

Aspects of Observational Seismology

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

Notes for Caltech class
January–March, 2010

Preface

These notes are to accompany a course that I will be teaching during my January to March 2010 visit to Caltech. The topics are related to my own seismology research projects and are rather diverse. Thus, each chapter is largely independent of the others. 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

Introduction

Seismology is the study of earthquakes and seismic waves and what they tell us about Earth structure. Seismology is a data-driven science and its most important discoveries usually result from analysis of new data sets or development of new data analysis methods. Most seismologists spend most of their time studying *seismograms*, which are simply a record of Earth motion at a particular place as a function of time. Fig. 1.1 shows an example seismogram.

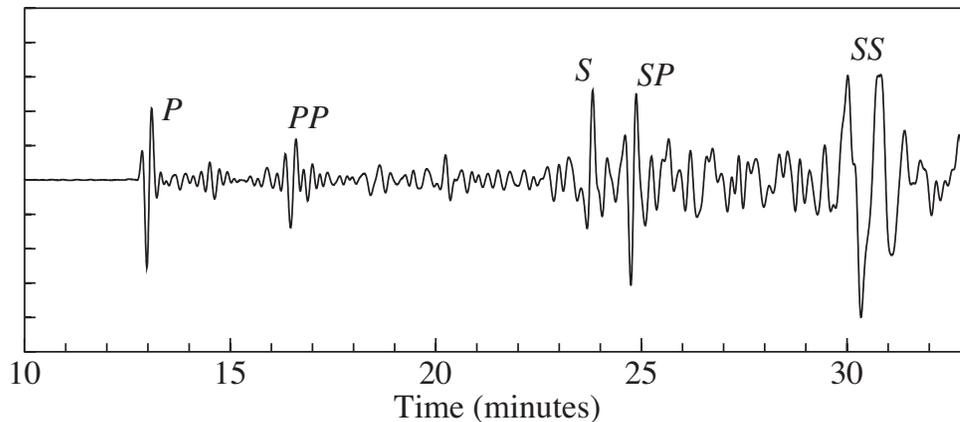


Figure 1.1: The 1994 Northridge earthquake recorded at station OBN in Russia. Some of the visible phases are labeled.

Modern seismograms are digitized at regular time intervals and analyzed on computers. Many concepts of time series analysis, including filtering and spectral methods, are valuable in seismic analysis. Although continuous background Earth “noise” is sometimes studied, most seismic analyses are of records of discrete sources of seismic wave energy, i.e., earthquakes and explosions. The appearance of these

seismic records varies greatly as a function of the source-receiver distance. The different distance ranges are often termed:

1. *Local* seismograms occur at distances up to about 200 km. The main focus is usually on the direct P waves (compressional) and S waves (shear) that are confined to Earth's crust. Analysis of data from the Southern California Seismic Network for earthquake magnitude and locations falls into this category. Surface waves are not prominent although they can sometimes be seen at very short periods.
2. *Regional* seismology studies examine waveforms from beyond ~ 200 up to 2000 km or so. At these distances, the first seismic arrivals travel through the upper mantle (below the Moho that separates the crust and mantle). Surface waves become more obvious in the records. Analysis of continent-sized data sets is an example of regional seismology, such as current USArray project to deploy seismometers across the United States.
3. *Global* seismology is at distances beyond about 2000 km ($\sim 20^\circ$), where seismic wave arrivals are termed *teleseisms*. This involves a multitude of body-wave phase arrivals, arising from reflected and phase-converted phases from the surface and the core-mantle boundary. For shallow sources, surface waves are the largest amplitude arrivals. Data comes from the global seismic network (GSN).

The appearance of seismograms also will vary greatly depending upon how they are filtered. What seismologists term short-period records are usually filtered to frequencies above about 0.5 Hz. What seismologists term long-period records are filtered to below about 0.1 Hz (above 10 s period). Examples of short- and long-period waveform stacks are shown in Figs. 1.2 and 1.3. Note the different phases visible at the different periods. I will discuss how these images are constructed a little later in this chapter.

You can learn a lot from a single seismogram. For example, if both P and S arrivals can be identified, then the $S - P$ time can be used to estimate the distance

Short-period (vertical)

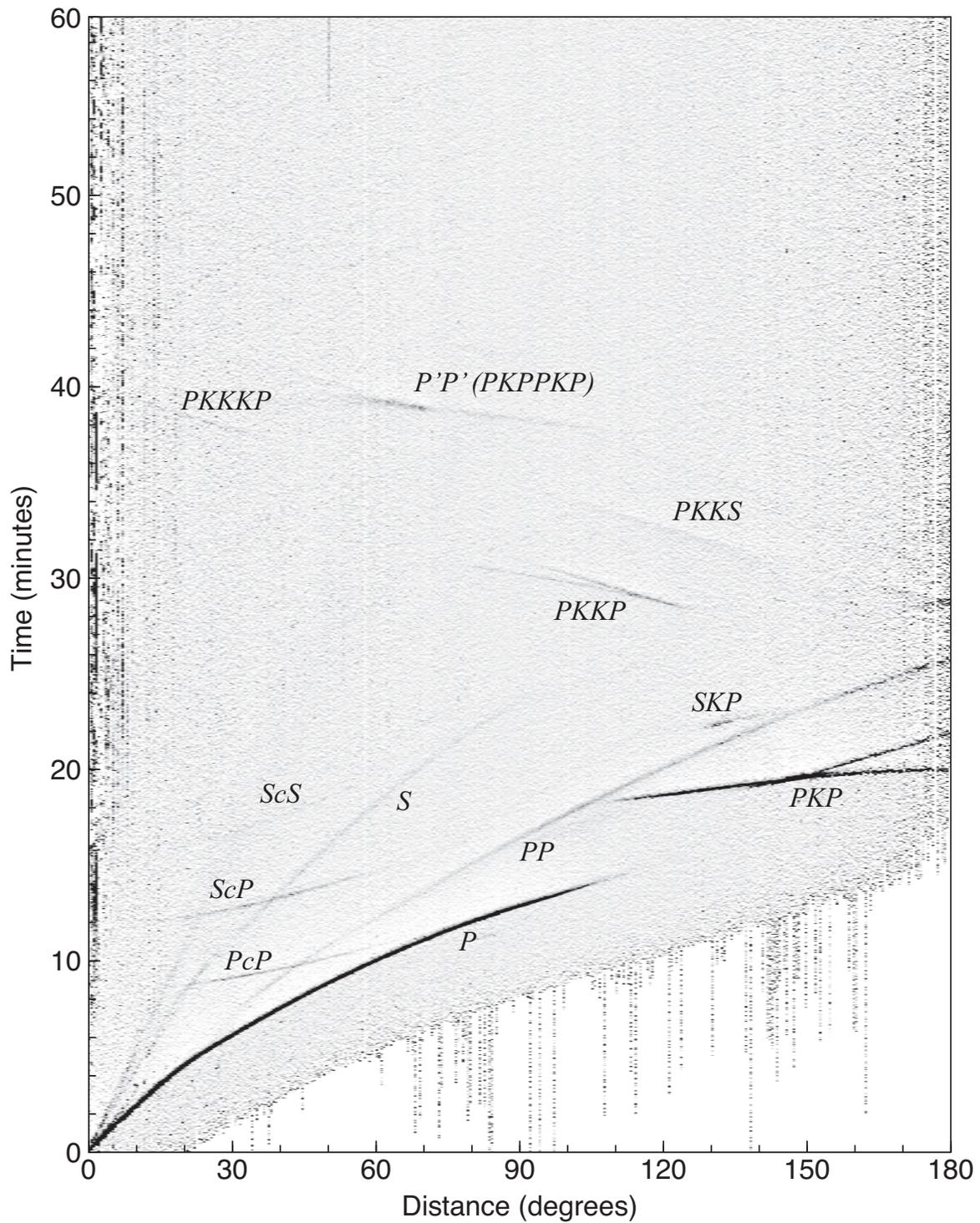


Figure 1.2: A stack of short-period (<2 s), vertical component data from the global networks between 1988 to 1994. (From Astiz et al., 1996.)

Long-period (vertical)

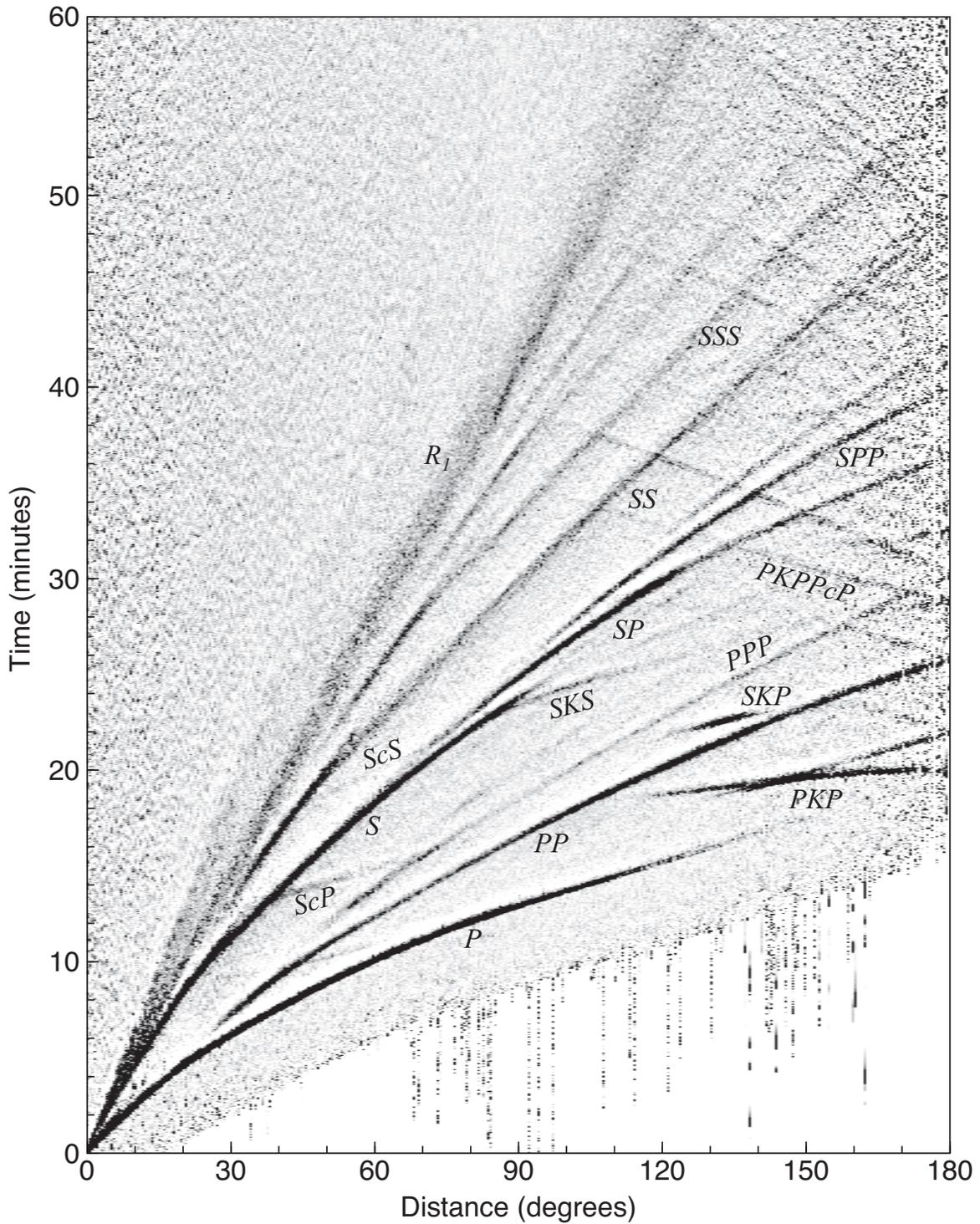


Figure 1.3: A stack of long-period (>10 s), vertical component data from the global networks between 1988 to 1994. (From Astiz et al., 1996.)

to the source. A rule of thumb in local seismology is that the distance in kilometers is about 8 times the $S - P$ time in seconds. Once the distance is known, the P - or S -wave amplitude can be used to estimate the event magnitude (S waves are usually used for local records, P waves in the case of teleseisms). In global seismology, surface wave *dispersion* (arrival times varying as a function of frequency) can be used to constrain seismic velocity as a function of depth.

You can learn more if you have a three-component (vertical and two orthogonal horizontal sensors) seismic station that fully characterizes the vector nature of ground motion. In this case, the angle that a seismic wave arrives at the station can be estimated, which permits an approximate source location to be determined. One can then separate the surface waves into Rayleigh waves (polarized in the source direction) and Love waves (polarized at right angles to the source direction). Sometimes the S arrival will be observed to be split into two orthogonal components of motion (see Fig. 1.4). This is called *shear-wave splitting* and is diagnostic of seismic anisotropy, in which wave speed varies as a function of direction in a material. The orientation and magnitude of the anisotropy can be estimated from shear-wave splitting observations. Sometimes, a weak S -wave arrival will be observed to follow the teleseismic P -wave arrival, which is caused by a P -to- S converted phase at the Moho (the crust-mantle boundary). The timing of this phase can be used to estimate crustal thickness.

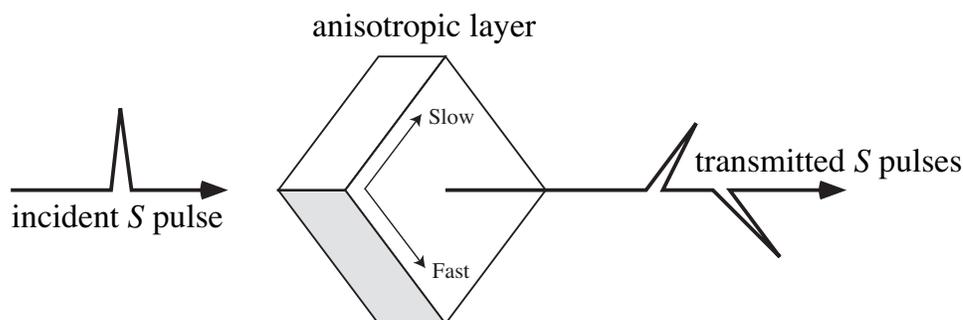


Figure 1.4: An S -wave that travels through an anisotropic layer can split into two S -waves with orthogonal polarizations; this is due to the difference in speed between the qS waves in the anisotropic material.

