ABSTRACT

Nonuniqueness in traveltime tomographic velocity analysis is largely the result of insufficiency in model parameterization. Traditional tomography parameterizes the model with nonoverlapping cells. Such single-scale tomography (SST) from inverting traveltimes only may produce underdeterminacy at places of insufficient ray coverage. To cope with the poor and uneven ray coverage, a multiscale tomography (MST) method is devised, which is a simultaneous application of many overlapping SSTs with different cell sizes. The MST decomposes velocity anomalies into components in a set of submodels of different cell sizes. At each model location the cells containing higher consistency between data contributions gain more from the inversion. The final MST model is a postinversion superposition of all submodel solutions. The MST method is applicable to the determination of interval velocity and interface geometry using turning rays or reflection rays. In comparison with the SST, superior results of the MST are seen in synthetic examples. The MST method is used to construct the 3D crustal velocities of southern California using the first arrivals from local earthquakes. While the SST and MST models may fit the data equally well, the MST model is smoother and geologically more plausible than the SST model.

INTRODUCTION

Seismic traveltime tomography has become an effective way for determining wave velocities in the presence of lateral velocity variations; however, serious limitations and resultant nonuniqueness exist in tomographic velocity analysis. Some of the limitations, such as noise in traveltime picking and errors in positioning the sources and receivers, can be improved by acquiring higher quality data. However, for given survey geometry and model grid size, it is usually impossible to improve the ray coverage, which is a combination of ray hit count and angular coverage in a given model. The consequences of poor ray coverage include limited resolution and increase of nonuniqueness of the solution model. For example, crosswell transmission tomography usually has poor horizontal resolution (e.g., Michelena, 1993; Zhou et al., 1993). In reflection seismology, angle-limited ray coverage results in the velocity and depth ambiguity (Bickel, 1990; Lines, 1993; Tieman, 1994) because of the competing effects of wave velocities and reflector depth on traveltimes.

This paper illustrates a multiscale tomography (MST) method that is introduced (Zhou, 1996) to reduce the nonuniqueness in tomography. Overcoming the nonuniqueness in solving ill-posed inverse problems is a central goal for research in geophysical inverse. An example is the classic work by Backus and Gilbert (1970) to support their inversion by optimization of the resolution matrices. More recent efforts to reduce nonuniqueness include applying continuum inverse theory and validating the inverse problem for an infinite number of unknowns (Delprat-Jannaud and Lailly, 1993) and using certain model parameterization (e.g., Vesnaver and Böhme, 2000). In the following, the idea of the methodology of the MST is developed and illustrated with several synthetic examples using turning rays or reflection rays. A field data example of the MST is given for determining the 3D crustal structure of southern California.

SINGLE-SCALE AND MULTISCALE TOMOGRAPHY

Single-scale tomography

Seismic traveltime tomography uses traveltimes of various seismic arrivals to constrain velocity variations in the portion of the subsurface that is traversed by the seismic raypaths. Generally, an observed data vector \( \mathbf{d} \) is a discretely sampled function of some specific model vector \( \mathbf{m} \):

\[
\mathbf{d} = \mathbf{f}(\mathbf{m}).
\]

The \( i \)th data sample is \( d_i = f_i(\mathbf{m}) \), where \( f_i(\cdot) \) stands for one sampled function, such as that using the \( i \)th shot and receiver pair. Using a reference model \( \mathbf{m}^{ref} \), equation (1) can be
\[ \delta \mathbf{d} = C \delta \mathbf{m}, \]  

(2)

where \( \delta \mathbf{d} = \mathbf{d} - f(\mathbf{m}^{\text{ref}}) \) is the traveltime residual vector, \( \delta \mathbf{m} = \mathbf{m} - \mathbf{m}^{\text{ref}} \) is the model perturbation vector, and \( C \) is the differential kernel matrix whose \( i \)th row and \( j \)th column element is \( c_{ij} = \partial f_j / \partial m_i \). Linearized tomography is an inversion of equation (2) using given data residuals \( \delta \mathbf{d} \) and kernel matrix \( C \) of a reference model to solve for the model perturbation corrections \( \delta \mathbf{m} \) at places of sufficient ray coverage.

