52

Probing the Earth's Interior with Seismic Tomography

Andrew Curtis
Schlumberger Cambridge Research, Cambridge, UK
Roel Snieder
Colorado School of Mines, Golden, USA

1. Introduction

Seismological data recording the oscillations of seismic waves at distinct locations on the Earth's surface can be used to infer variations in elastic properties and density within the Earth. Creating images or models of these variations is called seismic tomography. Figure 1 shows examples of different seismic wave types that are used routinely to estimate Earth structure. Interface waves propagate along interfaces within the Earth. They oscillate with an amplitude that decays with distance from the interface and are effected by (and hence contain information about) structure around that interface. Surface
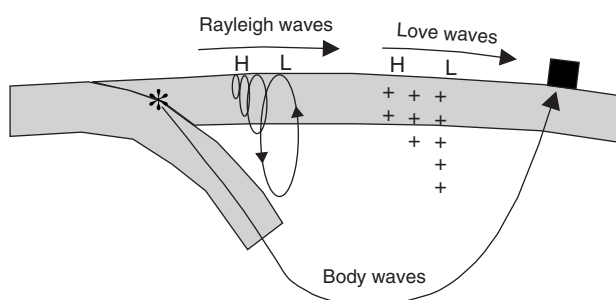

FIGURE 1 Various wave types excited by an earthquake (star) in a subduction zone and recorded at a station (square) that are often used for seismic tomography. Lithospheric plates are shaded in grey. For Rayleigh and Love waves, the general particle oscillation direction beneath the surface is drawn, while the propagation direction is shown above the surface; for the body wave only the path of energy propagation from earthquake to station is shown. "+" denotes particle oscillation perpendicular to the page, "H" denotes high frequency and "L" low frequency components. Note that low frequency surface waves generally travel faster than high frequency surface waves.

waves are a particular case that are generally classified as either Love waves which oscillate transversely (perpendicular to the path of propagation) or Rayleigh waves which oscillate elliptically in the vertical–longitudinal (parallel to the propagation path) plane. The longer the wavelength of surface waves, the deeper in the Earth they oscillate. The very long wavelength limit of surface waves are normal modes which vibrate the whole circumference of the Earth simultaneously, much like a ringing bell; the resonant frequencies with which they vibrate contain information about the structure of the whole Earth averaged in different ways. Many different body waves are distinguished (at the very least, P-compressional and S-shear waves), but all propagate through the Earth's interior and either refract, diffract or reflect energy due to elastic parameter or density variations. Hence body-wave data contain information about both interface and bulk properties of the Earth.

This chapter describes in detail the relationships between the character of body or surface wave propagation or of normal mode oscillations, and variations in subsurface Earth structure (the Earth model). It shows how, and under what conditions these relationships can be linearised; this allows linearised inverse theory to be used to estimate the structure from data recording the oscillations of such waves. However, an estimated model of Earth structure may be of little use unless we know the extent to which that estimate reflects the true Earth structure. We discuss how the inverse theory used and how the physics of seismic wave propagation limit the quality of the retrieved Earth model. We will see that constructing model estimates requires excellent data quality and stringent conditions to be fulfilled on the spatial distribution of the data. We show that these conditions can be used to *design* ideal seismological surveys. Finally, we show examples of surface-wave and body-wave tomography and illustrate that

only recently are similar Earth structures "visible" using these different data types.

2. A Brief History of Seismic Tomography

The development of seismic tomography began with attempts to estimate the radial structure of the Earth. The travel times of seismic waves for an Earth model that depends on depth only are related by an Abel transform to the seismic velocity. The inverse Abel transform was used early in this century as a tool to determine the depth-dependent velocity in the Earth from travel-time measurements (Herglotz, 1907; Wiechert, 1910; Schlichter, 1932). Using the WKB-approximation for the structure of surface wave modes, Takahashi (1955, 1957) developed an inversion method to determine the seismic velocity as a function of depth from the phase velocity of Love waves or Rayleigh waves. This method was also based on the inverse Abel transform.

Thus, the field of seismic tomography began with a method where an analytical technique, the inverse Abel transform, was used to infer the seismic velocity. In the Earth sciences, analytical techniques for creating tomographic images are not very useful. The reason for this is that in a realistic tomographic experiment, sources and receivers are distributed inhomogeneously and the reference model used in the tomography is often also inhomogeneous.

This situation is very different from examples in medical tomography where the source–receiver geometry has many degrees of symmetry and where the reference model is usually homogeneous. The resulting symmetry of the wave paths in medical tomography allows analytical methods, such as the Radon transform (Barrett, 1984), to be used to reconstruct the medium (Natterer, 1986). This high degree of symmmetry also makes it possible for the medical equipment to carry out much of the data processing automatically. Because of the lack of symmetry in most seismic tomographic experiments, the reconstruction of the medium requires the solution of very large systems of equations (see Section 4). For this reason, computational challenges have been a central focus of early developments in seismic tomography.

The earliest applications of seismic tomography in its modern form employed array data (Aki and Lee, 1976; Aki *et al.*, 1977). In these studies the local lithospheric structure under the array was inferred from observed anomalies of seismic waves. In contrast to medical imaging, in seismic tomography one has no control over the source locations if earthquakes are used to generate the seismic waves. This has led, for example, Aki and Lee (1976) and Spencer and Gubbins (1980) to include the source location and origin time in the tomographic inversion. Up to the early 1980s, seismic tomography for the 3D Earth structure was based almost exclusively on the arrival times or dispersion properties of seismic waves. The amplitude of seismic waves is influenced by the spatial derivatives of the seismic velocity since these derivatives determine the focusing of seismic waves. This effect was used by Thomson and Gubbins (1982) who developed a method of seismic tomography that treated the amplitude of seismic waves as extra relevant data.

It has been customary to solve the large systems of equations required for seismic tomography iteratively. Such methods are particularly attractive because the linear equations that must be solved are sparse when the seismic velocity is parameterized with a cell model (see Section 4). Initially, row-action methods were used where one row of the matrix at a time is used to update the model vector. Nolet (1985) introduced the more efficient conjugate gradient method to seismic tomography, where search directions in model space are updated in an efficient way. An overview of the numerical methods used to solve the linear equations of seismic tomography is given by van der Sluis and van der Vorst (1987).

With the growth in computing power and in the amount of seismological data available, it soon became possible to carry out seismic tomography on a global scale. Global images of the lower mantle were constructed by Dziewonski (1984) and by Comer and Clayton (1984). A global image for the upper mantle was presented in the same year by Woodhouse and Dziewonski (1984). These studies played an important role in the development of global tomography.

After 1984 the development of seismic tomography was extremely rapid. Refinements and improvements were made to tomographic inversion schemes. These developments include more detailed tomography on a global scale (Tanimoto and Anderson, 1985), the use of algorithms for waveform tomography (e.g., Nolet *et al.*, 1986; Tanimoto, 1987; Snieder, 1988a,b; Nolet, 1991) and the incorporation of ray-bending effects in seismic travel-time tomography (e.g., Bregman *et al.*, 1989; Sambridge, 1990). An essential element in the growth and use of seismic tomography has been the dramatic increase in the size of the tomographic problems that could be solved. Spakman *et al.* (1993) used more than half a million travel times to estimate the subsurface structure under Europe and the Mediterranean with an unprecedented resolution.

This brief overview of the history of seismic tomography cannot do justice to the numerous contributrions to this field that have been made by a large number of researchers. Review articles have appreared that focus on global tomography (Dziewonski and Woodhouse, 1987) as well as regional tomography (Romanowicz, 1991) and on the implications of seismic tomography on the dynamics of the Earth's mantle (Montagner, 1994). In addition, a number of textbooks give a comprehensive overview of the different aspects of seismic tomography (Nolet, 1987; Iyer and Hirahara, 1993; Morelli *et al.*, 1996; Dahlen and Tromp, 1998). The interested reader is encouraged to consult these texts for further information.

3. Linearising the Model–Data Relationships

In general, seismological data depend in a complicated, nonlinear way on the Earth model about which we seek information. The associated nonlinear inverse problem (see Section 4) of constraining the model by the data is usually difficult to solve, and there are no foolproof systematic algorithms that find correct solutions. It is for this reason that linear approximations play a central role in seismic tomography: after linearising the model–data relationship, one can draw on results from linear algebra to constrain the model (Section 4).

In this section three important linearisations are described. In Section 3.1 the Born approximation is introduced; this gives the first-order distortion of the entire wave-field due to model perturbations. The first-order effect of slowness perturbations on the arrival time of waves has its root in Fermat's principle as described in Section 3.2. The linearised imprint of Earth structure perturbations on its normal mode eigenfrequencies is introduced in Section 3.3.

In all cases, linearisations are approximations to the true nonlinear model–data relationships; these approximations are valid when the difference between the true Earth structure and our best initial guess at the structure (our reference model) are relatively small. Whether the true Earth structure is sufficiently close to our current best estimates that correct solutions can be found using linearised methods is still an open question.

3.1 The Born Approximation

The Born approximation is introduced here using the example of the Helmholtz equation that describes acoustic wave propagation:

$$\left(\nabla^2 + \frac{\omega^2}{c^2(\mathbf{r})}\right)p(\mathbf{r}) = F(\mathbf{r}) \tag{1}$$

In this expression $F(\mathbf{r})$ describes the source signal at angular frequency ω that excites the wave-field $p(\mathbf{r})$ which then propagates through a medium with velocity $c(\mathbf{r})$. This equation is not appropriate to describe the propagation of elastic waves in the Earth, but it contains the essential elements to convey the principle of the Born approximation with fewer complications.

