



## Earthquake recurrence on the south Hayward fault is most consistent with a time dependent, renewal process

Tom Parsons<sup>1</sup>

Received 2 September 2008; revised 1 October 2008; accepted 3 October 2008; published 1 November 2008.

[1] Elastic rebound and stress renewal are important components of earthquake forecasting because if large earthquakes can be shown to be periodic, then rupture probability is time dependent. While renewal models are used in formal forecasts, it has not been possible to exclude the alternate view that repeated large earthquakes can happen in rapid succession without requiring time for stress regeneration. Here a consistency test between time dependent and time independent recurrence distributions is made using a Monte Carlo method to replicate the paleoseismic series on the south Hayward fault. Time dependent distributions with recurrence interval of 210 years and coefficient of variation of 0.6 reproduce the event series on the south Hayward 5 times more often than any exponential distribution: a highly significant difference as determined using a two-tailed Z-test for relative proportions. Therefore large Hayward fault earthquakes are quasi-periodic, and are most consistent with a stress renewal process. **Citation:** Parsons, T. (2008), Earthquake recurrence on the south Hayward fault is most consistent with a time dependent, renewal process, *Geophys. Res. Lett.*, 35, L21301, doi:10.1029/2008GL035887.

### 1. Introduction

[2] Is large earthquake probability on major faults time dependent or constant? Time dependence emerges from the concept of elastic rebound, where some period is required to recharge stresses released in the last large earthquake. Under a time dependent process, probability is lowest just after a large earthquake, but rises with time (Figure 1). While this theoretical concept has been accepted as a component of earthquake forecasting [*Working Group on California Earthquake Probabilities (WGCEP)*, 2003, 2008], it has been difficult to prove from observations [*Davis et al.*, 1989; *Kagan and Jackson*, 1999]. Therefore it is common practice to also report time independent probability (constant vs. time) in forecasting [*WGCEP*, 2003, 2008; *Frankel et al.*, 2002; *Console et al.*, 2008; *Petersen et al.*, 2008], which increases overall uncertainty (Figure 1). In this paper, a remarkable paleoseismic sequence developed by *Lienkaemper and Williams* [2007] for the southern Hayward fault in the San Francisco Bay area of California (Figure 2) is examined quantitatively for consistency with time dependent and time independent recurrence models.

[3] Optimally, we would have enough recorded earthquake intervals to unequivocally define the shape of recurrence distributions on faults, which could then be used to

establish whether large earthquakes are periodic or temporally random. However, doing that requires at least ~25–50 intervals to gain the necessary resolution [e.g., *Matthews et al.*, 2002], and there are presently no well-resolved continuous series of large earthquakes that long. A large sample is needed because earthquake recurrence-time distributions are skewed asymmetrically about their means [e.g., *Nishenko and Buland*, 1987; *Hagiwara*, 1974; *Kagan and Knopoff*, 1987]. For example, if one wants to characterize mean recurrence from a limited sample, that sample has highest probability of being drawn from the mode (most frequent value) of the distribution, which is effective for normal parent distributions where the mode equals the mean (Figure 3). However, if one averages the small sample drawn from a skewed parent distribution, then the resulting value tends to fall somewhere between the mode and the actual mean. Tests with commonly used earthquake recurrence distributions show that this effect causes sample means to be ~60%–70% of the parent mean value [*Parsons*, 2008] (Figure 3).

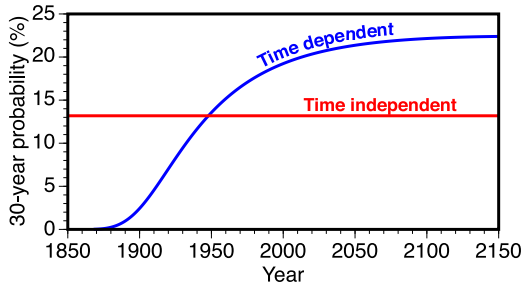
[4] A distribution that attempts to characterize earthquake recurrence has to be skewed (asymmetric about its mean) to prevent nonsensical negative recurrence times (Figure 3a). This complexity has made it difficult to define the very basic question of periodicity. However, if we specify general functional forms of recurrence distributions in advance, then it is possible overcome the limited sampling problem using Monte Carlo techniques [*Parsons*, 2008]. In this paper it is shown that time dependent and time independent distributions have different enough shapes (Figure 3) such that they can be compared for consistency against an observed earthquake series.

