert for 315,000 unknowns using a backpro-

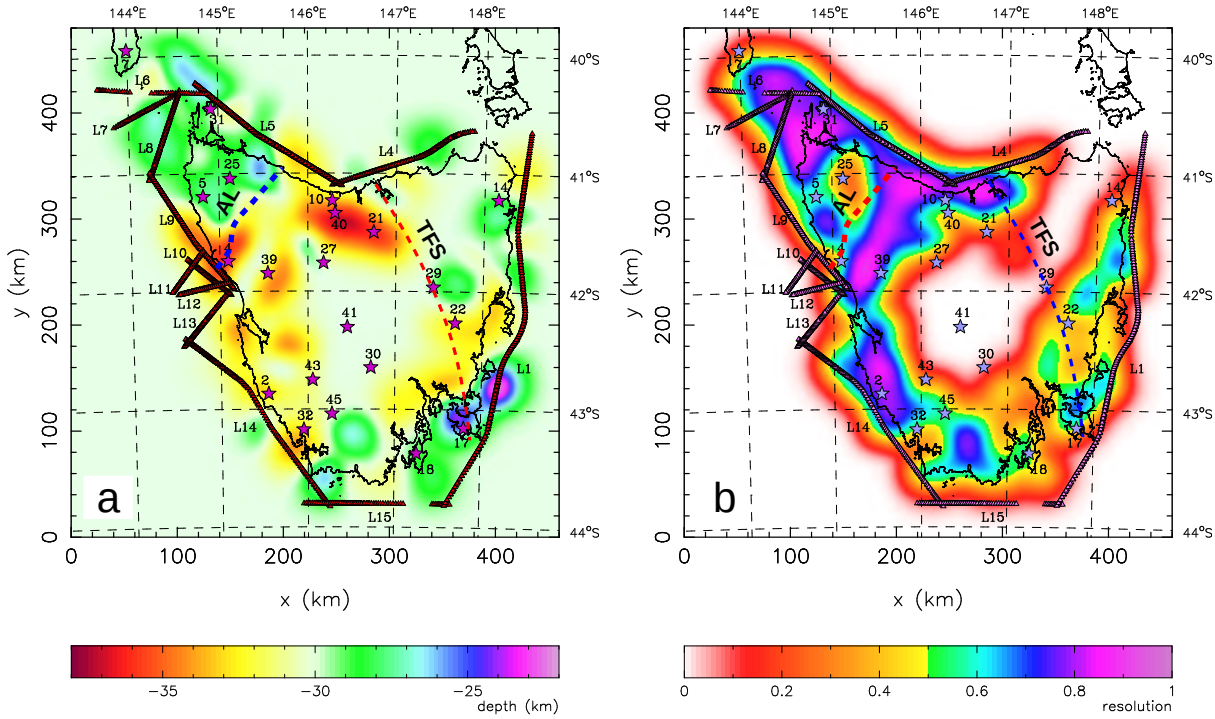

Figure 34: (a) Moho depth and (b) diagonal elements of the resolution matrix of the solution model for Tasmania produced from the TASGO wide-angle dataset. In both diagrams, stars indicate recorders and small triangles indicate shot points from which data were picked. Contiguous triangles form shot lines except where data gaps occur (e.g. lines 4 and 15). Thick dashed lines denote tectonic element boundaries: Arthur Lineament (AL) and Tamar Fracture System (TFS). [From Rawlinson et al. (2001b). Copyright 2001 American Geophysical Union. Reproduced by permission of American Geophysical Union]

jection method and 35,000 unknowns using a regularized least squares method. Studies of a similar magnitude were recently carried out by Day et al. (2001) and Zelt et al. (2001).

In summary, wide-angle tomography has proven to be a useful tool for revealing information about crustal structure. Its strength is its ability to resolve the trade-off between interface depth and lateral velocity variation within a layer given sufficient data coverage. This is in contrast to coincident reflection tomography. The main weaknesses of wide-angle tomography are the difficulty of picking reflection phases, which are always later arrivals, and the presence of extreme lateral variations in crustal structure, which can cause significant ray-path perturbations, and hence violate the local linearity assumption upon which iterative inversion methods are based.

3.3 Local Earthquake Tomography

Local earthquake tomography (LET) is a common tool for imaging subsurface structure in seismically active areas. Studies have been carried out in many different regions including California (Aki & Lee, 1976; Thurber, 1983; Eberhart-Phillips, 1986; Walck & Clayton, 1987; Walck, 1988; Zhao & Kanamori, 1992; Eberhart-Phillips & Michael, 1993; Scott et al., 1994), Taiwan (Kao & Rau, 1999), Japan (Nakanishi, 1985; Zhao et al., 1992), South America (Bosch, 1997; Graeber & Asch, 1999) and the Mediterranean (Chiarabba et al., 1997; Luca et al., 1997; Papazachos & Nolet, 1997; Haslinger et al., 1999). Apart from the source-receiver geometry, one of the distinguishing features of LET compared to the other types of tomography discussed here, is the need to relocate earthquake

hypocenters (spatial location and origin time) as part of the image reconstruction. This is because accurate hypocenter location requires an accurate knowledge of the velocity structure in the region occupied by the earthquakes and the recorders whose data are used in the location process. Adequate data are therefore required to resolve the trade-off between variations in structure and variations in hypocenter locations.

One of the original LET studies was carried out by Aki & Lee (1976). In their method, 3-D crustal structure is represented by constant slowness blocks. The initial model is completely homogeneous, and the inversion is linear, so rays consist of straight line paths between sources and receivers. Slowness perturbations for each block and hypocenter corrections are simultaneously inverted for using a DLS approach (damping but no smoothing). The method was applied to data collected at 60 stations in Bear Valley, California, from 32 local earthquakes. The $x \times y \times z$ block size was set at $3 \times 4 \times 5$ km^3, with a total of 264 blocks used to span a volume of $24 \times 44 \times 15$ km^3 which included all receivers and sources. 1218 first-arrival times were inverted for all unknowns. Estimates of covariance and resolution from linear theory were used to analyze the robustness of the velocity solution and relocated hypocenters. A feature of the solution model is a narrow low velocity band oriented along the San Andreas fault zone in the upper 5 km.

The conceptual basis of LET has not really altered since the study of Aki & Lee (1976). However, some refinements have been made. For example, Thurber (1983) used approximate 3-D ray tracing (which allows iterative improvement of the solution model) and a continuously varying velocity field via linear interpolation of a uniform grid of velocity nodes. Eberhart-Phillips (1990) introduced a pseudo bending scheme to find more accurate two-point paths starting from rays calculated using the approximate ray tracing method of Thurber (1983). These refinements are common to most implementations of LET in recent times and are encapsulated in the commonly used computer package SIMULPS12 (Evans et al., 1994).

A relatively recent application of LET was carried out by Graeber & Asch (1999) to image the subduction zone structure beneath the southern central Andes. In this study, local earthquakes were recorded by a network of 31 recorders deployed in northern Chile between the Coastal Cordillera and the Western Cordillera. A total of 16,488 P-arrivals from 764 events were inverted for P-velocity structure and hypocenter location using the SIMULPS12 algorithm. The initial model required by the inversion scheme was based upon a model produced by a separate inversion of the data for 1-D structure and hypocenter location only. The optimum damping factor for the DLS solution was obtained by visual inspection of the data variance vs. solution variance trade-off curve approximated by performing several single-iteration inversions with a range of damping values. Station delay terms were also included in the inversion as unknowns to help offset the effects of near surface structure. The solution model was obtained after 12 iterations of the DLS method, which reduced the data variance by 63% with respect to the optimum 1-D model. Fig. 35 shows a series of vertical slices through the resulting P-wave velocity model, which is described by a total of 2496 nodes. Relocated hypocenters are also shown in this figure. Only those parts of the velocity model that are relatively well resolved according to the diagonal elements of the resolution matrix are shown in Fig. 35. The solution model clearly suggests the presence of a subduction zone involving the easterly subduction of oceanic crust beneath continental crust, with the geometry of the subducted plate also indicated by the relocated hypocenters.

Apart from calculating the resolution matrix to investigate solution robustness, Graeber & Asch (1999) also performed several synthetic resolution tests, including a checkerboard test. The results of this test at two cross-sections are shown in Fig. 36. As expected, the reconstruction below the

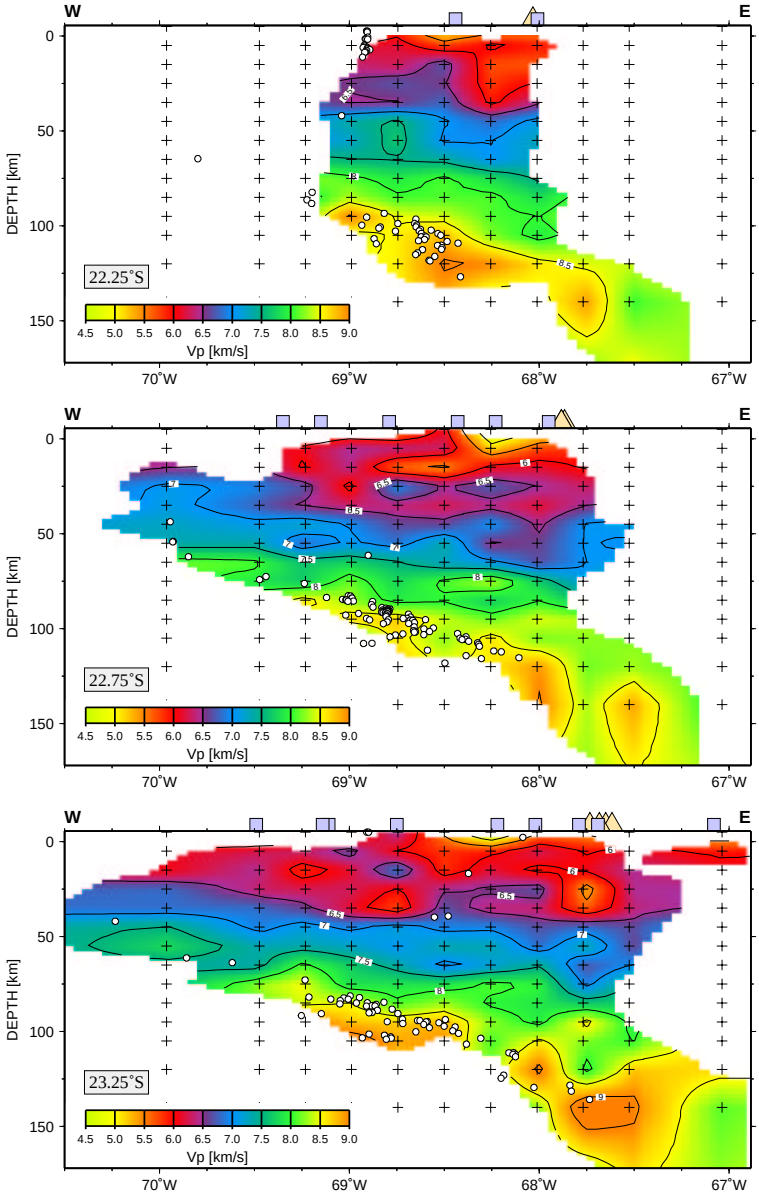

Figure 35: Cross-sections through a 3-D velocity model of the southern central Andes produced by inversion of local earthquake arrival times. Absolute velocities shown are a sum of initial velocities and calculated perturbations from an iterative non-linear inversion technique. Stations, recent volcanos and hypocenters adjacent to the sections are denoted by squares, triangles and open circles respectively. Velocity grid nodes are denoted by crosses. [From Graeber & Asch (1999). Copyright 1999 American Geophysical Union. Reproduced by permission of American Geophysical Union]

hypocenters is generally poor, but in the region between the hypocenters and receivers the reconstruction is generally good.