Central to the Born approximation is the separation of the model into a reference model and a perturbation. In this example, $1/c^2(\mathbf{r})$ is decomposed into a constant reference velocity c_0 and a perturbation:

$$\frac{1}{c^2(\mathbf{r})} = \frac{1}{c_0^2}\left[1 + n(\mathbf{r})\right] \tag{2}$$

Heterogeneity in the model is therefore described by the perturbation $n(\mathbf{r})/c_0^2$. The second ingredient needed for the Born approximation is the response of the unperturbed medium to a point force excitation at any arbitrary location $\mathbf{r}'$. This particular response is called the Green's function; by definition it satisfies the wave Eq. (1) for the reference medium and a point force excitation at $\mathbf{r}'$:

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2}\right)G_0(\mathbf{r}, \mathbf{r}') = \delta(\mathbf{r} - \mathbf{r}') \tag{3}$$

This Green's function can be used to find the unperturbed wave $p_0(\mathbf{r})$ that propagates through the reference medium c_0. The unperturbed wave satisfies Eq. (1) with the velocity $c(\mathbf{r})$ replaced by the unperturbed velocity:

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2}\right)p_0(\mathbf{r}) = F(\mathbf{r}) \tag{4}$$

The wave-field $p_0(\mathbf{r})$ at point $\mathbf{r}$ which is excited by the excitation F and which propagates through the reference medium c_0 can then be expressed in terms of this Green's function

$$p_0(\mathbf{r}) = \int G_0(\mathbf{r}, \mathbf{r}')F(\mathbf{r}')\,dV' \tag{5}$$

where the integral is over the volume containing all sources of excitation. The life-history of the unperturbed wave can be followed by reading this expression from right to left. At the source location $\mathbf{r}'$ the wave is excited by the excitation $F(\mathbf{r}')$. The unperturbed Green's function $G_0(\mathbf{r}, \mathbf{r}')$ describes how this wave propagates from the point of excitation $\mathbf{r}'$ to the point $\mathbf{r}$.

In order to find the waves scattered by perturbations to the reference model we write the total wave field as the sum of the unperturbed wave p_0 and the scattered waves p_S:

$$p(\mathbf{r}) = p_0(\mathbf{r}) + p_S(\mathbf{r}) \tag{6}$$

Inserting this expression and the decomposition (2) in Eq. (1) gives

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2}\right)p_0(\mathbf{r}) + \left(\nabla^2 + \frac{\omega^2}{c_0^2}\right)p_S(\mathbf{r}) + \frac{\omega^2}{c_0^2}n(\mathbf{r})p_0(\mathbf{r})$$
$$+ \frac{\omega^2}{c_0^2}n(\mathbf{r})p_S(\mathbf{r}) = F(\mathbf{r}) \tag{7}$$

Using Eq. (4) for the unperturbed wave this result can be written as

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2}\right)p_S(\mathbf{r}) = -\frac{\omega^2}{c_0^2}n(\mathbf{r})p_0(\mathbf{r}) - \frac{\omega^2}{c_0^2}n(\mathbf{r})p_S(\mathbf{r}) \tag{8}$$

The last term in the right hand side contains the cross-product of the scattererd wave $p_S(\mathbf{r})$ and the perturbation $n(\mathbf{r})$, hence this term is nonlinear in the perturbation of the medium. In the

Born approximation one linearises the relation between the scattered waves and the perturbation of the medium, which means that the last term in Eq. (8) is ignored. Physically this corresponds to ignoring the multiple scattering effects. The resulting expression for the scattered waves is denoted with $p_B(\mathbf{r})$, where B stands for *Born*. In the Born approximation the scattered waves thus satisfy

$$\left(\nabla^2 + \frac{\omega^2}{c_0^2}\right)p_B(\mathbf{r}) = -\frac{\omega^2}{c_0^2}n(\mathbf{r})p_0(\mathbf{r}) \tag{9}$$

This expression is of the same form as Eq. (4), the only difference is that the right hand side is not given by the excitation $F(\mathbf{r})$ but by a term proportional to $n(\mathbf{r})p_0(\mathbf{r})$. Physically this corresponds to the fact that the perturbation of the medium can be seen as the source of the scattered waves and that this excitation is proporional to the wave field at the location of the scatterers. Equation (9) can also be solved using the unperturbed Green's function to give:

$$p_B(\mathbf{r}) = -\int G_0(\mathbf{r},\mathbf{r}')\frac{\omega^2 n(\mathbf{r}')}{c_0^2}p_0(\mathbf{r}')dV' \tag{10}$$

Just as for Eq. (5), this expression can be interpreted by reading it from right to left. The wave field $p_0(\mathbf{r}')$ is excited and propagates through the unperturbed medium to the perturbation of the medium at location $\mathbf{r}'$. At this point the wave is scattered with a scattering coefficient $-\omega^2 n(\mathbf{r}')/c_0^2$ that depends linearly on the perturbation. The Green's function then describes how the scattered wave propagates to location $\mathbf{r}$.

It should be emphasised that Eq. (10) is an approximation. In reality the wave field interacts repeatedly with the perturbations of the medium. Physically this corresponds to the double scattered waves, triple scattered waves, etc. The Born approximation retains only the single scattered waves. This is both its strength and its weakness; it leads to a linear relation for the effect of a model perturbation on the recorded data but the approximation has a limited range of validity. A full derivation of the multiple scattering series including the Born approximation is given by Snieder (2001).

The reasoning of this section is not specific to the Helmholtz equation. The only essential elements that are needed are: (1) the wave equation is linear, (2) the model can be divided in a natural way into a reference model and a perturbation and (3) the Green's function of the unperturbed problem can be computed. Exactly the same arguments can be applied to acoustic waves and electromagnetic waves (de Hoop, 1995), quantum mechanics (Mertzbacher, 1970), elastic body waves (Hudson, 1977), elastic surface waves (Snieder, 1986a,b) and the Earth's normal modes (Dahlen and Tromp, 1998). The latter applications are of great importance to seismology.

3.2 Fermat's Theorem

Travel-time tomography uses the arrival times of elastic waves to infer the velocity in the Earth. The basic principle is very simple; the travel time is the integral of the slowness along the ray-path:

$$T = \int_{\mathbf{r}[\mathbf{u}]} u(\mathbf{r})ds \tag{11}$$

In this expression u is the slowness which is defined as the inverse of the velocity: $u = 1/c$. The travel time T can be measured from recorded data and the aim of travel-time tomography is to infer the slowness u from these data. Equation (11) suggests that this is a simple problem because the travel time and the data are related linearly. There is, however, a catch. Rays are curves that render the travel time measured along the curve stationary with respect to the curve position. This means that the ray along which the slowness in Eq. (11) is integrated also depends on the (unknown) slowness. This makes the relation between the travel time and the slowness in Eq. (11) nonlinear.

Fortunately, the relation between the travel-time perturbation and the slowness perturbation can be linearised in a simple way. Just as in Section 3.1 the model is divided into a reference model and a perturbation

$$u(\mathbf{r}) = u_0(\mathbf{r}) + u_1(\mathbf{r}) \tag{12}$$

and the travel time is decomposed into the travel time T_0 through the reference medium u_0 and a travel-time perturbation δT:

$$T = T_0 + \delta T \tag{13}$$

The traditional way to derive the linearised relation between the travel-time perturbation and the slowness perturbation given in Eq. (14) is to invoke Fermat's theorem. This theorem states that the travel time to first order does not depend on changes in the ray position (e.g., Ben-Menahem and Singh, 1981; Nolet, 1987). The proof of this theorem relies on variational calculus and is unnecessarily complex. An alternative derivation of the linearised relation between δT and u_1 is given in a very short and elegant derivation by Aldridge (1994) who shows that to first order

$$\delta T = \int_{\mathbf{r}_0[u_0]} u_1(\mathbf{r}_0)ds_0 \tag{14}$$

The derivation does not use Fermat's theorem explicity, but does use the fact that the travel time is stationary under ray perturbations. This derivation can also be extended to include higher-order travel-time perturbations (Snieder and Sambridge, 1992; Snieder and Aldridge, 1995). The crux of Eq. (14) is that the integration of the slowness perturbation is now taken over the reference ray $\mathbf{r}_0$ in the reference medium. Since the rays in the reference medium are known (because the reference medium is known), Eq. (14) constitutes a linear relation between the travel-time perturbation and the slowness perturbation.
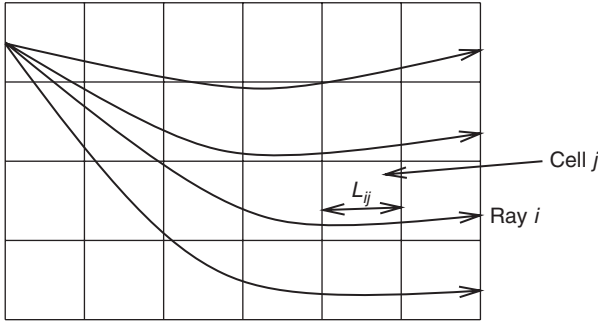
FIGURE 2 Diagram of a tomographic experiment.

