doi: 10.1111/j.1365-246X.2006.03078.x

A practical grid-based method for tracking multiple refraction and reflection phases in three-dimensional heterogeneous media

M. de Kool,* N. Rawlinson and M. Sambridge

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

Accepted 2006 May 23. Received 2006 May 8; in original form 2005 August 21

SUMMARY

We present a practical grid-based method in 3-D spherical coordinates for computing multiple phases comprising any number of reflection and transmission branches in heterogeneous layered media. The new scheme is based on a multistage approach which treats each layer that the wave front enters as a separate computational domain. A finite-difference eikonal solver known as the fast-marching method (FMM) is reinitialized at each interface to track the evolving wave front as either a reflection back into the incident layer or a transmission through to the adjacent layer. Unlike the standard FMM, which only finds first arrivals, this multistage approach can track those later arriving phases explicitly caused by the presence of discontinuities. Notably, the method does not require an irregular mesh to be constructed in order to connect interface nodes to neighbouring velocity nodes which lie on a regular grid. To improve accuracy, local grid refinement is used in the neighbourhood of a source point where wave front curvature is high. The method also provides a way to trace reflections from an interface that are not the first arrival (e.g. the global *PP* phase). These are computed by initializing the multistage FMM from both the source and receiver, propagating the two wave fronts to the reflecting interface, and finding stationary points of the sum of the two traveltime fields on the reflecting interface.

A series of examples are presented to test the efficiency, accuracy and robustness of the new scheme. As well as efficiently computing various global phases to an acceptable accuracy through the *ak*135 model, we also demonstrate the ability of the scheme to track complex crustal phases that may be encountered in coincident reflection, wide-angle reflection/refraction or local earthquake surveys. In one example, a variety of phases are computed in the presence of a realistic subduction zone, which includes several layer pinch-outs and a subducting slab. Our numerical tests show that the new scheme is a practical and robust alternative to conventional ray tracing for finding various phases in layered media at a variety of scales.

Key words: fast-marching method, finite-difference methods, ray tracing, seismic phases, seismic wave propagation, traveltime.

1 INTRODUCTION

Grid based (or 'Eulerian') traveltime prediction schemes are now firmly established as viable, if not preferable, alternatives to conventional ray-tracing (or 'Lagrangian') methods for computing first arrival seismic traveltimes in the presence of heterogeneous 2-D or 3-D media. One of the first practical grid-based schemes was developed by Vidale (1988), who used finite-difference solutions of the eikonal equation to compute traveltimes throughout a gridded velocity field. Since this seminal paper, numerous schemes for solving the eikonal equation with finite differences have appeared (Podvin & Lecomte 1991; van Trier & Symes 1991; Qin *et al.* 1992; Cao & Greenhalgh 1994; Kim & Cook 1999; Afnimar & Koketsu 2000; Qian & Symes 2002).

An alternative grid-based scheme for computing seismic traveltimes, which is less widely used than finite-difference eikonal solvers, is Shortest Path Raytracing or SPR. This approach appeals directly to Fermat's principle of stationary time by using Dijkstra-like algorithms to locate minimum traveltime paths through a network or graph which connects the nodes of a gridded velocity field (Nakanishi & Yamaguchi 1986; Moser 1991; Fischer & Lees 1993; Cheng & House 1996). Finite-difference eikonal solvers and, to a lesser extent, SPR are now commonly used in several classes of seismic imaging; for example, in refraction tomography (Hole 1992; Toomey *et al.* 1994; Zelt & Barton 1998; Day *et al.* 2001; Zelt *et al.* 2001), and in the migration of coincident reflection sections (Gray & May 1994; Bevc 1997; Zhao *et al.* 1998; Buske 1999; Popovici & Sethian 2002).

[*] Now at: Geoscience Australia, Symonston ACT 2609, Australia.

Compared to conventional shooting and bending methods of ray-tracing (e.g. Julian & Gubbins 1977; Cassell 1982; Um & Thurber 1987; Sambridge 1990; Koketsu & Sekine 1998) grid-based traveltime solvers have several distinct advantages. First, most are capable of computing traveltimes to all points of a velocity medium, and will correctly find diffractions in ray shadow zones (Vidale 1988). Secondly, while shooting and bending schemes may fail to converge to the true two point path, even in mildly heterogeneous media, many grid-based schemes are highly stable and will converge to the correct solution. Thirdly, grid-based schemes can be very efficient in computing traveltimes to the level of accuracy that is required in practical problems, particularly when there is a large number of sources or receivers. Often due to the difficulty of locating two point paths, ray tracing can be quite inefficient. Finally, finite-difference eikonal solvers and SPR are designed to consistently find the first-arrival traveltime in a continuous medium. Ray-tracing schemes also tend to only find single arrivals, but there is usually no guarantee as to whether the selected arrival is a first or later arrival.

First-arrival traveltimes are frequently used in seismic studies, but it is often the case that later arrivals carry most of the seismic energy and hence dominate the recorded signal. One of the current challenges in seismic imaging is to exploit such arrivals in order to produce more detailed and robust earth models. For this to happen, efficient numerical schemes are required to predict all arrivals of significant amplitude. The most practical and well developed of such schemes to date is wave front construction (Vinje *et al.* 1993; Lambaré *et al.* 1996; Vinje *et al.* 1999; Xu & Lambaré 2004), which propagates a wave front in a series of time steps using local ray tracing and interpolation. However, complex data structures are required to keep track of the wave front, and parts of the wave front with high curvature can be poorly resolved.

A number of grid-based schemes have been proposed for tracking later arrivals. One approach is to partition the multivalued solution into a series of single-valued solutions and apply first-arrival solvers in each of the segregated regions (Benamou 1996; Symes & Qian 2003). Although computationally efficient, it is hard to see how such a method can successfully locate all multipaths in complex media. Recently, several schemes based on the level set method have been proposed for tracking wave fronts in phase space (e.g. Fomel & Sethian 2002; Osher *et al.* 2002). While their potential for robustly computing all arrivals has been demonstrated, further development is required, in particular to make them more computationally efficient.

Although a practical Eulerian scheme for computing all multiple arrivals is yet to emerge, it should be possible to construct efficient and robust grid-based schemes to track a limited but useful class of later arriving phases associated with the presence of interfaces, which separate media of different wavespeed. So far, relatively few studies have attempted this, and ray tracing still appears to be the method of choice in layered media, for example, when global phases (pP, PcP, *PP* etc.) or crustal reflections, refractions and multiples are required. One approach for computing reflections with a grid-based technique is to track first-arrival traveltime fields from both source and receiver to the entire interface, and then use Fermat's principle of stationary time to locate reflection points along the interface (Podvin & Lecomte 1991; Riahi & Juhlin 1994). Using the 3-D finite-difference scheme of Vidale (1990), Hole & Zelt (1995) assumed that the impinging wave front and interface are sufficiently smooth to validate a local planar approximation, which allows reflected traveltimes to nodes adjacent to the interface to be computed using Snell's law. SPR can also be used to find reflections by requiring that the shortest path visit a specified set of nodes that lie on the interface (Moser 1991).

In an attempt to find a more robust and general approach to tracking later arrivals caused by the presence of interfaces, Rawlinson & Sambridge (2004a,b) developed a multistage fast-marching method (FMM) for complex layered media in 2-D Cartesian coordinates. FMM tracks the evolution of first-arrival wave fronts via finite-difference solution of the eikonal equation (Sethian 1996; Sethian & Popovici 1999), and distinguishes itself by combining computational efficiency with a high degree of stability (unconditional for first-order accurate traveltimes). The multistage FMM treats each layer that the wave front enters as a separate computational domain. A wave front is propagated through a layer until all points of a bounding interface are intersected. A reflected wave front can then be tracked by reinitializing FMM from the narrow band of interface traveltimes back into the incident layer; a transmitted wave front can be tracked by reinitializing FMM into the adjacent layer. Phases composed of any number of reflection and refraction events can therefore be built up by this multistage approach, although sequential reflections from the same interface cannot be tracked. To facilitate the tracking of wave fronts to and from an interface, a triangulation routine is used to locally suture interface nodes to neighbouring velocity nodes. Local grid refinement about the source is also implemented to improve accuracy.

In this paper, we build on the ideas of Rawlinson & Sambridge (2004a,b) to develop a practical grid-based scheme for tracking multiple reflection and refraction phases in complex 3-D layered media in spherical coordinates. A multistage FMM approach is used to track wave fronts from one interface to the next, but the new method avoids the need to construct a local grid about the interface to suture interface nodes to neighbouring velocity nodes. The use of spherical coordinates means that traveltime prediction problems at a variety of scales can be solved. Reflections from an interface that are not the first ones to arrive can be found by initiating the multistage FMM from both the source and receiver and using Fermat's principle of stationary time to find the reflection point. This involves locating the stationary points on a potentially convoluted traveltime surface, and allows, for example, global phases such as *pP* and *PP* to be tracked. In addition to describing the method, several numerical experiments are carried out to demonstrate the viability of the new scheme at a variety of scales. We aim to show that ray-tracing schemes should not be considered the default approach for computing traveltimes in layered media, even at global scales, as Eulerian schemes are generally more robust, and can be at least as flexible and efficient.

2 INTRODUCTION TO THE FAST-MARCHING METHOD

2.1 FMM in continuous media

We begin by giving a brief description of FMM in continuous media on a regular grid. For more details, refer to Sethian (1996), Sethian & Popovici (1999), Sethian (1999) and Popovici & Sethian (2002). The eikonal equation states that the magnitude of the traveltime gradient at any point along a wave front is equal to the inverse of the wavespeed at that point and may be written as

$$|\nabla_{\mathbf{x}} T| = s(\mathbf{x}), \qquad (1)$$

where ∇_x is the gradient operator, T is traveltime and $s(\mathbf{x})$ is slowness.

A significant obstacle for finite-difference methods that seek to solve the eikonal equation for the first-arrival traveltime field is that self-intersections of the wave front that arise in heterogeneous media cause discontinuities in the gradient. The eikonal equation cannot be easily solved in the presence of gradient discontinuities because the equation itself requires $\nabla_{\mathbf{x}} T$ to be defined.

This problem can be overcome by taking into account the direction of flow of information when evaluating the gradient in eq. (1). Only upwind finite differences are used, that is, only nodes over which the wave front has already passed are used to compute the arrival time of the front at a given node. The upwind scheme that is usually employed (Sethian & Popovici 1999; Chopp 2001; Popovici & Sethian 2002) may be expressed as

$$\begin{bmatrix} \max\left(D_a^{-x}T, -D_b^{+x}T, 0\right)^2 + \\ \max\left(D_c^{-y}T, -D_d^{+y}T, 0\right)^2 + \\ \max\left(D_e^{-z}T, -D_f^{+z}T, 0\right)^2 \end{bmatrix}_{ijk}^{1/2} = s_{i,j,k} \tag{2}$$

where T is traveltime, (i, j, k) are grid increment variables in any orthogonal coordinate system (x, y, z), and the integer variables a, b, c, d, e, f define the order of accuracy of the upwind finite-difference operator used in each of the six cases. For example, in a Cartesian coordinate system the first two upwind operators for $D^{-x}T_i$ are

$$\begin{aligned} D_1^{-x}T_{i,j,k} &= \frac{T_{i,j,k} - T_{i-1,j,k}}{\delta x} \\ D_2^{-x}T_{i,j,k} &= \frac{3T_{i,j,k} - 4T_{i-1,j,k} + T_{i-2,j,k}}{2\delta x}, \end{aligned} \tag{3}$$

where δx is the grid spacing in x. These operators are easily derived by appropriate summation of the Taylor series expansions for T_{i-1} and T_{i-2}. Which operator is used in eq. (2) depends on the availability of upwind traveltimes and the maximum order allowed. First-order accurate schemes only use D_1 operators and second-order accurate schemes preferentially use D_2 operators. Strictly speaking, the second-order method is really mixed order because it will use first-order approximations when causality does not permit the use of the required operator. For example, if we implement a second-order method and $T_{i-1} > T_{i-2}$, then the operator used would be $D_2^{-x}T_i$; while if $T_{i-1} < T_{i-2}$ we would have to resort to $D_1^{-x}T_i$. This typically only occurs when the wave front is highly curved compared to the resolution of the grid. Another case where D_1 operators are used is close to the boundary of the grid, when the nodes that would be required to construct the second-order operator lie outside the grid. For the remainder of this paper, a 'second-order scheme' implies that the second-order operator is used when allowed, and the first order otherwise.

