--DRAFT – August 12, 2004 --

Forecasting the Next Great San Francisco Earthquake

John B. Rundle[1], Paul B. Rundle[2] , Andrea Donnellan[3], Don Turcotte[4], Peggy Li[5] , Bruce Malamud[6], Lisa Grant[7], Geoffrey Fox[8], Dennis McLeod[9], Gleb Morein[2], Jay Parker[5], and W Klein[10]

Abstract. At 5:12 am PDT on April 18, 1906 the great San Francisco earthquake and fire killed more than 3000 persons, destroying the city in the process. This was the worst natural disaster to strike the continental United States in its history. As we approach the 100 year anniversary of that event, the question of when the next great San Francisco earthquake will occur is of critical concern. In this paper we present a new probabilistic risk analysis for future great earthquakes, based on state of the art computer simulation technology, augmented by recent statistical physics insights into the basic dynamical processes. Our results indicate that there is a 55% chance of an earthquake with magnitude 7.0 or larger, and a 13% chance of an earthquake with magnitude 7.3 or larger, occurring on the San Andreas fault near San Francisco during the next 50 years. During the next 5 years we find a 5% chance of such an earthquake with magnitude 7.0 or larger, and a 1.5% chance of a magnitude 7.3 or larger event.

---

[1]Center for Computational Science and Engineering, University of California, Davis, Davis, CA 95616; and Distinguished Visiting Scientist, Earth & Space Sciences Division Jet Propulsion Laboratory, Pasadena, CA 91125

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

[3]Earth & Space Science Division, Jet Propulsion Laboratory, Pasadena, CA 91125

[4]Department of Geology, University of California, Davis, CA 95616

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

[6]Department of Geography, Kings College, London, UK

[7]Department of Environmental Health, Science and Policy, University of California, Irvine, CA 92697

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

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

[10]Department of Physics, Boston University, Boston, MA 02215

1. Introduction

The great San Francisco earthquake occurred on a 470 km segment of the San Andreas fault from San Francisco Bay north to Cape Mendocino. With the acceptance of the plate tectonic hypothesis in the 1960's, it was recognized that 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. Since the observations of surface displacements across the fault near the city during the San Francisco earthquake was in the range of 2.0-5.0 m [1], the time required to accumulate this displacement is approximately 40-100 years. The simplest hypothesis for the recurrence of great San Francisco earthquakes is that they will occur at approximately these time intervals, indicating that the next earthquake may be imminent.

However, there are two problems with this simple 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. A variety of studies indicate that the mean displacement rate on the northern San Andreas fault itself is closer to 24 mm/yr. The second and more serious problem with the simple periodic hypothesis involves the existence of complex interactions between the San Andreas fault and other faults in the system. It is now recognized 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, but an approach based purely on a statistical formulation cannot properly account for the effects of the interactions. It is the purpose of this paper to make such a forecast, utilizing direct numerical simulations of fault system physics that include these complex fault interactions.

Our simulation, *Virtual California* [2], has been developed to include stress accumulation, fault interactions, and stress release on active earthquake fault systems. The major faults in the model are only those that have been the most active in recent geologic history. Earthquake activity data and slip rates on these models faults are obtained from geologic data bases maintained by the US Geological Survey [3], the Southern California Earthquake Center [4], and the QuakeSim project [5]. Similar types of simulations have been developed by other workers [6].

The *Virtual California* model is a *backslip* model in that 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 vertically-oriented rectangular fault segments are embedded in an elastic half space, and interact by means of quasi-static elastic interactions. The basic physics is stick-slip, in which coefficients of static and kinetic friction are used along with the space- and time-dependent shear stress on fault segments. Both shear and normal stresses on segments are computed by means of Boundary Element Methods (BEM) [7]**.** To set the stick-slip friction coefficients, we use the history of known major earthquakes having $M \geq 5.0$ in California during the last ~ 200 years. A consequence of our finite fault segment sizes is that our simulations do not generate small earthquakes having magnitudes M less than about $M \cong 5.8$.

