Pure and Applied Geophysics

Space- and Time-Dependent Probabilities for Earthquake Fault Systems from Numerical Simulations: Feasibility Study and First Results

Jordan Van Aalsburg,[1,2] John B. Rundle,[1,2] Lisa B. Grant,[3] Paul B. Rundle,[1,2] Gleb Yakovlev,[2] Donald L. Turcotte,[4] Andrea Donnellan,[5] Kristy F. Tiampo,[6] and Jose Fernandez[7]

Abstract—In weather forecasting, current and past observational data are routinely assimilated into numerical simulations to produce ensemble forecasts of future events in a process termed "model steering". Here we describe a similar approach that is motivated by analyses of previous forecasts of the Working Group on California Earthquake Probabilities (WGCEP). Our approach is adapted to the problem of earthquake forecasting using topologically realistic numerical simulations for the strike-slip fault system in California. By systematically comparing simulation data to observed paleoseismic data, a series of spatial probability density functions (PDFs) can be computed that describe the probable locations of future large earthquakes. We develop this approach and show examples of PDFs associated with magnitude $M > 6.5$ and $M > 7.0$ earthquakes in California.

Key words: Earthquakes, forecasting, California seismicity, earthquake hazard.

1. Introduction

In a series of reports, the Working Group on California Earthquake Probabilities (WGCEP) have computed probabilities of major earthquakes on California faults over a 30-year period.[1,2] These forecasts are used to set insurance rates and by emergency response planners and policymakers. A review of the reports (FIELD, 2007) describes common features, differences and assumptions of these studies. FIELD (2007) concludes by advocating the use of numerical simulation-based approaches to the problem of multi-decadal earthquake forecasting. An analogy may be drawn to weather and climate forecasting. Weather and seismicity are both complex, chaotic phenomena. Current weather patterns are routinely extrapolated to forecast several days into the future. These forecasts utilize numerical simulations of atmospheric behavior. Here we develop a similar approach by using *Virtual California*, a topologically realistic numerical simulation of strike-slip faults in California, to develop a series of spatial probability density functions (PDFs) that describe the probable locations of future large earthquakes.

2. The WGCEP Approach

As summarized by FIELD (2007), the WGCEP approach has been to (1) define a series of geological fault segments; (2) use paleoseismic and other data to determine the mean earthquake recurrence interval on each segment; (3) assume a set of statistical distributions to describe the recurrence statistics; (4) compute the probability of multi-segment ruptures,

[1] Department of Physics, University of California, Davis, USA. E-mail: jvan@cse.ucdavis.edu
[2] Computational Science and Engineering Center, University of California, Davis, USA.
[3] Program in Public Health, University of California, Irvine, CA 92697-3957, USA.
[4] Geology Department, University of California, Davis, USA.
[5] Jet Propulsion Laboratory, Pasadena, CA, USA.
[6] Department of Earth Science, University of Western Ontario, London, ON, Canada.
[7] Instituto de Astronomía y Geodesia (CSIC-UCM), Facultad CC Matemáticas, Ciudad Universitaria, Plaza de Ciencias, 3, 28040 Madrid, Spain.

[1] http://www.pubs.usgs.gov/of/2003/of03-214/WG02_OFR-03-214_ExecSummary.pdf.
[2] http://www.wgcep.org/.

assuming statistical independence of fault segments; and (5) adjust the results to reflect the time-dependence of the earthquake cycle (FIELD, 2007). The result is a set of probabilities for the occurrence of earthquakes $M > 6.5$ over the next thirty years.

The WGCEP approach assumes that earthquakes occur on geologically-defined fault segments, that earthquake ruptures rarely jump between fault segments, and that earthquake clustering can be discounted (FIELD, 2007). However, earthquake clustering is an established consequence of earthquake dynamics (MARCO et al. 1996; ZHUANG et al., 2004) and there are recent examples of earthquake ruptures jumping between fault segments, for example the 1992 M 7.3 Landers earthquake (WALD and HEATON 1994) and the 2002 M 7.9 Denali earthquake (EBERHART-PHILLIPS et al., 2003). The methods for including uncertainty in the modeled probabilities are problematic (PAGE and CARLSON, 2006).

3. The Virtual California Simulation Approach

Here we propose a method for computing probabilities using the type of simulation-based approach (RUNDLE, 1988; RUNDLE et al., 2001, 2002, 2004, 2005, 2006; VAN AALSBURG et al., 2007) advocated by FIELD (2007). Virtual California (VC) is a topologically realistic numerical simulation of earthquakes occurring on the San Andreas fault system. It includes the major strike-slip faults in California (Fig. 1). The approach using simulations such as VC is similar to the WGCEP approach. It begins with a series of faults divided into interacting fault elements, and uses paleoseismic and other data to set the frictional properties on each element. We then conduct a series of numerical simulations that attempt to reproduce the statistics and variability of the actual fault system. We search through the simulations to identify sequences of earthquakes that optimally represent the known earthquake history; and use the simulation data to measure the statistics and probabilities for future earthquake occurrence in space and time. The result is a set of probabilities for the occurrence of earthquakes of any size larger than the cutoff over user-selected future time-intervals. The probabilities determined by the simulations are
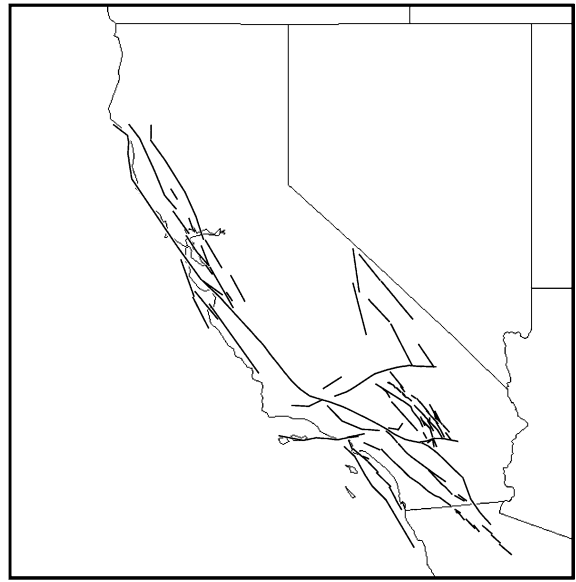

Figure 1
Map of California with the faults used in the Virtual California simulations as shown

time-dependent, implicitly include the effects of fault interactions, and are based on the same published data available to the WGCEP.