The simplest way to implement Eq. (14) in tomographic inversions is to divide the models into cells and to assume that the slowness perturbation is constant in every cell. When the travel-time data are denoted with the subscript i and the cells with the subscript j the discretized form of the integral (14) is given by:

$$\delta T_i = \sum_j L_{ij} u_j \qquad (15)$$

where u_j now denotes the slowness perturbation in cell j, and L_{ij} denotes the length of ray number i through cell number j as defined in Figure 2. This means that the inverse problem reduces to the problem of solving a system of linear equations.

By solving Eq. (15) we perform a single linear inversion. The solution comprises a set of slowness perturbations u_j, one in each cell, which can then be added to the slownesses of the reference model. Thus we obtain an updated reference model. Ray paths can be recalculated in the updated medium leading to a new set of linear equations of the form in Eq. (15). Solving these provides a second set of slowness perturbations to be added to the updated reference model, and so on. This iterated, linearized inversion process is often used in ray-path tomography in the hope that the solution accuracy increases with each iteration (e.g., Bijwaard and Spakman, 2000). Note, however, that there is no guarantee that this will be true.

3.3 Rayleigh's Principle

Other data that constrain the internal structure of the Earth are the normal-mode eigenfrequencies of the Earth and the phase or group velocity measurements of surface waves. These data can all be seen as the eigenvalues of a system of differential equations for the motion in the Earth with appropriate boundary conditions. In this section we refer to the normal-mode frequencies or the surface-wave dispersion measurements as λ_i.

In general, the relation between these eigenvalues and the associated velocity model is nonlinear, which complicates the inverse problem. As in Sections 3.1 and 3.2 the Earth model can be decomposed into a reference model m_0 and a perturbation m_1. The reference model has eigenvalues $\lambda_i^{(0)}$ that are perturbed with a perturbation $\delta\lambda_i$. The relation between the perturbations of the eigenvalues and the perturbations of the Earth model can be linearised using Rayleigh's principle (Woodhouse, 1976). This states that, to first order the perturbation of the eigenvalues depends linearly on the perturbation of the Earth model and on the modes in the reference Earth model. Since the modes in the reference Earth model are assumed to be known, the inverse problem can be seen as the solution of a linear system of equations.

For example, for surface waves it can be shown (e.g., Takeuchi and Saito, 1972; Aki and Richards, 1980) that the phase velocity perturbation of surface waves is related to the perturbations in shear velocity $\delta\beta$, compressional velocity $\delta\alpha$ and density $\delta\rho$ by the relation

$$\frac{\delta c}{c} = \int K_\beta(z) \frac{\delta\beta(z)}{\beta(z)}\, dz + \int K_\alpha(z) \frac{\delta\alpha(z)}{\alpha(z)}\, dz$$
$$+ \int K_\rho(z) \frac{\delta\rho(z)}{\rho(z)}\, dz \qquad (16)$$

which holds at each location on the Earth's surface. In this expression the kernels $K(z)$ depend only of the reference model and can be assumed to be known (e.g., see Fig. 5). Again, the relation between the data δc and the model perturbations $\delta\beta$, $\delta\alpha$ and $\delta\rho$ is linear. After discretisation this leads to a linear system of equations of the form of Eq. (15) that is relatively easy to solve. Expressions similar to Eq. (16) hold for the Earth's normal modes leading to an equally simple system of equations to solve (Woodhouse, 1980; Woodhouse and Dahlen, 1978).

4. Tomographic Inversion

Performing tomography inevitably involves the use of inverse theory. Since the number of data and model parameters is often very large, linearised physics and the relatively efficient linearised form of inverse theory from the field of linear algebra are usually used. However, to evaluate the quality of any tomographic solution found, it is important that we understand the strengths and limitations of this theory which we illustrate below. Further details can be found in Matsu'ura and Hirata (1982), Tarantola and Valette (1982a,b), Tarantola (1987) or Parker (1994).

4.1 Linearised Inverse Theory

The true Earth structure has more detail than any finite set of observations can resolve. Say we construct an infinite and complete set of basis functions $\{\mathbf{B}_j(\mathbf{x}): j = 1, \ldots, \infty\}$ such that

any possible model $\mathcal{M}$ of Earth structure at location $\mathbf{x}$ could be expressed as:

$$\mathcal{M}(\mathbf{x}) = \sum_{j=1}^{P} m_j \mathbf{B}_j(\mathbf{x}) \qquad (17)$$

where $P = \infty$. The tomographic problem then consists of estimating coefficients $\mathbf{m}' = [m_1, \ldots, m_P]$. Unfortunately only a finite amount of data is ever measured, so usually we restrict the model to a finite subset of the basis functions (i.e., P finite); these functions span only those features of Earth structure that are important to us and which might be resolvable by the available data. For this purpose, basis functions commonly employed in seismology are mutually exclusive polygonal cells ($B_j(\mathbf{x}) = 1$ if $\mathbf{x} \in$ cell j, 0 otherwise), orthogonal splines, or spherical harmonics for global models.

Suppose seismological data are recorded at a set of seismic receivers, possibly using artificial seismic sources. In the previous section we showed that in some situations such data can be related approximately to an appropriately parameterised Earth model using linearised physical relationships for models close to an initial model estimate $\mathbf{m}_0$. Let the recorded data be denoted by $\mathbf{d}_{\text{rec}}$, and let $\mathbf{F}(\mathbf{m}')$ represent the forward function that predicts the recorded data for a given model $\mathbf{m}'$. The recorded data differ from the data $\mathbf{d}_0 = \mathbf{F}(\mathbf{m}_0)$ calculated for the initial model estimate $\mathbf{m}_0$ because the true Earth structure differs from the initial model, and because the recorded data are contaminated with errors $\mathbf{e}$. In the remainder we refer to the difference between the recorded data and the reference data $\mathbf{d}_0$ for the initial model as "the data" $\mathbf{d}$:

$$\mathbf{d} \equiv \mathbf{d}_{\text{rec}} - \mathbf{d}_0 \qquad (18)$$

Let the true Earth model be denoted $\mathbf{m}_\oplus$ [using the infinite-dimensional basis in Eq. (17) with $P = \infty$] and let its projection onto the (usually finite-dimensional) space of basis functions used for $\mathbf{m}_0$ be denoted $\mathbf{m}_t$. The difference between $\mathbf{m}_t$ and the reference model $\mathbf{m}_0$ is referred to as "the model" $\mathbf{m}$:

$$\mathbf{m} \equiv \mathbf{m}_t - \mathbf{m}_0 \qquad (19)$$

Using the relation $\mathbf{d}_{\text{rec}} = F(\mathbf{m}_\oplus) + \mathbf{e}$ the relation between the data and model as defined in Eqs. (18) and (19) can be linearised:

$$\begin{aligned} \mathbf{d} = \mathbf{d}_{\text{rec}} - \mathbf{d}_0 &= F(\mathbf{m}_\oplus) - F(\mathbf{m}_0) + \mathbf{e} \\ &\approx \mathbf{A}(\mathbf{m}_t - \mathbf{m}_0) + \mathbf{e} \\ &= \mathbf{A}\mathbf{m} + \mathbf{e} \end{aligned} \qquad (20)$$

The matrix $\mathbf{A}$ describes the linearisation of the full physical model–data relationship $\mathbf{F}$ (elements are $A_{ij} = \partial F_i/\partial m_j$) and depends on the data types used, the basis functions and the seismic source and receiver locations (the survey design).

The generalised least-squares solution of Eq. (20) is defined to be the one that minimises the squared misfit function E:

$$E = (\mathbf{d}_{\text{rec}} - \mathbf{F}(\mathbf{m}))^T \mathbf{W}_d (\mathbf{d}_{\text{rec}} - \mathbf{F}(\mathbf{m})) + \lambda \mathbf{m}^T \mathbf{W}_m \mathbf{m} \qquad (21)$$

$$\simeq (\mathbf{d} - \mathbf{A}\mathbf{m})^T \mathbf{W}_d (\mathbf{d} - \mathbf{A}\mathbf{m}) + \lambda \mathbf{m}^T \mathbf{W}_m \mathbf{m} \qquad (22)$$

E has two components: the first term measures the squared misfit between the observed data (perturbation) $\mathbf{d}$ and the predicted perturbations $\mathbf{A}\mathbf{m}$ that are consistent with model $\mathbf{m}$, where different datum misfits can be weighted with a matrix $\mathbf{W}_d$ to reflect data uncertainty. This is added to λ times the norm of a quadratic combination of model $\mathbf{m}$ weighted by the matrix $\mathbf{W}_m$ (second term). Here $\mathbf{W}_d$ and $\mathbf{W}_m$ are positive definite, and λ is a trade-off parameter: minimising E with λ small gives a solution that preferentially fits the weighted measured data; with λ large the solution preferentially minimises the weighted model norm. The matrix $\mathbf{W}_m$ can be defined such that it causes the model to be either as smooth or as small (damped) as possible for instance. Adding this term is known as regularising the solution (Tikhonov, 1963). A simple piece of algebra (see Snieder and Trampert, 2000) shows that the minimum E occurs at

$$\mathbf{m}_{\text{est}} = (\mathbf{A}^T \mathbf{W}_d \mathbf{A} + \lambda \mathbf{W}_m)^{-1} \mathbf{A}^T \mathbf{W}_d \mathbf{d} \qquad (23)$$

When no regularisation is applied ($\lambda = 0$) the matrix inverse taken in Eq. (23) usually does not exist. This reflects the fact that the inverse problem is ill-posed (there are insufficient data to constrain the model). The amount of regularisation that one applies should ensure that the inverse matrix in Eq. (23) is well behaved. As an alternative one can replace the matrix inverse in Eq. (23) by an approximation of the matrix inverse that is devoid of small eigenvalues (see Matsu'ura and Hirata, 1982; Menke, 1989; Parker, 1994). However, stabilising the inverse problem by an explicit regularisation has the advantage that one controls and understands better the effects of regularisation.

