A simulation-based approach to forecasting the next great San Francisco earthquake


*Center for Computational Science and Engineering, and †Department of Geology, University of California, Davis, CA 95616; ‡Earth Space Science Division, and §Exploration Systems Autonomy Section, Jet Propulsion Laboratory, Pasadena, CA 91125; †Department of Geography, Kings College London, London WC2R 2LS, United Kingdom; **Department of Environmental Health, Science, and Policy, University of California, Irvine, CA 92697; ¶¶Department of Computer Science and Physics, Indiana University, Bloomington, IN 47405; ¶Department of Computer Science, University of Southern California, Los Angeles, CA 90089; §§Department of Physics, Boston University, Boston, MA 02215; and ¶¶Departments of Earth Science and Biological and Geological Sciences, University of Western Ontario, London, ON, Canada N6A 5B7

Contributed by D. L. Turcotte, August 31, 2005

In 1906 the great San Francisco earthquake and fire destroyed much of the city. As we approach the 100-year anniversary of that event, a critical concern is the hazard posed by another such earthquake. In this article we examine the assumptions presently used to compute the probability of occurrence of these earthquakes. We also present the results of a numerical simulation of interacting faults on the San Andreas system. Called Virtual California, this simulation can be used to compute the times, locations, and magnitudes of simulated earthquakes on the San Andreas fault in the vicinity of San Francisco. Of particular importance are results for the statistical distribution of recurrence times between great earthquakes, results that are difficult or impossible to obtain from a purely field-based approach.

hazards | Weibull distribution

The great San Francisco earthquake (April 18, 1906) and subsequent fires killed >3,000 persons and destroyed much of the city, leaving 225,000 of 400,000 inhabitants homeless. The 1906 earthquake occurred on a 470-km segment of the San Andreas fault that runs from the San Juan Bautista north to Cape Mendocino (Fig. 1) and is estimated to have had a moment magnitude $M \approx 7.9$ (1). Observations of surface displacements across the fault ranged from 2.0 to 5.0 m (2). As we approach the 100th anniversary of the great San Francisco earthquake, timeless questions to ask are to what extent the hazard is posed by another such event and how can this hazard be estimated.

The San Andreas fault is the major boundary between the Pacific and North American plates, which move past each other at an average rate of 49 mm yr$^{-1}$ (3), implying that to accumulate 2.0–5.0 m of displacement 40–100 years are needed. One of the simplest hypotheses for the recurrence of great earthquakes in the San Francisco area is that they will occur at approximately these 40–to 100-year time intervals. This idea would indicate that the next earthquake may be imminent. However, there are two problems with this simple “periodic” hypothesis. The first is that it is now recognized that only a fraction of the relative displacement between the plates occurs on the San Andreas fault proper. The remaining displacement occurs on other faults in the San Andreas system. Hall et al. (4) concluded that the mean displacement rate on just the northern part of the San Andreas Fault is closer to 24 mm yr$^{-1}$. With the periodic hypothesis this would imply recurrence intervals of 80–200 years.

The second and more serious problem with the periodic hypothesis involves the existence of complex interactions between the San Andreas Fault and other adjacent faults. It is now recognized (5–7) that these interactions lead to chaotic and complex nonperiodic behavior so that exact predictions of the future evolution of the system are not possible. Only probabilistic hazard forecasts can be made. For the past 15 years a purely statistical approach has been used by the Working Group on California Earthquake Probabilities (WGCEP) (8–11) to make risk assessments for northern California. Its statistical approach is a complex, collaborative process that uses observational data describing earthquake slips, lengths, creep rates, and other information on regional faults as inputs to a San Francisco Bay Regional fault model. Using its forecast algorithm, the WGCEP (11) found that the conditional probability for the occurrence of an earthquake having $M \geq 6.7$ during the 30-year period 2002–2031 is 18.2%.

As described in the WGCEP report (11), the critical assumption in computing the hazard probability is the choice of a probability distribution, or renewal model. The WGCEP study used the Brownian passage time (BPT) distribution. Previous studies used the log normal (LN) (8–10) and the Weibull distributions. The means and standard deviations of the distributions for event times on the fault segments were constrained by geological and seismological observations.

