Bulletin of the Seismological Society of America, Vol. 96, No. 6, pp. 1995–2007, December 2006, doi: 10.1785/0120050183

Simulation-Based Distributions of Earthquake Recurrence Times on the San Andreas Fault System

by Gleb Yakovlev, Donald L. Turcotte, John B. Rundle, and Paul B. Rundle

Abstract Earthquakes on a specified fault (or fault segment) with magnitudes greater than a specified value have a statistical distribution of recurrence times. The mean recurrence time can be related to the rate of strain accumulation and the strength of the fault. Very few faults have a recorded history of earthquakes that define a distribution well. For hazard assessment, in general, a statistical distribution of recurrence times is assumed along with parameter values. Assumed distributions include the Weibull (stretched exponential) distribution, the lognormal distribution, and the Brownian passage-time (inverse Gaussian) distribution. The distribution of earthquake waiting times is the conditional probability that an earthquake will occur at a time in the future if it has not occurred for a specified time in the past. The distribution of waiting times is very sensitive to the distribution of recurrence times. An exponential distribution of recurrence times is Poissonian, so there is no memory of the last event. The distribution of recurrence times must be thinner than the exponential if the mean waiting time is to decrease as the time since the last earthquake increases. Neither the lognormal or the Brownian passage time distribution satisfies this requirement. We use the "Virtual California" model for earthquake occurrence on the San Andreas fault system to produce a synthetic distribution of earthquake recurrence times on various faults in the San Andreas system. We find that the synthetic data are well represented by Weibull distributions. We also show that the Weibull distribution follows from both damage mechanics and statistical physics.

Introduction

The purpose of this article is to study the recurrence-time statistics of earthquakes in California. To do this we present the results of a 10^6-yr simulation of earthquake occurrence on a specified set of interacting faults in California. This simulation, "Virtual California," gives the statistical distribution of recurrence times between earthquakes on the faults and fault segments. An essential question is whether these simulation results can be used to quantify probabilistic earthquake-hazard assessments in California.

Clearly, the primary seismogenic fault in California is the San Andreas. Great earthquakes occur on the northern section, that is, 1906, and on the southern section, that is, 1857. The simplest model for the behavior of major faults is based on the hypothesis that a fault, or fault segment, experiences a periodic repetition of "characteristic" earthquakes (Youngs and Coppersmith, 1985). These ideas have been quantified by the slip-predictable and time-predictable models of earthquake occurrence (Shimazaki and Nakata, 1980). It is now well established, however, that earthquakes rarely occur without some aperiodicity. In addition, neither the slip-predictable nor the time-predictable models for earthquake recurrence have been successful. According to available evidence there is considerable variability in both the recurrence times and in the magnitudes of characteristic earthquakes. This variability can be attributed to the interactions between faults and fault segments. One example is the role of "stress shadows" following an earthquake or volcanic inflation (Rundle and Whitcomb, 1984; Harris, 2003). A relevant question is whether the reduced number of $m \approx 6$ earthquakes in northern California since 1906 can be attributed to the stress shadow of that earthquake.

Despite the extreme complexity of seismicity in California, well-known scaling laws are applicable (Turcotte, 1997). One example is the Gutenberg-Richter frequency-magnitude scaling. Another is Omori's law for the temporal decay of aftershocks. The question to be addressed in this article is whether recurrence-time statistics of earthquakes also satisfy a universal scaling law. Ideally, this question would be answered by using actual data. However, observed sequences of earthquakes on faults are so small that is it not possible to establish statistical patterns. For this reason we utilize numerical simulations of earthquake occurrence. With our simulation using Virtual California, we generate thousands of earthquakes on each fault segment considered.

Several statistical distributions have been proposed for the recurrence times between earthquakes. In this article we show good agreement between the Weibull distribution and our simulations. For this reason we discuss the Weibull distribution in some detail, but we also considered the lognormal and the Brownian passage time distributions and compare our results with the three distributions.

Distributions

It is standard practice to relate recurrence times between earthquakes on a fault or fault segment to statistical distributions (Matthews *et al.*, 2002). Once such a distribution has been specified the "waiting times" to future earthquakes can also be obtained. The basic statistical distribution that we consider is the statistical distribution of recurrence times t between earthquakes with magnitudes greater than a specified value on a fault or fault segment. The relevant probability distribution function (PDF) for recurrence times $p(t)$ is given by

$$p(t) = \frac{1}{N_{\mathrm{eqt}}} \frac{\delta N_{\mathrm{eq}}}{\delta t} \tag{1}$$

where t is the time measured forward from the time of the last earthquake, δN_{eq} is the number of earthquakes that occur in the interval between t and $t + \delta t$, and N_{eqt} is the total number of earthquakes considered. The PDF defined in equation (1) satisfies the normalization condition $\int_0^\infty p(t)\,dt = 1$. The corresponding cumulative distribution function (CDF) is given by

$$P(t) = \int_0^t p(t')\,dt' \tag{2}$$

This is the fraction of the recurrence times that is shorter than t.

A second statistical distribution that is important is the distribution of waiting times Δt for an earthquake on a fault if the time since the last earthquake is t_0. The waiting time is measured forward from the present, thus $t = t_0 + \Delta t$. We first consider the hazard function $h(t_0)$. This is the probability that an earthquake will occur at the time t_0 if it has not previously occurred and it is given by

$$h(t_0) = \frac{p(t_0)}{1 - P(t_0)}. \tag{3}$$

The conditional PDF that an earthquake will occur between times Δt and $\Delta t + \delta t$ if the last earthquake occurred at a time t_0 in the past can be written

$$p(\Delta t, t_0) = \frac{p(t_0 + \Delta t)}{1 - P(t_0)}. \tag{4}$$

The corresponding conditional CDF is the probability that an earthquake will occur in a waiting time Δt if the time since the last earthquake is t_0. This is given by

$$P(\Delta t, t_0) = \frac{P(t_0 + \Delta t) - P(t_0)}{1 - P(t_0)} = 1 - \frac{1 - P(t_0 + \Delta t)}{1 - P(t_0)}. \tag{5}$$

This is the fraction of the waiting times that is shorter than Δt.

The approach is best illustrated by using a specific distribution. For this purpose we choose the Weibull distribution, which is one of the most widely used lifetime distributions in a wide range of engineering applications (Weibull, 1951; Meeker and Escobar, 1991). An example is the distribution of times to failure of fibers. The Weibull distribution has also been widely used for specifying the distribution of earthquake recurrence times (Rikitake, 1982). The PDF for a Weibull distribution of recurrence times is given by (Patel *et al.*, 1976)

$$p(t) = \frac{\beta}{\tau}\left(\frac{t}{\tau}\right)^{\beta-1} \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right], \tag{6}$$

where β and τ are fitting parameters. The mean μ, standard deviation σ, and the coefficient of variation C_v of the Weibull distribution are given by

$$\mu = \tau\Gamma(1 + 1/\beta) \tag{7}$$

$$\sigma = \tau\left\{\Gamma(1 + 2/\beta) - [\Gamma(1 + 1/\beta)]^2\right\}^{1/2} \tag{8}$$

$$C_v = \frac{\sigma}{\mu} = \left\{\frac{\Gamma(1 + 2/\beta)}{[\Gamma(1 + 1/\beta)]^2} - 1\right\}^{1/2} \tag{9}$$

where Γ is the gamma function of x.

The CDF for a Weibull distribution from equations (2) and (6) is given by

$$P(t) = 1 - \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right]. \tag{10}$$

The hazard function for the Weibull distribution from equations (3), (6), and (10) is

$$h(t_0) = \frac{\beta}{\tau}\left(\frac{t_0}{\tau}\right)^{\beta-1}. \tag{11}$$

If $\beta > 1$ the probability that an earthquake will occur increases as a power of the time t_0 since the last earthquake. This is the most important scaling property of the Weibull distribution. Substituting equation (10) into equation (5) we

find that the conditional CDF for the Weibull distribution is given by

$$P(\Delta t, t_0) = 1 - \exp\left[\left(\frac{t_0}{\tau}\right)^{\beta} - \left(\frac{t_0 + \Delta t}{\tau}\right)^{\beta}\right]. \quad (12)$$

This is the probability that an earthquake will occur in a waiting time Δt if the time since the last earthquake is t_0. The waiting time corresponding to a probability of 0.5, $P(\Delta t, t_0) = 0.5$, is given by

$$\Delta t_{1/2} = \tau \left[\left(\frac{t_0}{\tau}\right)^{\beta} + \ln 2\right]^{1/\beta} - t_0. \quad (13)$$

This is the median waiting time. A further discussion of the Weibull distribution and other distributions used for recurrence-time statistics is given in Appendix A.