4.2 Assessing Quality of the Linearised Solution

The inverse problem is usually treated first as a model estimation problem followed by an assessment of how the estimated model is related to the true Earth model (the appraisal problem, Snieder, 1998). Although the solution [Eq. (23)] solves the problem of estimating Earth structure $\mathbf{m}_{\text{est}}$ from data $\mathbf{d}$, it is necessary to assess how good an approximation to the true Earth structure $\mathbf{m}_\oplus$ that this provides. We now explore the factors that cause the accuracy of $\mathbf{m}_{\text{est}}$ to deteriorate.

First, the data in Eq. (23) are usually contaminated with errors about which our only information is statistical [Eq. (20)]. These errors are mapped onto the estimated model, so model statistics are needed to indicate features in the estimated model that may be caused by errors in the data. The way in which

errors in the data propagate into the model depends on the regularisation. This follows from Eq. (23): the solution $\mathbf{m}_{est}$ depends on the type and strength of regularisation applied. The form of matrix $\mathbf{W}_m$ defines the type of regularisation and often there is no objective reason to prefer any particular form. Even once a form for $\mathbf{W}_m$ has been chosen, there is generally no objective way to choose λ unless a physical meaning is attributed to the weight matrices in terms of the probability distributions of noise $\mathbf{e}$ and models $\mathbf{m}$ (Tarantola and Valette, 1982a; Matsu'ura and Hirata, 1982; Tarantola, 1987). However, often even this strategy is impossible to apply objectively (Scales and Snieder, 1997), so in practice a degree of subjectivity is inevitable in all tomographic Earth models.

The second source of error comes from the linearised inverse theory itself. Figure 3(a) illustrates the process of finding a solution to a linearised system for one model parameter for $\lambda = 0$. Beginning at the reference model $\mathbf{m}_0$ we calculate $\mathbf{d}_0 = \mathbf{F}(\mathbf{m}_0)$ and hence the required data perturbation $\mathbf{d}$, and the linear operator $\mathbf{A}$ (the gradient) as illustrated previously. Assuming the physical relationship $\mathbf{F}$ is approximately linear, the gradient is used to extrapolate away from $\mathbf{m}_0$ to estimate the true model $\mathbf{m}_0 + \mathbf{m}$ that minimises the quadratic data misfit E given by Eq. (22). The data uncertainty (here represented by $\pm$ bounds, but more generally a data probability distribution) is mapped into a linearised model uncertainty by projecting it through the (linearised) physics.

Notice, for later on, that for the same data uncertainty, the projected model uncertainty would diminish if $\mathbf{F}(\mathbf{m})$ was steeper (roughly speaking, if $\mathbf{A}$ was larger in some sense to be defined later). This can be shown explicitly when the data uncertainties and prior model distributions are Gaussian ($\mathbf{C}_m = \mathbf{W}_m^{-1}$ and $\mathbf{C}_d = \mathbf{W}_d^{-1}$ are covariance matrices). Then the model uncertainty after a truly linear Bayesian inversion is Gaussian with covariance

$$\mathbf{C} = (\mathbf{A}^T \mathbf{C}_d^{-1} \mathbf{A} + \mathbf{C}_m^{-1})^{-1} \qquad (24)$$

(Tarantola and Valette, 1982a). In the case of one model parameter, $\mathbf{C}$ is reduced if the (single) gradient in $\mathbf{A}$ increases.

Figure 3(b) illustrates what can happen when $\mathbf{F}$ is not truly linear: if the reference model $\mathbf{m}_0$ is "close" to the unknown true solution $\mathbf{m}_t$ then generally the linearised step produces a model $\mathbf{m}_1$ that is closer to $\mathbf{m}_t$. Repeated linearised steps usually converge toward the solution. If the reference model is too far from the true solution (e.g., $\mathbf{m}_{0'}$) or if the problem is far from linear, this procedure may converge to a secondary solution (at a different misfit minimum). The linearised uncertainty estimates in each case give only the projected uncertainties around each solution found (one or other of the shaded areas in Fig. 3). The true uncertainty in both cases is more accurately described by the union of the three shaded projections: this union reflects our true state of uncertainty about the model, representing the complete subset of model space within which the data cannot discriminate. If the misfit E has n distinct minima, then the true uncertainty may comprise the union of n distinct regions of model space. Unfortunately n is seldom known or sought due to the comparitively large computational effort required for such a search. Hence, this source of error may cause not only erroneous model estimates, but invalid model uncertainty estimates.

A third source of error is caused by inherent non-uniqueness in the data recorded. Even if the inverse problem is truly linear and there are no errors in recorded data, there are usually features of the model that cannot be resolved. For example,
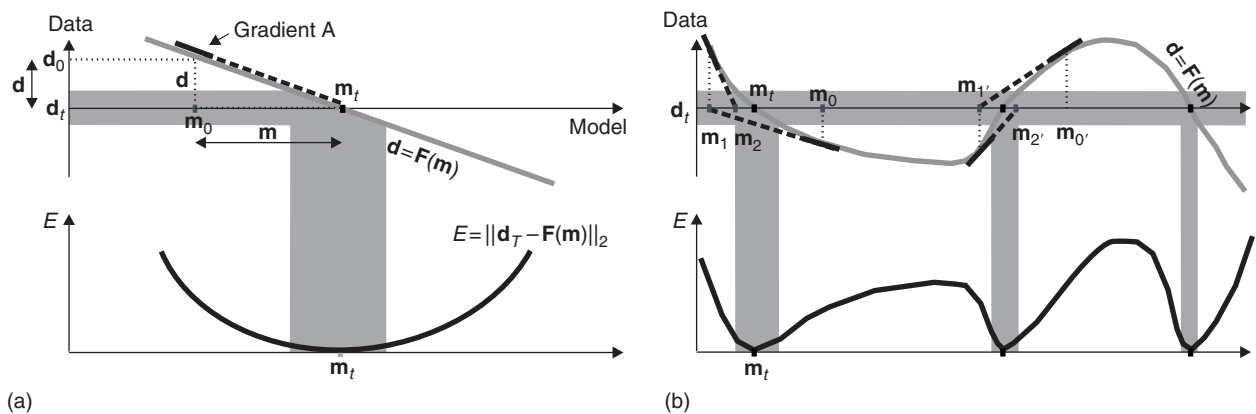

FIGURE 3 One-dimensional example of (a) linear and (b) linearised inversion. In each case, $\mathbf{m}_0$ represents the initial model chosen, and successive models found in the inversion process are shown as $\mathbf{m}_1$ and $\mathbf{m}_2$. $\mathbf{m}_t$ is the true model and the measured datum is $\mathbf{d}_t$ (with no errors in this case). The forward relationship $\mathbf{F}$ is shown in light grey and its gradient at specific points are represented by bold tangents with dashed linearised projections. The corresponding misfit function E [Eq. (21)] is shown in the lower plot of each pair. Horizontal axes always represent the model space (some labels omitted for clarity). Data uncertainties and their projections into the model space are represented by dark grey shading.

consider trying to constrain the seismic slowness of each cell in Figure 2 from a data set consisting of (noiseless) travel times along only the four ray paths shown; the data set is clearly insufficient to constrain the complete model.

In such cases, regularisation controls the otherwise unconstrained model features in our model estimate. Since the regularisation applied is usually chosen subjectively, these effects must be assessed. This is achieved by calculating the resolution matrix $\mathbf{R}$. This indicates how much information about the true Earth would be contained in estimate $\mathbf{m}_{est}$ for a given choice of $\mathbf{W}_m$, *if we had perfect data measurements and if the inversion was truly linear*. Hence, it represents a truly optimistic scenario! Its value lies in the fact that it describes the upper limit on our ability to estimate the true model due to the inherent deficiencies in the particular data set used. Substituting Eq. (20) into Eq. (23) one obtains the following relation between the estimated model and the true model:

$$\mathbf{m}_{est} = \mathbf{R}\,\mathbf{m}_t \qquad (25)$$

where

$$\mathbf{R} = (\mathbf{A}^T\mathbf{W}_d\mathbf{A} + \lambda\mathbf{W}_m)^{-1}\mathbf{A}^T\mathbf{W}_d\mathbf{A} \qquad (26)$$

If $\mathbf{R} = \mathbf{I}$ the estimated model is equal to the true model; as $\mathbf{R}$ deviates from $\mathbf{I}$ the estimate becomes an average of the true model weighted by resolution matrix $\mathbf{R}$. In nonlinear systems $\mathbf{R}$ is only valid locally around $\mathbf{m}_{est}$. Equation (26) seems to suggest that regularisation degrades the resolution; in the absence of regularisation ($\lambda = 0$) the resolution operator is given by $\mathbf{R} = (\mathbf{A}^T\mathbf{W}_d\mathbf{A})^{-1}\mathbf{A}^T\mathbf{W}_d\mathbf{A} = \mathbf{I}$. However, in practical problems in the absense of regularisation the inverse $(\mathbf{A}^T\mathbf{W}_d\mathbf{A})^{-1}$ does not exist, so one must use an approximate generalised inverse $(\mathbf{A}^T\mathbf{W}_d\mathbf{A})^{-g}$ instead. Since $(\mathbf{A}^T\mathbf{W}_d\mathbf{A})^{-g}(\mathbf{A}^T\mathbf{W}_d\mathbf{A})$ then differs from the identity matrix the resolution operator in that case also differs from the identity matrix. Usually a range of $\mathbf{W}_m$ is reasonable for any particular inversion so the resolution should be assessed for samples of $\mathbf{W}_m$ within this range (Rezovsky and Ritzwoller, 1999).