For this study, the VC fault model is composed of 650 fault boundary elements, each of 10 km width and 15 km depth. Elastic dislocation theory is applied to model fault element interactions. VC is a "back-slip" model. The accumulation of a slip deficit on each element is prescribed using available paleoseismic and instrumental data so that the long-term rate of slip is matched, on average, by the observed rate of stress accumulation on the faults (SAVAGE and PRESCOTT, 1978; RUNDLE and KANAMORI, 1987; RUNDLE, 1988). The mean recurrence time of earthquakes is determined using available data, to define friction law parameters. The friction law has several parts, including Mohr–Coulomb stick–slip properties; small amplitude, stable aseismic slip that increases as stress increases; and a stress-rate dependent failure criterion based upon laboratory studies of the functional form of the dynamic stress intensity factor. Fault interactions lead to complexity and statistical variability. Earthquake triggering, or initiation, is controlled by friction coefficients along with the space- and time-dependent stresses on fault elements which are computed by boundary element methods. Historical

earthquakes that have moment magnitudes $m \geq 5.0$ during the last 200 years are used to prescribe the friction coefficients. A consequence of the minimum size of the fault elements is that the simulations do not generate earthquakes having magnitudes less than about $m \geq 5.8$. Therefore, additional parameters must be selected by systematic tuning of the model, followed by a search for sequences of events that optimally reproduce the known history of large earthquakes. Similar to the WGCEP approach, accuracy of results in the simulation approach are explicitly constrained by the limited availability of historic and instrumental data on large earthquakes occurring on faults in the model.

4. VC and Assimilation of Paleoseismic Data

Virtual California is an example of a fault simulator, other examples can be found as published by WARD (1992, 1996, 2000), RICHARDS-DINGER and DIETERICH (2008), and ROBINSON (2004). The topology of VC faults is shown in Fig. 1. The San Andreas fault (SAF) is the longest continuous fault and the greatest source of seismic hazard in California. Paleoseismic data from the SAF system provide an unparalleled opportunity for documenting and understanding the multi-cycle rupture history of a major active fault. Paleoseismic data consist of geologic observations of faulting from paleo-earthquakes. Data most commonly reported include characteristics of surface ruptures, number of rupture "events" during a Holocene or Quaternary time interval (resulting in "average recurrence interval" for paleo-earthquakes), date of the most recent earthquake and/or sequence of paleo-earthquakes (with uncertainty), and measurements of surface displacement from paleo-earthquakes (GRANT, 2007).

The relatively rich paleoseismic data set from the SAF system provides an unparalleled opportunity for comparison with results of simulations. The best paleo-earthquake record in North America is from the Wrightwood site on the SAF in southern California (FUMAL et al., 2002; BIASI et al., 2002; WELDON et al., 2004, 2005). There are records of multiple ruptures at several other sites, including ten events at Pallett Creek (BIASI et al., 2002; SIEH et al., 1989), and

Bidart Fan (GRANT et al., 2005), also on the southern SAF. The record of paleo-earthquakes at these sites, which ruptured most recently in A.D. 1857, has formed the primary data set for probabilistic assessments of future southern San Andreas fault earthquakes, and for testing models of fault behavior and earthquake recurrence (WELDON et al. 2005; BIASI et al., 2002). Paleo-earthquake data are also available from the northern SAF and other faults in the SAF system, such as the San Jacinto and Garlock.

Paleoseismic data were compiled and formatted for assimilation into VC simulations in an initial feasibility study (VAN AALSBURG et al., 2007; GRANT, 2007). For this study, we used the same data set as VAN AALSBURG et al. (2007). Our goal is to obtain the statistical distribution of waiting times for simulated large earthquakes on specified faults and fault elements of the SAF system. We advance the VC model in 1 year increments, and simulate 40,000 years of earthquakes on the SAF system. Average slip on the fault elements and average recurrence intervals are tuned to match observed average rupture intervals at paleoseismic study sites. Due to fault interactions, slip events in the simulations display highly complex behavior, with no obvious regularities or predictability. For distinct groups of fault elements, the Weibull distribution represents the statistics of the largest earthquakes in a number of cases reasonably well, with fits to the empirical distribution functions having regression coefficients in excess of 0.99 (YAKOVLEV et al., 2006).

5. "Data Scoring" Methods

VAN AALSBURG et al. (2007) describe a "data scoring" method for identifying time windows in a simulation record that are most similar to the actual paleoseismic record. In any simulation, there are intervals of simulated data that resemble the recent past few hundred years of earthquakes, and periods that are different. If we identify the intervals of simulation data that optimally resemble the recent past, we might hypothesize that the time intervals following these optimal intervals might then possibly characterize future activity on the actual San Andreas fault system.

