ARE EARTHQUAKES PREDICTABLE?

Thomas H. Heaton
U.S. Geological Survey
525 S. Wilson Ave.
Pasadena, CA 91106

Introduction

When, where, and how big will the next earthquake be? Answering these questions would dramatically reduce the terror and suffering caused by earthquakes. The problem of predicting earthquakes has been actively pursued for more than a decade in many earthquake-prone countries, and numerous prediction methodologies have been suggested. Unfortunately, a prediction technique has yet to emerge that has gained general acceptance.

Earthquake prediction research has evolved in two general directions: 1) the search for precursory phenomena, and 2) attempts to recognize when and where the conditions are right for the occurrence of a significant earthquake (see Agnew and Ellsworth, 1991, for a review of recent research). Although numerous and varied precursory phenomena have been reported (usually after the occurrence of significant earthquakes), there is still great uncertainty about which quantities to measure and how to interpret them. Some suggested precursors are: changes in the rate of occurrence of small earthquakes, electromagnetic signals of several kinds, changes in water levels, changes in gases emitted from the Earth, changes in strain, animal behavior, and several others. Although I have found many of the reports of precursors to be fascinating, I have also found them to be confusing, ambiguous, and vague. Describing the plethora of prediction schemes that rely on precursory phenomena is not the purpose of my discussion here.

The primary focus of my discussion is the question of recognizing when conditions are right for the occurrence of a significant earthquake. Some questions that I will discuss are: 1) what is the stress on faults and how does it change in time? 2) what physical properties determine the strength of faults? 3) what actually happens on the fault surface during earthquakes? and 4) does the underlying physics of earthquake occurrence lead to a regular, repeating (predictable) behavior, or does it lead to a chaotic (unpredictable) behavior?

Characteristic Earthquake Hypothesis

One simple model for the occurrence of earthquakes is that stress slowly builds on a fault until a critical level of stress is reached, whereupon an earthquake occurs which decreases stress to some new level. The stress slowly in-
creases again until the critical stress is again achieved, and the earthquake repeats itself. If the change in stress (stress drop) is the same from one earthquake to another, and if the rate at which stress builds is constant, then we expect to see earthquakes that repeat over some characteristic time period called the recurrence interval. This model is often called the characteristic earthquake model and it is shown schematically in Figure 1. In this model, the earthquakes are characteristic because the same earthquake tends to repeat itself in a characteristic fashion. In practice, it is usually assumed that the stress drop during the earthquake is proportional to the fault slip and that the rate of stress buildup is proportional to the long-term geologic slip rate on the fault. If we know the geologic slip rate on the fault, the size of the characteristic slip during earthquakes, and the time of the last earthquake, then we can use this model to estimate the time of the next earthquake. This model has been used recently to estimate the probabilities of earthquakes on the major faults in California (Working Group on California Earthquake Probabilities, 1988 and 1990).

Unfortunately, it has been extremely difficult to test the characteristic earthquake model; we need to observe many repeats of earthquakes on a particular fault segment to see if they are indeed similar earthquakes and to see whether the timing is actually regular. The problem is that the repeat time for major earthquakes is long compared with the short history of instrumental recording of earthquakes. Most studies of earthquake occurrence rely on either pre-instrumental historical descriptions of damage patterns, or on geologic studies of when prehistoric slip occurred on a fault segment. Unfortunately, many questionable assumptions must be made when interpreting either of these types of data. As an example of the type of problem encountered, consider the M 7.1 1989 Loma Prieta earthquake that occurred on a strand of the San Andreas fault system in central California. Although this was a fairly major earthquake (one we would like to predict), there was no observed rupture at the Earth's surface; all of the fault slip occurred at depths between 5 and 18 kilometers. Therefore, there is no geologic recording of this earthquake at the Earth's surface. Furthermore, there is evidence that there was significant rupture at the Earth's surface during the M 7.9 1906 San Francisco earthquake, the previous rupture of this region of the San Andreas fault system. This raises the question of whether or not earthquakes along this stretch of the fault are in fact characteristic.

This dilemma has caused much debate within the seismological community. Ironically, the Working Group on California Earthquake Probabilities (1988) had used the characteristic earthquake model to estimate that this section of the San Andreas fault (the Santa Cruz mountains segment) was the most likely location for a M 6.5 earthquake. Thus, while the occurrence of the Loma Prieta earthquake in 1989
was seemingly a great success for the characteristic earthquake model, the character of the actual earthquake was quite surprising. It raised many questions about the validity of the characteristic earthquake model for this stretch of the fault.

