changes in mantle minerals (Section 5.4.2), which could produce slip only once on a fault surface, in contrast to frictional sliding, which can recur.

4.7.3 *Earthquake probabilities*

A natural use of earthquake statistics is to estimate the probability of future earthquakes. These probabilities are interesting from the standpoint of earthquake physics, and crucial for attempts to forecast the hazards due to large, damaging earthquakes (Section 1.2.5).

The challenge of estimating earthquake probabilities can be illustrated by a simple analogy. Problems in probability are often couched as games of chance, but earthquakes have the special feature that the game's rules are unknown. To see this, consider estimating the probability that particular playing cards will be dealt from a deck. If the game begins with a full deck, there is a 25% (13/52) chance of drawing a spade, an 8% (4/52) chance of an ace, and a 2% (1/52) chance of the ace of spades. These chances are analogous to the prospects of having a magnitude 6, 7, or 8 earthquake in a year. As play continues, there are several possible cases. If the deck is shuffled after every draw, the probabilities do not change. Alternatively, if the deck is not shuffled, the probabilities change depending on the cards that have been drawn. For example, if no aces have yet appeared, the probability of an ace increases with each draw. However, if cards are dealt from under the table, we do not know what cards the deck began with (there may be no aces or eight of them) and whether it is shuffled. We must infer what the deck contains, how it is shuffled, and what cards will appear, with no information except the cards already drawn. Hence if no aces have appeared after a large number of draws, the probability of an ace may be high (because the remaining cards contain several) or low (because the starting deck had few).

In the nomenclature of probability theory, the probability of events depends on the *probability density distribution* that is sampled and the sampling method. For earthquakes, we know neither because we do not have a theoretical model that successfully describes earthquake recurrence, so we adopt probability distributions based on the earthquake history which for most faults is short (only a few recurrences) and complicated. As a result, various distributions grossly consistent with the limited history are used and can produce quite different estimates.

The simplest model describes earthquake occurrence by a *Poisson distribution* often used to describe rare events.[4] We assume that the probability of n large earthquakes in an area or on a fault during time t is

$$p(n, t, \tau) = (t/\tau)^n e^{-t/\tau}/n!, \qquad (8)$$

where $1/\tau$ is the number expected in a year from the regional Gutenberg–Richter distribution or some variant, so τ is the mean recurrence time. The probability of one or more earthquakes is found from the probability that none will happen, using the certainty ($p = 1$) that an earthquake either will or will not happen, so

$$p(n \geq 1, t, \tau) = 1 - p(0, t, \tau) = 1 - e^{-t/\tau} \approx t/\tau, \qquad (9)$$

where the last step used the Taylor series expansion $e^x \approx 1 - x$, and so is valid for $t \ll \tau$. In this model, the probability that an earthquake will occur in an interval of time t starting from now does not depend on when "now" is, because a Poisson process has no "memory." On average, earthquakes are separated by time τ, but when the last earthquake occurred has no effect.

[4] Examples include volcanic eruptions, radioactive decay, and the number of Prussian soldiers killed by their horses.

The Poisson model is the simplest null hypothesis against which we can compare other models. However, its time-independence in which earthquakes are implicitly random events is not appealing, because almost all of our seismological instincts favor earthquake cycle models, in which strain builds up slowly from one major earthquake to the next.[5] In this case, the probability of a large earthquake should be small immediately after a large earthquake, and then grow with time. This is described by time-dependent models in which the probability of a large earthquake a time t after the past one is given by a probability density distribution $p(t, \tau, \sigma)$ that depends on the average and variability of the recurrence times, described by the mean τ and the standard deviation σ. In other words, p gives the probability that the recurrence time for this earthquake will be t, given an assumed distribution of recurrence times. The cumulative probability that the earthquake will occur by time T since the past earthquake is found by integrating the density function

$$P(T) = \int_0^T p(t, \tau, \sigma)dt. \qquad (10)$$

We seek to estimate how likely an earthquake is between now and some future time. Formally, this is the conditional probability that the earthquake will occur between time T_0 (now) and a future time T, given the condition that it has not yet happened by time T_0. To do this, we use *Bayes's theorem*, which states that $P(A|B)$, the conditional probability of event A given that event B has occurred, is the ratio of the joint probability $P(A, B)$ of both A and B to $P(B)$, the probability of event B:

$$P(A|B) = P(A, B)/P(B). \qquad (11)$$

In this case, the conditional probability $C(T, T_0)$ that the earthquake will occur between T_0 and T is the ratio of the probability that it will occur in that interval to the probability that it has not yet happened by T_0, which is just 1 minus the probability that it has. Hence

$$C(T, T_0) = (P(T) - P(T_0))/(1 - P(T_0)). \qquad (12)$$

The denominator is less than one, so the conditional probability is greater than the joint probability (numerator) because the fact that the earthquake has not happened makes it more likely.