### 2. Monte Carlo Comparison of Recurrence Distribution Parameters

[5] In this paper, two families of earthquake recurrence distributions are assumed up front: a time dependent form and a time independent form. These two distribution sets are compared with a Monte Carlo method that tests and assesses the relative success of different distributions against a set of observed intervals. The problem of small sampling is solved by drawing millions of synthetic earthquake sequences at random from time dependent and time independent distributions. Here, distributions that most often reproduce the south Hayward fault paleoseismic sequence are taken to be more representative of the actual (but unknown) recurrence-time distribution of large Hayward fault earthquakes.

[6] The goal is to compare the relative success of time dependent and time independent distributions at reproducing the south Hayward fault paleoseismic series. The method is described in detail by *Parsons* [2008], so just

<sup>1</sup>U.S. Geological Survey, Menlo Park, California, USA.



**Figure 1.** Comparison of time independent (red line) and time dependent (blue line) 30-year earthquake probability calculations for the south Hayward fault made using a 210-yr recurrence interval and a 0.6 coefficient of variation. The difference between the two results is ~7% by the end of 2008.

the specific application to south Hayward fault earthquake recurrence is described here. Time independence is represented by exponential distributions (Poisson process) as

$$f(t) = \lambda e^{-\lambda t}, \quad \text{for } t > 0 \quad (1)$$

where  $t$  is time, and  $\lambda$  is the mean rate (1/mean recurrence). Brownian Passage Time (inverse Gaussian) distributions [Kagan and Knopoff, 1987; Matthews *et al.*, 2002] are taken to represent time dependence as,

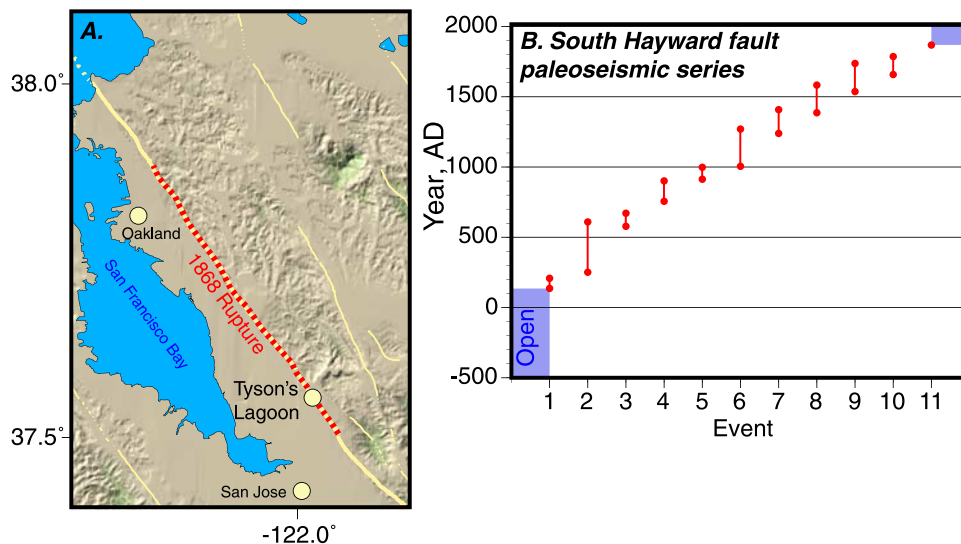
$$f(t, \mu, \alpha) = \sqrt{\frac{\mu}{2\pi\alpha^2 t^3}} \exp\left(-\frac{(t - \mu)^2}{2\mu\alpha^2 t}\right), \quad (2)$$

where  $\mu$  is recurrence interval and  $\alpha$  is coefficient of variation. The Brownian Passage Time distribution is nearly indistinguishable from the also commonly applied lognor-