Another important application of the characteristic earthquake model is the Parkfield Prediction Experiment (Bakun and Lindh, 1985). Moderate sized earthquakes (M 5.5 to 6.0) occurred in the Cholame Valley (apparently on the San Andreas fault) in 1966, 1934, 1922, 1901, and 1881. Furthermore, the Cholame Valley was apparently the northern terminus of the great 1857 rupture (M 8) of the San Andreas. In the hours before the 1857 earthquake there were three moderate sized earthquakes in the Cholame Valley. This sequence of historical earthquakes has been very important for the formulation of characteristic model. Assuming that each of these dates (including 1857) represent characteristic earthquakes, we conclude that the average recurrence time is 22 years. Since the last earthquake was in 1966, considerable resources have been concentrated in the Parkfield region in an attempt to predict the next earthquake.

The Parkfield earthquake is now almost 4 years beyond it's mean recurrence interval, giving plenty of time for scientists to raise serious questions about the validity of the Parkfield Prediction Experiment and the overall validity of the characteristic earthquake model. For instance, the little data available from the 1934 earthquake (Segal and others, 1990) suggests that significantly different stretches of the fault broke in 1934 and 1966 (the 1966 earthquake was relatively well recorded). Of course, very little is known about the character of the earlier historic events.

These problems with anticipating specific earthquakes on active faults illustrate why it is so important to formulate a deeper understanding of the physics of the earthquake process. It will be many generations before we have collected enough reliable data to observe the repetition of rupture along given fault strands.

The Stress Dilemma

One argument for the regular repeat of earthquakes goes like this, "tremendous stresses build in the Earth before a major earthquake, which then releases tremendous kinetic energy as waves radiated from the earthquake. Many years must pass before enough stress re-accumulates, and once it does, this tremendous energy must be released within some relatively short time window." There is a big problem with this line of reasoning. In actuality, many researchers have been struggling to understand why the shear stress on many active faults is so low and why the energy released in an earthquake is so little.

When almost any type of rock is confined at the pressures that must exist 5 to 15 km in the Earth (1.5 to 5
kilobars), it can withstand shear stresses of at least one half of the confining pressure before it yields. Since faults obviously yield during earthquakes, we expect that the shear stress must be on the order of 1 to 2 kilobars. However, it is clear that large earthquakes are not a process where faults decrease stress on a fault by a kilobar; and it's a good thing, because if they did, nothing would survive the experience. A kilobar stress drop on a circular fault rupture with a 15 km diameter would produce a slip of 100 meters and ground velocities of 10 meters/sec (about 25 MPH). If the your building was strong enough to survive this motion (a bomb shelter), then certainly you would be killed by a wall running into you at 10 meters/sec. The point is, that although earthquakes seem violent to humans, they actually involve stress changes that are small (typically about 15 bars) compared with the confining stress at the depth of earthquakes.

Another indication that the stress change in earthquakes is relatively small comes from the observation of strain buildup on major faults. For instance, the rate at which shear strain builds on the San Andreas fault is about $2 \times 10^{-7}$ per year, or less than 1 bar per decade (Savage, 1983). Average recurrence intervals of 100 to 200 years for large earthquakes are compatible with the notion that the stress drop is relatively small.

Although we have been able to establish that the average stress drop in earthquakes is relatively small, it has been more difficult to estimate the actual shear stress in the Earth's crust. Perhaps the shear stress is in the kilobar range, but only a very small fraction of the total stress is released in an earthquake. Estimating the shear stress level has been the subject of much work. For instance, there are indications that the shear stress is as large as a kilobar beneath large mountain ranges (Jeffreys, 1959); that is, the mountains would collapse under their own weight unless large shear stresses exist. Thus, at least some parts of the Earth's crust have strengths close to those expected from laboratory experiments. On the other hand, measurement of stress in deep holes such as the 3.5-km deep hole adjacent to the San Andreas fault in Cajon Pass, California, indicate that the shear stress on the San Andreas fault is relatively low, probably less than several hundred bars (Zoback and others, 1987).