This approach can be used with any assumed probability density function. The simplest is to assume that earthquake recurrence follows the familiar *Gaussian* or *normal* (bell curve) distribution (Section 6.5.1)

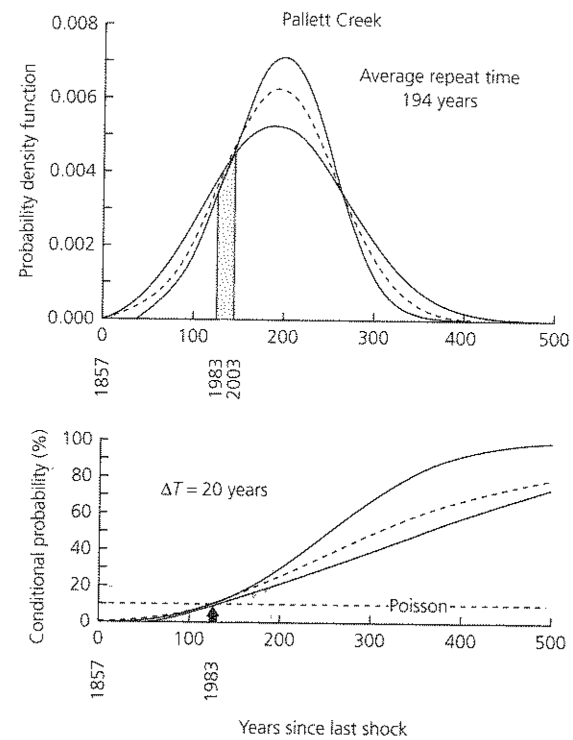

Fig. 4.7-9 Earthquake probability estimate for a segment of the San Andreas fault on which the last major earthquake occurred in 1857. *Top*: Probability density functions, with the interval 1983–2003 shaded. The dashed line is for a Gaussian distribution, with mean and standard deviations of 194 and 58 years, and the solid lines are for an alternative (Weibull) distribution. *Bottom*: Conditional probability that the next large earthquake will occur in the next 20 years, as a function of time since 1857. As of 1983 (arrow), the probabilities for the time-dependent models were comparable to those for a time-independent Poisson model. (Sykes and Nishenko, 1984. *J. Geophys. Res.*, *89*, 5905–27, copyright by the American Geophysical Union.)

$$p(t, \tau, \sigma) = \frac{1}{\sigma\sqrt{2\pi}} \exp\left[\frac{-1}{2}\left(\frac{t-\tau}{\sigma}\right)^2\right]. \qquad (13)$$

This distribution is often described using the normalized variable $z = (t - \tau)/\sigma$ describing how far, in terms of the standard deviation, t is from its mean.

Figure 4.7-9 shows such an analysis for the segment of the San Andreas fault including the Pallett Creek site (Fig. 1.2-15), on which the last major earthquake was the 1857 Fort Tejon earthquake. The analysis uses a Gaussian distribution with a mean and standard deviation of 194 and 58 years, corresponding to the most recent five major earthquakes. The upper panel shows the probability density function for this distribution (dashed line) and two others. These are used to estimate the conditional probability that a major earthquake would occur between 1983 (the study time) and 2003. These times are 126 and 146 years since 1857, and so correspond to normalized

[5] Of course, these instincts favoring determinism may ultimately prove incorrect — Einstein initially rejected quantum mechanics, arguing that "God does not play dice."

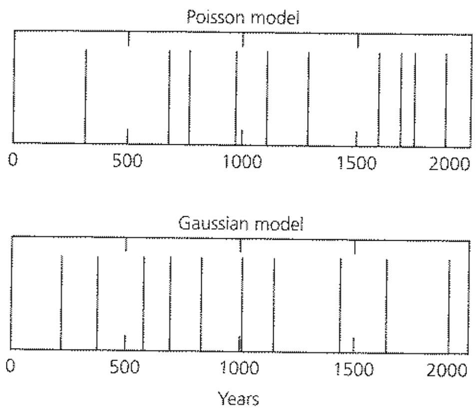

Fig. 4.7-10 Synthetic earthquake histories computed by sampling a Poisson model with the recurrence time of 194 years and a Gaussian model with this recurrence time and standard deviation 58 years. The Gaussian model yields a more periodic series, whereas the Poisson model yields clustering.

times of −1.17 and −0.83, with probabilities of 0.12 and 0.20. Thus the conditional probability (Eqn 12) is

$$C(2003, 1983) = (P(2003) - P(1983))/(1 - P(1983))$$
$$= (0.20 - 0.12)/(1.0 - 0.12) = 0.09, \qquad (14)$$

or 9%. The probability for successive 20-year intervals increases with time, and so is 29% if the earthquake has not occurred by 2057, and 56% if it has not occurred by 2157.

