Forecasting the Next Great San Francisco Earthquake

John B. Rundle[1,2*], Paul B. Rundle[1], Andrea Donnellan[2], Donald L. Turcotte[3], Robert Shcherbakov[1], Peggy Li[4], Bruce D. Malamud[5], Lisa Grant[6], Geoffrey Fox[7], Dennis McLeod[8], Gleb Morein[1], Jay Parker[4], and William Klein[9]

[1]Center for Computational Science and Engineering, University of California, Davis, CA 95616, USA

[2]Earth & Space Sciences Division, Jet Propulsion Laboratory, Pasadena, CA 91125, USA

[3]Department of Geology, University of California, Davis, CA 95616, USA

[4]Exploration Systems Autonomy Section, Jet Propulsion Laboratory, Pasadena, CA 91125, USA

[5]Department of Geography, Kings College London, WC2R 2LS, UK

[6]Department of Environmental Health, Science & Policy, University of California, Irvine, CA 92697, USA

[7]Departments of Computer Science and Physics, School of Informatics, Indiana University, Bloomington, IN 47405, USA

[8]Department of Computer Science, University of Southern California, Los Angeles, CA 90089, USA

[9]Department of Physics, Boston University, Boston, MA 02215, USA

[*]To whom correspondence should be addressed; E-mail: rundle@cse.ucdavis.edu.

In 1906 the great San Francisco earthquake and fire destroyed the city. As we approach this event's 100 year anniversary, a critical concern is when the next great San Francisco earthquake will occur. In this paper we present a new probabilistic risk analysis for future great earthquakes, based on simulations of earthquake fault systems. We find there is a 5% chance of an event with

magnitude $\mathrm{m} \geq 7.0$ occurring on the San Andreas Fault near San Francisco prior to 2009 and a 55% chance by 2054.

The great San Francisco earthquake (18 April 1906) and subsequent fires killed more than 3,000 persons, and destroyed much of the city leaving 225,000 out of 400,000 inhabitants homeless. As we approach the hundred year anniversary of this event, the question of when the next great San Francisco earthquake will occur is of great concern. In this paper we present probabilistic forecasts based on a new computer simulation approach. The 1906 San Francisco earthquake occurred on a 470 km segment of the San Andreas Fault that runs from the San Francisco Bay north to Cape Mendocino (Fig. 1) and is estimated to have had a moment magnitude $m = 7.9$ (*1*). Observations of surface displacements across the fault were in the range $2.0 - 5.0$ m (*2*). 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 San Francisco earthquakes is that they will occur at approximately these $40 - 100$ year time intervals. This 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, which in northern California is primarily in the east San Francisco Bay region, on the Hayward and Calaveras faults (see Fig. 1). A variety of studies (*4*) indicate 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 interval of 80 to 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*) that these interactions lead to chaotic and complex non-periodic behavior

so that exact prediction of the future evolution of the system is not possible. Only probabilistic hazard forecasts can be made. It is the purpose of this paper to make such a forecast, utilizing direct numerical simulations of fault system physics that include complex fault interactions.

Simulation-based approaches to forecasting and prediction of natural phenomena have been used with great success for weather and climate. The latter are often referred to as *General Circulation Models* (*6, 7*). Many of the phenomena are represented by parameterizations of the dynamics, and the equations are typically solved over spatial grids having length scales of a few degrees. Although even simple forms of the fluid dynamics equations are known to display chaotic behavior (*8*), general circulation models have repeatedly shown their value.

A simulation-based approach to earthquake risk assessment called *Virtual California* was developed by Rundle (*9*). This model was developed to include stress accumulation and release as well as stress interactions including the San Andreas and other adjacent faults. The model was based on a set of mapped faults with estimated slip rates, a prescribed plate tectonic motion, earthquakes on all faults, and elastic interactions. An updated version of Virtual California is used in this paper (*10–12*). 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 (*13–15*). A similar type of simulation has been developed by Ward (*16*).