But much, much more can be learned when data from many different seismic stations are available. In the early days of seismology, seismic stations were rare and

expensive and often operated separately by different institutions. But the importance of sharing data to determine accurate earthquake locations and Earth velocity structure was very clear. Thus seismology began a tradition of free and open sharing of data that continues to this day. This has been facilitated by centralized data repositories, initially just the arrival times and amplitudes measured by the different station operators, then to the actual seismograms themselves, as archived in film chip libraries, and eventually to modern digital seismic networks and data centers.

This data-sharing tradition is a very appealing part of being a seismologist. It's easy for any of us to obtain an abundance of data. All you have to do is go online to the appropriate data center (SCECDC for southern California, the IRIS DMC for GSN data). You don't have to know anybody there, you don't have to ask somebody to do you a favor to get the data, you don't have to collaborate with anybody. The data are simply there for you to use. Seismic networks are funded by the community for the entire community to use. Even data from PI-funded experiments to study particular regions typically are released 18 months after the experiment ends. Indeed, the National Science Foundation (NSF) requires that all data collected in NSF-funded experiments be made available through data centers. Furthermore, seismology is very data rich and most seismic data sets have not been fully analyzed. People are constantly discovering new things in old data sets.

This makes it possible for individual seismologists (and grad students) to make important contributions without having to get an experiment funded or to join part of some large team. Most published seismology papers have 1 to 4 authors. Compare that to the particle physics community! Other fields are starting to recognize the importance and power of data sharing. For example, the astronomy community is beginning to fund automated sky surveys where all of the data will immediately be available online. But seismology has a 100 year head start in open access data sharing.

In addition, increases in computer capabilities mean that individual seismologists can now access and analyze larger and larger data sets. The global seismogram archive is growing (see Fig. 1.5, but hard drive storage capacity is growing even faster. This means that within a few years you will be able to store many key

seismic data sets on your laptop.

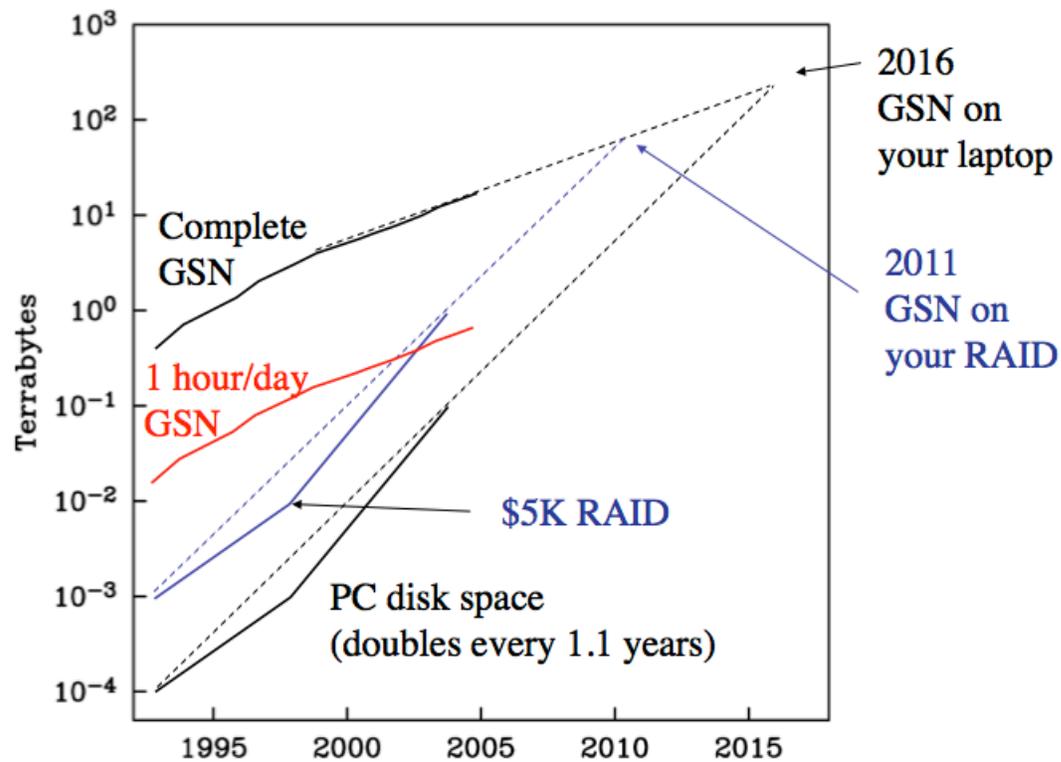


Figure 1.5: The global seismogram archive is growing rapidly, but not as fast as computer storage capabilities.

As more data have become available, more ambitious projects become possible. If you look at really old issues of BSSA (*Bulletin of the Seismological Society of America*), you will find articles about individual earthquakes where the author has simply located the earthquake and perhaps computed a magnitude and a focal mechanism. Now these are done as a routine part of network operations. Similarly, structural seismology studies have evolved from solving for 1-D (radially symmetric) Earth models to using a multitude of sources and receivers to solve for 3-D seismic velocity perturbations (often called seismic *tomography*).

Much of my research has involved trying to exploit the observational opportunities of large data sets. Some principles for efficient seismic data crunching include:

1. Analyze the entire dataset whenever possible.
2. Use simple methods to get a sense of the the data before doing complicated

inversions.

3. Consider using reflection seismology methods like stacking and back-projection.
4. Avoid any hand-processing of seismograms!

The global waveform images of Figs. 1.2 and 1.3 illustrate these principles. At the time they were made (Astiz et al., 1996), they were a result of processing all of the global seismograms available at the IRIS Data Center for earthquakes above M 5.7. At the time this was over 33,000 seismograms for sources shallower than 50 km. How can we condense all of this information into a single image? We decided to make a time versus distance image of the seismic wavefield. This is sometimes called the record-section format for displaying seismograms. Because Earth's seismic velocity variations are generally much greater in the vertical (radial) direction than in the horizontal direction, this is a very useful format. To first order, seismic arrivals at a given source-receiver distance will arrive at the same time (assuming identical source depths).

Because the earthquake magnitudes and source-receiver distances were all different, the amplitude of the seismograms varied enormously. In addition, some of the seismic arrivals were much higher amplitude than others. To handle this, we decided to apply a form of *automatic gain control* (AGC) to the seismograms to normalize their amplitudes. This has the great advantage that it removes the need for any other kind of amplitude correction (e.g., for distance, source size, receiver gain, etc.). The penalty is that we throw away any absolute amplitude information. We retain only timing and some relative amplitude information.

The AGC filter that we applied is very simple and based on the ratio of the average squared amplitude (i.e., power) in a short-term window to the squared amplitude in a longer term window. These are termed the short-term average and long-term average (STA and LTA). Specifically, we compute

$$Y_i = \frac{\sum_{j=0, m-1} X_{i-j}^2 / m}{\sum_{k=0, n-1} X_{i-m-k}^2 / n} \quad (1.1)$$

where X is the original time series, Y is the STA/LTA-filtered time series, m is the number of points for the short-term average and n is the number of points for the

long-term average. For optimal performance, it is important that the seismogram be demeaned (or high-pass filtered) before application of the filter so that it has a mean value of zero.

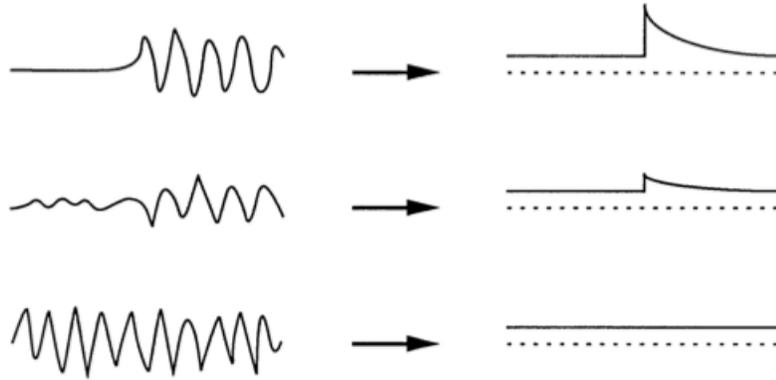


Figure 1.6: An STA/LTA filter turns the seismograms on the left into the records on the right. A value of one (STA = LTA) is shown as the dotted line.

Figure 1.6 is a cartoon that shows the effect of this filter on some example seismograms. Notice that it provides a measure of the local signal-to-noise ratio of the phase arrivals. The performance of the STA/LTA filter can be tuned by varying the length of the windows. The short-period image of Fig. 1.2 was made using an STA of 1 s and an LTA of 9 s. The long-period image of Fig. 1.3 was made using an STA of 3 s and an LTA of 30 s. An example of the operation of this filter on of the contributing seismograms to Figure 1.3 is shown in Figure 1.7.

Because they discard absolute amplitude information, only rarely can STA/LTA images of this type be used directly to learn new things about the Earth, at least in global seismology¹. However, they are extremely useful as a guide to the visibility and signal-to-noise of various seismic arrivals that may be present in a data set, thus providing guidance as to the most fruitful areas of future investigation. STA/LTA filters are also the underlying principle behind many automatic phase picking (timing) algorithms (e.g., Earle and Shearer, 1994).

I started my career largely working on *structural seismology*, which involves using seismic waves to learn about Earth structure, i.e., to obtain seismic velocity models.

¹The situation is very different in reflection seismic imaging of the crust where AGC-filtering is widely used. Here interpreters generally care much more about the timing of reflectors than their amplitude.

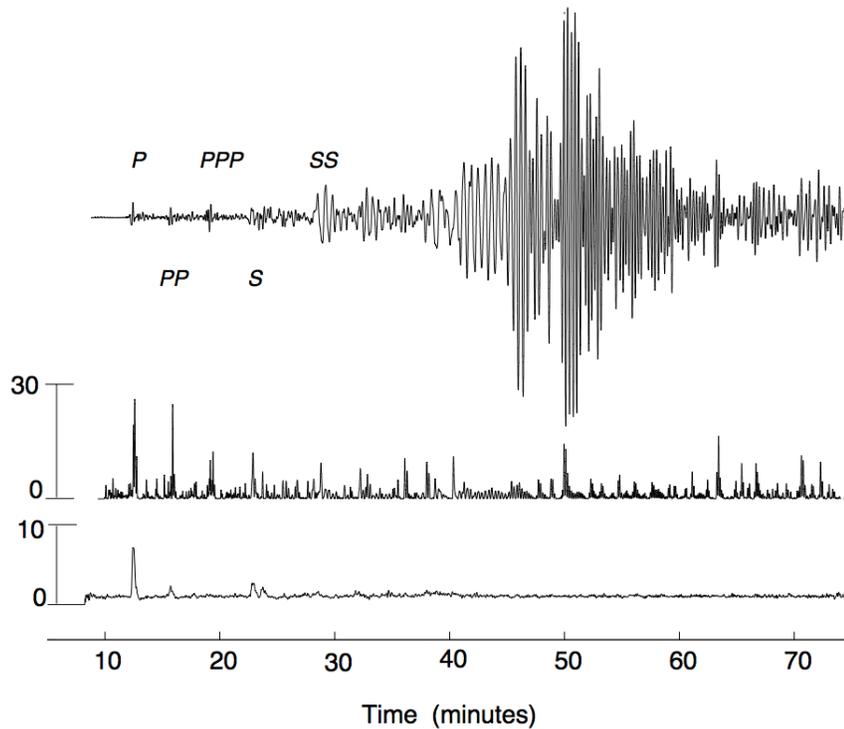


Figure 1.7: The top trace is a vertical-component seismogram (low-pass filtered at 10 s) of a M_W 6.5 earthquake on 26 April 1992, recorded at station OBN, located 83.5° away. The middle trace is the computed STA/LTA ratio function for this seismogram. The lower trace shows the average STA/LTA function resulting from all traces within a 0.5° distance bin. From Astiz et al., 1996.

This involves understanding the mechanics of seismic wave propagation. The basic physics for this is simply $F = ma$, generalized to a continuum, and is more than 100 years old. We are fortunate in seismology that the deformations involved are so small that linear strain theory is adequate for this problem. This makes our life much simpler and means that contributions from different wave types can be summed (the superposition principle). However, solving for the seismic response is difficult when Earth structure becomes complicated. Exact numerical solutions are now possible but are computationally intensive. These solutions are valuable because they are exact. However, the use of approximations (e.g., ray theory) remains useful because they provide insight regarding the origin of the different seismic phases. Given a specified source and Earth model, solving for the ground motion at a particular site involves computing *synthetic seismograms* and is called the *forward problem*. Much

more difficult is the *inverse problem* of determining source properties and Earth structure from the available seismograms, in which case questions of model resolution and uniqueness become very important. Nonetheless, we now have very accurate models of Earth's seismic velocity structure, which are continually improved as more data become available and new analysis techniques are developed.

More recently I have become more involved in studying earthquakes and their properties. There are many ways to study faults and earthquakes, including geological mapping and geodetic analyses. However, our primary source of information about earthquakes is obtained from the seismic waves they generate and thus the field of earthquake seismology cannot entirely be separated from structural seismology. But earthquake properties are known much more poorly than general Earth structure and earthquakes are intrinsically a “messier” problem than simply solving for wave propagation effects. This is because we don't know all the physics, fluids and chemistry that could be important. Many of the most fundamental aspects about earthquakes are unknown, including how they are triggered and how their ruptures propagate and come to a stop.

Before moving onto specific subjects, let's look again at the seismogram of Fig. 1.1 and the waveform images of Figs. 1.2 and 1.3. The main seismic phases are visible and clearly have a lot of information about Earth structure. But what about the other parts of the image and its contributing seismograms? The part of the seismograms before the initial P arrival is generally termed “noise” because it represents background Earth motions not caused by the earthquake. Seismic noise is mostly generated by wind, either directly on Earth's surface or through the waves that wind generates in the ocean. Until quite recently it was thought that this noise was only useful in learning about wind, waves, and storms. But it has now been shown that seismic noise travels as seismic waves and can be analyzed to learn about Earth structure. This is an amazing development and is revolutionizing some parts of seismology. Remarkably it provides a way that we may sometime learn about the internal structure of planets that lack quakes (Mars, Venus?).