It is interesting to compare these time-dependent probabilities to those predicted by the time-independent Poisson model. For an assumed mean recurrence time of 194 years, the probability in 20 years is 10%. Thus for times since the previous earthquake less than about 2/3 of the assumed recurrence interval, the Poisson model predicts higher probabilities. At about 2/3 of the interval, in this case about 1986, the models predict comparable probabilities. At later times the Gaussian model predicts progressively greater probabilities. This comparison illustrates the seismic gap concept: a gap exists when it has been long enough since the last major earthquake that time-dependent models predict an earthquake probability much higher than expected from time-independent models.

The differences between the models can be illustrated by comparing the earthquake histories that each predicts. Figure 4.7-10 shows synthetic earthquake histories generated by randomly sampling probability distributions with the parameters used in Fig. 4.7-9. In the simulation, both models yield ten earthquakes after an earthquake at time zero. The earthquakes from the Poisson model have a mean recurrence of 189 years and a standard deviation of 107 years, whereas those for the Gaussian model have a mean and standard deviation of 191 and 58 years, respectively. The difference results from the fact that the Poisson process is time-independent, so there are both shorter and longer intervals between earthquakes than for the Gaussian process, which is more regular. The Poisson process thus shows clustered earthquakes resulting from the random sampling. In the limit of very long histories, the Poisson process has a standard deviation of recurrence intervals equal to its mean. Thus a recurrence history with standard deviation close to the mean favors a Poisson process, whereas a standard deviation significantly smaller than the mean suggests a Gaussian or other time-dependent process. How to interpret the limited earthquake histories available is an interesting question, as illustrated by this simple example with ten recurrences, which is longer than usually available.

These examples bear out that estimates of earthquake probabilities depend significantly on both the probability distribution used and the parameters for that distribution, which are generally not well constrained by observations. For example, the analysis in Fig. 4.7-9 used a Gaussian distribution with a mean and standard deviation of 194 and 58 years, corresponding to the most recent five major earthquakes at Pallett Creek. Alternatively, the past ten earthquakes there yield a recurrence with a mean and standard deviation of 132 and 105 years (Section 1.2.5). Other probability distributions give different probability estimates, as illustrated by the curves in Fig. 4.7-9 corresponding to Poisson and Weibull distributions. Similarly, different estimates would result from using a log-normal distribution in which the natural logarithm of recurrence time is normally distributed, so recurrence intervals longer than the mean are more likely than shorter ones.

Hence earthquake forecasts are easy to make, but hard to test. Because the estimates must be tested using data that were not used to derive them, hundreds or thousands of years

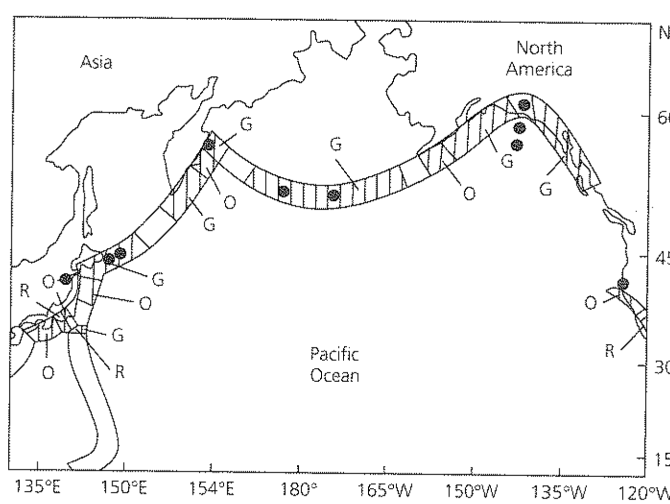

Fig. 4.7-11 Portion of the seismic gap map (McCann *et al.*, 1979) used by Kagan and Jackson (1991) to test the gap hypothesis. The shaded segments of the plate boundaries had been assigned seismic potentials of high (red, R), intermediate (orange, O), and low (green, G). Unshaded segments were regarded as having uncertain potential. During the ten years following the map's publication, ten large (M > 7) earthquakes (dots) occurred in these regions. None were in the high- or intermediate-risk segments, and five were in the low-risk segments. (Stein, 1992. Reproduced with permission from *Nature*.)