The topology of the Virtual California model is shown in figure 1 superposed on a LandSat image. The 650 strike-slip fault segments are represented by lines, the majority of which are red. The blue and yellow line represents the San Andreas fault, stretching from Cape Mendocino in the north to the Salton trough in the south. The yellow line represents the "San Francisco section" of the fault, a distance of approximately 250 km, and is the portion of the fault whose rupture would be strongly felt in San Francisco. Our goal is to forecast an estimate of the waiting time until the next earthquake on the yellow section of the fault for two examples, with magnitudes 1) an $M_{SF} \geq 7.0$ event; or 2) an $M_{SF} \geq 7.3$ event. We will define such an event as a "Great Earthquake". Using standard seismological relationships [8], we can compute that an event with $M_{SF} \geq 7.0$, having average slip of 4 m and depth of faulting of 15 km, would rupture approximately 20 km length of fault. With similar conditions, an event of $M_{SF} \geq 7.3$ would rupture 66 km of fault. Earthquakes like these would produce great damage in San Francisco.

Using Virtual California, we simulated 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 primarily a result of the fault interactions. Slip events in the simulations display highly complex behavior, with no obvious regularities or predictability.

For the section of the northern San Andreas fault shown in yellow on figure 1, we compiled data on all the events occurring on those segments that had 1) $M_{SF} \geq 7.0$ event; or 2) $M_{SF} \geq 7.3$ event. For the former, we obtained 395 events having an average recurrence interval of 101 years, and for the latter, 159 events having an average recurrence interval of 249 years. We then measured the time interval between successive events in each case.

From these time intervals, we construct probabilistic forecasts of when the next great earthquake will occur if the last great earthquake occurred a time interval T_{NO} years ago. In order to do this, we remove the events that have occurred during the period T_{NO} after the earthquake and consider only the remaining events in constructing the statistical distribution. This distribution is then conditioned on the assumption that no earthquake has occurred in the period T_{NO} after the last earthquake. We then plot the cumulative number of earthquakes that will have occurred at a time T_F in the future. This gives the cumulative conditional probability $P(T_F \mid T_{NO})$ that a great earthquake will occur during a time interval T_F in the future, if the last earthquake occurred at a time interval T_{NO} in the past.

Our results for great earthquakes with $M_{SF} \geq 7.0$ are given in figure 2, and our results for great earthquakes with $M_{SF} \geq 7.3$ are given in figure 3. In figures 2a and 3a the cumulative probabilities $P(T_F \mid T_{NO})$ are given as a function of the time since the last great earthquake T_{NO}.

Immediately after a great earthquake, i.e., in 1906, we have $T_{NO} = 0$. At that time, figure 2a indicates that there was a 50% chance of having an earthquake $M_{SF} \geq 7.0$ in the next t = 90 years, i.e., in 1996. Also at that time ($T_{NO} = 0$), there was a 50% chance of having an earthquake with $M_{SF} \geq 7.3$ in the next t = 249 years, as shown in figure 3a. Note that the vertical dashed lines in figures 2a and 2b give the mean time intervals between great earthquakes based on our simulations. However, it is now 98.4 years since the last great earthquake that occurred in 1906. The cumulative distributions for forecasts as shown in red in figures 2a and 3a correspond to this case (T_{NO} = 98.4 years). We see from figure 2a that there is a 50% chance of having a great earthquake ($M_{SF} \geq 7.0$) in the next T_F = 45 years (i.e., total time since 1906 would be $T_{NO} + T_F = 98.4 + 45$ years = 144.4 years). It can also be seen that there is a 25% chance for such an earthquake in the next T_F = 20 years, and a 75% chance in the next T_F = 75 years. To a good approximation, there is thus a 1% chance of having such an earthquake during each year in this period. If the 1989 $M_{SF} \cong 7.0$ Loma Prieta earthquake is considered to fall into this category, then the appropriate curve in figure 2a would correspond to T_{NO} = 15.4 years.

Similarly, figure 3a indicates that there is a 50% chance of having a great earthquake with $M_{SF} \geq 7.3$ in the next 180 years, a 25% chance of in the next 75 years, and a 75% chance in the next 250 years. To a good approximation, there is about a .3% chance of having such an earthquake during each year in this period.

