EARTH STRUCTURE & DYNAMICS

Gillian R. Foulger (Michaelmas Term)

Jeroen van Hunen (Easter Term)

Recommended Texts (Michaelmas Term)

Stein, S. and M. Wysession, *Introduction to Seismology, Earthquakes, and Earth Structure*, Blackwell Publishing, pp 498, 2003.

Lay, T. and T.C. Wallace, *Modern Global Seismology*, Academic Press, pp 521+xii, 1995.

1. Introduction

The Earth deforms on a broad range of timescales, from a fraction of a second to millions of years. Using earthquake seismology, the deformation of the Earth on timescales ranging from very short to a few thousand years can be studied.

Earthquakes occur mostly in well-defined belts around the Earth, the boundaries of plates. Their depths vary according to the boundary type. Beneath spreading plate boundaries they are generally shallower than 10 km. They tend to become deeper and larger as spreading rate decreases. Beneath transform plate boundaries they often extend down to 50 km depth but beneath subduction zones they occur as deep as 700 km. The seismicity of subduction zones varies. For example, a double zone exists in the Japan subduction zone. Earthquakes are sparse at intermediated depths in subduction zones. This may be related to the warming of the plate as it descends. The thermal re-equilibrium time of subducting slabs is of the order of the time it takes them to reach the bottom of the upper mantle at 650 km depth. Thus, the deepest earthquakes, which occur in this vicinity, may be related to a different process from the shallower earthquakes. They may be associated with the arrival of the slab at the partial barrier to convection at 650 km, and the deflection of the slab sideways. These very deep earthquakes have unusual focal mechanisms and it is not fully understood what causes them.

The shallower earthquakes in subduction zones are varied in type. Those in the lithosphere at the surface are associated with movement of the slab on thrust faults, bending of the lithosphere as it approaches the subduction zone, and deformation as it starts to subduct. Earthquakes at intermediate depths are attributed to the release of down-dip tensional and, at larger depth, extensional stresses.

There is strong interest in understanding the timescale of occurrence of large earthquakes, in order to mitigate public hazard. The main hazardous zones are the “Pacific ring of fire” and the Himalaya-Alpine belt. These are the regions where most of the world’s subduction zones lie. Scientists have studied the sequences of large earthquakes but so far this had not bourn much fruit. The “seismic gap hypothesis” has also been shown to be false.

Studying earthquake focal mechanisms was powerful in determining the nature of slip on transform fault plate boundaries. It was understood early on that the orientations of the transform faults, and their inactive extensions into the plates (fracture zones) indicate the directions of motion of the plates. When they were originally discovered, there was controversy regarding whether they displaced sideways portions of the spreading plate boundary and continually lengthened with time to progressively offset the ridge portions by longer and longer distances (the “transcurrent” fault situation). The alternative hypothesis was that they are of constant length and the offset remains essentially stable (the “transform” fault situation). These hypotheses predicted opposite senses of slip on the faults. Focal mechanisms solutions showed that the transform hypothesis was the correct one. It is interesting to note that the great Pacific fracture zones record no large change in motion of the Pacific plate at ~ 50 Ma, indicating that the 50˚ “bend” in the Emperor-Hawaiian volcanic chain did not come about by change in motion of the plate over a fixed "hot spot" or plume beneath.

Plate boundaries can be studied up to a point out at sea but the fact that they are submerged beneath several kilometres of water limits what can be done. There are, however, some places where they are exposed on land and although these boundaries are clearly anomalous in some sense by simple virtue of the fact that they are subaerial, it is still helpful to study them. The San Andreas fault is an example of a great transform fault that is largely on land. It is complex tectonically, with transpression and transtension features. Its palaeoseismicity has been studied intensively because of the importance of mitigating public earthquake hazard in California. "Trenching" has been conducted at various sites, the most famous of which is Pallett Creek. There, sedimentary sections exposed in cross sections cut through the river bed have been interpreted in terms of a series of earthquakes going back in time ~ 2,000 years. This research approach has not developed as much as was once expected, however, because of the ambiguities and difficulties in interpreting the data.

Earthquake recurrences and seismic gaps are still watched carefully on the San Andreas fault zone, which is continually monitored by the U.S. Geological Survey at Menlo Park, just south of San Francisco. A large research unit, including several dozen scientists, focuses on the fault. Among other things, the probabilities of large earthquakes occurring on different parts of the fault are estimated, in an effort to provide information that will result in building codes being strengthened.

2. Earthquake waves, rays & locations

2.1 Traditional methods

The seismic waves generated by earthquakes are either compressional ("P" or "primary") waves or shear ("S" or "secondary") waves. The designations P and S arose because P waves travel faster than S waves and thus arrive earlier on seismograms. The waves may be subdivided further into body waves, which travel through the interior of the Earth, and surface waves which propagate along the surface and are only sensitive to near-surface structure. These are designated Love and Rayleigh waves and are named after the scientists who identified them.

A seismogram comprises a record of ground motion at some distance from the earthquake hypocentre. It is made up of a train of many waves that have traveled along different paths from the source to the receiver. These paths have different lengths and pass through materials with different seismic wave speeds, thus arriving at different times at any given station. Although many different waves are studied by seismologists, the most important is the direct, earliest-arriving P wave. Because this arrives before any other wave, it is uncorrupted by the wave train from earlier-arriving phases and its arrival time can be picked most accurately.

The arrival times of various phases have been used to develop models of the structure of the interior of the Earth.

It is important to understand that seismology is essentially the only method that can image the interior of the Earth in any kind of detail. Although other methods in Earth science have been applied to try to gain insights into interior structure, *e.g.*, gravity, those methods are extremely weak compared with seismology. Geochemistry in particular has virtually no ability to reveal the interior structure of the Earth, despite the fact that geochemical data are commonly interpreted in terms of quite elaborate models of the interior of the Earth. Models of Earth structure erected using geochemical arguments are ambiguous and highly speculative.

In many situations, especially as regards seismic hazard, the location of an earthquake is of extreme importance. The simplest method to shed some light on the location of a local or regional earthquake is simply to multiply the S-P time by 8 to obtain an estimate of the epicentral distance from a seismic station. In this way, in emergency situations, an operator can instantly estimate the location of an earthquake before the full wave train has been recorded, merely by glancing at the earliest part of the seismogram.

The seismic waves recorded at stations are a function of source, path, and receiver effects. They are strongly affected by the structures through which they propagate. Body waves propagate in the Earth in the same manner as seismic waves from explosions. At velocity discontinuities they are refracted, reflected, and transformed into different waves (P into S and vice versa). Their amplitudes are increased or decreased by focusing and the attenuating effects of the material through which they pass. The shallowest layers of the Earth are particularly inhomogenous and can radically affect seismic recordings, masking potentially more interesting information about the deeper structure. For this reason, seismometers are often placed in boreholes a few hundred metres deep, in order that they may record earthquake waves before they have been corrupted by passage through the near-surface layers.

The "ray parameter" p is mathematically the horizontal slowness of a wave as it propagates across the surface of the Earth. Slowness is the reciprocal of velocity. p is a useful parameter and key in mathematical approaches to earthquake locations. Most computer programs that locate earthquakes do so in the following way:

1. start with an initial guess;

2. calculate the predicted arrival times at the seismic stations;

3. calculate the misfits;

4. move the hypocentre to reduce the misfits;

5. repeat 2. - 4. until the misfits reduce below a pre-declared threshold.

Examination of the equations for the change in travel time as a function of latitude, longitude and hypocentral depth shows that arrival-time observations are sensitive to changes in the former two but not to changes in hypocentral depth. For this reason, calculated depths tend to have much larger errors associated with them than latitudes and longitudes. The same principle applies to positioning using the Global Positioning System, where the position of a satellite receiver is calculated using signals transmitted by several satellites. In this case, the GPS receiver is analogous to the earthquake and the satellites are analogous to the seismic stations. In general, the horizontal position of a GPS receiver is calculated more accurately than the vertical position. It is not unusual for a hand-held GPS receiver to tell claim that it is below sea level!

The take-off angle of rays at an earthquake hypocenter is the angle between the downward vertical and the ray. A ray taking off vertically downwards has a take-off angle of 0˚. Diving rays emerge again at Earth's surface at progressively larger distances from the source as the take-off angle decreases. The richness and complexity of seismograms is built by the large diversity of arriving waves, which include ones that have bounced off Earth's surface one or more times and even reflected from the surface of the core.

Arrivals at seismic stations various distances from a source are complex as a result of Earth structural complexities including zones where the velocity increases or decreases rapidly. Such zones result in focusing and defocusing, which increase and reduce the amplitudes of arrivals. An extreme case of this is antipodal focusing, which increases the amplitudes of waves arriving at exactly 180˚ from a source by an order of magnitude (Rial, J.A., *Geophys. J. R. astr. Soc.*, **55**, 737-743, 1978).

A traditional labeling convention is used in earthquake seismology to denote waves that travel by different paths (Table 2.1). Innovative experiments have been designed using particular waves, to study problems of special interest. For example, wave speeds in the mantle directly beneath Hawaii were studied using waves from a large earthquake in Hawaii that reverberated between the surface and the core. Small-scale structure of the core-mantle boundary has been studied using waves that diffract along the surface of the core. Insight into the complexities of wave behaviour in regions of the Earth with rapidly-changing structure have been greatly aided by plots that show the trajectories of rays through the Earth. This kind of diagram was pioneered by Bruce R. Julian at the California Institute of Technology.

A detailed picture of the internal layered structure of the Earth was built up by plotting the times of arrivals of clear seismic phases as a function of distance from the source. Such plots revealed many new features which were used to further refine models of Earth structure. They may be used to calculate the epicentral distances of earthquakes from seismic stations.

Various different parameterizations of Earth structure can be used, including laterally homogeneous and laterally heterogeneous types, depending on the requirements of the work. Whole-Earth models include the Jeffreys-Bullen, PREM, and IASP91 models and these are used for general global earthquake locations. More specific study of particular regions of the Earth have shown that a widespread global low-velocity zone lies in the depth range ~ 50 - 200 km. This is thought to result from partial melt and this layer is interpreted as the asthenosphere. The core-mantle boundary region has been found to be particularly heterogeneous and is thought to contain patches of high-degree partial melt. This region is known as D" ("D double primed"). This is the only remaining usage of a terminology invented by Bullen, whereby different parts of the Earth are designated by letters. Confusion occasionally arises because Bullen placed the base of the upper mantle at the base of his "layer C", which is at ~ 1,000 km depth, whereas in general the base of the upper mantle is now considered to be the seismic velocity discontinuity at 650 km depth. Since the advent of modern, powerful computers and the accumulation of large amounts of arrival time data over the years, it has been possible to produce three-dimensional images of internal Earth structure using tomography.