The success of tomography depends on how the model is parameterized. In conventional single-scale tomography (SST), the model space is divided into some nonoverlapping cells (Nolet, 1987). One can then define a corresponding differential kernel and solve for the model values over these cells. Taking \( \delta s(x) \) as the slowness perturbation at spatial location \( x \), the traveltime residual of the \( i \)th ray \( \delta t_i \) is an integration of the slowness perturbation along this raypath:

\[ \delta t_i = \int_{\text{ray}_{\text{ray}}} d \mathbf{x} \delta s(\mathbf{x}). \]  

(3a)

Discretizing the model with a finite number of model cells, the conventional SST attempts to invert the following equation:

\[ \delta t_i = \sum_j \ell_{ij} \delta j, \]  

(3b)

where \( J \) is the total number of slowness cells, \( \delta s_j \) is the slowness perturbation of the \( j \)th cell, and \( \ell_{ij} \) is the length of the \( i \)th ray in the \( j \)th cell. Equation (3b) is a specific example of equation (2).

In real applications, the raypath coverage, represented by \( \{\ell_{ij}\} \) here, depends not only on the positions of shots and receivers but also on the velocity gradients of the reference velocity model. Even with regularly spaced shots and receivers, unevenness in ray coverage will coexist with velocity heterogeneities. Though SST is widely used in tomographic applications, its single model cell size can hardly fit the geometry of all anomalies.

Let us examine this problem with a simple example of decomposing a 1D velocity profile into interval velocities. Here, a true interval velocity profile is shown in Figure 1a. In an ideal case, if we know the thickness of all model layers such as that shown in Figure 1b, we would have a good chance to represent the velocity profile uniquely with a minimum number of model variables (five interval velocities in this case). In real situations without knowledge of the layer thickness, we somewhat arbitrarily choose the geometry of the layers, such as the SST model shown in Figure 1c with 17 equal-thickness layers. Such an SST model may have too many model variables at some parts of the model but still too few at other parts of the model. The fourth interface, which is dashed in Figure 1c, for example, requires finer model cell size. This example illustrates the most common situation of tomography with a mixture of under- and overdeterminacy.

Multiscale tomography

The multiscale tomography (MST) method is devised to cope with the poor determinacy resulting from unevenness in ray coverage (Zhou, 1996). The MST consists of three steps. First, it defines a set of submodels with different cell sizes; each submodel covers the whole model region. The submodel with the smallest cell size may be denoted as the first-order submodel; a higher order submodel consists of cells of greater size. For the 1D traveltime problem, Figure 2 shows five MST submodels. Each has layers that are multiples of the layers of the SST model. Hence, the first-order submodel is the same as the SST model shown in Figure 1c, and the fifth-order submodel has only one variable, the average value of the whole model region. The submodels allow a decomposition of velocity anomalies into components of different cell sizes.

As the second step, the MST determines all submodel values simultaneously rather than progressively. Progressive multiscale inversion has been proposed in waveform studies (e.g., Bunks et al., 1995) and data integration (e.g., Yoon et al., 2001). Many previous tomographic studies have also attempted with multiple sets of cell sizes progressively (e.g., Fukao et al., 1992; Bijwaard et al., 1998). In contrast, a simultaneous determination of the parameters of all submodels in the MST minimizes the spread of model values between different model cells. Each model location belongs to all submodels, while at the same location the consistency among the data contributions,
such as the apparent slowness of different rays, may vary over different submodels. A smaller size cell would make less impact on the ray traveltimes because of its limited raypath coverage but would be easier to fit the data with the local velocity variation. A large-size cell would potentially have a greater impact on the ray traveltimes but would be harder to fit the data with the velocity variation. At each location, the submodel with a higher level of consistency among traveltimes of all traversing rays will gain more from the inversion. Figures 2b–2f illustrate possible solutions of the MST.