The results given in figures 2a and 3a are presented in a different way in figures 2b and 3b. The median waiting times to the next great earthquake are given by the green dashed line and blue dots, the latter correspond to the conditional distributions shown in figures 2a and 3a. The median waiting times (T_F for 50% probability) are shown as a function of the time T_{NO} since the last great earthquake. The forecast waiting time for today is given by the red diamonds. As described above for $T_{NO} = 0$, the forecast median waiting time for a $M_{SF} \geq 7.0$ earthquake from figure 2b is 90 years, and for $M_{SF} \geq 7.3$, figure 3c similarly gives a value of 180 years. The yellow bands given in figures 2b and 3b represent the middle 50% values, showing waiting times with probabilities 25% (lower boundary) and 75% (upper boundary).

It is important to note that our results represent an initial estimate for the occurrence of the next great San Francisco earthquake. The accuracy of our forecasts depends on the degree to which our simulations include the essential features of the fault interactions, but are limited primarily by the quality and quantity of the data available to set the parameters for each fault segment in the model. It should be recognized that the latter problem will afflict all earthquake forecast methods.

For the past 15 years a purely statistical approach has been used by the Working Group on California Earthquake Probabilities (the most recent of these is "WGCEP 2003"; [9]) 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 uses 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, the WGCEP 2003 emphasized the Brownian Passage Time probability distribution function. The mean and standard deviations of the distributions for event times on the fault segments were obtained from geological and seismological observations. The latter variations are thought to be associated with a depressed rate of large earthquake occurrence in the Bay region in the time period since the 1906 earthquake occurred, relative to the more elevated rate of occurrence in the decades leading up to the 1906 event. In the WGCEP 2003 method, choice of the appropriate PDF, as well as statistical weighting factors and observations selected for use, are determined by "expert opinion" formed through a consensus-building procedure involving group voting (refer to [9] for a discussion of this process). 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 divide the section that ruptured in 1906 into four parts on the basis of geological data. These parts were “San Andreas South” (SAS), from roughly San Juan Bautista to San Jose; “San Andreas Peninsula” (SAP) from San Jose to San Francisco; “San Andreas North” (SAN) from San Francisco to just north of Pt Arena; and “San Andreas Ocean” from Pt Arena to Cape Mendocino. Geological data from the 1906 earthquake suggest that each part had an approximately constant slip. We note that the “Northern San Andreas fault” in our analysis (yellow line) includes approximately the SAS + SAP + southern half of SAN, so that 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 years 2002-2031, the mean rupture probabilities on their fault segments are: SAS part 11.3%; SAP part, 13.3%; SAN part, 11.6%; and SAO part, 10.7%. In addition, the WGCEP 2003 finds that SAS has a 2.6% chance of rupturing in a $M = 7.03$ event; SAP has a 4.4% chance of rupturing in a $M = 7.15$ event; SAN has a .9% chance of rupturing in a $M = 7.45$ event; and SAO has a .9% chance of rupturing in a $M = 7.29$ event. The WGCEP 2003 also finds that for the combination SAS+SAP+SAN+SAO, the mean rupture probability for events with $M \geq 6.7$ is 23.5%; for events with $M \geq 7.0$ it is 18.2%; and for events with $M \geq 7.5$ it is 9%. For a 30 year period our forecast for a rupture with $M_{SF} \geq 7.0$ from figure 2 is 35% and our forecast for a rupture with $M_{SF} \geq 7.3$ is 5%. For a 5 year period, which is of interest to engineering planners, the forecast for a rupture with $M_{SF} \geq 7.0$ from figure 2 is 5%, and for a rupture with $M_{SF} \geq 7.3$ is .5%

It should be emphasized that there are major differences between the simulation-based forecasts given in this paper, and the statistical forecasts previously developed 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 (measured) directly from our simulations, which include the physics of fault interactions and dynamics. Both methods give approximately equal mean inter-event times, because both use the same data base for mean fault slip on fault segments. The major difference between the two models 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. Both methods must also treat earthquakes occurring on un-modeled fault segments as perturbations on the statistical distributions used for the forecasts.