Two data sets are used for scoring: VC simulation data, and paleoseismic data from the natural SAF system. For simulation data, we are interested in the event—the location corresponding to the latitude and longitude of the fault element, time of slip measured in simulation years, and the amount of slip. The analysis which follows used a catalog containing 200,000 events spanning 40,000 simulation years. The second data of paleoseismic sites and as discussed above, consists of observations dating back 1000 years on the SAF system in California. Unlike simulation data, paleo-earthquake times are known only within a time window ranging from a few years to several hundred years. This data is stored in XML format similar to *VC* simulation data, with each paleoseismic site containing one or more events defined by a minimum and maximum time value. There are 21 paleoseismic sites used to score the Virtual California catalog and a total of 119 observed events (see VAN AALSBURG et al., 2007; Table 1).

The first step in scoring is to associate paleoseismic sites with fault elements in the *VC* model. The association can be as single site-element pair (nearest-neighbor) or can include all *VC* elements within a specified radius (long-range neighborhood). The variable-range neighborhood is implemented because *VC* is only a simple representation of actual faults.

To score a particular simulation year, we consider the "current time" t_{sim} in the simulation record to represent the "present day", $t = 2009$. We then compare the time history prior to t_{score}, i.e., $t < t_{sim}$, to the known history from paleoseismic data. The scoring algorithm proceeds as follows: (1) For each paleoseismic site, we examine each Virtual California element and compare its slip times to the slip times recorded at that paleoseismic site. (2) A score is assigned based on the method described above, using the scoring function defined below. (3) If a *Virtual California* element occurred within a time window, the total score is incremented using one of the methods described above. The score for a particular simulation year is combined contributions from each paleoseismic site.

In this study, the *VC* simulation data is scored using a unit-height Gaussian function. The time $t_{P,j}(\boldsymbol{x})$ are the time of the jth paleoseismic event in years before "actual present" $t = 2009$ at the site $\boldsymbol{x}$. Time $t_{S,i}(\boldsymbol{x})$ is the time of the ith "simulation paleoseismic event" in years before "simulation present" t_{sim} at the "simulation paleosite" $\boldsymbol{x}$. $\sigma^2_{P,j}(\boldsymbol{x})$ is the quoted squared error of the actual paleoseismic event at the paleosite $\boldsymbol{x}$. At each value of "simulation present time" t, we compute a score for that year for fault element i by summing over all paleoseismic events by using the event scoring function:

At location $\boldsymbol{x}$, the contribution to the score $S_{i,j}(t, x)$ from the ith simulation event at time $t_{S,i}$, with respect to the jth paleoseismic event at time $t_{P,j}$, is given by

$$S_{i,j}(t,x) = \exp\left[-\left[t_{S,i}(x) - t_{P,j}(x)\right]^2 / \sigma^2_{P,j}(x)\right] \quad (1)$$

This scoring function assigns a higher score to events which occur closer to the mean paleoseismic value, and a smaller score for simulation events further removed in time from the actual paleoseismic event. A Gaussian is constructed for each paleoseismic event, centered about the mean event date so that about 90% of its area lies within the error bounds. This diminishes the importance of simulation events that occur far from the mean time of the actual paleo-earthquake. VAN AALSBURG et al. (2007) describe this procedure in more detail, and show examples of a "high scoring" time and a "low scoring" time.

The scoring system does not invoke a penalty if there are more *VC* earthquakes near a paleoseismic site than there are observed paleo-earthquakes. The rationale for this choice is that not all earthquakes can

Table 1

Fault probabilities and fault lengths for the next $M > 6.5$ earthquake corresponding to the spatial probabilities shown in Fig. 2a

Fault	Eq. probability (%)	Fault length (km)
Bartlett Springs	12.2	85.0
Calaveras	74.3	154.0
Concord-Green Valley	1.4	55.0
Death Valley	5.4	248.0
Greenville	0.7	73.0
Maacama	2.0	179.0
Rodgers Creek	2.0	62.0
San Gregorio	0.7	89.0
Sargent	0.7	53.0
San Andreas South	0.7	580.0

Faults not listed in the table had less than 0.1% probability of occurrence for the next $M > 6.5$ earthquake

be observed using paleoseismic techniques, and thus the paleoseismic data represent a minimum number of paleo-earthquakes (GRANT, 2002).

6. Forecasting Feasibility Methods

Once a simulation time history has been scored year by year to obtain the time series Score(t_{sim}), we use the scored simulation in a forecasting experiment to forecast large earthquakes. The basic principle is that the higher the score at time t_{sim}, the more closely the seismic history leading up to t_{sim} resembles the actual seismic history of California. The assumption is that the seismic activity at "future" time intervals for times $t > t_{sim}$ will more closely resemble the seismic future in California if the score value is high. In addition, if we stack the time series data from "future" intervals, we can further surmise that the statistics of these stacked intervals may represent the statistics of future events on the real fault system. By using only the set of high-scoring times, together with their immediate future time intervals, we optimize VC to forecast seismic activity on the SAF system.

We select the "high scoring" years by applying a *decision threshold*. What constitutes a high score varies by method, radius of neighborhood, etc. Typically, we select a score so that approximately the top 0.37% of the simulated events are "high scoring". For each of these events we then compute the time until the next large events, either $m > 6.5$ or $m > 7.0$ (VAN AALSBURG *et al.*, 2007). Although the paleoseismic data have been used as part of the procedure to set the model friction parameters and long-term offset rates (e.g., RUNDLE *et al.*, 2001), using them to score the data is not redundant. The friction parameters use only the long-term, average recurrence intervals. The data scoring use the details of inter-event times, meaning that the scoring algorithm uses the variability of the data, rather than just long-term average rates.