Table 2.1 Naming convention of seismic phases in the Earth

	Compressional waves	**Shear waves**	
mantle	P	S	
core	K	-	
inner core	I	J	
Reflections from E's surface	PP	SS	double letters
Reflections from core surface	PcP	ScS	c
Reflections from inner core	PKiKP	SKiKS	i
Deep earthquakes reflected from E's surface (depth phases)	pP	sS	preceding lower case letter
Rays traveling only in the upper crust	Pg	Sg	
Rays traveling in the lower crust	P*	S*	
Rays traveling along the top of the mantle	Pn	Sn	
Waves can be converted	PS	SP	

Surface waves and normal modes may also be used to obtain information about internal Earth structure. Surface waves are dispersed. Thus, waves with longer periods arrive earliest in the surface-wave-train. Surface waves are sensitive to structure only in the upper few hundred kilometers of the Earth. The longer the period of the waves, the deeper they sense, but long-period waves can only sense large-scale structure. Traditionally, the relationship between period and apparent velocity is used to determine structure, and this approach has been applied to study variations in structure, *e.g.*, between aseismic ridges and "normal" oceanic lithosphere, between oceanic regions with different ages, and between the oceans and the continents.

Surface waves from large earthquakes travel around the entire globe, and thus may be recorded multiple times at a single station. Waves traveling in both directions arrive at a station, and the wave-trains may circumnavigate the globe several times in the cases of extremely large earthquakes. This can result in standing waves being excited everywhere on Earth's surface. The M ~ 9 Sumatra "Boxing Day" earthquake of 2004 was extremely large (it was a "great" earthquake) and provides an excellent example of this.

Great earthquakes also excite powerful normal modes. The Boxing Day earthquake provided the best recordings ever on the current, dense, modern network of seismic stations, and these data will enable many years of remarkable new findings. Normal modes are extremely long-period excitations of the planet. The longest periods are as long as the entire circumference of the Earth. Both spheroidal and torsional normal modes are excited.

A very large number of higher harmonics are detectable from an earthquake as powerful as the Boxing Day event, and these can be used to study the structure of the whole planet. They have the unusual ability, amongst seismic methods, to reveal density (instead of simply wave speed) and were recently used to show that the low-velocity "superplumes" in the lower mantle beneath the Pacific and Atlantic oceans are dense, chemical bodies and not low-density bodies that are buoyant as a result of high temperature.

2.2 Master event locations

This technique is used to compensate for errors in the travel time tables or regional/global crustal model used. It is assumed that the residuals of a "master event" are caused by unmodelled structure only. Then the differences in relative arrivals from a second event may then be attributed to the difference in relative location only.

This technique can only be used if the earthquakes are close to each other relative to the stations, *e.g.*, in the case of a mainshock and aftershocks recorded at distant stations.

There are three approaches to "master event"-type relocations:

1. Apply simple travel time corrections obtained from a "master event";

2. Relatively relocate a large group of earthquakes by clustering them and differencing arrival times amongst the entire set, and;

3. Spectral cross-correlation of multiplets (earthquakes that have almost identical waveforms) to obtain highly-accurate differential arrival times. *e.g.*, the Got method, and then application of method 2. This is a very recent method that has been applied with great success to earthquakes in Hawaii and is currently being used applied in the Department to volcanic and geothermal earthquakes.

Bibliography

Waldhauser, F. & W. L. Ellsworth (2000) A double-difference earthquake location algorithm: Method and application to the northern Hayward Fault, California. *Bull. seismol. Soc. Am.* **90**, 1353-1368.

3. Magnitudes & fractals

3.1 Magnitudes

Magnitude determination is an empirical method of estimating the size of an earthquake from the strength of ground motion. There are many different magnitude scales, *e.g.*, the local magnitude scale (popularly, but incorrectly known as the Richter scale):

$$M_L = \log_{10} A - \log_{10} A_0(\Delta)$$

where M_L is local magnitude, A is the maximum amplitude in mm and A_0 is the maximum amplitude at distance Δ for a "standard" (zero-magnitude) earthquake. Richter arbitrarily assigned a magnitude of 0 to an earthquake that gave an amplitude of 0.001 mm at 100 km, so as to avoid generating negative magnitudes. This ploy ceased to work when seismologists started to study very small earthquakes.

To determine a magnitude in practice, the seismologist first measures the amplitude of the earthquake of interest on the seismogram, then calculates Δ. The value of

$$\log_{10} A_0(\Delta)$$

is looked up in tables.

Other important magnitude scales include the surface-wave magnitude scale (M_s), the body wave magnitude scale (m_b) and the duration magnitude scale (M_d). It is important to note that *magnitude is not a fundamental property of earthquakes*. There may be a lot of variation in magnitude estimates because instruments of the same type may respond differently because of site effects, path attenuation and many other reasons. There is also great variation between magnitudes calculated for the same earthquake using different magnitude scales. For these reasons it is not meaningful to talk about the "accuracy" of a magnitude.

Magnitude has been related empirically to energy, and this has yielded the following relationship:

$$\log E = 11.8 + 1.5 M_S \ .$$

There is a great deal of scatter in the data, however, and the energy released by earthquakes that apparently have the same magnitude may vary by an order of magnitude. The above equation shows that an increase in one unit in magnitude corresponds to an increase in energy release by a factor of ~ 30.

3.2 Intensity

This is a measure of the strength of ground shaking and is quantified, for example, using the Mercalli scale. The Mercalli scale has been modified with time to update it to reflect the effects on modern objects, *e.g.*, cars and skyscrapers. Different scales have been developed that are suitable to different environments, *e.g.*, westernised urban, or rural Polynesian. For pre-instrumental, historic earthquakes, maps of intensity may be the only information available that enable an estimate of the location of the earthquake epicentre.

3.3 Seismic moment

Seismic moment is a measure of the work done at the source, and it is a measure of the energy released by a seismic event. It is measured in units of Newton metres (dyne cm in centimetre-gram-second (cgs) units). It is the most fundamental parameter related to earthquake size as only source parameters are involved and not instrument or propagation parameters. It is defined as:

$$M_0 = \mu A \bar{u}$$

where M_0 is seismic moment, μ is the shear modulus, A is fault area and u is the average slip.

Seismic moment may be determined by a) field observation of the length of the surface break and mapping of the size of the aftershock sequence, *e.g.*, in the case of the 1906 San Francisco earthquake, or b) using spectral analysis. An earthquake source may be approximated by the convolution of two boxcars, and the spectrum of such a source will have three distinct trends, separated by 2 corner frequencies.

For seismic waves with period T:

- if the rise time < rupture time < T (*i.e.*, long-period waves), then the spectrum will be flat,
- if rise time < T < rupture time (intermediate-period waves), then the spectrum will decay as $1/\omega$
- if T < rise time < rupture time (short-period waves), then the spectrum will decay as $1/\omega^2$

The corner frequency is usually defined as the intersection of the flat and the $1/\omega^2$ segments. An idealised, simplified body wave spectra, corrected for all propagation effects will have the following features:

- constant spectral level at low frequencies,
- a corner frequency ω_o, and
- a steep roll-off for $\omega > \omega_o$

The constant spectral level is related to seismic moment, and the corner frequency to stress drop. A magnitude scale known as the moment magnitude (M_w) has been developed, which is a best-fit of moment to commonly used magnitude scales (see derivation in next section). It yields a number that is magnitude-like, and thus satisfies end-users such as the general public, journalists, and science practitioners who are accustomed to understanding earthquakes in terms of magnitude. On the other hand, it is a fundamental measure of the amount of energy released by an earthquake, and in theory the same moment magnitude should be obtained by all scientists, regardless of what instrument or seismogram is used. It thus does not suffer from the unsatisfactory and misleading problems suffered by conventional magnitudes.

3.4 Stress drop

Stress drop is:

$$\Delta\sigma = C\mu\frac{\bar{u}}{L}$$

where *L* is the characteristic dimension of fault. *u*, the average slip, is quasi proportional to L and thus Δσ is quasi constant! It usually lies between ~ 10 and 100 bars. The discovery that the stress drop in earthquakes is roughly constant was one of the big surprises of seismology. It can be shown that the radiated seismic energy is proportional to stress drop:

$$E_s = \frac{1}{2}\Delta\sigma\bar{u}A$$

Thus:

$$E_s = \frac{\Delta\sigma}{2\mu}M_0$$

Using:

$$\log E = 11.8 + 1.5M_S$$

and assuming a typical stress drop of 30 bars, we get:

$$\log E = 11.8 + 1.5M_S$$

in units of dyne cm. Dividing through by 1.5, we can define a new magnitude scale:

$$M_W = \left(\frac{\log M_0}{1.5}\right) - 10.73$$

or:

$$M_W = \frac{2}{3}\log M_0 - 6.0$$

in units of N m. (1 Nm = 10^7 dyne cm).

There have been a lot of studies of the empirical relationships between M_w, M_o, fault area, different magnitude scales, source time function, surface rupture length and average displacement. Linear relationships are generally obtained, but only in a log-log sense. It is important to realise that almost any data that are significantly distributed look linearly correlated on a log-log plot, and thus there is, in truth, rather little correlation. Given one of the above-listed parameters, using these empirical relationships, another could only be predicted to within an order of magnitude (*i.e.*, a factor of 10) of the correct value.

3.5 The fractal nature of earthquakes and b.

Earthquakes are "fractal", or "self-similar" in many ways. "Fractal" means "a shape made of parts similar to the whole in some way". An example is the Koch curve. The fractal dimension is given by:

$$Nr^{D} = 1$$

where D is the fractal dimension, N is the number of parts into which the line is split and r is the ratio of similarity, *i.e.* of sizes of elementary units in successive iterations. This means that it is not possible to determine the scale by examining a sample. Earthquakes are self-similar in a number of ways, *e.g.*, fault area and moment, fault length and moment, and magnitude and number.