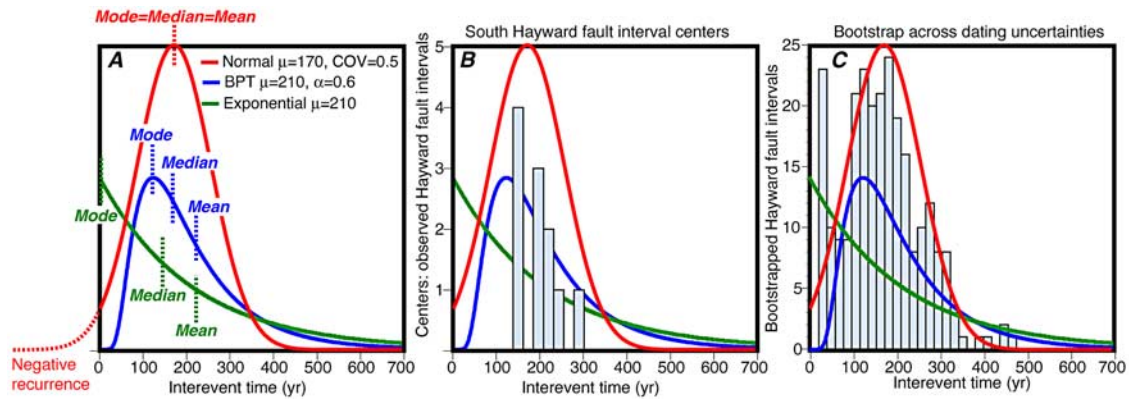
mal distribution for earthquake recurrence. It is unknown whether any of these distributions really characterizes south Hayward fault large earthquake recurrence. However, the significant difference in their shapes (Figure 3) captures the two hypotheses of periodic vs. temporally random recurrence. The largest difference near  $t = 0$  is a necessary feature that distinguishes the two model classes, and is the key difference that enables a test against the observed paleoseismic series.

[7] To make the test, a series of distributions that covers all reasonable mean recurrence intervals is developed (10 yrs to 5000 yrs for the south Hayward fault case). Time dependent distributions are characterized by two parameters (equation (2)), and are thus also constructed across coefficient of variation values between 0.01 and 0.99 for each mean recurrence interval. Groups of earthquake times are randomly drawn millions of times from each possible recurrence distribution and assembled into sequences. With this method, sequence means are identified directly from the parameters of parent distributions rather than from taking arithmetic means of sequences. Thus this process overcomes the sampling bias problem of skewed distributions and is more effective than bootstrapping over the dating uncertainty intervals, which repeats and stacks means from insufficient samples (Figure 3c).

[8] Every recurrence distribution is randomly sampled 15 million times for groups of 11 Hayward fault earthquakes. Those sequences that have one earthquake occurring in order during each observed event window (range of possible event times as constrained by radio carbon dating), and no earthquakes in the intervals between event windows are tallied, and a ranking of matches vs. mean recurrence and coefficient of variation to the observed record is produced (Figure 4). The examples shown in this paper use a uniform distribution for the event time-window defined by 95% confidence intervals on dating uncertainty