As the third and final step, the final MST model is a superposition of all submodel values. In other words, at spatial location \( \mathbf{x} \), the final model perturbation vector \( \delta \mathbf{m}(\mathbf{x}) \) is a superposition of all submodel perturbation vectors at that location:

\[
\delta \mathbf{m}(\mathbf{x}) = \sum_{k=1}^{K} w^{(k)} \delta \mathbf{m}^{(k)}(\mathbf{x}),
\]

where \( K \) is the total number of submodels, \( \delta \mathbf{m}^{(k)}(\mathbf{x}) \) denotes the value of the \( k \)th order submodel at location \( \mathbf{x} \), and \( w^{(k)} \) is a weighting factor that can be defined prior to the inversion to modulate the contributions of different submodels. Typically the weighting factors are normalized as

\[
\sum_{k=1}^{K} w^{(k)} = 1.
\]

The default value of \( w^{(k)} \) is 1/\( K \), as used in all examples in this paper. Combining equations (2) and (4), we obtain the general forward equation for the MST:

\[
\delta \mathbf{d} = \sum_{k=1}^{K} w^{(k)} \mathbf{C}^{(k)} \delta \mathbf{m}^{(k)}(\mathbf{x}),
\]

where \( \mathbf{C}^{(k)} \) is the kernel matrix of the \( k \)th submodel. The inversion variables for the MST in equation (6) are the values of all submodels \( \delta \mathbf{m}^{(k)}(\mathbf{x}) \), which is usually discretized into values on a set of cells. For example, following equation (4), the final slowness perturbation vector \( \delta \mathbf{s}(\mathbf{x}) \) may be taken as a superposition of all submodel slowness vectors \( \delta \mathbf{s}^{(k)}(\mathbf{x}) \) at the same location:

\[
\delta \mathbf{s}(\mathbf{x}) = \sum_{k=1}^{K} w^{(k)} \delta \mathbf{s}^{(k)}(\mathbf{x}).
\]

Hence, the forward multiscale equation for the \( i \)th ray is

\[
\delta t_i = \sum_{k=1}^{K} w^{(k)} \int_{\text{ray}_i} d\mathbf{x} \delta \mathbf{s}^{(k)}(\mathbf{x}).
\]

Just like equation (3), a discretization of the model with the submodel cells means replacing equation (8a) by

\[
\delta t_i = \sum_{k=1}^{K} \sum_{j} w^{(k)} \ell_{ij} \delta s_{ij}^{(k)},
\]

where \( J_i \) is the number of model parameters in the \( k \)th-order submodel, \( \delta s_{ij}^{(k)} \) is the slowness perturbation of the \( j \)th cell of the \( k \)th-order submodel, and \( \ell_{ij}^{(k)} \) is the length of the \( i \)th ray in the \( j \)th cell of the \( k \)th-order submodel. After inverting system (8b) for \( \delta s_{ij}^{(k)} \), the final MST solution is a superposition of solutions of all multiscale submodels according to equation (7). Because of this superposition, the number of model variables and cell size of the final solution of the MST are identical to the SST with the dimension of the smallest MST submodel.

In summary, the MST may be regarded as a simultaneous application of several overlapping SSTS using different cell sizes. The use of overlapping submodels in the MST is similar to wavelet decomposition, since the model values are defined locally in the model space. However, there is no orthogonality among the multiscale submodels. Because of the unevenness in raypath coverage in traveltime tomography, at each model location the multiscale submodels allow the capture of velocity inhomogeneities of the wavelengths sampled by the traversing rays at the location. For the rare situation when raypath coverage is perfect, the MST should give the same solution as the SST. Because the inversion kernels for the MST can be approximated using the kernels of the SST, the computation overhead of the MST is only marginally higher than that of the SST.