Many authors have applied the Weibull distribution to distributions of recurrence times between earthquakes. One of the earliest was Hagiwara (1974). Rikitake (1976, 1982) applied the Weibull distribution to the observed recurrence times of great earthquakes at six subduction zones. The derived values of the coefficient of variation C_v were 0.10, 0.22, 0.29, 0.33, 0.41, and 0.65. The corresponding values of the power-law exponent β are 12, 5.2, 3.9, 3.3, 2.6, and 1.6. Utsu (1984) considered both the recurrence times and waiting times for the Weibull, gamma, lognormal, and exponential distributions.

Rikitake (1991) fit the Weibull distribution to great earthquakes in the Tokyo area and obtained $C_v = 0.722$ ($\beta = 1.40$). Paleoseismicity studies at Pallett Creek, California, by Sieh *et al.* (1989) showed that the intervals between great southern California earthquakes on the San Andreas fault have the values 44, 63, 67, 134, 200, 246, and 332 yr. These values give $C_v = 0.645$. These authors fit a Weibull distribution to these data and found that $\tau = 166.1 \pm 44.5$ yr and $\beta = 1.50 \pm 0.80$. Substitution of these values into equation (11) taking $t_0 = 149$ yr since the last great earthquake in 1857 gives $h(149 \text{ yr}) = 0.0086 \text{ yr}^{-1}$, the estimated risk of such an earthquake in the next year is about 1%. Biasi *et al.* (2002) have also quantified the paleoseismic recurrence statistics at the Pallett Creek site as well as at the Wrightwood site on the San Andreas fault. For the Pallett Creek site they find $C_v = 0.594$ and for the Wrightwood site, $C_v = 0.532$. They fit their data to the exponential and lognormal distributions.

Authors have used several distributions to fit observed earthquake recurrence-time statistics. Ogata (1999) has considered in some detail the lognormal, gamma, exponential, and doubly exponential distributions as well as the Weibull distribution. Of these the lognormal distribution has probably found the widest use. Arguments in its favor have been presented by Nishenko and Buland (1987). It was also used in three formal assessments of future earthquake probabilities in California (Working Group, 1988, 1990, 1995).

Savage (1991) presented an extensive criticism of the use of the lognormal distribution. Davis *et al.* (1989) considered the lognormal distribution of waiting times for the next earthquake and pointed out the paradox that the longer it has been since the last earthquake, the longer the mean waiting time until the next earthquake. Sornette and Knopoff (1997) have also considered the paradox that for many statistical distributions that have been proposed for recurrence times, the mean waiting time for the next earthquake increases with an increase in the time since the last earthquake. They point out that this increase is associated with a thick tail of the distribution. The transition is the exponential distribution associated with a Poisonian distribution of intervals between earthquakes on a fault. The Poisonian distribution has no memory and the mean waiting time to the next event does not depend on the interval since the last event. Distributions that have "fatter" tails than the exponential distribution have an increase in mean waiting times with an increase in the time since the last event; an example is the lognormal distribution. Distributions that have "thinner" tails than the exponential have a decrease in mean waiting times with an increase in the time since the last event; an example is the Weibull distribution with an exponent β greater than one. The increasing risk of an earthquake with increasing t_0 is also quantified by the hazard function defined in equation (3). From equation (11) it is seen that the hazard function for the Weibull distribution will increase with increasing t_0 if $\beta > 1$.

Another statistical distribution that has been used to forecast earthquake recurrence and waiting times is the Brownian passage-time distribution (Matthews *et al.*, 2002; Ogata, 2002). This distribution is also known as an inverse Gaussian distribution. The motivation for the use of this distribution is that it models a steady loading with Brownian walk perturbations. This distribution was used in the most recent assessment of the earthquake hazard in the San Francisco area (Working Group, 2003). The distribution functions associated with both the lognormal and the Brownian passage-time distributions are given in Appendix A.

We have discussed three alternative statistical distributions for the recurrence times between earthquakes on a fault. These are the Weibull, the lognormal, and the Brownian passage-time distributions. Ideally, it should be possible to select a preferred distribution based on observed sequences of earthquakes on one or more faults. However, Savage (1994) and Stein and Newman (2004) have argued convincingly that actual sequences of earthquakes on specified faults are not long enough to establish the statistics of recurrence times with any reliability. Goes (1996) has approached this problem by superimposing sequences from many faults, but the fact that the coefficients of variation for faults are variable make this approach questionable. In this article we consider the use of numerical simulations to establish the statistics of recurrence times on faults.

Simulations

Simulation-based approaches to forecasting and prediction of natural phenomena have been used with considerable success for weather and climate. When carried out on a global scale, these simulations are referred to as general circulation models (Covey *et al.*, 2003; Kodera *et al.*, 2003). Turbulent phenomena are represented by parameterizations of the fluid dynamics, and the equations are typically solved over spatial grids having length scales of tens to hundreds of kilometers. Although even simple forms of the fluid dynamics equations are known to display chaotic behavior (Lorenz, 1963), general circulation models have repeatedly shown their value. In many cases ensemble forecasts are carried out; simulations are made by using several models to test the robustness of the forecasts.

Ideally similar simulations could be carried out to assess the earthquake hazard, but tectonic models involving finite strain are not feasible at this time. Such models would require both block motions (displacements on faults) and continuum deformation (granulation). It is not possible to evolve block motions because of geometrical incompatibilities; there will be either overlaps or gaps.

A more limited simulation model for distributed seismicity on the San Andreas and adjacent faults in southern California was given by Rundle (1988). This model included stress accumulation and release as well as stress interactions between the San Andreas and adjacent faults. An updated version of the Rundle (1988) model has been developed (Rundle *et al.*, 2001, 2002, 2004, 2005, 2006). This version, called Virtual California, includes the major strike-slip faults in all of California. The model is based on a set of mapped faults with estimated slip rates, a prescribed plate tectonic motion, earthquakes on all faults, and elastic interactions. The faults in the model are those that have been active in recent geologic history. Earthquake activity data and slip rates on these model faults are obtained from geologic databases. At present, Virtual California includes only strike-slip faults, which are responsible for a large fraction of the seismic-moment release.

Virtual California is a three-dimensional earthquake simulation code that includes the static interactions of stress between faults utilizing dislocation theory. Stress accumulation is by means of "backslip;" a linearly increasing displacement is applied to each fault at a prescribed rate. A model earthquake occurs on a fault when the increasing stress adjacent to the fault exceeds the prescribed static coefficient of friction on the fault. This is a quasi steady-state model in that faults do not grow or die. The backslip model produces a cyclic variation of stress on each fault. But because of the interactions between the faults the behavior of each fault is chaotic rather than cyclic. There is no long-term increase in either strain or stress.

The topology of Virtual California is shown in Figure 1 superimposed on a LandSat image. The 650 fault segments are approximately 10 km in length along strike with a 15-km depth. The major strike-slip faults are included and are shown as dark lines. Further details of the Virtual California model are given in Appendix B.

A similar model, standard physical earth model (SPEM), was developed by Ward (1992) and applied to seismicity associated with subduction at the Middle America trench. This model was further developed and applied to the entire San Andreas system by Goes and Ward (1994), to the San Andreas system in southern California by Ward (1996), and to the San Andreas system in northern California by Ward (2000).

Although the statistics of the simulated earthquakes produced by SPEM are similar to those produced by Virtual California, several differences between the models exist. Whereas Virtual California involves rectangular fault elements in an elastic half-space, SPEM is a plane strain (or plane stress) computation in an elastic plate of thickness H. In both codes, the dip of the fault is considered to be vertical. In addition, very different treatments of friction are included.

In this article we utilize the Virtual California model to obtain the statistical distributions of recurrence and waiting time for simulated large earthquakes on specified faults and fault segments. The simulation begins after the removal of an initial transient. We advance our model in 1-yr increments, and simulate 1,000,000 yr of earthquakes on the entire San Andreas fault system. We note that although the average slip on the fault segments and the average recurrence intervals are tuned to match the observed averages, the variability in the simulations is a result of the fault interactions. Slip events in the simulations display highly complex behavior, with no obvious regularities or predictability. Because of the size of the elements considered, earthquakes with magnitudes less than about 5.5 are not obtained. The frequency-magnitude statistics of the larger earthquakes on all faults satisfy Gutenberg-Richter statistics to a reasonably good approximation. Further modeling results have been given by Rundle *et al.* (2004, 2005, 2006).