The upwind scheme given by eq. (2) describes how to calculate new traveltimes using known traveltimes from adjacent grid points. Successful implementation of this scheme requires that the order in which nodes are updated be consistent with the direction of flow of information, that is, from smaller values of T to larger values of T. To achieve this, FMM systematically constructs traveltimes T in a downwind fashion from known values upwind by employing a *narrow-band* approach. The narrow-band concept is illustrated in Fig. 1(a), which for simplicity is restricted to a 2-D propagation grid rather than the full 3-D case which is harder to visualize. *Alive* points have their values correctly calculated, *close* points lie within the narrow band and have trial values, and *far* points have no values calculated. Trial values are calculated using eq. (2) with *alive* points only, which by definition lie upwind of the *close* points. The shape of the narrow band approximates the shape of the first-arrival wave front, and the idea is to propagate the band through the grid until all points become *alive*.

The narrow band is evolved by identifying the *close* point with minimum traveltime and tagging it as an *alive* point. All neighbours of this point that are not *alive* are tagged as *close* and have their traveltimes computed (or recomputed) using eq. (2). Repetition of this process allows the traveltime to all grid points to be calculated. Fig. 1(b) illustrates several evolution steps of the narrow-band method from a source point. Choosing the *close* points with minimum traveltime guarantees that *alive* points are not evaluated with partial information. In general, adjacent *close* points must lie upwind or downwind of each other; the *close* point with minimum traveltime will always lie upwind of adjacent *close* points.

2.2 Multistage FMM in layered media

Rawlinson & Sambridge (2004a,b) adapted the FMM scheme described above to allow phases other than the absolute first-arrival to be tracked in layered 2-D media, with emphasis on reflections from the layer boundaries. The two principal difficulties involved in the introduction of reflections are as follows. (i) The solution of the eikonal equation using the FMM yields the absolute first arrival time of the wave front to all points in the computational domain. To model reflected waves, the wave front has to pass through the computational domain at least twice. (ii) Layer boundaries that vary with depth generally do not conform to the regular velocity grid, and can thus not be accurately represented.

The problem of the representation of layer boundaries was solved by the introduction of a set of irregularly positioned nodes that lie exactly on the interface that is the layer boundary. A connectivity between the irregular interface nodes and the regular grid nodes was established with a local triangulation in the neighbourhood of the interface, and a special update procedure was used to calculate arrival times at the interface nodes from the nodes connected to them. This procedure yields the first arrival times at all the interface nodes.

By using a multistage FMM approach, the limitation of yielding only absolute first arrival times can be overcome for those later arriving phases that are caused by the presence of discontinuities. After the first stage of propagating the wave front to the reflecting interface, a copy of the region already traversed by the wave front is created. This copy is treated as a part of space not yet traversed by the wave front, that is, the nodes in the copy start with a node status *far*, except the ones on the reflecting interface that retain their arrival times from the first stage and start with a node status *close*, thus forming a new narrow band. In the second stage, the reflected phase is obtained by applying the FMM to the copy of the region. The narrow band at the start of the second stage 'surrounds' the interface and not the wave front, but causality is not breached because the first-arrival reflection wave front only impinges on the interface once at every point.

Although it is not strictly required by the causal nature of FMM, the same principle can also be applied to phases that refract at a layer boundary. In this case the arrival times of the wave front at the interface are calculated in a first stage just as in the reflection case, but instead of propagating the wave front back into a copy of the region from which it came, it is propagated into the region of space that lies on the other side of the interface. There are advantages to this method in terms of accuracy (avoiding very large local wave front curvature close to the interface), memory usage and program logic.

The method of reinitialization of the wave front at an interface outlined above can be repeated several times to obtain the first arrival time of a seismic phase that is characterized by a specific sequence of interactions (reflections or refractions) with layer boundaries. Fig. 2

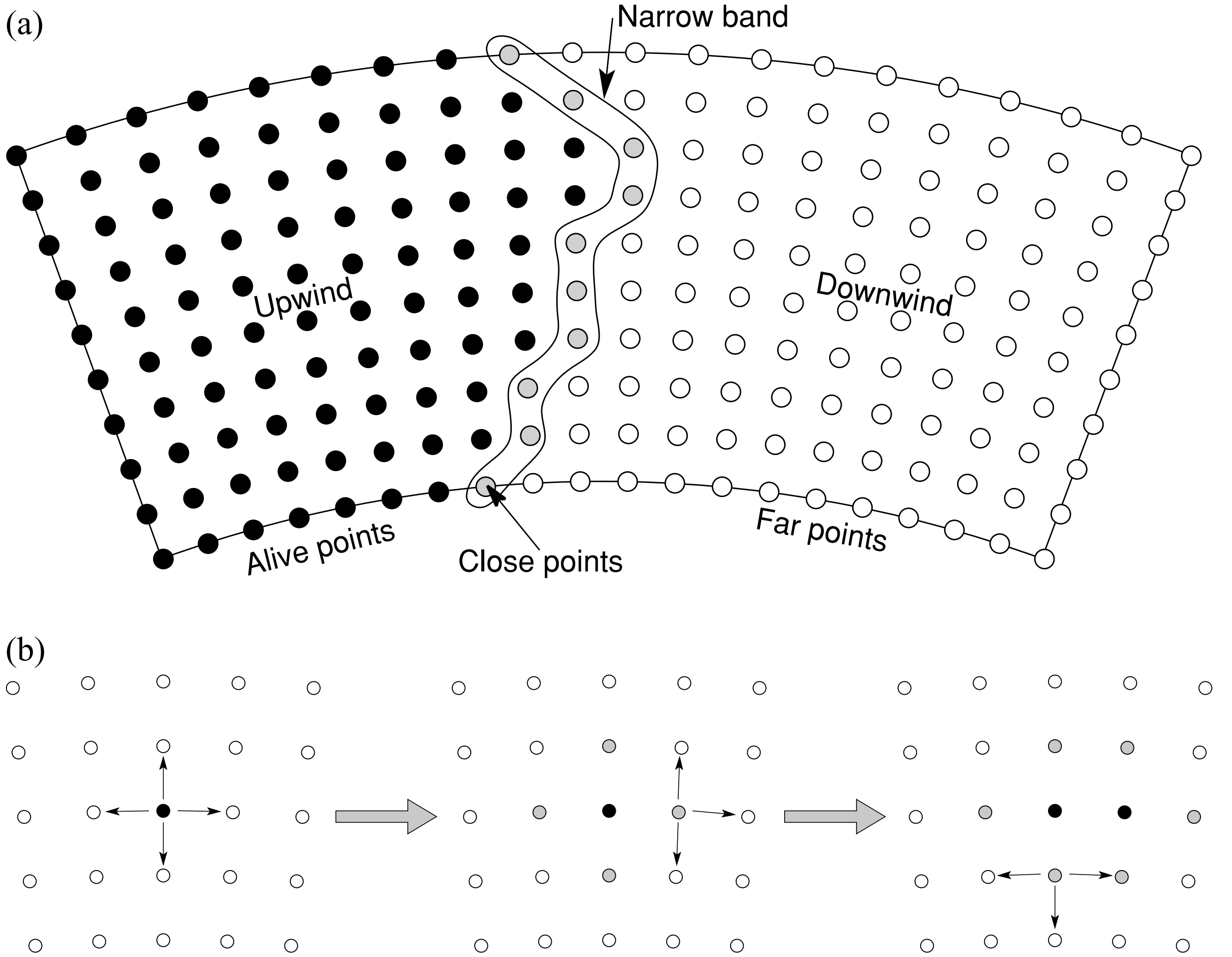

Figure 1. (a) The principle of the narrow-band method. (b) Example of how the narrow band evolves from a source point. See the text for more details.

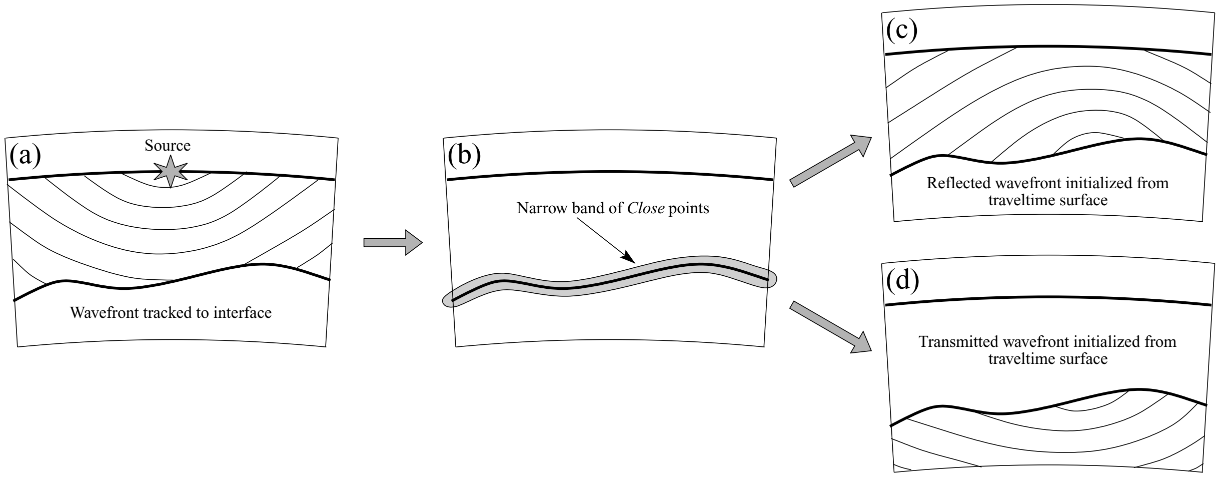
(a)
Source
Wavefront tracked to interface
(b)
Narrow band of *Close* points
(c)
Reflected wavefront initialized from traveltime surface
(d)
Transmitted wavefront initialized from traveltime surface

Figure 2. Illustration of how a reflected or refracted wave front can be tracked by partitioning the computational domain into separate regions in which the FMM is applied. (a) A wave front is generated at a point source; (b) the nodes on the interface define a narrow band; (c) the reflected wave front is propagated back into the initial region; (d) the refracted wave front is propagated into the adjacent region. See the text for more details.

illustrates the principle of this multistage scheme. Note that paths representing headwaves or diffractions will be correctly computed if they represent the first arrival of the specified phase. However, later arrivals that result from continuous velocity variation (e.g. swallow-tails) cannot be tracked.

Since multistage FMM is built around the initialization of wave fronts from interfaces, it becomes very easy to implement the modelling of teleseismic wave fronts that impinge on the bottom of a local model region and then propagate through it. Given a method of calculating the arrival time of the teleseismic front at the bottom interface, this interface is simply initialized as the narrow band and the fast marching is initiated.

In this paper, we extend the multistage FMM to 3-D spherical coordinates. In terms of methodology, the main differences with the 2-D multistage FMM are the treatment of nodes lying on the interface in terms of connectivity and update method and the addition of a facility to obtain reflection phases from a layer boundary that are not the first ones to arrive (e.g. *PP*, *pP*, *PKKP*).

3 DESCRIPTION OF THE NEW METHOD

In the following section we will describe in some detail how the multistage FMM in layered media is implemented in three dimensions and spherical coordinates.

3.1 Organization of the computational domain

Several grids (sets of nodes) are used in the computations that we will first define here.

3.1.1 The fast-marching grids

The propagation of wave fronts with the fast-marching method takes place on a set of nodes that exhibit both regular and irregular spacing. The regular nodes lie on a 3-D grid in spherical coordinates (radius r, latitude θ and longitude ϕ) that we will refer to as the *propagation grid*. This propagation grid is cut by a number of interfaces. The evaluation of the arrival time and direction of the wave front at the interface with the fast-marching method and the reinitialization of the wave front into the next region in multistage FMM requires the presence of nodes on the interface itself. Therefore, each interface is associated with a collection of nodes that lie exactly on the interface, at the positions where the connections of the regular propagation grid (i.e. lines parallel to its axes) intersect with the interface. We call this set of irregular nodes representing an interface an *intersection*.

