Aspects of Observational Seismology

Peter M. Shearer
Institute of Geophysics and Planetary Physics
Scripps Institution of Oceanography
University of California, San Diego

Lectures at the Earthquake Research Institute
31 August — 16 September, 2009

Preface

These notes are to accompany a short course that I will be teaching during my August/September 2009 visit to ERI. The topics are related to my own seismology research projects and are rather diverse. Thus, each lecture is largely independent of the others. My intention is to give a chalkboard description of the material in these notes, as background material, before giving a PowerPoint presentation of the accompanying research results. Some, but not all, of the material in these notes is from my book, *Introduction to Seismology* or from my papers listed in the "Additional Reading" list. These papers can be downloaded from my web site at: http://igppweb.ucsd.edu/ shearer/mahi/publist.html

Chapter 1

Upper mantle discontinuities

The upper mantle discontinuities provide important constraints on models of mantle composition and dynamics. The most established seismic discontinuities occur at mean depths near 410, 520, and 660 km and will be the focus of this lecture. The term "discontinuity" has traditionally been applied to these features, although they may involve steep velocity gradients rather than first-order discontinuities in seismic velocity. The velocity and density jumps at these depths result primarily from phase changes in olivine and other minerals, although some geophysicists, for geochemical and various other reasons, argue for small compositional changes near 660 km.

Before discussing observations of these features, I will review some of the aspects of ray theory that will help in the analysis.

1.1 Ray theory and triplications in 1-D Earth models

To first order, the Earth is spherically symmetric, as can be seen in a global stack of long-period seismograms (Fig. 1.1). A variety of seismic body-wave phases result from the P and S wave types and reflections and phase conversions within the Earth. If 3-D heterogeneity were very large, then these phases would not appear so sharp in a simple stack that combines all source-receiver paths at the same distance.

I will use the term "1-D Earth model" for spherically symmetric models in which velocity varies only as a function of radius. In this case, the ray parameter or horizontal slowness p is used to define the ray and can be expressed as:

$$p = u(z)\sin\theta = \frac{dT}{dX} = u_{\text{tp}} = \text{constant for given ray,} \qquad (1.1)$$

Long-period (vertical)
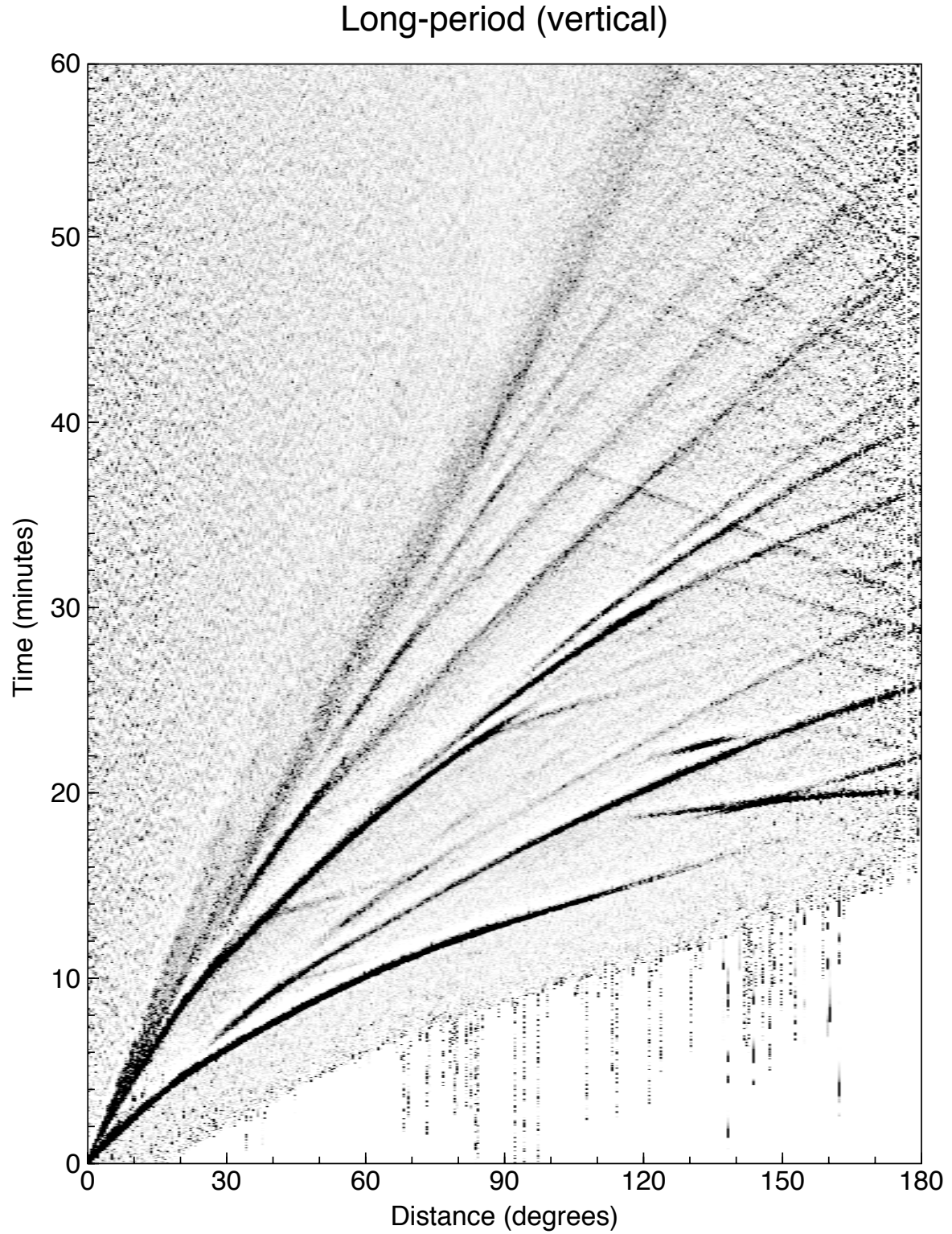

Figure 1.1: Global seismic network (GSN) vertical-component seismograms from 1988 to 1994 from shallow (< 50 km depth) earthquakes of $M > 5.7$, filtered to below 10 s period, normalized using a STA/LTA filter, and stacked in $0.5°$ epicentral distance bins. From Astiz et al. (1996).

where $u = 1/v$ is the slowness, z is depth, θ is the ray incidence angle (from vertical), T is the travel time, X is the horizontal range, and u_{tp} is the slowness at the ray turning point.

Generally in the Earth, $X(p)$ will increase as p decreases; that is, as the takeoff angle decreases, the range increases, as shown in Figure 1.2. In this case the derivative dX/dp is negative. When $dX/dp < 0$, we say that this *branch* of the travel time curve is *prograde*. Occasionally, because of a rapid velocity transition in the Earth, $dX/dp > 0$, and the rays turn back on themselves (Fig. 1.3). When $dX/dp > 0$ the travel time curve is termed *retrograde*. The transition from prograde to retrograde and back to prograde generates a *triplication* in the travel time curve.
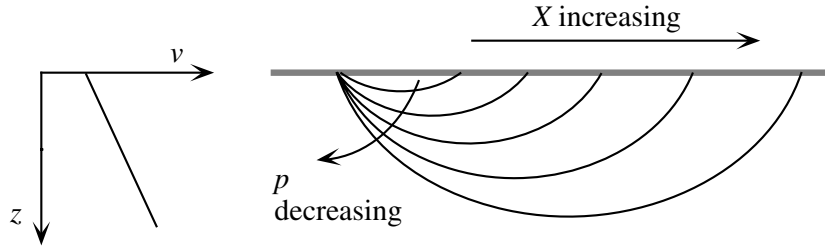
Figure 1.2: A gentle velocity increase with depth causes rays to travel further when they leave the source at steeper angles.
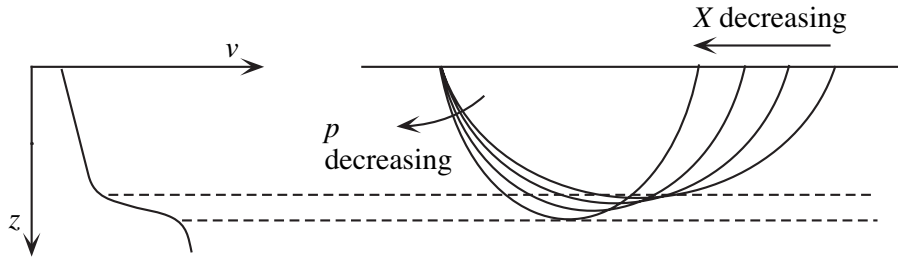
Figure 1.3: A steep velocity increase with depth causes steeper rays to fold back on themselves toward the source.

Triplications are very diagnostic of the presence of a steep velocity increase or discontinuity. The 410- and 660-km discontinuities cause a double triplication near 20 degrees (Fig. 1.4 and 1.5), which can be seen in both P wave and S waves. This is how these discontinuities were first discovered in the 1960s. Older studies of the triplications analyzed the timing (and sometimes the slopes, if array data were available) of the different branches of the travel-time curves. However, because the
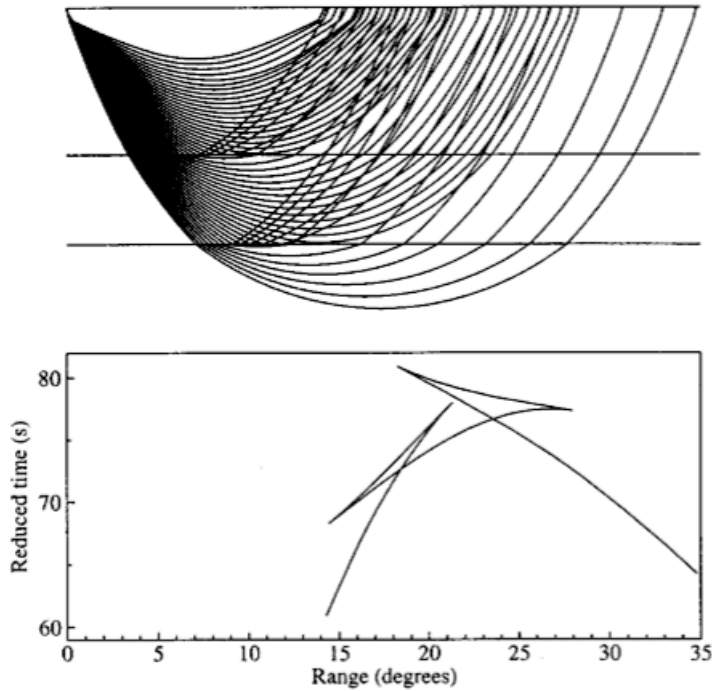
Figure 1.4: The seismic velocity increases near 410 and 660 km depth create a double triplication in the P-wave travel time curve near 20° epicentral distance, as predicted by the IASP91 velocity model (Kennet, 1991). A reduction velocity of 10 km/s is used for the lower plot. From Shearer (2000).

first arriving waves do not directly sample the discontinuities, and the onset times of secondary arrivals are difficult to pick accurately, these data are best examined using synthetic seismogram modeling. The goal is to find a velocity-depth profile that predicts theoretical seismograms that match the observed waveforms. This inversion procedure is difficult to automate, and most results have been obtained using trial-and-error forward modeling approaches.

An advantage of this type of modeling is that it often provides a complete velocity versus depth function extending from the surface through the transition zone. Thus, in principle, some of the tradeoffs between shallow velocity structure and discontinuity depth that complicate analysis of reflected and converted phases (see below) are removed. However, significant ambiguities remain. It is difficult to derive quantitative error bounds on discontinuity depths and amplitudes from forward modeling results. Tradeoffs are likely between the discontinuity properties and velocities im-
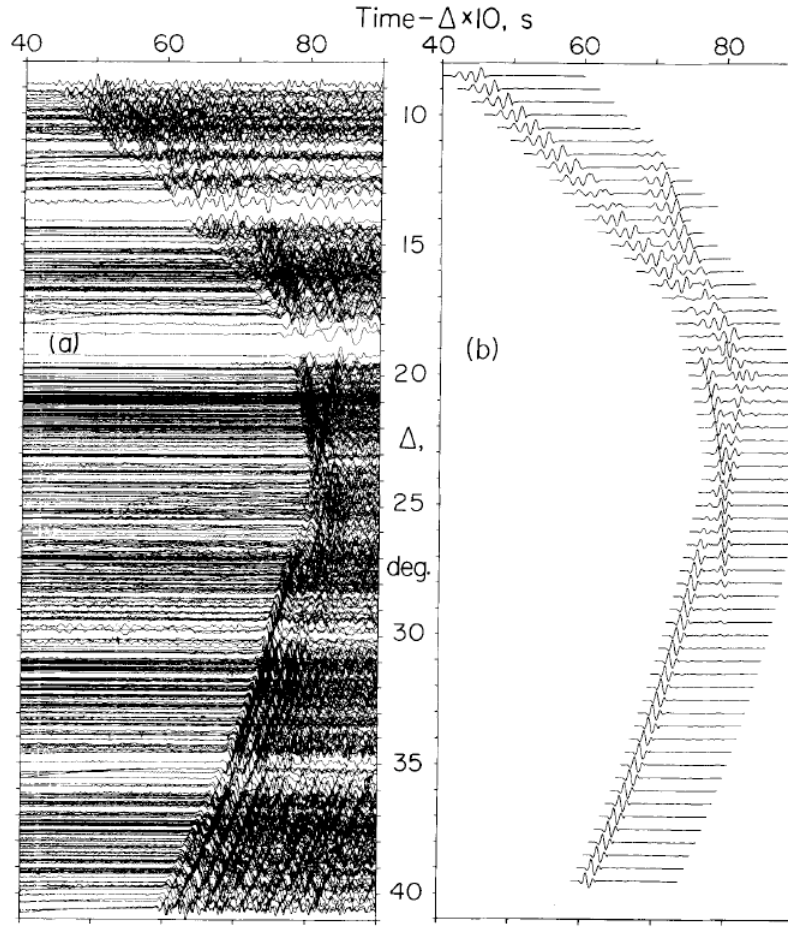
Figure 1.5: Record section of P waves from Mexican earthquakes recorded by southern California seismic stations (left) compared to synthetic seismograms. From Walck (1984).

mediately above and below the discontinuities—regions that are not sampled with first-arrival data. The derived models tend to be the simplest models that are found to be consistent with the observations. In most cases, the *410* and *660* discontinuities are first-order velocity jumps, separated by a linear velocity gradient. However, velocity increases spread out over 10 to 20 km depth intervals would produce nearly identical waveforms (except in the special case of pre-critical reflections), and subtle differences in the velocity gradients near the discontinuities could be missed. The data are only weakly sensitive to density; thus density, if included in a model, is typically derived using a velocity versus density scaling relationship.
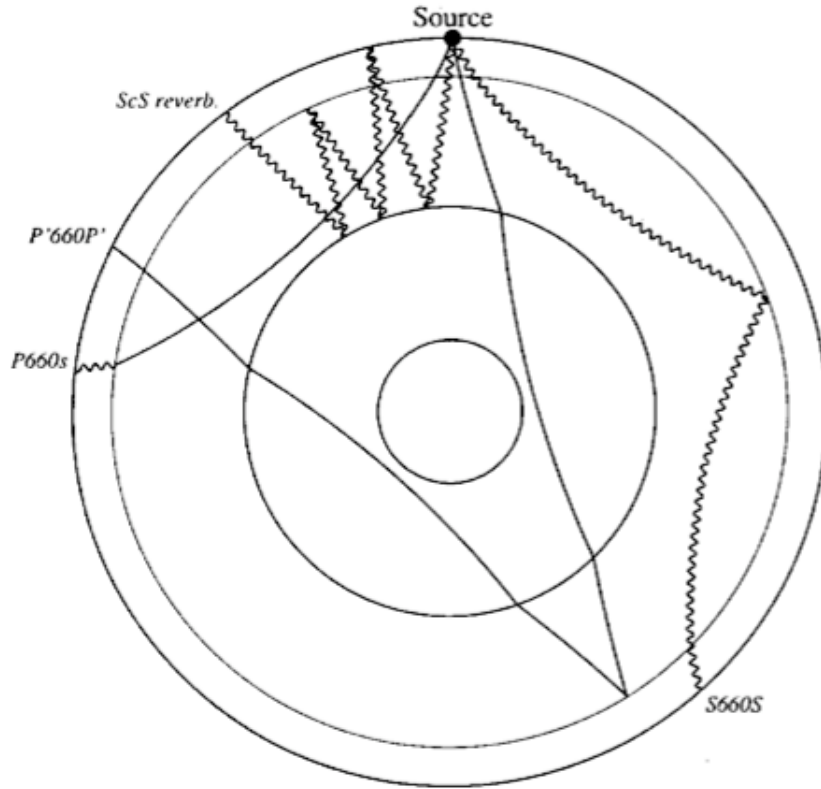
Figure 1.6: Example ray paths for discontinuity phases resulting from reflections or phase conversions at the 660-km discontinuity. P waves are shown as smooth lines, S waves as wiggly lines. The ScS reverberations include a large number of top- and bottom-side reflections, only a single example of which is plotted. From Shearer (2000).

1.2   Discontinuity phases

An alternative approach to investigating upper mantle discontinuity depths involves the study of minor seismic phases that result from reflections and phase conversions at the interfaces. These can take the form of P or S topside and bottomside reflections, or P-to-S and S-to-P converted phases. The ray geometry for many of these phases is shown in Figure 1.6. Typically these phases are too weak to observe clearly on individual seismograms, but stacking techniques (the averaging of results from many records) can be used to enhance their visibility. Analysis and interpretation of these data have many similarities to techniques used in reflection seismology.

Note that these reflected and converted waves are much more sensitive to dis-
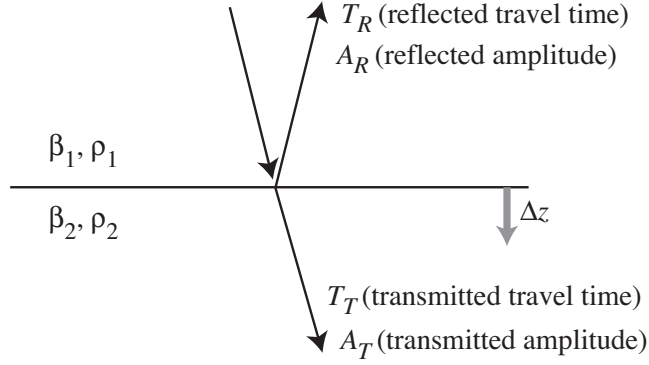
Figure 1.7: Ray geometry for near-vertical S-wave reflection and transmission.

continuity properties than directly transmitted waves. For example, consider the reflected and transmitted waves for an S-wave incident on a discontinuity (Fig. 1.7). For near-vertical incidence, the travel time perturbation for the reflected phase is approximately

$$\Delta T_R = \frac{2\Delta z}{\beta_1} \tag{1.2}$$

where Δz is the change in the layer depth and β_1 is the velocity of the top layer. The travel time perturbation for the transmitted wave is

$$\begin{aligned} \Delta T_T &= \frac{\Delta z}{\beta_1} - \frac{\Delta z}{\beta_2} = \Delta z \left(\frac{1}{\beta_1} - \frac{1}{\beta_2} \right) = \Delta z \left(\frac{\beta_2 - \beta_1}{\beta_1 \beta_2} \right) = \frac{\Delta z}{\beta_1} \frac{\beta_2 - \beta_1}{\beta_2} \\ &= \frac{1}{2} \frac{\beta_2 - \beta_1}{\beta_2} \Delta T_R \end{aligned} \tag{1.3}$$

where β_2 is the velocity in the bottom layer. Note that for a 10% velocity jump, $(\beta_2 - \beta_1)/\beta_2 \approx 0.1$, and the reflected travel time T_R is 20 times more sensitive to discontinuity depth changes than the transmitted travel time T_T.

Now consider the amplitudes of the phases. At vertical incidence, assuming an incident amplitude of one, the reflected and transmitted amplitudes are given by the S-wave reflection and transmission coefficients are

$$A_R = \grave{S}\acute{S}_{\text{vert}} \quad = \quad \frac{\rho_1 \beta_1 - \rho_2 \beta_2}{\rho_1 \beta_1 + \rho_2 \beta_2}, \tag{1.4}$$

$$A_T = \grave{S}\grave{S}_{\text{vert}} \quad = \quad \frac{2\rho_1 \beta_1}{\rho_1 \beta_1 + \rho_2 \beta_2}. \tag{1.5}$$

where ρ_1 and ρ_2 are the densities of the top and bottom layers, respectively. The product $\beta\rho$ is termed the shear impedance of the rock. A typical upper-mantle

discontinuity might have a 10% impedance contrast, i.e., $\Delta\rho\beta/\overline{\rho\beta} = 0.1$. In this case, $\grave{S}\acute{S} = -0.05$ (assuming $\beta_1\rho_1 < \beta_2\rho_2$) and $\grave{S}\grave{S} = 0.95$. The transmitted wave is much higher amplitude and will likely be easier to observe. But the reflected wave, if it can be observed, is much more sensitive to changes in the discontinuity impedance contrast. If the impedance contrast doubles to 20%, then the reflected amplitude also doubles from 0.05 to 0.1. But the transmitted amplitude is reduced only from 0.95 to 0.9, a 10% change in amplitude that will be much harder to measure. Because the reflected wave amplitude is directly proportional to the impedance change across the discontinuity, I will sometimes refer to the impedance jump as the "brightness" of the reflector.
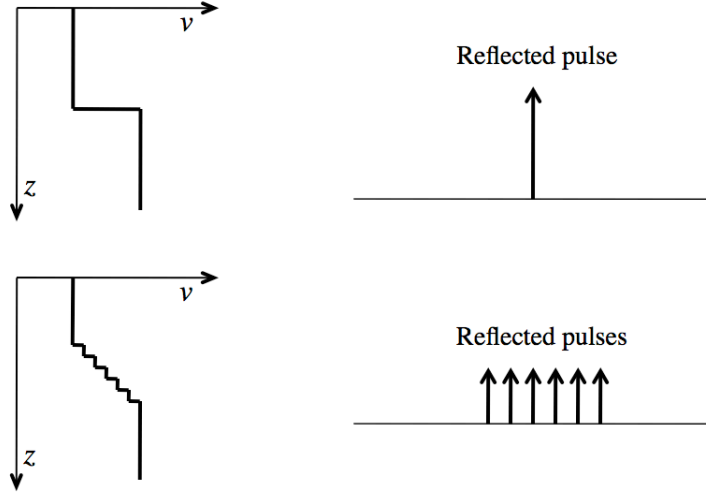
Figure 1.8: A step velocity discontinuity produces a delta-function reflected pulse. A series of velocity jumps produces a series of delta-function reflections.

Another important discontinuity property is the sharpness of the discontinuity, that is over how narrow a depth interval the rapid velocity increase occurs. This property can be detected in the possible frequency dependence of the reflected phase. A step discontinuity reflects all frequencies equally and produces a delta-function reflection for a delta-function input (Fig. 1.8, top). In contrast, a velocity gradient will produce a box car reflection. To see this, first consider a staircase velocity depth function (Fig. 1.8, bottom). Each small velocity jump will produce a delta function reflection.
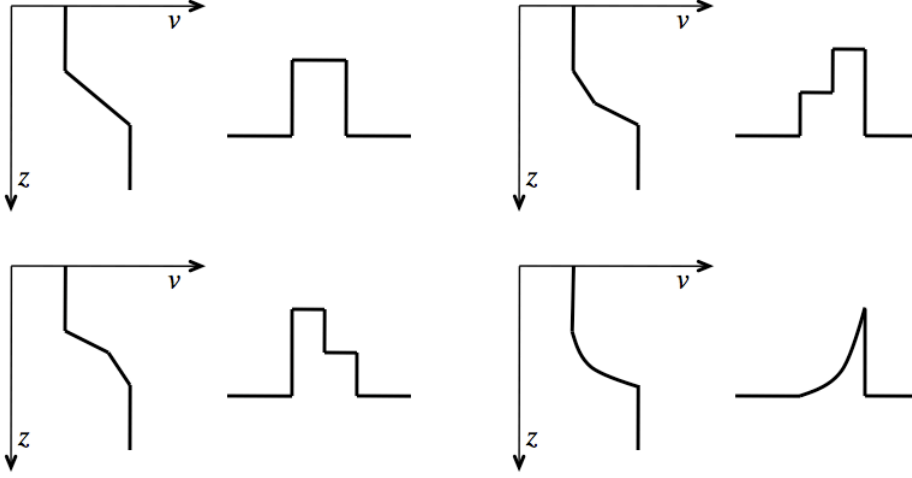
Figure 1.9: Different velocity-depth profiles and their top-side reflected pulses.

In the limit of small step size, the staircase model becomes a continuous velocity gradient, and the series of delta functions become a boxcar function (top left of Fig. 1.9). This acts as a low pass filter that removes high-frequency energy. Thus, the sharpness of a discontinuity can best be constrained by the highest frequencies that are observed to reflect off it. The most important evidence for the sharpness of the upper-mantle discontinuities is provided by observations of short-period precursors to $P'P'$. Underside reflections from both the *410* and *660* discontinuities are visible to maximum frequencies, f_{max}, of ∼1 Hz (sometimes slightly higher). The 520-km discontinuity is not seen in these data, even in data stacks with excellent signal-to-noise (Benz and Vidale, 1993), indicating that it is not as sharp as the other reflectors.