As specific examples we give our results for the northern San Andreas fault and the Calaveras fault in some detail. We first consider earthquakes on the northern San Andreas fault with magnitudes greater than m 7.5. We consider all earthquakes from San Jose to the northwestern termination of the fault. This includes 46 fault segments with a total length of about 460 km. We consider only earthquakes larger than m 7.5 because these are the dominant "characteristic" earthquakes on this section of the San Andreas fault and they are of primary concern. A magnitude m 7.5 earthquake includes 15 to 25 fault segments (lengths of 150 to 250 km). The PDF of recurrence times on the northern San Andreas fault for all 4606 earthquakes with magnitude greater than m 7.5 that occur on this fault is given in Figure 2. In making this plot we have used 20-year bins. The mean recurrence time is $\mu = 217$ yr, the standard deviation is $\sigma = 114.7$ yr, and the coefficient of variation is $C_v = 0.528$. Also included in this figure are the corresponding PDFs for the Weibull, lognormal, and Brownian passage-time distributions from

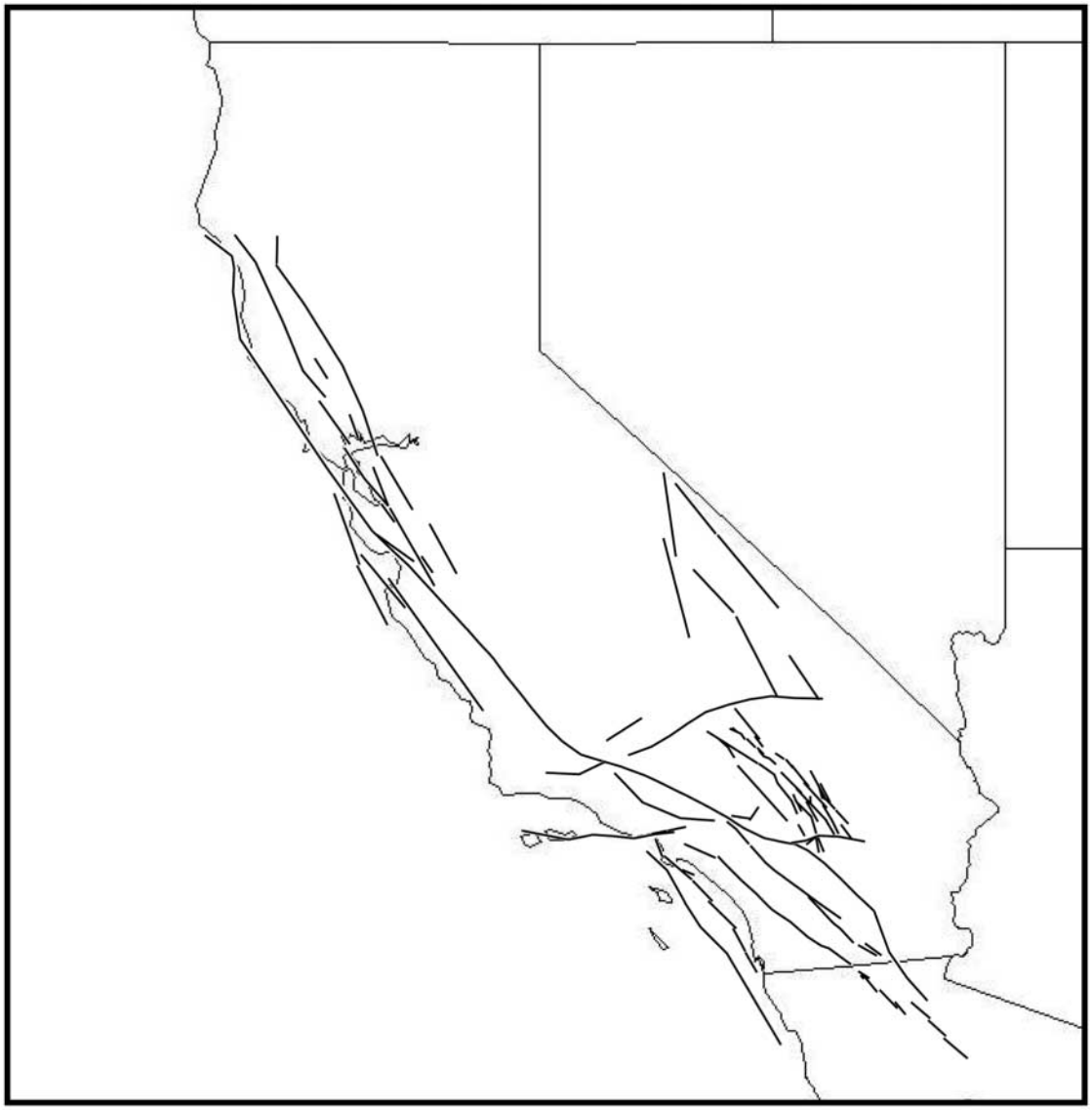

Figure 1. Faults segments making up Virtual California. The model has 650 fault segments, each approximately 10 km in length along strike and 15 km in depth.

equations (6), (A.1), and (A.7). In each case the model distributions have the same mean μ and standard deviation σ as the simulation results. It is clear that the Weibull distribution with $\tau = 245$ yr and $\beta = 1.976$ is in better agreement with the simulation results than either the lognormal or Brownian passage-time distributions. Using the preceding values or τ and β the hazard function for the Weibull distribution given in equation (11) can be determined. Taking $t_0 = 100$ yr since the great San Francisco earthquake gives $h(100 \text{ yr}) = 0.00364 \text{ yr}^{-1}$. This is the simulation forecast that a $m \geq 7.5$ earthquake will occur in the San Francisco region in the next year.

We next quantify the goodness of fit of the three distributions to the simulation as shown in Figure 2. To do this we use the Kolmogorov-Smirnov test (Press *et al.*, 1995). To use this test it is necessary to compare a cumulative distribution obtained from a simulation and a trial distribution. We compare each of the three theoretical cumulative distributions, Weibull, lognormal, and Brownian passage time, with the cumulative simulation data given as a PDF in Figure 2. First, the maximum value of the absolute difference between the two cumulative distributions D is determined. We find that $D = 0.0238$ for the Weibull distribution, $D = 0.0847$ for the lognormal distribution, and $D = 0.0865$ for the Brownian passage-time distribution. The corresponding values of the significance level Q are then determined. We find $Q = 0.010$ for the Weibull distribution, $Q = 3.3 \times 10^{-29}$ for the lognormal distribution, and $Q = 1.7 \times 10^{-30}$ for the Brownian passage-time distribution. Press *et al.* (1995) state: "It is not uncommon to deem acceptable values to be $Q > 0.001$." Thus our value for the Weibull distribution can be considered acceptable. Press *et al.* (1995) also

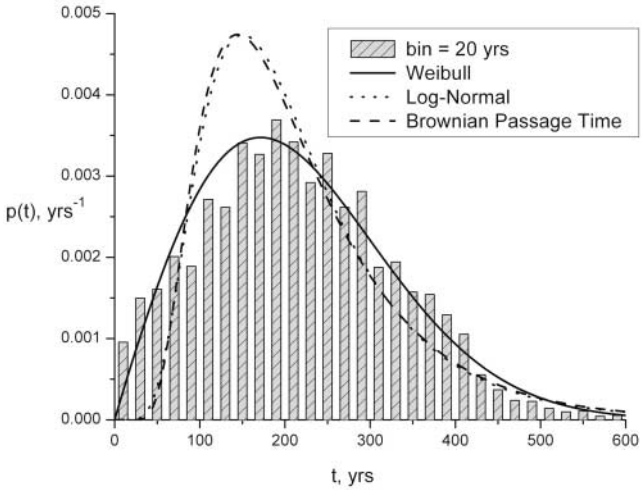

Figure 2. The PDF $p(t)$ of recurrence times t for 4606 simulated earthquakes on the northern San Andreas fault with $m > 7.5$. Also shown are the corresponding PDFs for the Weibull, lognormal, and Brownian passage-time distributions.