The production of heat from frictional sliding also provides another important constraint on the shear stress. If the friction during sliding is more than several hundred bars, then large enough quantities of heat should be produced on active faults that we should clearly be able to observe high heat flow through the crust adjacent to major active faults. Despite extensive research on this problem, no such heat flow anomalies have been detected. This observation almost certainly implies that the friction is less than several hundred bars during slip on faults (Brune and others, 1969; Lachenbruch, 1980).
To summarize the stress dilemma, the yield stress of rocks at earthquake depths is inferred to be in the kilobar range from laboratory measurements, the stress drop in earthquakes averages about 15 bars, the shear stress on major active faults is less than several hundred bars, and the friction during sliding is less than several hundred bars. Clearly there is a problem. From the preceding remarks, it is impossible that stress on the San Andreas fault steadily builds until a yield stress of more than a kilobar is reached. Why are faults so weak? In laboratory experiments, we have yet to observe yield in rocks at such low shear stresses and such high confining pressures.

There are currently two candidate solutions to the stress dilemma: 1) the effective confining pressure on rocks in fault zones is dramatically reduced by the presence of fluids under very high pressure, and 2) the physics of dynamic rupture are such that earthquakes can occur at relatively low stress. As I will discuss, both of these mechanisms seriously complicate the simple picture assumed for the characteristic earthquake model.

The Role of Fluids

Extremely high fluid pressures are required to reduce the strength of the San Andreas fault to only several hundred bars. The fluid pressure would have to be larger than the pressure from a fluid column extending to earthquakes depths (the hydrostatic pressure); it would need to be comparable to the pressure exerted by the weight of the overlying rocks (lithostatic pressure). Furthermore, there is strong evidence that the shear stress on the San Andreas fault plane is less than the shear stress on adjacent faults with different orientations. If the fluid pressure was homogeneously high in a region, then the strength of all planes should be low. It has therefore been hypothesized that there is a zone of very high fluid pressure concentrated only along some active faults such as the San Andreas (Rice, 1991). Unfortunately, this model requires a rather fortuitous distribution of porosities in order to confine such high pressures to only the fault zone.

The direct examination fault zones that have been uplifted and then deeply eroded provides important evidence that fluids are present in fault zones; veins of hydrothermally deposited minerals are often observed (Sibson and others, 1988). Since it is clear that buoyancy precludes the possibility of downward migration of fluids, it is hypothesized that fluids migrate upwards from a source deep in the crust, probably due to the metamorphism of hydrated minerals. It may be that fluids migrate upward through veins in the fault zone rather rapidly since the pressure gradients are inferred to be extremely large. Upward migration of fluids may trigger the occurrence of earthquakes. If this is the case, it would be necessary to understand the
plumbing of a fault zone in order to predict the occurrence of earthquakes. Understanding the role of fluids in fault zones is an important problem for which we currently have few certain answers.

The Role of Rupture Dynamics

Up to now, I have limited the discussion of earthquake occurrence to the accumulation of stress and the yield stress of faults. However, an earthquake is by its very nature a dynamic phenomenon, and there is far more to the physics than simply initiating rupture. Let me turn the prediction problem around slightly. We currently record an average of 40 earthquakes per day in southern California, and there is every reason to believe that there are many more that are too small for us to record. My question is not when these earthquakes will occur (they happen all the time), but instead, which of these will be a big earthquake? In order to answer this, we need to understand rupture dynamics.

From many observations of waves radiated by earthquakes, it is clear that they usually begin relatively abruptly. That is, rupture initiates at some point and then spreads over the fault surface at speeds slightly less than the shear-wave velocity (typically about 2.8 km/sec). It is also clear that big earthquakes are ones in which the rupture area is large. Thus predicting which of the many small earthquakes will be a big one is a matter of predicting which rupture will propagate a large distance (Brune, 1979).

Figure 2 shows a schematic of a rupture at time t propagating steadily along a fault in the x-direction with a rupture velocity \( V_r \). The slip \( D(x - V_r t) \) and shear stress \( \sigma(x - V_r t) \) are given as a function of time and distance along the fault plane. The shear stress on the fault plane is large in the region directly ahead of the rupture front. The assumption that the continuum behaves as a linear elastic solid leads to solutions that have a square root singularity in the stress at the tip of the propagating rupture (usually called a crack tip) (Freund, 1979). Although the stress in real materials must remain finite, it is quite easy to generate the kilobar stress levels at the crack tip that are necessary to cause the rocks to yield. That is, once a crack is propagating, the stress at which a material begins to yield is rather irrelevant.