Virtual California is a *backslip* model in that the loading of each fault segment occurs due to the accumulation of "backwards slip", or *slip deficit*, at the prescribed slip rate of the fault segment. The vertical rectangular fault segments are embedded in an elastic half space and interact elastically. Earthquake initiation is controlled by the coefficients of static and kinetic friction along with the space- and time-dependent shear and normal stresses on fault segments. Both shear and normal stresses on segments are computed by means of boundary element methods (*17*). To prescribe the friction coefficients, we use historical earthquakes having moment magnitudes $m \geq 5.0$ in California during the last ~ 200 years. A consequence of our fault

segmentation is that our simulations do not generate earthquakes having magnitudes less than about $m \approx 5.8$.

The topology of Virtual California is shown in Fig. 1 superimposed on a LandSat image. The 650 strike-slip fault segments are represented by red, blue, and yellow lines. The blue and yellow lines represent the San Andreas fault, stretching from the Salton trough in the south to Cape Mendocino in the north. The yellow line represents the "San Francisco section" of the San Andreas fault, about 250 km in length, and is the section of the fault whose rupture would be strongly felt in San Francisco. Our goal is to forecast waiting times until the next great earthquake on the yellow section of the fault for two minimum magnitudes: (i) $m_{SF} = 7.0$ and (ii) $m_{SF} = 7.3$. We use the subscript SF to indicate that the earthquake is on the San Andreas fault near San Francisco. Using standard seismological relationships (*18*), we estimate that an earthquake with $m_{SF} = 7.0$, an average slip of 4 m, and a depth of 15 km, would rupture approximately a 20 km length of fault. With similar conditions, an earthquake with $m_{SF} = 7.3$ would rupture a 66 km length of fault. Earthquakes like these would produce a large amount of damage in San Francisco.

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. It is important to 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.

Synthetic aperture radar interferometry (InSAR) is routinely used to obtain the coseismic displacements that occur after earthquakes (*19*). The displacements associated with two sets of our model earthquakes are illustrated in Fig. 2 as interferometric patterns. Each interferometric fringe corresponds to a displacement of 28 mm.

We will now consider only the model earthquakes on the section of the northern San Abdreas fault shown in yellow in Fig. 1. Over the 40,000 year simulation, we obtained 395 $m_{SF} \geq 7.0$ events having an average recurrence interval of 101 years, and 159 $m_{SF} \geq 7.3$ events having an average recurrence interval of 249 years.

Using the time intervals between successive events, we construct probabilistic forecasts of when the next great San Francisco earthquake will occur. The time t is measured forward from the time of occurrence of the last great earthquake. The time t_0 is the time since the last great earthquake. For San Francisco $t_0 = 2005 - 1906 = 99$ years. The waiting time Δt is measured forward from the present, thus $t = t_0 + \Delta t$. We 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 (*20*). We will compare our results with the Weibull distribution which gives the cumulative distribution of interval times between earthquakes as

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

where β and τ are fitting parameters. Although many distributions have been applied to earthquake interval times, the Weibull distribution is one of the most widely used (*21*). Sieh et al. (*22*) fit this distribution to the interval times of great earthquakes on the southern San Andreas fault with $\tau = 166.1 \pm 44.5$ years and $\beta = 1.5 \pm 0.8$. For the Weibull distribution, the cumulative conditional probability is given by (*23*)

$$P(t, t_0) = 1 - \exp\left[\left(\frac{t_0}{\tau}\right)^{\beta} - \left(\frac{t}{\tau}\right)^{\beta}\right] . \tag{2}$$

This is the cumulative conditional probability that an earthquake will have occurred at a time t after the last earthquake if the earthquake has not occurred at a time t_0 after the last earthquake.

From our simulations we first determine the distribution of interval times t between earthquakes for a specified minimum earthquake magnitude. The cumulative distribution of interval

times $P(t,0)$ for $m_{SF} \geq 7.0$ is given in Fig. 3a and for $m_{SF} \geq 7.3$ in Fig. 3b. These are the curves that pass through $t = 0$. Also included are the Weibull distributions from Eq. (1) that best fit the data. For $m_{SF} \geq 7.0$ in Fig. 3a our best fit requires $\beta = 1.67$ and $\tau = 114$ years, for $m_{SF} \geq 7.3$ in Fig. 3b our best fit requires $\beta = 2.17$ and $\tau = 289$ years.