\textbf{Partition of the MST submodels}

It is convenient to partition the submodels of the MST with cell sizes that are multiples of the SST cells. Hence, the differential kernels of the MST can be approximated using the kernel of the SST. Taking a 2D example, suppose that the cell size of the SST model is 1 \times 1; then the first-order MST submodel may use exactly the SST model dimension, the second-order MST submodel may have cell size 2 \times 2, the third-order submodel may have cell size 3 \times 3, and so on. With such a partition of the MST submodels in two-dimensions, if the SST has \( N^2 \) cells, the total number of MST cells would be \( N^2 + (N/2)^2 + (N/3)^2 + (N/4)^2 + \ldots + 1 < N^2 \times 65\% \). This means that the increase of the number of inversion variables from the SST to such a 2D MST partition is less than 65\%. For a 3D case of similar model partition, the increase in the number of inversion variables from SST to MST is only 20.2\%. In other words, the increase of model variables from the SST to MST is very moderate for both 2D and 3D applications.

Just like the conventional tomography, there are infinite ways to partition the MST model. Another way, for instance, is to partition the submodels using cell sizes that are prime-number multiples of the cell size of the SST. In other words, if the SST uses square cells of 1 \times 1 area size, the lengths of the sides of MST cells will be 1, 2, 3, 5, 7, 11, 13, 17, and so on, and the largest MST cell will be the whole model space.

Nevertheless, the most important consideration in partitioning the MST submodels is to have cell geometry as close as possible to the geometry of the expected anomalies. This point is the main reason for the MST. For instance, in many real applications the depth variation of the velocity may be significantly greater than the lateral variation. Hence, in a 3D case it makes more sense to have the 2D MST model partition for each layer rather than for the whole 3D model. This is the case for the southern California example shown later. Also in the 2D reflection tomography example shown later, because we try to determine the geometry of each model interface and there is no connection between any two interfaces, it would be reasonable to have MST model partition for each layer rather than the whole model.
A SYNTHETIC EXAMPLE OF CROSSWELL TOMOGRAPHY

In the synthetic and field examples shown in the following, the inverse operator is the LSQR algorithm (Paige and Saunders, 1982). The upper-left panel in Figure 3 shows a 2D synthetic velocity model with a crosswell shot and receiver configuration. The model is created by taking a smoothed middle portion of the Marmousi velocity model (Versteeg and Grau, 1991). In this case 43 shots in the left well and 43 receivers in the right well were used to survey the $1 \times 3$-km² model area, giving a total of 1849 raypaths. The depth interval is 70 m for shots and receivers.

Crosswell transmission tomography has poor horizontal resolution because of limited angular ray coverage (Michelena, 1993; Zhou et al., 1993). Nevertheless, the purpose of this test is to compare the SST and MST solutions under the same conditions. The cell size of the SST model here is $40 \times 40$ m², giving a total of $25 \times 75 = 1875$ inversion variables for the SST. Figure 3 also shows the geometry of the MST submodels. The cell sizes of the first nine MST submodels are prime-number multiples of the cell size of the SST model, starting from the SST model as the first-order MST submodel. The tenth, which is also the last MST submodel, has the whole $1000 \times 3000$ m² model area as one cell. Combining the cells of all submodel cells, there are 2765 inversion variables for the MST model. Depending on the ray coverage, the actual number of inversion variables might be less. After the postinversion superposition of all MST submodels, the final MST solution has the same 1875 cells as the SST solution.

Figure 4 shows raypaths in the true velocity model and the solutions of the SST and MST methods. The paths of the first-arrival rays tend to stay along fast velocities and avoid slow velocities. This is the main cause of nonlinearity in traveltime tomography. Starting from the same constant-velocity initial reference model shown in Figure 4c and using the same data, both SST and MST were carried out by iterating the processes of ray tracing in the reference model, inverting for slowness perturbations, and updating the reference velocity model. The solutions shown in Figures 4d and 4e were after four iterations of the SST and the MST, respectively. These solutions are chosen because they achieved a similar level of data fitting. The mean and standard deviation of traveltime residuals are 0.1 and 1.3 ms for the MST solution and 0.4 and 1.5 ms for the SST solution. The similar level of data fitting for these two solutions manifests the nonuniqueness in tomographic inversion. Nevertheless, the MST solution is much smoother than the SST solution. By the principle of parsimony, the simpler (or smoother here) among the two equally fitting solutions is the best solution. Compared with the true solution in this synthetic example, the MST solution outperforms the SST solution.