What about the parts of the seismic wavefield between the main arrivals? Average amplitudes here are higher than the pre-event noise, but the arrivals are not

organized into coherent arrivals. This is termed *signal generated noise* and results from scattering off small-scale features in the Earth. An example is the coda that typically follows seismic arrivals, particularly at high frequencies. We have no hope of modeling this energy *deterministically* by creating an exact Earth model that predicts every wiggle in the seismograms. However, we can create statistical models of random heterogeneity that predict the average behavior of the scattered waves (i.e., their power as a function of space and time).

1.1 References

- Astiz, L., Earle, P., and Shearer, P. (1996). Global stacking of broadband seismograms, *Seismol. Res. Lett.*, **67**, 8–18.
- Earle, P.S. and P.M. Shearer (1994). Characterization of global seismograms using an automatic picking algorithm, *Bull. Seismol. Soc. Am.*, **84**, 366–376.

Chapter 2

Discontinuity phases

The upper-mantle discontinuities provide important constraints on models of mantle composition and dynamics. The most established mantle seismic discontinuities occur at mean depths near 410, 520, and 660 km and will be the focus of this chapter. For want of better names, they are identified by these depths, which are subject to some uncertainty (the older literature sometimes will refer to the 400, 420, 650, 670, etc.). Furthermore it is now known that these features are not of constant depth, but have topography of tens of kilometers. This leads to oxymoronic statements like “The 660-km discontinuity is at 680 km depth in the northwest Pacific.” 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.

The mantle is mainly composed of olivine (Mg_2SiO_4), which undergoes phase changes near 410, 520 and 660 km (see Fig. 2.1). The sharpness of the seismic discontinuities is related to how rapidly the phase changes occur with depth. Generally the *660* is thought to be sharper than the *410*. The *520* is likely even more gradual, so much that it does not yet appear in most standard 1-D Earth models. Before discussing observations of these features, I will review some of the aspects of ray theory that will help in the analysis.

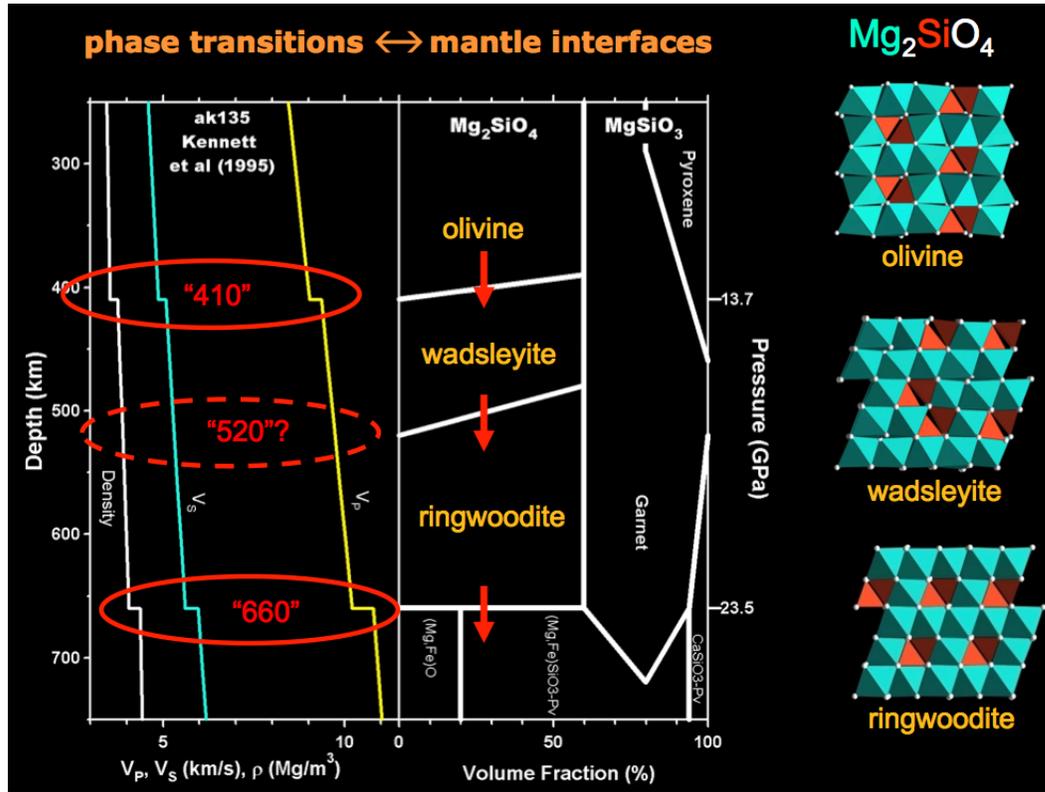


Figure 2.1: Seismic velocity and density (left) compared to mantle composition (right). Figure courtesy of Rob van der Hilst.

2.1 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. 1.3). 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:

$$p = u(z) \sin \theta = \frac{dT}{dX} = u_{\text{tp}} = \text{constant for given ray}, \quad (2.1)$$

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 2.2. 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. 2.3). When $dX/dp > 0$ the travel time curve is termed *retrograde*. The transition from prograde to retrograde and back to prograde generates a *triplcation* in the travel time curve.

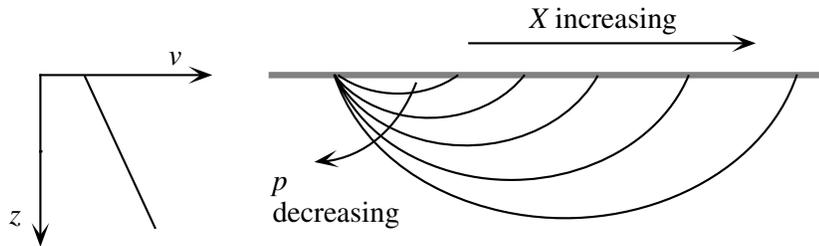


Figure 2.2: A gentle velocity increase with depth causes rays to travel further when they leave the source at steeper angles.

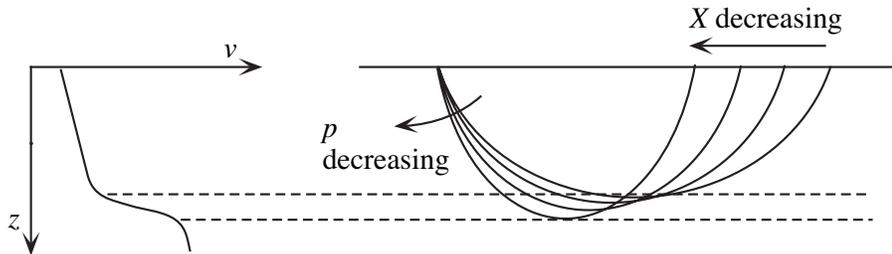


Figure 2.3: A steep velocity increase with depth causes steeper rays to fold back on themselves toward the source.

Triplcations are very diagnostic of the presence of a steep velocity increase or discontinuity. The 410- and 660-km discontinuities cause a double triplcation near 20 degrees (Fig. 2.4 and 2.5), which can be seen in both P waves and S waves. This is how these discontinuities were first discovered in the 1960s. Older studies of the triplcations analyzed the timing (and sometimes the slopes, if array data were available) of the different branches of the travel-time curves. However, because the 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

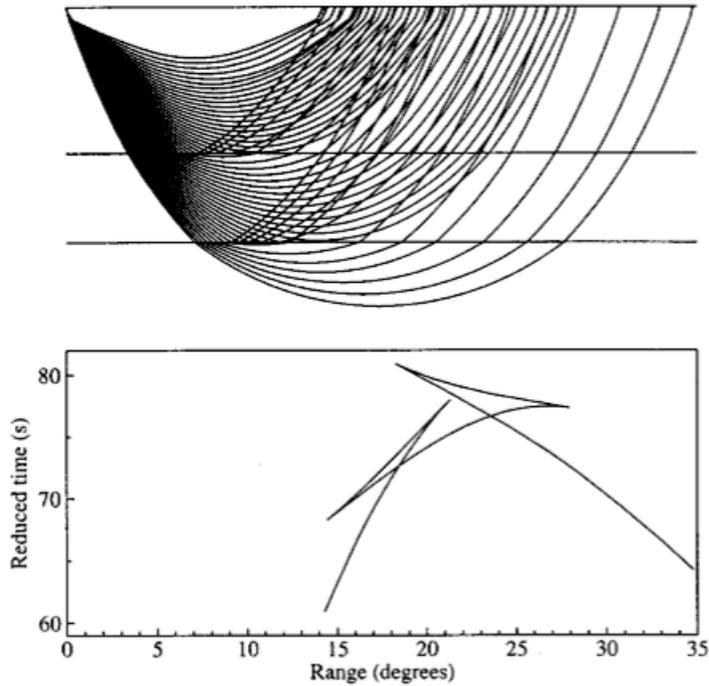


Figure 2.4: The seismic velocity increases near 410 and 660 km depth create a double triplication in the P -wave travel time curve near 20° epicentral distance, as predicted by the IASP91 velocity model (Kennet, 1991). A reduction velocity of 10 km/s is used for the lower plot. From Shearer (2000).

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 immediately 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

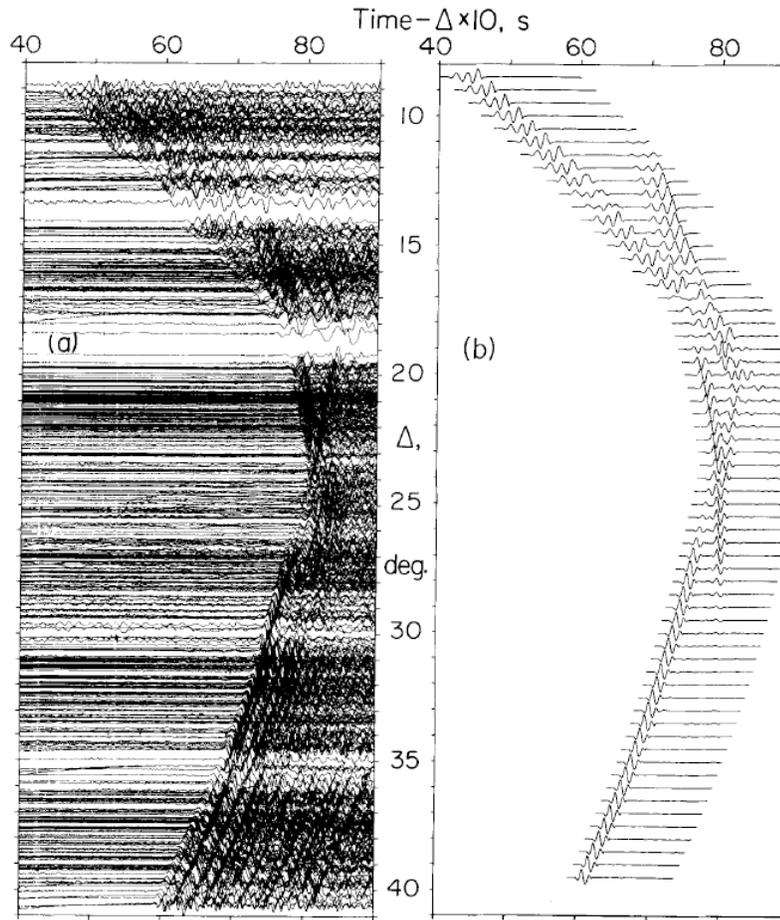


Figure 2.5: Record section of P waves from Mexican earthquakes recorded by southern California seismic stations (left) compared to synthetic seismograms. From Walck (1984).

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.

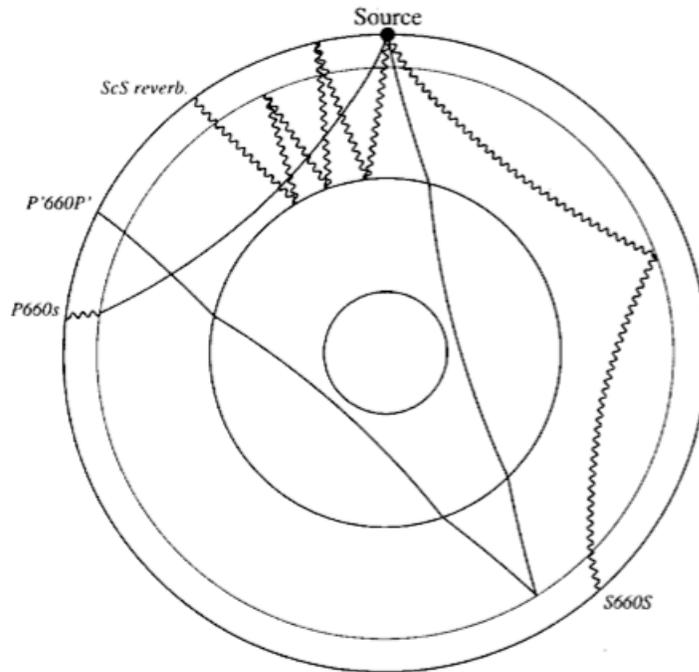


Figure 2.6: Example ray paths for discontinuity phases resulting from reflections or phase conversions at the 660-km discontinuity. P waves are shown as smooth lines, S waves as wiggly lines. The ScS reverberations include a large number of top- and bottom-side reflections, only a single example of which is plotted. From Shearer (2000).

2.2 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 2.6. 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 discontinuity properties than directly transmitted waves. For example, consider the reflected and transmitted waves for an S -wave incident on a discontinuity (Fig. 2.7).

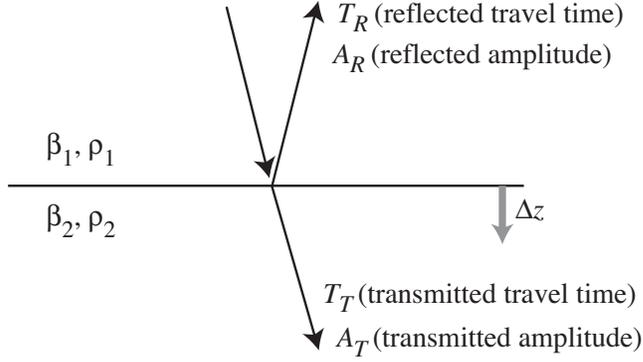


Figure 2.7: Ray geometry for near-vertical S -wave reflection and transmission.