Simulation-based approaches to forecasting and prediction of natural phenomena have been used with great success for weather and climate events. The latter are often referred to as *General Circulation Models* [11]. 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 dynamics [12], General Circulation

Models have repeatedly shown their value in forecasting the nightly weather [13], landfalls of hurricane storm tracks [14], and the onset of ENSO events months in advance[12].

Acknowledgements. 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.

References

[1] Thatcher, W, Strain accumulation and release mechanism of the 1906 San Francisco earthquake, *J. Geophys. Res*., **80**, 4862-4872 (1975).

[2] Rundle, JB, A physical model for earthquakes, 2. Application to southern California, *J. Geophys. Res*., **93**, 6255 – 6274 (1988). Rundle, P.B., Rundle, J.B., Tiampo, K.F., Martins, J.S.S., McGinnis, S., and Klein, W., Nonlinear network dynamics on earthquake fault systems, *Phys. Rev. Lett*., **87**, 148501(1-4) (2001). JB Rundle, PB Rundle, W Klein, J Martins, KF Tiampo, A Donnellan and LH Kellogg, GEM plate boundary simulations for the Plate Boundary Observatory: Understanding the physics of earthquakes on complex fault systems, *PAGEOPH*, **159**, 2357-2381 (2002). Rundle, JB, PB Rundle, A Donnellan and GC Fox, Gutenberg-Richter statistics in topologically realistic system-level earthquake stress-evolution simulations, *Earth, Planets and Space,* in press (2004).

[3] US Geological Survey earthquake fault data http://geohazards.cr.usgs.gov/eq/faults/fsrpage01.html.

[4] Southern California Earthquake Center (SCEC) home page, http://www.scec.org/

[5] Jet Propulsion Laboratory QuakeSim project home page, http://www-aig.jpl.nasa.gov/public/dus/quakesim/

[6] Ward, SN and SDB Goes, How regularly do earthquakes recurr? A synthetic seismicity model for the San Andreas fault, *Geophys. Res. Lett*., **20**, 2131-2134 (1993). Ward, S.N., San Francisco bay area earthquake simulations, a step towards a standard physical model, *Bull. Seism. Soc. Am*., **90**, (2000) 370-386.

[7] See for example: Boundary Elements XXIV. 24th International Conference on Boundary Element Methods. *WIT Press*, 754 pp., Southampton, UK. (2002). Denda M, Dong YF, A unified formulation and error estimation measure for the direct and the indirect boundary element methods in elasticity, Eng. Anal. with Boundary Elements, **25**, 557-64, (2001). Funken SA, Coupling of mixed finite element methods and boundary element methods in elasticity, *Zeitsch. Angewandte Math. Mech*., **80**, suppl., 3, S833-4, pp.S833-4, (2000). Jorge AB, Ribeiro GO, Fisher TS, New approaches for error estimation and adaptivity for 2D potential boundary element methods, *Int. J. Num.Meth. Eng*. **56**, 117-44, (2003). Mackerle J. FEM and BEM in the context of information retrieval, *Comptrs. & Struct*., **80**, 1595-604, (2002).

[8] Kanamori, H. and D.L. Anderson, Theoretical basis of some empirical relations in seismology, Bull. Seism. Soc. Am., 65, 1073-1096, (1975).

[9] Toppozada TR, Branum DM, Reichle MS, Hallstrom CL, San Andreas fault zone, California: M>or=5.5 earthquake history, *Bull. Seism. Soc. Am*., **92**, 2555-601 (2002).

[10] Working Group on California Earthquake Probabilities, Probabilities of Large Earthquakes Occurring in California on the San Andreas fault, US Geological Survey Open-File Rept. 88-398, 1988. Working Group on California Earthquake Probabilities, Probabilities of Large Earthquakes in the San Francisco Bay Region, California, US Geological Survey Circular 1053, 1990. Working Group on California Earthquake Probabilities, Earthquake Probabilities in the San Francisco Bay Region: 2000 to 2030-a Summary of Findings, US Geological Open File Report 99-517, 1999. Working Group on California Earthquake Probabilities, Earthquake Probabilities in the San Francisco Bay Region: 2002-2031, US Geological Open File Report 03-214, 2003.