Figure 5 is the anatomy of the contributions from the MST submodels. Because the inversion is done to recover the slowness perturbations, the panels in this figure are plotted as

![Diagram](image-url)
a percentage ratio of the slowness perturbations over the mean velocity of 2.7 km/s. The final MST model shown in the upper-left panel is a spatial summation of the 10 submodels shown. Clearly, the submodels of different wavelengths all contributed significantly to the final model. Interestingly, patterns of the $x$-shaped stretches in the first-order submodel in Figure 5 are similar to those of the SST model (Figure 4e). These stretches are along-raypath smearing artifacts from a lack of crossing rays for small cell sizes; this is a major reason for using the MST approach.

**FIG. 4.** (a) True velocity model. (b) Raypaths in the true model from shots (purple stars) to receivers (green triangles). (c) Initial reference model. (d) Fourth-iteration MST solution. (e) Fourth-iteration SST solution. Average and standard deviation of traveltime residuals (in milliseconds) are indicated on top of panels (c)—(e). See caption of Figure 3 for scale.

**FIG. 5.** The upper-left panel is a MST solution, which is a summation of the solutions of the submodels shown in the other panels. Plotted are the ratio (in percent) between the velocity perturbations and the mean velocity $V_m = 2.7$ km/s.
EXTENSION TO REFLECTION TOMOGRAPHY

Theory

The generalized MST formulation in equation (6) can be applied to determine variables other than velocity. Here, an example is shown for a simultaneous determination of interval velocities and interface geometry using reflection data. Research on reflection traveltime tomography has been conducted for nearly 20 years (e.g., Bishop et al., 1985; Farra and Madariaga, 1988; Stork, 1992). Kosloff et al. (1996) demonstrate a tomographic method to determine both velocities and interface depths using reflection traveltimes. Similar inversions for 2D velocity and interface structures are reported by Zelt and Smith (1992).

The description of reflection tomography here follows the work of Zhou (1997). The velocity structure is represented by depth-varying velocity layers separated by piece-wise planar interfaces. Hence, the geometry of all interfaces can be modified by varying the depths of the corresponding interface nodes. As an extension of equation (3b), the forward SST equation for this case is

$$\delta t_i = \sum_j k_{\xi_{ij}} \delta x_j + \sum_k k_{\zeta_{ik}} \delta z_k,$$

where $J$ is the total number of slowness cells and $L$ is the total number of interface nodes to be upgraded by the inversion. Kernel $k_{\xi_{ij}}$, which is the same as $\ell_{ij}$ in equation (3b), places constraint of the $i$th ray on the $j$th slowness cell. Kernel $k_{\zeta_{ik}}$ places constraint of the $i$th ray on the $k$th interface node; this can be a reflection constraint if the ray is reflected from the interface or a transmission constraint if the ray is transmitted through the interface (Zhou, 1997). Hence, the SST is an inversion of equation (9) to determine $\delta z_k$. The conventional velocity tomography using equation (3b) would be a special case of equation (9) by dropping the interface term.

The basic idea of the MST is to apply the superposition and decomposition principles to the model variables. As an extension of equation (8b), the forward expression for the reflection MST is

$$\delta t_i = \sum_m w_{ij}^{(m)} \sum_j k_{\xi_{ij}}^{(m)} \delta x_j^{(m)} + \sum_n w_{ik}^{(n)} \sum_{\ell} k_{\zeta_{i\ell}}^{(n)} \delta z_\ell^{(n)},$$

where $M$ is the number of slowness submodels, $\delta x_j^{(m)}$ is the slowness perturbation of the $j$th cell of the $m$th slowness submodel, $N$ is the number of interface submodels, and $\delta z_\ell^{(n)}$ is depth perturbation at the $\ell$th grid point of the $n$th interface submodel. There are $J_m$ slowness cells in the $m$th-order slowness submodel and $L_n$ gridpoints in the $n$th-order interface submodel. The terms $w_{ij}^{(m)}$ and $w_{ik}^{(n)}$ are the weighting coefficients satisfying the conditions

$$\sum_m w_{ij}^{(m)} = 1,$$  \hspace{1cm} (11a)

$$\sum_n w_{ik}^{(n)} = 1.$$  \hspace{1cm} (11b)