One final source of error must be considered, and this is derived from truncating expansion Eq. (17) to a finite set of P basis functions. If the true Earth structure $\mathcal{M}_t(\mathbf{x})$ contains some features that affect the recorded data but that cannot be represented within the truncated basis, these features will cause spurious errors in the finite-dimensional model solution $\mathbf{m}_{est}$ (an example is given later, see Color Plate 22). This process is called spectral leakage, and is caused by an inappropriate choice of the set of basis functions (Snieder *et al.*, 1991). Trampert and Snieder (1996) presented a solution to this problem under certain conditions, but the best way to solve it approximately is to ensure that the P chosen basis functions allow sufficient detail to model those features of $\mathcal{M}$ that have any significant influence on the data (if this is possible).

4.3 Computational Issues

Notice that each linearised iteration i in Figure 3(b) requires the calculation of the synthetic data and its gradient at model $\mathbf{m}_i$, plus a matrix inversion [Eq. (23)] or an approximation to solution (23) if this matrix is too large to be inverted (e.g., Paige and Saunders, 1982; Nolet, 1985; van der Sluis and van der Vorst, 1987). Assessing the true uncertainty in non-linear problems (e.g., the union of the shaded regions in Fig. 3) requires computationally intensive nonlinear inversion methods such as repeated linearised inversion from many different starting models (Deng, 1997), Monte Carlo (e.g., Hammersley and Handscomb, 1964), or other types of stochastic nonlinear inversion methods (Sen and Stoffa, 1991, 1992; Vinther and Mosegaard, 1996; Douma *et al.*, 1996; Sambridge, 1998; Devilee *et al.*, 1999). All of these typically require the function $\mathbf{F}$ to be evaluated many thousands of times. This is currently impossible for common large inverse systems with complex physics which is the main reason why linearised physics, inversion, and uncertainty estimates are so widely used, despite their shortcomings described above.

5. Optimal Survey Design

We now show how seismic surveys can be designed such that the data collected allow the model to be constrained with minimum uncertainty using linearised inverse methods. We showed above how the effectiveness (quality) of tomography is usually estimated using the model resolution $\mathbf{R}$ and uncertainty $\mathbf{C}$, both of which depend on $\mathbf{W}_m$, $\mathbf{W}_d$, and $\mathbf{A}$ and both of which are valid only locally around any misfit minimum. Once the regularisation $\mathbf{W}_m$ and a value of λ has been fixed, and assuming that the model basis functions have been chosen carefully, the tomographic quality approximately depends only on the number of misfit minima n, on matrices $\mathbf{A}(\check{\mathbf{m}}_i)$ (the gradient of $\mathbf{F}(\check{\mathbf{m}}_i)$) at each misfit minimum $\check{\mathbf{m}}_i$: $i = 1, \ldots, n$), and on data weights $\mathbf{W}_d$. In turn, all of these can only be influenced by varying the survey design.

Let $\mathbf{S} = [\mathbf{r}_1, \ldots, \mathbf{r}_s]$ represent the survey design where the variables $\mathbf{r}_i$ might describe the source or receiver locations, the data bandwidth, data types recorded, the make of equipment used, the number of repeated measurements, etc., all of which might affect n, $\mathbf{A}(\check{\mathbf{m}}_i)$ or $\mathbf{W}_d$. To design surveys optimally requires that we have some measure Φ of survey quality. The design $\mathbf{S}$ can then be varied such that Φ is maximised.

Let us begin by assuming that the number of minima $n = 1$. Therefore, iterated linearised tomographic inversion should be robust (Fig. 3). The easiest way to measure survey quality for linearised inversion would be to define $\Phi[\mathbf{C}, \mathbf{R} - \mathbf{I}]$ to measure some balance between the linearised uncertainty estimate $\mathbf{C}$ and how closely the resolution matrix $\mathbf{R}$ resembles the identity $\mathbf{I}$. These, and related measures are discussed by, for example, Barth and Wunsch (1990), Atkinson and Donev

(1992), Maurer and Boerner (1998), and Curtis (1999a). Increasing the complexity slightly, these linearised techniques can be adapted to measure how information is distributed between particular targets of interest within the model space; by varying the design to alter this distribution, focused surveys can be designed (Curtis, 1999b).

Examples of unfocused and focused crosswell tomographic surveys are shown in Figure 4 and Color Plate 21(a) respectively. Figure 4 shows a tomographic survey and corresponding ray paths through a homogeneous reference model that give a high degree of information about the slowness structure across the whole interwell model space (Curtis, 1999a). Notice that the average density of sources and receivers generally increases down the length of each well, and across the surface the average density is half that within the wells. The survey shown in Color Plate 21(a) uses fewer sources and receivers and hence cannot resolve velocities in all of the model cells. Instead, this plot shows a survey that provides most information about (is focused on) the region of interwell space marked by the crosses (Curtis, 1999b). Notice that the ray paths are mainly focused within the relevant region (but also note that the focusing of ray paths alone is a poor criterion on which to design surveys, see Curtis and Snieder, 1997).

A crucial subtlety exists here: survey design, by definition, is carried out before data have been collected and hence before either a misfit function E or any inversion solution $\mathbf{m}_e$ exist. Hence, instead of evaluating $\mathbf{C}$ and $\mathbf{R}$ at misfit minima, they must be evaluated at our best prior estimate $\mathbf{m}_p$ of the solution. Clearly if the gradients $\mathbf{A}(\mathbf{m}_p)$ and $\mathbf{A}(\mathbf{m}_t)$ differ, the quality estimate will be in error; this is why such methods rely intrinsically on pseudo-linearity of the forward function $\mathbf{F}$.

One method to compensate partially for this problem is offered in the field of Bayesian statistical experimental design (e.g., Atkinson and Donev, 1992; Maurer and Boerner, 1998). Say all of our knowledge about the Earth model prior to inversion can be summarised by a prior probability distribution $\rho(\mathbf{m})$, then our best estimate of survey design quality is given by:

$$\overline{\Phi} = \int_{\mathcal{M}} \Phi(\mathbf{m})\rho(\mathbf{m})d\mathbf{m} \qquad (27)$$

where $\Phi(\mathbf{m})$ is any measure of survey design quality evaluated at a specific model $\mathbf{m}$, similar to those measures discussed above for example. $\overline{\Phi}$ is called the Bayesian estimator of survey design quality. Designing experiments to maximise $\overline{\Phi}$ is the best we can do, given our (lack of)
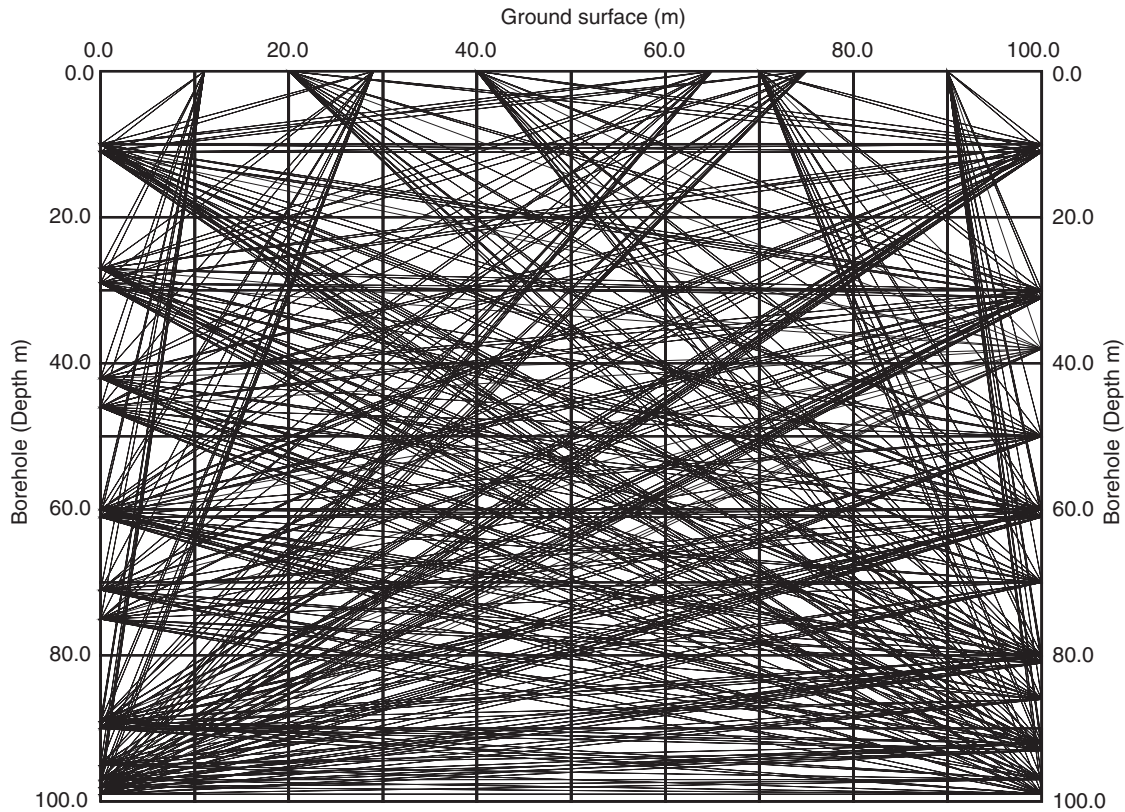

FIGURE 4 Best unfocused crosswell survey geometry found. The left and right sides of the diagram represent vertical boreholes, the top represents the ground surface. Squares represent model cells, ray paths between sources and receivers are shown by straight lines. Twenty sources and twenty receivers were used in this example.

knowledge about the Earth model before we conduct the survey.