The dynamic friction is the key parameter for deciding whether or not a rupture will continue to propagate. As long as the elastic energy released by sliding exceeds the work necessary to slide the faults surfaces past each other, the rupture will continue to propagate. For a given point on the fault, this can be stated mathematically as,

\[
\int_{0}^{\infty} \Delta \sigma(t) D(t) \, dt > 0,
\]
where $\omega(t)$ is the dynamic stress drop on the fault which is defined to be the stress on the fault before the earthquake began minus the frictional stress on the fault plane. As long as this energy integral is positive, the rupture will continue to radiate energy, and when the integral is negative the rupture will absorb energy and will die out.

Very little is known about the friction while two rock surfaces are sliding at high speeds (greater than a meter per second) and at high confining pressure. Because no excess heat from frictional sliding is observed on the San Andreas fault, we know that this sliding frictional stress is low, less than several hundred bars. Several mechanisms may make this stress low. If the frictional stress is high (1 to 2 kilobars) as sliding begins, then tremendous heat would instantly be generated. After a mere centimeter of slip, the fault surface would melt, thereby dramatically reducing the friction (Richards, 1976). Although there is evidence from exhumed faults that melting sometimes occurs, it appears to be a relatively rare phenomenon. If fluids are present in the vicinity of the rupture, they would suddenly heat and expand (Sibson, 1973; Lachenbruch, 1980). This would rapidly decrease the effective confining stress on the fault, thereby decreasing the dynamic friction. If the grinding of the fault surfaces generates intense acoustic waves (a loud noise), then this would also decrease the effective confining stress on the fault (Melosh, 1979). Given these considerations, I believe that it is not surprising that ruptures can propagate with relatively low driving shear stresses.

Another interesting clue to the problem of dynamic friction comes from the tire industry. Several decades ago researchers were puzzled about the relatively low stress required for frictional slip of rubber. By viewing rubber slipping on a transparent piece of glass, Schallamach (1971) discovered that the rubber was not directly sliding on the glass surface. Instead, linear wrinkles propagated over the slip surface with net slip occurring between the leading and trailing edge of the wrinkle. This mechanism allows the rubber to slip at much lower stresses than would be inferred for the smooth frictional slip of rubber over glass. Theoretical investigations of this behavior lead to the discovery of a new class of dislocation waves that can propagate over a surface (Comninou and Dundurs, 1987). These pulses of slip are similar to a wrinkle in a rug. Carpet installers have long realized that it is much easier to move a rug by introducing a wrinkle than it is to slide the entire surface. The theoretical calculations for slip in a continuum show that if a pulse of slip, with low frictional stress is introduced into a medium, it will tend to propagate. These theoretical solutions have only recently been recognized by seismologists, and we are currently studying their applicability to the Earth (Brune and others 1991).

I have discussed mechanisms for low dynamic friction that all imply that the sliding friction is controlled by
slip on the fault. On the other hand, slip on the fault is controlled by the tectonic stress and the sliding friction. This is a feedback system that is likely to be unstable. That is, dynamic friction and the slip may be expected to vary (perhaps chaotically) as the rupture propagates down the fault.

Propagation of Pulses of Slip

Figure 3 shows the distribution of slip (in centimeters) as a function of position on the rupture surfaces of the 1984 Morgan Hill (M 6.2) and 1979 Imperial Valley (M 6.5) earthquakes (Hartzell and Heaton, 1983 and 1985). These slip models were constructed to produce the seismic waves that were recorded at strong-motion seismometers located close to these earthquakes. The slip history on the fault surface is actually modeled as a pulse of slip that propagates down the fault plane at velocities just below the shear wave velocity. The shaded regions in these plots are the regions that slip at a particular time. In these models, only a small part of the total rupture surface is slipping at any given time. These slip models that were derived from waves radiated by earthquakes have characteristics that are very similar to the theoretical solutions for propagating pulses of slip.

In order to produce these propagating slip pulses, it is necessary that dynamic friction on the fault is low near the crack front, and it must also increase away from the crack front, causing the rupture to heal itself (Heaton, 1990). Notice that the spatial distribution of slip is also quite heterogeneous. Such a heterogeneous distribution of slip would either require a very heterogeneous distribution of material properties on the fault, or it could be produced by variations in the dynamic friction. As was just mentioned, such dynamic variations in dynamic friction are to be expected.