7. Results

VAN AALSBURG *et al.* (2007) present several scoring algorithms, including the unit-height Gaussian scoring function described in Eq. 1, and showed examples of temporal cumulative probability functions (CDFs) obtained by stacking data from high scoring years. These temporal CDFs represented the probability of the next event larger than $M > 7.0$ as a function of time until the next event. The CDFs can generally be characterized as having a Poisson appearance because statistics from many fault elements were stacked. The median time to the next event was found to be generally around eight years. Applied to the present time (2009), this would indicate a 50% probability of an $M > 7.0$ event in California by 2016. VAN AALSBURG *et al.* (2007) also gave examples of the fit to the paleoseismic data set for a representative low-scoring simulation year, and a representative high-scoring simulation year.

Here we focus on identifying the probable locations of the next $M > 6.5$ and $M > 7.0$ earthquakes in California that may occur on the fault system shown in Fig. 1. To compute these locations, we use the top 0.37% of the highest scoring years as determined from the scoring function Eq. 1. Using these 148 highest-scoring years as "the present", we then determine the boundary element(s) that participate in the "next" $M > 6.5$ or $M > 7.0$ events. A boundary element is considered to have participated if it is within 40 km of the latitude–longitude coordinates of an actual, observed paleoseismic event. Compiling these statistics, we obtain results shown in Figs. 2a, b and 3a, b. These results demonstrate feasibility of the method only, and should not be taken as a statistically validated forecast.

Examination of Fig. 2a indicates that most of the probability for the next $M > 6.5$ event is associated with the Calaveras fault in northern California, with lower probability scattered among other faults in northern California, including the Rodgers Creek and Green Valley—Bartlett Springs fault system. Figure 2b indicates that most of the probability for the next $M > 7.0$ earthquake is associated with the Carrizo section of the San Andreas fault, the Garlock fault, the northern San Andreas fault, the Hunting Creek-Berryessa fault, and to a lesser extent the Hayward and Rodgers Creek faults. The probability for each of these faults is given in Tables 1 and 2.

In Fig. 3a and b, we address the question: "During the fixed time interval consisting of the next thirty years from now, on which fault locations are at least 1
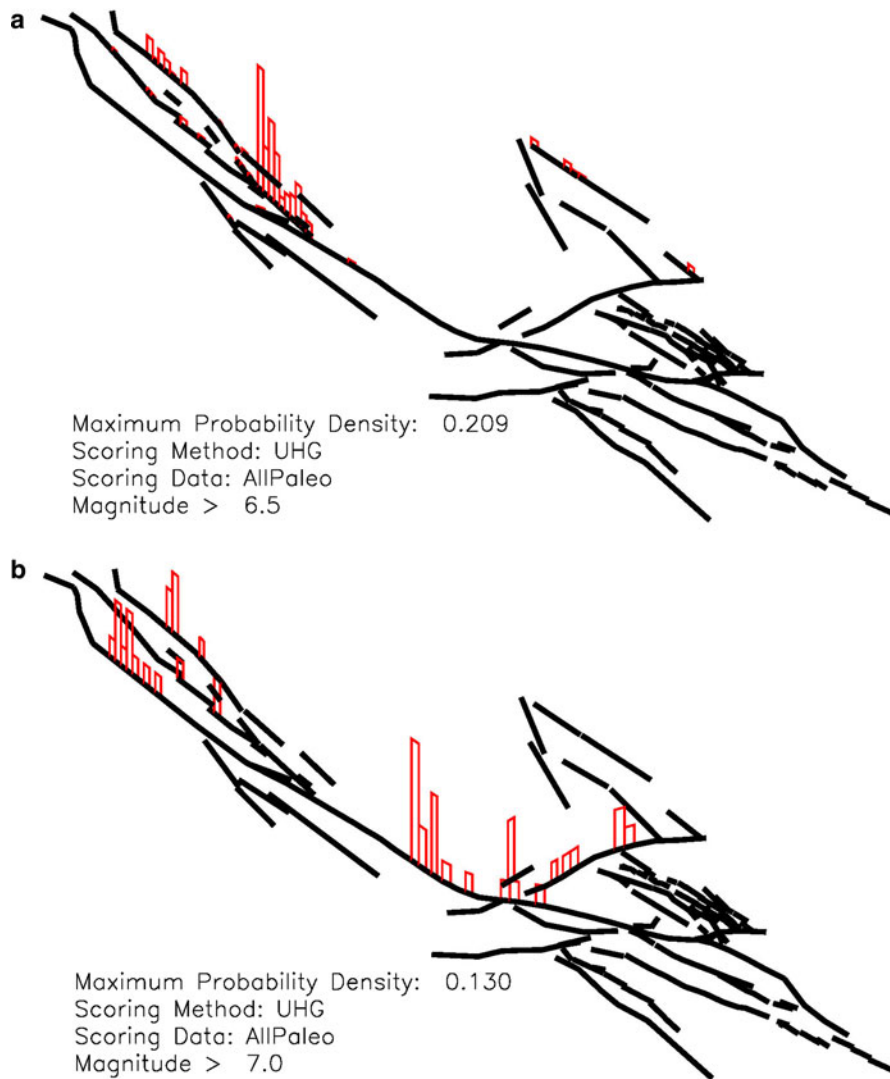
Figure 2

a Map showing the fault boundary element relative probabilities for participation in the next $M > 6.5$ earthquake. The corresponding fault probabilities are tabulated in Table 1. **b** Map showing the fault boundary element relative probabilities for participation in the next $M > 7.0$ earthquake. The corresponding fault probabilities are tabulated in Table 2

$M > 6.5$ (Fig. 3a) or at least 1 $M > 7.0$ events most likely to occur"? Figure 3a shows that the relative probability of $M > 6.5$ earthquakes is widely distributed spatially among many faults (30-year probability per $M > 6.5$ event). Figure 3b shows that for $M > 7.0$ earthquakes, probability is concentrated on the northern San Andreas fault between San Francisco and Mendocino, on the Carrizo section of the southern San Andreas fault, the Garlock and White Wolf faults, the northern San Andreas fault, the Rodgers Creek-Maacama faults, and the Hunting

Creek-Berryessa faults. The probabilities for each fault are given in Tables 3 and 4.