The interfaces divide the propagation grid into several parts. We will define a *region* as the set of nodes that is the combination of all the regular propagation grid nodes that lie between two consecutive interfaces, and the irregular nodes of the intersections that constitute the top and bottom. Regions are numbered from top to bottom, with region 1 lying just below the surface, region 2 below the first subsurface boundary layer, and so on. The relation between propagation grid, interface, intersection and region is illustrated in Fig. 3, which shows a small section of propagation grid cut by an interface.

Sometimes an interface will cut through a propagation grid cell very close to one of the regular nodes. This leads to very closely spaced intersection nodes that can cause numerical difficulties. To remedy this, a small tolerance length (1/200th of the grid spacing) is introduced. If an intersection node lies closer to a regular grid node than this tolerance, it is assumed to lie exactly on it.

3.1.2 Velocity and interface grids

A continuous velocity field covering the computational domain is defined by applying tricubic B-spline interpolation to a set of prescribed velocity values on a regular grid of nodes. These nodes lie on a 3-D grid in spherical coordinates. This grid is referred to as a *velocity grid*. To obtain the velocity at an arbitrary point in space, for example, a node of the fast-marching grids, we simply sample this continuously defined velocity field.

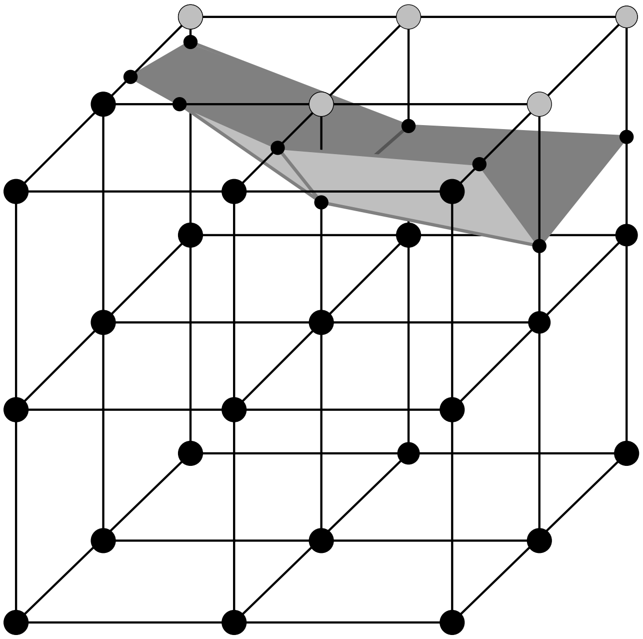

Figure 3. This figure illustrates the relation between the propagation grid, intersections and regions. The large circles are propagation grid nodes, the small circles intersection nodes. Small and large black nodes constitute the region below the interface; grey nodes lie above the interface, and together with the small black intersection nodes constitute the region above the interface.

It is also possible to specify the B-spline velocity field as a perturbation to a pre-defined spherically symmetric background model such as *ak*135 (Kennett *et al.* 1995). In this case, the velocities of the background model are transferred directly to the nodes on which FMM is applied, and afterwards the corrections obtained from the B-spline interpolation are applied. There is an advantage to this if the background model contains structure that is finer than can be resolved by the B-spline interpolation on the velocity grid, since the resolution on the fast-marching grid tends to be significantly higher than that of the velocity grid.

The depth (or radius) of the interfaces at a given location are derived by bicubic interpolation on a regular 2-D grid in latitude and longitude, the *interface grid*. This way of defining interfaces does not allow them to be multivalued in depth, and we can only consider interfaces that are defined by a single-valued function of the form $r_{\text{int}} = f(\theta, \phi)$. The interfaces are not allowed to intersect, but can be pinched together if required. They are numbered from top to bottom starting with the surface at index 1. Each region of the computational domain between two consecutive interfaces has its own velocity grid associated with it to allow for discontinuous transitions in wave speed.

3.2 Grid refinement around the source

Rawlinson & Sambridge (2004b) showed that, if the velocity distribution in the medium in which the wave propagates is not too complex, the error in the arrival time of a wave front at a distant location is dominated by the error introduced very close to the source. This is simply because close to the source the radius of curvature of the wave front is not much larger than the spacing of the grid nodes, and therefore the conditions under which FMM provides a good approximation of the arrival time are not satisfied. Rawlinson

& Sambridge (2004b) showed that by introducing a refined grid around the source, accuracy can be greatly increased without the large increase in computation time associated with high resolution over the entire grid.

This source grid refinement scheme is also implemented in the new method. A local regular grid is constructed around the source with a preset increase in resolution and size (the *source grid*). Typical values are an increase in resolution by a factor of 5, and a refined grid extending 50 nodes away from the source in each direction. If the source is not close to an interface or a propagation grid boundary, this creates a refined grid with about one million nodes.

If the source lies inside a region, it is first checked whether the top or bottom interfaces of this region cut the source grid. If one of these does cut through the source grid, an intersection is created containing nodes that lie on the interface, in exactly the same way as intersections are created on the main propagation grid. Finally, a refined source region is defined, containing the nodes of the regular refined source grid that lie inside the source region, and the intersection nodes of the top and bottom source region interfaces if these exist. The FMM is then applied to this refined source region just as it is applied to a region of the main propagation grid. In order to preserve causality (Rawlinson & Sambridge 2004b), the propagation of the wave front is stopped as soon as the wave front reaches a node that lies on the refined grid boundary but is not an intersection node or a node on the boundary of the main source region. The *alive* nodes (both regular and irregular) of the refined source region that coincide with a node in the much coarser main source region then transfer their arrival times and node status to the source region in the main propagation grid. A new narrow band is created in the main source region around the set of nodes that were initialized from the refined source region, and the FMM is restarted to complete the wave front propagation through the entire main source region.

Note that if the source lies in a region but close to an interface, the refined grid will only extend to that interface and accuracy for waves travelling in that direction may be reduced. If a source lies *exactly* on an interface, then in order to maintain accuracy refined source grids are created both in the region above and below the interface.

3.3 Three-dimensional fast marching in spherical coordinates

The FMM is easily adapted to a regular spherical coordinate system r, θ, ϕ (Sethian 1999). To do this one simply replaces the upwind gradient operators (eq. 3) with their equivalent for a spherical coordinate system:

$$\begin{aligned}
D_1^{-r} T_{i,j,k} &= \frac{T_{i,j,k} - T_{i-1,j,k}}{\delta r} \\
D_2^{-r} T_{i,j,k} &= \frac{3T_{i,j,k} - 4T_{i-1,j,k} + T_{i-2,j,k}}{2\delta r} \\
D_1^{-\theta} T_{i,j,k} &= \frac{T_{i,j,k} - T_{i,j-1,k}}{r\delta\theta} \\
D_2^{-\theta} T_{i,j,k} &= \frac{3T_{i,j,k} - 4T_{i,j-1,k} + T_{i,j-2,k}}{2r\delta\theta} \\
D_1^{-\phi} T_{i,j,k} &= \frac{T_{i,j,k} - T_{i,j,k-1}}{r\cos\theta\delta\phi} \\
D_2^{-\phi} T_{i,j,k} &= \frac{3T_{i,j,k} - 4T_{i,j,k-1} + T_{i,j,k-2}}{2r\cos\theta\delta\phi}
\end{aligned} \tag{4}$$

and similarly for the D^{+} operators. In our current implementation it is not possible to include the poles of the spherical coordinate system in the computational domain or to take periodicity into account.

3.4 Three-dimensional fast marching in the presence of irregular nodes

The standard implementation of the fast-marching method relies on the regularity of the grid to identify the neighbours of a given node. Grid regularity also allows a particularly simple finite-difference solution of the eikonal equation (i.e. eq. 4).

In the case of an irregular grid, a different method has to be used both to identify node neighbours and to update the traveltimes at each node. For 2-D problems, some of these methods are discussed in the literature (Sethian 1999; Rawlinson & Sambridge 2004a,b). Here we describe the methods we have developed for our 3-D problem.

3.4.1 Connectivity of the irregular grid nodes

The way in which we find neighbours is by considering each node to be part of a grid cell. In the regular part of the grid, each cell is a block with eight propagation grid nodes as vertices (Fig. 3). Each of these eight nodes is considered to belong to the cell. Similarly, each regular node belongs to eight cells, except on the grid boundaries.

If such a regular cell is cut by an interface, some (usually three or four) edges of the cell will intersect with the interface. Since we have constrained the intersection nodes to lie on the connections of the regular grid, the intersection associated with the interface will have a node at each of these points (the small black nodes in Fig. 3). These intersection nodes are also considered to belong to the cell. Each intersection node typically belongs to four grid cells.

In order to find the intersection nodes that belong to a given cut cell, and the cut cells a given intersection node belongs to, we create two sets of pointers that contain this information.

If we now want to identify the neighbours of a given node, which can be either a propagation grid node or an intersection node, we first use the pointers from the node to the cells it is part of to construct a list of connected cells. For this list of cells, we use the pointers from the cells to the nodes that belong to the cell to create a list of neighbouring nodes.

This way of finding neighbouring nodes fits naturally with the update procedure for FMM. On a regular 3-D grid the update proceeds by considering possible updates from each octant around the node and choosing the one that yields the lowest value for the arrival time. In our case the octants are generalized to connected cells, which may be regular or irregular. If the arrival time at a node needs to be updated, we consider every connected cell separately. If the cell contains only regular nodes, the normal regular grid update procedure can be used to calculate an update from that octant. If, however, the connected cell contains intersection nodes, the special update procedure for irregular nodes described below is used. Referring again to the example in Fig. 3, the central propagation grid node in this figure is connected to eight cells. The four cells in the bottom layer are regular, whereas the four cells in the top layer are all cut by the interface, and updates from these cells are done with the irregular node update procedure.

This method of defining the connectivity of the nodes in the irregular part of the grid is much simpler to implement than the 3-D equivalent (constrained tetrahedralization) of the explicit acute triangularization of nodes near an interface employed by Rawlinson & Sambridge (2004b).

3.4.2 Traveltime updates from irregular nodes

The update procedure for the arrival time at a given node using the known arrival times at a number of irregularly positioned neighbouring nodes is based on making the assumption that the wave front can be approximated locally as planar, and using the values of the arrival times at the neighbours to obtain the best possible estimate of the wave front normal that is certain not to underestimate the arrival time T_{u} at the updated node. The assumption of a plane wave front means that this method is of first order, and hence is generally not as accurate as the second-order schemes that are easily applied to regular grids. This does not have a significant impact on the overall accuracy of the results, however, since the irregular update method is only used in the direct vicinity of the interfaces that are described by irregular nodes. Since a plane wave approximation is much more easily applied in Cartesian coordinates, we first convert the positions of the nodes used for an irregular update to a local Cartesian coordinate system. A second-order method is used for the updates from regular connected cells.

To first order, the difference in arrival times δT between two points with relative position $\delta\mathbf{x}$ is given by

$$\delta T = \nabla T \cdot \delta\mathbf{x}. \tag{5}$$

By definition, $\nabla T \equiv s\mathbf{n}$, where $\mathbf{n}$ is the wave front normal. The arrival time update procedure for irregular cells essentially consists of solving the finite-difference form of eq. (5). When updating arrival times using information from the *alive* nodes that are part of an irregular connected cell, the number of *alive* nodes in that cell can be either 1, 2, 3 or >3. If the arrival times are known at exactly three neighbouring nodes, these times define a unique plane wave that can be calculated and then used to find the arrival time at the update node. When only one or two nodes are available, extra constraints have to be added to find a solution for the plane wave, and when more than three nodes are available there is redundant information. We consider these four cases separately.