Virtual California

In this article, we present the results of a topologically realistic numerical simulation of earthquake occurrence on the San Andreas fault in the vicinity of San Francisco. This simulation, called Virtual California, includes fault system physics, such as the complex elastic interactions between faults in the system, and friction laws developed with insights from laboratory experiments and field data. 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 (12, 13). Turbulent phenomena are represented by parameterizations of the 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 (5), general circulation models have repeatedly shown their value. In many cases ensemble forecasts are carried out, which use simulations computed with multiple models to test the robustness of the forecasts.

The Virtual California simulation, originally developed by Rundle (14), includes stress accumulation and release, as well as stress interactions between the San Andreas and other adjacent faults. The model is based on a set of mapped faults with estimated slip rates, prescribed long-term rates of fault slip, parameterizations of friction laws based on laboratory experiments and historic earthquake occurrence, and elastic interactions. An updated version of Virtual California (15–17) is used in this article. The faults in the model are those that have been

Abbreviations: WGCEP, Working Group on California Earthquake Probabilities; BPT, Brownian passage time; LN, log normal.

§To whom correspondence may be addressed. E-mail: rundle@cse.ucdavis.edu or turbott@geology.ucdavis.edu.

© 2005 by The National Academy of Sciences of the USA
active in recent geologic history. Earthquake activity data and slip rates on these model faults are obtained from geologic databases of earthquake activity on the northern San Andreas fault. A similar type of simulation has been developed by Ward and Goes (18) and Ward (19). A consequence of the size of the fault segments used in this version of Virtual California is that the simulations do not generate earthquakes having magnitudes less than about $m \approx 5.8$.

Virtual California is a backslip model, the loading of each fault segment occurs because of the accumulation of a slip deficit at the prescribed slip rate of the segment. The vertical rectangular fault segments interact elastically, and the interaction coefficients are computed by means of boundary element methods (20). Segment slip and earthquake initiation are controlled by a friction law that has its basis in laboratory-derived physics (17). Onset of initial instability is controlled by a static coefficient of friction. Segment sliding, once begun, continues until a residual stress is reached, plus or minus a random overshoot or undershoot of typically 10%. To prescribe the friction coefficients we use historical displacements in earthquakes having moment magnitudes $m \geq 5.0$ in California during the last 200 years (17).

The topology of Virtual California is shown in Fig. 1 superimposed on a LandSat image. The 650 fault segments are represented by lighter and darker lines. The lighter lines represent the San Andreas fault, stretching from the Salton trough in the south to Cape Mendocino in the north. The “San Francisco section” of the San Andreas fault, $\approx 250$ km in length, is the section of the fault whose rupture would be strongly felt in San Francisco and is considered here. Using standard seismological relationships (21), we estimate that an earthquake having $m_{SF} = 7.0$ with an average slip of 4 m and a depth of 15 km would rupture a 20-km length of fault. Earthquakes like these would produce considerable damage, destruction, and injury in San Francisco.

Our goal is to forecast waiting times until the next great earthquake on the yellow section in Fig. 1 of the fault for a minimum magnitude $m_{SF} = 7.0$. Using Virtual California, we advance our model in 1-year increments and simulate 40,000 years of earthquakes on the entire San Andreas Fault system. We note that although the average slip on a fault segment 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. In Fig. 2, we show examples of the distribution of earthquakes on the San Francisco section of the San Andreas fault for a 3,000-year period. Fig. 2 Left shows the slip in each earthquake as a function of distance along the fault from Fort Ross in the north to San Juan Bautista in the south. Fig. 2 Right shows the moment magnitudes of each of the simulated earthquakes.

One output of our simulations is the distribution of surface displacements caused by each model earthquake. Synthetic aperture radar interferometry is routinely used to obtain the coseismic displacements that occur after earthquakes (22). The displacements associated with two sets of our model earthquakes are illustrated in Fig. 3 as interferometric patterns. Each interferometric fringe corresponds to a displacement along the line of sight to the hypothetical spacecraft of 56 mm.

**Earthquake Risk**

A quantitative output of our simulations is the statistical distribution of recurrence times $t$ between successive great earthquakes on a given fault. For the northern section of the San Andreas fault near San Francisco, this distribution is required if the risk of future earthquakes on the fault is to be specified. We associate the properties of this distribution directly with the elastic interactions between faults, which are an essential feature of our model. Current estimates of risk are based on the observed statistics of time intervals. However, Savage (23) has argued convincingly that actual sequences of earthquakes on specified faults are not long enough to establish the statistics of recurrence times with the required reliability. We argue that it is preferable to use numerical simulations to obtain applicable statistics. We illustrate this approach by using numerical simulations to obtain recurrence statistics for synthetic earthquakes on the San Francisco section of the San Andreas fault over 40,000 years.