We next determine the cumulative conditional probabilities that an earthquake will occur at a time t after the last earthquake if it has not occurred at a time t_0. To do this we remove interval times that are less than or equal to t_0 and plot the cumulative distribution of the remaining interval times. The resulting distributions $P(t,t_0)$ are given in Fig. 3a for $m_{SF} > 7.0$ with $t_0 = 25$, 50, 75, 100, 125, and 150 years and in Fig. 3b for $m_{SF} > 7.3$ with $t_0 = 50$, 100, 150, 200, 250, and 300 years. With the fitting parameters β and τ used to fit Eq. (1) to the cumulative distributions of waiting times $P(t,t_0)$, we compare the predictions of the Weibull distribution from Eq. (2) with our simulations in Fig. 3. We see that the agreement is quite good.

The results given in Fig. 3 are presented differently in Fig. 4. The stars in Fig. 4 are the median waiting times Δt, $P(t_0, t_0 + \Delta t) = 0.5$, 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 red lines with $P(t,t_0) = 0.5$ with the cumulative distributions in Fig. 3. Also given as circles in Fig. 4 are the waiting times for $P(t, t_0 + \Delta t) = 0.25$ (lower limit of the yellow band) and for $P(t, t_0 + \Delta t) = 0.75$ (upper limit of the yellow band). The dashed red lines are the forecasts of risk based on the Weibull distributions from Eq. (2).

Immediately after a great earthquake, i.e., in 1906, we have $t_0 = 0$ years. At that time, Figs. 3a and 4a indicate that there was a 50% chance of having earthquake $m_{SF} \geq 7.0$ in the next $t = 90$ years, i.e., in 1996. Also at that time ($t_0 = 0$ years), there was a 50% chance of having an earthquake with $m_{SF} \geq 7.3$ in the next $t = 249$ years, as shown in Figs. 3b and 4b. However, 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. 3a and 4a that there is a 50% chance of having a great earthquake ($m_{SF} \geq 7.0$) in the next $\Delta t = 45$ years ($t = 145$ years). This is the red star in Fig. 4a. 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 $\Delta t = 250$ years ($t = 350$ years). During each year in this period, to a good approximation, there is a 1% chance of having such an earthquake.

Similarly, Figs. 3b and 4b indicate that there is a 75% chance of having a great earthquake with $m_{SF} \geq 7.3$ in the next $\Delta t = 250$ years, a 50% chance in the next $\Delta t = 180$ years (the red star in Fig. 4b), and a 25% chance in the next $\Delta t = 75$ years. To a good approximation, there is a 0.3% chance of having such an earthquake during each year in this period.

For the past fifteen years a purely statistical approach has been used by the Working Group on California Earthquake Probabilities to make risk assessments for northern California. Their statistical approach is a complex 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. This model is used in turn as an input to a procedure which utilizes an assumed probability density function to characterize the segments of each fault that is likely to rupture in an earthquake, as well as the timing and frequency of rupture on the segments.

In their most recent study (*24*) the WGCEP 2003 utilized the Brownian passage time probability distribution (also known as an inverse Gaussian distribution). The mean and standard deviations of the distributions for event times on the fault segments were constrained by geological and seismological observations. The statistical weighting factors and observations selected for use are determined by "expert opinion" formed through a consensus-building procedure involving group voting. The fundamental assumption is that the correct forecast is likely to lie among the various ideas and opinions expressed by the group.

In applying these methods to the northern San Andreas Fault, the WGCEP 2003 divided

the section that ruptured in 1906 into four parts on the basis of geological data: (i) SAS (San Andreas South) from roughly San Juan Bautista to San Jose; (ii) SAP (San Andreas Peninsula) from San Jose to San Francisco; (iii) SAN (San Andreas North) from San Francisco to just north of Pt Arena; (iv) SAO (San Andreas Ocean) from Pt Arena to Cape Mendocino. We note that the "Northern San Andreas Fault" in our analysis (yellow line) includes approximately SAS + SAP + southern half of SAN, so the forecast probabilities found by the WGCEP 2003 are not directly comparable to the forecasts computed by our method. Using their forecast algorithm, the WGCEP 2003 found that, for earthquakes having $M \geq 6.7$ during the 30 year period 2002–2031 is 18.2%.