To calculate the update to the arrival time at a node where the arrival times are known at three neighbouring nodes, say, T_1, T_2, T_3 at positions $\mathbf{x}_1$, $\mathbf{x}_2$, $\mathbf{x}_3$, the updated time is found by solving the set of equations

$$\begin{aligned} T_2 - T_1 &= s_{12}\mathbf{n}\cdot(\mathbf{x}_2 - \mathbf{x}_1)\\ T_3 - T_1 &= s_{13}\mathbf{n}\cdot(\mathbf{x}_3 - \mathbf{x}_1)\\ |\mathbf{n}| &= 1 \end{aligned} \tag{6}$$

for the wave front normal $\mathbf{n}$. Here s_{12} and s_{13} represent the average slowness between points 1 and 2 and 1 and 3, respectively. The first equation constrains the wave front to have a certain angle with respect to the line connecting points 1 and 2; the second gives a constraint on the angle between the wave front and the line connecting points 1 and 3. Once $\mathbf{n}$ is known, the updated arrival time T_{u} at $\mathbf{x}_{\mathrm{u}}$ can be found from

$$T_{\mathrm{u}} = T_1 + s_{1u}\mathbf{n}\cdot(\mathbf{x}_{\mathrm{u}} - \mathbf{x}_1). \tag{7}$$

An additional constraint has to be applied to the solution (Sethian 1999) to ensure the causality of the update method: the update must take place from a direction that is contained within the solid angle subtended by the triangle formed by the three nodes with known arrival time as seen from the node that is updated. If the vector $-\mathbf{n}$, where $\mathbf{n}$ is the solution of eq. (6), lies outside this solid angle the estimated arrival time calculated is rejected. Note that the enforcement of acute triangulation of nodes in the interface neighbourhood employed by Rawlinson & Sambridge (2004a) is a way of ensuring this causality, and obviates the need for an explicit causality test.

If only two neighbouring nodes have known arrival times, the wave front normal is not uniquely defined by the known nodes. An extra constraint on the wave front normal is added, ensuring that the updated arrival time is an upper limit, which is geometrically equivalent to the constraint that the wave front normal lies in the plane that contains the two known nodes and the node to be updated. The set of equations for $\mathbf{n}$ in this case is

$$\begin{aligned} T_2 - T_1 &= s_{12}\mathbf{n}\cdot(\mathbf{x}_2 - \mathbf{x}_1)\\ 0 &= \mathbf{n}\cdot[(\mathbf{x}_2 - \mathbf{x}_{\mathrm{u}})\times(\mathbf{x}_1 - \mathbf{x}_{\mathrm{u}})]\\ |\mathbf{n}| &= 1. \end{aligned} \tag{8}$$

Again a causality test has to be applied to the solution: in the plane that contains the two known nodes and the node to be updated, $-\mathbf{n}$ must lie within the angle subtended by the two known nodes as seen from the update point.

If only a single known node neighbours an update node, the upper limit on the arrival time at the update node is simply derived by multiplying the distance with the slowness:

$$T_{\mathrm{u}} = T_1 + s_{1u}|(\mathbf{x}_{\mathrm{u}} - \mathbf{x}_1)|. \tag{9}$$

No causality check is required when updating from a single known neighbouring node.

Because of the partial regularity of the set of nodes on which the wave front propagation is calculated, it frequently occurs that for a three-node update the three known nodes and the update node are coplanar. In this case, the three-node update procedure will fail, and a two-node update has to be applied instead.

If a connected cell contains more than three nodes with known arrival times, we simply consider all possible combinations of known nodes in single node, two-node and three-node update mode. It is not possible to consider only three-node combinations since the causality test may reject all such solutions, and the best (and only allowable) solution may be a single-node one. Potentially the number of combinations can be very large: $m + {}_2C_m + {}_3C_m$ where ${}_iC_j$ is the number of different combinations of *i* objects that can be formed out of a set of *j* objects, and *m* is the number of *alive* nodes in the connected cell. It is, however, significantly reduced to $m + {}_2C_{m-1}$ if one makes use of the causal nature of the fast-marching method, which ensures that any update from a combination of nodes that does not contain the node that most recently became alive has already been performed and does not have to be repeated.

3.5 Path sequences in multistage FMM

As outlined above, the multistage fast-marching method consists of finding fast-marching solutions to the traveltime field for a sequence of regions between interfaces. The first step in the sequence calculates the traveltime field in the region in which the source is located, or in the regions above and below it if the source is exactly located on an interface. Each consecutive step in the sequence takes the nodes and associated arrival times of either the top or bottom intersections of the region from the previous step in the sequence as the initial narrow band for the fast-marching wave front propagation into the next region. If the front is propagated back into the same region as the previous step in the sequence, it represents a reflection from the interface. If the front propagates into a region adjacent to the region of the previous step, it represents a refraction. This is illustrated schematically in Fig. 4.

A specific later arriving phase or *path* (we use these terms interchangeably) to be computed is defined as a sequence of interfaces and propagation directions. We can represent each step in such a sequence with a naming convention of the form P_{nm}, where *n* is the starting interface and *m* is the region in which the propagation takes place. For example, propagation of the wave front from interface

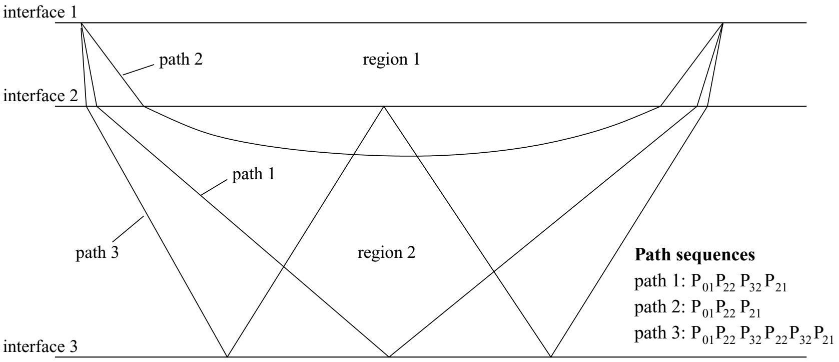
interface 1
path 2
region 1
interface 2
path 1
path 3
region 2
interface 3
Path sequences
path 1: $P_{01}P_{22}P_{32}P_{21}$
path 2: $P_{01}P_{22}P_{21}$
path 3: $P_{01}P_{22}P_{32}P_{22}P_{32}P_{21}$

Figure 4. An illustration of possible path sequences in multistage FMM. Paths 1 and 2 are relatively simple reflection and refraction phases, but sequences can be arbitrarily complex, as illustrated by the reflection multiple of path 3. See the text for more information on the path-sequence-naming convention.

3 upwards into region 2 would be represented by P_{32}. For traveltime fields that are obtained starting directly from the source, we take $n = 0$.

The first example (path 1) in Fig. 4 is a simple reflection from interface 3, with source and receiver located on the surface (interface 1). This path is represented by the sequence $P_{01}P_{22}P_{32}P_{21}$. Similarly, turning paths (path 2) can be specified by restarting the propagation from the same interface as in the previous step: $P_{01}P_{22}P_{21}$. The sequences can become arbitrarily complex, as illustrated by the third sequence (path 3) in Fig. 4, with the step sequence $P_{01}P_{22}P_{32}P_{22}\ P_{32}\ P_{21}$.

Several paths can be specified for a given source. In practice it occurs frequently that these paths differ from one another only after after a number of steps in the sequence. For example, every path will start with propagation through the region in which the source is located. In order to avoid having to repeat the calculations of the traveltime fields, we maintain a list of time fields together with a tree structure that can describe all possible path sequences (Fig. 5).

The 'branches' of this tree contain pointers to time fields corresponding to the step represented by the position of the branch in the tree. Every time field (parent) has four possible descendants (the next step in the sequence): propagation in the descendant can start from the same interface as where the parent started and go up or down, or start from the non-start interface of the parent and go up or down. When the calculation of a new path sequence starts, the corresponding path through the tree of possible sequences is followed. If the branch representing the step in the path sequence already exists, the number of the existing time field is added to the list of time fields describing the path. If a descendant is requested that does not yet exist, new time fields are created and added to the list of time fields, and corresponding branches are added to the tree representing the remainder of the sequence. Thus the final result of the FMM calculations for a given path sequence consists of a list of pointers to entries in the time field list. In the case of the three path examples in Fig. 4, it will be recognized that the first two steps of sequence 1 and 2, and the first three steps of sequence 1 and 3

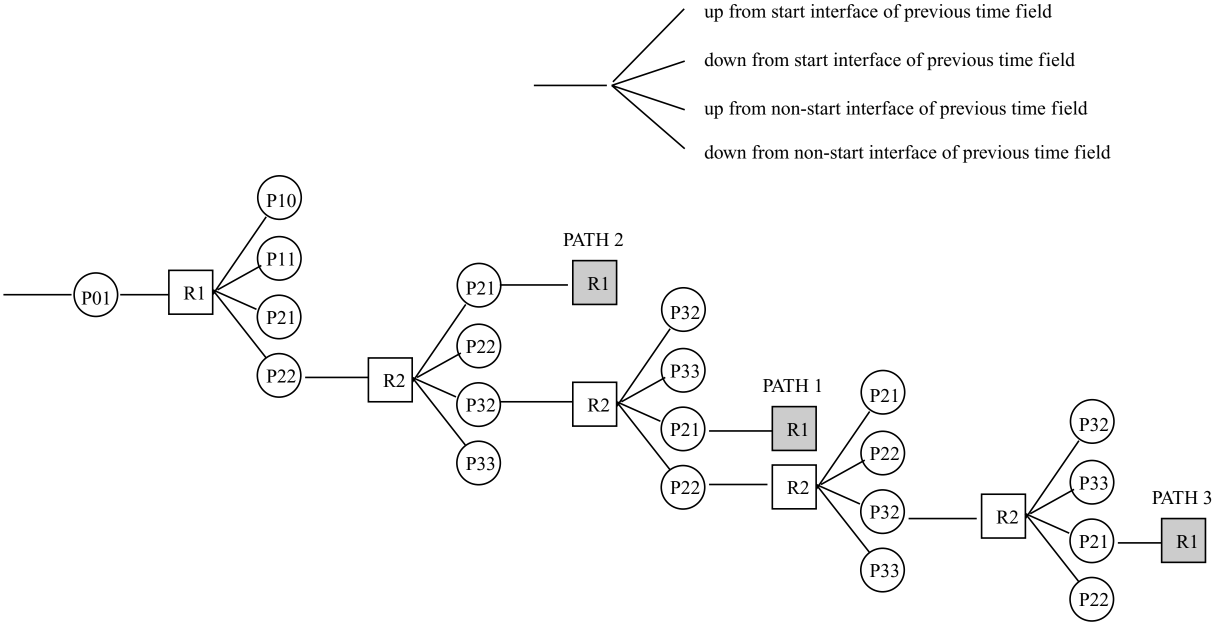
up from start interface of previous time field
down from start interface of previous time field
up from non-start interface of previous time field
down from non-start interface of previous time field
P01
R1
P10
P11
P21
P22
R2
P21
P22
P32
P33
PATH 2
R1
R2
P32
P33
P21
P22
PATH 1
R1
R2
P21
P22
P32
P33
R2
P32
P33
P21
P22
PATH 3
R1

Figure 5. A schematic representation of the tree structure in which the traveltime field sequences are stored to avoid repeating calculations for paths that have steps in common. The traveltime fields are represented by squares and labelled with the region they belong to. Each timefield can be followed by four others, as shown in the legend. The circles represent the propagation of the wave front through the next region, and are labelled according to the naming convention described in the text. The grey squares represent the final timefields of the three sequences illustrated in Fig, 4.

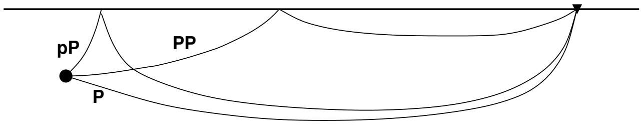

Figure 6. An illustration of late reflection phases such as *pP* and *PP* that can be obtained by using the reflection matching procedure described in Section 3.6.

are identical. This means that the second and third sequences only require the calculation of one and three regional traveltime fields respectively, instead of three and six for the full path.

3.6 Late reflections

The multistage FMM method described above can be used to evaluate the first arrival of a wave front that has been reflected from a layer boundary. Some very commonly observed seismic phases, however, correspond to reflections from boundaries that are not the first to arrive. In the following we will refer to such phases as *late reflections*.

Examples of late reflections in global seismology are *pP* and *PP* phases (Fig. 6). These can not be modelled by a straightforward application of multistage FMM: calculating the traveltime field sequence P_{01} P_{11} does not yield the *PP* arrival time since the arrival time of the wave front at the receiver obtained in the first P_{01} step (the normal *P* phase) is shorter than that of the *PP* phase, and the arrival time at the receiver will not be modified during the P_{11} step.