In addition to using first-arrival P-wave data, it is also quite common in LET to utilize S-wave data. A particularly useful property to map is the P-to-S velocity ratio (V_p/V_s). For example, regions with high levels of hydration are likely to have elevated V_p/V_s since the shear modulus will be reduced compared to the bulk modulus. One way of determining V_p/V_s is to separately invert for P and S velocity structure and then compare the results (e.g. Eberhart-Phillips, 1990). One problem with this is that the S-wave dataset is often smaller and of poorer quality than the P-wave dataset,

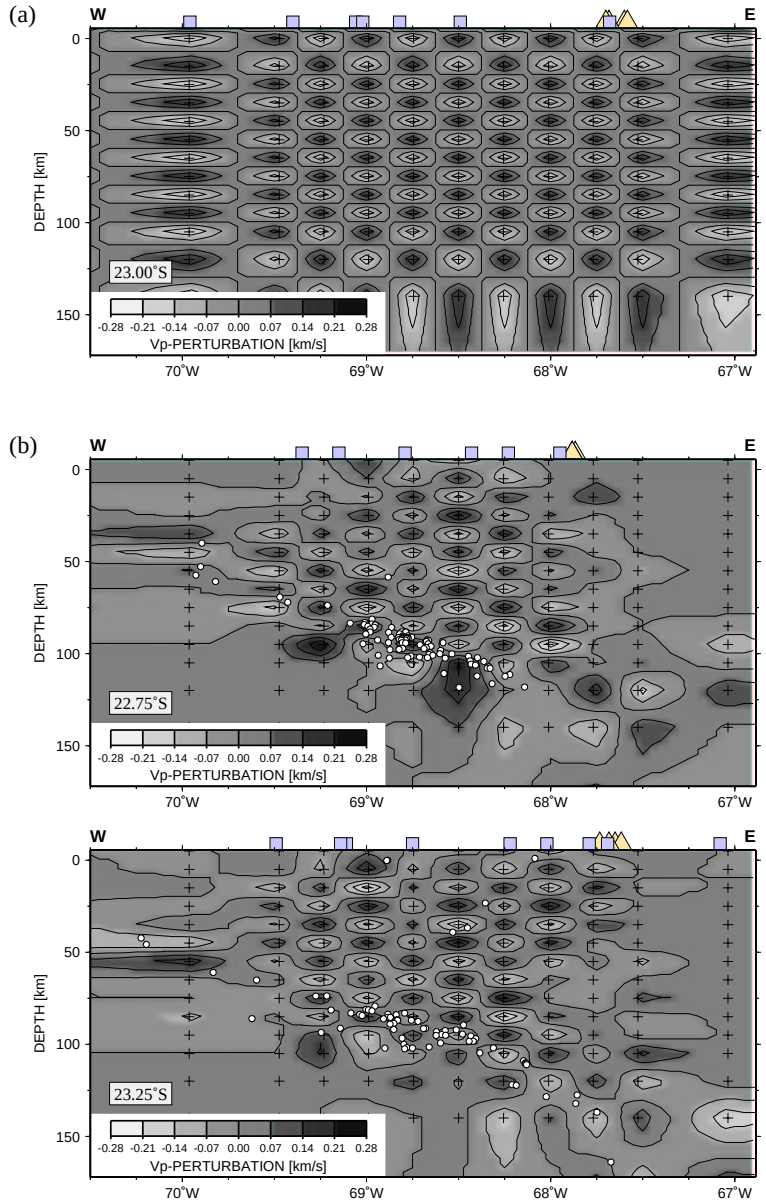

Figure 36: Checkerboard resolution test for the model shown in Fig. 35. (a) Synthetic checkerboard model, and (b) recovered checkerboard at two different latitudes. Symbols are defined as in Fig. 35. [From Graeber & Asch (1999). Copyright 1999 American Geophysical Union. Reproduced by permission of American Geophysical Union]

making detailed comparison of the two models difficult. An alternative is to directly invert the difference in P and S arrival times $T_s - T_p$ for V_p/V_s. Walck (1988) used this approach to construct a model of the 3-D variation in V_p/V_s in the Coso Region, California using backprojection. Graeber & Asch (1999) also construct an image of V_p/V_s by considering 8238 $T_s - T_p$ arrival time residuals. Their inversion for V_p/V_s was performed simultaneously with the inversion for V_p, hypocenter location and station delays. They found elevated V_p/V_s ratios in the zone above the subducting oceanic crust, suggesting the release of aqueous fluids from the oceanic crust into the overlying material.

Although restricted to regions of active seismicity, LET is a useful tool for mapping subsurface

structure at a variety of scales. Shallow earthquakes can be used to image the upper few kilometers of the crust (e.g. Walck, 1988; Chiarabba et al., 1997) while deep earthquakes from regions like subduction zones can be used to image lithospheric scale structure (e.g. Graeber & Asch, 1999). While relocation of earthquake hypocenters is a necessary part of the reconstruction of the velocity field from the data, their distribution provides valuable additional information on subsurface structure. Inclusion of S-wave data in the form of inversion of $T_s - T_p$ for V_p/V_s is not commonly considered in the other forms of tomography described in this paper, further distinguishing LET as a method worth considering provided local earthquake sources are available.

3.4 Teleseismic Tomography

Like wide-angle tomography and local earthquake tomography, teleseismic tomography has been used extensively to map the structure of the crust and lithosphere. Studies are often carried out on a variety of scales ranging from 10s of km (e.g. Rawlinson & Houseman, 1998; Steck et al., 1998) to 100s of km (e.g. Humphreys & Clayton, 1990; Dorbath & Paul, 1996). In the latter type of study, mantle beneath the lithosphere is also imaged. Usually, the horizontal extent of the receiver array and the source distribution determines the depth to which features may be resolved. Most teleseismic studies are carried out in 3-D (e.g. Aki et al., 1977; Oncescu et al., 1984; Humphreys & Clayton, 1990; Benz et al., 1992; Glahn & Granet, 1993; Achauer, 1994; Seber et al., 1996; Saltzer & Humphreys, 1997; Ritsema et al., 1998; Steck et al., 1998; Graeber et al., 2002), with relatively few 2-D studies (e.g. Evans, 1982; McQueen & Lambeck, 1996) probably because of the difficulty in lining up an array of recorders on roughly the same great circle as a set of teleseismic earthquakes with good angular coverage.

The data utilized by teleseismic tomography studies are the relative traveltime residuals of identifiable phases (usually the direct P-phase). These are obtained by calculating the difference between predicted and observed traveltimes and subtracting the mean for each source-receiver set. Predicted traveltimes are usually obtained by ray tracing through a 1-D reference model of the Earth, such as IASP91 (Kennett & Engdahl, 1991). Removing the mean from each set of traveltime residuals means that they are relatively insensitive to errors in hypocenter location. However, they also become insensitive to vertical variations in structure. One consequence of this is that the horizontally averaged velocity structure of the solution model will simply reflect the horizontally averaged velocity structure of the initial model. Interpretation of the resulting models must therefore be done with care (see Lévêque & Masson, 1999, for more details).

One of the first papers to describe a method of seismic tomography was that of Aki et al. (1977), who developed a scheme in the context of application to teleseismic data. In their method, the lithosphere is represented by layers composed of constant velocity blocks. Incoming wavefronts from teleseismic events are assumed to be locally planar beneath the model region and the initial model has constant velocity layers. Thus, the task of ray tracing from the incident wavefront to the receiver array is straightforward. Linear inversion is performed using either a stochastic inverse or a generalized inverse approach. The stochastic solution is essentially a DLS solution with $\epsilon = 1$, while the generalized inverse solution has $\epsilon = 0$, but identifies a minimum length solution using SVD. The method was applied to data collected by the Norwegian Seismic Array (Norsar) to image lithospheric structure in southeastern Norway. A total of 1496 traveltime residuals from 93 events recorded at 22 sites (each consisting of six seismometers) were inverted for the slowness perturbation of 405 constant velocity blocks contained in a five-layer model of the lithosphere (9×9 blocks per layer). The model region has a horizontal coverage of 180×180 km and extends to a depth of 126

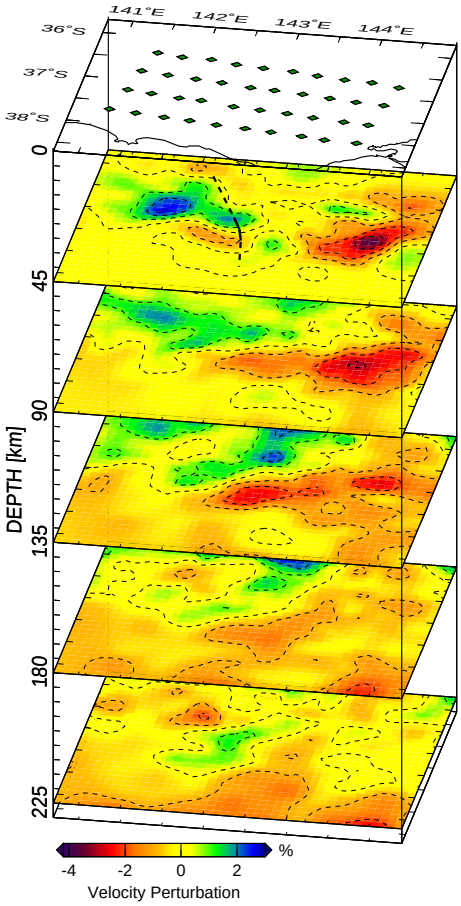

Figure 37: Horizontal sections taken through the 3-D solution model of western Victoria obtained by inversion of teleseismic traveltime residuals. Receivers are denoted by green diamonds and the dashed line on the 45 km depth slice indicates the approximate location of the Moyston Fault Zone. [From Graeber et al. (2002). Copyright 2002 Royal Astronomical Society. Reproduced by permission of Blackwell Science Ltd.]

km. Estimates of *a posteriori* model covariance and resolution obtained from linear theory were used to analyze the solution models. The encouraging results produced by this study undoubtedly catalyzed the numerous seismic tomography studies of the crust and lithosphere that soon followed.