We consider earthquakes on the section of the northern San Andreas fault shown in yellow in Fig. 1. Over the 40,000-year simulation, we obtained 395 simulated $m_{SF} \geq 7.0$ events having an average recurrence interval of 101 years. From the simulations, we measured the distribution of recurrence times $t$ be-
tween great earthquakes on the San Francisco segment. The time \( t \) is defined as the recurrence time between two successive great earthquakes (24).

A second important distribution that we will consider is the distribution of waiting times \( \Delta t \) until the next great earthquake, given that the time elapsed since the most recent great earthquake is \( t_0 \). If we take the time of the last great earthquake to be 1906 and the present to be 2005, we find for San Francisco \( t_0 = 2005 - 1906 = 99 \) years. The waiting time \( \Delta t \) is measured forward from the present, thus \( t = t_0 + \Delta t \). We will express our results in terms of the cumulative conditional probability \( P(t; t_0) \) that an earthquake will occur in the waiting time \( \Delta t = t - t_0 \) if the elapsed time since the last great earthquake is \( t_0 \) (25).

A probability distribution that has often been applied to recurrence statistics is the Weibull distribution (26–29), and it is used here for reasons that we will describe. For the Weibull distribution the fraction of the recurrence times \( P(t) \) that are \(< t \) can be expressed as

\[
P(t) = 1 - \exp \left[ -\left( \frac{t}{\tau} \right)^\beta \right],
\]

where \( \beta \) and \( \tau \) are fitting parameters. Sieh et al. (30) fit this distribution to the recurrence times of great earthquakes on the southern San Andreas fault obtained from paleoseismic studies with \( \tau = 166 \pm 44.5 \) years and \( \beta = 1.5 \pm 0.8 \). In its extension to the cumulative conditional probability the Weibull distribution is given by ref. 31

\[
P(t; t_0) = 1 - \exp \left[ \left( \frac{t}{\tau} \right)^\beta - \left( \frac{t_0}{\tau} \right)^\beta \right].
\]

Eq. 2 specifies the cumulative conditional probability that an earthquake will have occurred at a time \( t \) after the last earthquake if the earthquake has not occurred by a time \( t_0 \) after the last earthquake.

![Fig. 3](image1.png)

**Fig. 3.** Interferometric patterns of the coseismic deformations associated with two sets of model earthquakes. Each interferometric fringe corresponds to a displacement of 56 mm. (A) \( \bullet\bullet\bullet\). (B) \( \bullet\bullet\bullet\).

![Fig. 4](image2.png)

**Fig. 4.** \( \bullet\bullet\bullet\). (a) The wiggly line is the simulation-based cumulative probability \( P(t) \) that a great \( m_{gr} \geq 7.0 \) earthquake will have occurred on the San Andreas Fault near San Francisco at a recurrence time \( t \) years after the last great earthquake with \( m_{gr} \geq 7.0 \). For comparison, we plot three cumulative probability distributions having the same mean \( \mu = 101 \) years and standard deviation \( \sigma = 61 \) years as the simulation data. The solid line is the best-fitting Weibull distribution; the dashed line is the BPT distribution; and the dotted line is the LN distribution. (b) The wiggly line is the conditional probability \( P(t_0 + 30, t) \) that a magnitude \( m_{gr} \geq 7.0 \) event will occur in the next 30 years, given that it has not occurred by a time \( t \) since the last such event. The solid line is the corresponding conditional probability for the Weibull distribution; the dashed line is for the BPT; and the dotted line is for the LN.
We first consider the type of statistical forecast described in the WGCEP report (9). In Fig. 4a, the wiggly line is the cumulative probability \( P(t) \) that a simulated great \( m_{gF} \approx 7.0 \) earthquake will have occurred on the San Andreas Fault near San Francisco at a time \( t = t_0 + \Delta t \) years after the last great earthquake, if the last great earthquake occurred \( t_0 \) years ago in the past. Results are given for \( t_0 = 0, 25, 50, 75, 100, 125, \) and 150 years. Also included are fits to the data of the Weibull distribution.