In order to compute traveltimes for such late reflections, we split the path sequence into two parts: one part from the source to the intermediate reflecting surface, and the other part from the reflecting surface to the receiver. The second part of the sequence is then inverted, yielding the path from the receiver to the reflection point. By the principle of reciprocity, traveltimes from the receiver to all points on the intermediate reflecting interface for this inverted path will be identical to those from all points on the interface to the receiver following the original forward path. Using (multistage) FMM we compute the arrival times and the directions of the arriving wave front at the reflecting surface for the first part of the path starting at the source, and for the inverted path starting at the receiver. We now have two sets of arrival times and arrival directions of the wave fronts at the intermediate reflecting surface. The phases (rays) that we are looking for correspond to points on the interface where the two solutions can be matched so that one is a reflection of the other, or in other words that the components of the wave front normal to the incoming fronts parallel to the interface are equal and opposite. It is easily shown that this reflection condition is equivalent to Fermat's principle of stationary time, which states that rays correspond to extrema in the total traveltime from source to receiver when this total traveltime is evaluated as a function of position on the reflecting interface. If we consider the sum of the arrival times from source to surface and receiver to surface in an earth-like velocity model as a function of position on the surface, this function is generally found to have two extrema, a local minimum and a saddle point, where the first corresponds to a *pP* phase and the second to a *PP* phase.

Thus the problem of identifying this type of path comes down to locating the extrema in the distribution of the sum of the two traveltimes from source to interface (T_s) and receiver to interface (T_r) over an interface. Locating these extrema, however, presents several difficulties. In principle all local extrema represent a reflected phase, but in practice finite sampling, small errors and total traveltime distributions that have very small gradients can create artificial local extrema. Also, the position of saddle points (e.g. *PP*) can be much more difficult to determine than minima or maxima. We tried several algorithms to locate extrema, such as fitting polynomials to local surface patches and locating the extrema of the fitting function, but none of these was found to yield robust results. In practice, we therefore do not directly seek extrema in the traveltime distribution, but rather use the gradients of the two traveltime fields (i.e. the wave front normals). The reflection points correspond to surface points where the sum of the components of these gradients parallel to the interface vanishes:

$$\left|(\nabla T_{\mathrm{s}})_{||} + (\nabla T_{\mathrm{r}})_{||}\right| = 0. \tag{10}$$

In this way, we simplify our search to one for local minima with a value very close to zero, rather than extrema in general. First, the intersection nodes that lie on the reflecting interface, which are in general irregularly distributed, are connected by means of a Delaunay triangulation in latitude and longitude. This allows us to identify neighbours of a given node. We then compare the value of $|(\nabla T_{\mathrm{s}})_{||} + (\nabla T_{\mathrm{r}})_{||}|$ at each node with that of all nodes in an extended neighbourhood of the node: all vertices of the triangles connected to the primary node, and again all vertices of triangles neighbouring the first set. If the value at the node is smaller than at all other neighbourhood nodes, it is a local minimum. If the value is furthermore close to zero, that is, less than some fraction ϵ of the norm of the total time gradient *s*, it is considered to be a reflection point.

Although the reflection matching algorithm was primarily designed to find later arriving surface reflections like *PP*, the method can also be used to identify any later arriving reflection from an internal interface. If we use the normal multistage FMM to derive the arrival time of a phase reflected from an internal interface we will only obtain the first arriving reflection. If instead we specify a reflection match at the same interface, all reflections including later arriving ones will be found. Note, however, that if the path contains more than one reflection, only one interface can be used for reflection matching.

3.7 Constructing ray paths

Once a sequence of traveltime fields describing the propagation of the wave front from source to receiver through the specified sequence of refractions and reflections has been obtained, many applications will require the construction of the actual ray paths. These ray paths are an important result of the calculation if the multistage fast-marching method is used as the forward step in a seismic tomography problem, but they can also be used to validate the solution and to analyse the character of the ray, for example, does it contain sections that are diffracted or head waves.

To find the ray paths, we simply integrate along the time field gradient starting at the receiver end of the ray, going through the entire sequence of time fields specified for the phase in reverse order. In the first step, the section of the ray from the receiver to the starting interface of the last traveltime field in the path sequence is found. During each consecutive step, integration proceeds in the $-\nabla T$ direction until the ray hits the starting interface of the traveltime field, and the position where this happens is passed on as the start of the integration into the previous region. In the first time field of the sequence, describing the source region, integration ends when the ray gets close to the source. This procedure is slightly modified when the specified path includes a search for late arrivals from an intermediate reflecting surface as described in the previous section: in that case the ray paths are determined by starting at each reflection point found, and integrating down the time field gradient both towards the source and the receiver.

To integrate along the gradient of a traveltime field, the first step is the evaluation of the gradient itself. The normal FMM only yields the arrival time at each node of the grid, and for the integration we need the gradient of this field. On internal nodes of the regular grid (i.e. nodes that have neighbours in all six principal directions), the gradient can be easily derived by finite differencing. On the boundaries of the grid, however, we encounter two problems. The first is that we can only take one-sided differences in the directions crossing the interface, which can lead to quite significant errors in the gradient at the nodes on the interface. The second is that the intersection nodes are irregularly positioned. We have therefore taken the approach of storing the traveltime gradient at each node explicitly during the fast-marching calculations. The traveltime gradient at a node is given by a vector with a direction equal to the normal to the wave front used for the final update of the arrival time at the node, and a norm that is the local slowness. This approach significantly improves the accuracy of the ray paths near the interfaces.

An important reason to construct ray paths even if only arrival times are required is that the arrival times may represent paths that are not the intended ones: for some source/receiver combinations the requested path may not exist or represent a diffracted wave or a head wave. We illustrate this with some examples. In Fig. 7, a situation is shown of a source and five receivers at a range of distances from the hypocenter. The medium contains a constant-velocity layer over a layer in which velocity increases with depth. The path specified is $P_{01}P_{22}P_{21}$. For the three middle distance receivers *R*2, *R*3 and *R*4, this path is straightforward: the ray will be able to turn around within the thickness of region 2. For *R*5, the shortest path that follows the specified sequence (staying within region 2) is one that is diffracted along the bottom interface of the layer. Similarly, the path from the source to receiver *R*5 for the reflected phase $P_{01}P_{22}P_{32}\ P_{21}$ will yield a diffracted path because *R*5 lies in the shadow zone. For *R*1 a different problem is encountered: there is no path at all that penetrates into region 2, turns, and returns to this receiver, that is, the ray is non-existent. In this case, the traveltime returned by the FMM actually corresponds to a reflection from interface 2.

It is impossible to recognize these cases in the fast-marching phase of the calculations, but it is possible to detect them during ray path construction. In our implementation of ray path construction paths that do not exist for a specific source/receiver combination are recognized as invalid, and diffracted paths and headwaves are labelled as such.

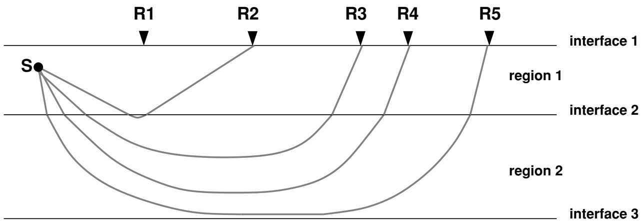

Figure 7. This figure illustrates some of the problems that can be encountered when reconstructing the ray paths from the arrival time fields calculated by FMM. If we specify a path that starts at the source, refracts through interface 2, turns in region 2, and refracts back through interface 2 to the receivers on the surface, it is found that such a path does not exist for receiver R1, and corresponds to a diffracted phase for receiver *R*5.

4 TEST RESULTS

4.1 Computational requirements

The computational resources used by the method in its current implementation depend somewhat on the complexity of the problem, but to first approximation they are linearly proportional to the number of nodes in the region on which the fast marching takes place. Virtually all CPU time is spent on the fast-marching phase of the calculations, with initialization of the grids and intersections, setting up the pointer arrays that define the nodal connectivity close to the interfaces and the ray tracing accounting for only a few per cent. CPU time use is about 22–25 s per million nodes on a 2.8-GHz Pentium 4 processor. Memory usage is about 250 MB per million nodes.

4.2 *P* wave arrivals for simple velocity models

As a first test of the method, we consider the case of a medium with a constant wave propagation speed of 8 km s^{-1}. The propagation grid covers a region of 20° in latitude and longitude and 1000 km in depth. The performance of the method is shown in Table 1 and as contour plots of the error in the arrival time on the surface in Fig. 8.

The first example (models A1–A4) is the direct arrival from a source located 100 km below the surface. Table 1 shows the average error in arrival time over the surface for a range of resolutions with and without grid refinement around the source. The second-order fast marching, in combination with the interaction of refined grid, normal grid and interfaces leads to a fairly complex pattern in the 3-D distribution of arrival time errors even for a constant wave speed.

Results are given for three models differing in resolution by a factor of 2, and one model at the coarsest resolution without grid refinement around the source. As in the 2-D cases presented by

Table 1. Summary of *P* wave arrival time errors for some simple velocity models. All cases model a region extending 20° in latitude and longitude and 1000 km in depth. The source is located 100 km below the surface.

Model[a]	Resolution	Refinement[b]	Mean error[c] (s)	CPU time (s)
A1	21 × 41 × 41	1/1	2.145	1.1
A2	21 × 41 × 41	5/10	0.511	6.4
A3	41 × 81 × 81	5/10	0.217	18
A4	81 × 161 × 161	5/10	0.095	65
B1	41 × 81 × 81	5/10	0.104	21
B2	81 × 161 × 161	5/10	0.046	101
C1	41 × 81 × 81	5/10	0.189	27
C2	81 × 161 × 161	5/10	0.061	124
D1	21 × 41 × 41	5/10	0.254	5.6
D2	41 × 81 × 81	5/10	0.148	15
D3	81 × 161 × 161	5/10	0.079	56

[a] The models are defined as follows: A—constant velocity of 8 km s^{-1}, source depth 100 km, direct arrival; B—constant velocity of 8 km s^{-1}, source depth 0 km, reflection from 1000 km depth; C—same as B, but with an extra interface at 500 km depth through which the wave front is transmitted twice; D—velocity inversely proportional to radius, surface velocity 8 km s^{-1}, source depth 0 km, direct arrival. [b] Refinement is defined as (number of refined cells per propagation grid cell)/(extent of refinement around the source expressed as the number propagation grid cells). [c] The error in the arrival time averaged over the surface.