A recent example of teleseismic tomography, which reflects the way in which the method has evolved since the paper of Aki et al. (1977), is that of Graeber et al. (2002). In this study, an array of 40 short period recorders were distributed throughout an area of about 270×150 km^2 in western Victoria, SE Australia. The tomographic scheme of VanDecar (1991) was used to invert 4067 teleseismic traveltime residuals for the velocity structure of the crust and upper mantle to a depth of 405 km. In this scheme, velocity structure is represented by a spherical grid of slowness nodes (27,898 of them in this case) interpolated by splines under tension. A shooting method of ray-tracing based on the method of Creager & Jordan (1984) finds source-receiver traveltimes and ray paths through the 3-D model. To solve the inverse problem, a gradient-based iterative non-linear scheme is employed. The objective function consists of a data misfit term and a number of regularization terms including damping, flattening (first-derivative smoothing) and smoothing (second-derivative smoothing). In addition, station and hypocenter corrections are included as unknowns. Station corrections help

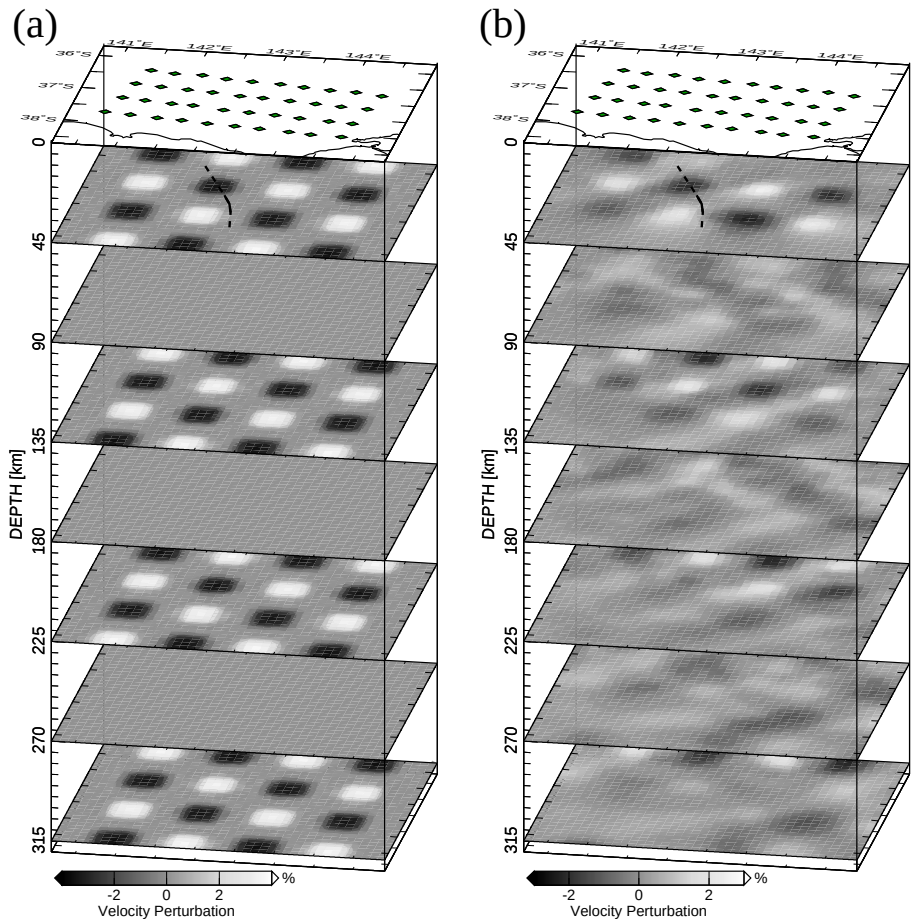

Figure 38: Synthetic checkerboard resolution test for the model shown in Fig. 37. (a) Original checkerboard model, and (b) reconstructed checkerboard. [From Graeber et al. (2002). Copyright 2002 Royal Astronomical Society. Reproduced by permission of Blackwell Science Ltd.]

reduce the mapping of near surface effects into deeper structure. Hypocenter adjustments are usually kept small because they are poorly constrained. The minimization of the objective function is achieved by assuming local linearization which allows the objective function to be written as a matrix equation (cf. Eq. 63). This matrix equation is then solved using a conjugate gradient method. The linear inversion and 3-D ray-tracing are performed iteratively to account for the non-linear nature of the inverse problem.

The solution model produced by Graeber et al. (2002) using this scheme is shown in Fig. 37 as a series of horizontal slices. One of the prominent features of the model is a relatively slow anomaly in the east which decreases in amplitude with depth. Graeber et al. (2002) interpret this feature, which lies beneath the Newer Volcanic Province, as a thermal anomaly related to Pliocene and Pleistocene hotspot-fueled volcanism. Synthetic tests were used to examine the robustness of the inversion solution. Fig. 38 shows the results of a checkerboard resolution test. In this version, alternating fast and slow anomalies are separated by zones of zero anomaly to permit smearing effects to be detected more easily (Fig. 38a). The reconstruction (Fig. 38b) recovers the checkerboard pattern more successfully nearer the surface. The amplitudes of the recovered checkerboard anomalies are generally less than the synthetic model, which is a typical feature of this type of test. The degraded structure of the recovered checkerboard is due partly to the source coverage - in this study, few

earthquakes from the south and west were detected.

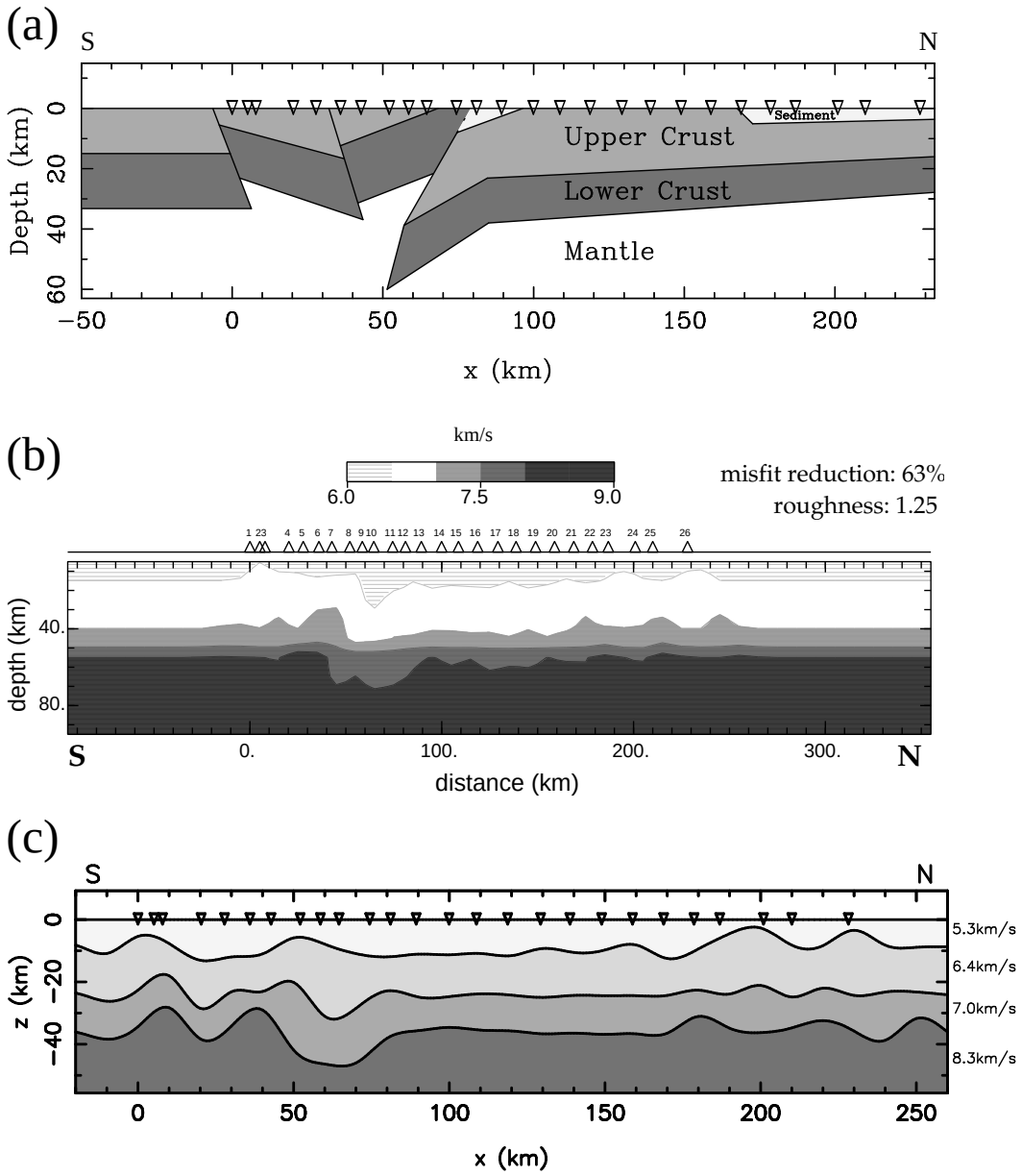

Figure 39: Three models derived from the Musgrave Line teleseismic dataset using different methods. (a) Crustal structure obtained by Lambeck & Burgess (1992) using forward modeling, (b) contours of velocity for the model obtained by inversion for slowness perturbation by McQueen & Lambeck (1996), and (c) crustal structure obtained by Rawlinson & Houseman (1998) using inversion for interface geometry. Note the differences in scale between (a), (b) and (c). [Modified from Rawlinson & Houseman (1998). Copyright 1998 Royal Astronomical Society. Reproduced by permission of Blackwell Science Ltd.]

In most teleseismic tomography studies, traveltime residuals are matched by adjusting velocity (or slowness) parameters that describe a continuous velocity field; interfaces are rarely included. There are several reasons for this. First, teleseismic data do not contain obvious signs of an interface, such as the “reflectors” seen in normal-incidence seismic sections, or the later arriving, large amplitude reflection curves evident in refraction studies. Interfaces may cause mode conversions

in teleseismic arrivals, but these are not easy to detect, particularly if the analysis relies on single component data. Second, methods that invert for both interface structure and continuous velocity variation must be able to address the trade-off between velocity and interface position. In teleseismic tomography, only transmitted rays are available, so the prospect of producing a well resolved combination of velocity and interface parameters is low. Finally the number and approximate depths of interfaces needs to be known *a priori* as there is little information in the data on this aspect. To some extent, this problem also occurs when a continuous velocity field is assumed, in that it is not possible to derive an initial model from the data itself. In such circumstances, the starting model may be obtained from other sources, such as wide-angle studies, or even global 1-D velocity models such as IASP91 (see Graeber et al., 2002). Surface wave inversion and teleseismic receiver function results may also be useful in locating approximate interface positions. Since the Earth is characterized by both continuous and discontinuous changes in structure, it is reasonable to investigate the other end member model type (i.e. consisting only of variable interfaces) when inverting teleseismic traveltime residuals for subsurface structure. This has been done by Davis (1991), Kohler (1997) and Rawlinson & Houseman (1998).

In the 2-D method of Rawlinson & Houseman (1998), structure is represented by sub-horizontal constant velocity layers. Interfaces are described by nodes interpolated by cubic splines. The depths of the nodes are the unknowns in the inversion. Rays are traced from a planar incident wavefront to the receiver array by means of a shooting method. A conjugate gradient method is used to minimize an objective function that consists only of a data residual term. The method is iterative, with ray-tracing applied after each application of the conjugate gradient method to account for the non-linear nature of the inverse problem. Synthetic tests showed that multiple interfaces may be reconstructed simultaneously in the presence of noise, and with relatively inaccurate initial models. For example, if the initial interfaces have an inaccurate average depth, their shape can still be recovered, but the average depth of the interface will not change. The method was applied to the Musgrave Line dataset collected in central Australia that traversed portions of the Musgrave Block and the Amadeus Basin. A three-interface crustal model was used to represent structure, and the data consisted of traveltimes from only five incoming wavefronts. The initial model was based on a 1-D refraction interpretation of wide-angle data collected nearby.

The Musgrave Line dataset was previously interpreted by Lambeck & Burgess (1992) and McQueen & Lambeck (1996) using different methods. Lambeck & Burgess (1992) used forward modeling to satisfy the data with structure represented in terms of constant velocity layers with piece-wise linear boundaries. McQueen & Lambeck (1996) used a SIRT backprojection method to invert the traveltime residuals for perturbations in slowness on a rectangular grid of constant slowness cells. Lateral structure was permitted to a depth of over 90 km. Both methods used a 2.5 D approach (i.e. models are defined in 3-D space but are restricted to 2-D structure), allowing additional out-of-plane sources to be considered.