For near-vertical incidence, the travel time perturbation for the reflected phase is approximately

$$\Delta T_R = \frac{2\Delta z}{\beta_1} \quad (2.2)$$

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

$$\begin{aligned} \Delta T_T &= \frac{\Delta z}{\beta_1} - \frac{\Delta z}{\beta_2} = \Delta z \left(\frac{1}{\beta_1} - \frac{1}{\beta_2} \right) = \Delta z \left(\frac{\beta_2 - \beta_1}{\beta_1 \beta_2} \right) = \frac{\Delta z}{\beta_1} \frac{\beta_2 - \beta_1}{\beta_2} \\ &= \frac{1}{2} \frac{\beta_2 - \beta_1}{\beta_2} \Delta T_R \end{aligned} \quad (2.3)$$

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

$$A_R = \dot{S}\dot{S}_{\text{vert}} = \frac{\rho_1 \beta_1 - \rho_2 \beta_2}{\rho_1 \beta_1 + \rho_2 \beta_2}, \quad (2.4)$$

$$A_T = \dot{S}\dot{S}_{\text{vert}} = \frac{2\rho_1 \beta_1}{\rho_1 \beta_1 + \rho_2 \beta_2}. \quad (2.5)$$

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, $\hat{S}\hat{S} = -0.05$ (assuming $\beta_1\rho_1 < \beta_2\rho_2$) and $\hat{S}\hat{S} = 0.95$. 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.

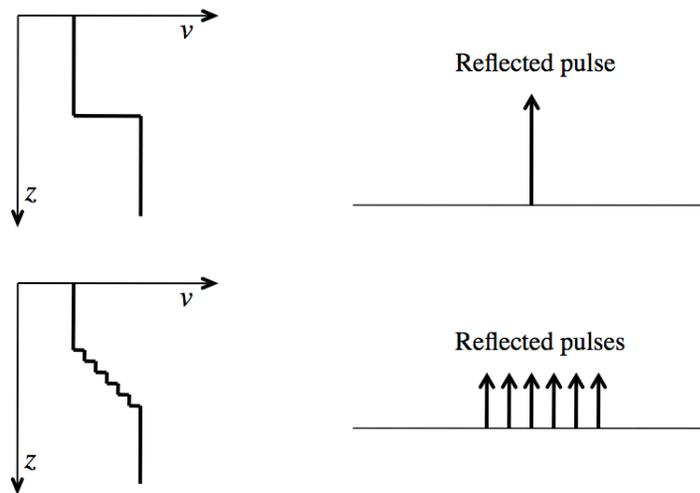


Figure 2.8: A step velocity discontinuity produces a delta-function reflected pulse. A series of velocity jumps produces a series of delta-function reflections.

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. 2.8, top). In contrast, a velocity gradient will produce a box car reflection. To see this, first consider a staircase velocity depth function (Fig. 2.8, bottom). Each small velocity jump will produce a delta function reflection.

In the limit of small step size, the staircase model becomes a continuous velocity

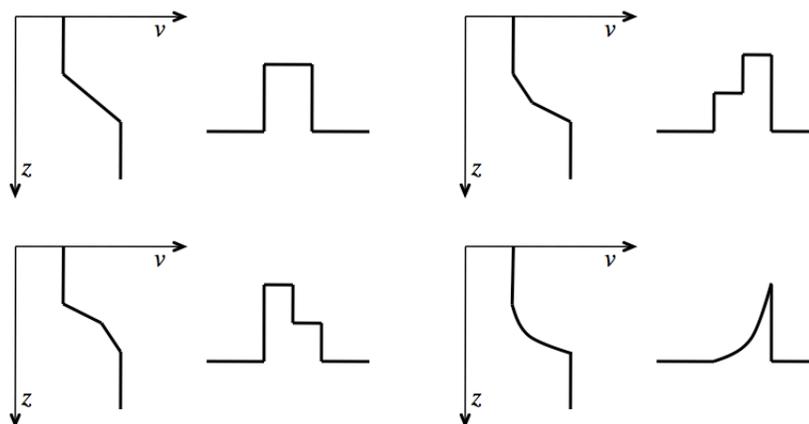


Figure 2.9: Different velocity-depth profiles and their top-side reflected pulses.

gradient, and the series of delta functions become a boxcar function (top left of Fig. 2.9). 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. 2.9). 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 7-km-thick gradient region immediately above a sharp discontinuity. Figure 2.10 shows some of their results, which indicate that the 660 is sharper (less than 2 km thick transition) than the 410 because it can be observed to higher frequency. The 520-km discontinuity is not seen in their results at all, indicating that any P impedance jump must occur over 20 km or more.

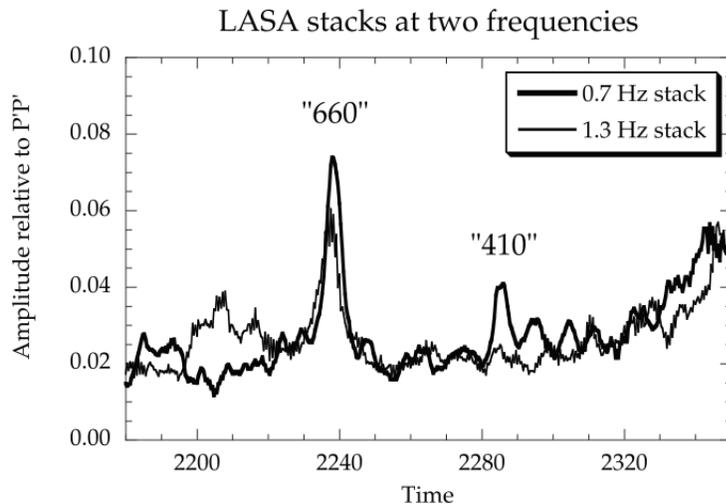


Figure 2.10: LASA envelope stacks of nine events at two different frequencies.

2.3 Reference phase stacking

I started making images like those in Figs. 1.2 and 1.3 about 20 years ago, when global seismic data first became conveniently available in bulk form (on CD-ROMs). They showed a lot of seismic phases, but all of them were previously known. Also, they did not show very clearly some well-known seismic phases, like the core reflections PcP and ScS . The AGC stacking method has trouble resolving phases that closely follow a stronger phase (because the LTA still has a large value) and also has limited time resolution because it discards polarity information. In an attempt to do better, I began experimenting with stacking the data using a reference phase. This is illustrated in Fig. 2.11 for long-period S waves recorded near 74 degrees. On the left the seismograms are aligned according to their predicted arrivals times based on a standard 1-D Earth model. Because of source radiation differences, instrument response variations, and depth phase interference, the seismograms do not line up. Summing and averaging the seismograms would not work very well because much of the energy would cancel out.

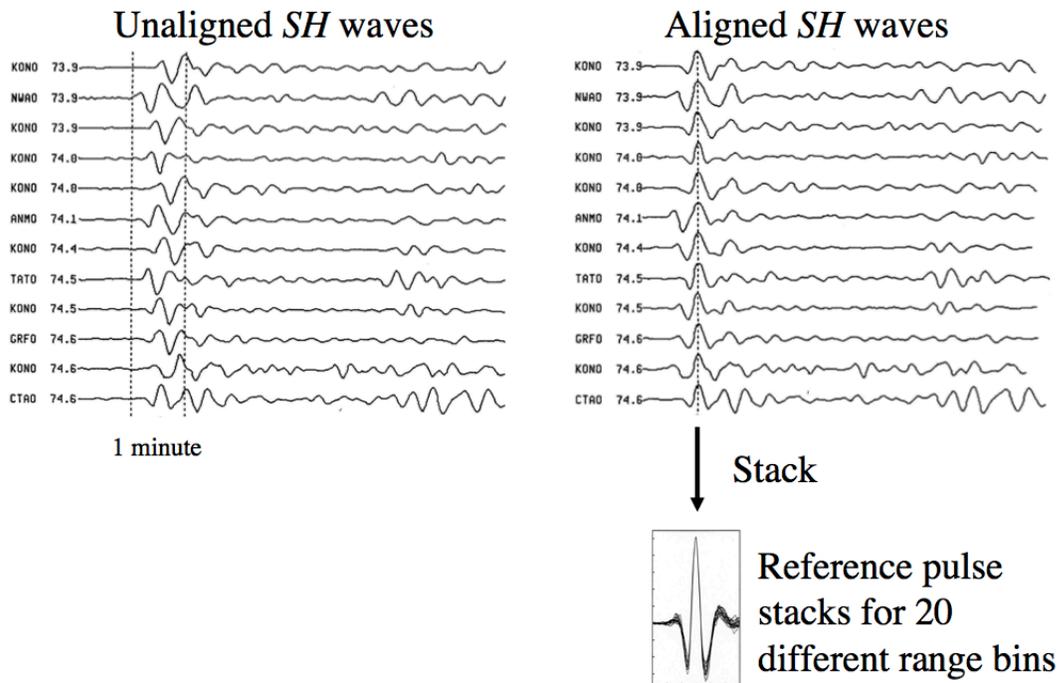


Figure 2.11: The reference phase stacking method aligns seismograms on the peak amplitude of the reference phase, thus ensuring a coherent stack.

To produce a coherent stack, I shifted the seismograms so that all their peak amplitudes lined up, flipping the polarity of some of the traces, and normalized them all to unit amplitude. This is shown on the right. Summing the records now produces a coherent stack. In fact, the stacked waveforms is quite repeatable among the different range bins if enough data are stacked. The reference phase stacking approach removes most of the information in the reference phase itself. Its great advantage, however, is that other phases that are shifted in time by a constant amount with respect to the reference phase will now also stack coherently. This can increase the signal-to-noise ratio of weaker phases so as to make them visible. In particular, upper-mantle discontinuity phases are now readily apparent in global waveform stacks (see Fig. 2.12). The phases visible in this figure roughly parallel to the P and S reference phases, but delayed by 1.5 to 4 minutes, are topside reverberations off the top of the discontinuities (peg-leg multiples). The 410, 520 and 660 km reflections can be seen. Note that there is no clear reflection from the often hypothesized 220-km discontinuity (e.g., in the PREM model).

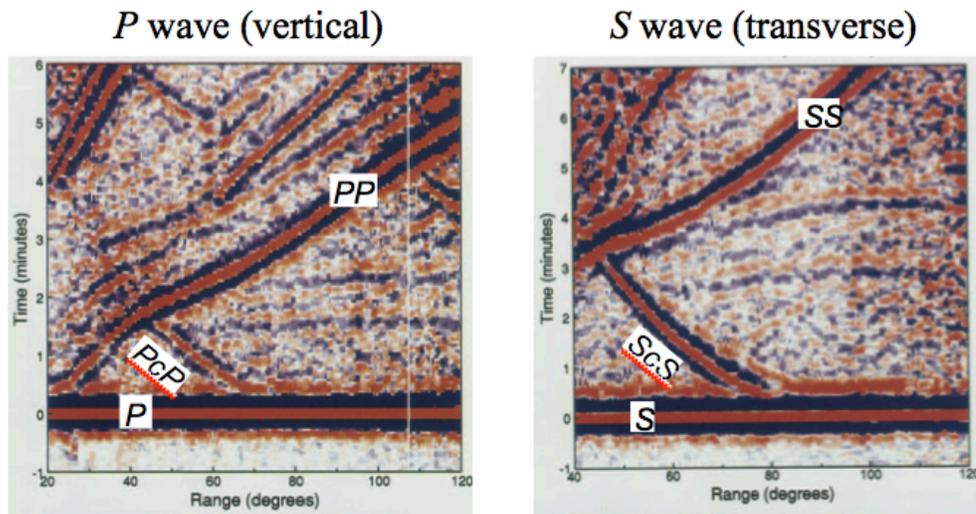


Figure 2.12: Stacked images of long-period GSN data from shallow sources (< 50 km) obtained using P as a reference (left) and S as a reference (right). From Shearer (1991).

2.4 *SS* precursors

The topside multiples visible in Fig. 2.12 are of limited use because they arise both from near-source and near-receiver structure. Better phases for global analysis of upper-mantle discontinuity properties are *SS* precursors. These are most clearly seen in global images when *SS* is used as a reference phase (see Fig. 2.13).

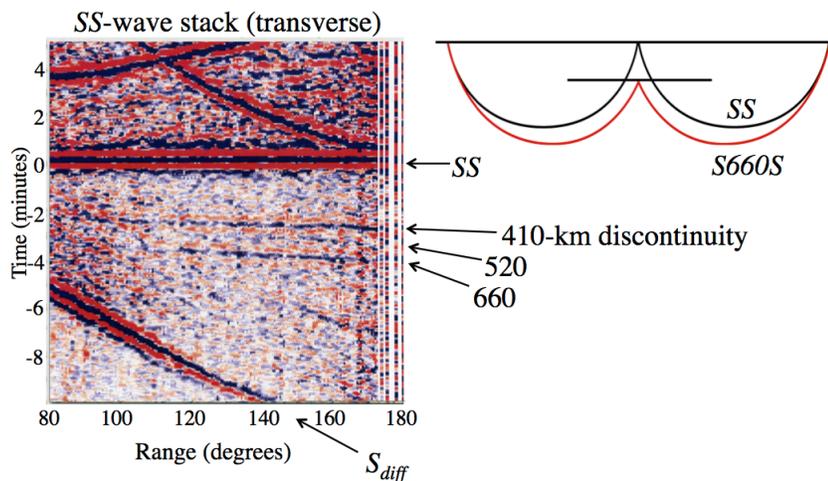


Figure 2.13: Stacked images of long-period GSN data from shallow sources (< 50 km) obtained using *SS* as a reference. From Shearer (1991).

SS precursors are especially valuable for global mapping because of their widely distributed bouncepoints, which lie nearly halfway between the sources and receivers (see Fig. 2.14). The timing of these phases is sensitive to the discontinuity depths below the *SS* surface bouncepoints. *SS* precursors cannot be reliably identified on individual seismograms, but the data can be grouped by bouncepoint to perform stacks that can identify regional variations in discontinuity tomography. The measured arrival times of the 410- and 660-km phases (termed *S410S* and *S660S*, respectively) can then be converted to depth by assuming a velocity model. The topography of the 660-km discontinuity measured in this way is shown in Figure 2.15. Only long horizontal wavelengths can be resolved because of the broad sensitivity of the long-period *SS* waves to structure near the bouncepoint. The most prominent features in this map are the troughs in the *660* that appear to be associated with subduction zones.