$P'P'$ precursor amplitudes are sensitive to the P impedance contrast across the discontinuities. Relatively sharp impedance increases are required to reflect high-frequency seismic waves. This can be quantified by computing the reflection coefficients as a function of frequency for specific models. If simple linear impedance gradients are assumed, these results suggest that discontinuity thicknesses of less than about 5 km are required to reflect observable P waves at 1 Hz (e.g., Richards, 1972; Lees et al., 1983), a constraint confirmed using synthetic seismogram modeling (Benz and Vidale, 1993). A linear impedance gradient of thickness h will act as a low

pass filter to reflected waves. At vertical incidence this filter is closely approximated by convolution with a boxcar function of width $t_w = 2h/v$, where t_w is the two-way travel time through the discontinuity and v is the wave velocity. The frequency response is given by a sinc function, the first zero of which occurs at $f_0 = 1/t_w$. We then have $h = v/2f_0 = \lambda/2$, where λ is the wavelength; the reflection coefficient becomes very small as the discontinuity thickness approaches half the wavelength.

Interpretation of $P'P'$ precursor results is further complicated by the likely presence of non-linear velocity increases, as predicted by models of mineral phase changes (e.g., Stixrude, 1997). The reflected pulse shape (assuming a delta-function input) will mimic the shape of the impedance profile (Fig. 1.9). In the frequency domain, the highest frequency reflections are determined more by the sharpness of the steepest part of the profile than by the total layer thickness. In principle, resolving the exact shape of the impedance profile is possible, given broadband data of sufficient quality. However, the effects of noise, attenuation and band-limited signals make this a challenging task. Recently, Xu et al. (2003) analyzed $P'P'$ observations at several short-period arrays and found that the *410* reflection could be best modeled as a 5-km-thick gradient region immediately above a sharp discontinuity.

1.3   Additional reading

Shearer, P. M. (2000). Upper mantle seismic discontinuities, *Earth's Deep Interior: Mineral Physics and Tomography from the Atomic to the Global Scale, AGU Geophysical Monograph 117*, 115–131.

Lawrence, J. F., and P. M. Shearer (2006). Constraining seismic velocity and density for the mantle transition zone with reflected and transmitted waveforms, *Geochem. Geophys. Geosys.*, **7**, doi: 10.1029/2006GC001339.

1.4   References

Astiz, L., Earle, P., and Shearer, P. (1996). Global stacking of broadband seismograms, *Seismol. Res. Lett.*, **67**, 8–18.

Benz, H.M., and J.E. Vidale (1993), Sharpness of upper-mantle discontinuities determined from high-frequency reflections, *Nature*, **365**, 147–150.

Kennett, B. N. L. (1991). *IASPEI 1991 Seismological Tables*, Res. School of Earth Sci., Aust. Natl. Univ., Canberra, Australia.

Lees, A.C, M.S.T. Bukowinski, and R. Jeanloz (1983). Reflection properties of

phase transition and compositional change models of the 670-km discontinuity, *J. Geophys. Res.*, **88**, 8145–8159.

Richards, P.G. (1972). Seismic waves reflected from velocity gradient anomalies within the Earth's upper mantle, *J. Geophys.*, **38**, 517–527.

Stixrude, L., (1997). Structure and sharpness of phase transitions and mantle discontinuities, *J. Geophys. Res.*, **102**, 14,835–14,852.

Walck, M. C. (1984). The P-wave upper mantle structure beneath an active spreading centre: the Gulf of California, *Geophys. J.R. Astron. Soc.*, **76**, 697–723.

Xu, F., J. E. Vidale, and P. S. Earle (2003). Survey of precursors to P'P': Fine structure of mantle discontinuities, *J. Geophys. Res.*, **108**, doi: 10.1029/2001JB000817.

Chapter 2

Seismic scattering

Most seismic analyses of Earth structure rely on observations of the travel times and waveforms of direct seismic waves that travel along ray paths determined by Earth's large-scale velocity structure. These observations permit inversions for radially averaged P-wave and S-wave velocity profiles as well as three-dimensional perturbations. However, smaller-scale velocity or density perturbations cause some fraction of the seismic energy to be scattered in other directions, usually arriving following the main phase as incoherent energy over an extended time interval. This later-arriving wavetrain is termed the coda of the direct phase. Scattering tends to be stronger at higher frequencies, which causes the coda to be most prominent in short-period records (see Fig. 2.1).
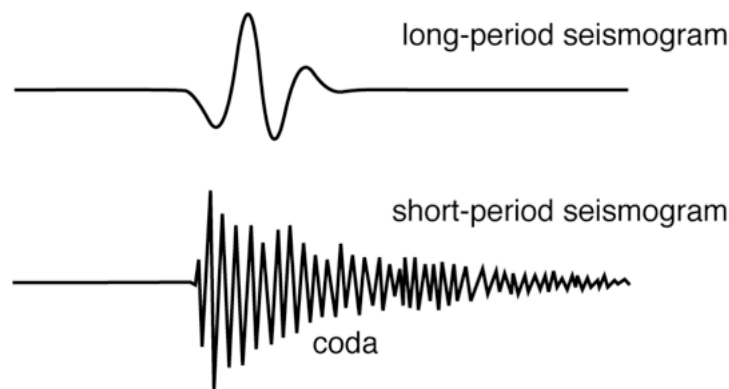

Figure 2.1: Long-period and short-period seismograms look very different.

Given the number of different scattering events and the complexity of the scattered wavefield, it is generally impossible to resolve individual scatterers. Instead,

coda-wave observations are modeled using random media theories that predict the average energy in the scattered waves as a function of scattering angle, given the statistical properties of the velocity and density perturbations. In this way, it is possible to characterize Earth's heterogeneity at much smaller scales than can be imaged using tomography or other methods.

The fact that direct seismic waves can be observed in the Earth indicates that this scattering must be relatively weak so that a significant fraction of the seismic energy remains in the primary arrivals. Scattering is much stronger in the Moon, making direct wave arrivals hard to resolve in lunar seismograms. In addition to facilitating observations of direct arrivals, weak (as opposed to strong) scattering also can simplify modeling by permitting use of single-scattering theory (i.e., the Born approximation). However, it is now clear that accurate modeling of scattering in the lithosphere, and possibly deeper in the mantle as well, requires calculations based on multiple scattering theories. Fortunately, increased computer power makes these calculations computationally feasible.

A good reference for seismic scattering is the book by Sato and Fehler (1998), henceforth termed S&F, which provides an extensive review of scattering theory and analysis methods, as well as a comprehensive summary of crustal and lithospheric studies.

2.1   Scattering theory

Wave scattering from random heterogeneities is a common phenomenon in many fields of science and theoretical modeling approaches have been extensively developed in physics, acoustics and seismology. Solving this problem for the full elastic wave equation (i.e., for both P and S waves) in the presence of strong perturbations in the elastic tensor and density is quite difficult, so various simplifying approximations are often applied. These include:

1. Assuming an isotropic elastic tensor
2. Using first-order perturbation theory in the case of weak scattering
3. Using the diffusion equation for very strong scattering
4. Assuming correlations among the velocity and density perturbations

2.1.1 Single-scattering theory and random media

For sufficiently weak velocity and density perturbations, most scattered energy will have experienced only one scattering event and can be adequately modeled using single-scattering theory. The mathematics in this case are greatly simplified if we assume that the primary waves are unchanged by their passage through the scattering region (the Born approximation). The total energy in the seismic wavefield therefore increases by the amount contained in the scattered waves and energy conservation is not obeyed. Thus this approximation is only valid when the scattered waves are much weaker than the primary waves, which is the case in the Earth when the velocity and density perturbations are relatively small (quantifying exactly how small depends upon the frequency of the waves and the source-to-receiver distance). Single-scattering theory is sometimes called Chernov theory after Chernov (1960). Detailed descriptions of Born scattering theory for elastic waves are contained in Wu and Aki (1985a,b), Wu (1989) and Sato and Fehler (1998).

A derivation of the Born equations for an isotropic medium can be found in section 13.2 of Aki and Richards (1980). It begins with the momentum equation for isotropic material:

$$\rho \ddot{u}_i = \partial_i(\lambda \partial_k u_k) + \partial_j[\mu(\partial_i u_j + \partial_j u_i)] \tag{2.1}$$

where $\mathbf{u}$ is the displacement vector, ρ is density, and λ and μ are the Lamé parameters. At this point we are assuming a general inhomogeneous medium, so the partial derivatives on the r.h.s. will apply to λ and μ as well as to the displacement. Now assume that the inhomogeneous medium consists of the sum of two parts, an "unperturbed" homogenous medium and the perturbations that make up the heterogeneity. Then the perturbed medium properties may be expressed as

$$
\begin{aligned}
\rho &= \rho_0 + \delta\rho \\
\lambda &= \lambda_0 + \delta\lambda \\
\mu &= \mu_0 + \delta\mu
\end{aligned} \tag{2.2}
$$

where ρ_0, λ_0 and μ_0 are for the unperturbed medium and are constant, and where $\delta\rho$, $\delta\lambda$ and $\delta\mu$ are the perturbations (functions of position but assumed to be much

smaller than the unperturbed values).

Now let us write the displacement $\mathbf{u}$ as the sum of primary waves $\mathbf{u}^0$ and scattered waves $\mathbf{u}^1$

$$\mathbf{u} = \mathbf{u}^0 + \mathbf{u}^1 \tag{2.3}$$

$\mathbf{u}^0$ is the solution for the unperturbed medium. The Born scattering equations are obtained by substituting (2.3) and (2.3) into (2.1), and dropping all terms involving squares or products of the perturbed medium terms and the scattered wave terms. Thus, both the primary wave is assumed to propagate cleanly through the unperturbed medium. Interactions between the primary wave and the medium perturbations generate scattered waves. These scattered waves, once generated, propagate cleanly through the unperturbed medium—there is no secondary or multiple scattering.

Single-scattering theory provides equations that give the average scattered power as a function of the incident and scattered wave types (i.e., P or S), the power of the incident wave, the local volume of the scattering region, the bulk and statistical properties of the random medium, the scattering angle (the angle between the incident wave and the scattered wave), and the seismic wavenumber ($k = 2\pi/\Lambda$, where Λ is the wavelength). A general random medium could have separate perturbations in P velocity, S velocity and density, but in practice a common simplification is to assume a linear scaling relationship among the perturbations (e.g., Sato, 1990) and/or to assume zero density perturbations. Performing the actual calculation for a specific-source receiver geometry involves integrating the contributions of small volume elements over the scattering region of interest. Each volume element will have a specific scattering angle and geometrical spreading factors for the source-to-scatterer and scatterer-to-receiver ray paths.

The nature of the scattering strongly depends upon the relative length scales of the heterogeneity and the seismic waves. Normally one has no hope of actually resolving all of the individual scatterers but only some statistical measure of their scale length and strength. A standard way to describe the spatial fluctuation of a random field is with the autocorrelation function (ACF). Let us define the fractional
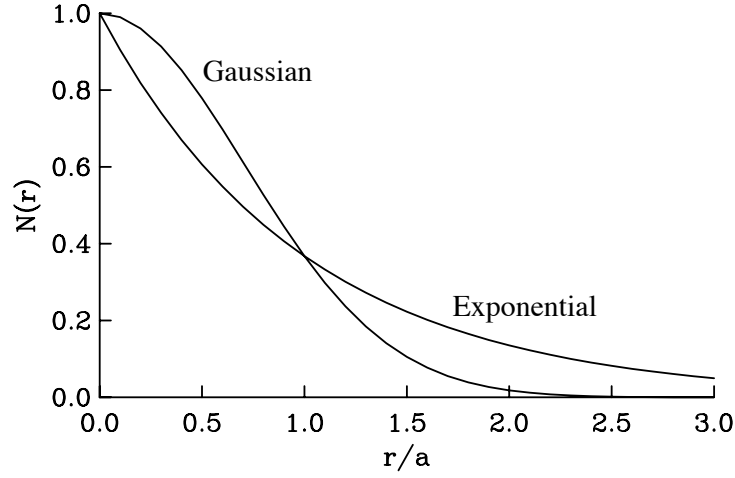
velocity perturbation as:

$$\mu = -\delta c / c_0 \tag{2.4}$$

(do not confuse this parameter with the shear modulus!) where we assume the fluctuation of μ is isotropic and stationary in space. The normalized autocorrelation function is

$$N(\mathbf{r}) = \frac{\langle \mu(\mathbf{r}')\mu(\mathbf{r}' + \mathbf{r})\rangle}{\langle \mu^2\rangle} \tag{2.5}$$

where $\langle\ \rangle$ is a spatial average over many statistically independent samples. Two specific forms for $N(\mathbf{r})$ are often modeled:

$$
\begin{aligned}
N(\mathbf{r}) &= e^{-|\mathbf{r}|/a} && \text{(exponential model)} & (2.6)\\
&= e^{-|\mathbf{r}|^2/a^2} && \text{(Gaussian model)} & (2.7)
\end{aligned}
$$

where a is called the *correlation distance*. Note that the Gaussian model will have "blobs" of relatively uniform size, whereas the exponential model will have greater heterogeneity structure at both smaller and larger wavelengths.

The correlation distance, a, provides a rough measure of the average size of the "blobs" in many commonly assumed forms for the ACF (e.g., Gaussian, exponential, van Kármán, etc.). Figure 2.2 shows examples of random realizations of the Gaussian and exponential ACF models.

If the velocity heterogeneity is large compared to the seismic wavelength ($a \gg \Lambda$, ka is large), then forward scattering predominates and becomes increasingly concentrated near the direction of the incident wave as ka increases. Back-scattered

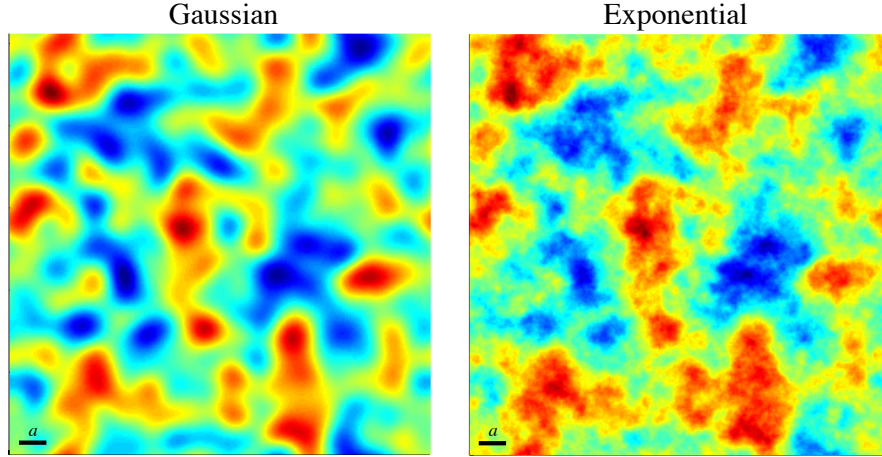Gaussian   Exponential

Figure 2.2: Examples of random media defined by a Gaussian ACF (left) and an exponential ACF (right). The correlation distance, a, is indicated in the lower-left corner.

power becomes very small, particularly for the Gaussian model. In the limit of large ka, the energy remains along the primary ray path and scattering effects do not need to be taken into account. However, this neglects the effect of density perturbations, which tend to increase the amount of backscattered power. Alternatively, if the blobs are small compared to the seismic wavelength ($a \ll \Lambda$, ka is small), then the scattering is often approximated as isotropic and the scattered power scales as $k^4 a^3$. In the limit of small ka, the scattering strength goes to zero and the medium behaves like a homogeneous solid. As discussed by Aki and Richards (1980, p. 749-750), scattering effects are strongest when a and Λ are of comparable size (i.e., when $ka \sim 2\pi$), in which case the scattering is highly directional.

Aki and Chouet (1975) presented an important application of single scattering theory to predict coda decay rates for local earthquakes. For a co-located source and receiver and homogeneous body-wave scattering in 3-D media, they obtained

$$A_C(t) \propto t^{-1} e^{-\omega t / 2Q_C} \tag{2.8}$$

where A_C is the coda amplitude at time t (from the earthquake origin time) and angular frequency ω. Q_C is termed the coda Q and there has been some uncertainty regarding its physical meaning, in particular whether it describes intrinsic attenuation, scattering attenuation, or some combination of both. Regardless of its

interpretation, this formula has proven successful in fitting coda decay rates in a large number of studies.

Single-scattering theory has also been important for modeling deep Earth scattering in terms of random heterogeneity models (see Shearer, 2007, for a review). Born theory has also been used to model expected travel time variations in direct arrivals that travel through random velocity heterogeneity (e.g., Spetzler and Snieder, 2001; Baig et al., 2003; Baig and Dahlen, 2004a,b). Although our focus here is largely on incoherent scattering from random media, it should be noted that the Born approximation can also be used to model the effect of specific velocity structures, provided their perturbations are weak compared to the background velocity field. In this case, true synthetic seismograms can be computed, not just the envelope functions. In addition, Born theory forms the basis for computing sensitivity kernels in finite-frequency tomography methods (e.g., Dahlen et al., 2000; Nolet et al., 2006).

2.1.2 Born equations for correlated velocity and density perturbations

In general, the Born scattering equations are quite complex. However, they become simpler if the material properties (velocity and density) obey fixed scaling relations. That is, assume that the P velocity α and S velocity β have have the same fractional velocity fluctuations (S&F 4.47):

$$\xi(\mathbf{x}) = \frac{\delta\alpha(\mathbf{x})}{\alpha_0} = \frac{\delta\beta(\mathbf{x})}{\beta_0} \tag{2.9}$$

where α_0 and β_0 are the mean P and S velocities of the medium. We further assume that the fractional density fluctuations are proportional to the velocity variations (S&F 4.48):

$$\frac{\delta\rho(\mathbf{x})}{\rho_0} = \nu\xi(\mathbf{x}) \tag{2.10}$$

where ν is the density/velocity fluctuation scaling factor.

The basic scattering patterns are given by (S&F 4.50):

$$X_r^{PP}(\psi,\zeta) \;=\; \frac{1}{\gamma_0^2}\left[\nu\left(-1+\cos\psi+\frac{2}{\gamma_0^2}\sin^2\psi\right)\right.$$
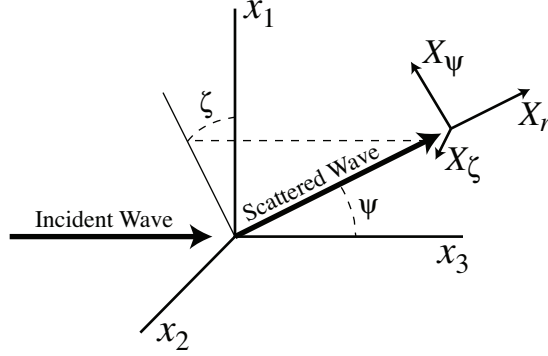
Figure 2.3: The ray-centered coordinate system used in the scattering equations. The incident ray is in the x_3 direction. For S waves the initial polarization is in the x_1 direction. The scattered ray direction is defined by the angles ψ and ζ. The scattered ray polarization is defined by X_r, X_ψ and X_ζ.

$$
\begin{aligned}
& \qquad\qquad\qquad\qquad -2 + \frac{4}{\gamma_0^2}\sin^2\psi \Bigg] \\
X_\psi^{PS}(\psi,\zeta) &= -\sin\psi\left[\nu\left(1 - \frac{2}{\gamma_0}\cos\psi\right) - \frac{4}{\gamma_0}\cos\psi\right] \\
X_r^{SP}(\psi,\zeta) &= \frac{1}{\gamma_0^2}\sin\psi\cos\zeta\left[\nu\left(1 - \frac{2}{\gamma_0}\cos\psi\right)\right. \\
& \qquad\qquad\qquad\left. -\frac{4}{\gamma_0}\cos\psi\right] \\
X_\psi^{SS}(\psi,\zeta) &= \cos\zeta\left[\nu(\cos\psi - \cos 2\psi) - 2\cos 2\psi\right] \\
X_\zeta^{SS}(\psi,\zeta) &= \sin\zeta\left[\nu(\cos\psi - 1) + 2\cos\psi\right]
\end{aligned}
\tag{2.11}
$$

where X_r^{PP} is the radial component of P-to-P scattering, $X\psi^{PS}$ is the ψ component of P-to-S scattering, etc. The angles ψ and ζ are defined as in Figure 2.3 and the velocity ratio $\gamma_0 = \alpha_0/\beta_0$.

Assuming a random media model, the scattered power per unit volume is given by the scattering coefficients for the various types of scattering (P to P, P to S, etc.) (S&F 4.52):

$$
\begin{aligned}
g^{PP}(\psi,\zeta;\omega) &= \frac{l^4}{4\pi}\left|X_r^{PP}(\psi,\zeta)\right|^2 P\left(\frac{2l}{\gamma_0}\sin\frac{\psi}{2}\right) \\
g^{PS}(\psi,\zeta;\omega) &= \frac{1}{\gamma_0}\frac{l^4}{4\pi}\left|X_\psi^{PS}(\psi,\zeta)\right|^2 \\
& \qquad P\left(\frac{l}{\gamma_0}\sqrt{1 + \gamma_0^2 - 2\gamma_0\cos\psi}\right) \\
g^{SP}(\psi,\zeta;\omega) &= \gamma_0\frac{l^4}{4\pi}\left|X_r^{SP}(\psi,\zeta)\right|^2 \\
& \qquad P\left(\frac{l}{\gamma_0}\sqrt{1 + \gamma_0^2 - 2\gamma_0\cos\psi}\right)
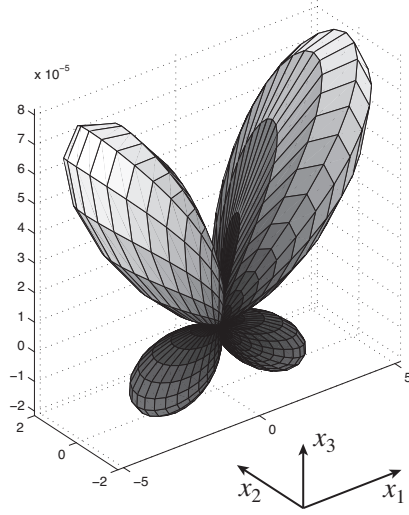\end{aligned}
$$

Figure 2.4: An example of a S-to-P scattering pattern as a function of ray angle, computed using the Born equations for a random media model. The distance from the origin gives the value of the coefficient g^{SP} in equation 2.12. The incident S wave is traveling in the x_3 direction and is polarized in the x_1 direction.

$$g^{SS}(\psi, \zeta; \omega) = \frac{l^4}{4\pi} \left(\left| X_{\psi}^{SS}(\psi, \zeta) \right|^2 + \left| X_{\zeta}^{SS}(\psi, \zeta) \right|^2 \right)$$
$$P \left(2l \sin \frac{\psi}{2} \right) \tag{2.12}$$

where $l = \omega/\beta_0$ is the S wavenumber for angular frequency ω, P is the power spectral density function (PSDF) for the random media model (see S&F p. 14–17). A popular choice for P is the exponential autocorrelation function, in which case we have (S&F 2.10):

$$P(m) = \frac{8\pi\epsilon^2 a^3}{(1 + a^2 m^2)^2} \tag{2.13}$$

where a is the correlation distance, ϵ is the RMS fractional fluctuation ($\epsilon^2 = \langle \xi(\mathbf{x})^2 \rangle$), and m is the wavenumber (i.e., the argument of the P functions in (2.12) above). For example,

$$P \left(2l \sin \frac{\psi}{2} \right) = \frac{8\pi\epsilon^2 a^3}{\left(1 + 4a^2 l^2 \sin^2 \frac{\psi}{2} \right)^2} \tag{2.14}$$

An example scattering pattern plotted in Figure 2.4 (the S-to-P coefficient g_{SP} in equation (2.12), computed for $\gamma_0 = \sqrt{3}, \nu = 0.8, \beta_0 = 6/\sqrt{3}$ km/s, $\omega = 2\pi, \epsilon = 0.01$, and $a = 1$ km), which is plotted at $6°$ increments in ψ and ζ.

The PSDF defines the strength of the heterogeneity as a function of its scale

length and controls how the amplitude of the scattering varies with seismic wavelength. Observations at a single frequency mainly constrain the heterogeneity at scale lengths near the seismic wavelength and cannot determine the PSDF very completely. Analysis of broadband data and consideration of scattered arrival amplitudes as a function of frequency is necessary to make quantitative estimates of the PSDF.

2.1.3   How To Write a Born Scattering Program

Most scattering programs are based on ray theory so you will need to be able to trace rays through your model and to compute travel time and geometrical spreading factors.

1. Define the background velocity vs. depth model, the source and receiver locations, and the ray paths to be modeled.

2. Decide on what type of random media (e.g., exponential, Gaussian, etc.) and what scattering equations you will use. This will determine what parameters you will need to specify the scattering part of the model. Determine the frequency (ω) at which you will model the scattering.

3. Determine where the scattering volume is in the Earth that you will use to model your observations. Specify the heterogeneity parameters that you will need, such as the scale length, the RMS velocity heterogeneity, the P-to-S scaling, etc.

4. Divide the scattering volume into cells that you will use to numerically integrate the scattered power.

5. For each source-receiver pair, initialize a time series to zero values.

6. For each cell in your scattering volume, compute the source-to-cell travel time and amplitude, A, of the incident wave. Compute the scattering angle, θ, the difference between the incident ray direction and the takeoff direction of the scattered ray that will land at the receiver (this is one of the trickier parts so be sure to thoroughly test this part of the code!). Compute the geometrical spreading factor for the scattered ray. Compute the local wavenumber k from ω and the average velocity in the cell.