In Fig. 4a and b, we address the question: "On the northern and southern San Andreas fault, *when* during the next thirty years are $M > 7$ earthquakes most likely to occur?" On the northern San Andreas fault, we focus on the spatial locations identified in Fig. 3b as being most likely to participate in a $M > 7$ earthquake. These locations can be recognized as having the red vertical bars, along the fault from Mendocino down to San Francisco. Figure 4a indicates that the
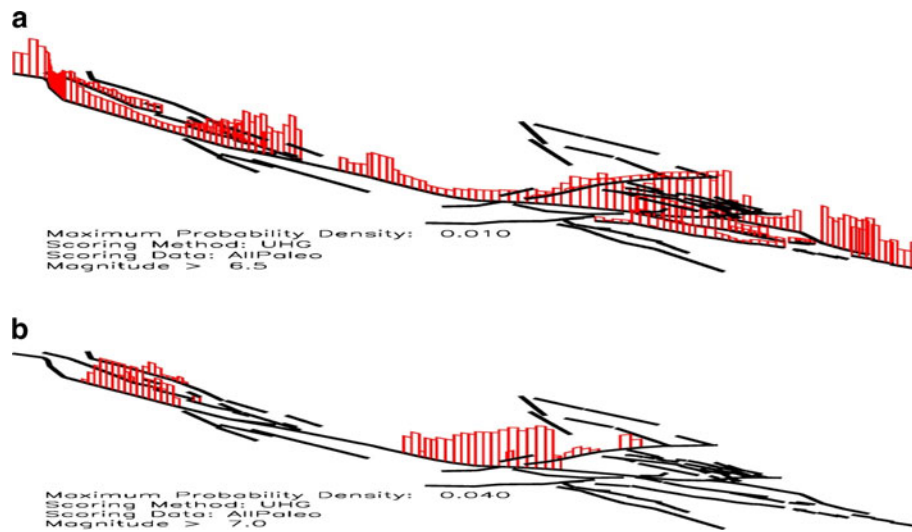
Figure 3

a Map showing the relative probabilities that at least 1 $M > 6.5$ earthquake will occur on the boundary element during the next 30 years. The corresponding fault probabilities are tabulated in Table 3. *Probability bars* are only plotted if the corresponding integrated fault probabilities are larger than 2%. **b.** Map showing the relative probabilities that at least 1 $M > 7.0$ earthquake will occur on the boundary element during the next 30 years. The corresponding fault probabilities are tabulated in Table 4

Table 2

Relative spatial probabilities that the next $M > 7.0$ earthquake will occur on a fault, corresponding to the spatial probabilities shown in Fig. 2b

Fault	Eq. probability (%)	Fault length (km)
Bartlett Springs	10.9	85.0
Hayward	4.3	111.0
Hunting Creek— Berryessa	2.2	59.0
Rodgers Creek	2.2	62.0
San Andreas North	23.9	467.0
San Andreas South	37.0	580.0
Garlock	13.0	234.0
White Wolf	6.5	47.0

Fault lengths are also listed. Faults not listed in the table had less than 0.1% probability of occurrence for the next $M > 7.0$ earthquake

Table 3

Relative spatial probabilities that at least 1 $M > 6.5$ earthquake will occur on a fault during the next 30 years (30-year probability per $M > 6.5$ event)

Fault	Eq. probability (%)	Fault length (km)
Calaveras	7.0	154.0
Hayward	2.3	111.0
Maacama	2.5	179.0
San Andreas North	15.9	467.0
San Andreas South	25.9	580.0
San Jacinto	7.1	291.0
Elsinore	3.2	236.0
Imperial Valley	11.1	162.0
Garlock	1.9	234.0
Brawley	1.9	52.0

Probabilities correspond to those shown in Fig. 3a, and fault lengths are also indicated

Faults not listed in the table had less than a 2% relative probability of occurrence for a $M > 6.5$ earthquake during the next 30 years

highest probability years are years 9 and 17 counting forward from the present, corresponding to 2018 and 2026. On the southern San Andreas fault, the most likely locations for an $M > 7$ earthquake during the next thirty years can be recognized by the vertical red bars located from the the Carrizon south to Fort Tejon. Figure 4b indicates that the most probable year for such an earthquake is year 26 counting forward from present, or 2035. However, on both the northern and southern San Andreas fault, there

remains significant, although lesser, probabilities for such an event in other years.

Demonstrating the accuracy of a forecast is a very difficult problem.[3] Figure 4a and b also partially

---

[3] http://www.bom.gov.au/bmrc/wefor/staff/eee/verif/verif_web_page.html.

Relative spatial probabilities that at least 1 M > 7.0 earthquake will occur on a fault during the next 30 years (30-year probability per M > 7.0 event)

Fault	Eq. probability (%)	Fault length (km)
Bartlett Springs	3.1	85.0
Hayward	0.6	111.0
Hunting Creek— Berryessa	0.6	59.0
Maacama	2.5	179.0
Rodgers Creek	0.6	62.0
San Andreas North	32.6	467.0
San Andreas South	54.0	580.0
Garlock	5.3	234.0
White Wolf	0.8	47.0

Probabilities correspond to those shown in Fig. 3a, and fault lengths are also indicated

Faults not listed in the table had less than a 0.1% relative probability of occurrence for a $M > 7.0$ earthquake during the next 30 years

answer the question: "For times identified as "optimal" during a *VC* simulation, do similar pasts imply similar futures?" This question bears on the accuracy of forecasts. The basic assumption in this paper is that similar pasts do imply similar futures. If this is not the case, use of simulations for earthquake forecasting will probably not be possible. Here, we have used a long history of simulations to identify optimal times whose preceding activity is similar to the actual paleoseismic events preceding the present, 2009. If the events following these optimal simulation times appear to be only a random sequence of earthquakes, uniformly distributed over the thirty year interval, this would suggest that past activity is not correlated with future activity. In that case, our proposed technique would probably not be useful.