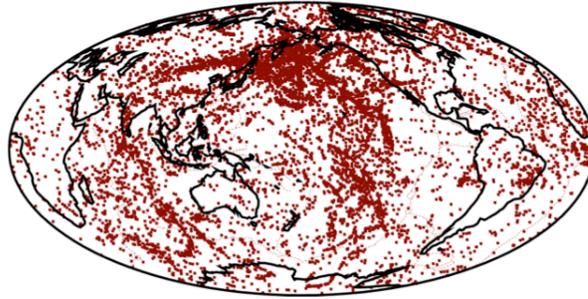


Figure 2.14: The global distribution of *SS* bouncepoints. From Flanagan and Shearer (1998).

High-pressure mineral physics experiments have shown that the olivine phase change near 660 km has what is termed a negative *Clapeyron slope*, which defines the expected change in pressure with respect to temperature. A negative slope means that an increase in depth (pressure) should occur if there is a decrease in temperature. Thus these results are consistent with the expected colder temperatures in subduction zones. It appears that in many cases slabs are deflected horizontally by the 660-km phase boundary (students should think about why the *660* tends to resist vertical flow) and pool into the transition zone. Tomography results show this for many of the subduction zones in the northwest Pacific (see Fig. 2.16). However, in other cases tomography results show that the slabs penetrate through the *660*. This would cause a narrow trough in the *660* that would be difficult to resolve with the *SS* precursors.

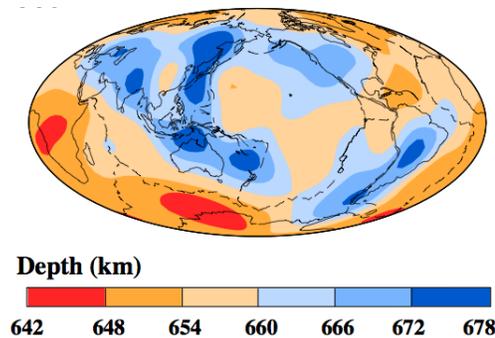


Figure 2.15: Long wavelength topography on the 660-km discontinuity as measured using *SS* precursors. From Flanagan and Shearer (1998).

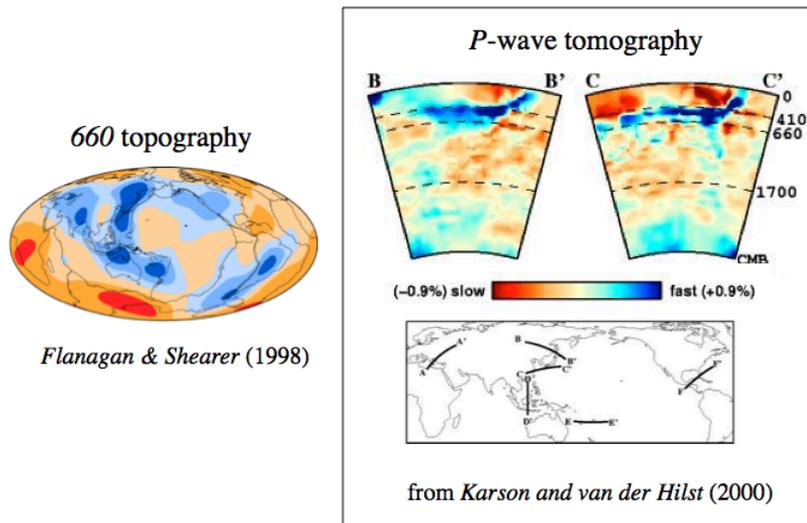


Figure 2.16: A comparison between long wavelength topography on the 660-km discontinuity as measured using SS precursors and transition zone velocity structure from seismic tomography.

2.5 Stacking and the importance of \sqrt{n}

Reflections and phase conversions from mantle discontinuities typically have much smaller amplitudes than direct seismic phases and are often invisible on individual seismograms. Thus, stacking methods are often employed to enhance their visibility by combining data from multiple records. Stacking is simply a sum or average of the seismograms. The strategy of stacking methods is to design them in such a way that noise will stack incoherently and cancel out, while the desired signal will stack coherently, thus increasing the signal-to-noise ratio of the stack compared to the individual records.

In seismology, the distinction between signal and noise is not always clearcut. We typically like to think of noise as a random process with no information content. When we analyze records from an earthquake or explosion, one measure of the noise is to examine records before the arrival of the first phase arrival (normally the P wave). But most of this “noise” in seismic records is Earth noise (i.e., microseisms) and is far from random. In the last few years, we have seen scores of new studies showing how analysis of microseism noise can constrain Earth structure. Within an earthquake record, we also often consider the energy between the main phases to

be noise in some sense, even though most of it is signal generated (e.g., scattered energy). This “noise” typically does not form coherent arrivals and is hard to model deterministically (i.e., wiggle for wiggle) using synthetic seismogram methods. An obvious example is the high-frequency coda that follows direct P and S arrivals. But even at long periods, where the main body wave phases appear as discrete waveforms, the energy level between phases is above the background pre-event level, implying that some scattering is occurring.

Whatever its origin, we typically assume that the noise in our seismograms is random and uncorrelated among the different records in the stack. In this case, we can apply standard statistical methods to estimate how noise will behave when it is stacked. Given a random variable X , the expected value, $E(X)$, is simply the mean, m , which may be expressed as

$$E(X) = m = \frac{1}{n} \sum_{i=1,n} X_i \quad (2.6)$$

where n is the number of values of X . The variance of X , $\text{var}(X)$, is often written as σ^2 and is defined¹ as

$$\text{var}(X) = \sigma^2 = E\left([X - E(X)]^2\right) = \frac{1}{n} \sum_{i=1,n} [X_i - E(X)]^2 \quad (2.7)$$

Note the following properties for two random variables X and Y :

$$E(X + Y) = E(X) + E(Y) \quad (2.8)$$

$$E(aX + b) = aE(X) + b \quad (2.9)$$

$$\text{Var}(aX + b) = a^2 \text{Var}(X) \quad (2.10)$$

If we assume that X and Y are uncorrelated we also have

$$E(X \cdot Y) = E(X) \cdot E(Y) \quad (2.11)$$

$$\text{Var}(X + Y) = \text{Var}(X) + \text{Var}(Y) \quad (2.12)$$

Now consider two random time series composed of the points contained in X and Y (see Fig. 2.17). If X and Y have the same variance σ^2 , then their sum will

¹Often the variance is defined with $1/(n-1)$ rather than $1/n$ to account for the reduced number of degrees of freedom once the mean is computed. Typically n is large enough that this distinction makes little difference.

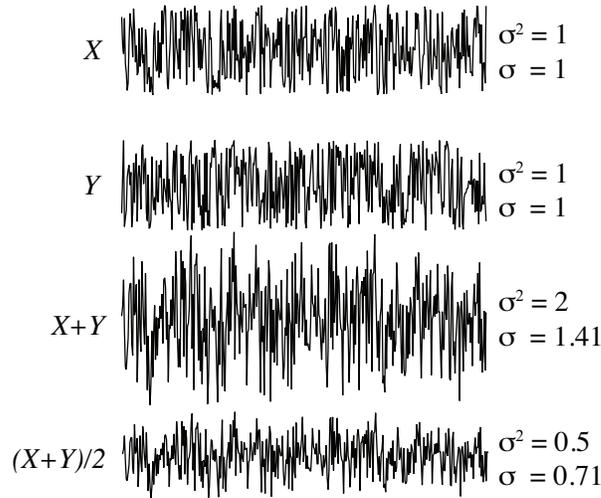


Figure 2.17: Example of stacking two random time series.

have variance $\text{Var}(X + Y) = \sigma^2 + \sigma^2 = 2\sigma^2$ and their average will have variance $\text{Var}[\frac{1}{2}(X + Y)] = (1/4)2\sigma^2 = \sigma^2/2$ (from eq. 2.10). The amplitude of the time series scales with the standard deviation, $\sigma = \sqrt{\sigma^2}$, and thus the amplitude of the average of the two traces is $1/\sqrt{2}$ times the amplitude of X and Y .

Note that the random values do not have to be normally distributed (i.e., Gaussian) for these relationships to hold. For example, the times series plotted in Fig. 2.17 have uniformly distributed random values. As an extreme example, consider two random time series consisting entirely of -1 and 1 (with equal probability). We can construct a table to determine the average outcome of their sum and average:

X	Y	$X + Y$	$\frac{1}{2}(X + Y)$
1	1	2	1
1	-1	0	0
-1	1	0	0
-1	-1	-2	-1

Note that in this case $\text{Var}(X) = \text{Var}(Y) = 1$ and $\text{Var}[\frac{1}{2}(X + Y)] = 1/2$, as predicted from (2.12).

Now let us generalize this result to stacking n time series:

$$\text{Var} \left[\frac{X_1 + X_2 + \dots + X_n}{n} \right] = \frac{1}{n^2} [\text{Var}(X_1) + \text{Var}(X_2) + \dots + \text{Var}(X_n)] \quad (2.13)$$

In the case where each time series has the same variance, σ^2 , then

$$\text{Var} \left[\frac{X_1 + X_2 + \dots + X_n}{n} \right] = \frac{1}{n^2} [n\sigma^2] = \frac{1}{n} \sigma^2 \quad (2.14)$$

The amplitude of the stack is proportional to the standard deviation (SD):

$$\text{SD} \left[\frac{X_1 + X_2 + \dots + X_n}{n} \right] = \sqrt{\frac{\sigma^2}{n}} = \frac{\sigma}{\sqrt{n}} \quad (2.15)$$

where σ is the standard deviation of the original time series. Notice that we expect the amplitude of the stack to decrease as $1/\sqrt{n}$.

We can get some physical insight about what is going on by considering the power in these time series. The power is proportional to the square of the amplitude (also true for seismic waves and seismograms!) and thus the variance of the records. Thus all of the above relationships can be thought of as simply representing the conservation of energy. When we stack n seismograms, we expect the power of the resulting sum to be n times larger. If we normalize the sum to represent the average of the traces, we divide by n and reduce the power by n^2 . Thus the power of the normalized stack decreases as $1/n$ and its amplitude as $1/\sqrt{n}$.

Now consider what happens if the traces contain a sum of random noise, R , and some coherent signal, S , that is identical for every trace.

$$\begin{aligned} X_1 &= R_1 + S \\ X_2 &= R_2 + S \\ &\cdot \\ X_n &= R_n + S \end{aligned}$$

The normalized stack (average trace) is

$$\text{avg}(X) = \frac{R_1 + R_2 + \dots + R_n}{n} + S \quad (2.16)$$

The amplitude of the signal is unchanged, but the amplitude of the noise will decrease as $1/\sqrt{n}$. Thus the signal-to-noise (STN) ratio will increase as \sqrt{n} . Thus if we have an initial STN of 2, we can improve this to 20 by stacking 100 records. Note that the initial STN does not have to exceed unity for this to work. By stacking

enough traces we can resolve coherent signals that are too small to be visible on any single trace. The key for this to work is that the noise has to be incoherent so that it tends to cancel out when it is stacked. It does not, of course, cancel out perfectly and it's important to remember that the amplitude and power in a simple sum will increase with n —it only goes down when we normalize the stack to the average trace value by dividing by n .

This \sqrt{n} behavior occurs in a wide range of situations:

1. In statistics, standard error estimates vary as $1/\sqrt{n}$ where n is the sample size.
2. In probability, if one tosses a coin n times, the expected difference between the number of heads and tails scales as \sqrt{n} .
3. The expected length of a random walk scales as \sqrt{n} where n is the number of walk segments.
4. The expected amplitude of an undamped harmonic oscillator subject to impulses at random times grows as \sqrt{n} (**reference?)

2.6 Using the bootstrap to estimate errors

Measurements without some estimate of their likely errors are scientifically questionable. Don't make the mistake of submitting your first paper without error bars on your data points. You are likely to get critical reviews. Of course there are many sources of possible errors. The easiest to model are random measurement errors. In this case, we typically assume that the errors are uncorrelated among the measurements (this is the best possible case and usually leads to the smallest error bars). You have probably had a course in statistics that treats this subject in some detail.

For example, if we have n samples of a random Gaussian variable X , we can easily compute their mean value, X_M . But this will not be the exact true mean of the underlying random distribution. One can show that the *standard error* of the mean is given by the computed standard deviation of X (the sample standard deviation), divided by the square root of the number of points:

$$SE(X_M) = \frac{SD(X)}{\sqrt{n}} \quad (2.17)$$

Note that \sqrt{n} appears again! This equation means that there is a 66% chance that the true mean is within $\pm SE(X_M)$ of X_M , and there is a 96% chance the true mean is within $\pm 2SE(X_M)$ of X_M . When people show error bars on their plots, they will often be asked whether they are one- or two-standard error bars (when they have large estimated errors, they tend to plot one-standard error bars...).

However, often the parameters for which we want to estimate likely uncertainties are the result of a complicated data-processing method. It is not always clear how to implement standard statistical theories in this case. A very powerful computer-based method to estimate standard errors is called bootstrap resampling. It is very simple to perform and should be in every geophysicists's toolbox. Let us suppose that we have n data of some kind. These could be travel time measurements, amplitude measurements, or even actual individual seismograms. From these data, we employ some processing scheme and compute a number. For example, this could be an estimated depth to a mantle discontinuity following a stack of n seismograms. We are interested in estimating a standard error for this number. This error should take into account that the data we use are not completely self-consistent. In other words, if we processed different subsets of the data we would obtain different results. The data that we actually have are samples from some random process. Because we only have a finite amount of data, we cannot completely sample the underlying random distribution.

The bootstrap method works by testing the effects of different random samples of the available data. My favorite implementation of the bootstrap will select n random samples from the n data for each of these tests. For example, if we have 10 data points and we randomly select 10, we might pick: 3, 9, 7, 1, 1, 3, 1, 2, 2, 8. Note that some points are sampled more than once (1 and 2) while others are not sampled at all (4, 5, 6). We then go ahead and process these 10 data and compute our desired parameter. This is the outcome of one bootstrap resample. Using the computer, we then repeat this process m times. To obtain reliable results m should be at least 100. This gives us m estimates of our computed parameter. One can show that the *standard deviation* of these 100 numbers provides an estimate of the *standard error* for their mean value. Note that we do not need to normalize by n .

The standard deviation of these m estimates will naturally go down as the number of original data points goes up.

It is possible that the m estimates will not be normally distributed. In this case, it is best to explicitly pick out the desired confidence limits from the distribution. This is common practice, although I recently noted that the Rice textbook on statistics says that this has not been rigorously shown to be correct. For all practical purposes in geophysics, however, it is likely to be close enough.