There are major differences between the simulation-based forecasts given in this paper, and the statistical forecasts given by the WGCEP 2003. In our approach, it is not necessary to prescribe a probability distribution of inter-event times. The distribution of event intervals is obtained directly from our simulations, which include the physics of fault interactions and dynamics. Since both methods use the same database for mean fault slip on fault segments, they give approximately equal mean inter-event times. The major difference between the two methods lies in the way in which inter-event times and probabilities for joint failure of multiple segments are computed. In our simulation approach, these times and probabilities come from the modeling of fault interactions. In the WGCEP 2003 statistical approach, times and probabilities are embedded in the choice of an applicable probability distribution function, as well as choices associated with a variety of other statistical weighting factors describing joint probabilities for multi-segment events.

It is important to note that our results represent an initial probabilistic forecast of the next great San Francisco earthquake. The accuracy of our forecast depends on the degree to which our simulations include the essential features of the fault interactions and by the quality and quantity of the data available to constrain the parameters for each fault segment in the model.

If different slip rates are prescribed the mean waiting times will change accordingly. However, the statistical distribution of waiting times is likely to remain unchanged. This variability is the result of fault interactions and the simulations reported here are the first time these interactions have been modeled. The statistical distribution of waiting times has long been a subject of controversy. The Weibull distribution utilized here is one of a number of distributions previously proposed (*23*). If $\beta = 1$ the Weibull distribution of waiting times reduces to a Poisson distribution. Waiting times are independent of the time since the last earthquake (no memory). In the limit $\beta \to \infty$ the intervals are constant and earthquakes are periodic. We find $\beta = 1.67$ for $m_{SF} \geq 7.0$ and $\beta = 2.17$ for $m_{SF} \geq 7.3$. The larger earthquakes are more periodic and less random. The validity of the Weibull distribution places important constraints on future probabilistic assessment of the earthquake hazard.

References and Notes

1. E. R. Engdahl, A. Villasenor, *International Handbook of Earthquake and Engineering Seismology*, W. H. K. Lee, H. Kanamori, P. C. Jennings, C. Kisslinger, eds. (Academic Press, Amsterdam, 2002), vol. Part A, pp. 665–690.

2. W. Thatcher, *J. Geophys. Res.* **80**, 4862 (1975).

3. C. Demets, R. G. Gordon, D. F. Argus, S. Stein, *Geophys. Res. Lett.* **21**, 2191 (1994).

4. N. T. Hall, R. H. Wright, K. B. Clahan, *J. Geophys. Res.* **104**, 23215 (1999).

5. D. L. Turcotte, *Fractals and Chaos in Geology and Geophysics* (Cambridge University Press, Cambridge, 1997), second edn.

6. K. Kodera, K. Matthes, K. Shibata, U. Langematz, Y. Kuroda, *Geophys. Res. Lett.* **30** (2003).

7. C. Covey, *et al.*, *Glob. Planet. Change* **37**, 103 (2003).

8. E. N. Lorenz, *J. Atmos. Sci.* **20**, 130 (1963).

9. J. B. Rundle, *J. Geophys. Res.* **93**, 6255 (1988).

10. P. B. Rundle, *et al.*, *Phys. Rev. Lett.* **8714**, Art. No. 148501 (2001).

11. J. B. Rundle, *et al.*, *Pure Appl. Geophys.* **159**, 2357 (2002).

12. J. B. Rundle, P. B. Rundle, A. Donnellan, G. Fox, *Earth Planets Space* **56**, 761 (2004).

13. Us geological survey earthquake fault data. Http://eqhazmaps.usgs.gov/html/faults.html.

14. Southern california earthquake center. Http://www.scec.org.

15. Jet propulsion laboratory quakesim project. Http://www-aig.jpl.nasa.gov/public/dus/quakesim.

16. S. N. Ward, *Bull. Seismol. Soc. Amer.* **90**, 370 (2000).

17. *Boundary Elements XXIV. 24th International Conference on Boundary Element Methods* (WIT Press, 2002), p. 754.