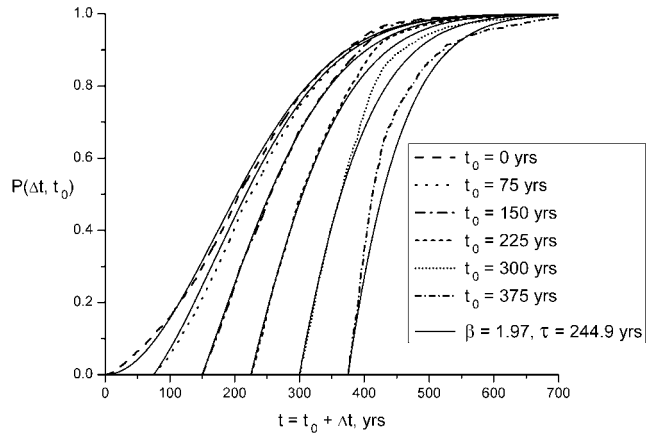

Figure 3. The conditional CDF $P(\Delta t, t_0)$ that a simulated $m > 7.5$ earthquake on the northern San Andreas fault will occur in a waiting time Δt in the future if the last such earthquake occurred t_0 years in the past. Results are given for several values of t_0. Also included are the conditional CDFs of waiting times for the Weibull distribution from equation (12).

state: "Truly wrong models are rejected with vastly smaller values of Q, say 10^{-18}." Clearly the lognormal and Brownian passage-time distributions fall into this category. The fit of the lognormal and Brownian passage-time distributions to the peak of the simulation data can be improved if a higher mean value is taken, but in this case there will be large misfits for the tail of the distributions. The Weibull distribution is the only one of the three that fits both the peak and the tail of the simulation data.

The corresponding conditional CDFs that an earthquake would occur in a waiting time Δt in the future if the last earthquake occurred t_0 years in the past is given in Figure 3. The result for $t_0 = 0$ is the CDF corresponding to the PDF given in Figure 2. Results are also given for $t_0 = 75$, 150, 225, 300, and 375 yr. In each case we have removed the recurrence times that are less than or equal to t_0 and have plotted the cumulative distribution of the remaining recurrence times $t > t_0$. Also included in Figure 3 is the conditional CDF of waiting times for the Weibull distribution from equation (12) using the same fitting parameters β and τ as in Figure 2.

In Figure 4 we give the dependence of the median waiting times $\Delta t_{1/2}$ to the next $m > 7.5$ earthquake on the northern San Andreas fault as a function of the time t_0 since the last $m > 7.5$ earthquake. These correspond to the values of Δt with $P(\Delta t, t_0) = 0.5$ in Figure 3. Also included are the median waiting times for the Weibull, lognormal, and Brownian passage-time distributions. Again we see that the Weibull prediction from equation (12) is in better agreement with the simulations than either the lognormal or the Brownian passage-time distributions. It is a characteristic of the lognormal distribution that the mean waiting time decreases and then increases. It is a characteristic of the Brownian passage-time distribution that the waiting time asymptotically approaches

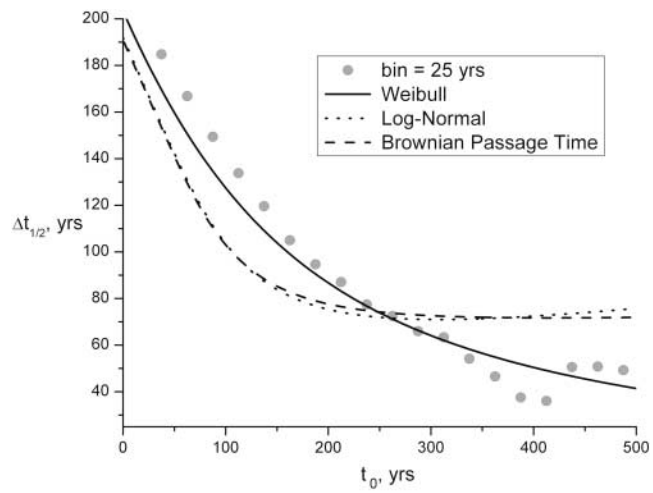

Figure 4. The median waiting time $\Delta t_{1/2}$ to the next $m > 7.5$ earthquake on the northern San Andreas fault as a function of the time t_0 since the last $m > 7.5$ earthquake. Also included are the median waiting times predicted by the Weibull, lognormal, and Brownian passage-time distributions.

a constant value corresponding to the exponential distribution. Neither of these results can be considered to be acceptable on a fault subjected to an increasing stress. On the other hand, the mean waiting time for a Weibull distribution systematically decreases with increasing t_0 in agreement with the simulations. This argument supporting the use of the Weibull distribution was previously given by Sieh *et al.* (1989).

Our simulations for the northern San Andreas fault with m_c 7.5 have a coefficient of variation $C_\nu = 0.528$. Goes and Ward (1994) carried out a 100,000-yr simulation of earth-

quakes on the San Andreas system and found that C_v = 0.50–0.55 for the northern section of the San Andreas fault for earthquakes with $m > 7.5$. The two simulations differ in many ways with different faults considered, different frictional and mean slip velocities, and different numerical approaches. Yet the variability of recurrence times as quantified by the coefficients of variation are very similar. We believe it is reasonable to conclude that this variability is robust and is not sensitive to details of fault-interaction simulations. Fault interactions lead naturally to Weibull distributions of recurrence times.

As our second specific example we consider the Calaveras fault in northern California. This fault includes 15 segments with a total length of about 150 km. Earthquakes on this fault are significantly smaller than earthquakes on the northern San Andreas fault considered previously. We consider only earthquakes larger than m 6.8 because these are the dominant characteristic earthquakes on this fault. A magnitude m 6.8 earthquake typically includes two to three fault segments (lengths of 20 to 30 km). The PDFs of recurrence times on this fault for all 8174 earthquakes with magnitudes greater than m 6.8 that occur on this fault are given in Figure 5. The mean recurrence time is $\mu = 122$ yr the standard deviation is $\sigma = 87.4$ yr, and the coefficient of variation is $C_v = 0.714$. Also included are the corresponding PDFs for the Weibull, lognormal, and Brownian passage-time distributions from equations (6), (A.1), and (A.7). In each case the model distributions have the same mean μ and standard deviation σ as the simulation results. Again it is clear that the Weibull distribution with $\tau = 135$ yr and $\beta = 1.42$ is in better agreement with the simulation results than either the lognormal or the Brownian passage-time distributions.

Once again we utilize the Kolmogorov-Smirnov test to quantify the goodness of fit. We find that $D = 0.0236$ and $Q = 0.00020$ for the Weibull distribution, $D = 0.0824$ and $Q = 9 \times 10^{-49}$ for the lognormal distribution, and $D = 0.0843$ and $Q = 4.9 \times 10^{-51}$ for the Brownian passage-time distribution. Clearly, the Weibull distribution has the best fit, although it is not as good as the fit to the northern San Andreas data.

The corresponding conditional CDFs that an earthquake would occur in a return time Δt in the future if the last earthquake occurred t_0 years in the past are given in Figure 6. The result for $t_0 = 0$ is the CDF corresponding to the PDF given in Figure 5. Results are again given for $t_0 = 75$, 150, 225, 300, and 375 yr. Also included in Figure 6 is the conditional CDF of waiting times for the Weibull distribution from equation (12) using the same fitting parameters $\tau = 134.5$ yr and $\beta = 1.42$ as in Figure 5.

In Figure 7 we give the dependence of the median waiting times $\Delta t_{1/2}$ to the next $m > 6.8$ earthquake on the Calaveras fault as a function of the time t_0 since the last $m > 6.8$ earthquake. These correspond to the values of Δt with $P(\Delta t, t_0) = 0.5$ in Figure 6. Also included are the median waiting times for the Weibull, lognormal, and Brownian passage-time distributions. Again we see that the Weibull prediction

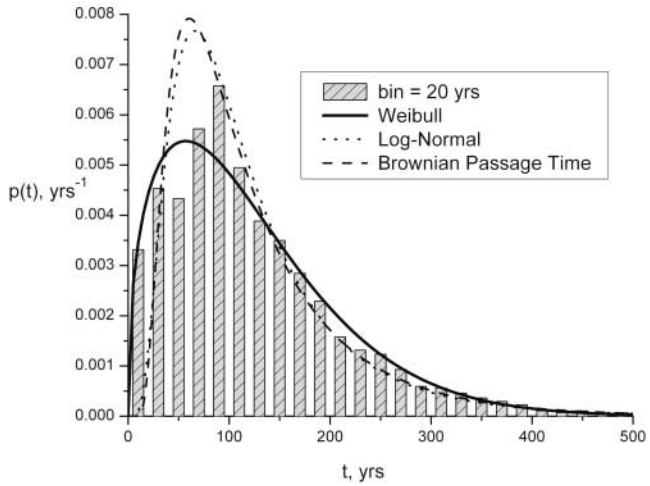

Figure 5. The PDF $p(t)$ of recurrence times t for 8174 simulated earthquakes on the Calaveras fault with $m > 6.8$. Also shown are the corresponding PDFs for the Weibull, lognormal, and Brownian passage-time distributions.