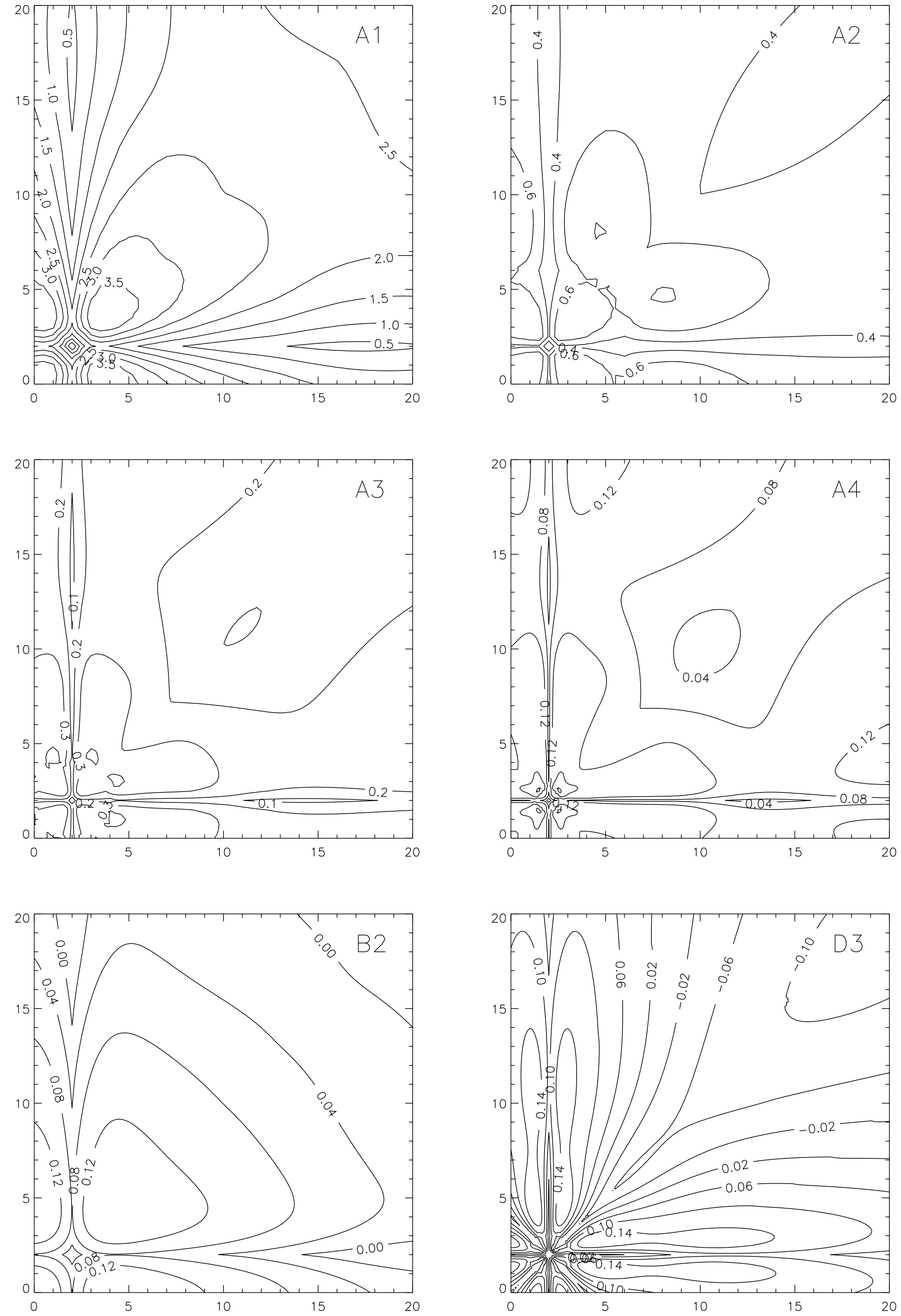

Figure 8. Contour plots of the arrival time error on the surface for models A1, A2, A3, A4, B2 and D3 from Table 1. Axes represent latitude (vertical) and longitude (horizontal), contour labels are in seconds.

Rawlinson & Sambridge (2004b), grid refinement around the source is seen to bring a large increase in accuracy, in this case a factor of 4. For a given source grid refinement, the average error is roughly proportional to the size of the grid cells, again in agreement with the 2-D results of Rawlinson & Sambridge (2004b). The quoted errors should be compared to the typical arrival time for this case of 200 s. Note that in three dimensions grid refinement around the source can be fairly costly in terms of CPU time: for all but the highest resolution model (81 × 161 × 161) the refined grid around the source contains many more nodes than the main propagation grid, which accounts for the CPU times listed for each model in Table 1.

The second example (models B1–B2) calculates the arrival time of the reflection from a layer at 1000 km depth from a source lying on the surface. This involves first propagating the wave front through the grid to the bottom interface, re-initializing the fast marching from the bottom interface and propagating back to the surface. The errors for this case are about half those in the direct arrival from a 100 km deep source. We attribute the smaller error to the fact that the wave front hits the interfaces at a much steeper angle and with less curvature. Thereby the effect of the lower-accuracy (first-order) layer directly adjacent to the interfaces is reduced.

In the third example (models C1–C2) we investigate the effect on accuracy of traversing an interface (refraction). We keep the velocity constant, but introduce an extra interface at 500 km depth. This should give the same result as the previous example, but stopping and restarting from the intermediate interface both on the way down and on the way up will obviously introduce additional errors. For the 41 × 81 × 81 grid, the error increases by 80 per cent, for the finer resolution of 81 × 161 × 161 by 30 per cent.

The next three models (D1–D3) address the slightly more complex case of velocity increasing with depth. We choose the case $v = v_0\ R_0/r$ where v_0 is the velocity at the surface and R_0 the radius of the earth. The traveltime between two points on the surface in this model is given by $t = R_0 \sin \Delta/v_0$, with Δ the angular distance between the two points. In order to use this simple analytical result, the source is positioned at the surface.

The average arrival time errors at the surface are smaller than in the constant-velocity case. Again we attribute this to the fact that a smaller fraction of each ray path lies in the first-order layer adjacent to the surface because the velocity gradient bends the ray path inward.

4.3 *P* wave traveltimes in the *ak*135 reference model

The next model we consider is considerably more complex than the previous cases for which analytical solutions are available. The velocity structure is assumed to be given by the *ak*135 reference model (Kennett *et al.* 1995). Accurate tables of arrival times in this 1-D model are available to compare with our solutions. *P* wave velocity discontinuities at 20, 35, 410 and 660 km depth are explicitly represented by interfaces as described in the previous section, and between these interfaces the velocity at the nodes is derived by linear interpolation from the *ak*135 velocity tables. In order to avoid the errors being dominated by the sampling of the velocity grids, the spacing of the velocity grid nodes was reduced from the usual case of four times the propagation grid node spacing to two times the propagation grid node spacing. The region modelled is the same size as in the previous examples, that is, 20° in latitude and 20° in longitude.

The errors in the arrival times are summarized in Table 2 and illustrated in Fig. 9. Remarkably, in spite of the presence of a velocity structure with significant velocity variations on a scale comparable to the grid spacing, the errors are comparable with the errors for the simple velocity fields considered above. Modelling *P* wave arrivals with explicit interfaces is more complicated than without, since depending on the distance between source and receiver the path of the ray that is the first *P*-arrival can differ. Labelling the interfaces with their depth, for example, I_0 is the surface, I_{410} the 410-km discontinuity, etc., the relevant interface sequences are as follows. For the closest receivers, the sequence of the path is *source*–I_{35}–I_{20}–I_0, for intermediate distance *source*–I_{410}–I_{410}–I_{35}–I_{20}–I_0 and for the longest distance in the grid *source*–I_{410}–I_{660}–I_{660}–I_{410}–I_{35}–I_{20}–I_0. In order to find the first arrival, each of these sequences is computed and the one that yields the shortest traveltime to a point on the surface is selected, which accounts for the considerably higher CPU usage in these models. This extra effort is compensated by the fact that all refraction branches are found, and not just the first arriving one. Note that the explicit use of interfaces leads to reasonable results even if some layers (in this case the crust) are thinner than the spacing of the propagation grid, and the use of a very fine (and expensive) grid spacing to resolve such thin layers can be avoided.

Model AK4 shows the arrival time errors if a continuous representation of the *ak*135 velocity model is used rather than explicit interfaces. In this case the non-smoothness and under-resolved structure of the *ak*135 velocity model greatly reduces the accuracy of the arrival times compared to the simple smooth distributions considered earlier. This demonstrates that the use of interfaces can

Table 2. Summary of *P* wave arrival time errors in the *ak*135 reference model. All cases model a region extending 20° in latitude and longitude and 1000 km in depth. The source is located 100 km below the surface. Models AK1, AK2 and AK3 represent the velocity discontinuities in the *ak*135 model explicitly with interfaces. Model AK4 uses a continuous representation of the velocity field without explicit interfaces.

Model	Resolution	Refinement	Mean error (s)	CPU time (s)
AK1	21 × 41 × 41	5/10	0.307	12
AK2	41 × 81 × 81	5/10	0.123	51
AK3	81 × 161 × 161	5/10	0.078	220
AK4	81 × 161 × 161	5/10	0.282	66

Table 3. Summary of *PP* wave arrival time errors. All cases model a region extending 20° in latitude and longitude and 1000 km in depth. The source–receiver distance is 22.416°. For the *PP* models, the source is located at the surface, for the PPAK models at 100 km depth. The *PP* models are based on the analytical velocity model with velocity inversely proportional to radius, models *PPAK1* and *PPAK2* on a continuous representation of the *ak*135 reference model, and *PPAK3* and *PPAK4* on the *ak*135 model with explicit discontinuities represented by interfaces.

Model	Resolution	Phase	Absolute error (s)	Relative error (per cent)
PP1	21 × 41 × 41	*P*	0.34	0.11
	21 × 41 × 41	*PP*	0.39	0.13
PP2	41 × 81 × 81	*P*	0.20	0.06
	41 × 81 × 81	*PP*	0.24	0.08
PP3	81 × 161 × 161	*P*	0.10	0.03
	81 × 161 × 161	*PP*	0.13	0.04
PPAK1	41 × 81 × 81	*P*	0.07	0.02
	41 × 81 × 81	*pP*	0.50	0.16
	41 × 81 × 81	*pPn*	0.32	0.10
PPAK2	81 × 161 × 161	*P*	0.29	0.10
	81 × 161 × 161	*pP*	0.42	0.14
	81 × 161 × 161	*pPn*	0.53	0.17
PPAK3	41 × 81 × 81	*P*	0.15	0.04
	41 × 81 × 81	*pP*	1.39	0.45
	41 × 81 × 81	*pPn*	0.39	0.13
PPAK4	81 × 161 × 161	*P*	0.07	0.02
	81 × 161 × 161	*pP*	0.78	0.25
	81 × 161 × 161	*pPn*	0.66	0.21

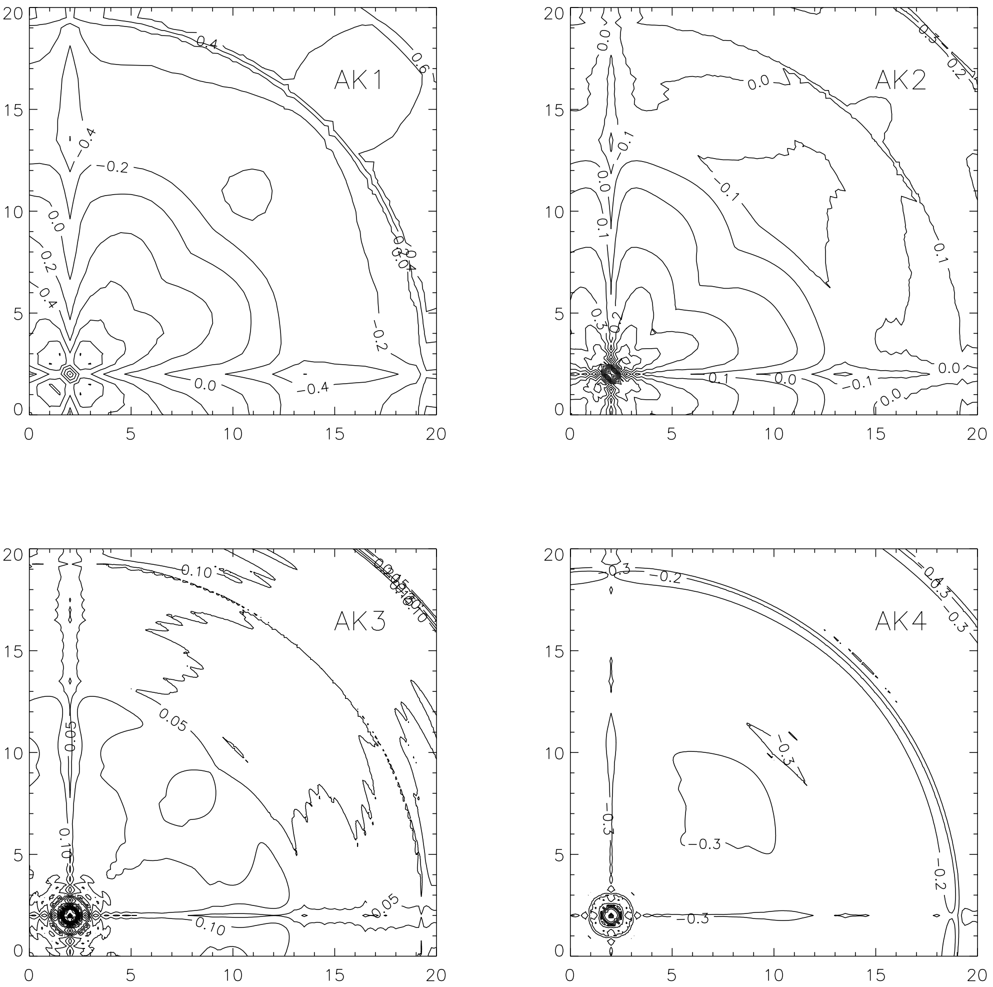

Figure 9. Contour plots of *P* wave arrival time error on the surface for models AK1–AK4 from Table 2. Axes represent latitude (vertical) and longitude (horizontal), contour labels are in seconds.

greatly reduce the error for a given propagation grid spacing compared to the case when the same velocity distribution is represented as a continuous velocity field. The reduction in error can be attributed to two effects. The first is the accurate representation of the position of the interface. The second is that because the wave front propagation by fast marching proceeds on a region-by-region basis, the wave front never passes over the abrupt changes in the velocity distribution. Propagating a wave front over such abrupt changes in velocity with fast marching can lead to the introduction of errors that are significantly larger than those for smooth velocity distributions.