The measure in Eq. (27) averages over the different values of Φ which arise due to nonlinearity of **F**. However, Curtis and Spencer (1999) illustrate that using a standard linearised quality measure for Φ similar to those described above does not circumvent the requirement that there is only a single misfit minimum – this is assumed in Eq. (27) since it is assumed for measure $\Phi(\mathbf{m})$. Hence, this expression only compensates for weak nonlinearity.

As explained earlier, in nonlinear inverse problems with multiple minima ($n > 1$), iterated linearised inversion converges to a single minimum and allows local uncertainty estimation, providing no indication of whether other possible solutions (other minima) exist. The only way to ensure the robustness of such inversions is therefore to change the misfit function such that it has a single minimum, and again this might be achieved by varying the survey design. This approach was demonstrated by Curtis and Spencer (1999) who varied the survey design such that the number of minima n was minimised in a seismic hypocenter estimation problem. However, these methods are computationally very expensive and so far they have only been applied to problems with few model parameters.

If we are prepared to pay the computational cost of using stochastic, nonlinear inversion methods as described earlier then we can relax the requirement that we have a unique misfit minimum since, in principle, multiple minima can be found and explored during the inversion and the true uncertainty can be assessed. In such cases, we might vary the survey design such that a measure of total expected information is maximized (Curtis and Spencer, 1999). However, this generally incurs the highest computational cost of any method proposed in this chapter. For geophysical problems with weak prior information about the model that we wish to estimate, this method is unlikely to be possible other than for low-dimensional problems [$P \sim 10$ in Eq. (17)].

6. Tomographic Applications

Tomography has been applied to many problems at length scales ranging from centimetres to the size of the Earth. There is too little space here to give an overview of all of these, so in the rest of this chapter we will focus on continental and cross-well scale tomography. We will show some examples that illustrate how the theory presented in this chapter has been applied, and how the results may be related to tectonically derived Earth structures.

6.1 Surface Wave Inversion Using the Born Approximation

Although traditionally the Born approximation has been used mostly to analyse reflection data (energy scattered from sharp velocity discontinuities), it can also be used for the analysis of transmission data since it accounts for the complete first-order perturbation of the data by changes in the model. However, it is true that the domain of applicability is larger for reflection data than for transmission data. There is a simple physical reason for this. Suppose the velocity is perturbed with a small perturbation of say 1%. In general the reflected waves will be of the same order of magnitude; they will be around 1% of the incident wave-field and are therefore very weak. Suppose a transmitted wave experiences this velocity perturbation of 1% over 100 wavelengths. This leads to a phase shift of a complete cycle. This is no longer a small perturbation of the wave field so linearised theory will not be applicable.

For surface waves in global seismology the wavelength is fairly large. For example, a Rayleigh wave with a period of 75 sec has a wavelength of about 300 km. This means that these waves propagate over a limited number of wavelengths when they are used for continental-scale inversions. It is for this reason that these waves can be analysed using the Born approximation. A description of this method of analysis, the linearised inversion, and a number of numerical examples is given by Snieder (1988a). This technique has been applied to infer the shear-velocity structure under Europe and the Mediterranean by Snieder (1988b) and under North America by Alsina *et al.* (1996). Shear-velocity perturbations from the latter study in three layers with depths between 25 and 300 km is shown in Color Plate 21(b). Red colours map slow velocities that may indicate high temperature and green colours map high velocities corresponding to low temperature. Note, though, that the identification of velocity anomalies with temperature perturbations should be treated with care. Apart from temperature perturbations the seismic velocity is also affected by variations in composition, the presence of volatiles, such as water, and possibly also pressure. A tectonic interpretation of this model is given by Alsina *et al.* (1996). Images such as this make it possible to see "snaphosts" of plate tectonics in action.

6.2 Body and Surface Wave Comparisons

It was shown earlier that linearizing approximations need to be made both for the estimation and assessment of tomographic models. In general we do not know how nonlinear these tomographic problems are. However, whatever errors are committed by linearisation, we would not expect them to be the same errors in different studies using independent data and different data types. Hence, agreement or disagreement between such studies may be a reasonable indicator of the solution quality.

In continental and global-scale seismology, models have been produced using surface waves (e.g., Trampert and Woodhouse, 1995, 1996; Ekström *et al.*, 1997; Ritzwoller and Levshin, 1998; Curtis *et al.*, 1997; Lee and Nolet, 1997; Laske and Masters, 1998), normal modes (Woodhouse

and Dziewonski, 1984), and body waves (Robertson and Woodhouse, 1995; van der Hilst *et al.*, 1997; Bijwaard and Spakman, 2000). However, it was not until recently that these models began to look alike (e.g., Ritzwoller and Lavely, 1995). Before this time, normal mode and most surface-wave models tended to be of very low lateral resolution, whereas body-wave models had very high resolution in some areas of the mantle and around seismically active zones in the lithosphere but very poor resolution elsewhere. Hence, there was no general agreement between the model structures found and it was not clear that the solutions found were really the only such structures that would fit the data.

More recently, surface-wave studies have produced models of continental structure with lateral resolution comparable to that of body-wave models. Color Plate 21(c) shows a comparison of 40 sec period, fundamental mode Rayleigh-wave phase velocity of Curtis *et al.* (1997) and a slice through the body wave *P*-velocity model of Bijwaard and Spakman (2000) at 53 km depth. To understand the phase-velocity model it is necessary to examine Figure 5 which shows the sensitivity of Rayleigh phase velocities to shear velocity at different depths in the Earth [these are the (normalised) kernels $K_\beta(z)$ in Eq. (16)]. Although *P* velocity and density also effect the phase velocity, shear velocity has a greater influence, hence these kernels show the dominant sensitivity.
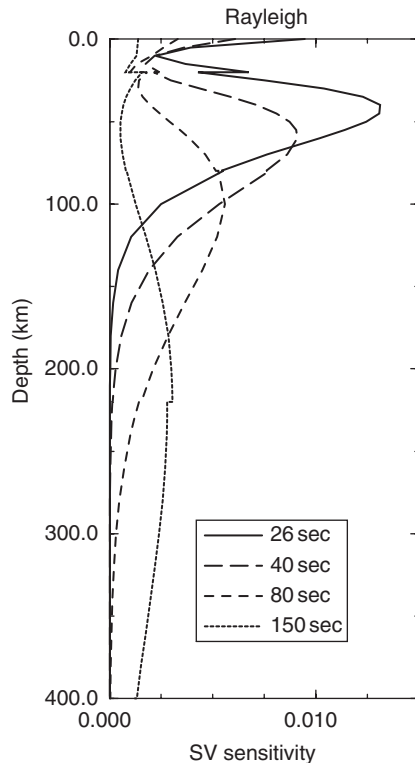
Equation 16 shows that phase velocity anomalies at any point on the Earth's surface are averages of shear velocity anomalies over depth, weighted by these kernels. Hence, the 40 sec period Rayleigh wave is most sensitive to shear velocity anomalies in the depth range 40–100 km with peak sensitivity around 55 km. It makes sense, then, to compare this with the *P*-velocity model of Bijwaard and Spakman (2000) around 55 km depth: if *P* and *S* anomalies are correlated (if they are due to the same Earth structures) then the anomalies should exhibit similar patterns.

Despite the approximations made in the above argument, several features common to both models in Color Plate 21(c) are immediately obvious. The low velocity anomaly running from the Aegean, through the Caucasus and North Iran to the Pamir and Tien Shan north of Tibet is visible in both studies. Low velocities exist beneath central Tibet in both studies, as detected in previous studies using various data types (Romanowicz, 1982; Molnar and Chen, 1984; Curtis and Woodhouse, 1997). Also, the high velocities of the Northwest Indian subcontinent are visible and extend beneath the west-central Himalayan arc in both studies. India is actively subducting beneath the Tibetan plateau but there is still a debate about how far north the subducted plate extends. Both of these studies would suggest that the answer is almost 0 km in the east, increasing to approximately the width of the Karakoram mountain chain in the west, at least around 55 km depth.

7. New Developments and Future Research

Tomographic models using body-wave and surface-wave data now appear to attain comparable resolution within some volumes of the Earth, but their sensitivity to aspects of Earth structure are tremendously different (Figs. 1 and 5). In this sense the two data types are complementary: for example, body wave models tend to have low resolution within much of the Earth's lithosphere but higher resolution in the mantle, whereas fundamental mode surface waves provide better resolution of the lithosphere but very poor resolution of the deeper mantle. Hence, some research effort is currently being spent trying to create tomographic models consistent with both data types simultaneously.

This is just one example of how a new data type can be used to compensate for deficiencies of our current data types. Such strategies would seem to be the most efficient way to add constraints to any current model, since it is often the case that it is impossible to compensate for these deficiencies simply by adding extra data of the same type. (For example, Fig. 5 shows that one could add any number of fundamental mode surface wave observations at periods lower than 40 sec and never be able to constrain the mantle at 400 km depth.) The natural extension to this is to include all possible data types, and

FIGURE 5 Sensitivity of fundamental mode Rayleigh-wave phase velocity to shear velocity variations at depth.

recently a Bayesian framework to enable this was presented by Bosch (1999).