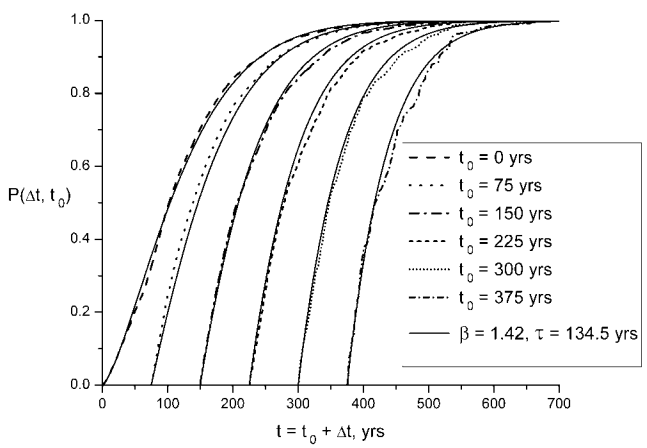

Figure 6. The conditional CDF $P(\Delta t, t_0)$ that a simulated $m > 6.8$ earthquake on the Calaveras fault will occur in a waiting time Δt in the future if the last such earthquake occurred t_0 years in the past. Results are given for several values of t_0. Also included are the conditional CDFs of waiting times for the Weibull distribution from equation (12).

from equation (12) is in better agreement with the simulations than either the lognormal or the Brownian passage-time distributions.

The results for the two examples given previously are also given in Table 1. In addition, results are given for the southern San Andreas, Hayward, San Gabriel, and San Jacinto faults. The recurrence times for the northern and southern San Andreas faults have very similar statistical properties. The coefficients of variation are $C_v = 0.530$ and $C_v = 0.556$, respectively. The secondary faults have somewhat larger coefficients of variation ranging from $C_v = 0.60$ for the Hayward fault to $C_v = 0.762$ for the San Gabriel fault.

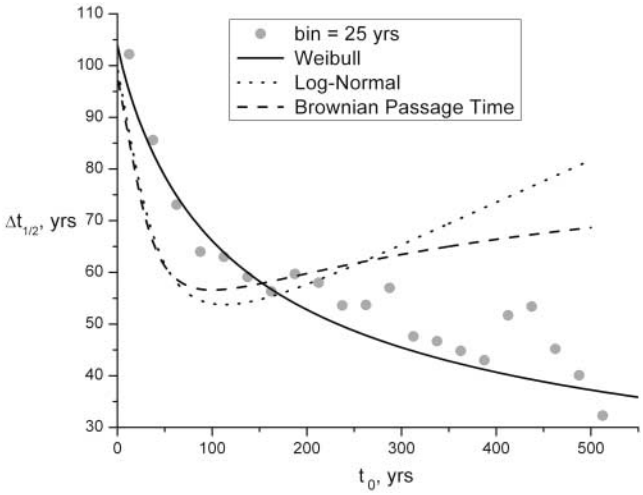

Figure 7. The median waiting time $\Delta t_{1/2}$ to the next $m > 6.8$ earthquake on the Calaveras fault as a function of the time t_0 since the last $m > 6.8$ earthquake. Also included are the median waiting times predicted by the Weibull, lognormal, and Brownian passage-time distributions.

All values deviate substantially from the value $C_v = 1.0$ expected for the exponential distribution of random recurrence times. Although tending toward the periodic value $C_v = 0$, all faults exhibit considerable variability.

Also included in Table 1 are two observational examples from the San Andreas fault. The classic example of characteristic (quasi periodic) earthquakes is the sequence of m 6+ earthquakes on the Parkfield section between 1857 and 2004. The seven recurrence times have a mean $\mu = 24.5$ yr, a standard deviation $\sigma = 9.25$ yr, and coefficient of variation $C_v = 0.38$. Although smaller than the values we have obtained for the northern and southern sections of the faults, the value is indicative of considerable aperiodicity that we would attribute to interactions with adjacent faults.

Paleoseismic studies of paleoearthquakes on the southern San Andreas fault (Sieh *et al.*, 1989) also gives seven recurrence times of m 7+ earthquakes at Pallett Creek. These recurrence times have a mean $\mu = 155$ yr, and a coefficient of variation $C_v = 0.70$. This is somewhat higher than the value $C_v = 0.556$ from our simulations of $m > 7.5$ earthquakes on the southern San Andreas fault. Note that the paleoseismic observations are at one point on a fault (i.e., Pallett Creek), whereas we have considered all events on a fault or fault segment.

Discussion

The statistical distribution of recurrence times plays an important role in terms of the probabilistic hazard assessment for earthquakes. Ideally the applicable distribution could be obtained from the historical record of earthquakes on a fault. In general, however, this record is too short to be of much value (Savage, 1994). An alternative approach is to use numeral simulations. In this article we report on a 10^6-yr simulation of earthquakes on the major strike-slip faults of the San Andreas fault system. The results correlate well with the Weibull distribution.

A measure of the variability of recurrence times on a fault or fault segment is the coefficient of variation C_v (the ratio of the standard deviation σ to the mean μ). For strictly periodic earthquakes on a fault or fault segment we would have $\sigma = C_v = 0$. For the random (exponential, no memory) distribution of recurrence times we would have $C_v = 1$ ($\sigma = \mu$). Typically we find C_v values in the range 0.5 to 0.75.

The variability of recurrence times on a specified fault or fault segment is attributed to fault interactions. An important question is whether our simulations provide recurrence-

Table 1
Virtual California Simulation Results for a 10^6 Year Run of Earthquake Activity

	Cutoff M_c	N	Mean μ (yr)	S.D. σ (yr)	C_v	Weibull τ (yr)	Weibull β
Virtual California simulation (10^6 yr)							
Northern San Andreas	7.5	4606	217	115	0.530	245	1.976
Southern San Andreas	7.5	5093	196	109	0.556	221	1.875
Hayward	7.0	2612	383	229	0.60	429	1.719
Calaveras	6.8	8174	122	87	0.715	135	1.42
San Gabriel	6.7	1913	522	398	0.762	568	1.325
San Jacinto	7.2	1075	929	562	0.605	1042	1.702
Observations (San Andreas)							
Parkfield	6.0	7	24.5	9.25	0.38	23.6	2.9
Pallett Creek (Sieh *et al.*, 1989)	7.0+	7	155	109	0.704	166	1.50

Results are given for six faults or fault segments with specified minimum earthquake magnitudes M_c. Included are the number of events N, the mean recurrence time μ, standard deviation (S.D.) σ, and the coefficient of variation C_v. Also given are the corresponding fitting parameters τ and β for the Weibull distribution. Results are also included for observed recurrence times on the Parkfield section and at Pallett Creek on the San Andreas fault.

time statistics similar to those on the actual faults. Evidence that this is in fact the case comes from the similarities between our simulations and the simulation results obtained by Goes and Ward (1994) and Ward (1996, 2000). The statistical results they obtained using the SPEM model are similar to those reported here using the Virtual California model. The two simulation models have many differences yet the statistical distributions of recurrence times are similar. This would indicate that recurrence-time scaling may be a universal feature of regional seismicity in the sense that Gutenberg-Richter frequency-magnitude statistics are universal. Ellsworth *et al.* (1999) analyzed 37 series of recurrent earthquakes and suggested a provisional generic value of the coefficient of variation $C_v \approx 0.5$.

Another important question is whether there is a scaling of seismicity that requires the validity of the Weibull distribution. The Weibull distribution is widely used in engineering to model the statistical distribution of failure times. Its applicability has been demonstrated by many actual tests (Meeker and Escobar, 1991). In Appendix C we have shown how the Weibull distribution can be derived from damage mechanics and fiber-bundle models.

Note that the stretched exponential (Weibull) distribution is found to be widely applicable in heterogeneous and homogeneous nucleation. In this context its applicability is known as Avrami's law (Avrami, 1940). The similarities between nucleation problems (i.e., droplet formation in supercooled steam) and the nucleation of earthquakes have been discussed previously (i.e., Rundle *et al.*, 2003).

The sequence of recurrence times on a fault is a time series. Bunde *et al.* (2003) and Altmann and Kantz (2005) have shown that time series that exhibit long-range correlations have recurrence-time statistics that satisfy stretched exponential (Weibull) distributions. Specifically, Bunde *et al.* (2003) argue against distributions in which the extreme event statistics become Poissonian, that is, like the Brownian passage-time distribution.

Acknowledgments

This work has been supported by DOE Grant DE-FG02-04ER15668 (to G.Y., J.B.R., and P.B.R.) and NSF Grant ATM 0327558 (to D.L.T.).