7. Use your preferred scattering equations to compute the amount of scattered power that will arrive at the receiver. Using the total source-to-scatterer-to-receiver travel time, add this contribution to your time series.

8. Repeat (6) and (7) for all the cells in your scattering volume.

9. Repeat (5)-(8) for all of your source-receiver pairs.

10. Your synthetics will give power as a function of time. If they are noisy looking, try using a longer sample interval dt for your time series or convolve the result with a realistic source-time function (in energy, not amplitude!).

11. Take the square root if you want the amplitude envelopes.

12. Often you will want to compare the scattered power to that in the direct arrival. To do so, simply compute the ray theoretical amplitude for each source-to-receiver ray path.

13. You can add in the effect of Q along the ray paths and reflection and transmission coefficients where the rays cross boundaries if you want to include these effects.

2.2   Multiple scattering theories

If the energy in the scattered wavefield is a significant fraction of the energy in the direct wave, then the Born approximation is inaccurate and a higher-order theory should be used that takes into account the energy reduction in the primary wave and the fact that the scattered waves may experience more than one scattering event. These effects are all naturally accounted for using finite-difference or finite-element calculations, but these are computationally intensive and there is a need for faster approaches that also provide physical insight into the scattering process. In the case of very strong scattering, the diffusion equation can be applied by assuming a random walk process. Although this approach preserves energy, it violates causality by permitting some energy to arrive before the direct P wave.

Most current approaches to synthesizing multiple scattering use radiative transfer theory to model energy transport. Radiative transfer theory was first used in seismology by Wu (1985) and Wu and Aki (1988) and recent reviews of the theory are contained in Sato and Fehler (1998) and Margerin (2004). Other results are detailed in Shang and Gao (1988), Zeng *et al.*(1991), Sato (1993), and Sato *et al.*(1997). Analytical solutions are possible for certain idealized cases (e.g., Wu, 1985; Zeng, 1991; Sato, 1993) but obtaining general results requires computer calculations.

Two analytical results are of particular interest (and can be used as tests of numerical simulations). For the case of no intrinsic attenuation, Zeng (1991) showed the coda power converges to the diffusion solution at long lapse times

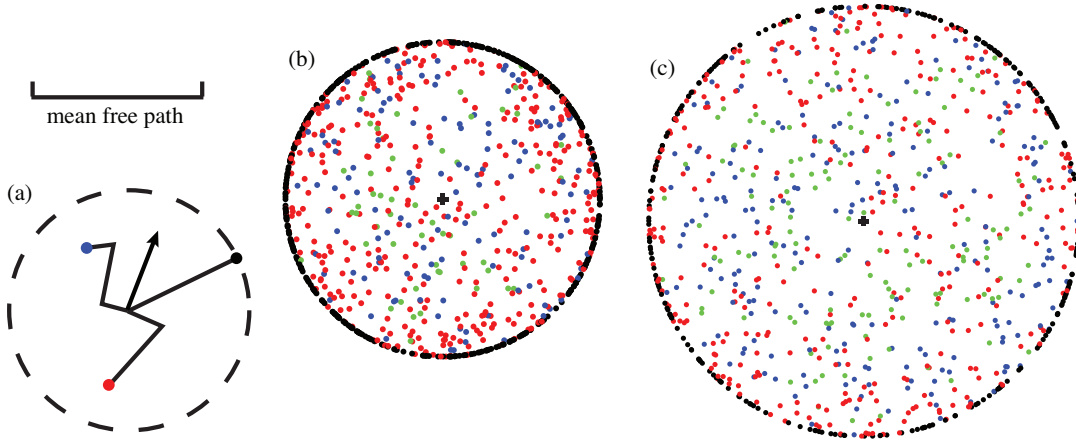$$P_C(t) \propto t^{-3/2} \tag{2.15}$$

Figure 2.5: Example of a Monte Carlo computer simulation of random scattering of seismic energy particles, assuming 2-D isotropic scattering in a uniform whole space. Particles are sprayed in all directions from the source with constant scattering probability defined by the indicated mean free path length, ℓ. As indicated in (a), black dots show particles that have not not been scattered, red dots show particles that have scattered once, blue dots show particles that have scattered twice, and green dots show particles scattered three or more times. (b) Results for 1000 particles after $t = 0.8\ell/v$, where v is velocity. (c) Results for 1000 particles after $t = 1.25\ell/v$. Note that the particle density is approximately constant for the scattered energy inside the circle defining the direct wavefront, as predicted by the energy flux model.

For elastic waves with no intrinsic attenuation, the equilibrium ratio of P and S energy density is given by (e.g., Sato, 1994; Ryzhik et al., 1996; Papanicolaou et al., 1996)

$$E_P/E_S = \tfrac{1}{2}(\beta/\alpha)^3 \tag{2.16}$$

Assuming a Poisson solid, this predicts about 10 times more S energy than P energy at equilibrium, a result of the relatively low efficiency of S-to-P scattering compared to P-to-S scattering (e.g., Malin and Phinney, 1985; Zeng, 1993). For media with intrinsic attenuation, an equilibrium ratio also exists but will generally differ from the purely elastic case (Margerin et al., 2001).

A powerful method for computing synthetic seismograms based on radiative transfer theory is to use a computer-based Monte Carlo approach to simulate the random walk of millions of seismic energy "particles" which are scattered with probabilities derived from random media theory. Figure 2.5 illustrates a simple exam-

ple of this method applied to 2-D isotropic scattering. Variations on this basic technique are described by Gusev and Abubakirov (1987), Abubakirov and Gusev (1990), Hoshiba (1991, 1994, 1997), Margerin *et al.*(2000), Bal and Moscoso (2000), Yoshimoto (2000), Margerin and Nolet (2003a,b), and Shearer and Earle (2004). All of these results suggest that body-wave scattering in the whole Earth can now be accurately modeled using ray theory and particle-based Monte Carlo methods. Although somewhat computationally intensive, continued improvements in computer speed make them practical to run on modest machines. They can handle multiple scattering over a range of scattering intensities, bridging the gap between the Born approximation for weak scattering and the diffusion equation for strong scattering. They also can include general depth-dependent or even 3-D variations in scattering properties, including non-isotropic scattering, without a significant increase in computation time compared to simpler problems.

2.3 Additional reading

Shearer, P. M., M. A. H. Hedlin, and P. S. Earle (1998). PKP and PKKP precursor observations: implications for the small-scale structure of the deep mantle and core, *The Core-Mantle Boundary Region, AGU Geodynamics Series*, **28**, 37–55.

Shearer, P. M. (2007). Seismic scattering in the deep Earth, in *Treatise on Geophysics, Volume 1: Deep Earth Structure*, Schubert, G. (ed.), Elsevier Ltd., Oxford, p. 695–730.

Shearer, P. M., and P. S. Earle (2008). Observing and modeling elastic scattering in the deep Earth, in *Advances in Geophysics, Volume 50: Earth Heterogeneity and Scattering Effects on Seismic Waves*, H. Sato and M. C. Fehler (ed.), Elsevier Ltd., Oxford, p. 167–193.

2.4 References

Abubakirov I R, Gusev A A 1990 Estimation of scattering properties of lithosphere of Kamchatka based on Monte-Carlo simulation of record envelope of a near earthquake. Phys. Earth Planet. Inter.64, 52-67

Aki K, Chouet B 1975 Origin of coda waves: source attenuation and scattering effects, J. Geophys. Res.80, 3322-3342

Aki, K. and P.G. Richards, *Quantitative seismology: theory and methods (volume 2)*, W.H. Freeman, San Francisco, 1980.

Baig A M, Dahlen F A 2004a Statistics of traveltimes and amplitudes in random media. Geophys. J. Int.158, 187-210

Baig A M, Dahlen F A 2004b Traveltime biases in random media and the S-wave discrepancy. Geophys. J. Int.158, 922-938

Baig AM, Dahlen F A, Hung S-H 2003 Traveltimes of waves in three-dimensional random media. Geophys. J. Int.153, 467-482

Bal G, Moscoso M 2000 Polarization effects of seismic waves on the basis of radiative transport theory. Geophys. J. Int.142, 571-585

Chernov, L.A., *Wave propagation in a random medium*, McGraw-Hill, New York, 1960.

Dahlen F A, Hung S-H, Nolet G 2000 Fréchet kernels for finite-frequency traveltimes— I. Theory. Geophys. J. Int.141, 157-174

Gusev A A, Abubakirov I R 1987 Monte Carlo simulation of record envelope of a near earthquake. Phys. Earth Planet. Inter.49, 30-36 bb Hoshiba M 1991 Simulation of multiple scattered coda wave excitation based on the energy conservation law. Phys. Earth Planet. Inter.67, 123-136

Hoshiba M 1993 Separation of scattering attenuation and intrinsic absorption in Japan using the multiple lapse time window analysis of full seismogram envelope. J. Geophys. Res.98, 15,809-15,824

Hoshiba M 1994 Simulation of coda wave envelope in depth dependent scattering and absorption structure. Geophys. Res. Lett.21, 2853-2856

Hoshiba M 1997 Seismic coda wave envelope in depth-dependent S wave velocity structure. Phys. Earth Planet. Inter.104, 15-22

Malin P E, Phinney R A 1985 On the relative scattering of P and S waves. Geophys. J.R. Astron. Soc.80, 603-618

Margerin L 2004 Introduction to radiative transfer of seismic waves. in *Seismic Data Analysis and Imaging with Global and Local Arrays, AGU Monograph Series*, American Geophysical Union

Margerin L, Campillo M, van Tiggelen B 2000 Monte Carlo simulation of multiple scattering of elastic waves. J. Geophys. Res.105, 7873-7892

Margerin L, Campillo M, van Tiggelen B 2001 Effect of absorption on energy partition of elastic waves. Bull. Seismol. Soc. Am.91, 624-627

Margerin L, Nolet G 2003a Multiple scattering of high-frequency seismic waves in the deep Earth: Modeling and numerical examples, J. Geophys. Res.108, B5, doi:10.1029/2002JB001974

Margerin L, Nolet G 2003b Multiple scattering of high-frequency seismic waves in the deep Earth:PKP precursor analysis and inversion for mantle granularity, J. Geophys. Res.108, B11, doi:10.1029/2003JB002455

Nolet G, Dahlen F A, Montelli R 2005 Travel times and amplitudes of seismic waves: a re-assessment. in *Seismic Earth: Analysis of broadband seismograms*, AGU Monograph Series, 37-48

Papanicolaou G, Ryshik L, Keller J 1996 On the stability of the P to S energy ratio in the diffusive regime. Bull. Seismol. Soc. Am.86, 1107-1115 (see also erratum in Bull. Seismol. Soc. Am.86, p. 1997)

Pekeris, C.L., Notes on the scattering of radiation in an inhomogeneous medium, *Physical Review*, **71**, 268, 1947.

Ryzhik L V, Papanicolaou G C, Keller J B 1996 Transport equations for elastics and other waves in random media. *Wave Motion* 24, 327-370

Sato H., Unified approach to amplitude attenuation and coda excitation in the randomly inhomogeneous lithosphere. *Pure Appl. Geophys.* 132, 93-121, 1990.

Sato, H 1993 Energy transportation in one- and two-dimensional scattering media: Analytic solutions of the multiple isotropic scattering model. Geophys. J. Int.112, 141-146

Sato, H 1994 Multiple isotropic scattering model including P-S conversions for the seismogram envelope formulation. Geophys. J. Int.117, 487-494

Sato, H. and M.C. Fehler, *Seismic wave propagation and scattering in the heterogeneous Earth*, Springer-Verlag, New York, 1998.

Sato H, Nakahara H, Ohtake M 1997 Synthesis of scattered energy density for non-spherical radiation from a point shear-dislocation source based on the radiative transfer theory. Phys. Earth Planet. Inter.104, 1-13

Shang T, Gao L 1988 Transportation theory of multiple scattering and its application to seismic coda waves of impulsive source. *Scientia Sinica* 31, 1503-1514

Shearer, P M, Earle P S 2004 The global short-period wavefield modelled with a Monte Carlo seismic phonon method. Geophys. J. Int.158, 1103-1117

Spetzler J, Snieder R 2001 The effect of small-scale heterogeneity on the arrival time of waves. Geophys. J. Int.145, 786-796

Wu R-S 1985 Multiple scattering and energy transfer of seismic waves separation of scattering effect from intrinsic attenuation I. Theoretical modelling. Geophys. J.R. Astron. Soc.82, 57-80

Wu, R. and K. Aki, Scattering characteristics of elastic waves by an elastic heterogeneity, *Geophysics*, **50**, 582–595, 1985.

Wu, R.S. and K. Aki, Elastic wave scattering by a random medium and the small-scale inhomogeneities in the lithosphere, *J. Geophys. Res.*, **90**, 10,261–10,273, 1985.

Wu R-S. Aki K 1988 Multiple scattering and energy transfer of seismic waves separation of scattering effect from intrinsic attenuation II. Application of the theory to Hindu Kush region. *Pure Appl. Geophys.* 128, 49-80

Yoshimoto K 2000 Monte Carlo simulation of seismogram envelopes in scattering media. J. Geophys. Res.105, 6153-6161

Zeng Y 1991 Compact solutions for multiple scattered wave energy in the time domain. Bull. Seismol. Soc. Am.81, 1022-1029

Zeng Y 1993 Theory of scattered P- and S-wave energy in a random isotropic scattering medium. Bull. Seismol. Soc. Am.83, 1264-1276

Zeng Y, Su F, Aki K 1991 Scattered wave energy propagation in a random isotropic scattering medium. J. Geophys. Res.96, 607-619

Chapter 3

Earthquake location methods

The problem of locating earthquakes from travel time data is one of the oldest challenges in seismology and continues to be an important component of seismic research. Earthquakes are defined by their *origin times* and *hypocenters*. The hypocenter is the (x, y, z) location of the event, while the *epicenter* is defined as the (x, y) point on the Earth's surface directly above the hypocenter. Earthquakes are generally treated as point sources in location methods. For large earthquakes that rupture from tens to hundreds of kilometers, the hypocenter is not necessarily the "center" of the earthquake. Rather it is the point at which seismic energy first begins to radiate at the beginning of the event. Since the rupture velocity is less than the P-wave velocity, the hypocenter can be determined from the first arrival times regardless of the eventual size and duration of the event. Earthquake information given in standard catalogs, such as the Preliminary Determination of Epicenters (PDE), is based on travel times of high-frequency body wave phases. These origin times and hypocenters should not be confused with long-period inversion results, which often give a *centroid* time and location for the event, representing the "average" time and location for the entire event.

Four parameters describe the origin time and hypocenter. Let's call these parameters the model, and define a model vector

$$\mathbf{m} = (m_1, m_2, m_3, m_4) = (T, x, y, z). \tag{3.1}$$

Now suppose we are given n observations of travel times, t_i, at individual seismic stations. In order to invert these times for the earthquake parameters, $\mathbf{m}$, we first

must assume a reference Earth model. For every value of $\mathbf{m}$ we can then calculate ranges to the ith station and compute predicted arrival times,

$$t_i^p = F_i(\mathbf{m}), \tag{3.2}$$

where $\mathbf{F}$ is the operator that gives the predicted arrival time at each station from $\mathbf{m}$. The difference between the observed and predicted times is

$$r_i = t_i - t_i^p = t_i - F_i(\mathbf{m}), \tag{3.3}$$

where r_i is the residual at the ith station. We wish to find the $\mathbf{m}$ that, in some sense, gives the smallest residuals between the observed and predicted times. Note that $\mathbf{F}$ is a function both of the Earth model and of the individual station locations. Most importantly, $\mathbf{F}$ is a *nonlinear* function of the model parameters (with the exception of the origin time T). In practice, for 1-D Earth models, $\mathbf{F}(\mathbf{m})$ is not particularly difficult to calculate, since the arrival times can be interpolated at the appropriate ranges from known travel time tables for the reference velocity model. However, the nonlinear dependence of the travel times on the earthquake location parameters greatly complicates the task of inverting for the best earthquake model. This nonlinearity is apparent even in the simple example of 2-D location within a plane of uniform velocity. The travel time from a station with coordinates $(x_i,\ y_i)$ to a point (x, y) is given by

$$t_i = \frac{\sqrt{(x - x_i)^2 + (y - y_i)^2}}{v}, \tag{3.4}$$

where v is the velocity. Clearly t does not scale linearly with either x or y in this equation. The result is that we cannot use standard methods of solving a system of linear equations to obtain a solution. Given a set of travel times to the stations, there is no single-step approach to finding the best event location.

Before discussing practical location strategies, it is instructive to consider what we might do if an infinite amount of computer power were available. In this case, we could perform a grid search over all possible locations and origin times and compute the predicted arrival times at each station. We could then find the particular $\mathbf{m}$ for which the predicted times t_i^p and the observed times t_i were in best agreement. How

do we define "best" agreement? A popular choice is least squares, that is, we seek to minimize

$$\epsilon = \sum_{i=1}^{n} [t_i - t_i^p]^2, \tag{3.5}$$

where n is the number of stations. The average squared residual, ϵ/n, is called the *variance*; thus we are trying to minimize the variance of the residuals. A common term that you may hear in describing models is *variance reduction* ("I got a 50% variance reduction with just two parameters" or "Their model only gives a 5% variance reduction in the raw data"). Here we use the term variance loosely to describe the spread in the residuals, independently of the number of free parameters in the fitting procedure. More formally, in statistics the variance is defined as ϵ/n_{df}, where n_{df} is the number of degrees of freedom (n_{df} is n minus the number of free parameters in the fit). For typical problems the number of fitting parameters is much less than the number of data, and so n and n_{df} are approximately equal.

Least squares is often used as a measure of misfit since it leads to simple analytical forms for the equations in minimization problems. It will tend to give the right answer if the misfit between t and t^p is caused by uncorrelated, random Gaussian noise in t. However, in many instances the errors are non-Gaussian, in which case least squares will give too much weight to the *outliers* in the data (a residual of 2 contributes 4 times more to the misfit than a residual of 1). As an alternative, we could use the sum of the differences

$$\epsilon = \sum_{i=1}^{n} |t_i - t_i^p| . \tag{3.6}$$

This measure of misfit is called the *L1 norm* and is considered more robust than the *L2 norm* (least squares) when excessive outliers are present in the data. For a distribution of numbers, the minimum *L2* norm yields the mean or average of the numbers, while the minimum *L1* norm gives the median value. The *L1* norm is not often used because the absolute value sign creates great complications in the equations. As an alternative to robust norms such as *L1*, it is possible to weight the residuals in the least squares problem using an iterative procedure that reduces the influence of the outlying points in subsequent steps. Of course in the case of our hypothetical "brute force" grid search it is straightforward to apply any norm that

Table 3.1: Percentage points of the χ^2 distribution.

n_{df}	$\chi^2(95\%)$	$\chi^2(50\%)$	$\chi^2(5\%)$
5	1.15	4.35	11.07
10	3.94	9.34	18.31
20	10.85	19.34	31.41
50	34.76	49.33	67.50
100	77.93	99.33	124.34

we desire. Once we have defined a measure of misfit, we can find the "best" $\mathbf{m}$ as the one with the smallest misfit, $\epsilon(\mathbf{m})$. The next step is to estimate the probable uncertainties in our location.

Some indication of these uncertainties can be seen in the behavior of the misfit function in the vicinity of its minimum. In our two-dimensional example, suppose that we contour $\epsilon(\mathbf{m})$ as a function of x and y, assuming that the origin time is known (since the t^p are a linear function of the origin time, determination of the best origin time for a given location is trivial). Clearly, if ϵ grows rapidly as we move away from the minimum point, we have resolved the location to better accuracy than when ϵ grows only very slowly away from its minimum.

How can we quantify this argument? By far the most common approach is based on least squares and the $L2$ norm, since the statistics of Gaussian processes are well understood. In this case we define

$$\chi^2 = \sum_{i=1}^{n} \frac{[t_i - t_i^p]^2}{\sigma_i^2}, \tag{3.7}$$

where σ_i is the expected standard deviation of the ith residual due to random measurement error. The expected value of χ^2 is approximately the number of degrees of freedom n_{df} (in our case $n_{df} = n - 4$ because $\mathbf{m}$ has 4 components) and 95% confidence limits may be obtained by consulting standard statistical tables (e.g., Table 3.1).

For example, if we locate an earthquake using 14 travel times, then $n_{df} = 10$ and there is a 90% probability that the value of χ^2 computed from the residuals at the best fitting hypocenter will be between 3.94 and 18.31. There is only a 5% chance that the value of χ^2 will exceed 18.31. The value $\chi^2(\mathbf{m})$ will grow as we move away

from the best-fitting location, and by contouring values of $\chi^2(\mathbf{m})$ we can obtain an estimate of the 95% error ellipse for the event location.

Note that the σ_i values are critical in this analysis—the statistics are based on the data misfit being caused entirely by random, uncorrelated Gaussian errors in the individual travel time measurements. However, the misfit in earthquake location problems is usually larger than would be expected from timing and picking errors alone. If the σ_i are set significantly smaller than the average residual, then the χ^2 measure may indicate that the solution should be rejected, most likely because unmodeled velocity structure is dominating the misfit. Alternatively, if the σ_i are set significantly larger than the average residual, then the best-fitting hypocenter could be rejected because it fits the data "too well."

To avoid these embarrassments, the estimated data uncertainties σ_i are often estimated from the residuals at the best location,

$$\sigma^2(\mathbf{m}_{\text{best}}) = \frac{\sum_{i=1}^{n} \left[t_i - t_i^p(\mathbf{m}_{\text{best}})\right]^2}{n_{df}}, \tag{3.8}$$

where $\mathbf{m}_{\text{best}}$ is the best-fitting location, and this constant value of σ^2 is used for all the σ_i^2 in (3.7), that is,

$$\chi^2(\mathbf{m}) = \frac{\sum_{i=1}^{n} \left[t_i - t_i^p(\mathbf{m})\right]^2}{\sigma^2}. \tag{3.9}$$

Note that $\chi^2(\mathbf{m}_{\text{best}}) = n_{df}$ so that the χ^2 value at the best-fitting hypocenter is close to the 50% point in the χ^2 distribution. By contouring $\chi^2(\mathbf{m})$, we can then obtain an estimate of the 95% confidence ellipse for the solution; that is, we can approximate the region within which there is a 95% chance that the true location lies.[2]

However, a serious problem with typical confidence limits is that they don't take into account the correlated changes to travel time residuals resulting from un-modeled lateral heterogeneity. For example, consider a model in which a vertical fault separates the crust into two blocks with slightly different velocities (Fig. 3.1). Events occurring on the fault will tend to be mislocated off the fault into the faster velocity block owing to a systematic bias in the travel times. This possibility is

---

[2]The error ellipse is only approximate because the uncertainties in the σ_i estimate are ignored.
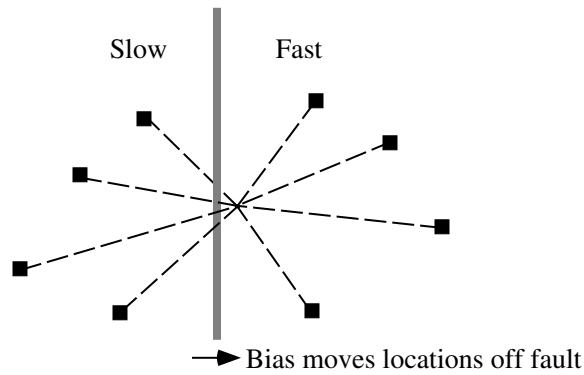
Figure 3.1: Earthquakes located along a fault will often be mislocated if the seismic velocity changes across the fault.
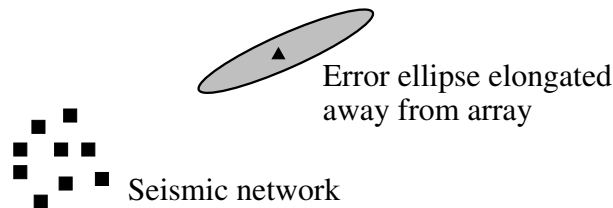

Figure 3.2: Earthquake locations for events outside of a network are often not well constrained.

not accounted for in the formal error analysis, which, in this case, incorrectly assumes that the travel time uncertainties are *uncorrelated* between different stations. The effects of unmodeled lateral heterogeneity are the dominant source of error for earthquake locations, provided a good station distribution is used in the inversion. Global locations in the ISC and PDE catalogs are typically off by about 25 km in horizontal position and depth (assuming depth phases such as pP are used to constrain the depth; if not, the probable depth errors are much greater). Techniques that can be used to improve earthquake locations include joint hypocenter velocity inversion and master event methods.

When a good station distribution is not available, location errors can be quite large. For example, the distance to events occurring outside of a seismic array is not well constrained, since there is a large tradeoff between range and origin time (Fig. 3.2). In this case, the location could be improved dramatically if a travel time was available from a station on the opposite side of the event. Generally it is best to have a good azimuthal distribution of stations surrounding an event to avoid these kinds of location uncertainties. Another problem is the tradeoff between

event depth and origin time that occurs when stations are not available at close ranges (Fig. 3.3). Since the takeoff angles of the rays are very similar, changes in the earthquake depth may be compensated for by a shift in the event origin time.