Magnitudes generally obey the Gutenberg-Richter law:

$$\log_{10} N = a - bM$$

where N is the cumulative number of events and M is magnitude, *i.e.,* a plot of N *vs.* M is a straight line with a constant slope. This slope is known as the *b-value*. b is ~ 1 for the world, and thus there is a 10-fold increase in the number of events for each decrease in magnitude unit. Since an earthquake one magnitude unit smaller than another releases only 1/30 of the energy, this means that 10 small earthquakes do not release the same amount of energy as one large one.

Substituting the equation:

$$\log M_0 = 1.5 M_S + 16.1$$

into the Gutenberg-Richter Law, we get:

$$N.\, M_0^{(b/1.5)} = A'$$

This has the same form as the equation that describes the Koch curve. Thus the relationship between earthquake number and magnitude is self-similar. Globally there is about one magnitude 8 earthquake per year, 10 magnitude 7 earthquakes and 100 magnitude 6 earthquakes, *etc.*

Variations in the value of b are of interest. b can vary between ~ 0.5 and 5.0 (in the extreme). b must be calculated using the method of maximum likelihood because the uncertainties of the points that make up the data distribution are larger for large-magnitude earthquakes, where there are few events. The full equation is:

$$\beta = \left[\overline{m} - \frac{(m_{MIN} - m_{MAX} \exp(-\beta(m_{MAX} - m_{MIN}))}{1 - \exp(-\beta(m_{MAX} - m_{MIN}))} \right]^{-1}$$

where $\overline{m}$ is the average magnitude, m_{MIN} and m_{MAX} are the maximum and minimum magnitudes used in the calculation and $b = \beta/\log_{10}e$. Approximations to this equation exist, which simplify calculations.

The magnitude-frequency plot shows interesting features at both small and large magnitudes. At small magnitudes the numbers of earthquakes fall short of predictions because the located set becomes incomplete as a result of deteriorating recordings. The numbers also fall off at high magnitude. This is thought to be because there is a natural upper limit to the size of a fault break, that is related to the thickness of the brittle lithosphere and the maximum length of geological faults. There may also be instrumental effects, *e.g.*, deep earthquakes do not excite powerful surface waves and thus tend to have relatively small M_s. The probabilities of very large earthquakes reduces with size.

The maximum theoretical magnitude for a seismogenic region can be calculated by fitting an asymptote to the frequency distribution of the largest earthquakes using Gumbel statistics. The maximum m_b calculated for the UK is 5.7, and the maximum M_L for the world is 9.2. It has also been suggested that the distribution is described by two different scalings. According to this view, there is a break point where the whole brittle crust is ruptured, but there is no theoretical limit to fault length. It is difficult to decide which of these theories is correct because the numbers of large earthquakes are so small that differences in the calculated goodnesses of fit to the data using the two theories are not statistically significant.

After a large earthquake, aftershocks almost always follow. Like other earthquakes, these are fractally distributed and the most powerful aftershocks are often just one magnitude unit smaller than the mainshock. Their numbers fall off exponentially with time according to Omori's law:

$$n = \frac{C}{(K+t)^P}$$

where n is the frequency of aftershocks at time t after the mainshock, and K, C and P are fault-dependent constants. The 2004 $M_W = 9.0$ Sumatra Boxing Day earthquake may be expected to generate extremely powerful aftershocks that form part of a sequence that will last for several decades.

3.6 Chaos

Chaos results from sensitivity to initial starting conditions and may result from extremely simple systems. Chaos is *deterministic* and does not imply randomness. It results from well-defined mathematical relationships and the results can be predicted exactly, but not very far in advance without computers larger than the universe. Well-known examples of chaotic systems are the weather and social interactions. Any one of us could be run over by a beer truck tomorrow, simply because the driver delayed his journey that morning by two seconds because of dropping his keys before getting into the truck.

Mathematical models have been developed of how crack/earthquake systems develop, and these have been used to study the dependence of D, b etc. on various parameters. The objective is to try to understand better the spatial and magnitude distribution of earthquakes in order to contribute to earthquake prediction. However, it has been suggested that "*the chaotic, highly nonlinear nature of the earthquake source process makes prediction an* *inherently unrealizable goal*" (Geller, 1997). This field is currently causing a fundamental re-evaluation of the possibility of earthquake prediction.

Bibliography

Geller, R.J., Earthquake prediction: A critical review, *Geophys. J. Int.*, **131**, 425-450, 1997.

Huang, J. and D.L. Turcotte, Are earthquakes an example of deterministic chaos? *Geophys. Res. Lett.*, **17**, 223-226, 1990.

Bak, P. and C. Tang, Earthquakes as self-organized critical phenomena, *J. Geophys. Res.*, **94**, 15,635-15,637, 1989.

4. Source mechanisms

4.1 Introduction

There are many causes of earthquakes, *e.g.,* faulting, volcanic eruptions, explosions, nuclear, chemical or natural and landslides. A common process that generates earthquakes is shear motion on geological faults. This can be of strike-slip, normal, or thrust type. Such motions occur in response to different orientations of the principal axes of stress, σ_1, σ_2 and σ_3.

As seismic waves travel outward from the source, the sense of the first-arriving P wave may be compressional or dilatational, depending on the direction of take-off from the source, the orientation of the fault plane, and the sense of slip. The Earth may be conceptually divided into four "quadrants" by the fault plane and an imaginary plane (the "auxilliary plane") normal to it. For waves travelling away from the source in a quadrant towards which ground motion initially occurred, the first P wave will be compressional. For waves travelling in quadrants away from which ground motion initially occurred, the first P wave will be dilatational.

The sense of motion of the first P-wave arrivals (compressional or dilatational) may be mapped on an imaginary sphere surrounding the hypocentre. This is known as the "focal sphere". For shear motion on a planar fault, in theory half the focal sphere will experience compressional first arrivals and the other half dilatational first arrivals. When mapped on the focal sphere, these produce the familiar "beach ball" pattern, with four quadrants, two of which are compressional (conventionally shown in black) and the other two of which are dilatational (conventionally shown in white). Such plots are known as fault plane solutions because they rely on the assumption that the source comprised shear motion on a planar fault.

In practice, traditionally, the seismic stations that recorded the earthquake are plotted on a stereonet, that may be an equal-area- or stereographic net, and the fault and auxiliary planes are deduced to be the orthogonal great-circle pair that separate the compressional and dilatational arrivals into four fields. The "pressure" (P) and "tension" (T) axes can also be deduced. These fall in the geometric centres of the dilatational and compressional quadrants respectively. In a freshly faulting rock, these are approximately the same as the principal stress axes σ_3 and σ_1. However, most earthquakes occur on pre-existing faults and thus this does not hold exactly. In these cases the only thing that can be said is that the direction of greatest compressive stress is somewhere in the dilatational quadrant, and the direction of least compressive stress is somewhere in the compressional quadrant.

The orientation of the fault plane and the direction of slip may be estimated from the fault plane solution, with a 90° ambiguity, because the fault and auxilliary planes cannot be distinguished from the focal mechanism alone. This ambiguity can often be eliminated, however, with reference to geological information, the orientation of a surface rupture or the geometry of an aftershock sequence.

Difficulties inherent in deriving focal mechanisms using the above method include:

- violations (inconsistent polarities) occur;
- a "best-fit" result is not obtained, only a suite of possible solutions; and
- if a shear-faulting mechanism is assumed *a priori*, then this will preclude any non-shear components from being detected.

The solution is to utilize more information from the seismogram to reduce ambiguity– amplitudes or the whole waveform. To include such data computers are used and the shear-faulting assumption can then be dropped. In addition to using P-wave polarity data, S waves, and full waveform inversion methods have been developed that can derive focal mechanisms for large earthquakes that are well recorded on digital, three-component stations.

The history of understanding the shear-faulting seismic source is ironical. It was originally thought that earthquakes that were caused by shear slip on a fault could be represented by a single couple. Other scientists thought that earthquakes were caused by volume collapse and they concluded that they had double couple source mechanisms. Ultimately it was discovered that both camps were half right. Many earthquakes are caused by shear slip but the body force equivalent is a double couple because moment must cancel or else rotation occurs.

4.2 Earthquake radiation patterns

The radiation pattern of seismic waves emanating from an earthquake is given by the equation:

$$u(x,t) = \frac{1}{4\pi\rho\alpha^3} A^{FP} \frac{1}{r} \dot{M}_o\left(t - \frac{r}{\alpha}\right) + \frac{1}{4\pi\rho\beta^3} A^{FS} \frac{1}{r} \dot{M}_o\left(t - \frac{r}{\beta}\right)$$

where

$$A^{FP} = \sin 2\theta \cos\phi \hat{r}$$
$$A^{FS} = \cos 2\theta \cos\phi \hat{\theta} - \cos\theta \sin\phi \hat{\phi}$$

u = displacement, **a** = P-wave speed, **b** = S-wave speed, and $\hat{r}, \hat{\theta}, \hat{\phi}$ are the unit vectors in the r, θ, ϕ directions.

Inspection of this equation reveals the radiation pattern of P (the 1st term) and S (the 2nd term) waves. P is proportional to $\sin 2\theta$ and S is proportional to $\cos 2\theta$. It can be seen from this that the amplitude of S-wave radiation is proportional to the gradient of the P-radiation because $\frac{\partial(\sin 2\theta)}{\partial\theta} \propto \cos 2\theta$.

4.3 The double-couple source

It is now known that the radiation patterns measured from earthquakes are not perfectly quadrantial in form, and that in many cases this results from non-shear source effects and not from path effects. Source processes that have some component of motion normal to the fault plane generate non-double-couple radiation patterns. Some possible processes that might cause such earthquakes are:

- Tensile crack formation in the presence of the inflow of high-pressure fluid. This is likely in geothermal areas,
- dyke injection,
- cavity collapse *e.g.,* in mines,
- other source effects *e.g.,* rupture propagation effects, anisotropy, and
- path effects that distort the radiation field.

Whatever the explanation, motion in three dimensions must be involved, *i.e.,* including perpendicular to the fault plane if one is involved.