References

Altmann, E. G., and H. Kantz (2005). Recurrence time analysis, long-term correlations, and extreme events, *Phys. Rev. E* **71,** 0,56,106.

Avrami, M. (1940). Kinetics of phase change. II. Transformation-time relations for random distributions of nuclei, *J. Chem. Phys.* **8,** 212–224.

Barnhard, T., and S. Hanson (1996). Fault parameters used in USGS 1996 seismic hazard maps, http://earthquake.usgs.gov/research/hazmaps/products_data/1996/faults.php (last accessed October 2006).

Biasi, G. P., R. J. Weldon, T. E. Fumal, and G. G. Seitz (2002). Paleosesmic event dating and the conditional probability of large earthquakes on the southern San Andreas fault, California, *Bull. Seism. Soc. Am.* **92,** 2761–2781.

Bunde, A., J. F. Eichner, S. Havlin, and J. W. Kantelhardt (2003). The effect of long-term correlations on the return periods of rare events, *Physica A* **330,** 1–7.

Chhikara, R. S., and J. L. Folks (1989). *The Inverse Gaussian Distribution: Theory, Methodology, and Applications*, Marcel Dekker, New York.

Coleman, B. D. (1958). Statistics and time dependence of mechanical breakdown in fibers, *J. Appl. Phys.* **29,** 968–983.

Covey, C., K. M. Achuta Rao, U. Cubasch, P. Jones, S. J. Lambert, M. E. Mann, T. J. Phillips, and K. E. Taylor (2003). An overview of results from the coupled model intercomparison project, *Global Planet. Change* **37,** 103–133.

Davis, P. M., D. D. Jackson, and Y. Y. Kagan (1989). The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seism. Soc. Am.* **79,** 1439–1456.

Deng, J. S., and L. R. Sykes (1997). Evolution of the stress field in southern California and triggering of moderate-size earthquakes: a 200-year perspective, *J. Geophys. Res.* **102,** no. B5, 9859–9886.

Ellsworth, W. L., M. V. Matthews, R. M. Nadeau, S. P. Nishenko, P. A. Reasenberg, and R. W. Simpson (1999). A physically based earthquake recurrence model for estimation of long-term earthquake probabilities, *U.S. Geol. Surv. Open-File Rept. 99-522*.

Goes, S. D. B. (1996). Irregular recurrence of large earthquakes: an analysis of historic and paleoseismic catalogs, *J. Geophys. Res.* **101,** 5739–5749.

Goes, S. D. B., and S. N. Ward (1994). Synthetic seismicity for the San Andreas Fault, *Ann. Geofis.* **37,** 1495–1513.

Hagiwara, Y. (1974). Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain, *Tectonophysics* **23,** 313–318.

Harris, R. A. (2003). Stress triggers, stress shadows, and seismic hazard, in *International Handbook of Earthquake and Engineering Seismology Part B*, W. H. K. Lee, H. Danamori, P. C. Jennings, and C. Kisslinger (Editors), Academic Press, Amsterdam, 1217–1232.

Kodera, K., K. Matthes, K. Shibata, U. Langematz, and Y. Kuroda (2003). Solar impact on the lower mesospheric subtropical jet: a comparative study with general circulation model simulations, *Geophys. Res. Lett.* **30,** no. 1315.

Krajcinovic, D. (1996). *Damage Mechanics*, Elsevier, Amsterdam.

Krajcinovic, D., and M. A. G. Silva (1982). Statistical aspects of the continuous damage theory, *J. Solid Struct.* **18,** 551–562.

Lorenz, E. N. (1963). Deterministic nonperiodic flow, *J. Atmos. Sci.* **20,** 130–141.

Matthews, M. V., W. L. Ellsworth, and P. A. Reasenberg (2002). A Brownian model for recurrent earthquakes, *Bull. Seism. Soc. Am.* **92,** 2233–2250.

Meeker, W. Q., and L. A. Escobar (1991). *Statistical Methods for Reliability Data*, John Wiley, New York.

Newman, W. I., and S. L. Phoenix (2001). Time dependent fiber-bundles with local load sharing, *Phys. Rev. E* **63,** 0,21,507.

Nishenko, S. P., and R. A. Buland (1987). A generic recurrence interval distribution for earthquake forecasting, *Bull. Seism. Soc. Am.* **77,** 1382–1399.

Ogata, Y. (1999). Estimating the hazard of rupture using uncertain occurrence times of paleoearthquakes, *J. Geophys. Res.* **104,** 17,995–18,014.

Ogata, Y. (2002). Slip-size-dependent renewal processes and Bayesian inferences for uncertainties, *J. Geophys. Res.* **107,** 2268.

Patel, J. K., C. H. Kapadia, and D. B. Owen (1976). *Handbook of Statistical Distributions*, Marcel Dekker, New York.

Petersen, M. D., W. A. Bryant, C. H. Cramer, T. Cao, M. Reichle, A. D. Frankel, J. J. Lienkaemper, P. A. McCrory, and D. P. Schwartz (1996). Probabilistic seismic hazard assessment for the state of California, *U.S. Geol. Surv. Open-File Rept. 96-706*.

Press, W. H., S. A. Teukolsky, W. T. Vetterling, and B. P. Flannery (1995). *Numerical Recipes in C*, Second Ed., Cambridge University Press, Cambridge, U.K.

Rikitake, T. (1976). Recurrence of great earthquakes at subduction zones, *Tectonophysics* **35,** 335–362.

Rikitake, T. (1982). *Earthquake Forecasting and Warning*, D. Reidel, Dordrecht, The Netherlands.

Rikitake, T. (1991). Assessment of earthquake hazard in the Tokyo area, Japan, *Tectonophysics* **199,** 121–131.

Rundle, J. B. (1988). A physical model for earthquakes, 2: Application to Southern California, *J. Geophys. Res.* **93,** 6255–6274.

Rundle, J. B., P. B. Rundle, A. Donnellan, and G. Fox (2004). Gutenberg-Richter statistics in topologically realistic system-level earthquake stress-evolution simulations, *Earth Planets Space* **56,** 761–771.

Rundle, J. B., P. B. Rundle, A. Donnellan, D. L. Turcotte, R. Shcherbakov, P. Li, B. D Malamud, L. B. Grant, G. C. Fox, D. McLeod, G. Yakovlev, J. Parker, W. Klein, and K. F. Tiampo (2005). A simulation-based approach to forecasting the next great San Francisco earthquake, *Proc. Natl. Acad. Sci.* **102,** 15,363–15,367.

Rundle, J. B., P. B. Rundle, W. Klein, J. S. Martins, K. F. Tiampo, A. Donnellan, and L. H. Kellogg (2002). GEM plate boundary simulations for the plate boundary observatory: a program for understanding the physics of earthquakes on complex fault networks via observations, theory and numerical simulation, *Pure Appl. Geophys.* **159,** 2357–2381.

Rundle, P. B., J. B. Rundle, K. F. Tiampo, J. S. Martins, S. McGinnis, and W. Klein (2001). Nonlinear network dynamics in earthquake fault systems, *Phys. Rev. Lett.* **87,** 1,48,501.

Rundle, J. B., P. B. Rundle, K. F. Tiampo, A. Donnellan, and D. L. Turcotte (2006). Virtual California: fault model, frictional parameters, applications, *Pure Appl. Geophys.* (in press).

Rundle, J. B., D. L. Turcotte, R. Shcherbakov, W. Klein, and C. Sammis (2003). Statistical physics approach to understanding the multiscale dynamics of earthquake fault systems, *Rev. Geophys.* **41,** no. 4, 1019.

Rundle, J. B., and J. H. Whitcomb (1984). A model for deformation in Long Valley, California, 1980–1983, *J. Geophys. Res.* **89,** 9371–9380.

Savage, J. C. (1991). Criticism of some forecasts of the national earthquake prediction council, *Bull. Seism. Soc. Am.* **81,** 862–881.

Savage, J. C. (1994). Empirical earthquake probabilities from observed recurrence intervals, *Bull. Seism. Soc. Am.* **84,** 219–221.

Shimazaki, K., and T. Nakata (1980). Time-predictable recurrence model for large earthquakes, *Geophys. Res. Lett.* **7,** 279–282.

Sieh, K., M. Stuiver, and D. Brillinger (1989). A more precise chronology of earthquakes produced by the San Andreas fault in southern California, *J. Geophys. Res.* **94,** 603–623.

Sornette, D., and L. Knopoff (1997). The paradox of the expected time until the next earthquake, *Bull. Seism. Soc. Am.* **87,** 789–798.