After inversion for $\delta x_j^{(m)}$ and $\delta z_\ell^{(n)}$ in all submodels, the final MST solutions for slowness and interfaces are the following two summations:

$$\delta s(x) = \sum_m w_j^{(m)} \delta s^{(m)}(x),$$  \hspace{1cm} (12a)

$$\delta z(y) = \sum_n w_i^{(n)} \delta z^{(n)}(y),$$  \hspace{1cm} (12b)

where $x$ is the spatial variable in the slowness model and $y$ is the spatial variable along the interfaces.

A comparison example

Let us test the above formulas of the reflection SST and MST for a 2D synthetic model shown in Figure 6a with one layer of water and five layers of sediments. The model simulates a 2D marine survey. This model has velocities ranging from 1.5 km/s near the surface to 5 km/s in the bottom layer. Lateral velocity variation exists in the second and fourth layers as well as in terms of variable interface geometry. The reflection raypaths from eight shots to nine receivers are shown in Figure 6b.

The multiscale submodels are applied for each layer to both the slowness cells as in equation (12a) and the interface nodes as in equation (12b). The interface nodes are located at the intersections between the interfaces and the vertical lines in these panels. Figures 6c to 6h show six examples of the slowness submodels. The submodel with the smallest cell size is exactly the SST model. The submodel with the largest cell size in Figure 6h represents the layer average values. Since there are 10 columns of cells in each layer, we use 10 velocity submodels and eleven interface submodels. The weighting factors are kept as constant, with $w_j^{(m)}=1/M$ and $w_i^{(n)}=1/N$, respectively.

Taking the reflection traveltimes through the true model as data, the objective of the reflection tomography is to recover the slowness and interface geometry of the true model. Figure 7 shows a series of models from applying the SST, with the true model interfaces shown as red dashed curves. Starting from the initial layer-cake reference model in Figure 7a, the tomographic inversion of equation (9) is carried out iteratively using the raypaths in the concurrent reference model and traveltime residuals to determine the velocity and interface perturbations and then to update the reference velocity and interface model. The velocity value of the bottom layer does not change since there is no ray traversing it. As indicated at the top of each solution panel, the standard deviation of traveltime residuals decreases from the initial reference model in Figure 7a to the fourth iteration model in Figure 7e but increases in the fifth iteration model in Figure 7f. Comparing with the true model in Figure 6a, the SST solution in iteration four is reasonable but with obvious mismatching.

Figure 8 shows the models from the MST inversion using equation (10). With the same data, same ray tracing, same inverse operator, and same starting reference model, the MST yields better solutions comparing with that of the SST in Figure 7. The standard deviation of traveltime residuals decreases monotonically with the MST iterations, reaching a level that is much smaller than the standard deviation of the corresponding SST iterations. The MST solution from the fifth iteration in Figure 8f matches very well with the true model in slowness and interface geometry. While the lateral velocity...
variations in the second and fourth layers are recovered well, small amount of erroneous lateral velocity variations exist in the third and fifth layers.

3D CRUSTAL TOMOGRAPHY IN SOUTHERN CALIFORNIA

This section provides a field data example of the MST method in mapping the 3D crustal velocity structure in southern California using the first-arrival data of local earthquakes. The $300 \times 480$ km study region (Figure 9) includes most of southern California and surrounding areas. This region has a well-distributed seismograph network for monitoring its high seismicity (Figure 9a). During the past 15 million years the crust in this region has been highly deformed by tectonic processes associated with the transition from a convergent to transpressional plate boundary as well as displacements along strike-slip faults (Crowell, 1962; Atwater, 1970; Allen, 1981). Variations in seismic velocities associated with different rock types are useful in understanding composition and evolution history of the crust (e.g., Bjorklund et al., 2002). There have been many regional P-wave tomographic studies in this region (e.g., Humphreys et al., 1984; Magistrale et al., 1992; Zhao and Kanamori, 1992) and some S-wave studies in smaller areas, such as in the Los Angeles basin and the central Transverse Ranges (Hauksson and Haase, 1997). The example shown here follows a regional study that used both P- and S-wave data (Zhou 1994a). Only the P-wave result is presented here.