The nodal surfaces of such radiation patterns are not orthogonal great circles. Because the constraining assumption that they are has to be dropped, such earthquakes cannot be studied effectively by using P-wave polarities only. Such data are very insensitive to non-shear (non-double-couple) components in the radiation patterns. P- and S-wave amplitudes, amplitude ratios, and whole waveforms can provide the data necessary to constrain non-double-couple components. For large earthquakes that are well recorded at three-component digital stations, waveform inversion methods have been developed to calculate such focal mechanisms. This method is not suitable for small earthquakes, *e.g.*, in geothermal areas, and for such earthquakes P:S wave amplitude ratios are used. The use of amplitude ratios instead of the amplitudes themselves causes major error sources, *e.g.*, focusing and de-focusing by path structural complications, to approximately cancel out.

Amplitude ratios may vary from zero to infinity. For the purposes of visual assessment of the quality of the results they may be effectively illustrated as the directions of unit-length arrows rather than symbols such as the areas of circles.

The best representation of a source where there are motions in three dimensions is the seismic moment tensor. This is a 3 x 3 tensor representing the general form of far-field displacements as the 9 possible couples in a Cartesian co-ordinate system. This representation expresses all the information that can be obtained from waves whose wavelengths are much greater than the fault dimensions, *i.e.,* the source is effectively a point source with an associated radiation pattern. Usually seismologists assume all body forces are couples, otherwise there would be a net force. The latter may be necessary to consider, however, *e.g.,* for landslides or volcanic earthquakes caused by sudden mass advection.

Moment may also be a function of time:

$$M_o = \mu \bar{u}(t) A$$

The "rise time" of a source is not zero, and this may be important for understanding large earthquakes. This may be expressed by higher-order moment tensors.

The far-field radiation from an instantaneous shear dislocation on a planar fault in an isotropic medium–a double-couple (DC)–corresponds to two orthogonal couples, *e.g.,* (1,2) + (2,1) *or* (1,3) + (3,1). Many natural earthquakes are thought to be well modelled in this way. The tensor elements vary with orientation of the source. The relationship with seismic moment is:

$$M = M_o \begin{pmatrix} 0 & 0 & 1 \\ 0 & 0 & 0 \\ -1 & 0 & 0 \end{pmatrix}$$

Non-double-couple sources have equivalent body forces that are not a diagonal pair. Examples are:

<u>An explosion:</u>

$$M = M_o \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

or:

$$M = \frac{4\pi}{3} a^3 (\lambda + \frac{2}{3}\mu) \Delta\theta \begin{pmatrix} 1 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}$$

where **a** = the radius of the spherical volume and $\Delta\theta$ = the fractional change in volume.

<u>A tensile crack:</u> For slip perpendicular to a crack in the 3 direction:

$$m = \begin{pmatrix} \mu\left[u_3(x,t)\right] & 0 & 0 \\ 0 & \mu\left[u_3(x,t)\right] & 0 \\ 0 & 0 & (\lambda + 2\mu)\left[u_3(x,t)\right] \end{pmatrix}$$

where $\lambda \cong \mu \cong 3 \times 10^{10}$ N m^{-2} (the Lame constants).

Thus a tensile crack is represented by elements of the moment density tensor in the ratio:

$$1:1:\frac{(\lambda+2\mu)}{\lambda}$$

An approximation is:

$$M = M_o\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & 3\end{pmatrix}$$

<u>A compensated linear vector dipole (CLVD):</u>

$$M = M_o\begin{pmatrix}1 & 0 & 0\\ 0 & 1 & 0\\ 0 & 0 & -2\end{pmatrix}$$

Real non-double-couple sources are more complicated, and have moment tensors of the general form:

$$M = M_o\begin{pmatrix}a & 0 & 0\\ 0 & b & 0\\ 0 & 0 & c\end{pmatrix}$$

with a+b+c not necessarily = 0 and possibly all different. It is known as "conserving volume" if a+b+c=0.

The objective of studying earthquake source processes is usually to try to understand processes in the Earth. In order to facilitate this, sources are often "decomposed" into sub-sources that can be understood in terms of physical processes. Unfortunately, there are many ways of doing this–the process is non-unique–and if not done carefully the results do little to help understanding. For example, a general source that "conserves volume" (*i.e.*, elements of the moment tensor sum to zero) can be decomposed into two orthogonal double couples:

$$\begin{pmatrix}a & 0 & 0\\ 0 & b & 0\\ 0 & 0 & c\end{pmatrix} = \begin{pmatrix}-c & 0 & 0\\ 0 & 0 & 0\\ 0 & 0 & c\end{pmatrix} + \begin{pmatrix}-b & 0 & 0\\ 0 & b & 0\\ 0 & 0 & 0\end{pmatrix}$$

where (a+b+c)=0. This represents simultaneous slip on orthogonal planes, with orthogonal directions of the principle stress axes. It is unlikely that the Earth experienced double, simultaneous failure at the same point in space, in response to stress fields orientated orthogonally to one another. The only geologically realistic way of decomposing general moment tensors is into sub-components with consistent directions of principal stress. This is achieved by decomposing them into a DC + CLVD.

Many sources do not "conserve volume" and common decompositions for those are:

1. explosion + two DCs; and

2. explosion + DC + CLVD.

The only geologically realistic case is 2.

Source type (*i.e.*, explosion, CLVD or DC) can be displayed on the "source-type plot", which is a probability equal-area plot. It contains no information on orientation, and so all DC sources plot at a single point. This is because a DC mechanism contains orientation information only. Sources that are combinations of shear faults and cracks plot in various regions depending on the relative orientations of the two components. The source-type plot is useful for studying variations in mode of failure accompanying geothermal injection experiments and volcanic episodes. Moment tensor analysis was recently used to investigate a mining disaster in Utah that cost the lives of several miners and rescuers. It was able to distinguish between the disaster having been precipitated by a mine gallery collapse and a natural earthquake.

Deep-focus earthquakes are also non-double-couple. They occur at depths where high temperatures preclude brittle failure. They were once attributed to the change from downdip compression to tension, but this theory was discredited by focal mechanism work.

The main differences between deep-focus and shallow earthquakes are:

1. There are fewer aftershocks;
2. the rise times and durations are shorter;
3. source time functions are more symmetrical–shallow earthquakes have most of the moment release early;
4. there is usually no spatial relationship with aftershocks;
5. aftershocks and subevents do not cluster on planes; and
6. most non-DC focal mechanisms have a small or no isotropic component (< 10% of the moment).

These observations suggest that deep-focus earthquakes are caused by different processes from shallow earthquakes. Two candidate explanations that fit the observations are:

1. the earthquakes comprise different sub-events with differently-orientated double-couple mechanisms; and
2. "transformational faulting is occurring. This is a rapid phase transformation (olivine –> β spinel) in a thin zone which facilitates shear faulting. The mechanism of such an event would not necessarily be a double couple and would not necessarily have a large volumetric component.

Bibliography

Foulger, G.R. & R.E. Long, Anomalous focal mechanisms; tensile crack formation on an accreting plate boundary, *Nature*, **310**, 43-45, 1984.

Julian, B.R., Evidence for dyke intrusion earthquake mechanisms near Long Valley caldera, California, *Nature*, **303**, 323-325, 1983.

Julian, B.R., A.D. Miller and G.R. Foulger, Non-double-couple earthquakes 1. Theory, *Rev. Geophysics*, **36**, 525-549, 1998.

Miller, A.D., G.R. Foulger and B.R. Julian, Non-double-couple earthquakes 2. Observations, *Rev. Geophysics*, **36**, 551-568, 1998.

5. Earthquake seismic instrumentation

5.1 The history of seismic recording

The history of seismic recording is summarised in the following table:

132 AD	The first seismoscope was made in China, a vessel with dragons heads and frogs.
Early 18th C	Italian seismoscopes.
1784	First attempt to record time of shaking.
1851	The speed of seismic waves moving across the surface was first measured.
1875	The first true seismography was invented in Italy. The relative motion between a pendulum and the Earth was recorded as a function of time.
1887	The oldest known seismogram. Instrumentation rapidly developed from there, with mechanical or optical amplification of mass motion, with friction providing damping.
1900	The first global array of 40 photographically recording horizontal-component seismographs.
1914	Electromagnetic seismometers were developed, where the mass is a magnet moving in an electric coil.

The first instruments were seismoscopes, *i.e.* instruments that detected earthquakes but did not produce records. The era of analogue recordings from inertial instruments began in the late 19th century, and the modern era of widespread electronic broadband instruments, GPS clocks and fully digital recording started in the 1990s.

5.2 Seismometers

Prior to the recent advent of powered, broadband instruments, most seismometers were based on the principle of the inertial pendulum. A large majority of the seismic data that has been recorded was obtained using instruments of this kind. A rigid frame is fixed to the ground, a pendulum swings, and its motion lags behind that of the ground because of inertia. Such mechanical systems may be represented by a mass on a spring, plus damping provided by a dashpot. The response of the mass to ground shaking is recorded.

The Milne-Shaw seismometer is one of the earliest instruments that came into widespread use. It dates from the early 20th century and was amongst the earliest deployed in observatories. Traditional observatory instruments commonly recorded on photographic paper (via a light beam that was deflected by movements of the mass), heat-sensitive paper (using a hot stylus), sooted paper (inscribed by a needle), or ordinary writing paper (inscribed by an ink pen). The paper is usually mounted on a rotating drum, and the inscriber moves slowly across the drum over a 24-hr period. In this way, a continuous record is written as a spiral. The paper must be changed daily. A drawback of photographic paper was that it was expensive. Ordinary paper was cheap but does not provide a very sharp trace. Sooted paper yielded a very sharp trace, but was messy to use.

Field portable instruments were developed from about the 1960s onward. They commonly comprised simple short-period vertical geophones or one-component seismometers with natural frequencies of one, or up to a few Hz. Deployment of three-component stations often involved installing three separate instruments about the size of a roll of kitchen paper (but a lot heavier!), limiting operations. Smaller, three-component geophones with natural frequencies of a few Hz later became available. Seismometers that could be deployed in spheres and lowered onto the sea floor (ocean-bottom seismometers) and very insensitive instruments that can record strong ground motion without going off scale, were developed later.

It is necessary to know the responses of instruments to ground motion, in order to deploy instruments suitable for the task at hand, and in order to make corrections and retrieve actual ground motion. For inertial instruments the response to ground motion was generally provided in graphical form. In the case of modern instruments, the relationship between ground motion and instrument response is usually expressed as a set of "poles" and "zeroes". These express the Laplace transform of the impulse response as the ratio of two polynomials, and describe the entire response.