In the preceding examples, we have assumed that only direct P-wave data are available. The addition of other phases recorded at the same stations can substantially improve location accuracy, since the use of *differential times* between phases removes the effect of the earthquake origin time. For example, S arrivals travel at a different speed than P arrivals and can be used to estimate the source–receiver range at each station directly from the $S - P$ time (a convenient rule of thumb for crustal phases is that the distance to the event in kilometers is about 8 times the $S - P$ time in seconds). Even better than S for determining earthquake depths from teleseismic data is the depth phase pP since the differential time $pP - P$ is very sensitive to the earthquake depth.

3.0.1 Iterative location methods

In our discussion so far we have assumed that the minimum ϵ could be found directly by searching over all $\epsilon(\mathbf{m})$. In practice, this often becomes computationally unfeasible and less direct methods must be employed. The standard technique is to *linearize* the problem by considering small perturbations to a target location

$$\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}, \tag{3.10}$$

where $\mathbf{m}_0$ is the current guess as to the best location and $\mathbf{m}$ is a new location a small distance away from $\mathbf{m}_0$. The predicted times at $\mathbf{m}$ may be approximated using the
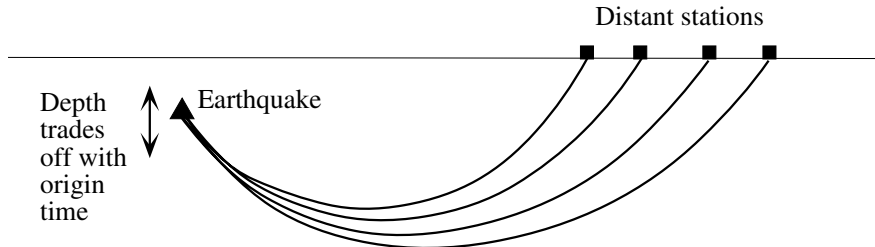

Figure 3.3: Earthquake depth can be hard to determine if only distant stations are available.

first term in the Taylor series expansion

$$t_i^p(\mathbf{m}) = t_i^p(\mathbf{m}_0) + \frac{\partial t_i^p}{\partial m_j} \Delta m_j. \tag{3.11}$$

The residuals at the new location $\mathbf{m}$ are given by

$$
\begin{aligned}
r_i(\mathbf{m}) &= t_i - t_i^p(\mathbf{m}) \\
&= t_i - t_i^p(\mathbf{m}_0) - \frac{\partial t_i^p}{\partial m_j} \Delta m_j \\
&= r_i(\mathbf{m}_0) - \frac{\partial t_i^p}{\partial m_j} \Delta m_j.
\end{aligned}
\tag{3.12}
$$

In order to minimize these residuals we seek to find $\Delta \mathbf{m}$ such that

$$r_i(\mathbf{m}_0) = \frac{\partial t_i^p}{\partial m_j} \Delta m_j \tag{3.13}$$

or

$$\mathbf{r}(\mathbf{m}_0) = \mathbf{G}\,\Delta \mathbf{m}, \tag{3.14}$$

where $\mathbf{G}$ is the matrix of partial derivatives $G_{ij} = \partial t_i^p / \partial m_j$, $i = 1, 2, ..., n$, $j = 1, ..., 4$. The best fit to Equation (3.14) may be obtained using standard least squares techniques to obtain the location adjustment $\Delta \mathbf{m}$. Next, we set $\mathbf{m}_0$ to $\mathbf{m}_0 + \Delta \mathbf{m}$ and repeat the process until the location converges. This iterative procedure generally converges fairly rapidly provided the initial guess is not too far from the actual location.

3.0.2   Relative event location methods

In the common situation where the location error is dominated by the biasing effects of unmodeled 3-D velocity structure, the relative location among events within a localized region can be determined with much greater accuracy than the absolute location of any of the events. This is because the lateral velocity variations outside the local region, which lead to changes in the measured travel times at distant stations, will have nearly the same effect on all of the events. In other words, the residuals caused by 3-D structure to a given station will be correlated among all of the events. If the ray path to a station is anomalously slow for one event, then it will be slow for the other events as well, provided the local source region is small

compared to the heterogeneity. However, the bias in the locations caused by the 3-D structure will vary among the events because they typically do not have picks from exactly the same set of stations.

The simplest way to improve relative location accuracy among nearby earthquakes is to consider differential times relative to a designated *master event*. The arrival times of other events relative to the master event times are

$$t^{\mathrm{rel}} = t - t^{\mathrm{master}}. \tag{3.15}$$

Setting the master event location to $\mathbf{m}_0$ in Equation (3.13), we see that the relative location $\Delta\mathbf{m}$ is given by the best-fitting solution to

$$t_i^{\mathrm{rel}} = t_i^p(\mathbf{m}) - t_i^p(\mathbf{m}_0) = \frac{\partial t_i^p}{\partial m_j}\Delta m_j, \tag{3.16}$$

where the solution will be valid provided $\Delta\mathbf{m}$ is small enough that the linear approximation holds. This approach works because the differential times subtract out any travel-time perturbations specific to a particular station. Note that the absolute location accuracy is limited by the location accuracy of the master event, which is assumed fixed. However, if the absolute location of the master event is known by other means (e.g., a surface explosion), then these relative locations can also be converted to absolute locations.

This approach can be generalized to optimally relocate events within a compact cluster with respect to the cluster centroid by projecting out the part of the travel-time perturbations that are common to particular stations, a method termed *hypocentroidal decomposition* by Jordan and Sverdrup (1981). A simpler technique is to compute station terms by averaging the residuals at each station, recompute the locations after correcting the observed picks for the station terms, and iterate until a stable set of locations and station terms is obtained (e.g., Frohlich, 1979). It can be shown that this iterative approach converges to the same solution as hypocentroidal decomposition (Lin and Shearer, 2005).

These ideas can be generalized to distributed seismicity where the effect of 3-D velocity structure on travel times will vary among different source regions. The *double-difference* location algorithm (Waldhauser and Ellsworth, 2000; Waldhauser,
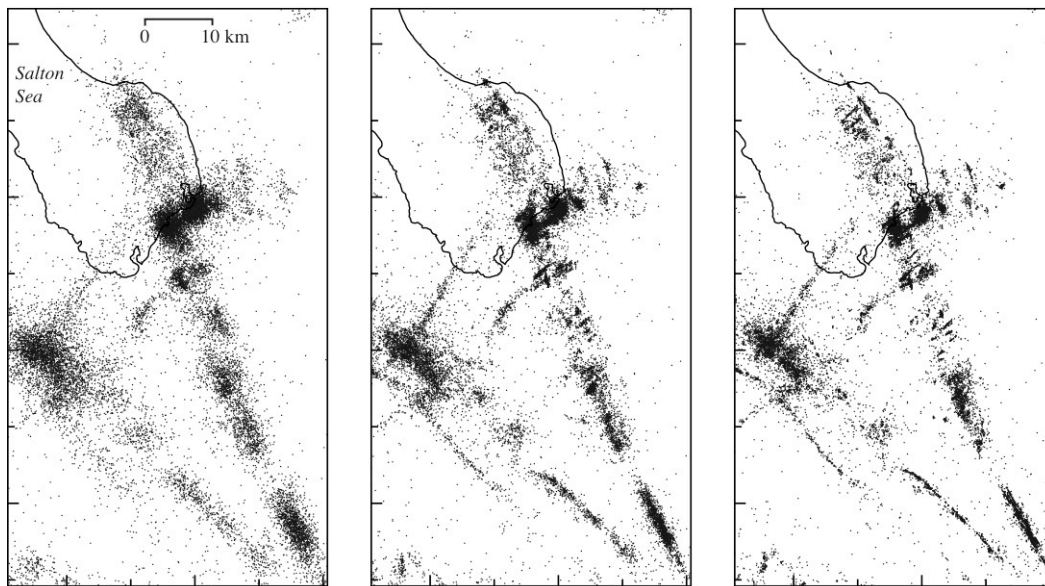
Figure 3.4: Earthquake locations for over 17,000 earthquakes in the Imperial Valley, California (1981–2005), as computed using: (left) single event location, (middle) source-specific station term (SSST) locations, and (right) waveform cross-correlation locations using results from Lin et al. (2007).

2001) performs simultaneous relocation of distributed events by minimizing the residual differences among nearby events. The *source-specific station term* (SSST) method (Richards-Dinger and Shearer, 2000; Lin and Shearer, 2006) iteratively computes spatially varying time corrections to each station. Further improvements in relative location accuracy can be achieved by using waveform cross-correlation to compute more accurate differential times among nearby events than can be measured using arrival time picks on individual seismograms. Figure 3.4 illustrates the improvement in local earthquake locations that can be achieved using these methods compared to classic single event location. Note the increasingly sharp delineation of seismicity features that is obtained using source-specific station terms and waveform cross-correlation.

3.0.3   How does the SSST method work?

The source-specific station term (SSST) method is an extension of the station term approach (e.g., Frohlich, 1979) to the case of distributed seismicity. First, consider a small cluster of events embedded within some unknown 3-D velocity structure,
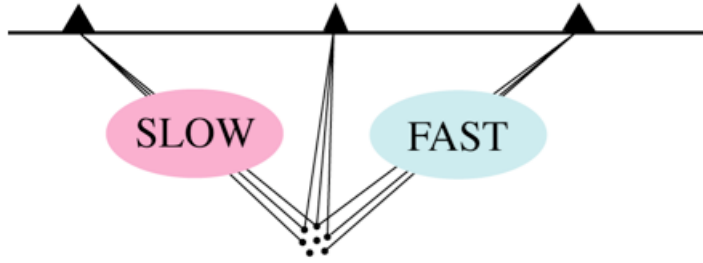
Figure 3.5: Events within a single small event cluster will have correlated travel time residuals to each station.
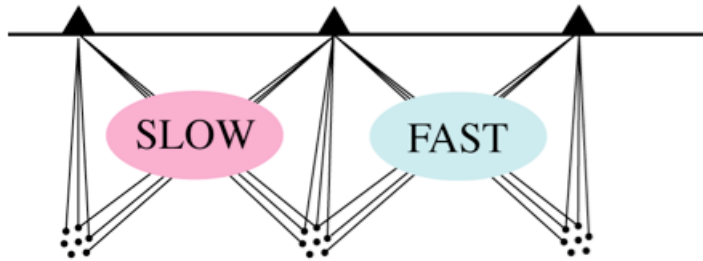

Figure 3.6: The residuals to each station will vary among different event clusters.

which are located using a 1-D velocity model (see Fig. 3.5). The 3-D structure will cause some rays to travel faster than the model and some slower than the model, which will bias the locations. Assuming the cluster is small compared to the 3-D structure, the relative locations of the events within the cluster can be improved through the use of station terms, which are computed by averaging the residuals at each station, recomputing the locations after correcting the observed picks for the station terms, and iterating until a stable set of locations and station terms is obtained.

Now consider several small clusters in different locations (Fig. 3.6). In this case, the biasing effects are different for each cluster and a different set of station terms would need to be computed for each cluster. This is straightforward, but the situation gets more complicated for distributed seismicity (Fig. 3.7). In this case the timing correction terms vary continuously as a function of event position. One approach would be to divide the crust into rectangular boxes and compute a different set of station terms for the events within each box. But this could cause edge artifacts as the locations change across the boundaries between boxes.
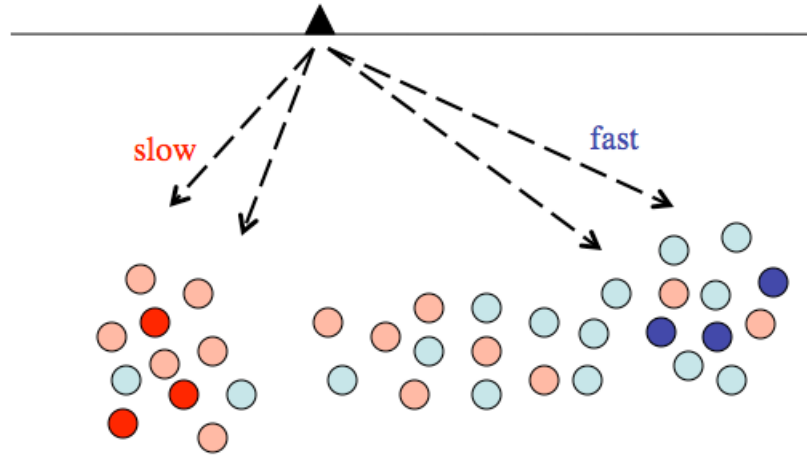
Figure 3.7: Travel time residuals to a single station will have some amount of random scatter and some systematic variation with event location. Negative (fast) residuals are shown in blue; positive (slow) residuals are shown in red. The average residual will be positive for paths through slow material and negative for paths through fast material.

A more flexible approach is to recognize that the residuals for each station contain uncorrelated contributions from random timing and picking errors and spatially correlated contributions from 3-D velocity structure. The goal of the SSST approach is to remove the effect of the spatially correlated residuals by subtracting a smoothed version of the residual field from the travel times.

The SSSTs are computed by smoothing the residual field over some specified smoothing radius r (see Fig. 3.8). There is a different set of SSSTs for every station, and every event will have a slightly different SSST (unless the events are in exactly the same location). The travel time data are then corrected for the SSSTs and the events relocated. Just as in the single station term method, the process is repeated until it converges to a stable set of locations and SSSTs. Note that in the limit of large r, the method is the same as the single station term approach (e.g., Frohlich, 1979). In the limit of small r, the method will have no effect because no smoothing of residuals among nearby events will take place.

In practice the best results seem to be obtained if r is set large for the first iteration and then is gradually shrunk with iteration number. This is called the "shrinking box" SSST method (Lin and Shearer, 2005).
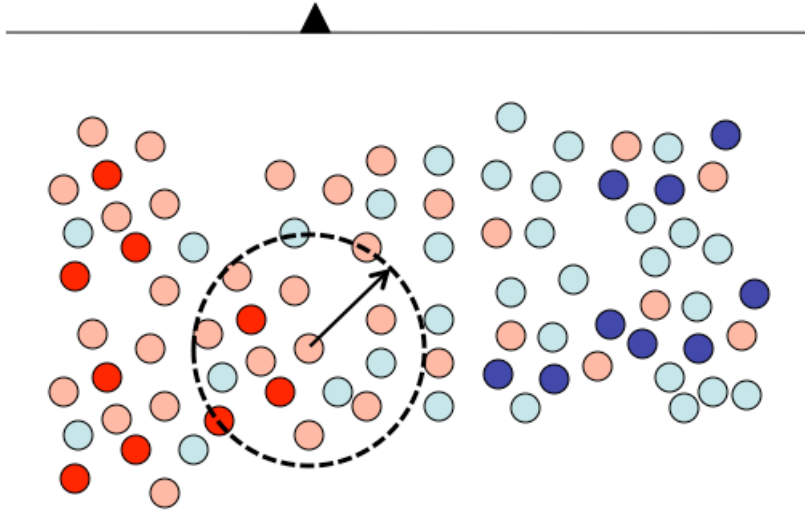
Figure 3.8: Source-specific station terms (SSSTs) are computed by smoothing the residual field over some specified smoothing radius.

3.1 EXERCISES

1. (COMPUTER) You are given P-wave arrival times for two earthquakes recorded by a 13-station seismic array. The station locations and times are listed in Table 5.2 and also given in the supplemental web material.

   (a) Write a computer program that performs a grid search to find the best location for these events. Try every point in a 100 km by 100 km array ($x = 0$ to 100 km, $y = 0$ to 100 km). At each point, compute the range to each of the 13 stations. Convert these ranges to time by assuming the velocity is 6 km/s (this is a 2-D problem, don't worry about depth). Compute the average sum of the squares of the residuals to each grid point (after finding the best-fitting origin time at the grid point; see below).

   (b) For each quake, list the best-fitting location and origin time.

   (c) From your answers in (b), estimate the uncertainties of the individual station residuals (e.g., σ^2 in 3.8) for each quake.

   (d) For each quake, use (c) to compute χ^2 at each of the grid points. What is χ^2 at the best fitting point in each case?

Table 3.2: P-arrival times for two earthquakes.

| | | Quake 1 | Quake 2 |
x (km)	y (km)	$t1$ (s)	$t2$ (s)
9.0	24.0	14.189	20.950
24.0	13.2	13.679	21.718
33.0	4.8	13.491	21.467
45.0	10.8	14.406	21.713
39.0	27.0	13.075	20.034
54.0	30.0	15.234	20.153
15.0	39.0	13.270	18.188
36.0	42.0	12.239	16.008
27.0	48.0	12.835	15.197
48.0	48.0	14.574	16.280
15.0	42.0	12.624	16.907
18.0	15.0	13.496	21.312
30.0	36.0	10.578	16.664

(e) Identify those values of χ^2 that are within the 95% confidence ellipse. For each quake, make a plot showing the station locations, the best quake location, and the points within the 95% confidence region.

(f) Note: Don't do a grid search for the origin time! Instead assume an origin time of zero to start; the best-fitting origin time at each grid point will be the average of the residuals that you calculate for that point. Then just subtract this time from all of the residuals to obtain the final residuals at each point.

3.2   Additional reading

Lin, G. and P. Shearer, Tests of relative earthquake location techniques using synthetic data, *J. Geophys. Res.*, 110, B4, B04304, doi:10.1029/2004JB003380, 2005.

Lin, G. and P. Shearer, The COMPLOC earthquake location package, *Seismol. Res. Lett.*, 77, 440-444, 2006.

Lin, G., P. M. Shearer and E. Hauksson, Applying a three-dimensional velocity model, waveform cross correlation, and cluster analysis to locate southern California seismicity from 1981 to 2005, *J. Geophys. Res.*, 112, B12309, doi: 10.1029/2007JB004986, 2007.

3.3 References

Frohlich, C. (1979). An efficient method for joint hypocenter determination for large groups of earthquakes, *Comp. Geosci.* **5,** 387-389.

Jordan, T. H., and K. A. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. Seism. Soc. Am.* **71,** 1105-1130.

Richards-Dinger, K., and P. Shearer (2000). Earthquake locations in southern California obtained using source-specific station terms, *J. Geophys. Res.* **105,** 10,939-10,960.

Waldhauser, F., and W. L. Ellsworth (2000). A double-difference earthquake location algorithm: method and application to the Northern Hayward Fault, CA, *Bull. Seism. Soc. Am.* **90,** 1353-1368.

Waldhauser, F. (2001). hypoDD: A program to compute double-difference hypocenter locations (hypoDD version 1.0, 3/2001), *U.S. Geol. Surv. open-File Rept. 01-113.*

Chapter 4

Seismic tomography

Observed travel times typically exhibit some scatter compared to the times predicted by even the best reference 1-D Earth model. The travel time *residual* may be computed by subtracting the predicted time from the observed time, $t_{\text{resid}} = t_{\text{obs}} - t_{\text{pred}}$. Negative residuals result from early arrivals indicative of faster-than-average structure, while positive residuals are late arrivals suggestive of slow structure. Residuals within a selected range window are often plotted as a histogram to show the spread in residuals. If the average residual is nonzero, as in the example below, this indicates that the reference 1-D velocity model may require some adjustment.

The spread in the residual histogram can be modeled as the sum of two parts: (1) random scatter in the times due to picking errors and (2) systematic travel time differences due to lateral heterogeneity. The goal of 3-D velocity inversion techniques is to resolve the lateral velocity perturbations. These techniques are now commonly called *seismic tomography* by analogy to medical imaging methods such as CAT scans. However, it is worth noting that 3-D seismic velocity inversion is much more complicated than the medical problem. This is due to several factors: (1) Seismic

47

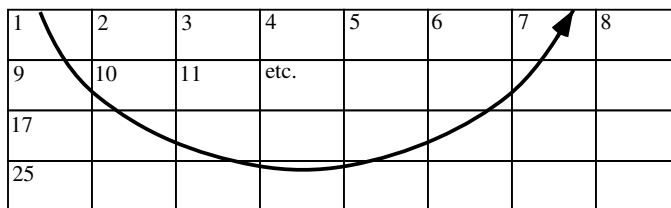| 1	2	3	4	5	6	7	8
9	10	11	etc.				
17							
25							
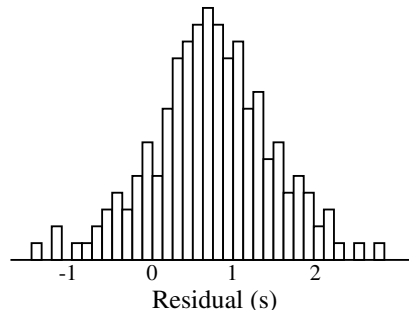
Figure 4.1: An example ray path and cell numbering scheme for a simple 2-D tomography problem.

ray paths generally are not straight and are a function of the velocity model itself, (2) the distribution of seismic sources and receivers is sparse and nonuniform, (3) the locations of the seismic sources are not well known and often trade off with the velocity model, and (4) picking and timing errors in the data are common.

Thus the analogy to medical tomography can be misleading when seismologists speak of *imaging* Earth structure, since the term "image" implies a rather direct measurement of the structure, whereas, in practice, seismic velocity inversion usually requires a number of modeling assumptions to deal with the difficulties listed above. It is comparatively easy to produce an image of apparent 3-D velocity perturbations; the more challenging task is to evaluate its statistical significance, robustness, and resolution.

4.0.1   Setting up the tomography problem

Assuming that a reference 1-D model is available, the next step is to parameterize the model of 3-D velocity perturbations. This is commonly done in two different ways: (1) the model is divided into blocks of uniform velocity perturbation or (2) spherical harmonic functions can be used in the case of global models to parameterize lateral velocity perturbations, with either layers or polynomial functions used to describe vertical variations.

As an example, we now illustrate the block parameterization in the case of body waves. Consider a two-dimensional geometry with the model divided into blocks as shown in Figure 4.1. For each travel time residual, there is an associated ray path that connects the source and receiver. Finding this exact ray path comprises the *two-point ray tracing problem*, and this can be a nontrivial task, particularly in the case of iterative tomography methods in which rays must be traced through

3-D structures. Methods for solving the two-point ray tracing problem include: (1) *ray shooting* in which slightly different take-off angles at the source are sampled in order to converge on the correct receiver location, (2) *ray bending* in which a nearby ray path is slightly deformed to arrive at the desired receiver location, or (3) finite difference or graph theory techniques that require a grid of points (e.g., Vidale, 1988; Moser, 1991). Fortunately, Fermat's principle suggests that we do not have to get precisely the right ray path to obtain the correct travel time—getting reasonably close should suffice, since, to first order, the travel times are insensitive to perturbations in the ray paths.

Once we have determined the geometry of the ray path, the next step is to find the travel time through each block that the ray crosses (although in principle this is straightforward, programming this on the computer can be a substantial chore!). The total travel time perturbation along the ray path is then given by the sum of the product of each block travel time with the fractional velocity perturbation within the block. In other words, the travel time residual r can be expressed as

$$r = \sum_k b_k v_k, \tag{4.1}$$

where b_k is the ray travel time through the kth block and v_k is the fractional velocity perturbation in the block (note that v_k is unitless, with $v_k = -0.01$ for 1% fast, $v_k = 0.01$ for 1% slow, etc.). The ray paths and the b_k values are assumed to be fixed to the values obtained from ray tracing through the reference model. Note that the velocity perturbations v_k are constant within individual blocks, but the velocity within each block may not be constant if the reference 1-D model contains velocity gradients. Since velocity perturbations will affect the ray paths, Equation (5.16) represents an approximation that is accurate only for small values of v_k.

If we set the ray travel times for the blocks not encountered by the ray to zero, we can express the travel time residual for the ith ray path as:

$$r_i = \sum_{j=1}^{m} b_{ij} v_j, \tag{4.2}$$

where m is the total number of blocks in the model. Note that most of the values of b_{ij} are zero since each ray will encounter only a small fraction of the blocks in the

model. For n travel time measurements, this becomes a matrix equation:

$$
\begin{bmatrix} r_1 \\ r_2 \\ r_3 \\ . \\ . \\ . \\ r_n \end{bmatrix} = \begin{bmatrix} 0 & 0 & 0 & 0 & 0.8 & \cdots \\ 0 & 0.6 & 0 & 1.3 & 0 & \cdots \\ 0.1 & 0 & 0 & 0 & 0 & \cdots \\ . & . & . & . & . & \cdots \\ . & . & . & . & . & \cdots \\ . & . & . & . & . & \cdots \\ 0 & 0 & 0.7 & 0 & 0 & \cdots \end{bmatrix} \begin{bmatrix} v_1 \\ v_2 \\ . \\ . \\ . \\ v_m \end{bmatrix},
\tag{4.3}
$$

where the numbers are examples of individual ray travel times through particular blocks. This can be written as

$$
\mathbf{d} = \mathbf{Gm}
\tag{4.4}
$$

using the conventional notation of $\mathbf{d}$ for the data vector, $\mathbf{m}$ for the model vector, and $\mathbf{G}$ for the linear operator that predicts the data from the model. The numbers in $\mathbf{G}$ are the travel times for each ray through each block. $\mathbf{G}$ will generally be extremely sparse with mostly zero elements. In the case shown, the number of travel time observations is greater than the number of model blocks ($n > m$), and, in principle, the problem is overdetermined and suitable for solution using standard techniques. The least squares solution to (5.1) is

$$
\mathbf{m} = (\mathbf{G}^T\mathbf{G})^{-1}\mathbf{G}^T\mathbf{d}.
\tag{4.5}
$$

In tomography problems this formula can almost never be used since the matrix $\mathbf{G}^T\mathbf{G}$ is invariably singular or extremely ill-conditioned. Some of the ray paths may be nearly identical while some of the blocks may not be sampled by any of the ray paths. These difficulties can be reduced in the case of small matrices with linear algebra techniques such as singular value decomposition (SVD). More commonly, however, m is so large that direct matrix inversion methods cannot be used. In either case, it will typically turn out that there is no unique solution to the problem—there are too many undersampled blocks and/or tradeoffs in the perturbations between different blocks.