A note of caution: Formal statistical measures of uncertainty such as standard error only provide a measure of the likely errors caused by finite data. They assume uncorrelated errors in the data and do not account for any systematic sources of bias (e.g., application of an incorrect velocity model). Thus they provide an absolute minimum estimate of the error. The true error is like to be larger, in many cases much larger.

2.7 Additional reading

Flanagan, M. P. and P. M. Shearer (1998). Global mapping of topography on transition zone velocity discontinuities by stacking SS precursors, *J. Geophys. Res.*, **103**, 2773–2692.

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.

2.8 Exercises

1. A file, traces200, can be found at <http://igppweb.ucsd.edu/~shearer/CIT/traces200>. It contains 200 seismograms from epicentral distances between 120 and 121 degrees and from sources shallower than 75 km. The traces have been lowpass filtered at 0.1 Hz and resampled to 1 sample/second. The file looks like:

```

          1          479
-379  0.6700E-02
-378  0.4190E-01
-377  0.5340E-01
.
.
98  0.3200E-02
```

```

99  0.3200E-02
      2          479
-379 0.3500E-01
-378 0.4060E-01
-377 0.4660E-01

```

where 1=trace number, 479=number of points to follow, -379 =time (s) relative to SS peak, 0.6700E-02=amplitude of trace, etc.

The traces have been windowed on the *SS* phase and time shifted and normalized so that the peak *SS* amplitude is one at zero time.

- (a) Stack and average these traces and plot the result. Use an amplitude scale between -0.1 and 0.1 so that the precursory peaks are visible. Label the *S660* and *S410S* arrivals. Can you identify any of the other peaks?
- (b) Use a bootstrap method to create additional lines for your plot, representing the 5% and 95% confidence limits for the amplitudes in your stack. In other words, perform 100 random resamples of the traces and obtain 100 stacked traces. At each time point, sort the traces by amplitude and identify the 5th and 96th largest values.
- (c) Experiment with reducing the number of traces by selecting and stacking a random subset of the original 200 traces. How many traces seem to be required to see the 410 and 660 peaks reliably?
- (d) These traces have been selected because the amplitude of *SS* is at least 8 times larger than the maximum amplitude within the precursory window. We can think of this as a local signal-to-noise ratio for *SS*. Experiment with increasing the required signal-to-noise. Does this seem to increase or decrease the “quality” of the stack? Experiment with weighting the traces by their local *SS* signal-to-noise, i.e., weighting traces with an STN of 20 twice as much as those with an STN of 10. How does this affect the stack?

2.9 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.
- Rice, J. A. (1995). *Mathematical Statistics and Data Analysis, Second Edition*, Duxbury Press, Belmont, California.

- 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 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

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

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 i th station and compute predicted arrival times,

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

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

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

where r_i is the residual at the i th 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

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

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

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

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

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

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

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

where σ_i is the expected standard deviation of the i th 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,

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

where \mathbf{m}_{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,

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

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 unmodeled 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

²The error ellipse is only approximate because the uncertainties in the σ_i estimate are ignored.

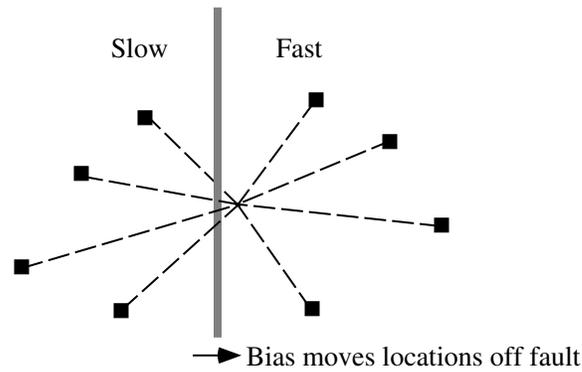


Figure 3.1: Earthquakes located along a fault will often be mislocated if the seismic velocity changes across the fault.

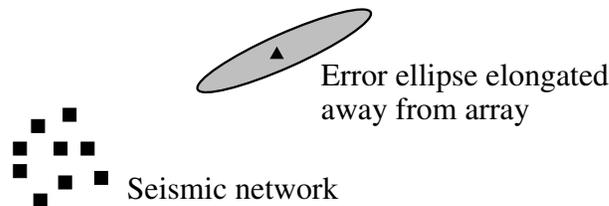


Figure 3.2: Earthquake locations for events outside of a network are often not well constrained.

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

$$\mathbf{m} = \mathbf{m}_0 + \Delta\mathbf{m}, \quad (3.10)$$

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

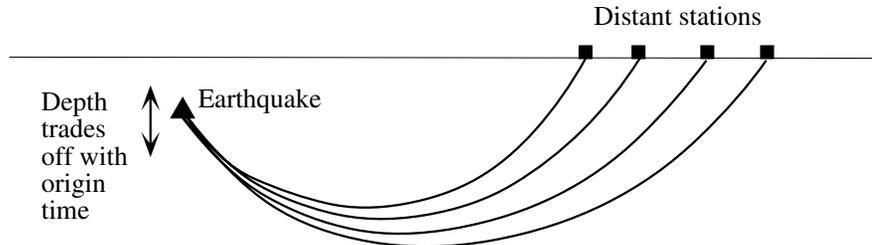


Figure 3.3: Earthquake depth can be hard to determine if only distant stations are available.

first term in the Taylor series expansion

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

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

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

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

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

or

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

where \mathbf{G} is the matrix of partial derivatives $G_{ij} = \partial t_i^p / \partial m_j$, $i = 1, 2, \dots, n$, $j = 1, \dots, 4$. 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

$$t^{\text{rel}} = t - t^{\text{master}}. \quad (3.15)$$

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

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

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,

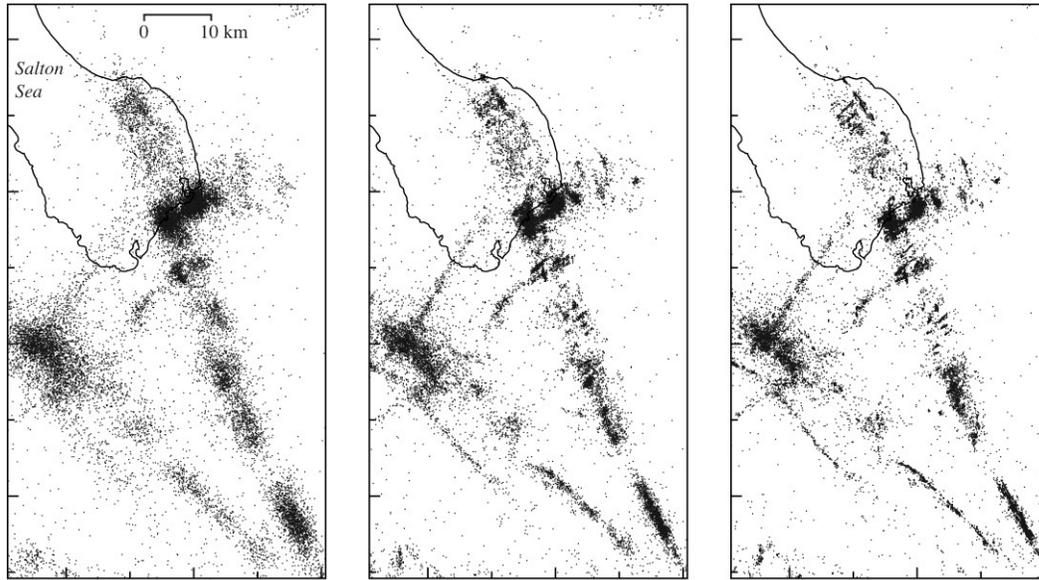


Figure 3.4: Earthquake locations for over 17,000 earthquakes in the Imperial Valley, California (1981–2005), as computed using: (left) single event location, (middle) source-specific station term (SSST) locations, and (right) waveform cross-correlation locations using results from Lin et al. (2007).

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,

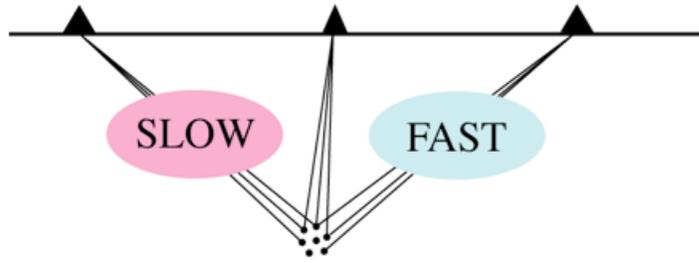


Figure 3.5: Events within a single small event cluster will have correlated travel time residuals to each station.

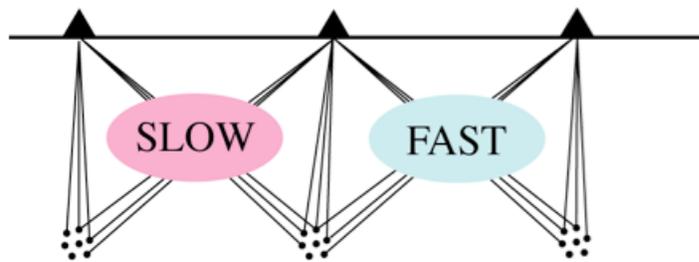


Figure 3.6: The residuals to each station will vary among different event clusters.

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.

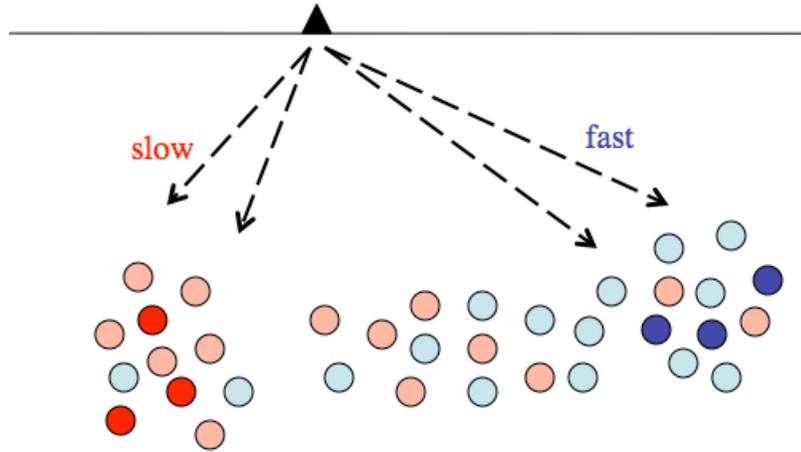


Figure 3.7: Travel time residuals to a single station will have some amount of random scatter and some systematic variation with event location. Negative (fast) residuals are shown in blue; positive (slow) residuals are shown in red. The average residual will be positive for paths through slow material 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).

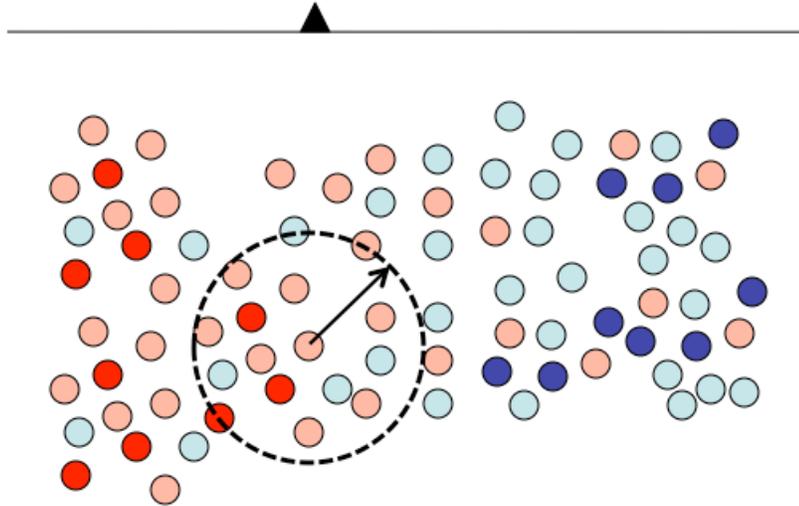


Figure 3.8: Source-specific station terms (SSSTs) are computed by smoothing the residual field over some specified smoothing radius.

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 on the class web site.
 - (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.

<i>x</i> (km)	<i>y</i> (km)	Quake 1	Quake 2
		<i>t</i> 1 (s)	<i>t</i> 2 (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.
2. (COMPUTER) Eight Caltech students go up to Mount Wilson to observe and time meteors during the Perseid Meteor Shower. They record the times of 100 meteors by using their digital watches. These times are given in a table (see class web site) that looks like this:

34.118	32.004	34.820	33.832	32.036	-99.000	32.765	29.798
-99.000	51.087	51.664	51.930	50.634	50.481	45.513	46.652
47.545	44.757	48.668	-99.000	44.439	45.447	43.019	41.277
-99.000	48.003	52.308	52.524	-99.000	-99.000	47.875	46.674
14.975	11.507	13.683	14.464	-99.000	11.054	-99.000	-99.000

Each line is for a single meteor and the columns are for the different students. Times are in seconds relative to the nearest minute of a reference time (not given). Sometimes students were not paying attention and they missed a measurement—these are indicated as -99. Unfortunately they did not synchronize their watches, so each student’s recorded times are likely systematically shifted by several seconds compared to the true times.

- (a) Estimate a time for each meteor by computing the average time of the (valid) measurements in each line. What is the standard deviation of the entire set of measurements relative to these mean times?
- (b) Using the same iterative approach as is used for station terms for the event location problem, compute time correction terms for each student, recompute the meteor times, recompute the correction terms, etc. How many iterations does it take for the problem to converge? What are the eight final correction terms that you obtain? After applying these correction terms, what is the standard deviation of the entire set of measurements relative to the times estimated for each event? Note that the correction terms are ambiguous with respect to an arbitrary constant time shift, i.e., you could add the same time to all of them and the fit to the data would be the same. To remove this ambiguity, constrain the mean time correction term to be zero.