The response function of a seismometer may always be expressed as the ratio of two polynomials:

$$\frac{x(s)}{u(s)} = \frac{b_0 + b_1 s + b_2 s^2 \ldots}{a_0 + a_1 s + a_2 s^2 \ldots}$$

where *x(s)/u(s)* is the ratio of the output to the input in the frequency domain. These polynomials may be expanded and the equation written:

$$\frac{x(s)}{u(s)} = \frac{A(s - z_1)(s - z_2)(s - z_3)\ldots(s - z_n)}{(s - p_1)(s - p_2)(s - p_3)\ldots(s - p_n)}$$

The $z_{1\text{-}n}$ are called the zeroes because if one of them equals s then $x(s)/u(s) = 0$.

The $p_{1\text{-}n}$ are called the poles because if one of them equals s then $x(s)/u(s) = \infty$.

The sharpness of the sensitivity roll-off is governed by how many poles and zeroes there are in a response function. The values of the poles and zeroes are provided by the instrument manufacturer.

Instruments are dominantly sensitive to displacement or acceleration, depending on the frequency of the ground motion compared with the natural frequency of the instrument:

- If the frequency of ground motion is high, the instrument output is proportional to displacement. This may be understood intuitively, since if the ground moves fast, the mass is left behind.

- If the frequency of ground motion is low, the instrument output is proportional to acceleration. This may be understood intuitively, since if the ground moves slowly, the mass follows ground motion with a slight lag.

Instruments designed to record displacement thus have relatively low natural frequencies, and instruments designed to record acceleration (accelerometers) have relatively high natural frequencies, perhaps as high as 5-10 Hz. Long period seismometers have very large masses (20-ton masses have been used!), or mechanical arrangements that behave the same way as a large mass would.

Modern seismometers are sensitive to ground velocity because motions of the pendulum mass are converted to voltage, which is proportional to mass velocity. Selecting with care the natural frequency of a seismometer enables electronic magnification of the frequencies of interest. The overall frequency response of the instrument is the product of the pendulum, transducer and galvanometer frequency responses. Traditional seismometers are purely mechanical instruments and do not require power.

Broadband seismometers have a flat response over a broad range of frequencies *e.g.*, 50 Hz to 100 s. Digital filtering is used to remove unwanted frequencies for specific applications. Such seismometers feature a force-feedback system, so the mass cannot make large excursions and the instrument can be small. This also increases the *bandwidth* (*i.e.* the frequency range) and the *linearity* of the seismometer. Force-feedback is achieved by a negative feedback loop whereby a restoring force is applied which is proportional to the inertial mass displacement so mass does not move significantly.

Broadband seismometers are electronic as opposed to purely mechanical and thus they must be powered. They produce purely digital output, often in very large quantities, depending on the sampling rate (*e.g.*, 10 samples per s (sps), 500 sps). The volume of data produced may be several tens of Mbyte/day per station, even using compression. Examples are the Strecheisen STS-1 and the Guralp 40T, 3ESP and 3T types. Their frequency sensitivity is much greater than any inertial pendulum seismometer.

Usually, two or more sets of traditional seismometers had to be deployed to record both the short- and long-period waves that are of interest to seismologists, whilst avoiding recording powerful oceanic microseismic noise that has a peak period of ~ 6 s and exceeds other signals in amplitude. These signals are caused by standing waves in the ocean and are particularly troublesome on islands surrounded by wide oceans. Noise is measured in decibels = $10\log_{10}$(signal_power), where:

signal_power is proportional to $(\text{signal\_amplitude})^2$

Thus 20 dB is equivalent to a factor of 10 in amplitude of ground acceleration.

Noise spectra commonly show that ground noise varies by up to a factor of 10^4 with the 6-s microseism peak swamping all other signals. Traditional seismic stations are designed to be insensitive to the 6-s microseism noise peak. This is vital if paper recording is used because in that case the seismograms cannot be filtered. Broadband seismometers have a flat response across all frequencies and thus the operator must usually filter the microseisms out, either with a low-pass or a high-pass filter. This is only possible with digital recording.

Broadband seismometers are now fairly compact. They have been miniaturised to about the size of the old 1-Hz field one-component seismometers. Their main problem in field deployment situations is that they are sensitive to rapid changes in temperature. They are thus typically insulated to ensure that they change temperature slowly.

In modern deployments, time-keeping is done using GPS clocks, instead of the old quartz-crystal clocks and radio signals used in the past. Data are either stored at the site, which must then be visited at intervals to change recording disk, or they are transmitted digitally to a recording centre via digital radio links or the internet. Data tend to be stored at central facilities, and to be made freely available. This is potentially a much more productive approach than the old system where the paper records, or the data from individual experiments, were kept at the host institute and thus largely unavailable to other scientists for analysis.

Permanent global seismic stations are interesting places to tour and you should enthusiastically seize any opportunity to visit one. They are typically built on massive bedrock, on which a concrete platform is laid. This platform is typically separated from the floor on which people walk by a layer of rubber ~ 1 cm thick. This reduces noise on the recorders from footsteps. Observatories often have several generations of different instruments on display, many of which have been retired. The installation of seismic stations in many developing countries by western powers, although such projects are funded by nuclear detection work, has fostered a lot of good will, brought employment and investment, and founded many international friendships.

5.3 Some seismic experiments

5.3.1 The WWSSN

The World Wide Standard Seismograph Network was installed in the early 1960s in order to monitor Russian nuclear testing. Stations were distributed as evenly as possible over the whole world. Each station comprised a three-component set of short-period seismometers (1 s pendulums, 0.7 s galvanometers) and a three-component set of long-period seismometers (15 or 30-s pendulums, 100-s period galvanometers). The system response of the sets was designed so that the system as a whole was insensitive in the 6-s microseism frequency band. Recording is analogue on paper on drums that rotate once an hour (long-period) or once per

15 mins (short-period) using photo-sensitive, heat-sensitive or ink paper. Timekeeping was done using a clock with radio clock corrections measured daily.

5.3.2 US Array

This 15-year experiment of extraordinary ambition started in the mid-2000s. A network of 70 permanent broadband stations at 300-km intervals has been deployed over the whole of the contiguous 48 states of the USA. This is supplemented by an array of 400 stations at 70-km spacings that were deployed first in a swathe extending from the Canadian border of Washington State to the Mexican border of California. This NS swathe is currently being progressively swept east by leap-frogging the westernmost stations to the eastern side. The data are stored at a data management centre in Seattle, WA, and freely available to all researchers.

5.3.3 Temporary networks

Temporary networks are commonly installed and operated for limited periods to address particular problems, *e.g.*, monitoring volcanoes or faults during periods of activity. In the 1970s stations often involved one- or rarely three-component, short-period, mechanical seismometers. Data transmission was generally done via frequency-modulated radio transmissions (analogue) or by wires stretched over the ground. Recording was analogue using tape recording. Timing used a clock and radio signals which were written onto the tape recorder.

In the 1990s more advanced equipment became available, involving digital recorders (DASs) and GPS timing. Examples are REFTEK or PDAS recorders. It has become standard to deploy three-component seismometers, often broadband instruments. Because recording is digital, the data must be transmitted digitally if it is desired to avoid frequent station visits to retrieve data. Digital radio links are currently expensive and thus on-site data recorders are commonly used.

6. Nuclear detection

6.1 From threshold test limits to a Comprehensive Test Ban Treaty

There is international pressure for a Comprehensive Test Ban Treaty (CTBT). Such a treaty bans all nuclear explosions for any reason whatsoever. A CTBT was adopted by the United Nations in 1996 but it has not yet come into force.

Advocates think that a CTBT will:

- Prevent the development of more powerful weapons (though testing is not necessary for development),
- De-emphasise the importance of nuclear weapons in national security, and;
- Reduce the discriminatory nature of the non-proliferation regime.

However, monitoring is necessary to check that everyone is complying, otherwise nations will not co-operate. Monitoring can be done in many ways, *e.g.*, by using photos from space, spying etc., but seismology is the primary tool.

6.2 Seismic monitoring of nuclear tests

6.2.1 History

The first nuclear explosion (Trinity) was conducted during World War II and was suspended above the ground. The second and third explosions ("Little Boy" and "Fat Man") were conducted above the Japanese cities Hiroshima and Nagasaki. They killed, respectively, 140,000 and ~ 50,000 civilians either instantly or after a very short time. Japan surrendered immediately after the dropping of the second bomb, not knowing that in fact this was the last bomb that the US had.

There has since been a lot of reflection on whether these bombings saved more lives than they cost. Questions include whether cities should have been targeted, whether the second bomb was necessary, and whether they would ever have been dropped on a European city. The destruction they did dwarfs the 9/11 terrorist attack that shocked the world in 2001. Even so, they were actually very small devices compared with current capabilities.

Subsequent to World War II, the USA rapidly developed its nuclear capability and a program of surface tests was conducted. In 1949 the USSR conducted its first test–an amazing achievement for a country with a much smaller population than the USA and, unlike the USA, with its infrastructure largely destroyed by the War. The first hydrogen (thermo-nuclear) bomb was detonated in 1952, another landmark step by our political leaders toward global destruction. All American nuclear testing is done at a test range in Nevada, which is interesting to look at using Google Earth.

From then, other nations began to join the race–Britain in 1956, France in 1960, China in 1964, India in 1974, Pakistan in 1998 and North Korea in 2009. It is almost certain that Israel has nuclear weapons, South Africa has been suspected, and there is currently political anxiety that Iran has ambitions to develop weapons.

A small test, "Rainier", was conducted in an underground tunnel in Nevada in 1957. This was an experiment to determine if it was possible to test underground and thus eliminate radioactive fallout. The test was positive. An additional result was the strong seismic signals observed, which resulted in seismology being adopted as a monitoring tool for the future. Since then seismology and nuclear warfare have had a symbiotic relationship.

The proportion of energy that is converted into seismic energy is called the "seismic efficiency". It is greatest for underwater explosions, smaller for underground explosions, smaller still for surface explosions and smallest for atmospheric explosions. Atmospheric tests are thus one way of avoiding detection using seismology, but they are more likely to be detected by atmospheric monitoring of radionucleides.