A common approach to dealing with ill-posed least squares problems is to impose additional constraints on the problem, a process referred to as regularization. One example of regularization is the *damped least squares* solution in which (5.1) is

replaced with

$$\begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \lambda\mathbf{I} \end{bmatrix} \mathbf{m}, \tag{4.6}$$

where $\mathbf{I}$ is the identity matrix and λ is a weighting parameter that controls the degree of damping. The least squares solution to this problem will minimize the functional

$$\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda^2\|\mathbf{m}\|^2,$$

where the first term is the misfit to the data and the second term is the variance of the model. By adjusting the parameter λ we can control the tradeoff between misfit and model variance. These constraints add stability to the inversion—perturbations in blocks that are not sampled by rays will go to zero; anomalies will be distributed equally among blocks that are sampled only with identical ray paths. However, the damped least squares solution will not necessarily lead to a smooth model, since it is the size of the model, not its roughness, that is minimized. Model perturbations in adjacent blocks can be quite different.

A common measure of model roughness for block models is the *Laplacian operator* ∇^2, which can be approximated with a difference operator in both 2-D and 3-D block geometries. To minimize ∇^2 we replace $\mathbf{I}$ with $\mathbf{L}$ in (4.6):

$$\begin{bmatrix} \mathbf{d} \\ 0 \end{bmatrix} = \begin{bmatrix} \mathbf{G} \\ \lambda\mathbf{L} \end{bmatrix} \mathbf{m}, \tag{4.7}$$

where $\mathbf{L}$ is the finite difference approximation to the Laplacian applied over all model blocks. Each row of $\mathbf{L}$ is given by the difference between the target block and the average of the adjacent cells.

For example, in a 2-D model the Laplacian becomes

$$\nabla_j^2 \simeq \tfrac{1}{4}(m_{\text{left}} + m_{\text{right}} + m_{\text{up}} + m_{\text{down}}) - m_j,$$

where ∇_j^2 is the Laplacian of the jth model point. In this case the least squares inversion will minimize

$$\|\mathbf{Gm} - \mathbf{d}\|^2 + \lambda^2\|\mathbf{Lm}\|^2,$$

where λ controls the tradeoff between misfit and model roughness. This type of regularization adds stability to the inversion in a different way than damped least squares. The resulting models will be smooth, but not necessarily of minimum variance. Blocks that are not sampled by ray paths will be interpolated between nearby cells, or, more dangerously, extrapolated when they are near the edge of the model.

Both damped least squares and minimum roughness inversions have advantages and disadvantages, and the best regularization method to use will vary from problem to problem. In general, one should distrust damped least squares solutions that contain significant fine-scale structure at scale lengths comparable to the block dimensions, whereas minimum roughness solutions are suspect when they produce large-amplitude anomalies in regions constrained by little data.

We have so far assumed that all of the data are weighted equally. This is not always a good idea in tomography problems since travel time residuals are often non-Gaussian and plagued with outliers. This difficulty has been addressed in different ways. Often the residuals are first windowed to remove the largest outliers. Travel time residuals from similar ray paths are commonly averaged to form *summary ray* residuals before beginning the inversion. In iterative schemes the influence of anomalous data points can be downweighted in subsequent steps, thus simulating a more robust misfit norm than used in least squares.

4.0.2   Solving the tomography problem

For "small" problems (number of blocks in model $m < 500$ or so), conventional linear algebra methods such as Gauss reduction or singular value decomposition can be used to obtain exact solutions to Equations (4.6) or (4.7). In these cases, we have a significant advantage in that it is also practical to compute formal resolution and model covariance matrices. However, more commonly m is too large for such calculations to be practical. For example, a 3-D model parameterized by 100 blocks laterally and 20 blocks in depth contains 200,000 model points. Clearly we are not going to be able to invert directly a 200,000 by 200,000 matrix! Indeed we could not even fit such a matrix into the memory of our computer.
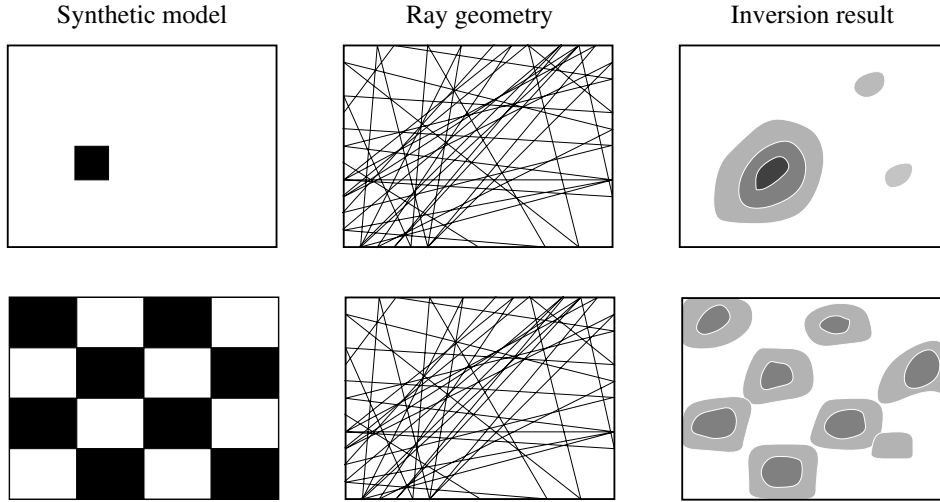
Synthetic model Ray geometry Inversion result

Figure 4.2: The resolution of tomographic models is often evaluated using the impulse response test (top) or the checkerboard test (bottom). In each, a synthetic set of travel times are created for a simple velocity model using the same ray paths present in the real data; then the synthetic times are inverted to see how well the starting model is recovered.

Thus, we must turn to iterative methods designed for large sparse systems of equations in order to solve these problems. Fortunately these have proven extremely useful in tomography problems and are found to converge fairly rapidly to close approximations to the true solutions. Examples of iterative methods include names such as ART-backprojection, SIRT, conjugate gradient, and LSQR (see Nolet, 1987, for a detailed discussion of many of these methods). Although it is instructive to see the form of equations such as (4.3) and (5.1), in practice we rarely attempt to construct $\mathbf{G}$ as a matrix. Rather we treat $\mathbf{G}$ as a linear operator that acts on the model to predict the data. On the computer, this often will take the form of a subroutine. Since the iterative techniques effectively use only one row of $\mathbf{G}$ at a time, they are sometimes given the name *row action methods*.

A disadvantage of these iterative solutions is that it becomes impossible to compute formal resolution and covariance matrices for the model. As substitutes for these measures, it has become common practice to conduct experiments on synthetic data sets. The synthetic data are generated by assuming a particular model of velocity perturbations and computing travel time anomalies using the same ray paths as the real data. The synthetic data are then inverted to see how well the test model is recovered (Fig. 4.2). One example of this procedure is the *impulse response*

test, in which a single localized anomaly is placed in a region of interest to see how well it can be resolved. Another method that is often applied is the *checkerboard test*, in which a model with a regular pattern of alternating fast and slow velocities is examined. In this case, the degree of smearing of the checkerboard pattern will vary with position in the model, giving some indication of the relative resolution in different areas.

It is not always clear that these tests give a reliable indication of the true resolution and uniqueness of the velocity inversions. Impulse response and checkerboard tests can be misleading because they typically assume uniform amplitude anomalies and perfect, noise-free data. In real tomography problems, the data are contaminated by noise to some degree and the velocity models that are obtained contain anomalies of varying amplitude. In these cases it is often only the higher amplitude features that are unambiguously resolved. In principle, some of these problems can be addressed using techniques that randomly resample the data (such as "jackknife" or "bootstrap" methods). However, these require repeating the inversion procedure up to 100 times or more, a significant obstacle in these computationally intensive analyses. Questions regarding the best way to evaluate resolution in tomographic inversions are not fully answered, and this continues to be an active area of research.

4.0.3   Tomography complications

In the preceding discussion it has been assumed that the source locations and origin times were precisely known. However, in the case of earthquakes this is rarely the case, and there is the potential for bias due to errors in the locations. Since the earthquakes are generally located using a reference 1-D velocity model, we would expect the locations to change given a 3-D velocity model, and indeed there is often a tradeoff between velocity anomalies and earthquake locations. This problem can be addressed by *joint hypocenter and velocity inversions* (JHV) that solve for both the earthquake locations and the velocity structure. In practice, for large inversions, this is often an iterative process in which initial earthquake locations are assumed, a velocity model is derived, the earthquakes are relocated using the new model, a new velocity model is derived, etc. Tradeoffs between quake locations and velocity
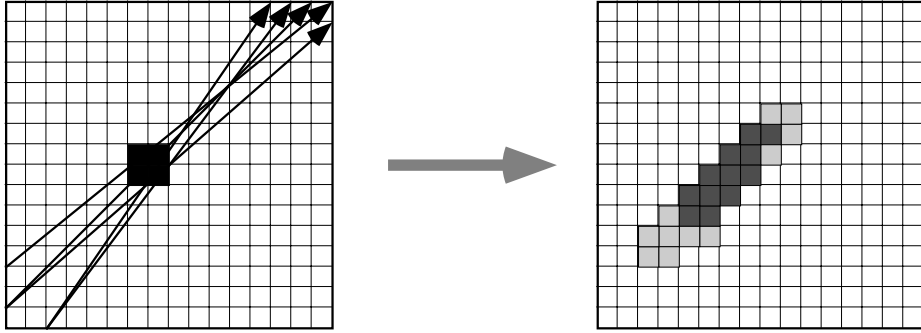
Figure 4.3: When only a limited range of ray angles are available, resolution of velocity anomalies is limited in the direction parallel to the rays.

structure will be minimized in this procedure, but only if a wide variety of ray paths are available to locate each quake (we will discuss the earthquake location problem in greater detail in the next section).

Another ambiguity in velocity inversions concerns the shallow structure at each seismic station. Rays generally come up at near-vertical angles beneath individual stations and sample only a very limited lateral area in the uppermost crust. Because of this, and the fact that no information is generally obtained for the shallow structure between stations, times to individual stations in large-scale inversions are usually adjusted using a *station correction*, a time for each station obtained by averaging the residuals from all ray paths to the station. As in the case of earthquake locations, it is important that the station correction be obtained from a wide range of ray paths, to minimize the biasing effect of travel time differences from deeper velocity anomalies.

Seismic tomography works best when a large number of different ray geometries are present and each cell in the model is crossed by rays at a wide range of angles. Unfortunately, this is often not the case, since the sources and receivers are unevenly distributed, and, at least in global tomography problems, largely confined to Earth's surface. Typically, this will result in many blocks being sampled at only a limited range of ray angles. When this occurs, anomalies are smeared along the ray path orientation (Fig. 4.3). This problem cannot be cured by regularization or other numerical methods—only the inclusion of additional ray paths at different angles can improve the resolution.

In some cases, there is the danger that the 3-D velocity perturbations could cause the source–receiver ray paths to deviate significantly from the reference model ray paths. If these ray-path deviations are large enough, then Fermat's principle may not save us and our results could be biased. This concern can be addressed by performing full 3-D ray tracing calculations on the velocity model and iterating until a stable solution is achieved. This requires significantly more work and has not generally been done in global tomography problems where the velocity perturbations are only a few percent. This effect is probably of greater importance in local and regional tomography problems where larger velocity anomalies are found and steep velocity gradients and/or discontinuities are more likely to be present.

There is also a tendency for rays to bend or be diffracted around localized slow anomalies, which may introduce a bias into tomographic inversions by making such features less resolvable than fast velocity anomalies (Nolet and Moser, 1993). More details concerning traditional seismic tomography techniques can be found in the books by Nolet (1987) and Iyer and Hirahara (1993).

4.0.4   Finite frequency tomography

"Classic" seismic tomography assumes the ray theoretical approximation, in which travel-time anomalies are accumulated only along the geometrical ray path. However, at realistic seismic wavelengths there will always be some averaging of structure adjacent to the theoretical ray path. Recently, seismologists have begun computing these finite-frequency effects in the form of kernels (sometimes called *Fréchet derivatives*) that show the sensitivity of the travel time or other observables for a particular seismic phase and source-receiver geometry to velocity perturbations throughout the Earth (e.g., Dahlen et al., 2000; Hung et al., 2000; Zhao et al., 2000). Examples of these kernels computed for a 1-D reference model for a P wave at $60°$ range are plotted in Figure 4.4. These are sometimes given the name *banana-doughnut kernels*, with "banana" describing the fact they are wider at the middle of the ray path than near its endpoints, and "doughnut" arising from the counterintuitive fact that their sensitivity is zero to velocity perturbations exactly along the geometrical ray path. The width of the kernels shrinks with the frequency of the waves and thus
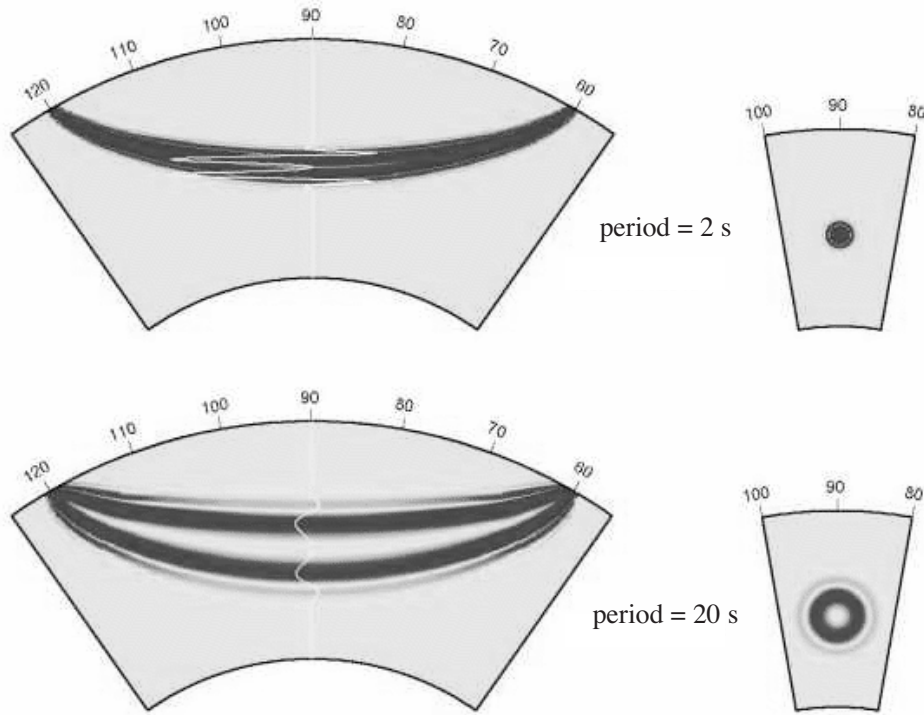
Figure 4.4: Banana-doughnut kernels showing the sensitivity of P-wave travel times at 60°
epicentral distance to velocity perturbations in the mantle. The right-hand plots show the
cross-section perpendicular to the ray direction at its midpoint. Note the much wider kernel
at 20 s period compared to 2 s period and the more pronounced "doughnut hole" along the
geometrical ray path. Figure from Dahlen et al. (2000).

the finite-frequency differences from geometrical ray theory are most important at
long periods.

In principle, the use of finite-frequency kernels should improve seismic tomog-
raphy by properly accounting for the effects of off-ray-path structure. There has
been some recent controversy as to how significant these improvements are for the
global mantle tomography problem with respect to the imaging of plumes, when
compared to differences arising from data coverage and regularization (see Montelli
et al., 2004; de Hoop and van der Hilst, 2005a,b; Dahlen and Nolet, 2005). How-
ever, it is clear that finite-frequency tomography represents a significant theoretical
advance and will eventually become common practice. Researchers are now com-
puting sensitivity kernels based on 3-D Earth models and developing sophisticated
algorithms for directly inverting waveforms for Earth structure (e.g., Zhao et al.,
2005; Liu and Tromp, 2006). These methods hold the promise of resolving structure

using much more of the information in seismograms than simply the travel times of
direct phases.

4.1   Additional reading

Houser, C., G. Masters, P. Shearer, and G. Laske, Shear and compressional velocity
   models of the mantle from cluster analysis of long-period waveforms, *Geophys.
   J. Int.*, **174**, 195212, doi:10.1111/j.1365-246X.2008.03763.x, 2008.

Lin, G., P. M. Shearer, E. Hauksson and C. H. Thurber, A three-dimensional crustal
   seismic velocity model for southern California from a composite event method,
   J. Geophys. Res., **112**, doi: 10.1029/2007JB004977, 2007.

4.2   References

Dahlen, F.A., Hung, S.-H. and Nolet, G. (2000). Frechet kernets for finite-frequency
   travel times—I. Theory, *Geophys. J. Int.*, **141**, 157–174.

Dahlen, F. and Nolet, G. (2005). Comment on the paper on sensitivity kernels for
   wave equation transmission tomography by de Hoop and van der Hilst, *Geophys.
   J. Int.*, **163**, 949–951.

de Hoop, M. and van der Hilst, R. (2005). On sensitivity kernels for wave equation
   transmission tomography, *Geophys. J. Int.*, **160**, 621–633.

de Hoop, M. and van der Hilst, R. (2005). Reply to a comment by F.A. Dahlen and
   G. Nolet on: On sensitivity kernels for wave equation tomography, *Geophys. J.
   Int.*, **163**, 952–955.

Hung, S.-H., Dahlen, F. and Nolet, G. (2000). Frechet kernets for finite-frequency
   travel times—I. Examples, *Geophys. J. Int.*, 141, 175–203.

Iyer, H. M. and Hirahara, K. (eds.) (1993). *Seismic Tomography: Theory and
   Practice*, Chapman and Hall, London.

Liu, Q. and Tromp, J. (2006). Finite-frequency kernels based on adjoint methods,
   Bull. Seismol. Soc. Am., **96**, 2383–2397.

Montelli, R., Nolet, G., Dahlen, F.A., Masters, G., Engdahl, E.R., Hung, S.-H.
   (2004). Finite-frequency tomography reveals a variety of plumes in the mantle,
   Science, **303**, 338–343.

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

Nolet, G. (ed.) (1987). *Seismic Tomography*, D. Reidel Publishing, Holland.

Nolet, G. and Moser, T.-J. (1993). Teleseismic delay times in a 3-D earth and a
   new look at the S discrepancy, *Geophys. J. Int.*, **114**, 185–195.

Vidale, J. (1988). Finite-difference calculation of travel times, *Bull. Seismol. Soc.
   Am.*, **78**, 2062–2076.

Zhao, L., Jordan, T. H. and Chapman, C. H. (2000). Three-dimensional Frechet
   differential kernels for seismic delay times, *Geophys. J. Int.*, **141**, 558–576.

Zhao, L., Jordan, T.H., Olsen, K.B. and Chen, P. (2005). Fréchet kernels for imaging regional earth structure based on three-dimensional reference models, *Geophys. J. Int.*, **95**, 2066–2080.

Chapter 5

Back-projection methods

Consider a linear set of equations relating observed data to a model:

$$\mathbf{d} = \mathbf{Gm} \tag{5.1}$$

using the conventional notation of $\mathbf{d}$ for the data vector, $\mathbf{m}$ for the model vector, and $\mathbf{G}$ for the linear operator that predicts the data from the model. Our goal in geophysical inverse problems is to estimate $\mathbf{m}$ from the observations, $\mathbf{d}$. Assuming there are more data points than model points, the standard way to solve this problem is to define a residual vector, $\mathbf{r} = \mathbf{d} - \mathbf{Gm}$, and find the $\mathbf{m}$ that minimizes $\mathbf{r} \cdot \mathbf{r}$. This is the least squares solutions and it can be shown that

$$\mathbf{m} = (\mathbf{G}^T \mathbf{G})^{-1} \mathbf{G}^T \mathbf{d}. \tag{5.2}$$

However often $\mathbf{G}^T \mathbf{G}$ is singular or ill-conditioned, or it may simply be too large to invert. What can be done is these cases? The simplest and crudest way to proceed is to make the approximation

$$(\mathbf{G}^T \mathbf{G})^{-1} \approx \mathbf{I} \tag{5.3}$$

in which case we can estimate the model as

$$\mathbf{m} \approx \mathbf{G}^T \mathbf{d}. \tag{5.4}$$

The transposed matrix $\mathbf{G}^T$ is the *adjoint* or back-projection operator. Each model point is constructed as the weighted sum of the data points that it affects. Can such a crude approximation be of any use? It's certainly easy to think of examples where (5.3) is completely invalid. However, in real geophysical problems it's

surprising how often this method works, particularly if a scaling factor is allowed to bring the data and model-predicted data into better agreement (i.e., assuming $(\mathbf{G}^T\mathbf{G})^{-1} \approx \lambda\mathbf{I}$, where λ is a constant). Indeed, it is sometimes observed that the adjoint works better than the formal inverse because it is more tolerant of imperfections in the data. Jon Claerbout discusses this in a wonderful set of notes (e.g., http://sepwww.stanford.edu/sep/prof/gee/ajt/paper_html/node1.html).

In seismology our data are typically a set of seismograms. In source inversions, we normally assume that the Earth's velocity structure is known and we solve for the locations and times of seismic wave radiators (e.g., solving for a slip model). In reflection seismology, we normally assume that the location and time of the source is known and we solve for the location of the reflector(s) that cause the observed arrivals. In each case, the model estimate at each model point is obtained by finding the times in the seismograms at which changes in the model will affect the seismogram. The model estimate from back-projection is obtained by simply summing or *stacking* the seismogram values at these points. The main thing to compute is the travel time between the model points and each recording station. These give the time shifts necessary to find the times in each seismogram that are sensitive to the model perturbations.

One way of thinking about this is that we have the computer perform a series of hypothesis tests over a time-space model grid. Is there a seismic radiator at this space-time point? If there is, we would expect it to show up in seismograms at these times. If we sum over the seismogram values at these times, we should get a large amplitude. Of course, it is possible that inference from radiation at other model points will cause us to have a biased estimate. But on average, we hope (expect) these other contributions to cancel out. This is the forward-time way of thinking about the problem.

But we could also think about this in reverse time. In this case we start with the seismograms and project their values backward in time through the model grid. As we do this, we accumulate the values in the model grid points. The model points that are likely sources will experience constructive interference as the time-reversed wavefields focus to these points. This is why this process is sometimes called back-

projection or reverse time migration. But the result is exactly the same as the forward modeling approach described in the previous paragraph.

Left unstated in this discussion is how the amplitudes in the seismograms should be scaled. If one wants to recover true model amplitudes, then geometrical spreading and other factors should be taken into account. Often, however, the goal is simply an image of the model and the absolute amplitude is not that important. For example, in reflection seismology automatic gain control is often used to equalize the contributions from different records and true amplitude information is lost. These amplitude normalization methods can make back-projection more robust with respect to noisy data or uncertainties in the velocity model.

5.1 Migration in reflection seismology

In reflection seismology, complicated structures will produce scattered and diffracted arrivals that cannot be modeled by simple plane-wave reflections, and accurate interpretation of data from such features requires a theory that takes these arrivals into account. Most of the analysis techniques developed for this purpose are based on the idea that velocity perturbations in the medium can be thought of as generating secondary seismic sources in response to the incident wavefield, and the reflected wavefield can be modeled as a sum of these secondary wavelets.

5.1.1 Huygens' principle

Huygens' principle, first described by Christiaan Huygens (c. 1678), is most commonly mentioned in the context of light waves and optical ray theory, but it is applicable to any wave propagation problem. If we consider a plane wavefront traveling in a homogeneous medium, we can see how the wavefront can be thought to propagate through the constructive interference of secondary wavelets (Fig. 5.1). This simple idea provides, at least in a qualitative sense, an explanation for the behavior of waves when they pass through a narrow aperture.

The bending of the ray paths at the edges of the gap is termed *diffraction.* The degree to which the waves diffract into the "shadow" of the obstacle depends upon the wavelength of the waves in relation to the size of the opening. At relatively
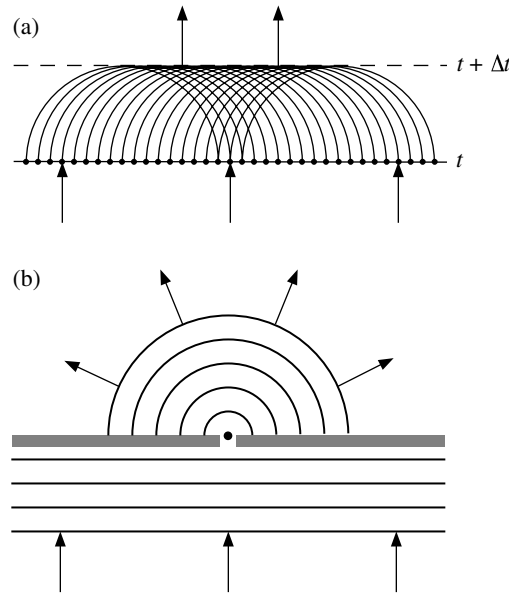
Figure 5.1: Illustrations of Huygens' principle. (a) A plane wave at time $t + \Delta t$ can be modeled as the coherent sum of the spherical wavefronts emitted by point sources on the wavefront at time t. (b) A small opening in a barrier to incident waves will produce a diffracted wavefront if the opening is small compared to the wavelength.

long wavelengths (e.g., ocean waves striking a hole in a jetty), the transmitted waves will spread out almost uniformly over $180°$. However, at short wavelengths the diffraction from the edges of the slot will produce a much smaller spreading in the wavefield. For light waves, very narrow slits are required to produce noticeable diffraction. These properties can be modeled using Huygens' principle by computing the effects of constructive and destructive interference at different wavelengths.