18. H. Kanamori, D. L. Anderson, *Bull. Seismol. Soc. Amer.* **65**, 1073 (1975).

19. D. Massonnet, *et al.*, *Nature* **364**, 138 (1993).

20. S. G. Wesnousky, C. H. Scholz, K. Shimazaki, T. Matsuda, *Bull. Seismol. Soc. Amer.* **74**, 687 (1984).

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

22. K. Sieh, M. Stuiver, D. Brillinger, *J. Geophys. Res.* **94**, 603 (1989).

23. D. Sornette, L. Knopoff, *Bull. Seismol. Soc. Amer.* **87**, 789 (1997).

24. Earthquake probabilities in the san francisco bay region, *Tech. Rep. USGS Open File Report 03-214* (2003).

25. This work has been supported by a grant from US Department of Energy, Office of Basic Energy Sciences to the University of California, Davis (JBR; PBR); and under additional funding from National Aeronautics and Space Administration under grants to the Jet Propulsion Laboratory, the University of California, Davis, and the University of Indiana (JBR; PBR; AD). We are also grateful to the Southern California Earthquake Center for partial support.

Fig. 1. Faults segments making up Virtual California. The model has 650 fault segments, each approximately 10 km in length along strike and a 15 km depth. The yellow and blue segments make up the San Andreas fault. In this paper we consider earthquakes only on the yellow "San Francisco" segment of the San Andreas fault.

Fig. 2. Interferometric patterns of the coseismic deformations associated with two sets of model earthquakes. Each interferometric fringe corresponds to a displacement of 56 mm.

Fig. 3. (a) The conditional cumulative probability $P(t, t_0)$ that a great $m_{SF} \geq 7.0$ earthquake will occur on the San Andreas Fault near San Francisco at a time 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 the fits to the data of the Weibull distribution. First, the best fit of Eq. (1) to the complete distribution of interval times ($t_0 = 0$) is obtained taking $\beta = 1.67$ and $\tau = 117$ years. These values are then substituted into Eq. (2) taking $t_0 = 25$, 50, ..., 150 years. The Weibull fits are shown as colored curves. (b) Results for $m_{SF} \geq 7.3$. In this case we take $t_0 = 0$, 50, 100, 150, 200, and 250 years. The best fit of Eq. (1) to the complete distribution of interval times ($t_0 = 0$) requires $\beta = 2.17$ and $\tau = 289$ years.

Fig. 4. The stars (corresponding to the 50% probability of the distributions in Fig. 3) and the green solid line give the median waiting times until the next great earthquake as a function of the time t_0 since the last great earthquake. The red triangle is the median waiting time (50% probability) from today. The yellow band represents waiting times with 25% probability (lower edge of yellow band) to 75% probability (upper edge of yellow band). The dashed red lines are the forecast using the Weibull distribution in Eq. (2) (a) for $m_{SF} \geq 7.0$ and (b) for $m_{SF} \geq 7.3$.