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 COMLOC 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

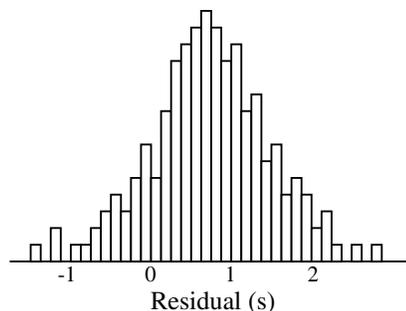
- Frohlich, C. (1979). An efficient method for joint hypocenter determination for large groups of earthquakes, *Comp. Geosci.* **5**, 387-389.
- Jordan, T. H., and K. A. Sverdrup (1981). Teleseismic location techniques and their application to earthquake clusters in the south-central Pacific, *Bull. 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



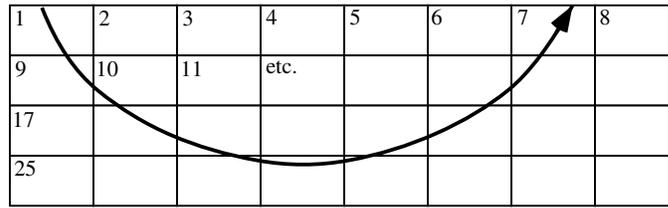


Figure 4.1: An example ray path and cell numbering scheme for a simple 2-D tomography problem.

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

$$r = \sum_k b_k v_k, \quad (4.1)$$

where b_k is the ray travel time through the k th 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 i th ray path as:

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

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:

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

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

$$\mathbf{d} = \mathbf{G}\mathbf{m} \quad (4.4)$$

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

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

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

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

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

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

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):

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

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

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

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

	m_{up}	
m_{left}	m_j	m_{right}
	m_{down}	

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

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.

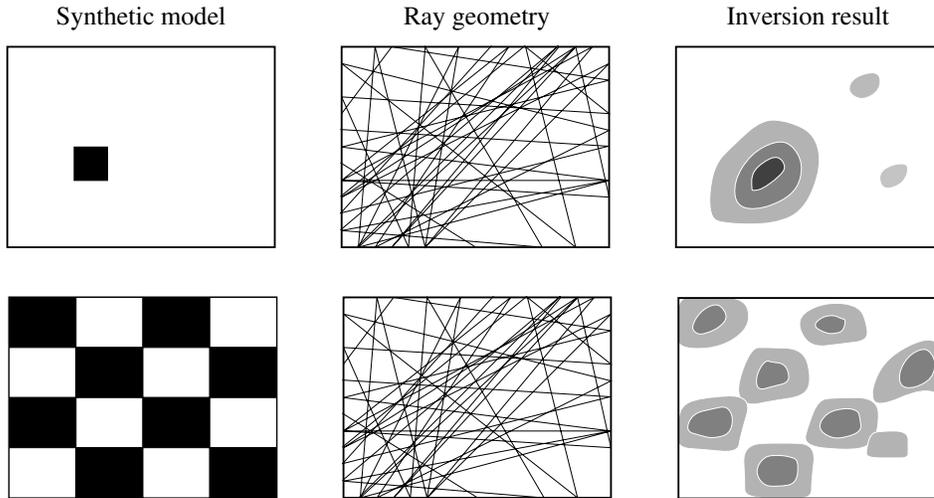


Figure 4.2: The resolution of tomographic models is often evaluated using the impulse response test (top) or the checkerboard test (bottom). In each, a synthetic set of travel times are created for a simple velocity model using the same ray paths present in the real data; then the synthetic times are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

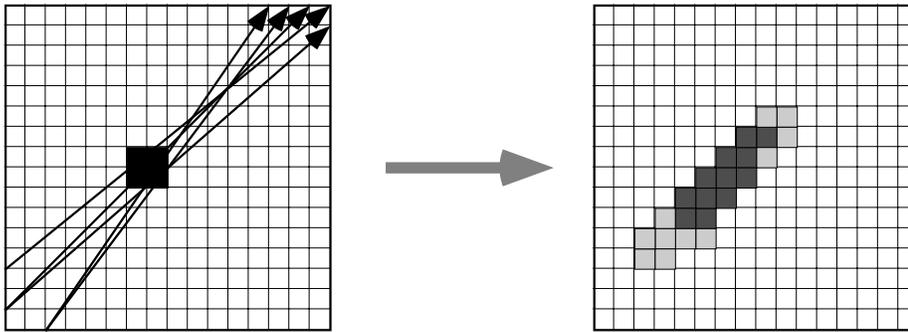


Figure 4.3: When only a limited range of ray angles are available, resolution of velocity anomalies is limited in the direction parallel to the rays.

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

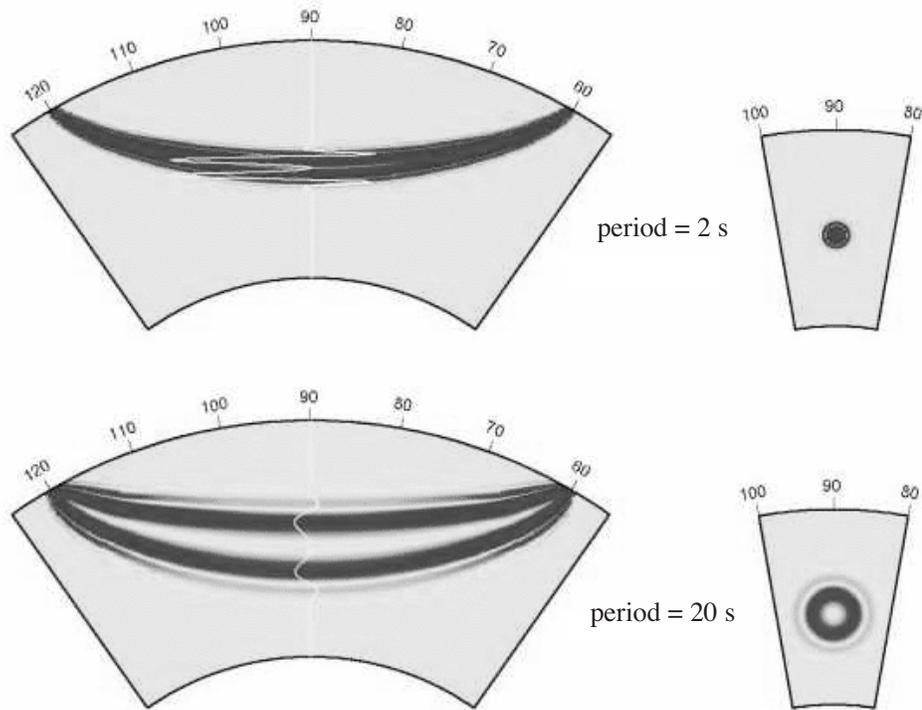


Figure 4.4: Banana-doughnut kernels showing the sensitivity of P -wave travel times at 60° epicentral distance to velocity perturbations in the mantle. The right-hand plots show the cross-section perpendicular to the ray direction at its midpoint. Note the much wider kernel at 20 s period compared to 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 tomography 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). However, it is clear that finite-frequency tomography represents a significant theoretical advance and will eventually become common practice. Researchers are now computing 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:

$$\mathbf{d} = \mathbf{G}\mathbf{m} \tag{5.1}$$

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, $\mathbf{r} = \mathbf{d} - \mathbf{G}\mathbf{m}$, 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 in 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

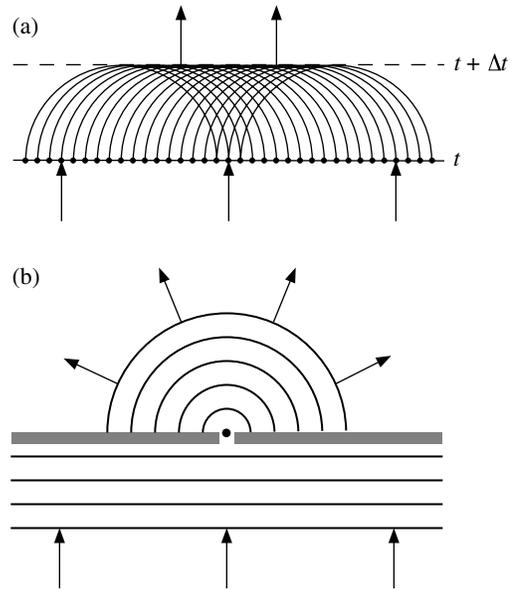


Figure 5.1: Illustrations of Huygens' principle. (a) A plane wave at time $t + \Delta t$ can be modeled as the coherent sum of the spherical wavefronts emitted by point sources on the wavefront at time t . (b) A small opening in a barrier to incident waves will produce a diffracted wavefront if the opening 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

$$t_0 = \frac{2h}{v}, \quad (5.5)$$

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

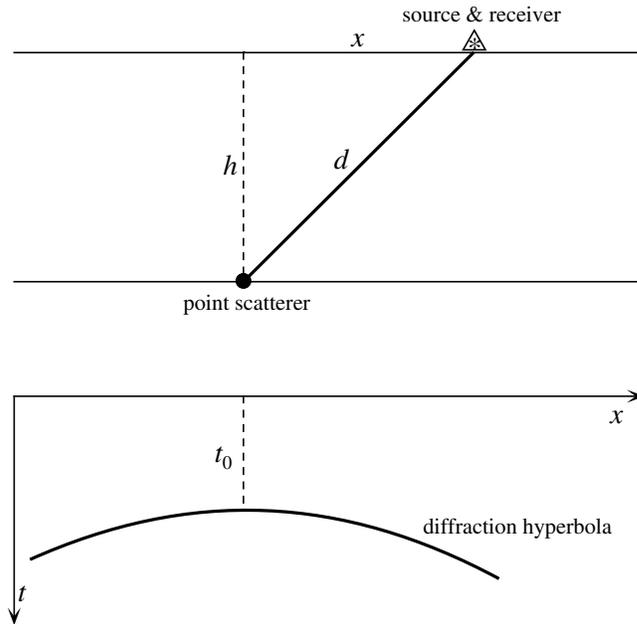


Figure 5.2: A point scatterer will produce a curved “reflector” in a zero-offset section.

given by

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

Squaring and rearranging, this can be expressed as

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

or

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

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

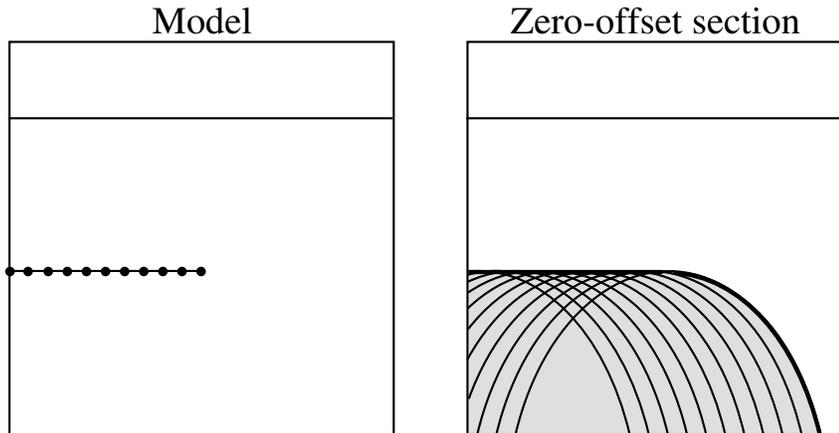


Figure 5.3: The endpoint of a horizontal reflector will produce a diffracted arrival in a zero-offset section. The reflector itself can be modeled as the coherent sum of the diffraction hyperbola from individual point scatterers. The diffracted phase, shown as the curved heavy line, occurs at the boundary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

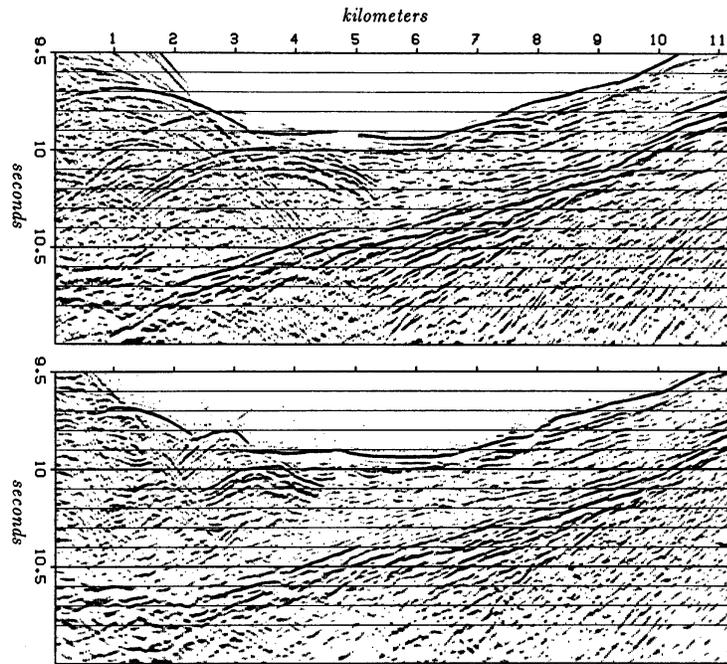


Figure 5.4: Original (top) and migrated (bottom) reflection data from a survey line across the Japan trench (figure modified from Claerbout, 1985; data from the Tokyo University Oceanographic Research Institute).

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 EXERCISES

1. (COMPUTER) Your 20-station seismic network recorded some high-frequency tremor. You may obtain 50 s of data from each station, sampled at 100 samples/s from the class website. The ascii file contains the (x, y) station location in km, followed by 5000 points, etc., for the 20 stations. You have reason to believe that the tremor is located at 30 km depth and that the crust has a uniform P velocity of 6 km/s. Use P-wave backprojection to locate likely source(s) of the tremor. (HINT: Do a grid search of possible tremor source locations, using x values from 1 to 100 km and y values from 1 to 100 km (i.e., try 10,000 possible source locations, all at 30 km depth). Stack the seismograms at each source grid point using the appropriate time shifts for the computed travel times to each station. Then compute the RMS for the time series, save this value for each of the 100x100 source locations, and plot the resulting image of source amplitudes.) What happens if you assume an incorrect value for the P velocity?

5.3 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.