The inclusion of explicit interfaces not only leads to increased accuracy for a given grid resolution, but also yields the arrival time of some later arriving phases. For example, for the longest source–receiver distances modelled here, we correctly find the arrival time of two *P* waves: one corresponding to the ray turning below the 660-km discontinuity (the first arrival), and the other in which the turning point lies between the 410- and 660-km discontinuities (the second arrival). The multistage FMM is capable of tracking any later arriving phase of the *ak*135 wavefield that explicitly arises due to the presence of an interface. In order to demonstrate this capability, refraction, reflection and diffraction traveltime curves for several of the Earth's shallow discontinuities are plotted in Fig. 10. As many as five of these phases arrive at any given angular distance. More later arrivals (generated by multiple bounces, for example) could be computed with the new method, but from an observational point of view, one has to consider whether they are likely to be detected in real data. Although *ak*135 is a spherically symmetric model, a desirable feature of the multistage FMM is that lateral heterogeneity in interface and velocity structure can easily be accommodated without compromising stability or significantly increasing computing time.

4.4 *PP* arrival times

As a test of the reflection matching algorithm used to obtain arrival times for late reflections (see Section 3.6), we computed the arrival times of *PP* type phases in the analytical velocity model with velocity inversely proportional to radius, and in the *ak*135 reference model. Since the reflection matching algorithm requires a wave front propagation from the receiver to the intermediate reflection point, arrival times are not available for the entire surface but only for a given receiver position. We therefore will not quote average arrival time errors but only specific ones for a receiver positioned on the surface at an angular distance of 22.416° from the source.

Table 3 summarizes the arrival time errors. For the simple velocity distribution with velocity inversely proportional to radius

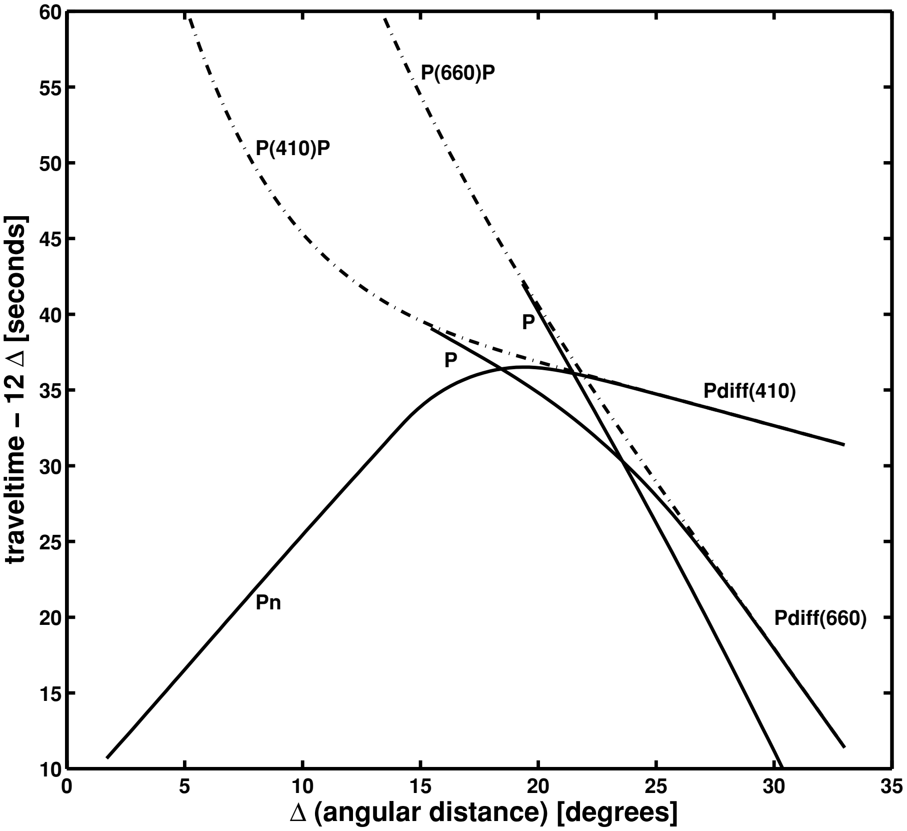

Figure 10. Traveltime curves computed by the multistage FMM for various phases that occur as a result of several of the Earth's shallow discontinuities (the Moho and 410 and 660 discontinuities), as defined by the global reference model *ak*135. Dashed lines represent reflections from the 410 and 660 discontinuities; these become diffractions at larger angular distances. The source is located on the surface of the Earth.

(*PP1–PP3*), the errors are only slightly larger than the direct *P*-arrivals, which is quite good considering the fact that it involves finding the reflection point and performing two independent wave front propagations whose errors will be superimposed.

For the models based on the *ak*135 reference model (*PPAK1–PPAK4*), the errors are quite variable, but can be understood as follows. Consider first only the errors in the direct *P*-arrivals. In the models with continuous velocity representation, the total error in arrival time is both due to a poor representation of the model (unavoidable because of the coarse grid spacing compared to the structure in the *ak*135 model near the surface), and due to errors induced by second-order FMM when the velocity distribution is very non-smooth on the scale of the grid resolution. By coincidence, these errors tend to cancel each other out (for the specific geometry considered) at a grid resolution of 25 km in the radial direction (41 × 81 × 81), leading to quite small errors in the direct *P*-arrival. That this is the case can be seen from the result of the model with twice the resolution, for which the direct *P*-arrival error is much larger instead of smaller. By contrast, in the velocity model representing the discontuities with explicit interfaces the errors converge as expected with an increase in grid resolution.

The ray paths of the *pP* and *pPn* phases are more complex and lie closer to the surface. In the continuous velocity models, this leads to an increase in error by a factor of about 2. The large errors for the *pP* and *pPn* phases in the discontinuous velocity model are due to a limitation in our current implementation of the method: the grid refinement around the source is limited to the region in which the source lies. In the layered *ak*135 model, the propagation of the wave front from the receiver to the intermediate reflection point starts at the surface, and the source region is the very thin (20 km) upper crust layer. This means that there is virtually no space for source grid refinement during the computation of the wave front propagation from the receiver to the reflection point, leading to large errors in that section of the path.

We also note here that the poor representation of the *ak*135 model when a continuous velocity distribution is used leads to the appearance of additional reflection components with traveltimes close to that of the *pPn* phase in the reflection point search.

Overall, we conclude from these tests that the algorithm used to find late reflection phases does not introduce errors that are significantly larger than those inherent in the fast-marching method.

4.5 Complex geometry

All of the examples shown above consider problems in which velocity distributions and interfaces are spherically symmetric. The reason is that for these cases we know the correct arrival times and we can perform a detailed study of the arrival time errors. As soon as the geometry becomes both radially and laterally inhomogeneous, it becomes very difficult to assess the accuracy of our results. Rawlinson & Sambridge (2004a) estimated the accuracy of their 2-D implementation of the multistage fast-marching method for media with strongly deformed interfaces separating regions with linear velocity gradients, by comparing with the results of an accurate 2-D Cartesian ray-tracing method. They concluded that although accuracy (but not stability) deteriorates as the complexity of medium

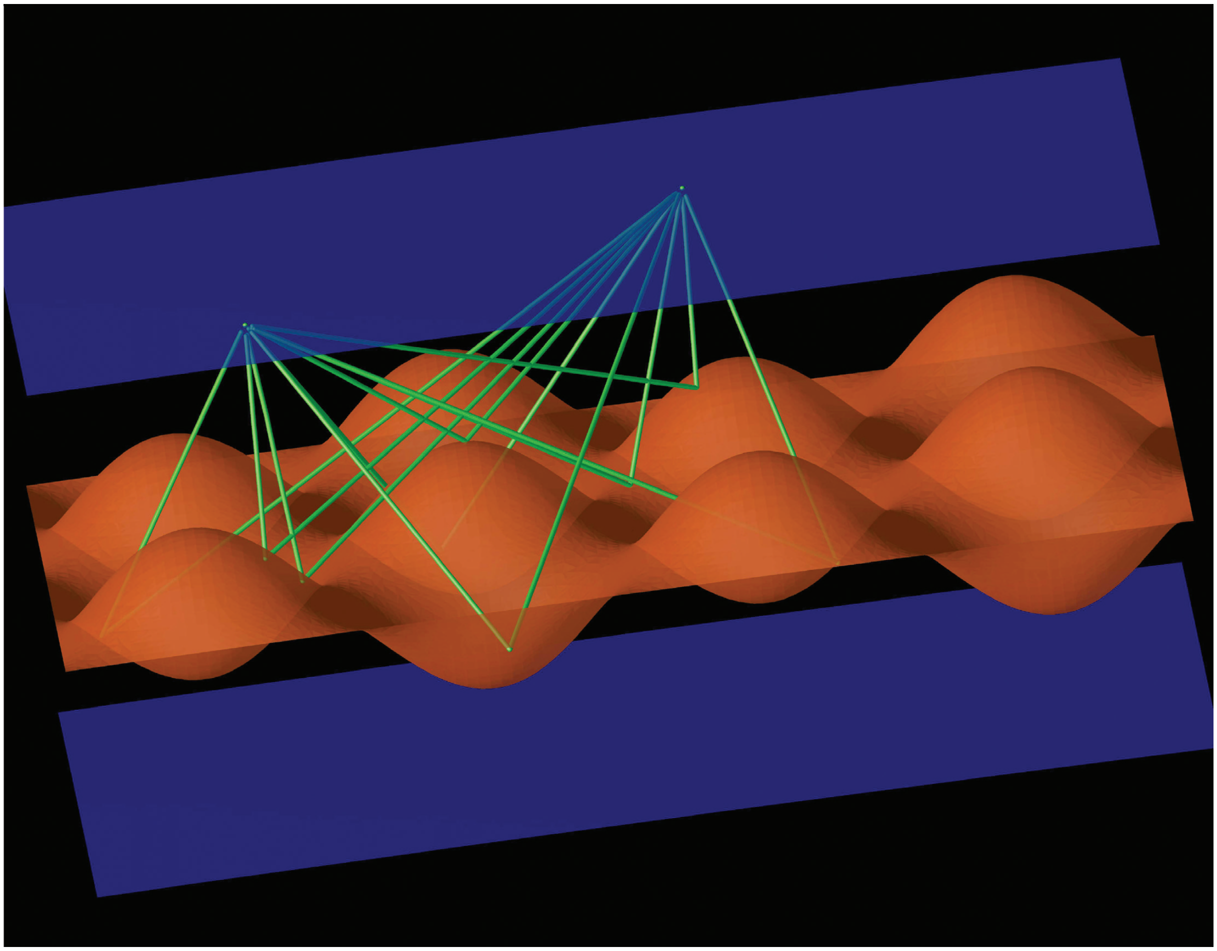

Figure 11. Ray paths representing late reflections from a highly deformed ('egg carton' shaped) interface. The reflection off the top of the interface bump roughly halfway between source and receiver is the shortest path, the other eight rays are late reflections.

increases, the errors were still comparable with those for simple cases with known solutions. This is consistent with our results, for example, when comparing constant-velocity medium direct arrivals with *P*-arrivals in the layered *ak*135 model. In the following, we examine some results for fully 3-D problems.