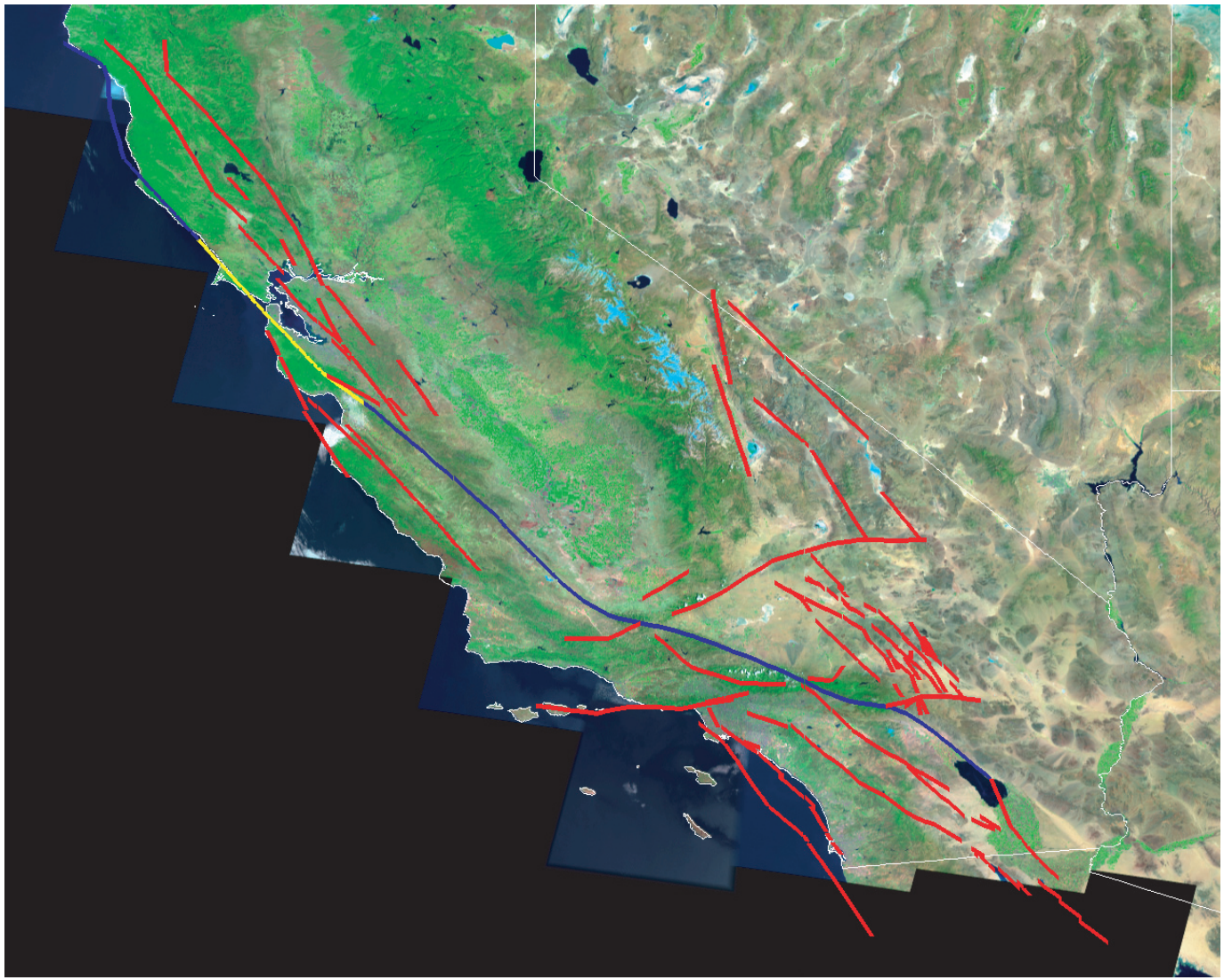

Figure 1. Rundle et al., Science 2004

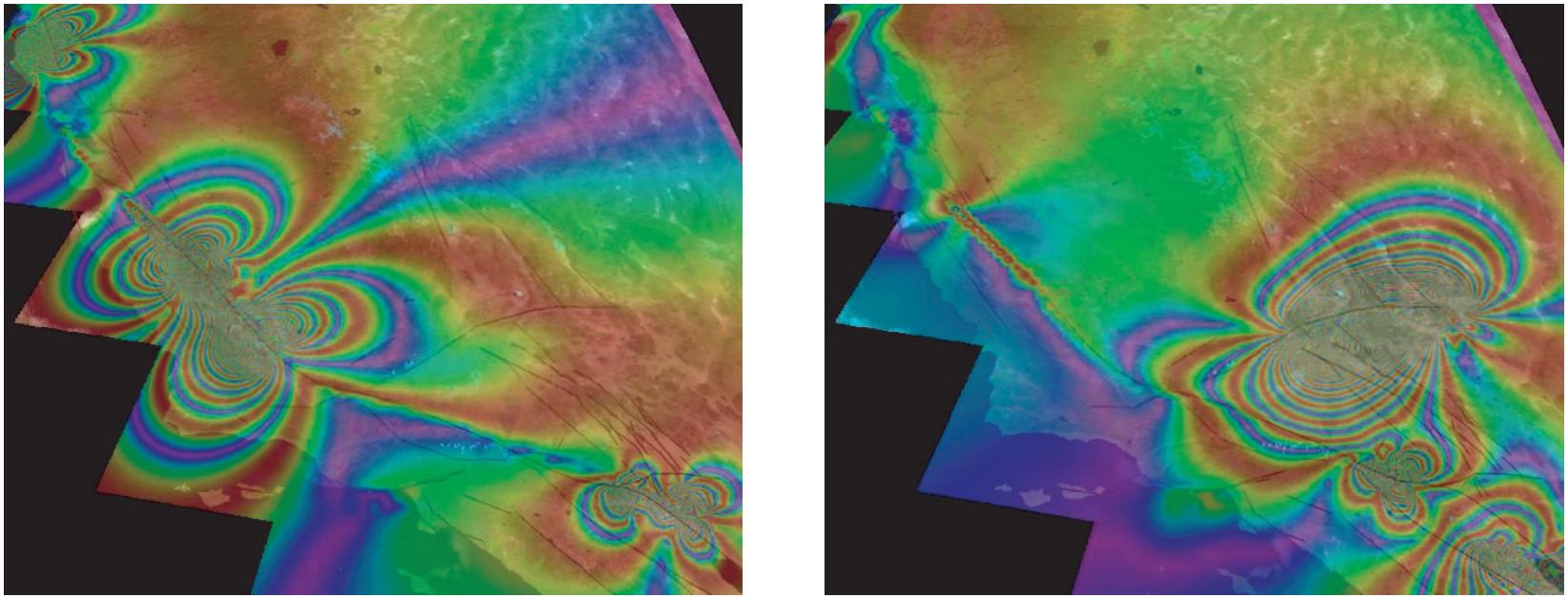