Fig. 39 compares all three models for the Musgrave Line that have been derived using different assumptions about structure and different approaches to satisfy the data. In terms of defining laterally slow and fast regions, all three models have a reasonable level of agreement. It is interesting to note that the two end member inversion models both satisfy the data to approximately the same level (63% misfit reduction for Fig. 39b and 62% misfit reduction for Fig. 39c), suggesting that inverting for interface structure and inverting for velocity variation are equally valid as far as the data are concerned. One advantage a tomographic inversion approach has over forward modeling is the ability to test the robustness of the solution. Fig. 40 shows two synthetic tests used to assess the

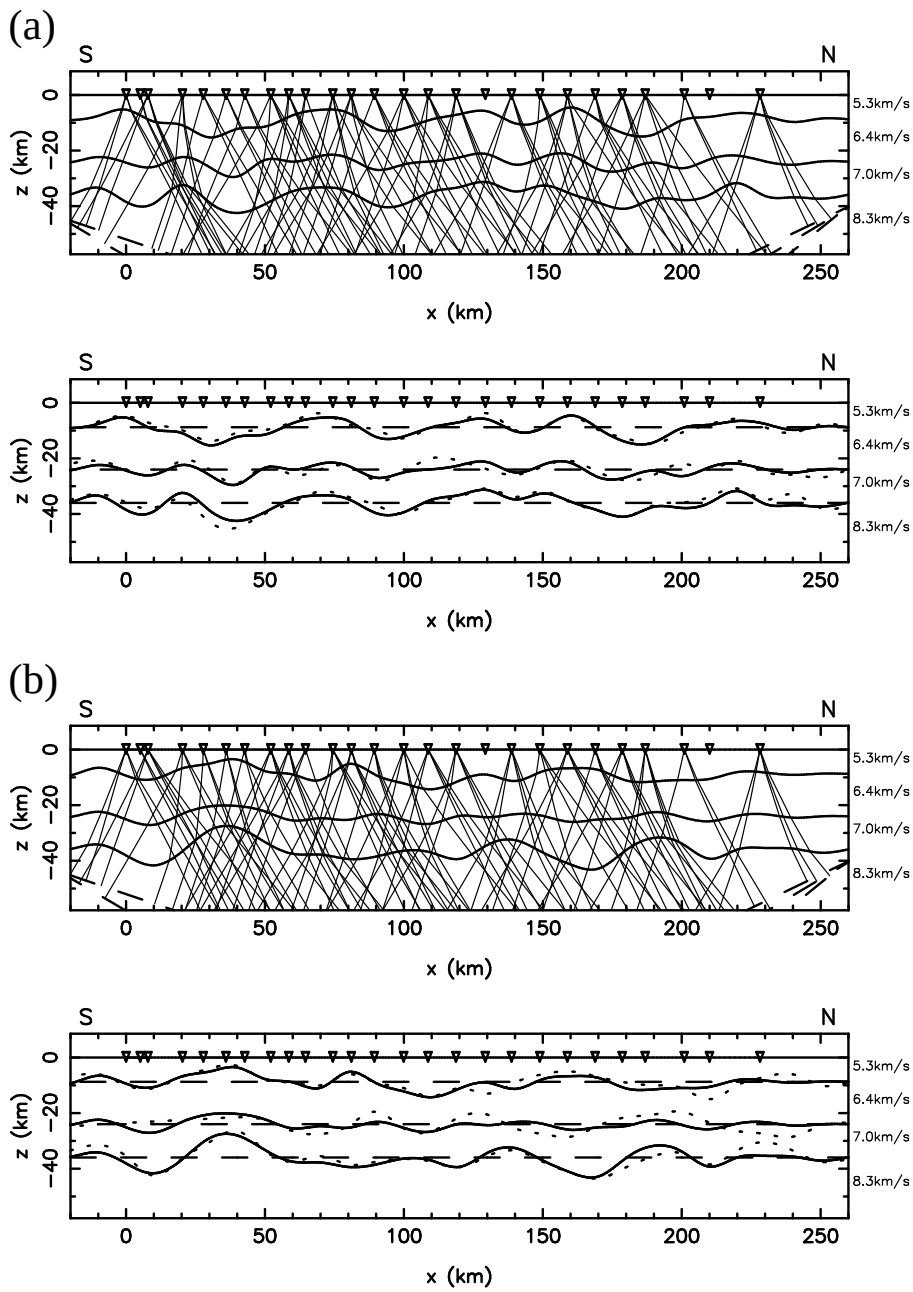

Figure 40: Two synthetic tests designed to analyze the robustness of the solution given in Fig. 39c. Ray diagrams show paths from the five wavefronts through the recovered structure. Dashed lines show the initial model, dotted lines the true model and solid lines the reconstruction. The synthetic models in (a) and (b) were chosen to be quite different in order to illustrate the effect of different ray geometries on the reconstruction.

quality of the solution shown in Fig. 39c. Both structures were chosen arbitrarily, and the tests use the same wavefront geometry and receiver array as the Musgrave Line experiment. The reconstruction of all three interfaces is generally quite accurate in both tests, suggesting that most of the significant structure contained in the solution model (Fig. 39c) is well constrained by the data.

Teleseismic tomography is a useful tool for investigating crust and upper mantle structure. Compared to local earthquake tomography, its main advantages are that it can be used to map structure

at greater depths, and that it is not restricted to regions of active seismicity. However, the fact that relative traveltime residuals must be used means that only relative perturbations in lateral structure can be resolved. In addition, the assumption that the model region is completely responsible for the teleseismic traveltime residuals will not always be valid. Also, the near-vertical incidence angle of ray-paths and the sometimes poor azimuthal source coverage may lead to significant smearing effects that in turn can lead to incorrect interpretations of results (see Keller et al., 2000). On the other hand, the interface structure shown in Fig. 39c demonstrates one advantage of the near-vertical transmission of rays: highly complex lateral structure may be resolved. Wide-angle tomography (for example) would have great difficulty imaging such structural complexity.

4 Future Developments

Traveltime tomography still remains the most popular approach for imaging subsurface structure on crustal and lithospheric scales. A wide array of schemes and algorithms can be found to tackle these problems, and there is little doubt that traveltime tomography will continue to play a role in the imaging of earth structure well into the future. Our aim here is to discuss, and to some degree speculate on, the possible directions in which seismic traveltime tomography methods may develop, and whether it will continue to be the predominant interpretation technique for large datasets.

A current limitation of many traveltime tomography techniques is that they only consider first-arrivals and ignore the potentially vast amount of information contained in later arriving phases. Exceptions include coincident reflection tomography and wide-angle tomography. In the case of reflection tomography, the number of phases picked usually corresponds to the number of interfaces contained in the model. Wide-angle tomography often includes later arriving reflection phases in a similar fashion. One of the principal difficulties in including traveltime information from later-arriving phases is that they are often obscured by the wavetrain of earlier arriving phases. This can make identification difficult or impossible. Another difficulty is that even if coherent later arrivals can be picked, they need to be related to model structure in order to be useful. For first-arrivals picked from a seismogram, this is not required because we just need to search for the transmission paths through the structures that correspond to first arrivals. Of course, this task is not necessarily straight forward, but the point is that we don't need to know anything about the ray path *a priori*. For wide-angle reflection phases, for example, we do need to know which interface they reflect from. If a layered parameterization is used, with the number of interfaces defined by the number of reflection phases identified, then the problem of association may not arise. If the interfaces are defined *a priori*, then it may not be easy to relate later arrivals with structure. One approach is to constrain the initial model using only first-arrivals plus any other phases that can be confidently identified, and then use the updated model to predict the traveltimes of other phases that weren't picked but may be evident in the data. A direct comparison with the observed data may allow additional phases to be picked with more confidence (Zelt, 1999). Multiples and phase conversions may also be identified in this manner.

As well as the problems of detection and association of later arrivals described above, yet another difficulty is that current methods of traveltime and ray path determination used in tomography are only capable of finding a small subset of all possible arrivals. Obvious examples are the grid-based methods (Section 2.2.2) and SPR (Section 2.2.3), both of which are limited to first-arrivals. Both methods can be made to find reflection arrivals but not later arrivals associated with caustics. Traditional ray tracing methods can find multiple arrivals - for example, by shooting an extremely dense

spread of rays, or by bending using a large variety of initial paths. However, in even moderately complex media, these approaches would be very computationally expensive. Recently, significant effort has been made in developing and implementing schemes for computing multivalued traveltime maps. One approach, often referred to as *wavefront construction*, involves the discrete propagation of a wavefront through a medium by using local ray tracing from each wavefront surface. Rays are traced for a given time step, with the end points of the rays describing the geometry of the new wavefront. New rays are introduced by interpolation if diverging rays cause parts of the wavefront to be poorly sampled. Multiple arrivals can be found using this scheme, including caustics. This basic idea for locating multiple arrivals has been developed by a number of authors (e.g. Vinje et al., 1993; Ettrich & Gajewski, 1996; Lucio et al., 1996; Lambaré et al., 1996; Vinje et al., 1996, 1999). Other schemes for finding multi-valued traveltime fields include Big Ray tracing (Benamou, 1996; Abgrall, 1999), which is a two-step procedure involving relatively coarse ray tracing followed by solution of the eikonal equation in spatial domains defined by adjacent ray paths, and Dynamic Surface Extension (Steinhoff et al., 2000), which shares similarities with wavefront construction.

The use of these schemes within a tomographic inversion may allow the traveltimes of more later arrivals to be utilized, thus resulting in a potentially better constrained and more detailed solution. With the advent of modern high fidelity broad band digital seismometers, much more information on the later arriving wavefield is being recorded, and new methodology such as wavefront construction is required to better exploit this type of data.

Another area in which traveltime tomography may see more development is in the combining of different datasets in a simultaneous inversion. Usually, tomographic studies constrain structure using just one type of data (reflection, wide-angle, local earthquake, teleseismic), but this need not be the case provided more than one overlapping dataset is available. For example, wide-angle or LET arrays are likely to detect teleseismic arrivals during their deployment, which could be exploited together with the principal dataset. Studies that combine different datasets have been carried out in the past. Thurber (1983) and Ankeny et al. (1986) included refraction traveltimes from several explosive sources in a LET study, Parsons & Zoback (1997) simultaneously inverted local earthquake, explosive source and airgun traveltimes for the crustal structure of the San Francisco Peninsula, and Sato et al. (1996) simultaneously inverted local earthquake and teleseismic traveltime data to image structure beneath the northeastern Japan arc. Methods that combine coincident reflection and wide-angle data in a simultaneous inversion have been developed by Wang & Braile (1996) and McCaughey & Singh (1997).

An often important component of subsurface velocity structure (especially in regions exhibiting significant deformation) that is frequently ignored in traveltime tomography studies is anisotropy. If the effects of anisotropy are severe enough, reconstructions that assume isotropy may produce erroneous results (Watanabe et al., 1996). The principal reason for ignoring anisotropy is that the data can usually be satisfied adequately with an isotropic model. Introducing anisotropy effectively increases the number of unknowns considerably for the same spatial resolution, meaning that the inverse problem is likely to become under-determined for most realistic datasets. As a consequence, traveltime tomography studies that have considered anisotropy usually make several assumptions about structure to reduce the number of unknowns. For example, Hearn (1996) inverted P_n traveltimes for lateral variations in both velocity and anisotropy in the uppermost mantle. The Moho surface along which the ray paths propagate is assumed to be planar, and is parameterized in terms of cells. Each cell is associated with three unknowns: slowness and two coefficients of azimuthal anisotropy. Using multiple traveltime datasets and later arriving phases as suggested above may

make anisotropy tomography a more realistic proposition as data coverage will be greatly improved.