5.1.2   Diffraction hyperbolas

We can apply Huygens' principle to reflection seismology by imagining that each point on a reflector generates a secondary source in response to the incident wavefield. This is sometimes called the "exploding reflector" model. Consider a single point scatterer in a zero-offset section (Fig. 5.2). The minimum travel time is given by

$$t_0 = \frac{2h}{v}, \tag{5.5}$$

where h is the depth of the scatterer and v is the velocity (assumed constant in this case). More generally, the travel time as a function of horizontal distance, x, is
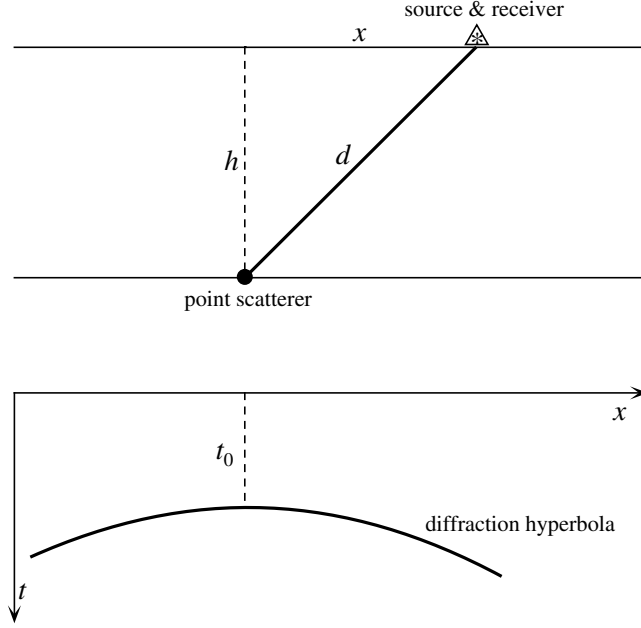
Figure 5.2: A point scatterer will produce a curved "reflector" in a zero-offset section.

given by

$$t(x) = \frac{2\sqrt{x^2 + h^2}}{v}. \tag{5.6}$$

Squaring and rearranging, this can be expressed as

$$\frac{v^2 t^2}{4h^2} - \frac{x^2}{h^2} = 1 \tag{5.7}$$

or

$$\frac{t^2}{t_0^2} - \frac{4x^2}{v^2 t_0^2} = 1 \tag{5.8}$$

after substituting $4h^2 = v^2 t_0^2$ from (5.5). The travel time curve for the scattered arrival has the form of a hyperbola with the apex directly above the scattering point. This equation describes travel time as a function of distance away from a point scatterer at depth for zero-offset data (the source and receiver are coincident).

5.1.3 Migration methods

Consider a horizontal reflector that is made up of a series of point scatterers, each of which generates a diffraction hyperbola in a zero-offset profile (Fig. 5.3). Following Huygens' principle, these hyperbolas sum coherently only at the time of the main reflection; the later contributions cancel out. However, if the reflector vanishes at

<div style="text-align: center;">Model   Zero-offset section</div>
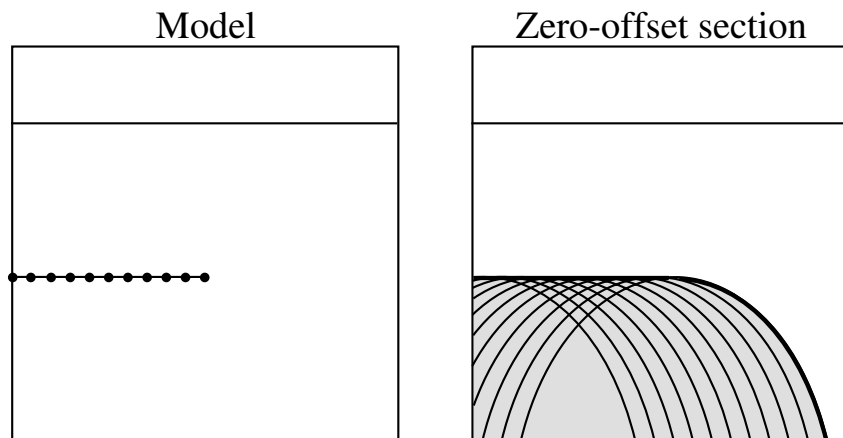

Figure 5.3: The endpoint of a horizontal reflector will produce a diffracted arrival in a zero-offset section. The reflector itself can be modeled as the coherent sum of the diffraction hyperbola from individual point scatterers. The diffracted phase, shown as the curved heavy line, occurs at the boundary of the region of scattered arrivals.

some point, then there will be a diffracted arrival from the endpoint that will show up in the zero-offset data. This creates an artifact in the section that might be falsely interpreted as a dipping, curved reflector.

Techniques for removing these artifacts from reflection data are termed *migration* and a number of different approaches have been developed. The simplest of these methods is termed *diffraction summation migration* and involves assuming that each point in a zero-offset section is the apex of a hypothetical diffraction hyperbola. The value of the time series at that point is replaced by the average of the data from adjacent traces taken at points along the hyperbola. In this way, diffraction artifacts are "collapsed" into their true locations in the migrated section. In many cases migration can produce a dramatic improvement in image quality (e.g., Fig. 5.4).

A proper implementation of diffraction summation migration requires wave propagation theory that goes beyond the simple ideas of Huygens' principle. In particular, the scattered amplitudes vary as a function of range and ray angle, and the Huygens secondary sources are given, for a three-dimensional geometry, by the time derivative of the source-time function (in the frequency domain this is described by the factor $-i\omega$, a $\pi/2$ (90 degree) phase shift with amplitude proportional to frequency). In the case of a two-dimensional geometry, the secondary sources are the "half-derivative" of the source function (a 45 degree phase shift with amplitude
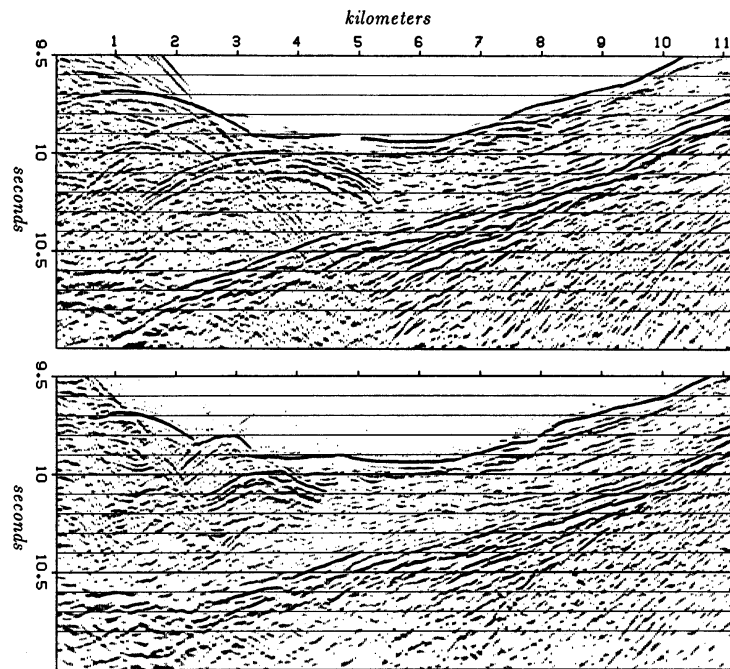
Figure 5.4: Original (top) and migrated (bottom) reflection data from a survey line across the Japan trench (figure modified from Claerbout, 1985; data from the Tokyo University Oceanographic Research Institute).

scaled by the square root of frequency). These details are provided by Kirchhoff theory, which is discussed later in this chapter. The diffraction hyperbola equation assumes a uniform velocity structure, but migration concepts can be generalized to more complicated velocity models. However, it is important to have an accurate velocity model, as use of the wrong model can result in "undermigrated" or "overmigrated" sections.

In common practice, data from seismic reflection experiments are first processed into zero-offset sections through common midpoint (CMP) stacking. The zero-offset section is then migrated to produce the final result. This is termed *poststack migration*. Because CMP stacking assumes horizontal layering and may blur some of the details of the original data, better results can be obtained if the migration is performed prior to stacking. This is called *prestack migration*. Although prestack migration is known to produce superior results, it is not implemented routinely owing to its much greater computational cost.

5.2   Additional reading

Allmann, B.P., and P.M. Shearer, A high-frequency secondary event during the 2004 Parkfield earthquake, *Science*, **318**, 1279, doi: 10.1126/science.1146537, 2007.

Ishii, M., P.M. Shearer, H. Houston and J.E. Vidale, Extent, duration and speed of the 2004 Sumatra-Andaman earthquake imaged by the Hi-Net array, *Nature*, **435**, doi:10.1038/nature03675, 2005.

Ishii, M., P.M. Shearer, H. Houston, and J.E. Vidale, Teleseismic P wave imaging of the 26 December 2004 Sumatra-Andaman and 28 March 2005 Sumatra earthquake ruptures using the Hi-net array, *J. Geophys. Res.*, **112**, B11307, doi:10.1029/2006JB004700, 2007.

Walker, K.T., M. Ishii and P.M. Shearer, Rupture details of the 28 March 2005 Sumatra Mw 8.6 earthquake imaged with teleseismic P waves, *Geophys. Res. Lett.*, **32**, L24303, doi: 10.1029/2005GL024395, 2005.

Walker, K. T., and P. M. Shearer, Illuminating the near-sonic rupture velocities of the intracontinental Kokoxili Mw 7.8 and Denali fault Mw 7.9 strike-slip earthquakes with global P wave back projection imaging, *J. Geophys. Res.*, **114**, doi: 10.1029/2008JB005738, 2009.

Chapter 6

Pulse shapes, spectra, and stress drop

The displacement that occurs on opposite sides of a fault during an earthquake is permanent; the Earth does not return to its original state following the event. Thus, the equivalent body force representation of the displacement field must involve a permanent change in the applied forces. In addition, the displacement is not instantaneous but occurs over some finite duration of rupture. We can accommodate these properties by generalizing the moment tensor source representation to be time dependent. For instance, one of the components of the moment tensor could be expressed as $M(t)$ and might have the form shown at the top left of Figure 6.1. This is what the near-field displacement would look like; for example, this might describe the path of a house near the San Andreas Fault during a large earthquake. These displacements are permanent and can be measured at some distance away from large earthquakes by geodetic means (such as surveying or GPS) after the shaking has subsided.

The expressions for the far-field displacements from isotropic or double-couple sources all involve the time derivative of the moment tensor. The time derivative of $M(t)$ is proportional to the far-field dynamic response (the middle panel of Figure 6.1), such as would be observed in a P- or S-wave arrival. Note that this is a displacement pulse and that there is no permanent displacement after the wave passes. Most seismometers measure velocity $\dot{u}(t)$ rather than displacement $u(t)$, in which case what is actually recorded will have an additional time derivative.
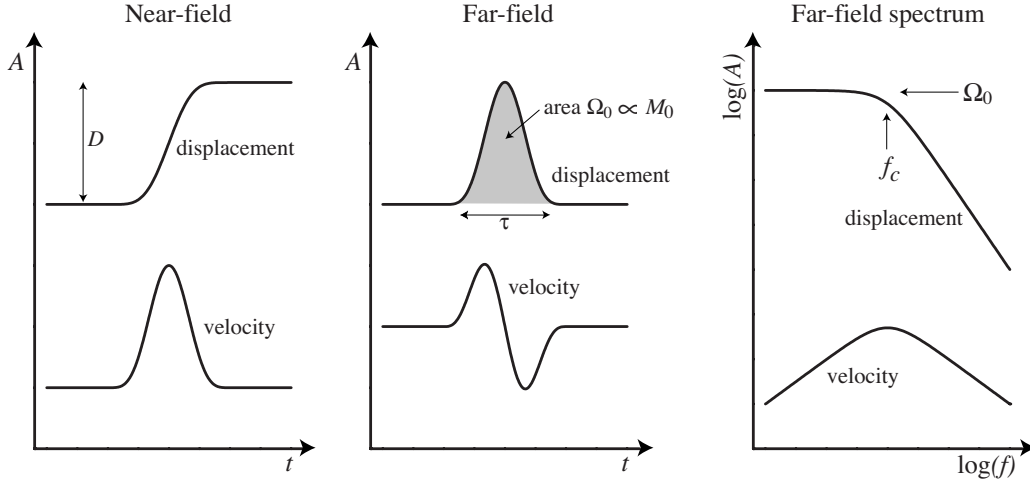
Figure 6.1: The relationships between near-field displacement and far-field displacement and velocity for time series (left two panels) and spectra (right panel).

In problems of Earth structure, it generally matters little whether we use velocity rather than displacement provided we assume an extra derivative for the source when we are modeling the waveforms. However, when studying seismic sources, velocity is almost always converted to displacement. This is done by integrating the velocity record and normally also involves a correction for the instrument response. The aim is to recover an unbiased record of $\dot{M}(t)$ at the source. We will assume for most of this section that we are measuring far-field displacement.

The spectrum of the far-field displacement pulse (see top right of Figure 6.1) at low frequencies will be flat at a level, Ω_0, equal to the area beneath the pulse. The displacement spectrum will then roll off at higher frequencies, with the *corner frequency*, f_c, inversely proportional to the pulse width, τ. In the frequency domain the effect of the time derivatives is to multiply the spectrum by f. Thus velocity records are enhanced in high frequencies relative to displacement records.

The long-period spectral level, Ω_0, is proportional to the scalar seismic moment, M_0. Recall that $M_0 = \mu A D$, where μ is the shear modulus, A is the fault area, and D is the displacement. In the case of body waves it can be shown that

$$M_0 = \frac{4\pi\rho c^3 r \Omega_0}{U_{\phi\theta}} \tag{6.1}$$

where ρ is the density, c is the wave velocity, r is the distance from the source, and $U_{\phi\theta}$ is the radiation pattern term. This equation is for spherical wavefronts

expanding in a whole space but can be applied to more complicated velocity models using ray theory if the r factor is replaced with the appropriate term for geometrical spreading. If Ω_0 is measured from a station at the Earth's surface, then corrections must be applied to account for the wave amplification that occurs from the surface reflections. There are analogous expressions for computing M_0 from surface waves. These equations are important because they show how a fundamental property of the earthquake source—scalar moment—can be obtained directly from seismic wave observations at great distances. Because Ω_0 is measured at the lowest possible frequency, it is relatively insensitive to the effects of scattering and attenuation, making scalar moment estimates more reliable than measurements of source properties that require higher frequency parts of the spectrum. However, note that Equation (6.1) does require knowledge of the focal mechanism owing to the $U_{\phi\theta}$ term. If a focal mechanism is not available, sometimes M_0 is estimated by averaging results from many stations and replacing $U_{\phi\theta}$ with the mean radiation term over the focal sphere (0.52 and 0.63 for P and S waves, respectively). Of course, the scalar moment is a simple function of the complete moment tensor if it is available.

Many different theoretical earthquake source models have been proposed and they predict different shapes for the body-wave spectra. Brune (1970) described one of the most influential models, in which the displacement amplitude spectrum is given by

$$A(f) = \frac{\Omega_0}{1 + (f/f_c)^2} \tag{6.2}$$

where f_c is the corner frequency. Note that the high-frequency fall-off rate agrees with the Haskell fault model. A more general model is

$$A(f) = \frac{\Omega_0}{[1 + (f/f_c)^{\gamma n}]^{1/\gamma}} \tag{6.3}$$

which was found by Boatwright (1980) with $\gamma = 2$ to provide a better fit to the sharper corners that he found in his data. Equations (6.2) and (6.3) with $n = 2$ are often called ω^{-2} source models. Some theoretical source models, particularly those which consider elongated fault geometries, predict ω^{-3} fall off at high frequencies. However, studies of both globally and locally recorded earthquakes over a wide range

of sizes have generally shown that their average high-frequency fall-off rate is close to ω^{-2}, although individual earthquakes often have quite different spectral behavior.

6.0.1   Empirical Green's functions

One of the challenging aspects of studying seismic spectra is separating out what originates from the source and what is caused by attenuation or other path effects. For example, for a simple constant Q model the spectra will drop off exponentially at high frequencies

$$A(f) = A_0(f)e^{-\pi ft/Q}. \tag{6.4}$$

In principle, this fall off has different curvature than the power law decay with frequency of theoretical source models and one approach has been to use (6.4) together with (6.2) or (6.3) to simultaneously solve for Q and f_c (and sometimes n and γ as well). However, with the irregular spectra and limited bandwidth of real data it can be difficult to separately resolve the source and attenuation contributions and there is often a tradeoff between them.

Another approach is to use records from a smaller earthquake near the target earthquake to compute an empirical path and attenuation correction. The assumption is that the second quake is small enough that its corner frequency is above the observation band and its spectrum is nearly flat, i.e., it is effectively a delta-function source. In this case one can either deconvolve its waveform from the target earthquake record in the time domain or simply correct the observed spectrum in the frequency domain. This is called the *empirical Green's function* or EGF method (e.g., Mueller, 1985; Hough, 1997) and is widely used in source studies. It does, however, require that there be a suitable event close enough to the target earthquake that the path effects will be approximately the same.

6.1   Stress Drop

The seismic moment, $M_0 = \mu \overline{D} A$, does not distinguish between an earthquake involving small slip on a large fault and one with large slip on a small fault, provided the product of the average slip $(\overline{D})$ and fault area (A) remains constant. However,

these earthquakes would change the stress on the fault by very different amounts. This change may be defined as the *stress drop*, which is the average difference between the stress[1] on a fault before an earthquake to the stress after the earthquake:

$$\Delta\sigma = \frac{1}{A} \int_S [\sigma(t_2) - \sigma(t_1)]\, dS \;, \tag{6.5}$$

where the integral is performed over the surface of the fault and A is the fault area. Analytical solutions for the stress drop have been derived for a few specialized cases of faults embedded within homogeneous material. For a circular fault in a whole-space, Eshelby (1957) obtained

$$\Delta\sigma = \frac{7\pi\mu\overline{D}}{16r} = \frac{7M_0}{16r^3} \;, \tag{6.6}$$

where r is the fault radius, μ is the shear modulus, and $\overline{D}$ is the average displacement. For strike-slip motion on a shallow, rectangular fault of length L and width w $(L \gg w)$, Knopoff (1958) obtained

$$\Delta\sigma = \frac{2\mu\overline{D}}{\pi w} = \frac{2M_0}{\pi w^2 L} \;. \tag{6.7}$$

More generally, we may write

$$\Delta\sigma = C\mu \left[\frac{\overline{D}}{\tilde{L}}\right], \tag{6.8}$$

where $\tilde{L}$ is a *characteristic rupture dimension* (r in the case of the circular fault, w for the long rectangular fault) and C is a nondimensional constant that depends upon the geometry of the rupture. Notice that physically it makes sense that the shear stress change on the fault will be proportional to the ratio of the displacement to the size of the fault. Large slip on a small fault will cause more stress than small slip on a large fault. It should be noted that these solutions assume smooth forms for the slip function on the fault surface and thus represent only approximations to the spatially averaged stress drop on real faults, for which the displacement and corresponding stress drop may vary in complicated ways owing to non-uniform elastic properties and initial stresses. A widely used result to obtain results for faults made up of arbitrary rectangular slip patches is the half-space solution of Okada (1992).

---

[1]In this section "stress" refers specifically to the shear stress across the fault plane.

For large earthquakes for which the fault geometry can be constrained from surface rupture or aftershock studies, the stress drop can then be estimated from the moment. For large, shallow earthquakes, $\Delta\sigma$ varies from about 1 to 10 MPa (10 to 100 bars in the units often used in older studies) with no observed dependence on moment for M_0 variations from 10^{18} to 10^{23} N m (Kanamori and Anderson, 1975; Kanamori and Brodsky, 2004). Earthquakes near plate boundaries (*interplate* events) generally have been observed to have somewhat lower stress drops than those that occur in the interior of plates (*intraplate* events) (e.g., Kanamori and Anderson, 1975; Kanamori and Allen, 1986). Average $\Delta\sigma$ for interplate quakes is about 3 MPa compared to about 6 MPa for intraplate events (Allmann and Shearer, 2008). This implies that intraplate faults are "stronger" in some sense than interplate faults and have smaller fault dimensions for the same moment release.

For small earthquakes, direct observations of the rupture geometry are not possible so the fault dimensions must be estimated from far-field observations of the radiated seismic waves. In this case it is necessary to make certain assumptions about the source properties. In particular, these methods generally assume that the source dimension is proportional to the observed body-wave pulse width (after correcting for attenuation). The first quantitative model for estimating stress drop in this way was derived by Brune (1970), who assumed a simple kinematic model for a circular fault with effectively infinite rupture velocity and showed that the expected high-frequency spectral fall-off rate is ω^{-2} and that the corner frequency is inversely proportional to the source radius. This result, together with a number of other proposed rupture models, predicts that the fault radius varies as

$$r = \frac{k\beta}{f_c} \, , \tag{6.9}$$

where r is the fault radius, f_c is the observed corner frequency (see Figure 6.1) and k is a constant that depends upon the specific theoretical model. Currently, perhaps the most widely used result is from Madariaga (1976), who performed dynamic calculations for a circular fault using finite differences. Assuming that the rupture velocity is 90% of the shear-wave velocity ($v_r = 0.9\beta$), he obtained $k = 0.32$ and 0.21 for the P- and S-wave corner frequencies, respectively, with an ω^{-2} high-
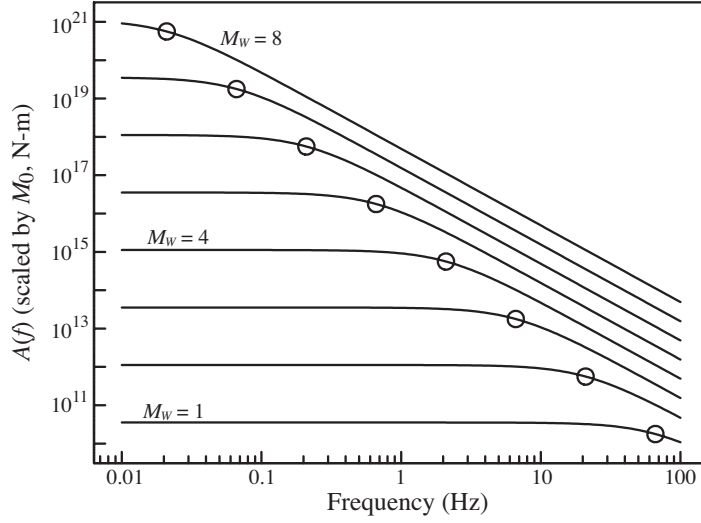
Figure 6.2: Predicted P-wave spectra from the Madariaga (1976) source model, assuming a constant stress drop of 3 MPa. The spectra have been scaled such that their amplitudes at low frequency are equal to their moments, M_0. The circles show the corner frequencies (f_c). Individual spectra are for moment magnitudes, M_W, from 1 to 8 (see (??) for the definition of M_W)

frequency fall-off rate. His model predicts a P-wave corner frequency about 50% higher than the S-wave corner frequency ($f_c^P \simeq 1.5 f_c^S$). Figure 6.2 plots predicted P-wave spectra for the Madariaga (1976) model for a wide range of M_0, assuming a constant stress drop of 3 MPa. Note that the corner frequency varies as $M_0^{-1/3}$, with higher corner frequencies for smaller earthquakes.

From (6.6) and (6.9), we have

$$\Delta\sigma = \frac{7}{16} \left(\frac{f_c}{k\beta} \right)^3 M_0 . \tag{6.10}$$

This is how stress drop can be estimated directly from far-field body wave spectra using corner-frequency measurements, together with measurements of M_0 (which can be computed from the low frequency part of the spectrum, see Ω_0 in Figure 6.1). Because this equation involves the cube of the $(f_c/k\beta)$ term, the computed $\Delta\sigma$ is extremely sensitive to differences in the assumed theoretical model (which determines the value of k and in general depends upon the assumed rupture velocity) and to variations in the estimated corner frequency f_c. The Brune (1970) model has a k value about 1.7 times larger than the Madariaga (1976) model, which translates to stress drop estimates about 5 times smaller. The corner frequency, f_c,

can be tricky to measure from individual spectra, which are rarely as smooth as the theoretical models predict, and are sensitive to corrections for attenuation effects. Published stress drop values exhibit considerable scatter and it can be difficult to determine what part of these variations are real and what part may be attributed to differences in the modeling assumptions and analysis methods. However, there are large variations in individual earthquake stress drops even within single studies, suggesting that much of the observed scatter is real. For example, Shearer *et al.* (2006) analyzed *P*-wave spectra from over 60,000 small earthquakes in southern California using the Madariaga (1976) model and obtained $\Delta\sigma$ values from 0.2 to 20 MPa, with the bulk of the events between 0.5 to 5 MPa.