Some of the main issues highlighted above were the deficiencies of linearised inversion schemes. Clearly, to remove problems associated with this it is necessary to use nonlinear schemes that explore much more of the misfit surface than just a single misfit minimum. However, for tomographic problems this is still computationally intractible (Curtis and Lomax, 2001). Indeed, it is only recently that iterated linearised inversion has been achieved in global tomography, and even this has only been attempted from a single initial reference model (Bijwaard and Spakman, 2000). Linearised inversion from different reference models and introducing fully nonlinear inversion are areas of future research that will surely follow as computer power increases and more efficient methodologies are found.

Two important areas of tomography that we have not had space to discuss in detail are industrial tomography and seismic anisotropy. Seismic anisotropy is the name used to describe the variation in seismic velocity for waves either travelling or oscillating in different directions. Many studies have been carried out to attempt to constrain anisotropy within the Earth (e.g., Nataf *et al.*, 1986; Montagner and Nataf, 1988; Laske and Masters, 1998), but the data geometries available using earthquake data do not always enable detailed anisotropic structure to be resolved distinctly from the isotropic structure across much of the Earth. Studies in which seismic anisotropy is most likely to be best constrained occur in industrial geophysics where surveys can be designed using the methods presented earlier to provide extremely dense data geometries that are fit specifically for this purpose. Detecting anisotropy is important for hydrocarbon production and in selecting sites for waste disposal since it tends to indicate the orientation of the prevelant fracture field. Such fracture networks produce anisotropic permeability and hence can dramatically change either the optimal network of wells required to drain a reservoir, or our estimates of the directions in which waste-infected groundwater might flow. For these reasons a great deal of industrial research is currently directed toward developing new methods for anisotropy detection.

Color Plate 22 shows an example of the isotropic components of seismic velocity found from cross-well tomography at a site in Europe (see Miller and Chapman, 1991; Miller and Costa, 1992). The upper plot was obtained when only isotropic velocity components were included in model vector **m** in Eq. (17); the lower plot was obtained when **m** included both isotropic and anisotropic velocity components. The Earth structure is clearly fairly constant horizontally between the wells, but the left-hand tomogram shows a superimposed "X" structure. This is characteristic of cases where the data were generated by wave propagation through an anisotropic medium but where the model basis functions [Eq. (17)] cannot represent this anisotropy. The cross appears by spectral leakage of the anisotropic structure into the isotropic model basis

components (Section 4). This effect is removed by using a more appropriate choice of basis functions (Color Plate 22, lower plot) that includes the anisotropy of the medium.

It has been impossible to describe all the methods now used for tomographic applications, and there are many more research areas being investigated than the few mentioned here. We have tried to cover those techniques that are common in, or may be useful for, regional or global seismology. For techniques and applications not covered here, we refer the reader to Aki and Richards (1980) or Dahlen and Tromp (1998).

Acknowledgments

References

Aki, K., A. Christoffersen, and E.S. Husebye (1977). Determination of the three-dimensional structure of the lithosphere. *J. Geophys. Res.* **82**, 277–296.

Aki, K. and W.H.K. Lee (1976). Determination of three-dimensional velocity anomalies under a seismic array using first p arrival times from local earthquakes, part 1. A homogeneous initial model. *J. Geophys. Res.* **81**, 4381–4399.

Aki, K. and P.G. Richards (1980). "Quantitative Seismology, Theory and Methods," Vol. 1. W.H. Freeman, San Francisco.

Aldridge, D.F. (1994). Linearization of the Eikonal equation. *Geophysics* **59**, 1631–1632.

Alsina, D., R.L. Woodward, and R. Snieder (1996). Shear-wave velocity structure in North America from large-scale waveform inversions of surface waves. *J. Geophys. Res.* **101**, 15969–15986.

Atkinson, A.C. and A.N. Donev (1992). "Optimum Experimental Designs." Clarendon Press, Oxford.

Barrett, H.H. (1984). The radon transform and its applications. *Progr. Optics* **XXI**, 217–286.

Barth, N. and C. Wunsch (1990). Oceanographic experiment design by simulated annealing. *J. Phys. Oceanogr.* **20**, 1249–1263.

Ben-Menahem, A. and S.J. Singh (1981). "Seismic Waves and Sources." Springer-Verlag, New York.

Bijwaard, H. and W. Spakman (2000). Nonlinear global p-wave tomography by iterated linearized inversion. *Geophys. J. Int.* **141**, 71–82.

Bosch, M. (1999). Lithologic tomography: from plural geophysical data to lithological estimation. *J. Geophys. Res.* **104**(B1), 749–766.

Bregman, N., R. Bailey, and C. Chapman (1989). Crosshole seismic tomography. *Geophysics* **54**, 200–215.

Comer, R. and R. Clayton (1984). Tomographic reconstruction of lateral velocity heterogeneity in the earth's mantle (abstract). *EOS Trans. Am. Geophys. U.* **65**, 236.

Curtis, A. (1999a). Optimal experiment design: cross-borehole tomographic examples. *Geophys. J. Int.* **136**, 637–650.

Curtis, A. (1999b). Optimal design of focussed experiments and surveys. *Geophys. J. Int.* **139**, 205–215.

Curtis, A. and A. Lomax (2001). Prior information, sampling distributions and the curse of dimensionality. *Geophysics* **66**, 372–378.

Curtis, A. and R. Snieder (1997). Reconditioning inverse problems using the genetic algorithm and revised parameterization. *Geophysics* **62**(5), 1524–1532.

Curtis, A. and C. Spencer (1999). "Survey design strategies for linearized, nonlinear inversion". *Soc. of Expl. Geophys. Expanded Abstracts*, 1775–1778.

Curtis, A. and J.H. Woodhouse (1997). Crust and upper mantle shear velocity structure beneath the Tibetan plateau and surrounding regions from interevent surface wave phase velocity inversion. *J. Geophys. Res.* **102**(B8), 11789–11813.

Curtis, A., B. Dost, J. Trampert, and R. Snieder (1997). Eurasian fundamental mode surface wave phase velocities and their relationship with tectonic structures. *J. Geophys. Res.* **103**(B11), 26919–26947.

Dahlen, F.A. and J. Tromp (1998). "Theoretical Global Seismology." Princeton University Press, Princeton.

de Hoop, A.T. (1995). "Handbook of Radiation and Scattering of Waves: Acoustic Waves in Fluids, Elastic Waves in Solids, Electromagnetic Waves." Academic Press, San Diego.

Deng, H.L. (1997). A complexity analysis of generic optimization problems: Characterizing topography of high-dimensional functions. PhD thesis, Center for Wave Phenomena, Colorado School of Mines, Golden, Colorado.

Devilee, R., A. Curtis, and K. Roy-Chowdhurry (1999). An efficient, probabilistic, neural network approach to solving inverse problems: to inverting surface wave velocities for Eurasian crustal thickness. *J. Geophys. Res.* **104**(B12), 28841–28856.

Douma, H., R. Snieder, and A. Lomax (1996). Ensemble inference in terms of empirical orthogonal functions. *Geophys. J. Int.* **127**, 363–378.

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

Dziewonski, A.M. and J.H. Woodhouse (1987). Global images of the earth's interior. *Science* **236**, 37–48.

Ekström, G., J. Tromp, and E.W.F. Larson (1997). Measurements and global models of surface wave propagation. *J. Geophys. Res.* **102**(B4), 8137–8157.

Hammersley, J.M. and D.C. Handscomb (1964). "Monte Carlo Methods." Methuen, London.

Herglotz, G. (1907). Über das benndorfsche problem der fortplanzungsgeschwindigkeit der erdbebenstralen. *Z. Geophys.* **8**, 145–147.

Hudson, J.A. (1977). Scattered waves in the coda of P. *J. Geophys.* **43**, 359–374.

Iyer, H.M. and K. Hirahara (1993). "Seismic Tomography." Prentice-Hall, London.

Laske, G. and G. Masters (1998). Surface-wave polarization data and global anisotropic structure. *Geophys. J. Int.* **132**, 508–520.

Lee, S. and G. Nolet (1997). Upper mantle s velocity structure of North America. *J. Geophys. Res.* **102**, 22815–22838.

Matsu'ura, M. and N. Hirata (1982). Generalized least-squares solutions to quasi-linear inverse problems with a priori information. *J. Phys. Earth* **30**, 451–468.

Maurer, H. and D.E. Boerner (1998). Optimized and robust experimental design: a non-linear application to em sounding. *Geophys. J. Int.* **132**, 458–468.

Menke, W. (1989). "Geophysical Data Analysis: Discrete Inverse Theory" (Revised edn.), vol. 45 of International Geophysics Series. Academic Press, Harcourt Brace Jovanovich, New York.

Mertzbacher, E. (1970). "Quantum mechanics," 2nd edn. Wiley, New York.

Miller, D. and C.H. Chapman (1991). Incontrovertible evidence of anisotropy in crosswell data. Extended abstracts of the 61st Annual Meeting of the SEG, pp. 858–928.

Miller, D. and C. Costa (1992). Inversion for the devine anisotropic medium. Expanded abstract from the EAGE annual meeting, Paris, pp. 88–89.