The traveltimes of seismic phases constitute just one of the components of the recorded waveform. Combining traveltimes with amplitude information or using the shape of the waveform itself is likely to yield much more information than traveltimes alone. Wang & Pratt (1997) combined traveltime and amplitude data in reflection tomography. They showed that using both datasets in an inversion for velocity or interface structure results in more accurate solutions compared to inversions that invert only one data type. Neele et al. (1993) combined amplitude and traveltimes from a teleseismic dataset in a joint inversion for upper mantle structure. Pratt et al. (1996) demonstrated via 2-D synthetic tests that wavefield inversion of wide-angle seismic data produces velocity models of significantly greater resolution than traveltime inversion. Combining traveltime data with non-seismic data is also likely to be further developed in the future. Using both gravity data and traveltime data from wide-angle surveys to constrain crustal models has been done in the past. For example, Vogt et al. (1998) used a forward modeling approach to reconcile the two data types and produce a structural model for the Hatton Basin and continental margin. Lees & VanDecar (1991) simultaneously inverted local earthquake traveltimes and Bouger gravity anomalies for crustal structure in Western Washington.

Like most other areas of science, increases in computing power will also influence the way in which traveltime tomography will be approached in the future. Indeed, many of the potential future developments described above are reliant to some extent on the availability of greater computing power. Computationally intensive methods for solving the inverse step of the problem like genetic algorithms or simulated annealing (Section 2.3.3) will become more feasible for larger scale problems. Currently, they are limited to perhaps several hundred unknowns at most. Also, the results from other studies could be used to improve model accuracy. For example, a teleseismic model could be embedded within a lower resolution global model to help prevent the mapping of aspherical mantle structure into the local region of interest. This has been done in the context of regional-scale traveltime inversion by Widiyantoro & van der Hilst (1997); in this study, global and regional traveltimes were inverted simultaneously for a regional model on a $1^\circ \times 1^\circ$ grid embedded within a global model with a $5^\circ \times 5^\circ$ cell size. However, the non-linearity of the problem was not accounted for in the inversion, probably due in part to the significant computational expense involved.

In conclusion, there are still many new ways in which seismic traveltime information can be exploited in tomographic-style problems. Most current methods use only a small subset of the recorded information, such as first arrivals. The new generation of methods are likely to be able to represent continuous and discontinuous structure, use a wide variety of data geometries (e.g. coincident reflection, wide-angle etc.), incorporate traveltimes from later phases, account for anisotropy, use other aspects of the waveform such as amplitudes, and perhaps be able to use other data types such as gravity. The optimum method would use all the information contained in the recorded waveform to constrain structure, but practical implementation of such an approach is a long way off. Seismic traveltimes have played an important role in helping us to understand the structure and composition of the crust and lithosphere via tomographic inversion, and will continue to do so well into the future.

ACKNOWLEDGMENTS

We thank Fiona Darbyshire for kindly supplying Fig. 29 and Fig. 30, and Frank Graeber for kindly supplying Fig. 35, Fig. 36, Fig. 37 and Fig. 38. Greg Houseman is thanked for useful discussions on an early version of this work, and an anonymous referee is thanked for constructive comments on the original manuscript.

References

Aarts, E. & Korst, J., 1989. *Simulated Annealing and Boltzmann machines*, John Wiley & Sons, Chichester.

Abers, G. A., 1994. Three-dimensional inversion of regional P and S arrival times in the East Aleutians and sources of subduction zone gravity highs, *J. Geophys. Res.*, **99**, 4395–4412.

Abgrall, R., 1999. Big ray-tracing and eikonal solver on unstructured grids: Application to the computation of a multivalued traveltime field in the Marmousi model, *Geophysics*, **64**, 230–239.

Achauer, U., 1994. New ideas on the Kenya rift based on the inversion of the combined dataset of the 1985 and 1989/90 seismic tomography experiments, *Tectonophysics*, **236**, 305–329.

Afnimar & Koketsu, K., 2000. Finite difference traveltime calculation for head waves travelling along an irregular interface, *Geophys. J. Int.*, **143**, 729–734.

Aki, K. & Lee, W. H. K., 1976. Determination of the three-dimensional velocity anomalies under a seismic array using first P arrival times from local earthquakes 1. A homogeneous intial model, *J. Geophys. Res.*, **81**, 4381–4399.

Aki, K. & Richards, P. G., 1980. *Quantitative seismology: theory and methods*, W. H. Freeman, San Francisco.

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

Alkhalifah, T. & Fomel, S., 2001. Implementing the fast marching eikonal solver: Spherical versus Cartesian coordinates, *Geophysical Prospecting*, **49**, 165–178.

Ankeny, L. A., Braile, L. W., & Olsen, K. H., 1986. Upper crustal structure beneath the Jemez Mountains volcanic field, New Mexico, determined by three-dimensional simultaneous inversion of seismic refraction and earthquake data, *J. Geophys. Res.*, **91**, 6188–6198.

Asad, A. M., Pullammanappallil, S. K., Anooshehpoor, A., & Louie, J. N., 1999. Inversion of traveltime data for earthquake locations and three-dimensional velocity structure in the Eureka Valley area, eastern California, *Bull. Seism. Soc. Am.*, **89**, 796–810.

Bartels, R. H., Beatty, J. C., & Barsky, B. A., 1987. *An introduction to splines for use in computer graphics and geometric modelling*, Morgan Kaufmann, Los Altos.

Benamou, J. D., 1996. Big ray tracing: Multivalued travel time field computation using viscosity solutions of the Eikonal equation, *J. Comp. Phys.*, **128**, 463–474.

Benz, H. M. & Smith, R. B., 1984. Simultaneous inversion for lateral velocity variations and hypocenters in the Yellowstone region using earthquake and refraction data, *J. Geophys. Res.*, **89**, 1208–1220.

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.

Bijwaard, H. & Spakman, W., 2000. Non-linear global P-wave tomography by iterated linearized inversion, *Geophys. J. Int.*, **141**, 71–82.

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

Bishop, T. P., Bube, K. P., Cutler, R. T., Langan, R. T., Love, P. L., Resnick, J. R., Shuey, R. T., Spindler, D. A., & Wyld, H. W., 1985. Tomographic determination of velocity and depth in laterally varying media, *Geophysics*, **50**, 903–923.

Blundell, C. A., 1993. *Resolution analysis of seismic P-wave velocity estimates using reflection tomographic inversion*, Ph.D. thesis, Monash University.

Bosch, M., 1997. P wave velocity tomography of the Venezuelan region from local arrival times, *J. Geophys. Res.*, **102**, 5455–5472.

Boschetti, F., Dentith, M. K., & List, R. D., 1996. Inversion of seismic refraction data using genetic algorithms, *Geophysics*, **61**, 1715–1727.

Cao, S. & Greenhalgh, S., 1994. Finite-difference solution of the eikonal equation using an efficient, first-arrival, wavefront tracking scheme, *Geophysics*, **59**, 632–643.

Cara, M. & Lévêque, J. J., 1987. Waveform inversion using secondary observables, *Geophys. Res. Lett.*, **14**, 1046–1049.

Carroll, S. & Beresford, G., 1996. Combining reflection tomography with layer replacement for velocity analysis of near-surface reefs, *Geophysics*, **61**, 561–569.

Cassell, B. R., 1982. A method for calculating synthetic seismograms in laterally varying media, *Geophys. J. Royal Astr. Soc.*, **69**, 339–354.

Červený, V., 1987. Ray tracing algorithms in three-dimensional laterally varying layered structures, in *Seismic tomography: With applications in global seismology and exploration geophysics*, edited by G. Nolet, pp. 99–133, D. Reidel, Dordrecht.

Červený, V., 2001. *Seismic Ray Theory*, Cambridge University Press, Cambridge.

Červený, V. & Pšenčik, I., 1983. Gaussian beams and paraxial ray approximation in three-dimensional elastic inhomogeneous media, *J. Geophys.*, **53**, 1–15.

Červený, V., Klimeš, L., & Pšenčik, I., 1984. Paraxial ray approximations in the computation of seismic wavefields in inhomogeneous media, *Geophys. J. Royal Astr. Soc.*, **79**, 89–104.

Chapman, C. H. & Drummond, R., 1982. Body-wave seismograms in inhomogeneous media using Maslov asymptotic theory, *Bull. Seism. Soc. Am.*, **72**, S277–S317.

Cheng, N. & House, L., 1996. Minimum traveltime calculations in 3-D graph theory, *Geophysics*, **61**, 1895–1898.

Chiarabba, C., Amoto, A., & Meghraoui, M., 1997. Tomographic images of the El Asnam fault zone, *J. Geophys. Res.*, **102**, 24,485–24,498.

Chiu, S. K. L., Kanasewich, E. R., & Phadke, S., 1986. Three-dimensional determination of structure and velocity by seismic tomography, *Geophysics*, **51**, 1559–1571.

Claerbout, J. F. & Muir, F., 1973. Robust modelling with erratic data, *Geophysics*, **38**, 826–844.

Clowes, R. M., Zelt, C. A., Amor, J. R., & Ellis, R. M., 1995. Lithospheric structure in the southern Canadian Cordillera from a network of seismic refraction lines, *Can. J. Earth Sci.*, **32**, 1485–1513.

Constable, S. C., Parker, R. L., & Constable, C. G., 1987. Occam's inversion: A practical algorithm for generating smooth models from electromagnetic sounding data, *Geophysics*, **52**, 289–300.

Creager, K. C. & Jordan, T. H., 1984. Slab penetration into the lower mantle, *J. Geophys. Res.*, **89**, 3031–3049.

Darbyshire, F. A., Bjarnason, I. J., White, R. S., & Flórenz, O. G., 1998. Crustal structure above the Iceland mantle plume, imaged by the ICEMELT refraction profile, *Geophys. J. Int.*, **135**, 1131–1149.

Davis, P. M., 1991. Continental rift structures and dynamics with reference to teleseismic studies of the Rio Grande and East African rifts, *Tectonophysics*, **197**, 309–325.

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.

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.

Dijkstra, E. W., 1959. A note on two problems in connection with graphs, *Numer. Math.*, **1**, 269–271.

Dorbath, C. & Paul, A., 1996. Tomography of the Andean crust and mantle at 20°S: First results of the Lithoscope experiment, *Phys. Earth Planet. Inter.*, **97**, 133–144.

Drijkoningen, G. G. & White, R. S., 1995. Seismic velocity structure of oceanic crust by inversion using genetic algorithms, *Geophys. J. Int.*, **123**, 653–664.

Dueker, K., Humphreys, E., & Biasi, G., 1993. Teleseismic imaging of the western United States upper mantle structure using the simultaneous iterative reconstruction technique, in *Seismic tomography: theory and practice*, edited by H. M. Iyer & K. Hirahara, pp. 265–298, Chapman & Hall, London.

Eberhart-Phillips, D., 1986. Three-dimensional velocity structure in northern California coast ranges from inversion of local earthquake arrival times, *Bull. Seism. Soc. Am.*, **76**, 1025–1052.

Eberhart-Phillips, D., 1990. Three-dimensional *P* and *S* velocity structure in the Coalinga Region, California, *J. Geophys. Res.*, **95**, 15,343–15,363.

Eberhart-Phillips, D. & Michael, A. J., 1993. Three-dimensional velocity structure, seismicity, and fault structure in the Parkfield Region, central California, *J. Geophys. Res.*, **98**, 15,737–15,758.

Eberhart-Phillips, D. & Reyners, M., 1997. Continental subduction and three-dimensional crustal structure: The northern South Island, New Zealand, *J. Geophys. Res.*, **102**, 11,848–11,861.