In principle, stress drop, like moment, is essentially a static measurement of permanent changes caused by an earthquake. However, the methods for estimating stress drops for small earthquakes are derived from body-wave pulse shapes and assumptions about the dynamics of the source. Because these are not direct measurements of static stress drop, they are sometimes termed *Brune-type* stress drops, although they may not be computed exactly as in Brune (1970). It is important to remember that these measurements involve a number of modeling assumptions that may not be true for individual earthquakes. For example, variations in rupture speed will cause a change in corner frequency even if the stress drop remains constant. Finally, note that measurements of the stress drop do not constrain the absolute level of stress on faults. The absolute level of stress in the crust near faults has long been a subject of controversy, with heat flow constraints suggesting lower levels of stress for real faults than laboratory rock-sliding experiments seem to require.

6.2 Additional reading

Allmann, B.P., and P.M. Shearer, Spatial and temporal stress drop variations in small earthquakes near Parkfield, California, *J. Geophys. Res.*, **112**, B4, B04305, doi:10.1029/2006JB004395, 2007.

Allmann, B. B., and P. M. Shearer, Global variations of stress drop for moderate to large earthquakes, *J. Geophys. Res.*, **114**, doi: 10.1029/2009JB005821, 2009.

Shearer, P. M., G. A. Prieto, and E. Hauksson, Comprehensive analysis of earth-

quake source spectra in southern California, *J. Geophys. Res.*, **111**, B06303, doi:10.1029/2005JB003979, 2006.

6.3 References

Boatwright, J. (1980). A spectral theory for circular seismic sources: simple estimates of source dimension, dynamic stress drop and radiated energy, *Bull. Seismol. Soc. Am.*, **70**, 1–27.

Brune, J. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, **75**, 4997–5009.

Eshelby, J.D. (1957). The determination of the elastic field of an ellipsoidal inclusion, and related problems, *Prec. Roy. Soc. London*, **A241**, 376–396.

Hough, S.E. (1997). Empirical Green's function analysis: Taking the next step, *J. Geophys. Res.*, **102**, 5369–5384.

Kanamori, H., and Allen, C. R. (1986). Earthquake repeat time and average stress drop, in *Earthquake Source Mechanics*, S. Das and C. H. Scholz (editors), American Geophysical Union, Washington, C.C., 227–235.

Kanamori, H. and Anderson, D. L. (1975). Theoretical basis of some empirical relations in seismology, *BSSA*, **65**, 1073–1095.

Kanamori, H. and Brodsky, E. (2004). The physics of earthquakes, *Rep. Prog. Phys.*, **67**, 1429–1496.

Knopoff, L. (1958). Energy release in earthquakes, *Geophys. J.*, **1**, 44–52.

Madariaga, R. (1976). Dynamics of an expanding circular fault, *Bull. Seismol. Soc. Am.*, **66**, 639–666.

Müller, G. (1985). Source pulse enhancement by deconvolution of an empirical Green's function, *Geophys. Res. Lett.*, **12**, 33–36.

Okada, Y. (1992). Internal deformation due to shear and tensile faults in a half-space, *Bull. Seismol. Soc. Am.*, **82**, 1018–1040.

Chapter 7

Earthquake scaling and energy

The fact that earthquake stress drops appear to be at least approximately constant over a wide range of earthquake sizes has implications for earthquake scaling relationships. Aki (1967) proposed that the physics of of large and small earthquakes may be fundamentally similar, in which case we should expect *scale-invariance* or *self-similarity* of the rupture process. This implies that regardless of which theoretical earthquake source model is correct, the properties of the source will change in predictable ways as a function of earthquake size.

This is illustrated in Figure 7.1, which shows the expected change in pulse shape and spectrum when an earthquake rupture plane is increased in size by a factor b. Assuming the dimensions of the larger rupture are scaled proportionally, then the fault area, A, will increase by a factor b^2, the displacement, D, will increase by b, and the moment, $M_0 = \mu D A$, will increase by a factor of b^3. Stress drop remains constant because it is proportional to $DA^{-1/2}$. It follows that moment will scale with fault area as

$$M_0 \propto A^{3/2} \tag{7.1}$$

and such a scaling is observed to be approximately correct for large earthquakes (e.g., Kanamori and Anderson, 1975; Kanamori and Brodsky, 2004).

For an identical source-receiver geometry, no attenuation, and constant rupture velocity (predicted from self-similarity), the far-field displacement pulse will increase in duration by a factor of b and in amplitude by a factor of b^2. Note that the area under the pulse, Ω_0, also increases by b^3, as expected since Ω_0 is proportional to
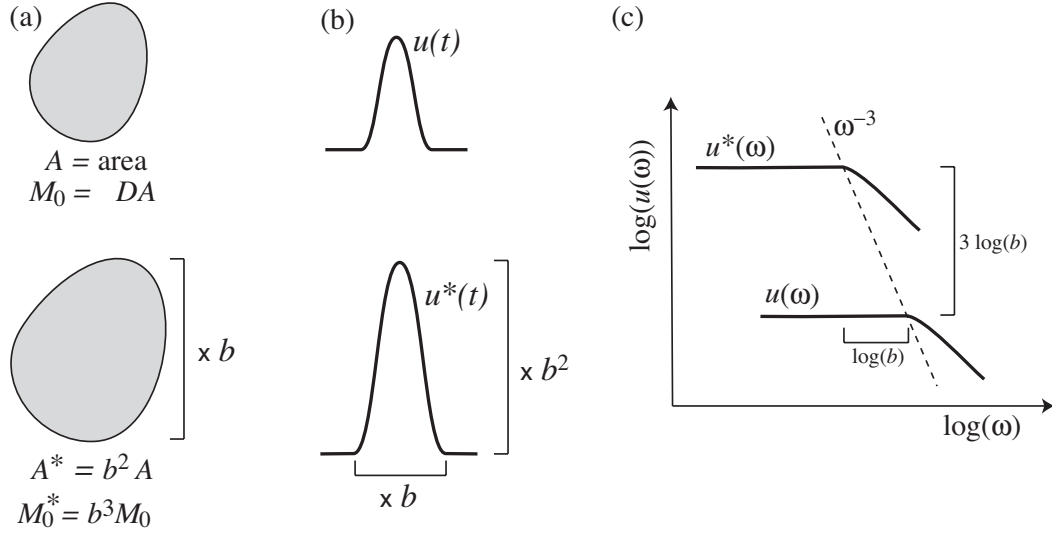
Figure 7.1: Illustration of the effects of self-similarity when an earthquake is increased in size by a factor b, showing the behavior of (a) rupture area and moment, (b) far-field displacement pulses, and (c) displacement spectra. Figure adapted from Prieto et al. (2004).

M_0. It follows that the displacement pulse, u^*, recorded by the second earthquake can be expressed as

$$u^*(t) = b^2 u(t/b) \tag{7.2}$$

where $u(t)$ is the recorded displacement pulse of the first earthquake. The radiated seismic energy, E_R, in the recorded pulse will be proportional to $\int \dot{u}^2(t)\, dt$ (the integrated square of the slope of the pulse), so the second pulse will contain a factor b^3 more energy than the first pulse. Thus the radiated seismic energy to moment ratio (E_R/M_0) remains constant.

Using the similarity theorem for the Fourier transform, it follows that the spectrum of the second earthquake is given by

$$u^*(\omega) = b^3 u(b\omega) \tag{7.3}$$

where $u(\omega)$ is the spectrum of the first earthquake. This relationship predicts that the shape of all spectra on a log-log plot will be identical, but offset along a line of ω^{-3} (Figure 7.1c). This means that corner frequency will vary as

$$f_c \propto M_0^{-1/3} \tag{7.4}$$

as is seen in Figure 6.2.

Self-similarity appears to be at least roughly true for average earthquake properties, although this has been a subject of considerable debate and there are large variations among individual earthquakes. It should be noted that self-similarity may break down for very large earthquakes that rupture through the entire seismogenic zone. In this case, ruptures are much longer than they are wide, with aspect ratios of 10 or more, which might make them behave differently than the less elongated rupture planes expected of smaller earthquakes (e.g., Scholz, 1982, 1997; Heaton, 1990). For example, the 1906 San Francisco earthquake ruptured for about 450 km to a depth of no more than 10 km (Thatcher, 1975).

7.1 Radiated Seismic Energy

Seismic moment and static stress drop are fundamental properties of the slip geometry of an earthquake, but they say nothing directly about the dynamics of the event, such as how fast the rupture propagated or how fast the two sides of the fault moved. This is why it is possible to estimate M_0 and $\Delta\sigma$ from geodetic measures of Earth deformation long after an earthquake; they are measures of the permanent static displacements across faults. Fault creep events that are too slow to radiate seismic energy at observable frequencies can nonetheless have significant moments and stress drops (although as noted in section 6.1, some methods of actually computing stress drops require seismic wave observations and make assumptions about source dynamics).

In contrast, one of the most fundamental measures of earthquake dynamics is the total radiated energy, E_R, which represents the seismic energy that would propagate to the far field in a whole space with no attenuation. Using the expressions for seismic energy flux, we have (e.g., Venkataraman et al., 2006)

$$E_R = \rho \int_S \int_{-\infty}^{\infty} \left[\alpha \dot{u}_\alpha^2(t,\theta,\phi) + \beta \dot{u}_\beta^2(t,\theta,\phi) \right] dt\, dS, \qquad (7.5)$$

where $\dot{u}_\alpha$ and $\dot{u}_\beta$ are velocity seismograms for P and S waves, respectively, and S is a spherical surface at a large distance around the source. Of course, we cannot integrate over the entire focal sphere; we must use seismic observations from a discrete number of seismic stations. Using ray theory, we can correct the observed

amplitudes for varying amounts of geometrical spreading and determine the ray takeoff angles, θ and ϕ, at the source. Because of radiation pattern effects, $\dot{u}_\alpha$ and $\dot{u}_\beta$ vary greatly over the surface of the sphere and thus a large number of observations from different seismic stations would be necessary to estimate E_R reliably from (7.5) directly. However, if the focal mechanism and thus the radiation pattern is known, then single station estimates are possible, i.e.,

$$E_R = E_R^P + E_R^S = 4\pi\rho\alpha r^2 \frac{\langle {}_P U_{\phi\theta}{}^2 \rangle}{{}_P U_{\phi\theta}{}^2} I_P + 4\pi\rho\beta r^2 \frac{\langle {}_S U_{\phi\theta}{}^2 \rangle}{{}_S U_{\phi\theta}{}^2} I_S \tag{7.6}$$

where ${}_P U_{\phi\theta}$ and ${}_S U_{\phi\theta}$ are the P and S radiation pattern terms and $\langle U_{\phi\theta}{}^2 \rangle$ is the mean over the focal sphere of $(U_{\phi\theta})^2$ ($\langle {}_P U_{\phi\theta}{}^2 \rangle = 4/15$ for P waves and $\langle {}_S U_{\phi\theta}{}^2 \rangle = 2/5$ for S waves), and I_P and I_S are the time-integrated values of $\dot{u}_\alpha^2$ and $\dot{u}_\beta^2$, as corrected for geometrical spreading and any near-receiver effects (e.g., free-surface reflections or amplifications from slow velocities in shallow layers) to what they would be at a uniform distance r in the absence of attenuation.

I_P and I_S are usually computed in the frequency domain from body-wave spectra because it is easier to correct for attenuation and instrument response effects, as well as to check for adequate signal-to-noise properties. From Parseval's theorem, we have

$$I = \int_{-\infty}^{\infty} |v(t)|^2 \, dt = \int_{-\infty}^{\infty} |v(f)|^2 \, df \tag{7.7}$$

In principal, the integration is performed to infinite frequency. However, the velocity spectrum peaks near the corner frequency (see Figure 6.1), and this peak becomes even stronger when the velocity is squared. For the ω^{-2} model, calculations have shown that 90% of the total energy is obtained if the integration is performed out to 10 times the corner frequency (Ide and Beroza, 2001). Often data do not have this much bandwidth, which can lead to underestimation of the energy. To correct for this, the integration can be extrapolated beyond the observed bandwidth of the data by assuming that the spectral fall off continues at a fixed rate. However, in this case the result is no longer a direct measurement from the data because it relies on assumptions about the nature of the source.

The ratio of S-wave energy to P-wave energy is defined as

$$q = E_R^S / E_R^P \tag{7.8}$$

For a point-source model in which the P and S-wave pulses have identical shapes (and thus identical corner frequencies f_c^P and f_c^S), it can be shown that $q = 1.5(\alpha/\beta)^5 \simeq 23.4$ for a Poisson solid. However, many theoretical finite source models predict that the P-wave pulse will be shorter in duration than the S-wave pulse (i.e., $f_c^P > f_c^S$), which will result in lower values for q. For example, the Madariaga (1976) model has $f_c^P \simeq 1.5 f_c^S$, from which one can compute (Boatwright and Fletcher, 1984) that q is about 7. Observations have generally suggested average q values between 9 and 25, with a large amount of scatter for individual earthquakes.

Measuring E_R is much more difficult than measuring M_0 and results among different groups for the same earthquakes often differ by factors of 2 or more. This is because E_R is derived from high-frequency parts of the source spectrum where corrections for attenuation are critically important. Most of the energy is radiated as S waves, which are particularly sensitive to attenuation. If only E_R^P measurements are available, E_R can still be estimated if a fixed value of q is assumed, but once again this detracts from the directness of the observation. Because energy is proportional to the square of the wave amplitudes, the effects of the radiation pattern are more severe for E_R calculations compared to M_0 calculations. The $U_{\phi\theta}$ terms in the denominators of (7.6) go to zero at the nodes in the radiation pattern. This can lead to artificially high energy estimates if measurable wave amplitudes are seen near the nodes, which can happen due to scattering, 3D structure, or inaccuracies in the focal mechanism. Finally, rupture directivity does not affect M_0 estimates (because Ω_0 is preserved despite changes in the pulse amplitudes) but produces large variations in I_P and I_S (e.g., Ma and Archuleta, 2006). If directivity effects are important, then (7.6) is incomplete and can produce biased results, depending upon whether the critical takeoff angles with the highest amplitudes are included in the available data.

The ratio of the radiated energy to the moment

$$\tilde{e} = \frac{E_R}{M_0} = \frac{1}{\mu} \frac{E_R}{\overline{D}A} \tag{7.9}$$

is called the *scaled energy* and is dimensionless (note that 1 joule = 1 N m). The parameter $\mu\tilde{e} = E_R/\overline{D}A$ has units of stress and has traditionally been called *appar-*
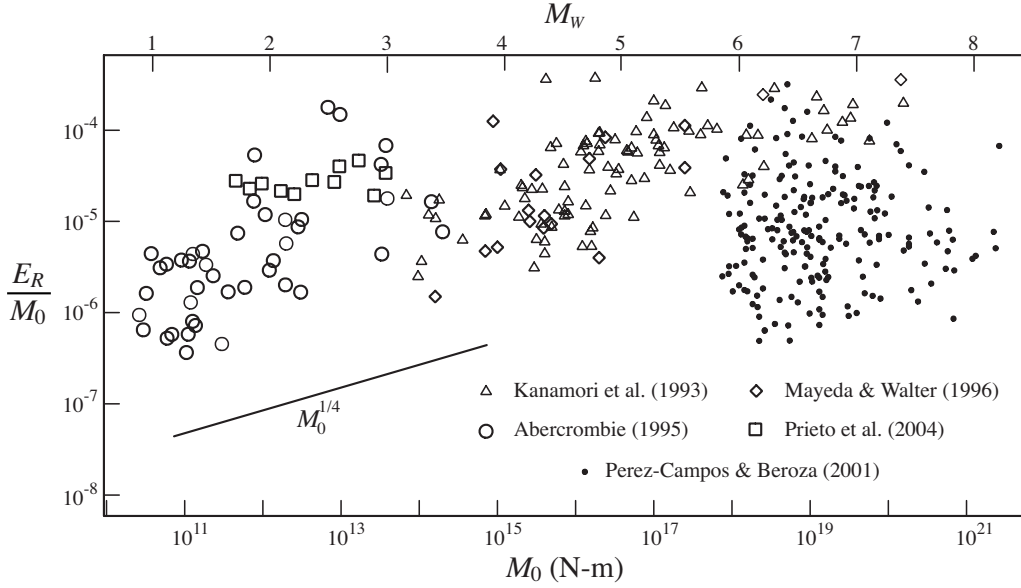
Figure 7.2: The observed radiated seismic energy to moment ratio, $\tilde{e} = E_R/M_0$, plotted as a function of moment. The $M_0^{1/4}$ trend noted in some studies is plotted for reference.

ent stress but this term can be confusing because it is not directly related to either absolute stress or stress drop. The scaled energy, $\tilde{e}$, is proportional to the energy radiated per unit fault area and per unit slip. As noted in the previous section, if earthquakes are self-similar then $\tilde{e}$ should be constant as a function of moment. Whether this is indeed the case has been the subject of some controversy (e.g., see recent review by Walter et al., 2006). Some have argued that average $\tilde{e}$ grows with moment approximately as $M_0^{1/4}$ (e.g., Mayeda and Walter, 1996) while others have maintained that average $\tilde{e}$ is seen to be nearly constant with M_0 when one carefully corrects for possible biases in the data analysis (e.g., Ide and Beroza, 2001). Figure 7.2 plots $\tilde{e}$ versus M_0, showing results from a number of different studies. Note that there is a great deal of scatter in the $\tilde{e}$ estimates, which span over an order of magnitude even at the same moment. However, there is some evidence for an increase in $\tilde{e}$ with moment, particularly for the smaller earthquakes. Ide and Beroza (2001) have argued, however, that this may be an artifact of the data selection method in the Abercrombie (1995) study. An important issue is the fact that energy estimates derived from teleseismic data tend to be about 10 times smaller than those obtained from local records (Singh and Ordaz, 1994; Mayeda and Walter, 1996). This can be

seen in Figure 7.2, noting that Perez-Campos and Beroza (2001) is the only tele-seismic study plotted. If these points are excluded, the $M_0^{1/4}$ trend becomes much clearer.

7.1.1 Earthquake energy partitioning

The total strain and gravitational energy released during an earthquake is given by

$$E = \tfrac{1}{2}\overline{(\sigma_1 + \sigma_2)D}A \qquad (7.10)$$

where σ_1 is the initial stress, σ_2 is the final stress, D is displacement, A is the fault area, and the overbar means the spatial average. Note that $\tfrac{1}{2}(\sigma_1 + \sigma_2) = \overline{\sigma}$ is the average shear stress on the fault so this is analogous to "work = force × distance" from basic physics. As discussed in Kanamori and Brodsky (2004) and Kanamori and Rivera (2006), this is usually approximated as

$$E = \overline{\sigma}\,\overline{D}A = \tfrac{1}{2}\Delta\sigma\overline{D}A + \sigma_2\overline{D}A \qquad (7.11)$$

where the average stress drop $\Delta\sigma = \sigma_1 - \sigma_2$. The total energy can be partitioned into three parts:

$$E = E_R + E_F + E_G \qquad (7.12)$$

where E_R is the radiated seismic energy, E_F is the frictional energy (often released as heat), and E_G is the energy used to fracture the rock, although the separation between E_F and E_G is not always clear cut. In principle, E_R and E_G can be estimated from seismic data. However, E_F cannot be measured from direct seismic wave observations and depends upon the absolute level of stress on the fault, which is difficult to determine.

This energy balance is shown graphically in Figure 7.3 for two idealized earthquakes on faults of unit area and total displacement D. In the first example, the *Orowan fault model* (e.g., Orowan, 1960; Kostrov, 1974), the stress on the fault, σ_f, drops abruptly to σ_2 as soon as the fault starts moving. In this case, there is no fracture energy, E_G, and σ_2 represents the dynamic frictional stress on the fault. The total energy released is the shaded trapezoid, which is the sum of E_R and E_F.
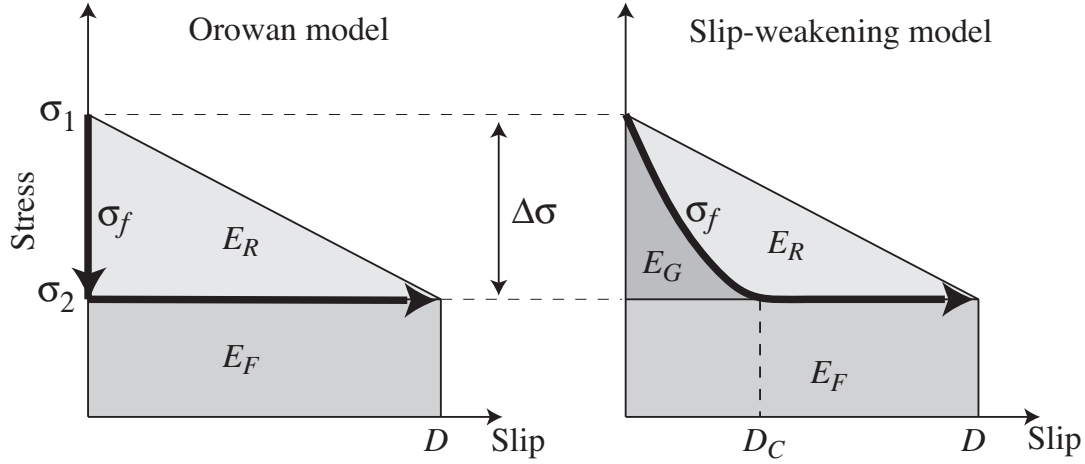
Figure 7.3: The shear stress, σ_f, on a point on a fault as a function of slip for the Orowan fault model and a simple example of a slip-weakening fault model. σ_1 and σ_2 are the initial and final stresses, D is the total slip, D_C is the critical slip, E_R is the radiated seismic energy, E_F is the frictional energy dissipated, and E_G is the fracture energy released.

Generalizing to a fault of area A, we have

$$E_R \quad = \quad \tfrac{1}{2}(\sigma_1 - \sigma_2)DA = \tfrac{1}{2}\Delta\sigma DA \qquad (7.13)$$

$$E_F \quad = \quad \sigma_2 DA \qquad\qquad\qquad\qquad (7.14)$$

In this case, the stress drop can be expressed as

$$\Delta\sigma(\text{Orowan}) = \frac{2E_R}{DA} = \frac{2\mu E_R}{M_0} = 2\mu\tilde{e} \qquad (7.15)$$

and we see that this model predicts a very simple relationship between stress drop and scaled energy, $\tilde{e}$. This is sometimes termed the *Orowan stress drop* to make clear that it only represents the true stress drop if the earthquake obeys this simple model. Assuming $\Delta\sigma = 3$ MPa and $\mu = 30$ GPa (typical values for crustal earthquakes), the Orowan model predicts $\tilde{e} = 5 \times 10^{-5}$, which is in rough agreement with direct observations of $\tilde{e}$ for large earthquakes (see Fig. 7.2).

In general, however, we expect the rupture process to be more involved than the Orowan model and the σ_f function may follow a complicated trajectory. In some models, σ_f rises above σ_1 at the onset of rupture to what is termed the yield stress before dropping as slip begins. It is also possible for σ_f to fall below σ_2 during part of the rupture and for σ_f to end at a value above or below the final stress state once

the earthquake is completely over (the latter phenomena are called overshoot and undershoot, respectively, and are predicted by some theoretical models).

The right part of Figure 7.3 shows an example of a *slip-weakening model* in which the stress drops from σ_1 to σ_2 over a distance D_C (sometimes called the *critical slip*) and then continues at a constant stress $\sigma_f = \sigma_2$. The radiated seismic energy, E_R, is reduced by the area to the left of the curve, which represents the fracture energy E_G. In this case we have

$$E_G = E - E_F - E_R = \tfrac{1}{2}\Delta\sigma DA - E_R = \frac{\Delta\sigma}{2\mu}M_0 - E_R \qquad (7.16)$$

and

$$\Delta\sigma = \frac{2\mu(E_R + E_G)}{M_0} \geq \Delta\sigma(\text{Orowan}) \qquad (7.17)$$

and we see that in principle we can estimate the fracture energy E_G if we are able to separately measure M_0, $\Delta\sigma$ and E_R, and that the Orowan stress drop represents the minimum possible stress drop, given values of E_R and M_0, at least for simple models in which $\sigma_f \geq \sigma_2$. It should be noted that σ_f for real earthquakes may follow more complicated trajectories than those plotted in Figure 7.3, in which case E_F is not determined by the final stress and the partitioning in (7.16) and (7.17) between E_F and E_G does not necessarily have physical significance in the faulting process.

The *radiation efficiency* is defined as the ratio

$$\eta_R = \frac{E_R}{E_R + E_G} \qquad (7.18)$$

and is an important measure of the dynamic properties of earthquakes. Note that $\eta_R = 1$ for the Orowan fault model. For our simple slip-weakening model, it can be expressed as

$$\eta_R = \frac{E_R}{\tfrac{1}{2}\Delta\sigma DA} = \frac{2\mu}{\Delta\sigma}\frac{E_R}{M_0} = 2\mu\frac{\tilde{e}}{\Delta\sigma}, \qquad (7.19)$$

and thus is proportional to the ratio between the scaled energy and the stress drop. As discussed in Kanamori and Brodsky (2004), the radiation efficiency can be related to the rupture velocity, v_r, in theoretical crack models:

$$\eta_R = 1 - g(v_r) \qquad (7.20)$$

where $g(v_r)$ is a function that depends upon the specific crack model and the ratio of v_r to the Rayleigh or shear wave velocity. For example, for Mode III (transverse shear) cracks,

$$g(v) = \sqrt{\frac{1 - v_r/\beta}{1 + v_r/\beta}}, \tag{7.21}$$

in which case η_R approaches one and the fracture energy, E_G, goes to zero as the rupture velocity approaches the shear wave velocity. For about 30 earthquakes of $6.6 < M_W < 8.3$, Venkataraman and Kanamori (2004) obtained radiation efficiency estimates generally between 0.25 and 1.0. One class of earthquakes that appear to have $\eta_R < 0.25$ are *tsunami earthquakes*, which involve slow rupture and generate large tsunamis relative to their moment.