Figure 2. Rundle et al., Science 2004

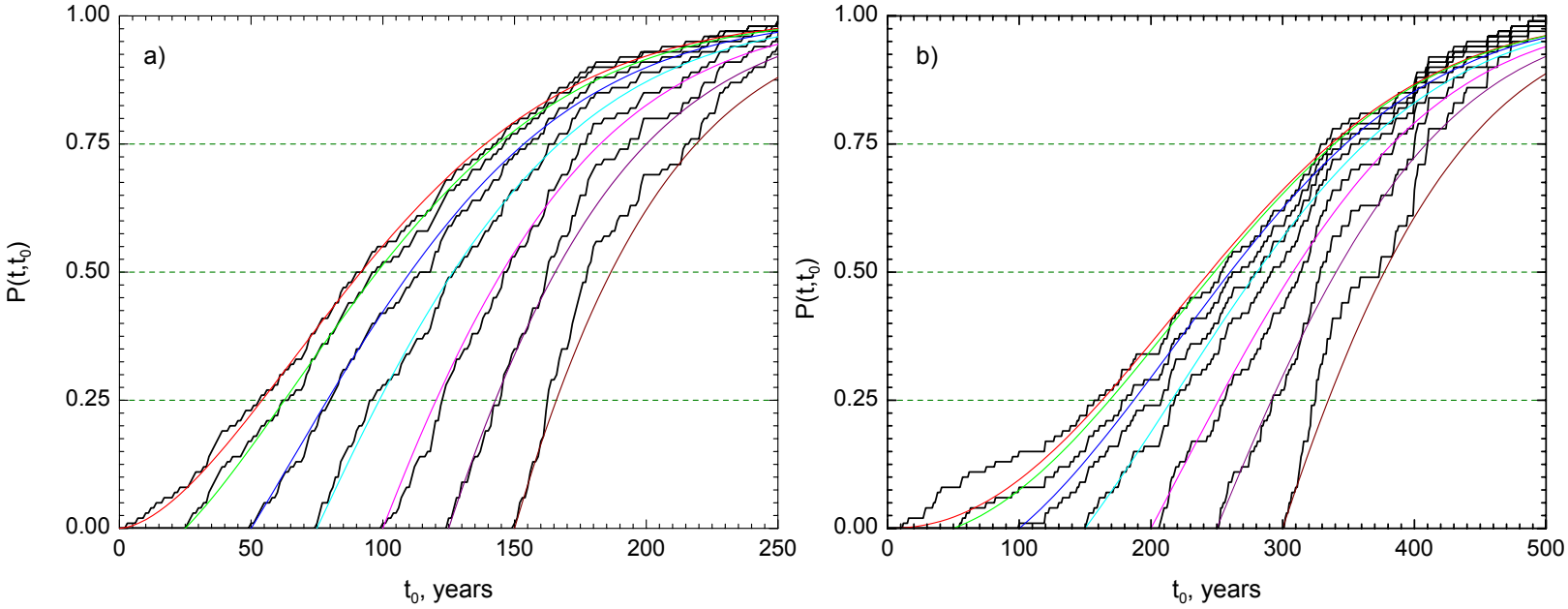