The Limited Test Ban Treaty of 1963 banned atmospheric, oceanic and deep space testing and restricted underground testing. There were 116 signatory nations. In the 1960s, the U.S. military funded the World Wide Standard Seismograph Network (WWSSN). 125 stations were installed in 31 countries and this contributed in a major way to the discovery of plate tectonics. The WWSSN has been run by the U.S. Geological Survey since 1973. The numbers of nuclear tests conducted up to the late 1980s is extraordinary. In 1974 the Threshold Test Ban Treaty banned underground testing of devices with yields larger than 150 Kt.

The effectiveness of imposing test bans is critically dependent on how effective monitoring is. Thus, the U.S. Dept. of Defense has funded large amounts of seismological research to:

- Improve methods to detect and locate;
- Discriminate explosions from earthquakes and industrial explosions, and;
- Estimate yield using similar methods to magnitude determination.

This helped to bring about the Threshold Test Ban Treaty (TTBT). Underground detonations of devices with yields of ~150 Kt, if tamped, produce a seismic signal equivalent to an earthquake with a magnitude of ~ 5.3. A 1 Kt explosion is equivalent to a seismic moment of ~ 4 x 10^{12} N m.

Throughout the 1970s there was debate about where the detection threshold was. Provisions in treaties are dependent on this. The future of the human race seemed to ride on how well seismic waves are transmitted through Eurasia. It is interesting to think about this in the context of the commonly-held view that 99.9% of the usefulness of seismology lies in seismic reflection, the major contribution of which is to keep the cost of petrol down.

It is easy to identify simple devices consisting of two pieces of material near the critical mass, as the minimum size of such devices is ~ 20 Kt. All first tests by nations have been of this type. A 10-Kt explosion is equivalent to ~ M 4.8. There are approximately 4 earthquakes per day of this magnitude. It is more difficult to monitor nuclear tests if nations buy bombs with yields of < 1 Kt. Without decoupling, such a bomb would be equivalent to ~ M 4.0. There are ~ 20 earthquakes per day of this magnitude. As the size of devices becomes smaller, or effective ways are found of muffling them, the task of studying all candidate seismic events rapidly becomes overwhelming.

6.2.2 Style of monitoring

Throughout the Cold War, the emphasis was on monitoring known test sites. Most of the USSR is a large landmass, and thus special arrays were placed by the USA in Scandinavia and Alaska to monitor the test sites in Novaya Zemlya and Kazakhstan. The USSR deployed sites in central Asia to monitor Nevada, the Pacifi, the Sahara (then belonging to France, now Algeria) and Lop Nor, Xinjiang (China). UK tests were done at the Nevada test site.

Now that western nations have apparently satiated their appetites for nuclear testing, there is pressure to prevent other nations from doing the same. Thus, a multilateral ban is sought, and monitoring style must change. Previously nations were watching a few known test sites only, with methods tailored to those sites. Now the whole planet must be monitored.

6.3 Arrays for monitoring

Arrays for are carefully-designed, geometric, regular networks of super-well calibrated/timed stations. Their purpose is nuclear bomb detection, and they are focused on particular sites *e.g.*, Novaya Zemlya. Initially, several small arrays were built in the USA and experience gained from these was used to design larger arrays:

LASA: Built in Montana, it cost US$6 billion in today's money. It was extraordinarily expensive because it was built using digital technology in the early 1960s. It operated from the mid-1960s to 1978, included 525 seismometers and had an aperture of 200 km. There were 21 seismometer clusters, each with 25 sensors covering 7 km^2. The sensors were vertical, high-frequency (> 3 Hz) instruments.

NORSAR: This array began operation is 1971. It contained a total of 22 subarrays with an aperture of 100 km. In 1976 this was reduced to 7 with an aperture of 50 km. Each subarray contained 5 x 1-10 Hz sensors measuring the vertical component, plus a central 3-component instrument, and was 10 km in lateral extent. The traces were summed to form a composite trace for each subarray, thus enhancing the signal-to-noise ratio and partially eliminating local effects.

Arrays were used to compute slownesses–the speed at which a wavefront sweeps across the array. This can also be used to find the azimuth of approach of a signal. The individual traces are delayed and summed and azimuth and slowness where power peaks is measured. This is known as "steering the array", but it is really a processing method, not a realignment of the network. This method can also be used to identify phases that are otherwise difficult to see, and was used by Bruce R. Julian to make the first ever identification of PKJKP.

Other arrays include:

- Tonto Forest, Arizona,
- BMD, Oregon,
- WMO, Oklahoma,
- CPO, Tennessee, and
- EKA, Eskdalemuir, Scotland.

There are other arrays in India, Canada, Brazil, Thailand and Taiwan, and 4 small arrays in Europe:

- NORESS, a subset of NORSAR,

- ARCESS, northern Norway,
- FINESSA, Finland, and
- GERESS, Germany.

Signals are recorded digitally, and there is automatic computation of direction of approach of the waves and the epicentral distances of the sources.

Array seismology is an obsolete approach now. What is now needed is monitoring of the whole world down to smaller magnitudes. A fundamental change in seismological strategy is needed for this. Worldwide networks of broadband, 3-component instruments are needed that can be used to reconstruct ground motion at a range of different frequencies. The world is now well on the way to achieving this.

6.4 How is detection done?

6.4.1 Basics

There are three main steps:

1. Detection–has an event occurred and if so where?
2. Identification–was it an explosion?
3. Yield estimation–if it was an explosion, how big was it?

Seismologists use body waves, surface waves and regional waves. A valuable phase is Lg which is a regional wave comprising multiple reflections of waves trapped in the low-velocity crustal layer. Lg is hard to interpret because detailed regional crustal structure is needed.

6.4.2 Location

Up to now, this has been the principal identification method. The vast majority of detonations occurred at known test sites whereas the vast majority of large earthquakes occur in subduction zones.

If the event occurred in the ocean, it can be identified on the basis of whether hydroacoustic signals were generated (T-waves). T-waves are waves trapped in the low-velocity layer in the interval 800-1,300 m below surface. Waves bounce to and fro in this layer (beyond critical angle) and there is little attenuation. T-arrivals may be larger than P- and S-arrivals and for explosions T may be 30 times larger than P and S. If the event occurred on land the possibility of it being an explosion may be eliminated in some cases if it located near human habitation or if there is no evidence of human activity nearby. A nuclear test is a major undertaking, and cannot be conducted without a great deal of equipment, traffic, and human activity.

6.4.3 Depth

Whereas many earthquakes occur deeper than 10 km, nuclear tests must be within drilling depth. All tests are made shallower than 2.5 km, and all tests > 150-Kt have been made shallower than 1 km. The difficulty of depth determination is a hindrance to this identification method. A conservative discriminator is to rule out the possibility of an event being an explosion if it is deeper than 15 km.

6.4.4 Aftershocks

These occur commonly after earthquakes, but uncommonly after explosions.

6.4.5 $M_S : m_b$

$M_S : m_b$ is greater for earthquakes than for explosions. This is because explosions are poor in long-period energy. M_S uses waves with periods of ~ 20 s, and m_b uses waves with periods of ~ 1 s. The source dimensions of a nuclear explosion are much smaller than those of an earthquake of same magnitude, and thus nuclear explosions are richer in high-frequency waves and poorer in long-period waves. The effect of this is often plainly seen in the seismogram.

The problem with this method is that long-period surface waves are often difficult to detect for explosions < M 4.5. Thus this method can be used for small earthquakes but not for small explosions. The method can be extended across the whole spectrum, *e.g.*, using the "variable frequency magnitude" (VFM) method, where $m_b(\text{frequency}_1)$ and $m_b(\text{frequency}_2)$ are compared.

6.4.6 P/S spectral ratio

Theoretically an explosion should generate no SH waves. However, a comparison of the earthquake in Nevada 16th August 1966 and the nuclear explosion "Greeley" shows that the SH components were very similar. This is because of "tectonic release". Thus the absence of S cannot be used to discriminate earthquakes from explosions. However, explosions are poorer generators of S waves than earthquakes and P:S for explosions is greater than for earthquakes.

6.4.7 Problems

Identification becomes difficult at smaller magnitudes because:

- There are more earthquakes;
- More industrial explosions;
- More ways to evade, and;
- A lower signal-to-noise ratio.

6.5 Discriminating between chemical and nuclear explosions.

Almost all chemical explosions are < M 4. Nations might try to muffle nuclear tests so they look like chemical explosions. Ways of discriminating nuclear from chemical explosions include:

1. Almost all chemical explosions > 0.1 Kt are really 100+ smaller explosions fired in sequence (“ripple firing”). Thus the source is spatially extended and it loses some of the features of a highly-concentrated nuclear explosion. Source multiplicity affects the spectrum and this is possible to detect seismically;

2. Chemical explosions are usually shallower than nuclear explosions, and they cause ground deformation and lack a radio-chemical signal, and;

3. If nations wanted, the problem could be dealt with by prior declaration and on-site inspection of large chemical explosions.

6.6 How successful is detection?

Global coverage varies, and is poorer in the southern hemisphere. It is much better in some places than others because of arrays that can be “steered”. Worldwide, detection is:

- reliable for explosions of 150+ Kt;
- good for 15-Kt explosions, and;
- adequate for most 1-Kt explosions.

6.7 What can be done to evade, or “spoof”?

The following have all been suggested. Which do you think are practical?

1. Test in deep space or behind the Sun.

2. Detonate a series of explosions to mimic an earthquake sequence.

3. Test during or soon after an earthquake.

4. Test in large underground cavities. This can theoretically reduce the amplitude of seismic waves by factor of 70. At the low end of the frequency range, the amplitude of seismic waves generated is approximately proportional to the volume of the new cavity. At high frequencies this is less effective. Thus this method is only partially effective. A decoupled explosion of 1 Kt (~ M 4.0) would have a reduced magnitude of only 0.015 Kt, equivalent to M ~ 2.5. There are 270 earthquakes per day of this magnitude and several thousand industrial explosions per year in the US larger than this. That problem is mitigated, however, because the latter are usually ripple fired. There are only 10-30 chemical explosions per year that produce seismic magnitudes > 3.

For full decoupling: where the cavity radius is in m and the yield in Kt:

$$cavity\ radius \simeq 20(yield)^{1/3}$$

It can be shown using this equation that the cavity sizes required make it impractical.

5. Test in non-spherical cavities.

6. Test in low-coupling material, *e.g.*, alluvium. This might be possible in Nevada but not in the USSR because of the lack of suitable geology. It would also suffer from the problem of causing a large surface crater.