Ettrich, N. & Gajewski, D., 1996. Wave front construction in smooth media for prestack depth migration, *Pageoph*, **148**, 481–502.

Evans, J., Eberhart-Phillips, D., & Thurber, C., 1994. User's manual for $SIMULPS12$ for imaging v_P and v_P/v_S: A derivative of the "Thurber" tomographic inversion $SIMUL3$ for local earthquakes and explosions, Open file report 431, U.S. Geological Survey.

Evans, J. R., 1982. Compressional wave velocity structure of the upper 350 km under the Eastern Snake River Plain near Rexburg, Idaho, *J. Geophys. Res.*, **87**, 2654–2670.

Faria, E. L. & Stoffa, P. L., 1994. Traveltime computation in transversely isotropic media, *Geophysics*, **59**, 272–281.

Farra, V. & Madariaga, R., 1988. Non-linear reflection tomography, *Geophys. J.*, **95**, 135–147.

Fischer, R. & Lees, J. M., 1993. Shortest path ray tracing with sparse graphs, *Geophysics*, **58**, 987–996.

Fletcher, R. & Reeves, C. M., 1964. Function minimization by conjugate gradients, *Comput. J.*, **7**, 149–154.

Frederiksen, A. W., Bostock, M. G., VanDecar, J. C., & Cassidy, J. F., 1998. Seismic structure of the upper mantle beneath the northern Canadian Cordillera from teleseismic travel-time inversion, *Tectonophysics*, **294**, 43–55.

Gjøystdal, H., Reinhardsen, J. E., & Ursin, B., 1984. Traveltime and wavefront curvature calculations in three-dimensional inhomogeneous layered media with curved interfaces, *Geophysics*, **49**, 1466–1494.

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.

Goldberg, D. E., 1989. *Genetic Algorithms in Search, Optimization, and Machine Learning*, Addison-Wesley, Reading.

Gorbatov, A., Widiyantoro, S., Fukao, Y., & Gordeev, E., 2000. Signature of remnant slabs in the North Pacific from P-wave tomography, *Geophys. J. Int.*, **142**, 27–36.

Gorbatov, A., Fukao, Y., Widiyantoro, S., & Gordeev, E., 2001. Seismic evidence for a mantle plume oceanwards of the Kamchatka-Aleutian trench junction, *Geophys. J. Int.*, **146**, 282–288.

Grad, M., Shiobara, H., Janik, T., Guterch, A., & Shimamura, H., 1997. Crustal model of the Bransfield Rift, West Antarctica, from detailed OBS refraction experiments, *Geophys. J. Int.*, **130**, 506–518.

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.

Granet, M. & Trampert, J., 1989. Large-scale P-velocity structures in the Euro-Mediterranean area, *Geophys. J. Int.*, **99**, 583–594.

Grevemeyer, I., Fleuh, E. R., Reichert, C., Biala, J., Kläschen, D., & Kopp, C., 2001. Crustal architecture and deep structure of the Ninetyeast Ridge hotspot trail from active-source ocean bottom seismology, *Geophys. J. Int.*, **144**, 414–431.

Gubbins, D., 1992. *Seismology and plate tectonics*, Cambridge University Press, Cambridge.

Guiziou, J. L., Mallet, J. L., & Madariaga, R., 1996. 3-D seismic reflection tomography on top of the GOCAD depth modeler, *Geophysics*, **61**, 1499–1510.

Gürbüz, C. & Evans, J. R., 1991. A seismic refraction study of the western Tuz Gölü basin, central Turkey, *Geophys. J. Int.*, **106**, 239–251.

Hammer, P. T. C., Dorman, L. M., Hildebrand, J. A., & Cornuelle, B. D., 1994. Jasper Seamount structure: Seafloor seismic refraction tomography, *J. Geophys. Res.*, **99**, 6731–6752.

Hasegawa, A., Umino, N., & Takagi, A., 1978. Double-planed deep seismic zone and upper-mantle structure in the northeastern Japan arc, *Geophys. J. Royal Astr. Soc.*, **54**, 281–296.

Haslinger, F., Kissling, E., Ansorge, J., Hatzfeld, D., Papadimitriou, E., Karakostas, V., Makropoulos, K., Kahle, H. G., & Peter, Y., 1999. 3D crustal structure from local earthquake tomography around the Gulf of Arta (Ionian region, NW Greece), *Tectonophysics*, **304**, 201–218.

Hearn, T. M., 1996. Anisotropic Pn tomography in the western United States, *J. Geophys. Res.*, **101**, 8403–8414.

Hestenes, M. & Stiefel, E., 1952. Methods of conjugate gradients for solving linear systems, *Nat. Bur. Stand. J. Res.*, **49**, 409–436.

Hildebrand, J. A., Dorman, L. M., Hammer, P. T. C., Schreiner, A. E., & Cornuelle, B. D., 1989. Seismic tomography of Jasper Seamount, *Geophys. Res. Lett.*, **16**, 1355–1358.

Hole, J. A., 1992. Nonlinear high-resolution three-dimensional travel-time tomography, *J. Geophys. Res.*, **97**, 6553–6562.

Hole, J. A. & Zelt, B. C., 1995. 3-D finite-difference reflection travel times, *Geophys. J. Int.*, **121**, 427–434.

Hole, J. A., Clowes, R. M., & Ellis, R. M., 1992. Interface inversion using broadside seismic refraction data and three-dimensional traveltime calculations, *J. Geophys. Res.*, **97**, 3417–3429.

Humphreys, E. & Clayton, R. W., 1988. Adaption of back projection tomography to seismic travel time problems, *J. Geophys. Res.*, **93**, 1073–1085.

Humphreys, E. D. & Clayton, R. W., 1990. Tomographic image of the Southern California Mantle, *J. Geophys. Res.*, **95**, 19,725–19,746.

Inoue, H., Fukao, Y., Tanabe, K., & Ogata, Y., 1990. Whole mantle P-wave travel time tomography, *Phys. Earth Planet. Inter.*, **59**, 294–328.

Iwasaki, T., Hirata, N., Kanazawa, T., Melles, J., Suyehiro, K., Urabe, T., Möller, L., Makris, J., & Shimamura, H., 1990. Crustal and upper mantle structure in the Ryukyu Island Arc deduced from deep seismic sounding, *Geophys. J. Int.*, **102**, 631–651.

Iwasaki, T., Ozel, O., Moriya, T., Sakai, S., Suzuki, S., Aoki, G., Maeda, T., & Iidaka, T., 1998. Lateral structural variations across a collision zone in central Hokkaido, Japan, as revealed by wide-angle seismic experiments, *Geophys. J. Int.*, **132**, 435–457.

Iyer, H. & Hirahara, K., 1993. *Seismic Tomography: Theory and Practice*, Chapman & Hall, London.

Jarchow, C. M., Catchings, R. D., & Lutter, W. J., 1994. Large-explosive source, wide-recording aperture, seismic profiling on the Columbia Plateau, Washington, *Geophysics*, **59**, 259–271.

Julian, B. R. & Gubbins, D., 1977. Three-dimensional seismic ray tracing, *J. Geophys.*, **43**, 95–113.

Kanasewich, E., Burianyk, M. J. A., Ellis, R. M., Clowes, R. M., White, D. J., Lôté, T., Forsyth, D. A., Luetgert, J. A., & Spence, G. D., 1994. Crustal velocity structure of the Omineca Belt, southwestern Canadian Cordillera, *J. Geophys. Res.*, **99**, 2653–2670.

Kanasewich, E. R. & Chiu, S. K. L., 1985. Least-squares inversion of spatial seismic refraction data, *Bull. Seism. Soc. Am.*, **75**, 865–880.

Kao, H. & Rau, R. J., 1999. Detailed structures of the subducted Philippine Sea plate beneath northeast Taiwan: A new type of double seismic zone, *J. Geophys. Res.*, **104**, 1015–1033.

Kawakatsu, H., 1985. Double seismic zone in Tonga, *Nature*, **316**, 53–55.

Keller, W. R., Anderson, D. L., & Clayton, R. W., 2000. Resolution of tomographic models of the mantle beneath Iceland, *Geophys. Res. Lett.*, **27**, 3993–3996.

Kennett, B. L. N., 1998. *Seismic wave propagation and seismic tomography*, Research School of Earth Sciences, Institute of Advanced Studies, The Australian National University, Canberra.

Kennett, B. L. N., 2002. *The Seismic Wavefield, Volume 2: Interpretation of Seismograms on Regional and Global Scales*, Cambridge University Press, Cambridge.

Kennett, B. L. N. & Engdahl, E. R., 1991. Traveltimes for global earthquake location and phase identification, *Geophys. J. Int.*, **105**, 429–465.

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.

Kim, S., 2001. An $O(N)$ level set method for eikonal equations, *SIAM J. Sci. Comput.*, **22**, 2178–2193.

Kirkpatrick, S., Gelatt, Jr, C. D., & Vecchi, M. P., 1983. Optimization by simulated annealing, *Science*, **220**, 671–680.

Klimeš, L. & Kvasnička, M., 1994. 3-D network ray tracing, *Geophys. J. Int.*, **116**, 726–738.

Kodaira, S., Mjelde, R., Gunnarsson, K., Shiobara, H., & Shimamura, H., 1998. Structure of the Jan Mayen microcontinent amd implications for its evolution, *Geophys. J. Int.*, **132**, 383–400.

Kohler, M. D., 1997. Three-dimensional velocity structure and resolution of the core-mantle boundary region from whole-mantle inversions of body waves, *Phys. Earth Planet. Inter.*, **101**, 85–104.

Kohler, M. D. & Davis, P. M., 1997. Crustal thickness variations in southern California from Los Angeles Region Seismic Experiment passive phase teleseismic travel times, *Bull. Seism. Soc. Am.*, **87**, 1330–1344.

Korenaga, J., Holbrook, W. S., Kent, G. M., Kelemen, R. S., Detrick, R. S., Larsen, H. C., Hopper, J. R., & Dahl-Jensen, T., 2000. Crustal structure of the southeast Greenland margin from joint refraction and reflection tomography, *J. Geophys. Res.*, **105**, 21,591–21,614.

Kosloff, D., Sherwood, J., Koren, Z., Machet, E., & Falkovitz, Y., 1996. Velocity and interface depth determination by tomography of depth migrated gathers, *Geophysics*, **61**, 1511–1523.

Kreyszig, E., 1993. *Advanced engineering mathematics*, John Wiley & Sons, Inc., New York.

Lambaré, G., Lucio, P. S., & Hanyga, A., 1996. Two-dimensional multivalued traveltime and amplitude maps by uniform sampling of a ray field, *Geophys. J. Int.*, **125**, 584–598.

Lambeck, K. & Burgess, G., 1992. Deep crustal structure of the Musgrave Block, central Australia: results from teleseismic traveltime anomalies, *Australian Journal of Earth Sciences*, **39**, 1–19.

Lambeck, K., Burgess, G., & Shaw, R. D., 1988. Teleseismic travel-time anomalies and deep crustal structure in central Australia, *Geophys. J.*, **94**, 105–124.

Langan, R. T., Lerche, I., & Cutler, R. T., 1985. Tracing of rays through heterogeneous media: An accurate and efficient procedure, *Geophysics*, **50**, 1456–1465.

Lee, W. H. K. & Pereyra, V., 1993. Mathematical introduction to seismic tomography, in *Seismic tomography: theory and practice*, edited by H. M. Iyer & K. Hirahara, pp. 9–22, Chapman & Hall, London.

Lees, J. M. & VanDecar, J. C., 1991. Seismic tomography constrained by Bouguer gravity anomalies: Applications in western Washington, *Pageoph*, **135**, 31–52.