Molnar, P. and W.-P. Chen (1984). *s–p* travel time residuals and lateral inhomogeneity in the mantle beneath Tibet and the Himalaya. *J. Geophys. Res.* **89**(B8), 6911–6917.

Montagner, J.-P. (1994). What can seismology tell us about mantle convection? *Rev. Geophys.* **32**, 115–137.

Montagner, J.-P. and H.-C. Nataf (1988). Vectorial tomography—i. Theory. *Geophys. J.* **94**, 295–307.

Morelli, A., E. Boschi, and G. Ekström (1996). "Seismic Modelling of the Earth's Structure." Editrice Compositori, Bologna.

Nataf, H.-C., I. Nakanishi, and D.L. Anderson (1986). Measurement of mantle wave velocities and inversion for lateral heterogeneities and anisotropy 3. inversion. *J. Geophys. Res.* **91**(B7), 7261–7307.

Natterer, F. (1986). "The Mathematics of Computerized Tomography." John Wiley, New York.

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

Nolet, G. (1987). Seismic wave propagation and seismic tomography. In: "Seismic Tomography" (G. Nolet, Ed.), pp. 1–23. Reidel, Dordrecht.

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

Nolet, G., J. van Trier, and R. Huisman (1986). A formalism for nonlinear inversion of seismic surface waves. *Geophys. Res. Lett.* **13**, 26–29.

Paige, C.C. and M.A. Saunders (1982). LSQR: an algorithm for sparse linear equations and sparse least squares. *ACM Trans. Math. Soft.* **8**, 43–71.

Parker, R.L. (1994). "Geophysical Inverse Theory." Princeton University Press, Princeton.

Rezovsky, J.S. and M.H. Ritzwoller (1999). Regularization uncertainty in density models estimated from normal mode data. *Grl.* **26**, 2319–2322.

Ritzwoller, M.H. and E. Lavely (1995). Three-dimensional seismic models of the earth's mantle. *Rev. Geophys.* **33**, 1–66.

Ritzwoller, M.H. and A.L. Levshin (1998). Eurasian surface wave tomography: group velocities. *J. Geophys. Res.* **103**(B3), 4839–4878.

Robertson, G.S. and J.H. Woodhouse (1995). Evidence for proportionality of p and s heterogeneity in the lower mantle. *Geophys. J. Int.* **123**, 85–116.

Romanowicz, B. (1982). Constraints on the structure of the Tibetan plateau from pure path phase velocities of Love and Rayleigh waves. *J. Geophys. Res.* **87**(B8), 6865–6883.

Romanowicz, B. (1991). Seismic tomography of the earth's mantle. *Annu. Rev. Earth Planet. Sci.* **19**, 77–99.

Sambridge, M. (1990). Non-linear arrival time inversion: constraining velocity anomalies by seeking smooth models in 3-D. *Geophys. J. R. Astron. Soc.* **102**, 653–677.

Sambridge, M. (1998). Exploring multidimensional landscapes without a map. *Inverse Problems* **14**(3), 427–440.

Scales, J.A. and R. Snieder (1997). To Bayes or not to Bayes? *Geophysics* **62**(4), 1045–1046.

Schlichter, L.B. (1932). The theory of the interpretation of seismic travel-time curves in horizontal structures. *Physics* **3**, 273–295.

Sen, M.K. and P.L. Stoffa (1991). Nonlinear one-dimensional seismic waveform inversion using simulated annealing. *Geophysics* **56**(10), 1624–1638.

Sen, M.K. and P.L. Stoffa (1992). Rapid sampling of model space using genetic algorithms: examples from seismic waveform inversion. *Geophys. J. Int.* **108**, 281–292.

Snieder, R. (1986a). 3D linearized scattering of surface waves and a formalism for surface wave holography. *Geophys. J. R. Astron. Soc.* **84**, 581–605.

Snieder, R. (1986b). The influence of topography on the propagation and scattering of surface waves. *Phys. Earth Planet. Inter.* **44**, 226–241.

Snieder, R. (1988a). Large-scale waveform inversions of surface waves for lateral heterogeneity, 1, theory and numerical examples. *J. Geophys. Res.* **93**, 12055–12065.

Snieder, R. (1988b). Large-scale waveform inversions of surface waves for lateral heterogeneity, 2, application to surface waves in Europe and the Mediterranean. *J. Geophys. Res.* **93**, 12067–12080.

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

Snieder, R. (2001). "A Guided Tour of Mathematical Methods for the Physical Sciences." Cambridge University Press, Cambridge, UK.

Snieder, R. and D.F. Aldridge (1995). Perturbation theory for travel times. *J. Acoust. Soc. Am.* **98**, 1565–1569.

Snieder, R. and M. Sambridge (1992). Ray perturbation theory for travel times and ray paths in 3-D heterogeneous media. *Geophys. J. Int.* **109**, 294–322.

Snieder, R. and J. Trampert (2000). Linear and nonlinear inverse problems. In: "Geomatic methods for the analysis of data in the earth sciences" (Dermanis, A. and Grün, A., and F. Sanso, Eds.), Springer, Berlin.

Snieder, R.K., J. Beckers, and F. Neele (1991). The effect of small-scale structure on normal mode frequencies and global inversions. *J. Geophys. Res.* **96**, 501–515.

Spakman, W., S. Van Der Lee, and R. van der Hilst (1993). Travel-time tomography of the European–Mediterranean mantle down to 1400 km. *Phys. Earth Planet. Inter.* **79**, 3–74.

Spencer, C. and D. Gubbins (1980). Travel-time inversion for simultaneous earthquake location and velocity structure determination in laterally varying media. *Geophys. J. R. Astron. Soc.* **63**, 95–116.

Takahashi, T. (1955). Analysis of the dispersion curves of Love waves. *Bull. Earthq. Res. Inst.* **33**, 287–296.

Takahashi, T. (1957). The dispersion of Rayleigh waves in heterogeneous media. *Bull. Earthq. Res. Inst.* **35**, 297–308.

Takeuchi, H. and M. Saito (1972). Seismic surface waves. In: "Seismology: Surface Waves and Earth Oscillations (Methods in Computational Physics, Vol. 11)" (B.A. Bolt, Ed.), Academic Press, New York.

Tanimoto, T. (1987). The three-dimensional shear wave structure in the mantle by overtone waveform inversion—i. Radial seismogram inversion. *Geophys. J. R. Astron. Soc.* **89**, 713–740.

Tanimoto, T. and D.L. Anderson (1990). Lateral heterogeneity and azimuthal anisotropy of the upper mantle: Love and Rayleigh waves 100–250 s. *J. Geophys. Res.* **90**, 1842–1858.

Tarantola, A. (1987). "Inverse Problem Theory." Elsevier Science, Amsterdam.

Tarantola, A. and B. Valette (1982a). Inverse problems = quest for information. *J. Geophys.* **50**, 159–170.

Tarantola, A. and B. Valette (1982b). Generalized nonlinear inverse problems solved using the least squares criterion. *Rev. Geophys.* **20**(2), 219–232.

Thomson, C.J. and D. Gubbins (1982). Three-dimensional lithospheric modelling at norsar: linearity of the method and amplitude variations from the anomalies. *Geophys. J. R. Astron. Soc.* **71**, 1–36.

Tikhonov, A.N. (1963). On the solution of improperly posed problems and the method of regularization. *Dokl. Akad. Nauk SSSR* **151**, 501.

Trampert, J. and R. Snieder (1996). Model estimations based on truncated expansions: possible artifacts in seismic tomography. *Science* **271**, 1257–1260.

Trampert, J. and J.H. Woodhouse (1995). Global phase velocity maps of Love and Rayleigh waves between 40 and 150 seconds. *Geophys. J. Int.* **122**, 675–690.

Trampert, J. and J.H. Woodhouse (1996). High resolution global phase velocity distributions. *Geophys. Res. Lett.* **23**(1), 21–24.

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

van der Sluis, A. and H.A. van der Vorst (1987). Numerical solution of large, sparse linear algebraic systems arising from tomographic problems. In: "Seismic tomography, with Applications in Global Seismology and Exploration Geophysics" (G. Nolet, Ed.), Reidel, Dordrecht.

Vinther, R. and K. Mosegaard (1996). Seismic inversion through Tabu search. *Geophys. Prospect.* **44**, 555–570.

Weichert, E. (1910). Bestimmung des weges der erdbebenwellen im erdinneren. i. theoretisches. *Phys. Z.* **11**, 294–304.

Woodhouse, J.H. (1976). On Rayleigh's principal. *Geophys. J. R. Astron. Soc.* **46**, 11–22.

Woodhouse, J.H. (1980). The coupling and attenuation of nearly resonant multiplets. *Geophys. J. R. Astron. Soc.* **61**, 261–283.

Woodhouse, J.H. and F.A. Dahlen (1978). The effect of a general aspherical perturbation on the free oscillations of the Earth. *Geophys. J. R. Astron. Soc.* **53**, 335–354.

Woodhouse, J.H. and A.M. Dziewonski (1984). Mapping the upper mantle: three-dimensional modeling of earth structure by inversion of seismic waveforms. *J. Geophys. Res.* **89**, 5953–5986.

Editor's Note

Please see also Chapter 10, Normal modes of the Earth and planets, by P. Lognonne and E. Clevede; Chapter 11, Inversion of surface waves: a review, by Romanowicz; Chapter 16, Probabilistic approach to inverse problems, by K. Mosegaard and A. Tarantola; Chapter 51, The Earth's interior, by T. Lay; and Chapter 53, Seismic anisotropy, by Cara.