As a first example, we calculate the ray paths for late reflections from a highly deformed interface. The velocity is assumed to be constant and the resolution of the propagation grid is 51 × 101 × 101. In Fig. 11, a source and a receiver lie on the surface, exactly on the diagonal of the surface grid so that it is easier to understand the resulting ray paths. We are looking for phases that are reflections from the strongly deformed interface below them, whose deformations (the product of two sine functions in latitude and longitude) are also symmetric around the diagonal plane of the grid. The shortest path, which is the path found when the standard multistage fast-marching method is applied, is the reflection off the top of the interface bump roughly halfway between source and receiver. Because of the symmetry, five reflections lie exactly in the diagonal plane: two reflections off the bump below the source, one off the bump between source and receiver, and two off the bump just below the receiver. In addition, four ray paths are found that lie outside the diagonal plane, and reflect off the sides of the interface deformations. The late reflections off the bump below the receiver and from the sides are likely to have higher amplitudes than the first arrival.

In the final example, we apply our method to a complex geometry representing a more realistic geophysical setting. We consider a local problem (with physical grid size 50 km depth and 1° ≈ 110 km in latitude and longitude) at a resolution of 51 × 101 × 101. A 3-D representation is given in Fig. 12. The model consists of a coast line (ocean and land), a sediment layer under the ocean, continental crust, oceanic crust subducting under the continent, and part of the upper mantle. Note that the water and sediment layers pinch out at the coast line and disappear under the land area on the right side of the image. The continental crust and the slab contain linear velocity gradients, and the continental crust has random velocity variations between +20 per cent and −20 per cent added to each node of the velocity grid (spaced 4 kilometers apart). The tricubic B-spline interpolation used to evaluate the velocity field has a support of 64 nodes so that significant smoothing occurs. The resulting distribution of relative velocity perturbations on the nodes of the region representing the continental crust has an average of 3.6 per cent, a maximum of 15 per cent, and 98 per cent of the nodes have a velocity perturbation smaller than 10 per cent.

Two of the seismic sources are on the sea surface representing survey experiments, the other two are located in the subducting slab representing local earthquakes. An array of eight receivers is distributed over the land area. For the ocean surface sources, we calculate traveltimes and construct the ray paths for phases that reflect from the top and the bottom of the subducting slab. For the

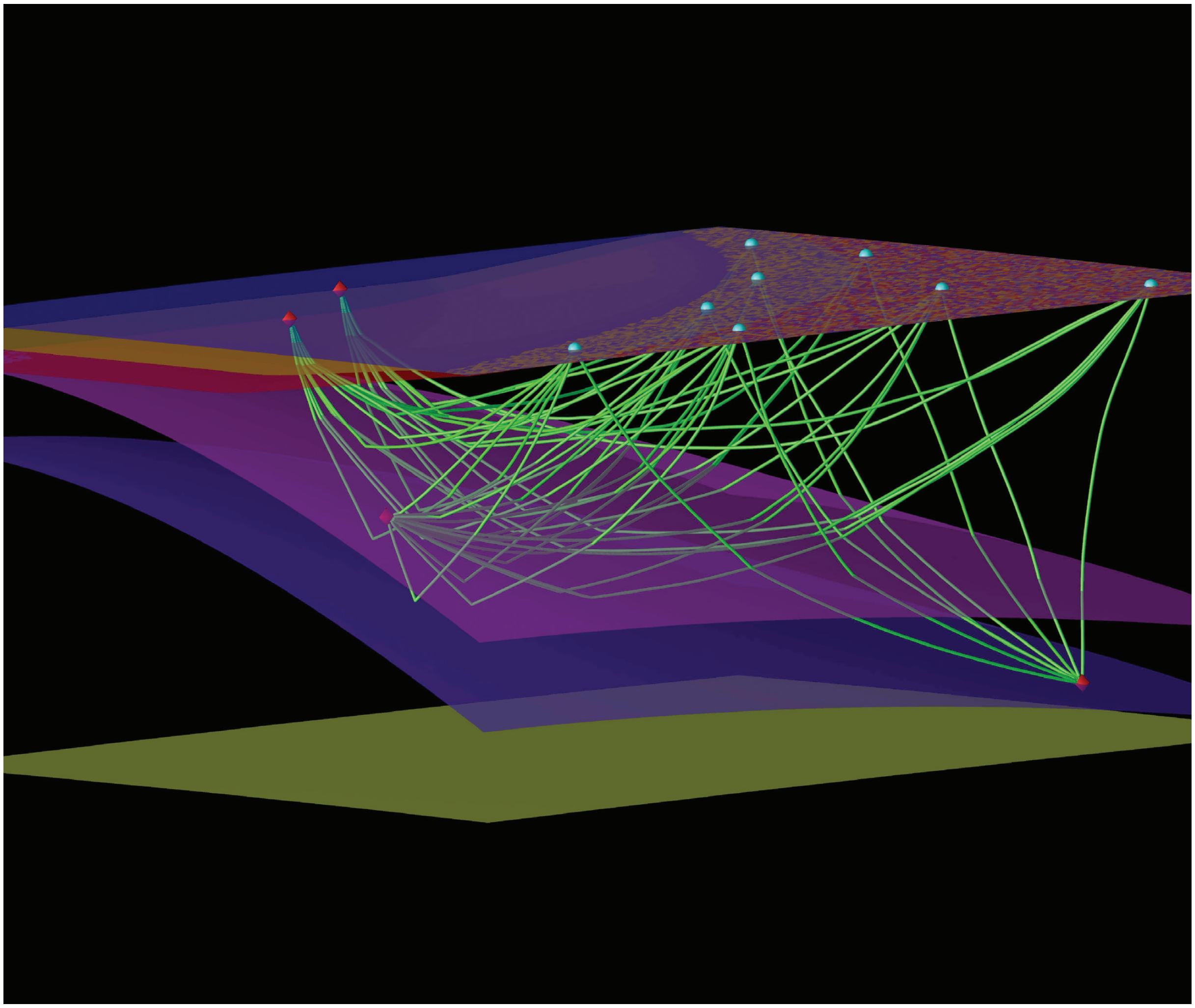

Figure 12. Ray paths in a complex layered geometry. The model consists of a coast line, a sediment layer beneath the ocean, continental crust, oceanic crust subducting under the continent, and part of the upper mantle. Seismic sources are represented by red diamonds, receivers by blue spheres. Two of the sources are on the ocean surface, another two are located in the subducting slab. An array of eight receivers is distributed over the land area. Velocity is laterally inhomogeneous in the continental crust. A selection of ray paths representing direct arrivals and reflections from the top and the bottom of the subducting slab is shown.

earthquake occurring near the top of the slab, we show the direct arrival and the reflection from the bottom of the slab, and for the earthquake near the bottom of the slab only the direct arrival. This problem requires the calculation of 38 regional traveltime fields. The computation took 225 s to complete.

If one compares the arrival times for different realizations of the random velocity fluctuations, the differences in arrival time are typically 1–2 per cent, which is at least an order of magnitude larger than the typical errors found for multistage FMM at comparable grid resolution in the previous section. This means that the method is sufficiently sensitive to detect lateral velocity variations like those used in this example.

5 CONCLUSIONS

In this paper, we have presented a new implementation of the multistage fast-marching method for calculating the arrival times and ray paths of multiple phases in complex layered media originally proposed in Rawlinson & Sambridge (2004a,b). It extends the original method to three dimensions, works in spherical coordinates and includes the possibility of finding late reflections. The accuracy and robustness of the results and the required CPU time per node are comparable to those of the earlier 2-D Cartesian implementation.

Compared to 3-D traveltime and ray path determination using shooting or bending methods, the multistage fast-marching method offers the advantage of stability: a result is always found and no iteration is necessary. Shooting and bending methods can have significant problems with convergence if the medium is complex. The method can also be computationally efficient compared to ray tracing, especially for cases in which there are few sources and many receivers, since the computation time scales only with the number of sources rather than with the product of the number of sources and receivers. After the traveltime fields have been obtained for each source, the additional cost of finding ray paths for each receiver is relatively insignificant. The same applies to cases with a few receivers and many sources, since the role of source and receiver can simply be interchanged in the calculations.

A disadvantage of eikonal-based methods, especially in 3-D where the resolution of the computational grids can not be very high, is that the accuracy of the results is generally not as good as can be obtained with a properly converged iterative ray-tracing method. The accuracy that can be easily obtained with current computational resources (i.e. using grids with millions to tens of millions of nodes) is, however, sufficient for many applications, including a range of 3-D tomographic imaging problems, for example, teleseismic tomography, wide-angle reflection/refraction tomography, local earthquake tomography and regional (continental scale) tomography. Because of the flexible parametrisation of the velocity fields and the interface positions in terms of nodal values on a grid and spline interpolation, it is in principle straightforward to treat the velocities in each region and the position of each interface separately as either fixed or parameters to be inverted for. In the case of teleseismic tomography, one could include interfaces that represent crustal structure that is constrained by other data; explicit representation of a dipping slab in local earthquake tomography may be advantageous. The new scheme offers the possibility of combining different classes of data sets (e.g. local earthquake, refraction and teleseismic). The ability to include potentially complex interface and velocity structure, yet retain computational stability and efficiency makes the new scheme practical for large seismic imaging problems.

We have shown that the multistage fast-marching method can provide arrival times with relative errors between 0.1 and 0.01 per cent for a range of synthetic examples, and that it performs robustly for very heterogeneous media approximating a realistic geophysical setting. Ultimately, the practicability of the method will be established by its application to real problems. The source code of our implementation is available for general use through a website at http://rses.anu.edu.au/seismology/fmmcode/.

ACKNOWLEDGMENTS

This work was supported by Australian Research Council Discovery Project DP0451133.

REFERENCES

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

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

Bevc, D., 1997. Imaging complex structures with semirecursive Kirchhoff migration, *Geophysics,* **62,** 577–588.

Buske, S., 1999. 3-D prestack Kirchhoff migration of the ISO89-3D data set, *Pure appl. Geophys.,* **156,** 157–171.

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

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

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

Chopp, D.L., 2001. Some improvements of the fast marching method, *SIAM J. Sci. Comput.,* **23,** 230–244.

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.

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

Fomel, S. & Sethian, J.A., 2002. Fast-phase space computation of multiple arrivals, *Proc. Nat. Acad. Sci.,* **99,** 7329–7334.

Gray, S.H. & May, W.P., 1994. Kirchhoff migration using eikonal equation traveltimes, *Geophysics,* **59,** 810–817.

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.

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

Kennett, B.L.N., Engdahl, E.R. & Buland, R., 1995. Constraints on seismic velocities in the earth from traveltimes, *Geophys. J. Int.,* **122,** 108–124.

Kim, S. & Cook, R., 1999. 3-D traveltime computation using second-order ENO scheme, *Geophysics,* **64,** 1867–1876.

Koketsu, K. & Sekine, S., 1998. Pseudo-bending method for three-dimensional seismic ray tracing in a spherical earth with discontinuities, *Geophys. J. Int.,* **132,** 339–346.

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.

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

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.

Osher, S., Cheng, L.-T., Kang, M., Shim, H. & Tsai, Y.-H., 2002. Geometric optics in a phase-space-based level set and Eulerian framework, *J. Comp. Phys.,* **179,** 622–648.

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.

Qian, J. & Symes, W.W., 2002. An adaptive finite-difference method for traveltimes and amplitudes, *Geophysics,* **67,** 167–176.

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. & Sambridge, M., 2004a. Wavefront evolution in strongly heterogeneous layered media using the Fast Marching Method, *Geophys. J. Int.,* **156,** 631–647.

Rawlinson, N. & Sambridge, M., 2004b. Multiple reflection and transmission phases in complex layered media using a multistage fast marching method, *Geophysics,* **69,** 1338–1350.

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

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.

Sethian, J.A., 1996. A fast marching level set method for monotonically advancing fronts, *Proc. Nat. Acad. Sci.,* **93,** 1591–1595.

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.

Symes, W.W. & Qian, J., 2003. A slowness matching eulerian method for multivalued solutions of eikonal equaions, *SIAM J. Sci. Comput.,* **19,** 501–526.

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

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

Xu, S. & Lambaré, G., 2004. Fast migration/inversion with multivalued rayfields: Part 1—Method, validation test, and application in 2-D to Marmousi, *Geophysics,* **69,** 1311–1319.

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, 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.

Zhao, P., Uren, N.F., Wenzel, F., Hatherly, P.J. & McDonald, J.A., 1998. Kirchhoff diffraction mapping in media with large velocity contrasts, *Geophysics,* **63,** 2072–2081.