7. Firing simultaneously with a large, legal, chemical explosion. This is probably the only plausible spoofing method.

6.8 Summary of the current situation

Identification capability will always be poorer than detection capability, in practice by about 0.5 magnitude units. There is a problem of numbers. There is a 10-fold increase in the number of earthquakes for every decrease in one magnitude unit. Thus the number of ambiguous events increases as size decreases and the cost of monitoring increases. Larger explosions are easy to identify.

The current state-of-the-art is that seismology can detect at the sub-Kt level for tamped explosions (~ magnitude 3 = 0.05 Kt), and at the 5-Kt level for decoupled explosions. Thus 5 Kt is the bottom line today.

The future: It is not possible to monitor a truly comprehensive test ban treaty. A low-yield threshold test ban requires monitoring programs that can give the yield, and this is becoming increasingly difficult. Furthermore, some scientists think that the greatest current nuclear threat is not governments developing weapons, but devices falling into the hands of terrorists. Some nuclear devices are no larger than a suitcase, and the event of 9/11 demonstrated that it is not necessary to command a powerful air force to deliver such devices.

Bibliography

Pasyanos, M.E. and Walter, W.R., Improvements to regional explosion identification using attenuation models of the lithosphere, *Geophys. Res. Lett.*, **36**, doi:10.1029/2009GL038505, 2009.

Richards, P.G. and J. Zavales, Seismic discrimination of nuclear explosions, Ann. Rev. Earth Planet. Sci., 18, 257-286, 1990.

Richards, P.G., Testing the nuclear test-ban treaty, Nature, 389, 781-782, 1997.

Richards, P.G., Kim, W.-Y., 2009. Monitoring for nuclear explosions. Scientific American March, 2009, 70-77.

Go to http://www.ctbto.org for up-to-date information on signatures, status of networks, and much besides.

Also, visit http://www.ldeo.columbia.edu/%7Erichards/PGRc.v.long.html – the website of Paul Richards, and look at publications nos. 126, 127, 128, 133, and 142.

Seismic tomography

6.9 Basics

Earthquake seismology can be used in many ways to determine Earth structure, but seismic tomography is particularly popular because it can provide three dimensional images of large volumes of the Earth. It involves inverting the arrival times of earthquake and explosion waves, recorded at a network, for the structure of the volume sampled by cross-cutting rays. Four main scales are potentially available:

- Local earthquake tomography, on the scale of a few 10s of km;
- Regional earthquake tomography, resolving structure on the scale of a few 10s of km, but using sources outside the network;
- Teleseismic tomography, on a scale of a few 100 km, and;
- Whole-mantle tomography, on the scale of the whole mantle.

6.10 Local-earthquake tomography

Local-earthquake tomography usually involves deploying a dense network, with station separations reflecting the scale of features sought. The seismic network should cover the area of interest, and ideally a set of earthquakes that fill the study volume will be recorded. Structure is parameterised as a grid with trilinear interpolation between nodes. In the data inversion, the earthquake locations and crustal structure are solved for separately in a series of damped least squares inversions.

Such experiments are commonly conducted in volcanic areas, often with the hope of imaging a magma chamber. This has never been reliably achieved, however, possibly because rays passing through low-velocity bodies arrive later than rays that pass around them, through higher-velocity material, or possibly because magma chambers are ephemeral features and do not exist for much of the lifetime of any volcano.

6.11 Regional earthquake tomography

The problem of regional earthquake tomography has not yet been solved. It is difficult to develop this method because structure between the sources and the network is generally not well known, cannot be calculated in the inversion, and the incoming wavefronts cannot be assumed to be planar.

6.12 Teleseismic tomography

The same problems also exist in teleseismic tomography, but are not so serious since the wavefronts may be approximated to planes because of the great distance of the sources. Typically stations are deployed at spacings of a few 10s of km and operated for up to a few years, to record a set of large teleseismic earthquakes. Network apertures are typically a few

100 km. It is a rule of thumb that good resolution may be achieved down to a depth approximately equal to the diameter of the network. As many different incoming phases as possible are picked and the data set carefully quality controlled to eliminate the worst outliers. A starting model must be chosen, and standard Earth models such as PREM or IASP91 are generally used. The volume of interest is traditionally parameterised as a stack of blocks a few 10s of km on a side, with uniform velocity in each. Inversions of both P- and S-waves separately are generally conducted.

The question of quality of the results is important. Four measures are commonly used:

- hit-count;
- resolution;
- volume metrics, and;
- ability to retrieve a theoretical structure.

The hit-count is simply the number of rays that sample a particular block, or pass near a particular node. The hit-count should be large for well-resolved regions. Resolution is measured by the diagonal elements of the resolution matrix **R**, defined as:

$$\hat{\mathbf{m}} = \mathbf{Rm}$$

where

$\hat{\mathbf{m}}$ = the "true" Earth model

$\mathbf{m}$ = the inversion Earth model result

$\mathbf{R}$ = the resolution matrix

The resolution matrix takes into account data uncertainty (measured as the *a posteriori* data variance) and geometric considerations, *e.g.*, whether all the rays formed a single, sub-parallel bundle, or whether they were well-distributed in trajectory. The mathematical description of **R** is described in detail by *Evans and Achauer* (1993). The volume metric is a measure of how smeared the anomaly is in space and is obtained from the columns of the resolution matrix.

Finally, a realistic idea of the power of a given experiment to image a particular expected structure may be obtained by erecting a theoretical model and generating a synthetic data set by calculating the delays expected if the real set of rays had passed through the theoretical structure. This synthetic data set is then inverted to see how well the theoretical model is retrieved.

Several teleseismic tomography experiments have been performed in Iceland, and the data collected have been processed by multiple groups both separately and combined using approaches of varying sophistication. A low-velocity body extending from near the surface down to the mantle transition zone underlies Iceland. All the experiments either require that the anomaly is restricted to the upper mantle, or they cannot resolve deep enough to say. The anomaly is elongated parallel to the mid-Atlantic ridge at depths of a few 100 km, suggesting influence from surface tectonic processes.

6.13 Whole-mantle tomography

Whole-mantle tomography is fundamentally the same as other kinds of tomography but the formulation is more complex. Computer programs must deal with the fact that the target volume is a hollow sphere. Core structure is not studied because it is relatively homogeneous. The outer core is liquid and thought to maintain homogeneity through rapid convection. Extremely long wavelength waves may be involved, and the Earth may be parameterised using spherical harmonics, not simple blocks. Such tomography is not able to image reliably bodies that are smaller than ~ 1,000 km and the resolution is sometimes even lower.

In recent years "finite frequency" tomography has been developed. This takes into account the fact that seismic waves do not sense only infinitesimally narrow zones along rays, but a finite volume around them. In fact, very surprisingly, mathematically, an arrival has zero sensitivity to structure precisely along the ray. The significance of this fact is well illustrated by the notorious case of the multiple ScS bounces from the M 6.3, 1973 earthquake in Hawaii. These were initially interpreted to show that the mantle has relatively high wave speeds beneath the Big Island, inconsistent with a mantle plume. When the sizes of the Fresnel zones were taken into account, it was found that this conclusion was unsafe because it became clear that the regions to which the arrivals were sensitive was too broad (~ 1,000 km) for a plume a few hundred kilometres wide to be readily detected. Furthermore, the ScS_2-ScS data that were used were rather insensitive.

Multiple whole-mantle tomography experiments using different data and approaches all confirm that the low seismic velocity anomaly beneath Iceland is confined to the upper mantle, albeit on the large spatial scale to which whole-mantle tomography is limited. A single claim that a whole-mantle plume was detected has been shown to be spurious, the image having been produced by manipulating the colour scale used and truncating the line of section to remove other, similar "plume-like" features underlying Canada and Scandinavia, where plumes are not expected.

Bibliography

Evans, J. R. and U. Achauer (1993). Teleseismic velocity tomography using the ACH method: theory and application to continental-scale studies. *Seismic Tomography: Theory and Applications*. H. M. Iyer and K. Hirahara. London, Chapman and Hall: 319-360.

Foulger, G. R., M. J. Pritchard, et al. (2001). Seismic tomography shows that upwelling beneath Iceland is confined to the upper mantle. *Geophysical Journal International* **146**: 504-530.

Iyer, H. M. and K. Hirahara, Eds. (1993). *Seismic tomography: theory and practice*. London, Chapman & Hall.

Rawlinson, N., S. Pozgay & S. Fishwick,Seismic tomography: A window into deep Earth, *Physics of the Earth and Planetary Interiors*, **178**, 101-135, 2010.

7. Earth structure

7.1 The fundamentals

A fundamentally important question in global geophysics is whether the mantle convects as one layer or as two separate layers, the upper and the lower mantle. The question is further complicated by the fact that some scientists consider the base of the upper mantle to lie at 650 km and others at ~ 1,000 km (Bullen's definition). The question of layered *vs*. whole-mantle convection is relevant to the debate regarding the existence of mantle plumes.

The theory that volcanism on Earth's surface is caused by shallow, upper-mantle processes considers vigorous convection to mostly be confined to the upper mantle and the lower mantle to be largely isolated and to have convective overturn times of the order of the age of the Earth. In this view, intraplate volcanism is a passive response to intraplate extension, which draws melt up from relatively shallow depth.

The view that some surface volcanism is attributable to mantle plumes requires diapirs of hot rock to rise from the core-mantle boundary to the surface, and for counter-flow to comprise subducting slabs that sink to the base of the lower mantle. In this view, intraplate volcanism is driven directly by heat loss from the core and the material that melted and erupted travelled up in plumes from the core-mantle boundary.

Seismology has been widely used to attempt to distinguish between these two hypotheses. However, this endeavour is fraught with problems that are generally not appreciated by non-seismologists, or simply ignored. One problem is resolution. Resolution deteriorates with depth in teleseismic tomography images and so whether structures extend down through the transition zone and into the lower mantle is difficult to determine. In general networks are not wide enough to resolve structures well at 600-700 km depth. Whole-mantle tomography resolves structure on much larger spatial scales and so the non-detection of plume-like features may not be accepted as evidence for their absence.

Whole mantle tomography is also plagued by the non-uniformity of earthquake sources and stations. This means that there are large regions of the mantle that are poorly sample by rays, or not sampled at all. In tomographic inversions, these regions will not be perturbed from their initial starting velocity values. However, in published images, it is often not clear if these regions have normal velocity or whether they were simply unilluminated by rays.