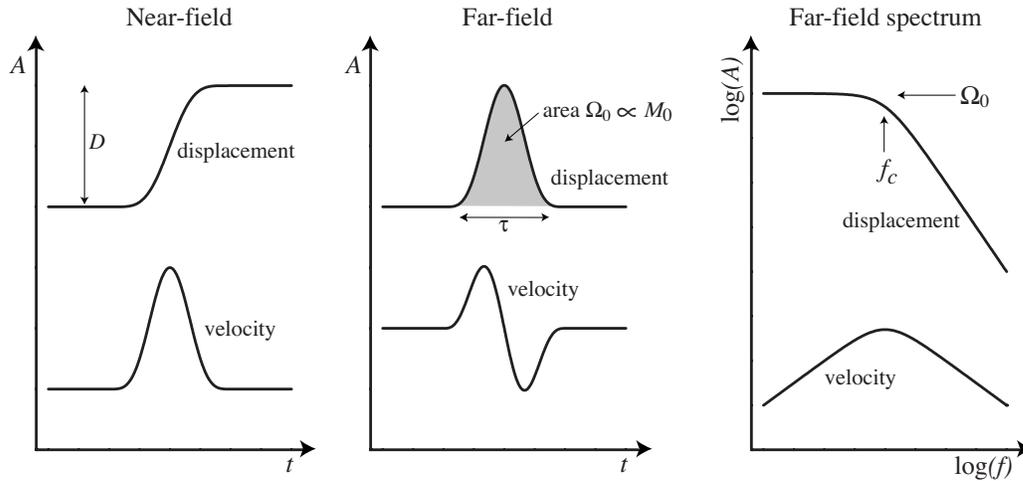


Figure 6.1: The relationships between near-field displacement and far-field displacement and velocity for time series (left two panels) and spectra (right panel).

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 $M_0 = \mu AD$, 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

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

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

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

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

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

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

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

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, $M_0 = \mu \bar{D}A$, 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 (\bar{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 stress¹ on a fault before an earthquake to the stress after the earthquake:

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

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

$$\Delta\sigma = \frac{7\pi\mu\bar{D}}{16r} = \frac{7M_0}{16r^3} , \quad (6.6)$$

where r is the fault radius, μ is the shear modulus, and \bar{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

$$\Delta\sigma = \frac{2\mu\bar{D}}{\pi w} = \frac{2M_0}{\pi w^2 L} . \quad (6.7)$$

More generally, we may write

$$\Delta\sigma = C\mu \left[\frac{\bar{D}}{\tilde{L}} \right] , \quad (6.8)$$

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).

¹In 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

$$r = \frac{k\beta}{f_c}, \quad (6.9)$$

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-

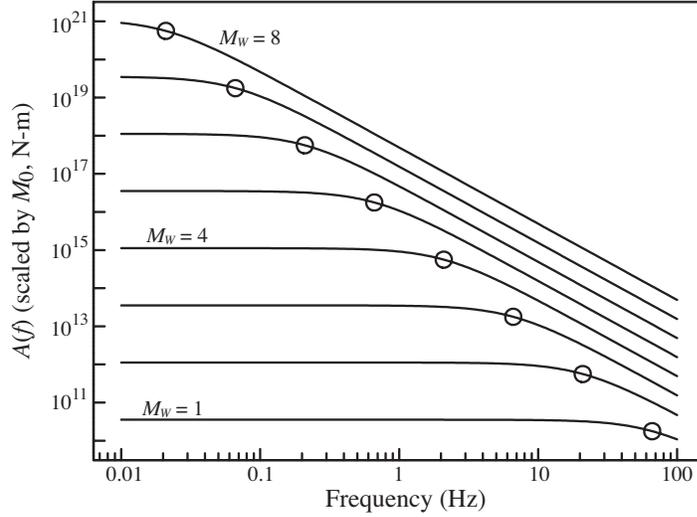


Figure 6.2: Predicted P -wave spectra from the Madariaga (1976) source model, assuming a constant stress drop of 3 MPa. The spectra have been scaled such that their amplitudes at low frequency are equal to their moments, M_0 . The circles show the corner frequencies (f_c). Individual spectra are for 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 ($f_c^P \simeq 1.5f_c^S$). 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

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

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 EXERCISES

1. Your borehole seismic experiment obtains the P -wave spectra plotted in Figure 6.3 at a distance of 10 km from an earthquake. Using a ruler, crudely estimate Ω_0 and f_c from the plot. Assuming that the density is 2700 kg/m³, the P velocity is 6 km/s, and the S velocity is 3.46 km/s, compute the moment,

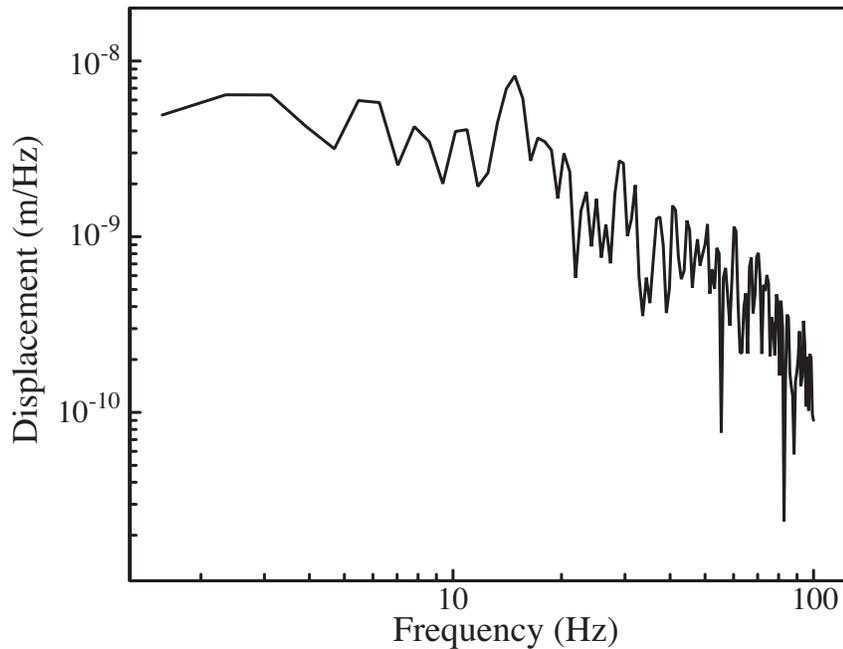


Figure 6.3: A P -wave spectrum from a borehole seismic experiment.

M_0 , the moment magnitude, M_W , the source radius, r , and the stress drop, $\Delta\sigma$, for this event. State any modeling assumptions that you make. Do not confuse the source radius r in Section 6.1 with the source-receiver distance r in Equation 6.1. The recording station is deep enough that you may assume that the effect of the free surface can be ignored and that the rock properties are uniform between the earthquake and the station. You may also assume that attenuation is negligible (or that the spectrum has already been corrected for attenuation) and that the radiation pattern term is simply its average P -wave value of 0.52.

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

6.4 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 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, $M_0 = \mu DA$, 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

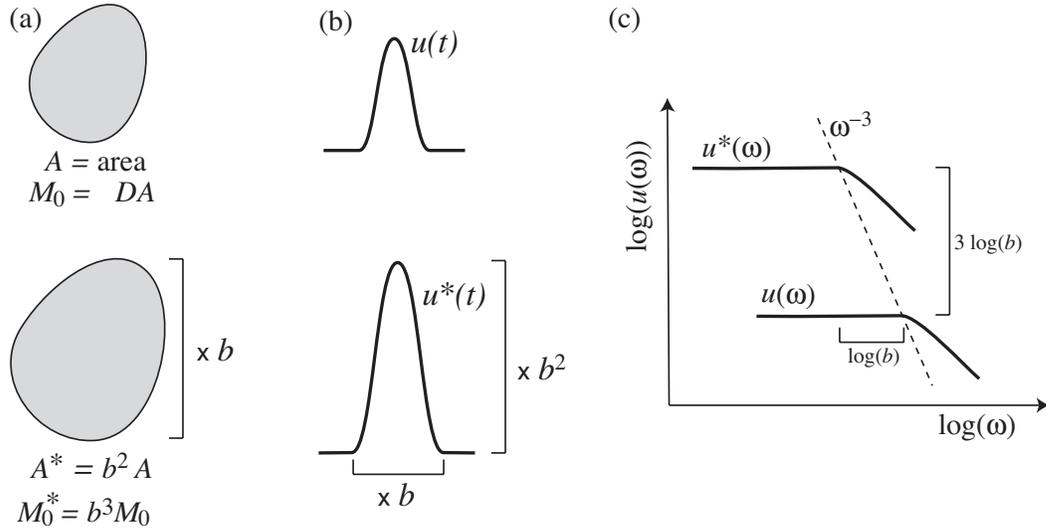


Figure 7.1: Illustration of the effects of self-similarity when an earthquake is increased in size by a factor b , showing the behavior of (a) rupture area and moment, (b) far-field displacement pulses, and (c) displacement spectra. Figure adapted from Prieto et al. (2004).

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

$$u^*(t) = b^2 u(t/b) \quad (7.2)$$

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

$$u^*(\omega) = b^3 u(b\omega) \quad (7.3)$$

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

$$f_c \propto M_0^{-1/3} \quad (7.4)$$

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)

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

where \dot{u}_α and \dot{u}_β 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}_α and \dot{u}_β 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.,

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

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}_α^2 and \dot{u}_β^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

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

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

$$q = E_R^S / E_R^P \quad (7.8)$$

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 $f_c^P \simeq 1.5f_c^S$, 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

$$\tilde{e} = \frac{E_R}{M_0} = \frac{1}{\mu} \frac{E_R}{\overline{DA}} \quad (7.9)$$

is called the *scaled energy* and is dimensionless (note that 1 joule = 1 N m). The parameter $\mu\tilde{e} = E_R/\overline{DA}$ has units of stress and has traditionally been called *appar-*

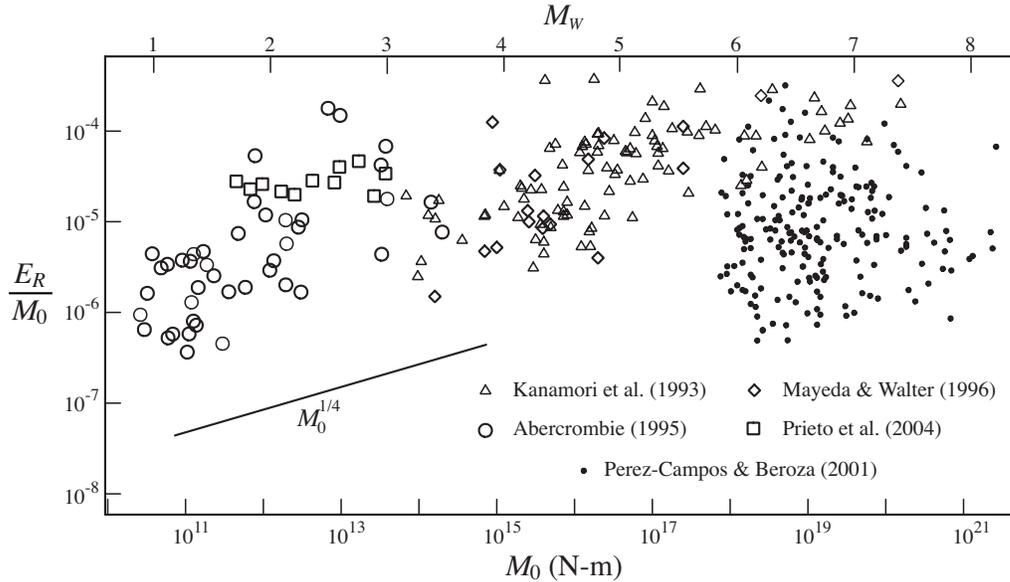


Figure 7.2: The observed radiated seismic energy to moment ratio, $\tilde{e} = E_R/M_0$, plotted as a function of moment. The $M_0^{1/4}$ trend noted in some studies is plotted for reference.

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 teleseismic 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

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

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 $\frac{1}{2}(\sigma_1 + \sigma_2) = \bar{\sigma}$ is the average shear stress on the fault so this is analogous to “work = force \times distance” from basic physics. As discussed in Kanamori and Brodsky (2004) and Kanamori and Rivera (2006), this is usually approximated as

$$E = \bar{\sigma} \bar{D} A = \frac{1}{2} \Delta \sigma \bar{D} A + \sigma_2 \bar{D} A \quad (7.11)$$

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

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

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 .

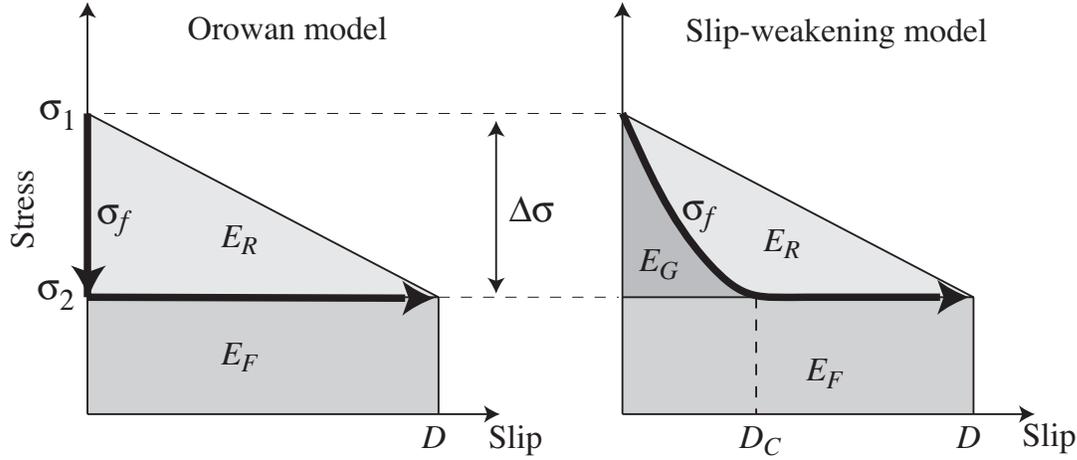


Figure 7.3: The shear stress, σ_f , on a point on a fault as a function of slip for the Orowan fault model and a simple example of a slip-weakening fault model. σ_1 and σ_2 are the initial and final stresses, D is the total slip, D_C is the critical slip, E_R is the radiated seismic energy, E_F is the frictional energy dissipated, and E_G is the fracture energy released.

Generalizing to a fault of area A , we have

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

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

In this case, the stress drop can be expressed as

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

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

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

and

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

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

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

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

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

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:

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

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,

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

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:

$$\eta = \frac{E_R}{E} = \frac{E_R}{\bar{\sigma} \bar{D} A} = \frac{\mu E_R}{\bar{\sigma} M_0} = \frac{\mu \tilde{e}}{\bar{\sigma}}. \quad (7.22)$$

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

$$E_{\min} = \frac{1}{2} \bar{\Delta} \bar{\sigma} \bar{D} A = \frac{\Delta \sigma}{2\mu} M_0 \quad (7.23)$$

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.