The heterogeneous nature of the slip distributions shown in Figure 3 also poses major problems for the characteristic earthquake model. If these earthquakes were to repeat themselves over several recurrence cycles, then large slips would accumulate in some regions, but other regions would have very little slip. This is clearly an untenable situation. It suggests that future earthquakes will have different slip distributions that have large slips in regions of small slip during past earthquakes. In this case, the earthquakes would not be characteristic.

Implications for Earthquake Prediction

I have shown that dynamic variations in stress during earthquakes are large compared with the changes in stress that accumulate slowly over decades. I have also discussed why understanding the nature of dynamic friction is critical
to understanding when a large earthquake will occur. In Figure 4, I show a schematic guess of what the stress on a fault plane might look like over a period of hundreds of years. Stress usually builds slowly on the fault (about 1 bar per decade for the San Andreas). The occurrence of earthquakes on adjacent fault segments may suddenly increase the stress, but by less than several tens of bars. However, if an earthquake rupture initiates at some location (perhaps from a local concentration of stress, or perhaps because of the upward migration of fluids), then the rupture front produces very large dynamic stress variations that can fracture even very strong parts of the fault. The actual propagation of the rupture is then controlled by the dynamic friction, which is inferred to be relatively low (probably less than several hundred bars). Predicting the time of a large earthquake requires that we must be able to predict the nature of this dynamic friction. Although there is much that is unknown about dynamic friction, we can guess that it is has interesting and complex physics.

According to these models, very long and very short recurrence intervals may be observed for any section of the fault. The key to knowing when a given section will rupture is the ability to predict the nature of dynamic rupture. This may be a very difficult task (perhaps impossible). In any case, it is clear that there is much to learn about the physics of earthquake rupture. What we learn will have a profound influence on our ability to predict when, where, and how large future earthquakes will be.

References


Figure Captions

Figure 1. Simple model of earthquake recurrence. Similar earthquakes repeat over regular time intervals when stress steadily builds to some critical level.

Figure 2. Model of a propagating rupture. Rupture travels in the x-direction with a velocity $V_r$ (typically 2.5 to 3.0 km/sec). Fault zone initially yields due to very large stresses which concentrate just forward of the leading edge of the rupture. Rupture will continue to grow as long as the energy from the release of elastic strain is larger than the energy required for frictional sliding. $\Delta\sigma$ is the dynamic stress drop which is defined as the stress before initiation of the rupture minus the stress on the sliding fault plane (the frictional stress). $\Delta\sigma$ is the final stress change after the completion of the earthquake.

Figure 3. Models of the distribution of slip (cm) on the fault planes of the 1984 Morgan Hill earthquake (M 6.2) and the 1979 Imperial Valley earthquake (M 6.5). These models are constructed to reproduce ground motions recorded within several tens of kilometers of these earthquakes. In the models, rupture actually is confined to a narrow zone of rupture (the stippled regions) that propagates down the fault. The size of slip in the pulse of rupture grows and shrinks as the pulse propagates down the fault. Note how uneven the slip is on the fault surface. Will future earthquakes have a similar distribution of slip, or will they have large slips in regions of small slip in the last event?

Figure 4. Model of shear stress for a given point on a fault when dynamic rupture is considered. This model is to be contrasted with Figure 1 which does not consider the importance of dynamic rupture stress. Stress usually increases slowly (about one bar per decade). Occasional nearby earthquakes may cause the stress to suddenly increase by several bars. When a large rupture propagates through the point being considered, large dynamic stress variations occur on the time scale of seconds. Occasionally, small slip may occur in a region adjacent to much larger slips and the stress may actually increase at a particular point as the result of an earthquake. As in Figure 3, dynamic friction plays a vital role in deciding when a rupture will propagate over a large fault area. In this model, the spatial distribution of slip is heterogeneous and varies from one event to another. The timing of events in this model may also be irregular.
REPEATING CHARACTERISTIC EARTHQUAKE

- Large earthquakes occur when Critical Stress is reached
- Large earthquake occurs only if rupture propagates large distances
- If \( \int_0^\infty \Delta \sigma_D D(t) \, dt > 0 \) rupture can propagate
Heterogeneous Slip

- Slip is spatially uniform on long time scales (thousands of years)
- Either there is creep between events,
- or, Slip distribution (and stress change) varies from event to event
HETEROGENEOUS-SLIP, DYNAMIC RUPTURE MODEL

Slip (m)

Shear Stress (bars)

Time (years)

Adjacent event
Positive stress drop

Fig. A.