Lévêque, J. J. & Masson, F., 1999. From ACH tomographic models to absolute velocity models, *Geophys. J. Int.*, **137**, 621–629.

Lévêque, J. J., Rivern, L., & Wittlinger, G., 1993. On the use of the checker-board test to assess the resolution of tomographic inversions, *Geophys. J. Int.*, **115**, 313–318.

Lizarralde, D. & Holbrook, W. S., 1997. U.S. mid-Atlantic margin structure and early thermal evolution, *J. Geophys. Res.*, **102**, 22,855–22,875.

Louden, K. E. & Fan, J., 1998. Crustal structures of Grenville, Makkovik, and southern Nain provinces along the Lithoprobe ECSOOT Transect: Regional seismic refraction and gravity models and their tectonic implications, *Can. J. Earth Sci.*, **35**, 583–601.

Luca, G. D., Filippi, L., Caccamo, D., Neri, G., & Scarpa, R., 1997. Crustal structure and seismicity of southern Tyrrhenian basin, *Phys. Earth Planet. Inter.*, **103**, 117–133.

Lucio, P. S., Lambaré, G., & Hanyga, A., 1996. 3D multivalued travel time and amplitude maps, *Pageoph*, **148**, 449–479.

Lutter, W. J. & Nowack, R. L., 1990. Inversion for crustal structure using reflections from the PASSCAL Ouachita experiment, *J. Geophys. Res.*, **95**, 4633–4646.

Lutter, W. J., Nowack, R. L., & Braile, L., 1990. Seismic imaging of upper crustal structure using travel times from the PASSCAL Ouachita experiment, *J. Geophys. Res.*, **95**, 4621–4631.

Lutter, W. J., Catchings, R. D., & Jarchow, C. M., 1994. An image of the Columbia Plateau from inversion of high-resolution seismic data, *Geophysics*, **59**, 1278–1289.

McCaughey, M. & Singh, S. C., 1997. Simultaneous velocity and interface tomography of normal-incidence and wide-aperture seismic traveltime data, *Geophys. J. Int.*, **131**, 87–99.

McMechan, G. A., 1983. Seismic tomography in boreholes, *Geophys. J. Royal Astr. Soc.*, **74**, 601–612.

McQueen, H. W. S. & Lambeck, K., 1996. Determination of crustal structure in central Australia by inversion of traveltime residuals, *Geophys. J. Int.*, **126**, 645–662.

Menke, W., 1989. *Geophysical data analysis: Discrete inverse theory*, Academic Press, New York.

Mjelde, R., Digranes, P., Shimamura, H., Shiobara, H., Kodaira, S., Brekke, H., Egebjerg, T., Sørenes, N., & Thorbjørnsen, S., 1998. Crustal structure of the northern part of the Vøring Basin, mid-Norway margin, from wide-angle seismic and gravity data, *Tectonophysics*, **293**, 175–205.

Morgan, R. P. L., Barton, P. J., Warner, M., Morgan, J., Price, C., & Jones, K., 2000. Lithospheric structure north of Scotland - I. P-wave modelling, deep reflection profiles and gravity, *Geophys. J. Int.*, **142**, 716–736.

Morozov, I. B., Smithson, S. B., Hollister, L. S., & Diebold, J. B., 1998. Wide-angle seismic imaging across accreted terranes, southeastern Alaska and western British Columbia, *Tectonophysics*, **299**, 281–296.

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

Moser, T. J., 1991. Shortest path calculation of seismic rays, *Geophysics*, **56**, 59–67.

Nakanishi, I., 1985. Three-dimensional structure beneath the Hokkaido-Tohoku region as derived from a tomographic inversion of P-arrival times, *J. Phys. Earth*, **33**, 241–256.

Nakanishi, I. & Yamaguchi, K., 1986. A numerical experiment on nonlinear image reconstruction from first-arrival times for two-dimensional island arc structure, *J. Phys. Earth*, **34**, 195–201.

Navin, D. A., Peirce, C., & Sinha, M. C., 1998. The RAMESSES experiment -II. Evidence for accumulated melt beneath a slow spreading ridge from wide-angle refraction and multichannel reflection seismic profiles, *Geophys. J. Int.*, **135**, 746–772.

Neele, F., VanDecar, J., & Snieder, R., 1993. The use of P wave amplitude data in a joint inversion with travel times for upper mantle velocity structure, *J. Geophys. Res.*, **98**, 12,033–12,054.

Nielsen, L., Thybo, H., & Solodilov, L., 1999. Seismic tomographic inversion of Russian PNE data along profile Kraton, *Geophys. Res. Lett.*, **26**, 3413–3416.

Nolet, G., 1985. Solving or resolving inadequate and noisy tomographic systems, *J. Comp. Phys.*, **61**, 463–482.

Nolet, G., 1987. Waveform tomography, in *Seismic tomography: With applications in global seismology and exploration geophysics*, edited by G. Nolet, pp. 301–322, D. Reidel, Dordrecht.

Nolet, G., 1987. *Seismic tomography: With applications in global seismology and exploration geophysics*, D. Reidel, Dordrecht.

Nolet, G., 1990. Partitioned waveform inversion and two-dimensional structure under the network of autonomously recording seismographs, *J. Geophys. Res.*, **95**, 8499–8512.

Nowack, R. L. & Lyslo, J. A., 1989. Fréchet derivatives for curved interfaces in the ray approximation, *Geophys. J.*, **97**, 497–509.

Oncescu, M. C., Burlacu, V., Anghel, M., & Smalbergher, V., 1984. Three-dimensional P-wave velocity image under the Carpathian Arc, *Tectonophysics*, **106**, 305–319.

Papazachos, C. & Nolet, G., 1997. P and S deep velocity structure of the Hellenic area obtained by robust nonlinear inversion of travel times, *J. Geophys. Res.*, **102**, 8349–8367.

Parker, R. L., 1994. *Geophysical inverse theory*, Princeton University Press, Princeton.

Parsons, T. & Zoback, M. L., 1997. Three-dimensional upper crustal velocity structure beneath San Francisco Peninsula, California, *J. Geophys. Res.*, **102**, 5473–5490.

Parsons, T., McCarthy, J., Kohler, W. M., Ammon, C. J., Benz, H. M., Hole, J. A., & Criley, E. E., 1996. Crustal structure of the Colorada Plateau, Arizona: Application of new long-offset seismic data analysis techniques, *J. Geophys. Res.*, **101**, 11,173–11,194.

Pereyra, V., Lee, W. H. K., & Keller, H. B., 1980. Solving two-point seismic-ray tracing problems in a heterogeneous medium, *Bull. Seism. Soc. Am.*, **70**, 79–99.

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

Popovici, A. M. & Sethian, J. A., 2002. 3-D imaging using higher order fast marching traveltimes, *Geophysics*, **67**, 604–609.

Pratt, R. G., Song, Z. M., Williamson, P., & Warner, M., 1996. Two-dimensional velocity models from wide-angle seismic data by wavefield inversion, *Geophys. J. Int.*, **124**, 323–340.

Press, W. H., Tuekolsky, S. A., Vetterling, W. T., & Flannery, B. P., 1992. *Numerical Recipes in FORTRAN*, Cambridge University Press, Cambridge.

Priestley, K., Cipar, J., Egorkin, A., & Pavlenkova, N., 1994. Upper-mantle velocity structure beneath the Siberian platform, *Geophys. J. Int.*, **118**, 369–378.

Prothero, W. A., Taylor, W. J., & Eickemeyer, J. A., 1988. A fast, two-point, three-dimensional raytracing algorithm using a simple step search method, *Bull. Seism. Soc. Am.*, **78**, 1190–1198.

Protti, M., Schwartz, S. Y., & Zandt, G., 1996. Simultaneous inversion for earthquake location and velocity structure beneath central Costa Rica, *Bull. Seism. Soc. Am.*, **86**, 19–31.

Pullammanappallil, S. K. & Louie, J. N., 1993. Inversion of seismic reflection traveltimes using a nonlinear optimization scheme, *Geophysics*, **58**, 1607–1620.

Qin, F., Luo, Y., Olsen, K. B., Cai, W., & Schuster, G. T., 1992. Finite-difference solution of the eikonal equation along expanding wavefronts, *Geophysics*, **57**, 478–487.

Rawlinson, N., 2000. *Inversion of Seismic Data for Layered Crustal Structure*, Ph.D. thesis, Monash University.

Rawlinson, N. & Houseman, G. A., 1998. Inversion for interface structure using teleseismic traveltime residuals, *Geophys. J. Int.*, **133**, 756–772.

Rawlinson, N., Houseman, G. A., & Collins, C. D. N., 2001a. Inversion of seismic refraction and wide-angle reflection traveltimes for 3-D layered crustal structure, *Geophys. J. Int.*, **145**, 381–401.

Rawlinson, N., Houseman, G. A., Collins, C. D. N., & Drummond, B. J., 2001b. New evidence of Tasmania's tectonic history from a novel seismic experiment, *Geophys. Res. Lett.*, **28**, 3337–3340.

Recq, M., Goslin, J., Charvis, P., & Operto, S., 1998. Small-scale crustal variability within an intraplate structure: The Crozet Bank (southern Indian Ocean), *Geophys. J. Int.*, **134**, 145–156.

Riahi, M. A. & Juhlin, C., 1994. 3-D interpretation of reflected arrival times by finite-difference techniques, *Geophysics*, **59**, 844–849.

Riahi, M. A. & Lund, C. E., 1994. Two-dimensional modelling and interpretation of seismic wide-angle data from the western Gulf of Bothnia, *Tectonophysics*, **239**, 149–164.

Riahi, M. A., Lund, C. E., & Pederson, L. B., 1997. Three-dimensional image of the Moho undulations beneath the Gulf of Bothnia using wide-angle seismic data, *Geophys. J. Int.*, **129**, 461–471.

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.

Ryberg, T., Wenzel, F., Mechie, J., Egorkin, A., Fuchs, K., & Solodilov, L., 1996. Two-dimensional velocity structure beneath northern Eurasia derived from the super long-range seismic profile Quartz, *Bull. Seism. Soc. Am.*, **86**, 857–867.

Saltzer, R. L. & Humphreys, E. D., 1997. Upper mantle *P* wave velocity structure of the eastern Snake River Plain and its relationship to geodynamic models of the region, *J. Geophys. Res.*, **102**, 11,829–11,841.

Sambridge, M. & Mosegaard, K., 2001. Monte Carlo methods in geophysical inverse problems, *Rev. of Geophys.*, **submitted**.

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.

Sambridge, M. S. & Drijkoningen, G., 1992. Genetic algorithms in seismic waveform inversion, *Geophys. J. Int.*, **109**, 323–342.

Sambridge, M. S. & Kennett, B. L. N., 1990. Boundary value ray tracing in a heterogeneous medium: A simple and versatile algorithm, *Geophys. J. Int.*, **101**, 157–168.

Sato, T., Kosuga, M., & Tanaka, K., 1996. Tomographic inversion for *P* wave velocity structure beneath the northeastern Japan arc using local and teleseismic data, *J. Geophys. Res.*, **101**, 17,597–17,615.

Scales, J. A., 1987. Tomographic inversion via the conjugate gradient method, *Geophysics*, **52**, 179–185.

Scales, J. A. & Snieder, R., 1997. To Bayes or not to Bayes, *Geophysics*, **62**, 1045–1046.

Scott, J. S., Masters, T. G., & Vernon, F. L., 1994. 3-D velocity structure of the San Jacinto fault zone near Anza, California-I. *P* waves, *Geophys. J. Int.*, **119**, 611–626.