Stein, S., and A. Newman (2004) Characteristic and uncharacteristic earthquakes as possible artifacts: applications to the New Madrid and Wabash seismic zones, *Seism. Res. Lett.* **75,** 173–187.

Turcotte, D. L. (1997). *Fractals and Chaos in Geology and Geophysics*, Second Ed., Cambridge University Press, Cambridge, U.K.

Turcotte, D. L., W. I. Newman, and R. Shcherbakov (2003). Micro and macroscopic models of rock fracture, *Geophys. J. Int.* **152,** 718–728.

Utsu, T. (1984). Estimation of parameters for recurrence models of earthquakes, *Bull. Earthquake Res. Inst. Univ. Tokyo* **59,** 53–66.

Ward, S. N. (1992). An application of synthetic seismicity in earthquake statistics: the Middle America Trench, *J. Geophys. Res.* **97,** 6675–6682.

Ward, S. N. (1996). A synthetic seismicity model for southern California: Cycles, probabilities, hazards, *J. Geophys. Res.* **101,** 22,393–22,418.

Ward, S. N. (2000). San Francisco Bay Area earthquake simulations: a step toward a standard physical earthquake model, *Bull. Seism. Soc. Am.* **90,** 370–386.

Weibull, W. (1951). A statistical distribution of wide applicability, *J. Appl. Mech.* **18,** 293–297.

Working Group on California Earthquake Probabilities (1988). Probabilities of large earthquakes occurring in California on the San Andreas fault, *U.S. Geol. Surv. Open-File Rept. 88-398*.

Working Group on California Earthquake Probabilities (1990). Probabilities of large earthquakes in the San Francisco Bay region, California, *U.S. Geol. Surv. Circular 1053*.

Working Group on California Earthquake Probabilities (1995). Seismic hazards in southern California: probable earthquakes, 1994 to 2024, *Bull. Seism. Soc. Am.* **85,** 379–439.

Working Group on California Earthquake Probabilities (2003). Earthquake probabilities in the San Francisco Bay region, *U.S. Geol. Surv. Open-File Rept. 03-214*.

Youngs, R. R., and K. J. Coppersmith (1985). Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates, *Bull. Seism. Soc. Am.* **75,** 939–964.

Appendix A

Statistical Distributions

In this article we compare simulation results with three statistical distributions: Weibull, lognormal, and Brownian passage time. The distribution functions that we consider are given in equations (1)–(5). The forms of the Weibull distribution that we consider are given in equations (6)–(11).

Examples of the Weibull distribution from equation (6) as a function of the nondimensional recurrence time t/τ are given in Figure A1 for $\beta = 1$, 2, and 4. For $\beta = 1$ the Weibull distribution reduces to the exponential distribution and has no memory; earthquakes occur randomly. For $\beta = 2$ the distribution is known as a Rayleigh distribution. In the limit $\beta \to \infty$ the distribution is a delta function and the earthquakes occur periodically with a period τ. For $0 < \beta < 1$ the Weibull distribution is known as the stretched exponential distribution in the physics literature.

In terms of the earthquake hazard, an important aspect of an applicable statistical distribution is its behavior for large t. This is best expressed in terms of the conditional probabilities that an earthquake will occur in a future interval

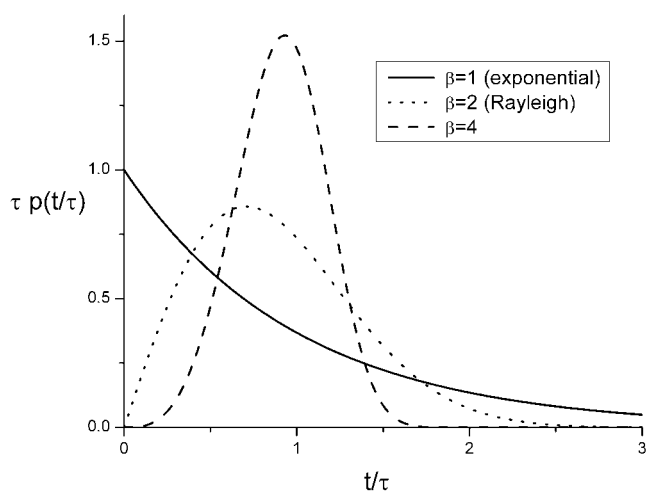

Figure A1. Dependence of the nondimensional PDF τp for recurrence times as a function of the nondimensional recurrence time t/τ for the Weibull distribution from equation (6). Distributions are given for $\beta = 1$ (Poisson distribution), $\beta = 2$ (Rayleigh distribution), and $\beta = 4$.

Δt if the time since the last earthquake is t_0. The relevant distributions have been given in equations (3) and (4). For the Weibull distribution the hazard function is given in equation (11). For $\beta = 1$ the hazard function is constant; that is, there is no memory of the last earthquake. The conditional CDF that an earthquake will occur in a waiting time Δt if the time since the last earthquake is t_0 is given in equation (10). An important measure of this behavior is the dependence of the median waiting time for the next earthquake $\Delta t_{1/2}/\tau$ on the time t_0 since the last earthquake. For the Weibull distribution this is given in equation (11). The dependence of the nondimensional median waiting time $\Delta t_{1/2}/\tau$ on the nondimensional time since the last earthquake t_0/τ is given in Figure A2 for $\beta = 1, 2, 4, \infty$. For the exponential distribution the waiting time is constant, $\Delta t_{1/2}/\tau = \ln 2$. The events are random so that there is no memory of the last event. In the limit $\beta \to \infty$ the events are periodic, $\mu = \tau$, and $\Delta t_{1/2}/\tau = 1 - t_0/\tau$ ($t_0 < \tau$). The waiting time decreases linearly with t_0 as expected.

The median waiting time until the next earthquake decreases as the time since the last earthquake only if $\beta > 1$. For this condition to be satisfied the tail of the distribution of recurrence times must be thinner than the exponential distribution associated with random earthquakes.

Two other distributions have been widely used to fit observed earthquake recurrence-time statistics; these are the lognormal and Brownian passage-time distributions. The lognormal is one of the most widely used statistical distributions in a wide variety of fields. The PDF for a lognormal distribution of recurrence times t is given by (Patel *et al.*, 1976)

$$p(t) = \frac{1}{(2\pi)^{1/2}\,\sigma_y t} \exp\left[-\frac{(\ln t - \bar{y})^2}{2\sigma_y^2}\right]. \quad \text{(A.1)}$$

The lognormal distribution can be obtained from the normal distribution by making the substitution $y = \ln t$; $\bar{y}$ and σ_y are the mean and standard deviation of this equivalent normal distribution. The mean μ, standard deviation σ, and coefficient of variation C_v for the lognormal distribution are given by

$$\mu = \exp\left(\bar{y} + \frac{1}{2}\sigma_y^2\right) \quad \text{(A.2)}$$

$$\sigma = \mu(e^{\sigma_y^2} - 1)^{1/2} \quad \text{(A.3)}$$

$$C_v = \sigma/\mu = (e^{\sigma_y^2} - 1)^{1/2}. \quad \text{(A.4)}$$

The corresponding CDF for the lognormal distribution from equations (2) and (A.1) is given by

$$P(t) = \frac{1}{2}\left[1 + \operatorname{erf}\left(\frac{\ln t - \bar{y}}{\sqrt{2}\sigma_y}\right)\right], \quad \text{(A.5)}$$

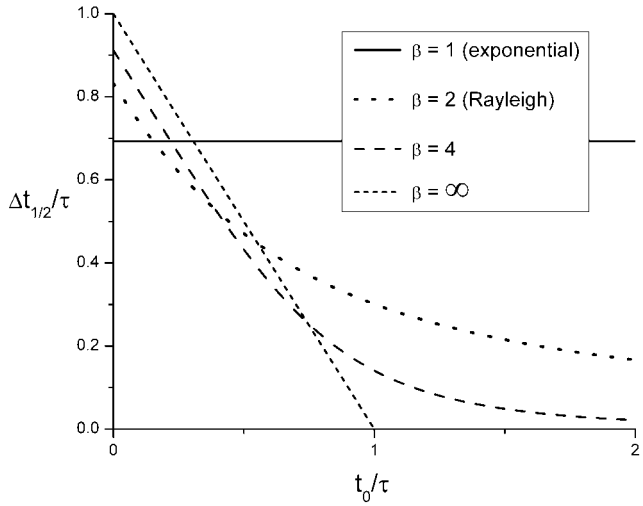

Figure A2. Dependence of the nondimensional median waiting time $\Delta t_{1/2}/\tau$ on the nondimensional time since the last earthquake t_0/τ from equation (13). Results are given for $\beta = 1$ (Poisson distribution), $\beta = 2$ (Rayleigh distribution), and $\beta = 4$, and $\beta = \infty$ (periodic earthquakes).

where $\operatorname{erf}(x) = 2/\pi^{1/2} \int_0^x e^{-y^2}\,dy$ is the error function. Substituting equation (A.5) into equation (5) we find that the conditional CDF hazard function for the lognormal distribution is given by

$$P(\Delta t, t_0) = 1 - \frac{\operatorname{erfc}\left(\dfrac{\ln[t_0 + \Delta t] - \bar{y}}{\sqrt{2}\sigma_y}\right)}{\operatorname{erfc}\left(\dfrac{ln\ t_0 - \bar{y}}{\sqrt{2}\sigma_y}\right)}, \quad \text{(A.6)}$$

where $\operatorname{erfc}(x) = 1 - \operatorname{erf}(x)$ is the complementary error function. This is the probability that an earthquake will occur in a waiting time Δt if the time since the last earthquake is t_0. The median waiting time $\Delta t_{1/2}$ corresponding the probability $P(\Delta t, t_0) = 0.5$ can be found from equation (A.6) by iteration.