[11] See for example: Kodera K, Matthes K, Shibata K, Langematz U, Kuroda Y, Solar impact on the lower mesospheric subtropical jet: a comparative study with general circulation model simulations, *Geophys. Res. Lett.*, **30**, 48-1-4, (2003); Covey C, AchutaRao CM, Cubasch U, Jones P, Lambert SJ, Mann ME, Phillips TJ, Taylor KE, An overview of results from the Coupled Model Intercomparison Project, *Global & Planetary Change*, **37**, 103-33 (2003).

[12] Lorenz, EN, Deterministic nonperiodic flow, J. Atmos. Sci., 20, 130-41 (1963).

[13] Hersbach H, Mureau R, Opsteegh JD, Barkmeijer J, Developments of a targeted ensemble prediction system, *Quart. J. Roy. Met. Soc.*, **129**, 2027-48, (2003); Fourrie N, Thepaut J-N. Evaluation of the AIRS near-real-time channel selection for application to numerical weather prediction, *Quart. J. Roy. Met. Soc*., **129**, 2425-39, (2003).

[14] Zhang Z, Krishnamurti TN, Ensemble forecasting of hurricane tracks, *Bull. Am. Met. Soc.*, **78**, 2785-95 (1997). Dengler K, A numerical study of the effects of land proximity and changes in sea surface temperature on hurricane tracks, Quart. J. Roy. Met. Soc., **123**, 1307-21 (1997).

[15] Penland, C and Sardeshmukh PD, The optimal growth of tropical sea surface temperature anomalies, *J. Climate*, **8**, 352-360 (1995); Pavia EG, Secondary forecast models-the ENSO example, *J. App. Met*., 39, 1952-5 (2000). Tziperman E, Zebiak SE, Cane MA, Mechanisms of seasonal-ENSO interaction, J. Atmos. Sci., **54**, 61-71 (1997).