Velocities in adjacent, sampled regions are likely to be perturbed, and this produces the visual effect of a structure the shape of the ray bundle. Depending on the angle of approach of rays, such a structure may resemble a subducting slab, a tilted plume, or some other structure that may seem geologically reasonable. It is clearly critical to be able to distinguish what parts of images are reliable and what parts are not. Which apparently imaged structures are reliable and may be interpreted in terms of Earth structure and dynamics, and which are not is a subject of great debate and controversy amongst informed seismologists but typically not amongst non-seismologists.

Whole-mantle tomography shows that the structure of the upper mantle is radically different from the lower mantle. In the upper mantle wave speeds vary strongly and correlate closely with the oceans and continents to a surprisingly great depth. A certain amount of correlation occurs throughout the upper mantle. In the lower mantle, variations in velocity are much more muted and the correlation with surface continents and oceans disappears. Heterogeneity begins to increase again at ~ 2,500 km depth. The core-mantle boundary region is much more heterogeneous than the upper- and mid-parts of the lower mantle and, curiously, also correlates with surface structure to some extent.

Detailed studies of the critical region in the depth range 550-800 km show that the change in structure at the base of the upper mantle at ~ 650 km depth is sharp and strong. This conclusion is confirmed by studies of the spectral power of the anomaly field on various scales, which shows a sharp change at the base of the upper mantle.

7.2 Downwellings

The depth to which subducting slabs sink is an important question on which much rests. Seismic evidence has been quoted to support both the view that slabs do not sink below ~ 1,000 km and perhaps not even below 650 km, and the view that they founder to the core-mantle boundary and accumulate in a "slab graveyard" there. Tomographic images of the subduction zones around the Pacific rim mostly show that the angle of subduction of slabs flattens in the mantle transition zone (410-650 km) and the older parts of slabs lay down on the floor of the transition zone. It is though that they are laid down like toothpaste on a brush, as the subducting plate migrates back, because the earthquake activity that would be expected if the plate were actively thrust horizontally is not observed. The fact that deep-focus earthquakes in subduction zones do not occur deeper than ~ 700 km is further evidence in support of an upper-mantle destination for slabs. There may nevertheless be some foundering beneath the 650-km seismic discontinuity, though the evidence for this is non-unique.

The main argument in favour of whole-mantle convection is tomographic images of a global-scale high-velocity body beneath north America. This has been interpreted as a "Farallon slab", *i.e.*, a slab that subducted for a long time beneath the west coast of north America. It has variable appearance in different inversions and cross sections, depending on the line of section and the saturation level of the colour scale. In some it has the form of an elongate, sloping feature that merges with an extensive high-velocity region that occupies the bottom half of the lower mantle. In other inversions it is less slab-like in shape and seems rather to be be just a part of an extremely extensive, global-scale, high-velocity region.

7.3 Upwellings

The so-called "superplumes" are vast regions of low velocity of the order of 10,000 km broad in the lowermost mantle. One underlies the south Pacific and the other the south Atlantic. They have been widely assumed to be relatively hot and thus to be associated with thermally buoyant deep mantle plumes. It has been speculated that they may themselves be plumes, or that plumes may rise from their upper surfaces. This development has been unfortunate because it has been shown conclusively using normal modes that these bodies are not hot, but derive their low seismic velocities from their anomalous chemistry (*Trampert et al.*, 2004). Normal modes are essentially the only seismic observations that can distinguish between temperature and chemistry as the source of seismic velocity variations, because they are sensitive to density. This is something that other seismic methods cannot generally do.

Three regions postulated to be underlain by deep mantle plumes have been studied in careful detail using ambitious, large-scale teleseismic tomography experiments. The data collected have also been subjected to auxilliary supporting analyses to test other predictions of the model. These three regions are Iceland, Yellowstone and Eifel.

In all three cases low-velocity bodies have been found to underlie the volcanic fields but they are all found to be confined to the upper mantle. The anomaly beneath Iceland appears to extend deep into the transition zone and to approach the 650-km discontinuity. The anomaly beneath Eifel is mostly confined to the region above ~ 410 km. The anomaly beneath Yellowstone is strong in the upper ~ 200 km but beneath that is so weak that it is disputed whether it is resolved, or merely an artifact of smearing of the shallow anomaly along the incoming ray bundle.

The current US Array experiment is unprecedented because it will enable detailed teleseismic tomography images to be obtained that extend to much greater depths than has been possible in the past, even in the most ambitious teleseismic tomography experiments. Results so far from US Array confirm that Earth structure is highly heterogeneous in the upper mantle, in particular above the transition zone, that anomalies weaken with increasing depth throughout the transition zone, and that they are weaker still in the topmost lower mantle. In the Yellowstone region various low-velocity bodies are imaged but a classical mantle plume is not one of them. Are tomograms Rorschach tests?

How should seismic wave speeds be interpreted? The common assumption that red (low velocity) = hot and blue (high velocity) = cold is untenable. Seismic velocity depends on mineral and physical phase (solid or liquid), composition and temperature, the latter being the weakest effect. Seismology cannot directly detect melt, slabs or plumes. It can only detect seismic parameters such as velocity and the interpretation of these is ambiguous. Reasonable variations in the chemistry of mantle material, *e.g.*, the MgO/(MgO+FeO) content, can explain virtually all variations in seismic wave speeds. Because of the coarse resolution and tendency of the tomographic inversion method to produce structurally smooth results, regions whose seismic velocities derive from different effects may appear to merge into single bodies.

7.4 The transition zone

A key question is whether geological structures such as subducting slabs or hot diapirs penetrate the mantle transition zone. The bounding seismic discontinuities of the transition zone at 410- and 650-km depths are mineralogical phase changes. They are not thought to be chemical boundaries. The Clausius-Clapeyron slopes of the phase change in olivine to a wadsleite structure at 410-km depth is positive. This means that cold, sinking material will transform to the denser phase at a depth slightly shallower than 410 km, which will encourage it to sink further. The reverse will hold for hot risers. A hot riser will transform to the less-dense phase somewhat deeper than 410 km, assisting its continuing rise.

The reverse situation exists at the 650-km phase change from ringwoodite to perovskite and magnesiowüstite structure. There, the Clausius-Clapeyron slope is negative. Thus, cold sinkers are unable to transform to the denser phase until they have penetrated some distance into the dense lower mantle. Such penetration will not occur if the slab is less dense than the underlying lower mantle. Likewise, the ascent of hot risers is impeded. Hot risers will not transform to the less dense phase until they have penetrated some distance into the transition zone, and dense material cannot rise into less dense material. For these reasons, the base of the transition zone is a barrier to convection. The question is, how strong a barrier? Is it a complete barrier, or only a partial barrier? Because of this behaviour, the temperature of the mantle actually drops across the 650-km discontinuity, and this is an argument against "upper mantle" plumes that are postulated to rise from the base of the transition zone.

Because of these issues, variations in the depth of the 410- and 650-km discontinuities is a subject of intense study. Hot risers are predicted to thin the transition zone by depressing the 410-km discontinuity and elevating the 650-km discontinuity. The reverse is predicted for through-going, cold subducting slabs, which are predicted to thicken the transition zone. It has thus been suggested that the thickness of the transition zone may be used as a mantle thermometer. Thick and thin transition zones have therefore been sought beneath subduction zones and intraplate volcanic regions.

Nevertheless, this endeavour is also plagued by ambiguity. The depth to the transition-zone-bounding seismic discontinuities is predicted to vary with composition, in particular Fe and H_2O content, as well as temperature. Difficulties arise also because seismological techniques can only resolve discontinuity depths to within a few kilometres, and the errors can only be reduced to these small levels by extremely painstaking work involving stacking many tens or even hundreds of carefully selected, suitable seismograms.

The two main methods used are to study a) precursors to the main arrivals on seismograms, *e.g.*, SS, that have reflected from the discontinuities, or b) transition-zone penetrating waves that have transformed from P to S or vice versa. For the most accurate work, allowance must also be made for seismic velocity variations in the transition zone. For example, a high-velocity anomaly in the transition zone would result in a short transition-zone traverse time, which is exactly equivalent to the effect of a thin transition zone. It is thus necessary to correct for the former in order to avoid erroneous results.

This subject is still quite volatile, and new publications often report very different results from earlier papers. Recent results have reported that topography of the 410-km discontinuity is little correlated with the continents and oceans, but 650-km-discontinuity topography is strongly correlated. It is surprising that the surface of the Earth seems to have such deep influence, and this is contrary to concepts that relatively thin tectonic plates on the surface of the Earth are essentially decoupled from the mantle beneath the asthenosphere. The transition zone is generally thinner than average under the oceans, and this presents some dilemmas for interpreting thin transition zones beneath ocean islands where plumes are being sought.

7.5 Summary

Earthquake seismology is essentially the only method available to Earth scientists for probing the interior of our planet with any degree of accuracy or precision. Disciplines such as geochemistry cannot do this as they only sample surface materials and there are extremely few constraints on their depth of origin. Nevertheless, seismological observations are ambiguous in their geological interpretation and cannot be used to determine some of the parameters scientists would ideally like to determine, such as temperature.

Seismology can nevertheless give some important information that may be relied upon. The transition zone is clearly a significant barrier to convection. In the cases of the three major volcanic regions that have been studied in detail seismically, low-velocity structures that penetrate the transition zone and pass through into the lower mantle are absent. It is a challenge for the future to find ways to use the robust findings of seismology to further our understanding of Earth structure and dynamics and to shed the tendency to try to wring out of seismology information that it is intrinsically unable to constrain reliably.

Bibliography

Gu, Y. J., A. M. Dziewonski, S. Weijia et al. (2001) Models of the mantle shear velocity and discontinuities in the pattern of lateral heterogeneities. *J. geophys. Res.* **106**, 11,169-111,199.

Gu, Y. J., A. M. Dziewonski & G. Ekstrom (2003) Simultaneous inversion for mantle shear velocity and topography of transition zone discontinuities. *Geophys. J. Int.* **154**, 559-583.

Trampert, J., F. Deschamps, et al. (2004) Probabilistic tomography maps chemical heterogeneities throughout the lower mantle. *Science* **306**, 853-856.