The PDF for a Brownian passage-time distribution of recurrence times t is given by (Chhikara and Folks, 1989)

$$p(t) = \left(\frac{\mu}{2\pi C_v^2 t^3}\right)^{1/2} \exp\left[-\frac{(t - \mu)^2}{2C_v^2 \mu t}\right], \quad \text{(A.7)}$$

where μ is the mean and C_v is the coefficient of variation of the distribution. The corresponding conditional CDF for the probability that an earthquake will occur in a waiting time Δt if the time since the last earthquake is t_0 can be obtained analytically along with the corresponding median waiting time. However, these expressions are lengthy and are not given here explicitly.

Appendix B

The Virtual California Model

The Virtual California model by Rundle *et al.* (2001, 2004, 2005, 2006) is a backslip model in that the loading of each fault segment occurs via the accumulation of slip deficit $s(x, t) - Vt$, where $s(x, t)$ is slip, V is the constant long-term slip rate, and t is time. The advantage of the backslip model is that it is repetitive and does not evolve in time. The deformations are infinitesimal and linear. Ideally plate velocities would be applied at the boundaries of the model. But this results in finite strains and faults must lengthen as repetitive earthquakes occur. The result in the earth is granulation and healing. At the present time, the Virtual California model includes only vertical strike-slip faults with the potential for $m \geq 6$ earthquakes. Seismic-moment release in California is dominated by these faults. Thrust earthquakes, such as the 1994 Northridge and 1971 San Fernando faults, are certainly damaging, but their repetitive behavior is poorly documented. We believe that the similarities of the outputs between Virtual California and SPEM indicate that the overall statistical behavior is not sensitive to the details of the simulations. The topology of the fault distribution used in this article is illustrated in Figure 1. It is composed of 650 fault segments, each with a width of 10 km and a depth of 15 km.

Two data inputs are required for the model. These are the mean slip rate on each fault segment and the mean recurrence interval between earthquakes on the segment. Slip rates have been determined from a variety of sources including Deng and Sykes (1997, their table 1), Barnhard and Hanson (1996), and Petersen *et al.* (1996). It is assumed that each fault segment has a periodic characteristic earthquake of a prescribed magnitude on it. Thus with the slip rate prescribed, either the mean recurrence time or the earthquake magnitude is required. Neither of these quantities are well constrained by either historical or paleo data for segments, however. Some data are available from Deng and Sykes (1997, their table 2). However, for many fault segments interpolations are required. It is assumed that the characteristic earthquake on one of these segments is similar to other segments in the vicinity. Details of this interpolation have been given by Rundle *et al.* (2006). To stabilize the computations it was necessary to introduce aseismic slip on all fault segments. The aseismic slip rate was generally taken to be 10% of the total prescribed slip rate on the fault segment.

The present version of Virtual California utilizes a Coulomb failure criterion based on prescribed values of the static and dynamic coefficients of friction on each fault segment. The dynamic coefficients are selected so that the mean regional stress level is 20 MPa. The static coefficients of friction are selected so that characteristic earthquakes occur at the prescribed intervals.

The key aspect of the Virtual California model is the interaction between fault segments, which are obtained using static Green's functions. Changes in both shear stress and normal stress on all other segments are determined when slip occurs on the segment under consideration. Backslip is driven on all fault segments until the weakest segment becomes unstable using the Coulomb failure criterion and the prescribed static coefficient of friction. The slip on this segment then occurs and the resulting changes in stress on all other segments are determined. Other segments are then checked for stability. This process is repeated until all segments are found stable.

To expedite slip on adjacent fault segments a reduction in the static coefficient of friction on a segment was associated with a rapid increase in the stress on the fault. This is a parameterization of the effects associated with a dynamic stress intensity factor. Quantitative details of these inputs have been given by Rundle *et al.* (2006).

Appendix C

Derivation of the Weibull Distribution

The Weibull distribution is widely used in engineering studies of the statistical distributions of catastrophic failures (Meeker and Escolar, 1991). A standard failure problem in engineering is the fiber bundle or stranded cable. The generally accepted approach to the dynamic time-dependent failure of a fiber bundle is to specify an expression for the rate of failure of fibers (Coleman, 1958; Newman and Phoenix, 2001). The form of this breakdown rule is given by

$$\frac{dN}{dt} = -Nv(\sigma), \tag{C.1}$$

where N is the number of unbroken fibers at time t and $v(\sigma)$ is known as the hazard rate, a function of the stress σ applied to the fibers.

To complete the specification of the problem it is necessary to prescribe this dependence. For engineering materials it is standard practice (Newman and Phoenix, 2001) to empirically assume the power-law relation

$$v(\sigma) = v_0\left(\frac{\sigma}{\sigma_0}\right)^p, \tag{C.2}$$

where v_0 is the reference hazard rate at the reference stress σ_0, and p is the power-law exponent.

Krajcinovic (1996) and Turcotte *et al.* (2003) have shown an equivalence between the fiber-bundle model given previously and the damage model. In damage mechanics a damage variable α is introduced according to

$$E = E_0(1 - \alpha), \tag{C.3}$$

where Young's modulus E is reduced from its undamage ($\alpha = 0$) value E_0 until failure occurs when $\alpha = 1$ ($E = 0$). The damage α in a fiber bundle is given by

$$\alpha = 1 - \frac{N}{N_0}, \tag{C.4}$$

where N_0 is the original number of fibers.

Because in the fiber-bundle model we assume that the stress σ on all fibers is equal, the fibers do not interact. Thus it is appropriate to assume an ergodic hypothesis. Instead of applying our model to a fiber bundle with N_0 fibers, we apply it to a single fiber that is replaced N_0 times. The single fiber is analogous to a fault and the failure of the fiber is analogous to an earthquake. The distribution of failure times of the fiber is analogous to the distribution of recurrence times on a fault. The CDF of failure times of fibers P is related to N by

$$P(t) = 1 - \frac{N(t)}{N_0}. \tag{C.5}$$

We further assume that after each fiber failure the stress on the fiber is reset to zero and increases linearly with time

$$\sigma(t) = \sigma_0 \frac{t}{\tau}. \tag{C.6}$$

This represents the linear increase in the tectonic stress on a fault after an earthquake.

To relate the constants v_0 and p in the fiber-bundle model equation (C.2) to the constants β and τ in the Weibull distribution equation (10) we write

$$p = \beta - 1 \tag{C.7}$$

$$v_0 = \frac{\beta}{\tau}. \tag{C.8}$$

Substitution of equations (C.5) to (C.8) into equations (C.1) and (C.2) gives

$$\frac{dP}{dt} = \frac{\beta}{\tau}\left(\frac{t}{\tau}\right)^{\beta-1}(1 - P). \tag{C.9}$$

Integrating with the initial condition that $P = 0$ at $t = 0$ gives

$$P(t) = 1 - \exp\left[-\left(\frac{t}{\tau}\right)^{\beta}\right]. \tag{C.10}$$

This is the cumulative form of the Weibull distribution given in equation (10). Krajcinovic and Silva (1982) have shown that damage mechanics also leads to Weibull distributions.

Center for Computational Science and Engineering
University of California
One Shields Avenue
Davis, California 95616
(G.Y., J.B.R., P.B.R)

Department of Geology
University of California
One Shields Avenue
Davis, California 95616
(D.L.T.)

Manuscript received 29 August 2005.