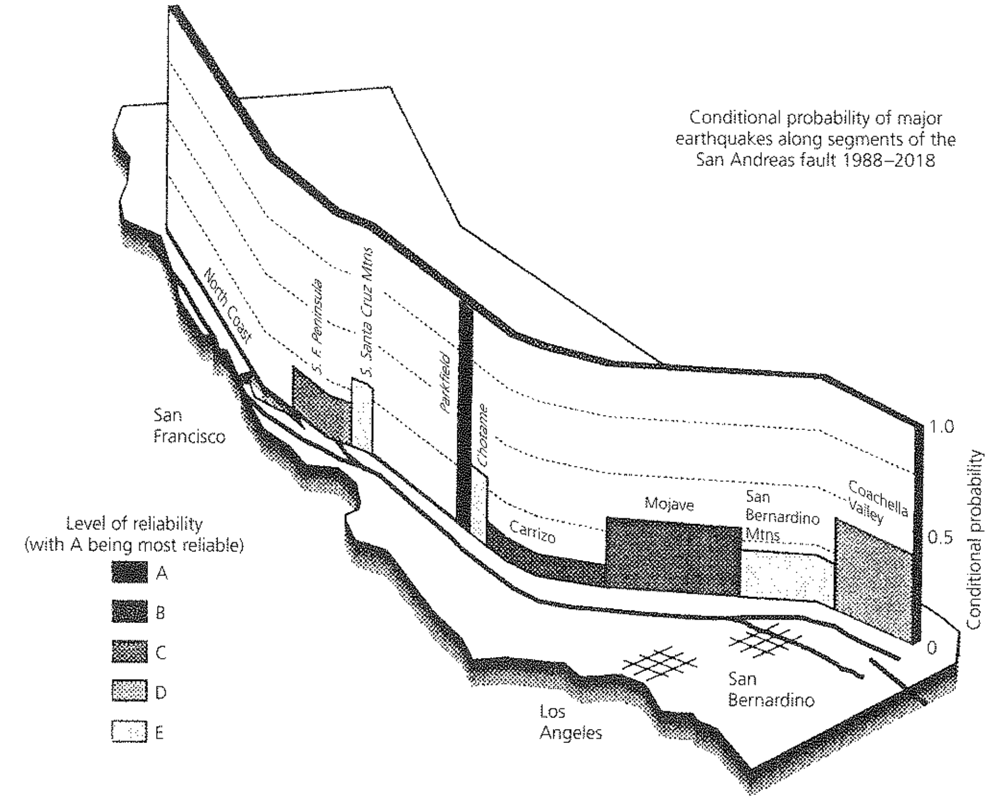

Fig. 4.7-12 Conditional probabilities of major earthquakes estimated for segments of the San Andreas fault for the period 1988–2018. (Agnew *et al.*, 1988. Courtesy of the US Geological Survey.)

(multiple recurrences) will be needed to assess how well various models predict large earthquakes on specific faults or fault segments. The first challenge is to show that a model predicts future earthquakes significantly better than the simple time-independent Poissonian model.

Given human impatience, attempts have been made to conduct alternative tests using smaller earthquakes or many faults over a short time interval. To date, the results are not encouraging. As discussed in Section 1.2.5, the history of relatively small (M 5–6) earthquakes near Parkfield, California, was used in 1985 to predict at 95% confidence level that the next one would occur by 1993, whereas the earthquake has not materialized to date (2002). Presumably the earthquake will occur eventually, although its conditional probability seems to have been overestimated and might even be assumed to be decreasing, because the longer the earthquake is delayed, the longer the mean recurrence interval inferred from the earthquake history becomes.[6] Moreover, a global test of the seismic gap hypothesis, which examined how well a gap map (Fig. 4.7-11) forecast the locations of major earthquakes, found that the map did no better than random guessing. In fact, many more large earthquakes occurred in areas identified as low risk than in the presumed higher-risk gaps. This result, which appears inconsistent with ideas of earthquake cycles and seismic gaps, has led to various interpretations, including that the gap model applies only to the largest events that break major portions of the plate boundary.

Perhaps the most sophisticated large-scale earthquake probability studies have been in California. Figure 4.7-12 shows conditional probabilities estimated along segments of the San Andreas fault. Such models can also include factors such as variable slip in earthquakes and stress changes due to nearby earthquakes (Section 5.7). Testing more complicated models with more adjustable parameters, however, will be even more challenging and take even longer.

Hence, at present, estimates of earthquake probabilities have large uncertainties. For example, using the complex Pallett Creek earthquake series (Fig. 1.2-15), in 1989 the range of probabilities for a major earthquake before 2019 was estimated as about 7–51%.[7] Thus it has been suggested that it is only meaningful to quote probabilities in broad ranges, such as low (<10%), intermediate (10–90%), or high (>90%).[8] However, despite these formidable difficulties, estimation of earthquake probabilities seems certain to remain an active research area. If some probability model is ultimately demonstrated to be reasonably successful, its use could advance efforts to estimate earthquake hazards.

[6] This situation, discussed by Davis *et al.* (1989), has been likened to waiting for a bus – the longer the bus fails to arrive, the less likely its arrival seems. A homework problem illustrates these issues.

[7] Sieh *et al.* (1989).

[8] Savage (1991).