![Multiscale Tomography Diagram](image)

FIG. 6. (a) A 2D synthetic model for the reflection tomographic experiments. The values in the panel denote average velocities of each model layer (in kilometers per second) (b) Raypaths in the true model are from eight shots (stars) to nine receivers (triangles). (c)–(h) Examples of the multiscale submodels.
The first-arrival data of earthquakes were compiled by the Southern California Earthquake Data Center (SCEDC). This study used all local events that occurred from 1981 to 1994. The hypocenters in the SCEDC bulletin were determined using a regional velocity model endorsed by the USGS. For this study the raw data are arrival-time picks which, unlike traveltimes, do not depend on source locations and origin times. Our objective is to determine new hypocentral parameters and 3D velocity model. However, the bulletin values were used to presort the data. Our selected events have bulletin hypocentral errors less than 1 km in spatial dimensions and less than 0.01 s in the origin time. The 195 seismograph stations selected provide good spatial coverage. A total of 37,451 local events, with 1,026,535 P-wave picks and 131,932 S-wave picks, were used for the hypocentral determination and tomographic inversion. During preliminary processing the outliers in the data picks are sorted out using a series of windows in focal depth and epicentral distance. In each small window large deviations from the L1-norm average of the arrival times were discarded.

The depth range of the 3D model is from 3 km above sea level down to 36 km below sea level, with thirteen 3-km-thick layers covering much of the crust. The actual thickness of the top layer varies from zero in the ocean to over 2 km in the mountainous areas. The MST partition of nine submodels is applied to each model layer. As shown in Figure 9b, the cell size of the first-order submodel is $10 \times 10 \times 3$ km$^3$. The lateral dimensions of the cells in the other submodels are $20 \times 20$, $30 \times 30$, $42.86 \times 48$, $60 \times 60$, $75 \times 80$, $100 \times 120$, $150 \times 160$, and $300 \times 480$ km$^2$. Hence, the MST inversion involves 25,356 model cells of different sizes. After the postversion superposition of all submodel solutions, the final MST solution has 18,720 model cells—the same number for the corresponding SST.

The first stage of the tomographic work consists of inversion for the 1D P- and S-wave velocities, with the initial reference 1D velocity profiles interpreted from the literature. Starting with the bulletin hypocenters and raypaths in the initial reference models, the picked arrival times are used to invert for the 1D velocities and new hypocentral positions and origin times. The average velocity and its slope are determined for each type of body wave. Tests show a little change in the P- and S-wave profiles when different initial reference models are used.

To determine 3D velocity variations, the following four steps are carried out iteratively: (1) Hypocentral determination using both P- and S-wave traveltimes and reference 3D velocity model (Zhou, 1994b); (2) 3D ray tracing using updated hypocenters and velocity models to update traveltime residuals; (3) inverting for 3D P- and S-wave velocity perturbations using new raypaths and traveltime residuals; and (4) updating the reference velocity models. The inverse operator is again the LSQR algorithm (Paige and Saunders, 1982).

**Fig. 7.** Initial reference model (upper-left panel) and five iterations of the SST solutions using reflection arrivals. Dashed curves indicate the interfaces of the true model in Figure 6a. Mean and standard deviation of traveltime residuals (in milliseconds) are shown at the top of each model. See Figure 6 for the velocity scale.
Fig. 8. Same initial reference model as for the SST and five iterations of the MST solutions. Dashed curves indicate the interfaces of the true model. Mean and standard deviation of traveltime residuals (in milliseconds) are shown at the top of each model. See Figure 6 for the velocity scale.