Figure 4a and b appear to indicate that while there is a lower level background of random times, due to statistical variations, there are nonetheless a few times that stand out as preferred occurrence times for future large earthquakes. For Fig. 4a (northern San Andreas fault), these are years 9 and 17. For Fig. 4b (southern San Andreas fault), year 26 stands out. As the simulation model, including faults, average recurrence times, average long-term slip rates, and other model data are more closely matched to the actual San Andreas fault system data, it is possible that statistical variation will be reduced. Because

there are nonetheless a few preferred times for future earthquakes that stand out above the relatively uniform background probability, Fig. 4a and b suggest the conclusion that similar earthquake pasts seem to be at least somewhat correlated with similar earthquake futures.

8. *Discussion and Conclusions*

We have presented a general method for using numerical earthquake fault system simulations to compute spatial forecast probabilities for earthquakes having magnitudes above a given, threshold. Our method utilizes catalogs of simulated earthquakes from the model *Virtual California*, together with a data scoring algorithm that identifies parts of simulation catalogs most similar to recent earthquake history in California as determined by paleoseismology. Optimal parts of the simulation catalogs are then used to compute statistical forecasts for future large events. While our results are preliminary, the probabilities we compute show the power of the method.

Our method can be compared to the recent methods developed by the Working Group on California Earthquake Probabilities (2002, 2008) (see notes [1, 2], The WGCEP assume that coherent geological fault segments exist and rupture repeatedly as a unit (characteristic earthquake assumption), that earthquake ruptures do not generally jump from one fault to another, that earthquake ruptures typically obey either Brownian Passage Time or log-normal statistics, and that earthquakes on different fault segments are independent and uncorrelated.

In contrast, *VC* is a physically, rather than statistically, motivated model that assumes earthquake faults interact elastically, that friction retards slip on fault surfaces, and that faults typically slip at their observed, long-term rates. *VC* uses topologically realistic models of fault systems to generate catalogs of simulated major earthquakes that can then be analyzed statistically for patterns and other information. Here we show how these simulated catalogs can be used in earthquake forecasting. While the average intervals between paleoearthquakes are used to assign the frictional parameters on the model faults, the
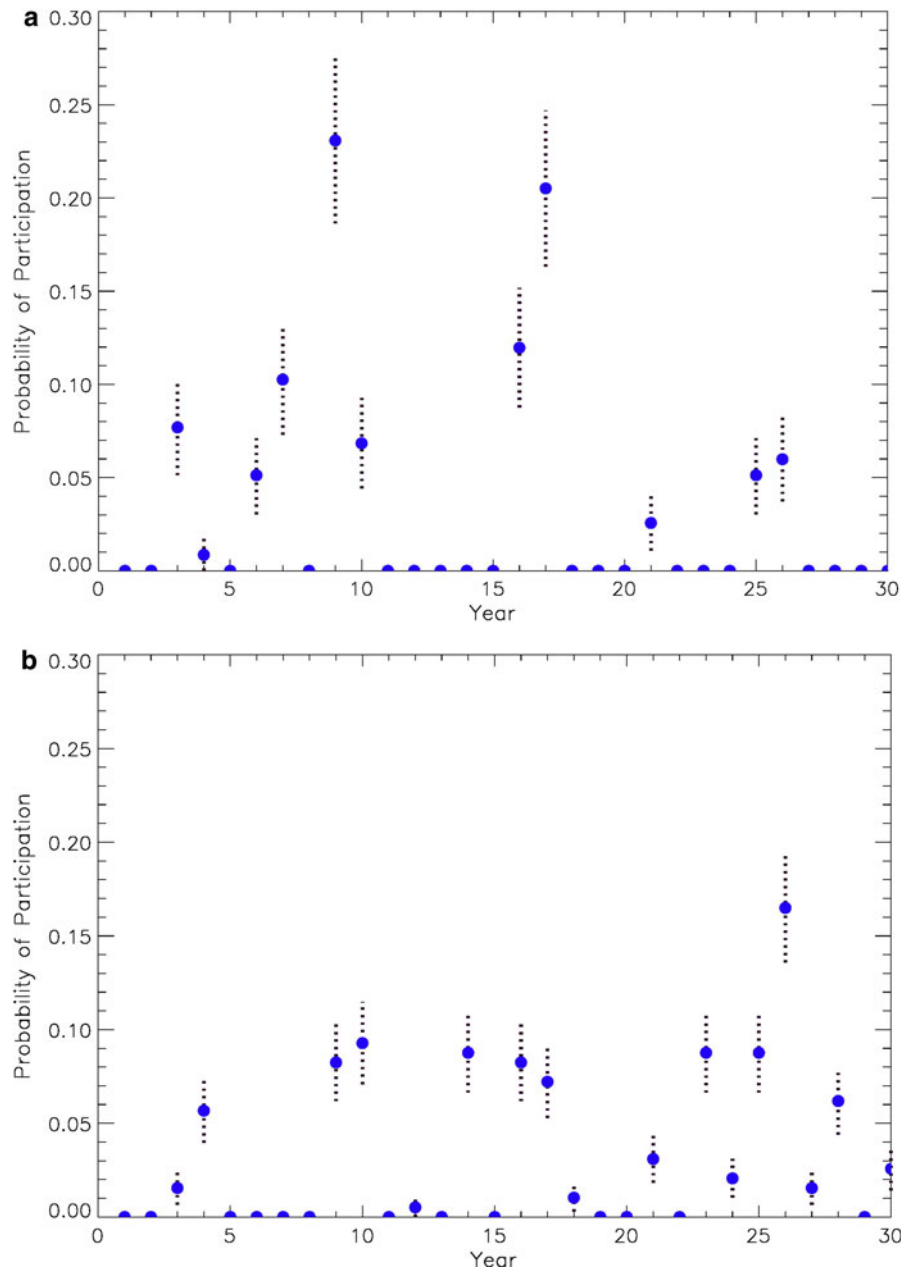
Figure 4