In Fig. 4b we show the same type of conditional probability computed by the WGCEP (9), obtained from the simulation data in Fig. 4a. In Fig. 4b the wiggly line is the simulation-based conditional probability \( P(t_0 + 30, t) \) that a magnitude \( m_{gF} \approx 7.0 \) event will occur in the next 30 years, given that it has not occurred during the time \( t_0 \) since the last such event. For comparison, in Fig. 4b the solid line is the corresponding conditional probability for the Weibull distribution; the dashed line is for the BPT; and the dotted line is for the LN.

From the results shown in Fig. 4, it can be seen that the Weibull distribution describes the simulation data substantially better than either the BPT or LN distributions. At least in Virtual California, we can conclude that among these three statistical distributions, the Weibull distribution is the preferred distribution to describe the failure of a group of fault segments interacting by means of elastic stress transfer.

The corresponding cumulative conditional distributions of waiting times \( \Delta t \) from our simulations are given in Fig. 5. These are the cumulative conditional probabilities that an earthquake will have occurred at a time \( t = t_0 + \Delta t \) after the last earthquake if it has not occurred at a time \( t_0 \). We remove recurrence times that are less than or equal to \( t_0 \) and plot the cumulative distribution of the remaining recurrence times. The left-most curve \( P(t_0) \) in Fig. 5 is the same as the distribution of recurrence times \( P(t) \) given in Fig. 4a. Cumulative conditional distributions \( P(t_0) \) are given in Fig. 5 with \( t_0 = 25, 50, 75, 100, 150 \), and 150 years. With the fitting parameters \( \beta \) and \( \tau \) used to fit Eq. 1 to the cumulative distributions of waiting times \( P(t) \), we again compare the predictions of the Weibull distribution for \( P(t_0) \) from Eq. 2, the smooth curves, with data from our simulations in Fig. 5, the irregular curves. Again good agreement is found.

The data given in Fig. 5 can also be used to determine the waiting times to the next great earthquake \( \Delta t = t - t_0 \) corresponding to a specified probability of occurrence as a function of the time since the last great earthquake occurred \( t_0 \). This dependence is given in Fig. 6. The small stars in Fig. 6 are the median waiting times \( \Delta t \), \( P(t_0 + \Delta t, t_0) = 0.5 \), to the next great earthquake as a function of the time \( t_0 \) since the last great earthquake. These stars are the intersections of the dashed line with \( P(t_0) = 0.5 \) with the cumulative distributions in Fig. 5. Also given as circles in Fig. 6 are the waiting times for \( P(t_0) = 0.25 \) (lower limit of the gray band) and \( P(t_0) = 0.75 \) (upper limit of the gray band). The dashed red lines in Fig. 5 are the forecasts of risk based on the Weibull distributions from Eq. 2.

Immediately after a great earthquake, e.g., in 1906, we have \( t_0 = 0 \) years. At that time, Figs. 5 and 6 indicate that there was a 50% chance of having an earthquake \( m_{gF} \approx 7.0 \) in the next \( t = 90 \) years, i.e., in 1996. In 2006 it will have been 100 years since the last great earthquake occurred in 1906. The cumulative conditional distributions corresponding to this case have \( t_0 = 100 \) years. We see from Figs. 5 and 6 that there is a 50% chance of having a great earthquake \( m_{gF} \approx 7.0 \) in the next \( \Delta t = 45 \) years (\( t = 145 \) years). This is indicated by the large star in Fig. 6. It can also be seen that there is a 25% chance for such an earthquake in the next \( \Delta t = 20 \) years (\( t = 120 \) years), and a 75% chance of having such an earthquake in the next \( \Delta t = 80 \) years (\( t = 180 \) years). During each year in this period, to a good approximation, there is a 1% chance of having such an earthquake. These estimates are consistent with the information in Fig. 4b, which indicates a 30% chance of an \( m_{gF} \approx 7.0 \) earthquake during the period 2006–2036.

We see from Figs. 3–6 that the Weibull distribution that fits the distribution of interval times also does an excellent job of fitting the conditional probabilities and the waiting times. In both simulations and in our Weibull fit, the median waiting times systematically decrease with increases in the time since the last great earthquake. This is not the case for other distributions that provide a good fit to interval times (9). Our results therefore support the use of Weibull distributions to carry out probabilistic hazard analyses of earthquake occurrences.

**Discussion**

There are major differences between the simulation-based forecasts given here and the statistical forecasts given by the WGCEP