The radiation efficiency should not be confused with the *seismic efficiency*, η, defined as the fraction of the total energy that is radiated into seismic waves:

$$\eta = \frac{E_R}{E} = \frac{E_R}{\overline{\sigma}\overline{D}A} = \frac{\mu E_R}{\overline{\sigma}M_0} = \frac{\mu\tilde{e}}{\overline{\sigma}}. \tag{7.22}$$

The seismic efficiency is more difficult to estimate than the radiation efficiency because it depends upon the poorly constrained absolute stress level on the fault.

In the extreme case where we assume that the earthquake relieves all of the stress on the fault, then $\sigma_2 = 0$ and we say that the stress drop is total. In this case, $E_F = 0$ and we have

$$E_{\min} = \tfrac{1}{2}\overline{\Delta\sigma}\overline{D}A = \frac{\Delta\sigma}{2\mu}M_0 \tag{7.23}$$

This represents the minimum amount of energy release for an earthquake with a given stress drop and moment.

The theories that describe how slip on a fault initiates, propagates and comes to a halt can be very complicated, even for idealized models with uniform pre-stress and elastic properties. Much of the recent work in this area has involved theory and observations of *rate and state friction* (e.g., Dieterich, 1994) in which the frictional properties are time and slip dependent. Because these models vary in their behavior and it is likely that real earthquakes span a range of different rupture properties, it is important to keep in mind the distinction between parameters that are more-or-less directly estimated (e.g., moment, geodetically-determined static stress drop, and

radiated energy) and those that depend upon modeling assumptions (e.g., Brune-type and Orowan stress drops) and thus are not truly independent measurements. For example, it would make little sense to use Equation (7.16) to estimate E_G if both $\Delta\sigma$ and E_R are derived from fitting the observed body-wave spectra to the same theoretical model.

7.2 Additional reading

Allmann, B.P., and P.M. Shearer, Spatial and temporal stress drop variations in small earthquakes near Parkfield, California, *J. Geophys. Res.*, **112**, B4, B04305, doi:10.1029/2006JB004395, 2007.

Allmann, B. B., and P. M. Shearer, Global variations of stress drop for moderate to large earthquakes, *J. Geophys. Res.*, **114**, doi: 10.1029/2009JB005821, 2009.

Prieto, G.A., Shearer, P.M., Vernon, F.L. and Kilb, D. (2004). Earthquake source scaling and self-similarity estimation from stacking P and S spectra, *J. Geophys. Res.*, **109**, doi: 1029/2004JB003084.

Shearer, P. M., G. A. Prieto, and E. Hauksson, Comprehensive analysis of earthquake source spectra in southern California, *J. Geophys. Res.*, **111**, B06303, doi:10.1029/2005JB003979, 2006.

7.3 References

Abercrombie, R.E. (1995). Earthquake source scaling relationships from -1 to 5 M_L using seismograms recorded at 2.5-km depth, *J. Geophys. Res.*, **100**, 24,015–24,036.

Aki, K. (1967). Scaling law of seismic spectrum, *J. Geophys. Res.*, **72**, 1217–1231.

Boatwright, J. and J.B. Fletcher (1984). The partition of radiated energy between P and S waves, *Bull. Seismol. Soc. Am.*, **74**, 361–376.

Brune, J. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes, *J. Geophys. Res.*, **75**, 4997–5009.

Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.*, **99**, 2601–2618.

Heaton, T. H. (1990). Evidence for and implications of self-healing pulses of slip in earthquake rupture, *Phys. Earth Planet. Inter.*, **64**, 1–20.

Ide, S. and Beroza, G.C. (2001). Does apparent stress vary with earthquake size?, *Geophys. Res. Lett.*, **28**, 3349–3352.

Kanamori, H. and Anderson, D. L. (1975). Theoretical basis of some empirical relations in seismology, *BSSA*, **65**, 1073–1095.

Kanamori, H. and Brodsky, E. (2004). The physics of earthquakes, *Rep. Prog. Phys.*, **67**, 1429–1496.

Kanamori, H. and Rivera, L. (2006). Energy partitioning during an earthquake, in *Earthquakes: Radiated Energy and the Physics of Faulting, Geophysical Monograph Series 170*, American Geophysical Union, Washington, doi: 10.1029/170GM03, 3–13.

Kostrov, B. (1974). Seismic moment and energy of earthquakes and seismic flow of rock, *Izv. Acad. Sci., USSR, Phys. Solid Earth (Engl. Transl.)*, **1**, 23–40.

Ma, S. and Archuleta, R.J. (2006). Radiated seismic energy based on dynamic rupture models of faulting, *J. Geophys. Res.*, **111**, doi: 10.1029/2005JB004055.

Madariaga, R. (1976). Dynamics of an expanding circular fault, *Bull. Seismol. Soc. Am.*, **66**, 639–666.

Mayeda, K. and Walter, W.R. (1996). Moment, energy, stress drop, and source spectra of western United States earthquakes from regional coda envelopes, *J. Geophys. Res.*, **101**, 11,195–11,208.

Orowan, E. (1960). Mechanism of seismic faulting, *Geol. Soc. Am. Bull.*, **79**, 323–345.

Pérez-Campos, X. and Beroza, G.C. (2001). An apparent mechanism dependence of radiated seismic energy, *JGR*, **106**, 11,127–11,136.

Scholz, C.H. (1982). Scaling laws for large earthquakes: Consequences for physical models, *Bull. Seismol. Soc. Am.*, **72**, 1–14.

Scholz, C.H. (1997). Size distributions for large and small earthquakes, *Bull. Seismol. Soc. Am.*, **87**, 1074–1077.

Singh, S.K. and M. Ordaz (1994). Seismic energy release in Mexican subduction zone earthquakes, *Bull. Seismol. Soc. Am.*, **84**, 1533–1550.

Thatcher, W. (1975). Strain accumulation and release mechanism of the 1906 San Francisco earthquake, *J. Geophys. Res.*, **80**, 4862–4872.

Venkataraman, A., Beroza, G.C., and Boatwright, J. (2006). A brief review of techniques used to estimate radiated seismic energy, in *Earthquakes: Radiated Energy and the Physics of Faulting, Geophysical Monograph Series 170*, American Geophysical Union, Washington, 15–24, doi: 10.1029/170GM04.

Venkataraman, A. and Kanamori, H. (2004). Observational constraints on the fracture energy of subduction zone earthquakes, *J. Geophys. Res.*, **109**, doi: 10.1029/2003JB002549.

Chapter 8

Earthquake triggering

The events in real earthquake catalogs occur at apparently random times, with the exception of aftershock sequences. Earthquake times over large regions can be modeled reasonably well as a Poisson process, that is, the probability of an earthquake at any given time is constant and independent of the time of the last event[1]. However, at small scales there is a noticeable clustering of earthquakes in time and space that violates the simple Poisson model. Much of this can be explained as aftershock triggering, but some features, such as swarms without a clear mainshock, appear to require other mechanisms. In general, it is difficult to completely separate aftershocks from other earthquakes because even small events increase the probability of future events to some extent.

The most obvious example of non-random earthquake occurrence is the existence of aftershock sequences after large earthquakes. Although the exact timing of individual events is still random, an increased rate of activity is observed that is temporally and spatially correlated with the mainshock. The seismicity rate decays with time, following a power law relationship, called *Omori's Law* after Omori (1894),

$$n(t) = \frac{K}{t + c} \tag{8.1}$$

where $n(t)$ is the number of aftershocks per unit time above a given magnitude, t is the time measured from the mainshock, and K and c are constants. This is often

---

[1]For example, see Gardner and Knopoff, 1974, "Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian?" This paper is famous for having the shortest abstract in the geophysical literature—it simply says "Yes."
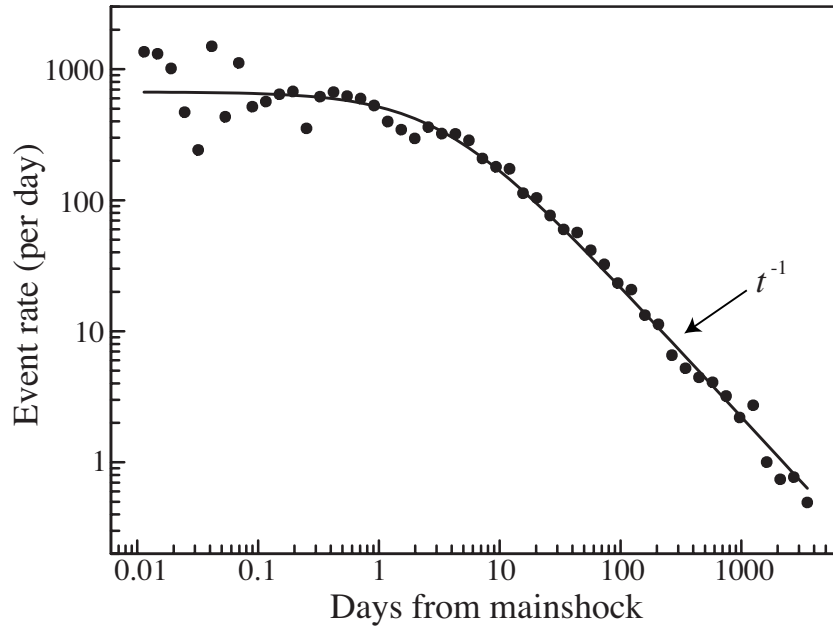
Figure 8.1: Aftershock rate for the 1994 Northridge, California, earthquake as a function of time after the mainshock. The line shows the Omori's law prediction for $K = 2230$, $c = 3.3$ days, and $p = 1$. Data are from the Southern California Seismic Network catalog within a lat/lon window of (34.2°, 34.45°, -118.75°, -118.3°).

generalized to the modified Omori's Law

$$n(t) = K(t + c)^{-p} \tag{8.2}$$

in which p is typically close to 1.

As an example, Figure 8.1 plots the aftershock rate following the 1994 Northridge, California, earthquake ($M_W = 6.7$), which is well fit by Omori's law with $c = 3.3$ days and $p = 1$. The parameter c is related to a relative deficit of aftershocks immediately following the earthquake compared to a simple uniform power law. For the Northridge example, this is mainly caused by the inability of the seismic network to detect and locate the large number of events occurring in the first few days after the earthquake (e.g., Kagan and Houston, 2005). When care is taken to obtain a more complete catalog, the deficit in early aftershocks lasts only a few minutes after the mainshock (Peng et al., 2007).

Earthquakes are thought to trigger aftershocks either from the dynamic effects of their radiated seismic waves or the resulting permanent static stress changes (for reviews, see Harris, 2002, and Freed, 2005). A common assumption based on

rock behavior in laboratory experiments is that earthquake occurrence on a fault is promoted by increases in the *Coulomb failure function* (CFF)

$$\text{CFF} = |\tau_s| + \mu(\tau_n + P), \tag{8.3}$$

where τ_s is the shear traction on the fault, τ_n is the normal traction (positive for tension), P is the pore fluid pressure, and μ is the coefficient of static friction (don't confuse this with the shear modulus!). The second term is negative because in our sign convention τ_n is negative for the hydrostatic compression forces at depth. Thus, increases in shear stress or decreases in fault normal compression (which "unclamp" the fault) will encourage failure, and the opposite changes will discourage failure. Numerous studies have searched for possible spatial correlations between aftershock occurrence and the sign of the CFF change predicted by mainshock slip models, and many have found that there tend to be more aftershocks in regions where the static stress changes should promote earthquakes (e.g., Reasenberg and Simpson, 1992; Harris and Simpson, 1992; Stein et al., 1992; Stein, 1999). However, these correlations are not perfect and some aftershocks occur even in areas where the CFF changes are negative. The relative importance of static and dynamic triggering for aftershocks is also not yet firmly established. Dynamic stress changes from seismic waves often trigger earthquakes at large distances from mainshocks and some have argued that dynamic effects could be the dominant triggering mechanism for near-field aftershocks as well (e.g., Kilb et al., 2000; Felzer and Brodsky, 2006).

An obvious and important aspect of aftershocks is that they don't all occur instantly at the time of the mainshock—they have a time dependence that is described by Omori's law. This indicates that whatever their triggering mechanism, it must initiate a time-dependent failure process that causes events to occur at a wide range of times following the mainshock. There cannot simply be a precise threshold stress that, when exceeded, immediately triggers earthquakes. Additional evidence for this come from the lack of an obvious correlation between earthquake occurrence time and the solid Earth tides. Daily variations in crustal stresses caused by the tides greatly exceed the daily accumulation of stress from tectonic loading. Thus, any threshold level of stress will be first exceeded only at certain times in the tidal cycle,

which might be expected to produce strong periodicities in earthquake occurrence times. Many researchers have searched for tidal signals in earthquake catalogs, but the most careful studies (e.g., Vidale et al., 1998) have found little or no correlation between earthquakes and tidal stresses.

Omori's law does not say anything about the magnitude distribution of the aftershocks or their spatial relationship to the mainshock. However, by combining Omori's law with the Gutenberg-Richter magnitude-frequency law[2] and other empirical relationships, one can develop general models that predict the probability of future events based on the record of previous seismicity. The most well-known of these is called the *Epidemic Type Aftershock-Sequences* or *ETAS* model (for reviews, see Ogata, 1999, and Helmstetter and Sornette, 2002). In the ETAS model, every earthquake, no matter how small, increases the probability of future nearby events. The increased probability is greatest immediately after an earthquake and then decreases following Omori's law until it reaches a background level of seismicity. These models do not require that aftershocks always be smaller than the triggering event. Sometimes mainshocks can be considered really big aftershocks of a *foreshock*, a smaller preceding earthquake that is spatially and temporally near the mainshock. Thus when any earthquake occurs, the possibility that it might be a foreshock increases the probability that a larger earthquake will soon follow. In California, for example, it has been estimated that an M 5.3 earthquake on the San Gorgonio Pass segment of the San Andreas Fault would produce a 1% chance of a much larger earthquake occurring within the next 3 days (Agnew and Jones, 1991).

The ETAS model in its original form does not include any spatial constraints on aftershock probabilities, that is the observed decay in aftershock density with distance from the mainshock. Felzer and Brodsky (2006) have explored this decay rate, and incorporating their result into the ETAS model of Ogata (1999), a general

---

[2]The G-R law is $\log_{10} N = a - bM$, where N is the number of events with magnitudes greater than or equal to M. In this equation, a describes the total number of earthquakes, while the parameter b is called the *b-value* and measures the relative number of large quakes compared to small quakes. The b-value is generally found to lie between 0.8 and 1.2 for a wide variety of regions and different magnitude scales (for a review, see Utsu, 2002a).

equation for estimated earthquake probability is

$$\lambda(\mathbf{x}, t) = \lambda_0 + \sum_i \kappa 10^{\alpha(m_i - m_0)} (t_i + c)^{-p} r_i^{-q} \qquad (8.4)$$

where $\lambda(\mathbf{x}, t)$ is the predicted event density (events per unit volume and unit time) at position $\mathbf{x}$ and time t, λ_0 is a background rate (untriggered), which in general may be spatially varying, κ is a triggering productivity parameter, the summation is taken over all events in the catalog prior to t, m_i is the magnitude of each earthquake, m_0 is the minimum magnitude of the counted events, α (≈ 1) accounts for the fact that larger earthquakes trigger more events, t_i is the time from the ith event to t, c and p (≈ 1) are the Omori decay constants, r_i is the distance from the ith event to $\mathbf{x}$, and q defines the decay with distance. This type of model is an attempt to quantify the clustering in time and space of seismicity, i.e., the common observation that earthquakes are most likely to occur near recent earthquake activity. Note that this equation approximates earthquakes as point sources and would require modification to accurately predict aftershock density around the extended rupture of a large earthquake. By including the Gutenberg-Richter b-value relation, these models can also be used to estimate the probability as a function of earthquake size. Most of these models are purely empirical, but there have also been attempts to create physical models based on time-dependent failure mechanisms, such as the rate-and-state friction laws of Deterich (1994). Although this work is unlikely to lead to deterministic predictions of individual events, understanding how earthquake occurrence relates to prior events is important for developing more accurate earthquake probability forecasts.

An example of this is the U.S. Geological Survey sponsored effort to provide realtime estimates of the probability of significant ground shaking in California. As shown in Figure 8.2, the probability of earthquake occurrence increases following every earthquake, but then decays back to the background rate. Large earthquakes increase the risk of future events more than small earthquakes. Using seismicity catalogs it is possible to use such a model to predict the instantaneous probability of earthquake occurrence of a given magnitude as a function of location. This can be combined with the known relationship between earthquake magnitude and shaking
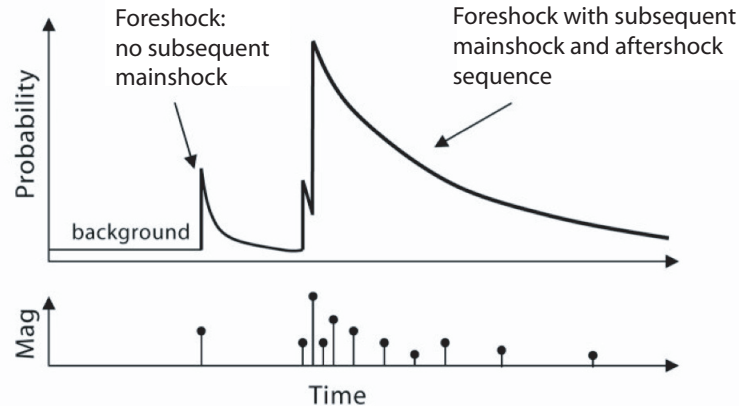
Figure 8.2: A cartoon illustrating how earthquake probability increases immediately after prior events, and then decays back to the background seismicity rate. Figure adopted from web material at: http://pasadena.wr.usgs.gov/step/.

intensity in California, as measured, for example, by the Mercalli scale (see Fig. **??**). Integrating over all locations, the result can be plotted as the probability of ground motion exceeding a specified Mercalli intensity within a 1-day period. An example of the results of this calculation is shown in Figure 8.3, comparing just before to just after the 28 September 2004 Parkfield earthquake (M 6). The probability of damaging shaking is greatly increased in a large region around Parkfield, reflecting the chance that the Parkfield earthquake might be a foreshock of a larger earthquake.

Although space-time clustering of earthquakes is clearly observed, the physics behind this clustering is not well understood. Some earthquake clusters, such as mainshock-aftershock sequences are most likely caused by triggering of events by previous events (although whether this occurs primarily as a result of static or dynamic stress changes is still debated). In other cases, such as earthquake swarms that lack an initiating mainshock, it seems more likely that the earthquakes are triggered by some underlying physical process, such as slow creep or fluid movement (e.g., Vidale and Shearer, 2006). These questions are also relevant to when foreshock sequences are observed prior to large earthquakes. Do the mainshocks occur simply because the foreshock activity itself increases the likelihood of a big earthquake, or could they be symptomatic of an underlying physical process that ultimately causes the larger event? The latter scenario provides more hope for prediction of large earthquakes, if the physical process could be understood more completely.
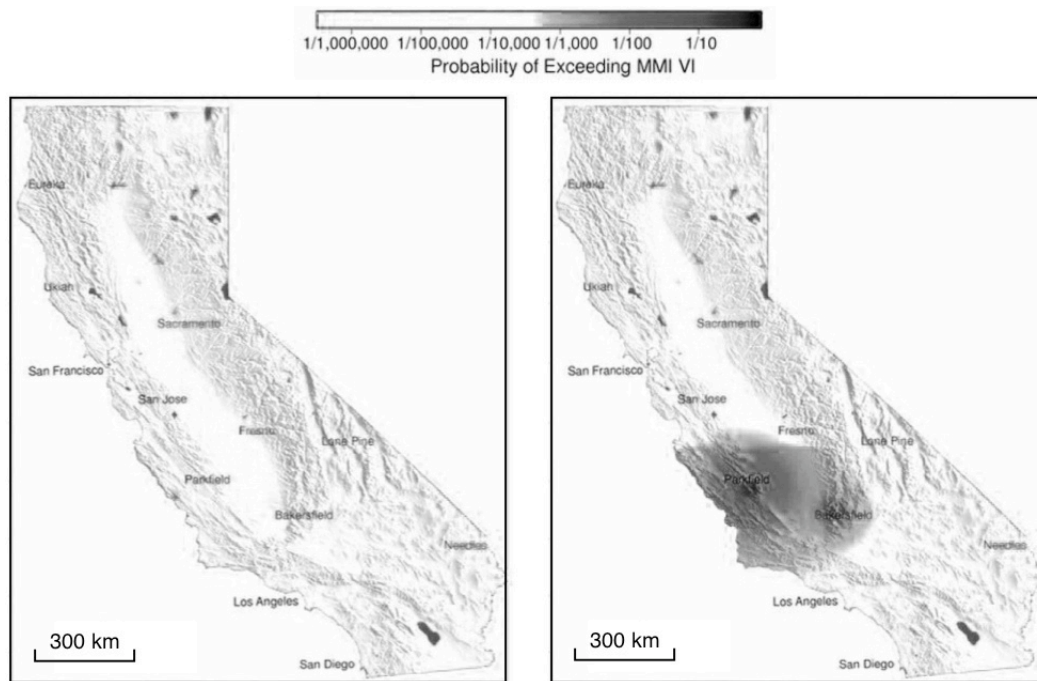
Figure 8.3: The probability of local ground motions of Modified Mercalli intensity 6 or greater within a 24-hour period, immediately before and after the 2004 Parkfield earthquake in California. Source: http://pasadena.wr.usgs.gov/step/

8.1 Additional reading

Shearer, P. M., and G. Lin, Evidence for Mogi doughnut behavior in seismicity preceding small earthquakes in southern California, *J. Geophys. Res.*, **114**, doi: 10.1029/2009JB005982, 2009.

Vidale, J.E., and Shearer, P.M. (2006). A survey of 71 earthquake bursts across southern California: Exploring the role of pore fluid pressure fluctuations and aseismic slip as drivers, *J. Geophys. Res.*, **111**, doi: 10.1029/2005JB004034.

8.2 References

Agnew, D.C., and Jones, L.M. (1991). Prediction probabilities from foreshocks, *J. Geophys. Res.*, **96**, 11,959–11,971.

Dieterich, J. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *J. Geophys. Res.*, **99**, 2601–2618.

Felzer, K.R. and Brodsky, E.E. (2006). Decay of aftershock density with distance indicates triggering by dynamic stress, *Nature*, **441**, 735–738.

Freed, A.M. (2005). Earthquake triggering by static, dynamic, and postseismic stress transfer, *Annu. Rev. Earth Planet. Sci.*, **33**, 335–367, doi: 10.1146/annurev.earth.33.092203.122505.

Gardner, J. K. and Knopoff, L. (1974). Is the sequence of earthquakes in southern

California, with aftershocks removed, Poissonian?, *Bull. Seismol. Soc. Am.*, **64**, 1363–1367.

Harris, R. (2002). Stress triggers, stress shadows, and seismic hazard, in *International Handbook of Earthquake and Engineering Seismology, Part A*, ed. W.H. Lee, Academic Press, San Diego, 1217–1232.

Harris, R. A. and Simpson, R. W. (1992). Changes in static stress on southern California faults after the 1992 Landers earthquake, *Nature*, **360**, 251–254.

Helmstetter, A., and Sornette, D. (2002). Diffusion of epicenters of earthquake aftershocks, Omori's law, and generalized continuous-time random models, *Phys. Rev. E*, **66**, 061104.

Kagan, Y.Y. and Houston, H. (2005). Relation between mainshock rupture process and Omori's law for aftershock moment release rate, —*GJI*, **163**, 1039–1048, doi: 10.1111/j.1365-246X.2005.02772.x.

Kilb, D., Gomberg, J., and Bodin, P. (2000). Triggering of earthquake aftershocks by dynamic stresses, *Nature*, **408**, 570–574, doi: 10.1038/35046046.

Ogata, Y. (1999). Seismicity analysis through point-process modeling: a review, *Pure Appl. Geophys.*, **155**, 471–507.

Omori, F. (1894). On the aftershocks of earthquakes, *J. Coll. Sci. Imp. Univ. Tokyo*, **7**, 111–216.

Peng, Z., Vidale, J.E., Ishii, M., and Helmstetter, A. (2007). Seismicity rate immediately before and after main shock rupture from high-frequency waveforms in Japan, *J. Geophys. Res.*, **112**, doi: 10.1029/2006JB004386.

Reasenberg, P.A. and Simpson, R.W. (1992). Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake, *Science*, **255**, 1687–1690.

Stein, R.S. (1999). The role of stress transfer in earthquake occurrence, *Nature*, **402**, 605–609, doi: 10.1038/45144.

Stein, R. S., King, G. C. P., and Lin, J. (1992). Changes in failure stress on the southern San Andreas fault system caused by the 1992 magnitude = 7.4 Landers earthquake, *Science*, **258**, 1328–1332.

Utsu, T. (2002a). Statistical features of seismicity, in *International Handbook of Earthquake and Engineering Seismology, Part A*, ed. W.H. Lee, Academic Press, San Diego, 719–732.

Vidale, J.E., Agnew, D.C., Johnston, M.J.S., and Oppenheimer, D.H. (1998). Absence of earthquake correlation with Earth tides: an indication of high preseismic fault stress rate, *J. Geophys. Res.*, **103**, 24,567–24572.