a Probability density function for the times during the next thirty years beginning from present (January 1, 2009) when a $M > 7.0$ earthquake is most likely to occur on the Northern San Andreas fault. Location on the fault corresponds to the high probability region on the NSAF shown in Fig. 3b. *Vertical bars* on data points indicate the 1σ Poisson counting uncertainty. **b** Probability density function for the times during the next thirty years beginning from present (January 1, 2009) when a $M > 7.0$ earthquake is most likely to occur on the Southern San Andreas fault. Location on the fault corresponds to the high probability region on the NSAF shown in Fig. 3b. *Vertical bars* on data points indicate the 1σ Poisson counting uncertainty

variability of the paleoearthquake occurrence times are used to determine which parts of the simulated catalogs are optimal for use in forecasting.

Finally, it is of interest to compare our forecasts to paleoseismic observations published by GRANT and SIEH (1994), GRANT (1996), and AKCIZ *et al.* (2009).

Their work suggested that recent ruptures of the SAF in the Carrizo were clustered in time ("uncharacteristic earthquakes") rather than more regularly spaced in time ("characteristic earthquakes"). So while the long-term average recurrence time might be several centuries (SIEH and JAHNS, 1984; WGCEP 1988 and 1995), their data showed evidence for as many as four major earthquake ruptures between 1218 A.D. and 1510 A.D. The recent work by AKCIZ et al. (2009) reveals shorter average intervals.

The results shown in Fig. 2b suggest that the next major $M > 7.0$ earthquake could occur on the Carrizo reach of the SAF, possibly within thirty years from 2009. Under the "characteristic earthquake" scenario, with the most recent major rupture having occurred in 1857, it would be unlikely for another major rupture to occur in the near future. However, under a temporally clustered, "uncharacteristic earthquake" scenario, a major rupture in the Carrizo Plane might be expected in the near future.

With respect to $M > 6.5$ earthquakes, the most likely fault to rupture appears to be the Calaveras fault. Evidence from Coulomb stress transfer calculations (REASENBERG and SIMPSON, 1992) indicates that although the 1989 M 7.1 Loma Prieta earthquake might have raised the stress on the Calaveras fault by less than 1 bar, the seismicity rate nevertheless declined in the years following 1989. For that reason, the high probability on the Calaveras fault as shown in Fig. 2a is somewhat unexpected, if direct stress transfer from the Loma Prieta earthquake is assumed to be a triggering event.

Acknowledgments

REFERENCES

AKCIZ, S. O., GRANT LUDWIG, L., and ARROWSMITH, J. R. (2009), *Revised dates of large earthquakes along the Carrizo section of the San Andreas Fault, California, since A.D. 1310 ± 30,* J. Geophys. Res. *114*, B01313, doi:10.1029/2007JB005285.

BIASI, G. P., WELDON, R. J. II, FUMAL, T. E., and SEITZ, G. G., (2002), *Paleoseismic event dating and the conditional probability of earthquakes on the southern San Andreas fault, California*, Bull. Seism. Soc. Am. *92*, 7, 2,761–2,781.

EBERHART-PHILLIPS, D. *et al.* (2003), *The 2002 Denali Fault Earthquake, Alaska: A Large Magnitude, Slip-Partitioned Event*, Science *300*, 1113–1118, doi:10.1126/science.1082703.

FIELD, E. H. (2007), *A summary of previous working groups on California earthquake probabilities*, Bull. Seismol. Soc. Am. *97*, 1033–1053.

FUMAL, T. E., WELDON, R. J. II, BIASI, G. P., DAWSON, T. E., SEITZ, G. G., FROST, W. T., and SCHWARTZ, D. P. (2002), *Evidence for large earthquakes on the San Andreas fault at the Wrightwood, Californa, paleoseismic site: A.D. 500 to present,* Bull. Seismol. Soc. Am. *92*, 7, 2,726–2,760.

GRANT, L. B., and SIEH K. E. (1994), *Paleoseismic evidence of clustered earthquakes on the San Andreas fault in the Carrizo Plain, California*, J. Geophys. Res., *99*, 6819–6841.

GRANT, L. B. (1996), *Uncharacteristic earthquakes on the San Andreas fault*, Science, *272*, 826–827.

GRANT, L. B., *Paleoseismology,* Chapter 30. In *IASPEI International Handbook of Earthquake and Engineering Seismology* (W. H. Lee, H. Kanamori, and P.C. Jennings, eds.) (Internatl. Assoc. Seismol. Phys. Earth's Inter. 2002) v. 81A, 475–489.

GRANT, L., *Paleoseismology*. In *Treatise on Geophysics* (G. Schubert, ed.), Volume 4, *Seismology* (H. Kanamori, ed.), (Elsevier, 2007), 567–589.