Fig. 9. Maps of the study region in southern California. (a) Small crosses are earthquake foci, and triangles are seismographic stations used in this study. (b) For each of the nine MST submodels, a single cell is plotted and labeled by the order of its submodel. The lateral cell size of the first-order submodel is $10 \times 10$ km$^2$, and the lateral cell size of the ninth submodel is the whole model region. Cells of all submodels are overlapped on every 3-km-thick layer of the model. Major faults are outlined in (b).
Figure 10 compares the P-wave solutions from the SST and MST for crustal southern California at four depth ranges. Plotted are velocity perturbations, which are the difference between the P-wave velocity $V_p$ and the mean layer velocity $V_m$ in kilometers per second. As denoted below each panel, the mean layer velocities clearly increase with depth. Both the SST and MST solutions are derived using the same data and same initial reference model. The LSQR operator is controlled so both solutions fit the data equally well. The only cause for the difference between the solutions is the different model parameterizations between the SST and the MST. As shown in Figure 10, the two velocity solutions have similar long-wavelength patterns and amplitudes. However, the SST solution is much spikier than the MST solution. The two solutions are most similar in the shallowest depth range of 0–3 km where the ray coverage is the best. In the central part of the shallowest panels, the MST recovers most of the short-wavelength anomalies depicted by the SST solution. At places of poor ray coverage such as at depth or near the edges of the model, the spiky anomalies in the SST solution are likely to be the along-raypath smearing artifacts at places of poor ray coverage. Once again by the principle of parsimony for the two solutions that fit the data equally well, the smoother MST solution is deemed to be the more plausible.

Figure 10. P-wave tomograms from the MST (upper row) and SST (lower row) solutions at four depth ranges. Coastal line and state boundaries are shown in green; surface traces of major faults are shown in black. Denoted below each panel are the depth range and $V_m$, the mean velocity of each layer. Color patterns show velocity perturbation $dV_p = V_p - V_m$, where $V_p$ is the velocity value at each location.

Regularization of the inversion through theoretical constraints or using data and model statistics is a common way to smooth and stabilize tomographic solutions. We may regard the MST as another regularization of the model space to directly cope with unevenness in data coverage and to increase geometrical matching between model variables and real anomalies. Compared with common regularization schemes such as damping or minimizing solution perturbations, the MST penalizes those parts of the solution with poor data coverage. As demonstrated in Figure 10, the MST better preserves solution amplitude and resolution as permitted by data coverage.

CONCLUSIONS

An adequate model parameterization holds the key to reducing the nonuniqueness in traveltime tomography. This paper describes an MST method as a viable way to treat the problem of uneven and poor ray coverage in tomography. By this method the solution space is composed of a series of overlapping submodels with different cell sizes. The model variables of all submodels are inverted simultaneously, and the final solution is a superposition of all submodel solutions. The superior result of the MST in comparison with the SST is confirmed by several synthetic examples using transmission and reflection
In the field data example of constructing the 3D crustal velocity model in southern California, the MST solution is much smoother than the corresponding SST solution, even though both solutions fit the data equally well. This indicates that the MST solution is more plausible because it is the simpler of the two equal-fitting models.

There are several reasons behind the superiority of the MST with respect to the SST. First, the multitude of MST model cells available at each location improves the chance to realize the most suitable cell size with respect to raypath coverage and anomaly geometry. Second, the multitude of MST model cells means MST is more broadband, or capable of better resolution, than SST. Finally, the MST bundles the neighboring cells into a hierarchical spectrum of solution wavelengths, allowing a more effective application of damping or other regularizations of inversion. The SST and MST will yield similar solutions when the ray coverage is perfect. The computation overhead of the MST is only marginally greater than that of the SST because the inversion kernels for the MST can be approximated using the kernels of the SST.

ACKNOWLEDGMENTS

Field data were provided by the Southern California Earthquake Data Center. Critical review by Paul Docherty and two anonymous reviewers has been very helpful.

REFERENCES