Seber, D., Barazangi, M., Tadili, B. A., Ramdani, M., Ibenbrahim, A., & Sari, D. B., 1996. Three-dimensional upper mantle structure beneath the intraplate Atlas and interplate Rif mountains of Morocco, *J. Geophys. Res.*, **101**, 3125–3138.

Sethian, J. A., 1999. *Level Set Methods and Fast marching Methods*, Cambridge University Press, Cambridge.

Sethian, J. A. & Popovici, A. M., 1999. 3-D traveltime computation using the fast marching method, *Geophysics*, **64**, 516–523.

Shalev, E., 1993. Cubic B-slines: Strategies of translating a simple structure to B-spline parameterization, *Bull. Seism. Soc. Am.*, **83**, 1617–1627.

Shaw, P. R. & Orcutt, A., 1985. Waveform inversion of seismic refraction data and applications to young Pacific crust, *Geophys. J. Royal Astr. Soc.*, **82**, 375–414.

Smith, W. H. F. & Wessel, P., 1990. Gridding with continuous curvature splines in tension, *Geophysics*, **55**, 293–305.

Snieder, R., 1998. The role of nonlinearity in inverse problems, *Inverse Problems*, **14**, 387–404.

Snieder, R. & Sambridge, M., 1993. The ambiguity in ray perturbation theory, *J. Geophys. Res.*, **98**, 22,021–22,034.

Spakman, W., 1991. Delay-time tomography of the upper mantle below Europe, the Mediterranean and Asia Minor, *Geophys. J. Int.*, **107**, 309–332.

Spakman, W., 1993. Iterative strategies for nonlinear travel-time tomography using global earthquake data, in *Seismic tomography: theory and practice*, edited by H. M. Iyer & K. Hirahara, pp. 190–226, Chapman & Hall, London.

Staples, R. K., White, R. S., Brandsdóttir, Menke, W., Maguire, P. K. H., & McBride, J. H., 1997. Färoe-Iceland Ridge experiment 1. Crustal structure of northeastern Iceland, *J. Geophys. Res.*, **102**, 7849–7866.

Steck, L. K., Thurber, C. H., Fehler, M., Lutter, W. J., Roberts, P. M., Baldridge, W. S., Stafford, D. G., & Sessions, R., 1998. Crust and upper mantle P wave velocity structure beneath Valles caldera, New Mexico: Results from the Jemez teleseismic tomography experiment, *J. Geophys. Res.*, **103**, 24,301–24,320.

Steinhoff, J., Fan, M., & Wang, L., 2000. A new Eulerian method for the computation of propagating short acoustic and electromagnetic pulses, *J. Comp. Phys.*, **157**, 683–706.

Tarantola, A., 1987. *Inverse problem theory*, Elsevier, Amsterdam.

Telford, W. M., Geldart, L. P., Sheriff, R. E., & Keys, D. A., 1976. *Applied Geophysics*, Cambridge University Press, Cambridge.

Telford, W. M., Geldart, L. P., & Sheriff, R. E., 1990. *Applied Geophysics*, Cambridge University Press, Cambridge.

Thomson, C. J. & Gubbins, D., 1982. Three-dimensional lithospheric modelling at NORSAR: linearity of the method and amplitude variations from the anomalies, *Geophys. J. Royal Astr. Soc.*, **71**, 1–36.

Thurber, C. H., 1983. Earthquake locations and three-dimensional crustal structure in the Coyote Lake area, central California, *J. Geophys. Res.*, **88**, 8226–8236.

Thurber, C. H., 1993. Local earthquake tomography: Velocities and V_p/V_s - theory, in *Seismic tomography: theory and practice*, edited by H. M. Iyer & K. Hirahara, pp. 563–583, Chapman & Hall, London.

Thurber, C. H. & Ellsworth, W. L., 1980. Rapid solution of ray tracing problems in heterogeneous media, *Bull. Seism. Soc. Am.*, **70**, 1137–1148.

Toomey, D. R., Solomon, S. C., & Purdy, G. M., 1994. Tomographic imaging of the shallow crustal structure of the East Pacific Rise at 9°30′N, *J. Geophys. Res.*, **99**, 24,135–24,157.

Um, J. & Thurber, C., 1987. A fast algorithm for two-point seismic ray tracing, *Bull. Seism. Soc. Am.*, **77**, 972–986.

van der Hilst, R. D., Widiyantoro, S., & Engdahl, E. R., 1997. Evidence for deep mantle circulation from global tomography, *Nature*, **386**, 578–584.

van Trier, J. & Symes, W. W., 1991. Upwind finite-difference calculation of traveltimes, *Geophysics*, **56**, 812–821.

VanDecar, J. C., 1991. *Upper-mantle structure of the Cascadia subduction zone from non-linear teleseismic travel-time inversion*, Ph.D. thesis, University of Washington.

VanDecar, J. C. & Snieder, R., 1994. Obtaining smooth solutions to large, linear, inverse problems, *Geophysics*, **59**, 818–829.

VanDecar, J. C., James, D. E., & Assumpção, M., 1995. Seismic evidence for a fossil mantle plume beneath South America and implications for plate driving forces, *Nature*, **378**, 25–31.

Vidale, J. E., 1988. Finite-difference calculations of traveltimes, *Bull. Seism. Soc. Am.*, **78**, 2062–2076.

Vidale, J. E., 1990. Finite-difference calculations of traveltimes in three dimensions, *Geophysics*, **55**, 521–526.

Vinje, V., Iversen, E., & Gjøystdal, H., 1993. Traveltime and amplitude estimation using wavefront construction, *Geophysics*, **58**, 1157–1166.

Vinje, V., Iversen, E., Åstebøl, K., & Gjøystdal, H., 1996. Estimation of multivalued arrivals in 3D models using wavefront construction - Part I, *Geophysical Prospecting*, **44**, 819–842.

Vinje, V., Åstebøl, K., Iversen, E., & Gjøystdal, H., 1999. 3-D ray modelling by wavefront construction in open models, *Geophysical Prospecting*, **64**, 1912–1919.

Vogt, U., Makris, J., O'Reilly, B., Hausser, F., Readman, P. W., Jacob, A. W. B., & Shannon, P. M., 1998. The Hatton Basin and continental margin: Crustal structure from wide-angle seismic and gravity data, *J. Geophys. Res.*, **103**, 12,545–12,566.

Walck, M. C., 1988. Three-dimensional V_p/V_s variations for the Coso region, California, *J. Geophys. Res.*, **93**, 2047–2052.

Walck, M. C. & Clayton, R. W., 1987. P wave velocity variations in the Coso region, California, derived from local earthquake travel times, *J. Geophys. Res.*, **92**, 393–405.

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.

Wang, W. & Houseman, G. A., 1994. Inversion of reflection seismic amplitude data for interface geometry, *Geophys. J. Int.*, **117**, 92–110.

Wang, Y. & Houseman, G. A., 1995. Tomographic inversion of reflection seismic amplitude data for velocity variation, *Geophys. J. Int.*, **123**, 355–372.

Wang, Y. & Pratt, R. G., 1997. Sensitivities of seismic traveltimes and amplitudes in reflection tomography, *Geophys. J. Int.*, **131**, 618–642.

Watanabe, T., Hirai, t., & Sassa, K., 1996. Seismic traveltime tomography in anisotropic heterogeneous media, *J. Appl. Geophys.*, **35**, 133–143.

Weiland, C. M., Steck, L. K., Dawson, P. B., & Korneev, V. A., 1995. Nonlinear teleseismic tomography at Long Valley caldera, using three-dimensional minimum travel time ray tracing, *J. Geophys. Res.*, **100**, 20,379–20,390.

White, D. J., 1989. Two-dimensional seismic refraction tomography, *Geophys. J.*, **97**, 223–245.

Whitley, D. L., 1994. A genetic algorithm tutorial, *Stats. and Comput.*, **4**, 65–85.

Widiyantoro, S. & van der Hilst, R., 1997. Mantle structure beneath Indonesia inferred from high-resolution tomographic imaging, *Geophys. J. Int.*, **130**, 167–182.

Wiggins, S. M., Dorman, L. M., Cornuelle, B. D., & Hildebrand, J. A., 1996. Hess deep rift valley structure from seismic tomography, *J. Geophys. Res.*, **101**, 22,335–22,353.

Williamson, P. R., 1990. Tomographic inversion in reflection seismology, *Geophys. J. Int.*, **100**, 255–274.

Ye, S., Flueh, E. R., Klaeschen, D., & von Huene, R., 1997. Crustal structure along the EDGE transect beneath the Kodiak shelf off Alaska derived from OBH seismic refraction data, *Geophys. J. Int.*, **130**, 283–302.

Yilmaz, O., 1987. *Seismic data processing*, Society of Exploration Geophysicists, Tulsa.

Zelt, B. C., Ellis, R. M., Clowes, R. M., & Hole, J. A., 1996. Inversion of three-dimensional wide-angle seimic data from the southwestern Canadia Cordillera, *J. Geophys. Res.*, **286**, 209–221.

Zelt, B. C., Ellis, R. M., Zelt, C. A., Hyndman, R. D., Lowe, C., Spence, G. D., & Fisher, M. A., 2001. Three-dimensional crustal velocity structure beneath the Strait of Georgia, British Columbia, *Geophys. J. Int.*, **144**, 695–712.

Zelt, C. A., 1998. Lateral velocity resolution from three-dimensional seismic refraction data, *Geophys. J. Int.*, **135**, 1101–1112.

Zelt, C. A., 1999. Modelling strategies and model assessment for wide-angle seismic traveltime 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.

Zelt, C. A. & White, D. J., 1995. Crustal structure and tectonics of the southeastern Canadian Cordillera, *J. Geophys. Res.*, **100**, 24,255–24,273.

Zelt, C. A., Hojka, A. M., Flueh, E. R., & McIntosh, K. D., 1999. 3D simultaneous seismic refraction and reflection tomography of wide-angle data from the central Chilean margin, *Geophys. Res. Lett.*, **26**, 2577–2580.

Zhang, J. & Toksöz, M. N., 1998. Nonlinear refraction traveltime tomography, *Geophysics*, **63**, 1726–1737.

Zhao, D. & Kanamori, H., 1992. P-wave image of the crust and uppermost mantle in southern California, *Geophys. Res. Lett.*, **23**, 2329–2332.

Zhao, D., Hasegawa, A., & Horiuchi, S., 1992. Tomographic imaging of P and S wave velocity structure beneath Northeastern Japan, *J. Geophys. Res.*, **97**, 19,909–19,928.

Zhao, D., Hasegawa, A., & Kanamori, H., 1994. Deep structure of Japan subduction zone as derived from local, regional, and teleseismic events, *J. Geophys. Res.*, **99**, 22,313–22,329.

Zhao, D., Kanamori, H., & Humphreys, E., 1996. Simultaneous inversion of local and teleseismic data for the crust and mantle structure of southern California, *Phys. Earth Planet. Inter.*, **93**, 191–214.

Zhao, Z., Kubota, R., Fumio, S., & Iizuka, S., 1997. Crustal structure in the southern Kanto-Tokai region derived from tomographic method for seismic explosion survey, *J. Geophys. Res.*, **45**, 433–453.

Zhu, H. & Ebel, J. E., 1994. Tomographic inversion for the seismic velocity structure beneath northern New England using seismic refraction data, *J. Geophys. Res.*, **99**, 15,331–15,357.

Zielhuis, A. & van der Hilst, R. D., 1996. Upper-mantle shear velocity beneath eastern Australia from inversion of waveforms from SKIPPY portable arrays, *Geophys. J. Int.*, **127**, 1–16.