GRANT, L. B., GOULD, M. M., DONNELLAN, A., McLEOD, D., Yun-An CHEN, A., SUNG, S.-S., PIERCE, M., FOX, G. C., and RUNDLE P. (2005), *A Web services-based universal approach to heterogeneous fault databases*, Comput. Sci. Eng. *7*, 4, 51–57.

MARCO, S., STEIN, M., AGNON A., and RON, H. (1996), *Long-term earthquake clustering: A 50,000-year paleoseismic record in the Dead Sea Graben*, J. Geophys. Res. *101*, 6179–6192.

PAGE, M. T., and Carlson, J. M. (2006), *Methodologies for earthquake hazard assessment: Model uncertainty and the WGCEP-2002 forecast*, Bull. Seismol. Soc. Am. *96*, 1624–1633.

REASENBERG, P. A., and SIMPSON, R. W. (1992), *Response of regional seismicity to the static stress change produced by the Loma Prieta earthquake*, Science *255*, 5052, 1687–1690.

RICHARDS-DINGER, K., and DIETERICH, J. H., (2008), *A regional scale earthquake simulator for faults with rate- and state-dependent frictional properties*, Eos Trans. AGU *87*(52), Fall Meet. Suppl., Abstract S34A-08.

ROBINSON, R. (2004), *Potential earthquake triggering in a complex fault network: the northern South Island, New Zealand*, Geophys. J. Int. *159*, 734–738.

RUNDLE, J. B., and KANAMORI, H. (1987), *Applications of an inhomogeneous stress (patch) model to complex subduction zone*

earthquakes: A discrete interaction matrix approach, J. Geophys. Res. 92, 2606–2616.

RUNDLE, J. B. (1988), A physical model for earthquakes, 2, Application to southern California, J. Geophys. Res. 93, 6255–6274.

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

RUNDLE, J. B., RUNDLE, P. B., DONNELLAN, A., TURCOTTE, D. L., SCHERBAKOV, R., LI, P., MALAMUD, B. D., GRANT, L. B., FOX, G. C., MCLEOD, D., YAKOLEV, G., PARKER, J., KLEIN, W., and TIAMPO K. F. (2005), A simulation-based approach to forecasting the next great San Francisco earthquake, Proc. Natl. Acad. Sci. 102, 15363–15367. http://www.pnas.org/cgi/doi/10.1073/pnas.0507528102.

RUNDLE, J. B., TIAMPO K. F., KLEIN W., and MARTINS J. S. S. (2002), Self-organization in leaky threshold systems: The influence of near-mean field dynamics and its implications for earthquakes, neurobiology, and forecasting, Proc. Natl. Acad. Sci. USA 99, 2514–2521., Suppl. 1.

RUNDLE, P. B., RUNDLE, J. B., TIAMPO, K. F., MARTINS, J. S. S., MCGINNIS, S., and KLEIN, W. (2001), Nonlinear network dynamics on earthquake fault systems, Phys. Rev. Lett. 87, 148501.

RUNDLE, P. B., RUNDLE, J. B., TIAMPO, K. F., DONNELLAN, A., and TURCOTTE, D. L. (2006), Virtual California: Fault model, frictional parameters, applications, Pure Appl. Geophys. 163, 1819–1846, doi:10.1007/s00024-006-0099-x.

SAVAGE, J. C., and PRESCOTT W. H. (1978), Asthenosphere readjustment and the earthquake cycle, J. Geophys. Res. 83, 3369–3376.

SIEH, K. E., and JAHNS, R. H. (1984), Holocene activity of the San-Andreas Fault at Wallace-Creek, California, Geolog. Soc. Am. Bull. 95, 883–896.

SIEH, K., STUIVER, M., and BRILLINGER, D. (1989), A more precise chronology of earthquakes produced by the San Andreas fault in southern California, J. Geophys. Res. 94, B1, 603–623.

VAN AALSBURG, J., GRANT, L. B., YAKOVLEV, G., RUNDLE, P. B., RUNDLE, J. B., YAKOVLEV, G., TURCOTTE, D. L, and DONNELLAN, A. (2007), A feasibility study of data assimilation in numerical simulations of earthquake fault systems, Phys. Earth Planet Int. 163, 149–162.

WALD, D. J., and HEATON, T. H. (1994), Spatial and temporal distribution of slip for the 1992 Landers, California, earthquake, Bull. Seismol. Soc. Am. 84, 668–691.

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, 22393–22418.

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

WELDON, R. J. II, SCHARER, K. M., FUMAL, T. E., and BIASI, G. P. (2004), Wrightwood and the earthquake cycle: what a long recurrence record tells us about how faults work, GSA Today 14, 9, 4–10.

WELDON, R. J. II, FUMAL, T. E., BIASI, G. P., and SCHARER, K. M. (2005), Past and future earthquakes on the San Andreas Fault. Science 308, 5724, 966–967.

WORKING GROUP on CALIFORNIA EARTHQUAKE PROBABILITIES (WGCEP). (1988), Probabilities of large earthquakes occurring in California on the San Andreas fault, U.S. Geol. Surv. Open-File Rept., 62 pp.

WORKING GROUP on CALIFORNIA EARTHQUAKE PROBABILITIES (WGCEP). (1995), Seismic hazards in southern California: probable earthquakes, 1994–2024, Bull. Seismol. Soc. Am. 85, 379–439.

YAKOVLEV, G., TURCOTTE, D. L., RUNDLE, R. B., and RUNDLE P. B. (2006), Simulation based distributions of earthquake recurrence times on the San Andreas fault system, Bull. Seismol. Soc. Am. 96, 1995–2007.

ZHUANG, J., OGATA, Y., and VERE-JONES, D. (2004), Analyzing earthquake clustering features by using stochastic reconstruction, J. Geophys. Res., 109, B05301, doi:10.1029/2003JB002879.