Figure 3. Rundle et al., Science 2004

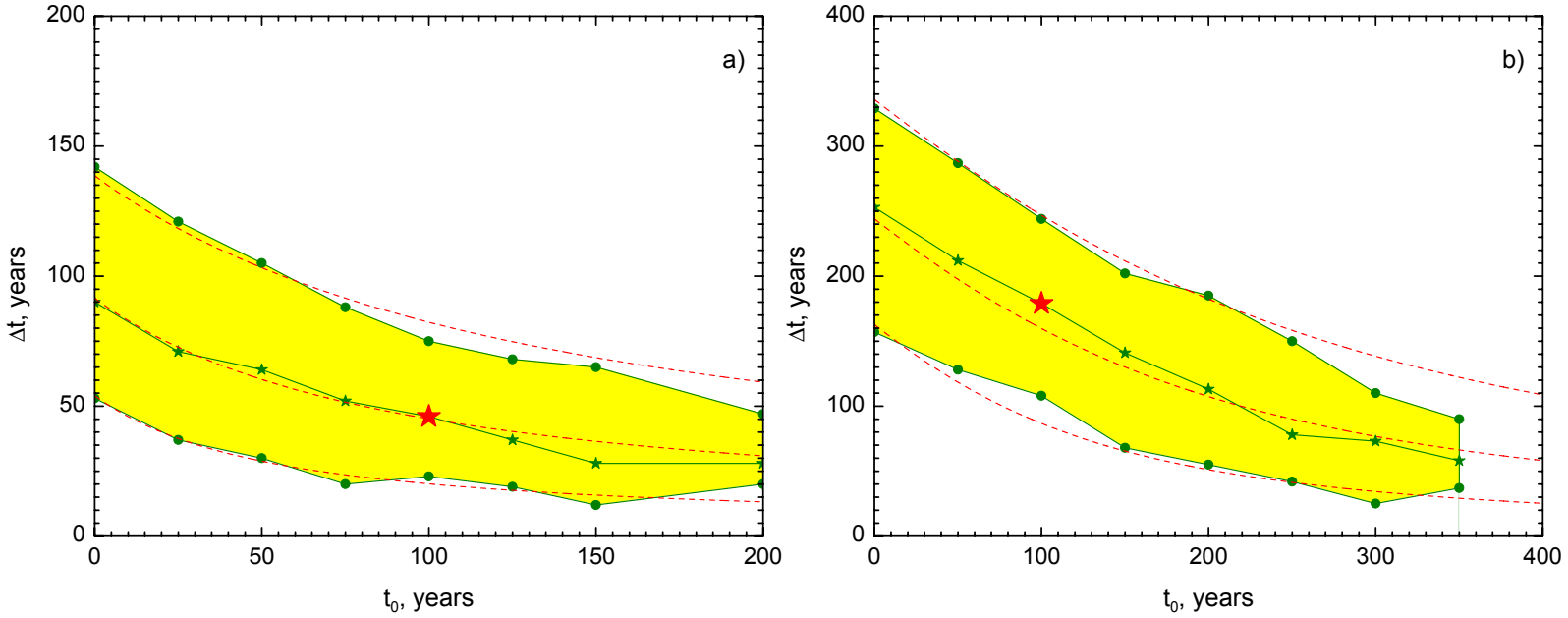

Figure 4. Rundle et al., Science 2004