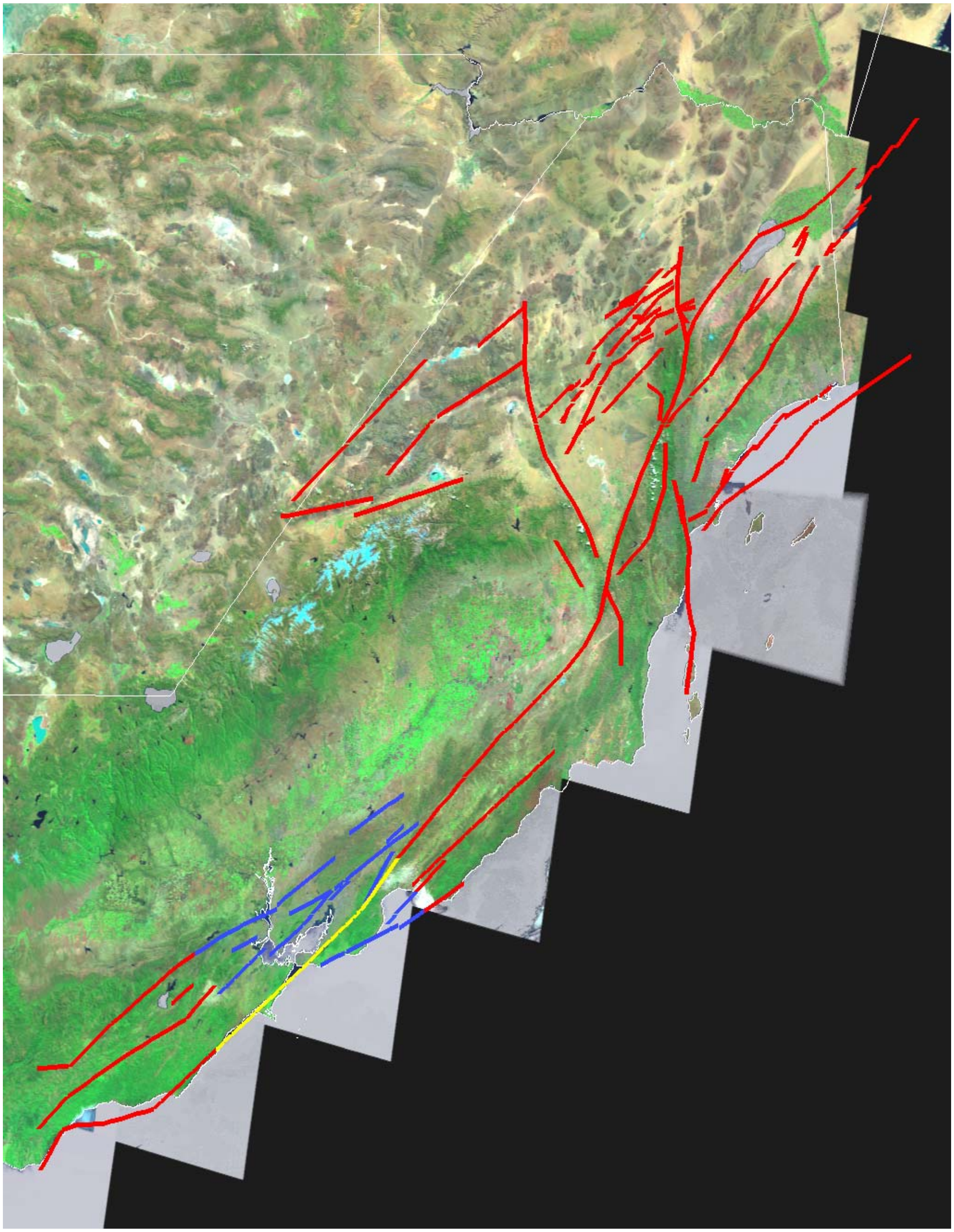
Figure Captions

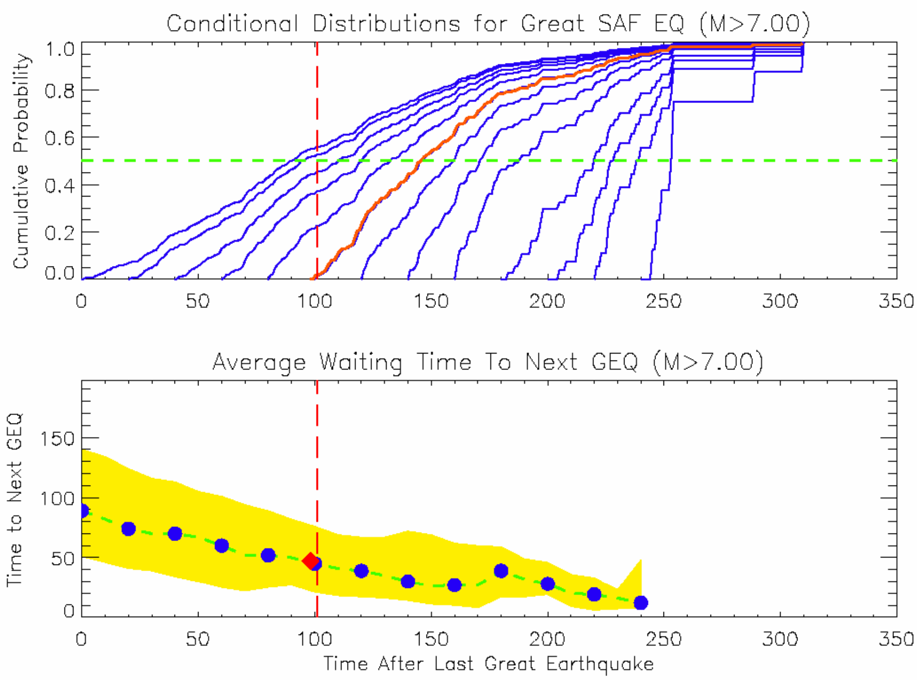
Figure 1. Faults segments making up the Virtual California model. Model has 650 fault segments, each approximately 10 km in length along strike, and 15 km depth. Blue faults are Bay Area faults, yellow fault is Bay Area trace of the San Andreas fault, and red faults are the remaining faults in the model.

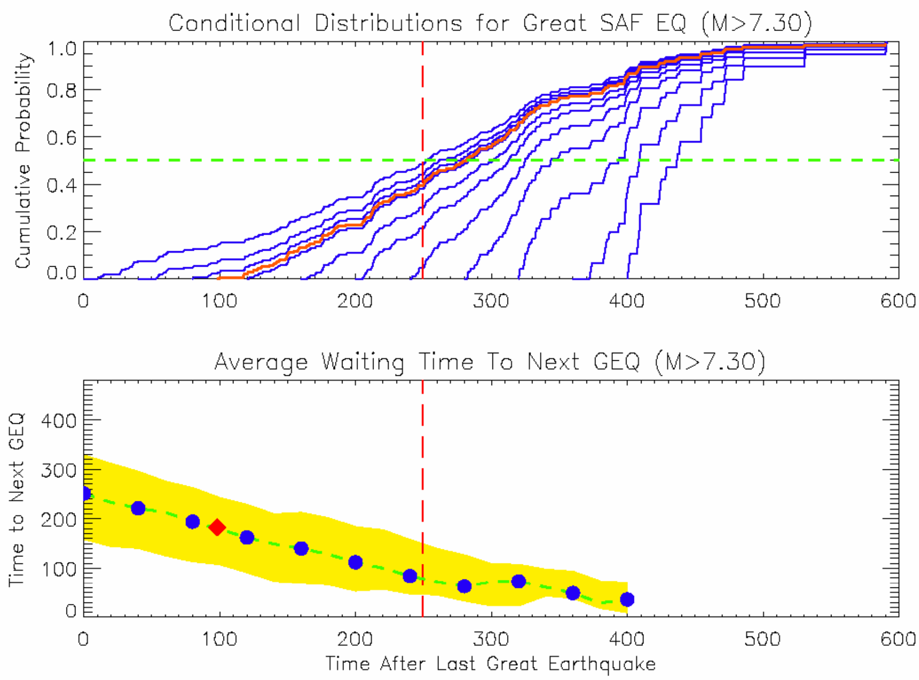
Figure 2. (a) The conditional cumulative probability $P(T_F|T_{NO})$ that a great $M_{SF} \geq 7.0$ earthquake will occu on the San Andreas fault near San Francisco during a time interval T_F in the future if the last great earthquake occurred at a time interval T_{NO} in the past. Results are given for $T_{NO} = 0$ (left distribution), 20, 40, 60, 80, 98.4 (red distribution), 120, 140, 160, 180, 200, 220, and 240 years. The red distribution is applicable to today, because the last great earthquake that indisputably occurred on this section of San Andreas fault was the earthquake of 1906, 98.4 years ago. The vertical red line is the mean waiting time for such a great earthquake when the previous earthquake has just occurred ($T_{NO} = 0$). The intersection of the horizontal green line with each distribution can be used to compute the time corresponding to a cumulative probability value of 50%, which is the median waiting time. If there were such a great earthquake yesterday, there would be a 50% probability that another such earthquake would occur within 90 years.

(b) The blue dots (corresponding to the 50% probability of the distributions in (a) above) and the green dashed line shows the median waiting time until the next great earthquake as a function of the time T_{NO} since the last great $M_{SF} \geq 7.0$ earthquake. The red triangle is the median waiting time from today. The yellow band represents the middle 50% of waiting times, from 25% probability (lower edge of yellow band) to 75% probability (upper edge of yellow band).

Figure 3. Same information as in figure 2, but for great earthquakes with $M_{SF} \geq 7.3$.




Conditional Distributions for Great SAF EQ (M>7.00)
Cumulative Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
50
100
150
200
250
300
350
Average Waiting Time To Next GEQ (M>7.00)
Time to Next GEQ
150
100
50
0
0
50
100
150
200
250
300
350
Time After Last Great Earthquake


Conditional Distributions for Great SAF EQ (M>7.30)
Cumulative Probability
1.0
0.8
0.6
0.4
0.2
0.0
0
100
200
300
400
500
600
Average Waiting Time To Next GEQ (M>7.30)
Time to Next GEQ
400
300
200
100
0
0
100
200
300
400
500
600
Time After Last Great Earthquake