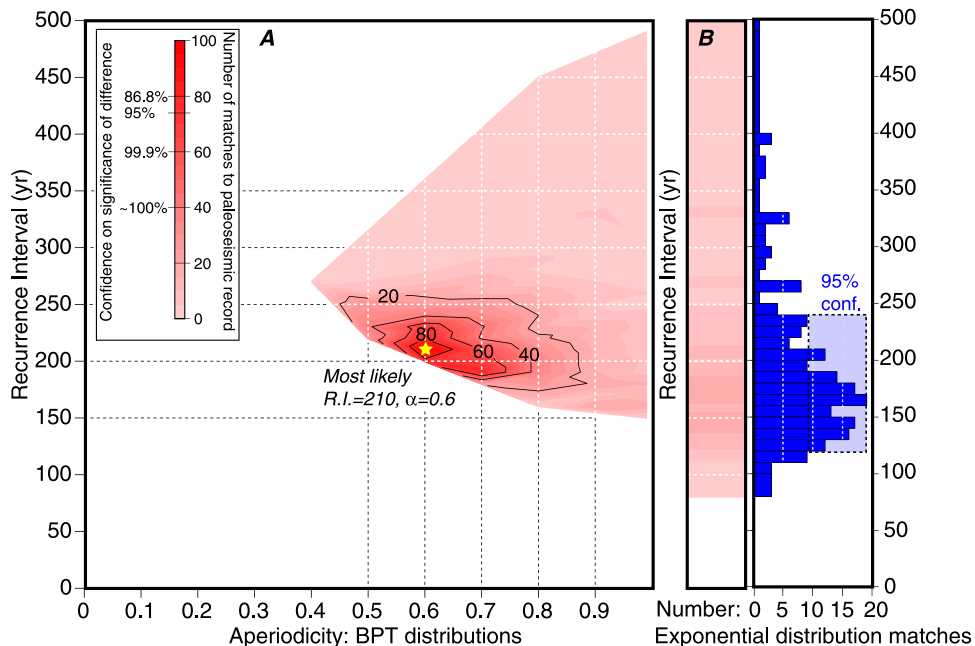
**Figure 2.** (a) Hayward fault location east of San Francisco Bay with the paleoseismic site at Tyson’s Lagoon identified. The extent of the 1868 earthquake rupture is shown according to *Yu and Segall* [1996]. (b) The 95% confidence bounds on event times are shown from *Lienkaemper and Williams* [2007]. Open intervals before and after the sequence are shaded in blue.



**Figure 3.** (a) Example earthquake recurrence distributions are compared with a normal distribution. The blue curve is Brownian Passage Time (BPT) [Kagan and Knopoff, 1987; Matthews et al., 2002], which is a commonly used time-dependent distribution. The green curve is an exponential distribution and is used for time independent earthquake probability calculations. Earthquake recurrence distributions are skewed such that the modes are not the same as the means as in the normal distribution; thus a small sampling will most likely have some bias towards the mode, and the sample mean will often underestimate the parent distribution mean. The normal distribution is unacceptable for earthquake recurrence because it can allow negative recurrence times. (b) A histogram of south Hayward fault paleoseismic event times (interval centers) appears most consistent with time dependent distributions. (c) A histogram of a bootstrap over full dating uncertainties is shown that suggests possible consistency with all distributions.

(Figure 2b), and an event that happens at any time within the window is considered a match. The Monte Carlo sequences begin with an event that is given freedom to happen any time prior to the first observed earthquake time window. The extra event contributes nothing other than a

starting point for the sampling. This is needed because the first observed time window has some range within which the event might have happened, whereas Monte Carlo simulation must begin at a point in time. It is expected that that an earthquake occurred prior to the first identified event



**Figure 4.** Contours of matches to south Hayward fault paleoseismic event series of different (a) time dependent (Brownian Passage Time) and (b) time independent (exponential) recurrence distributions. The best-fit distributions are time dependent, with recurrence intervals of  $\mu \sim 210$  yr, and coefficient of variation  $\alpha \sim 0.6$ . Confidence (Z-test) on the significance of relative proportions is keyed to the contour intervals. The best-fit exponential distributions have significantly fewer matches, leading to the conclusion that earthquake recurrence on the south Hayward fault is time dependent, possibly from a stress renewal process. A histogram of exponential distribution matches is shown in Figure 4b, with 95% confidence of significance shaded, which gives the same information as the adjacent contour mapping, but in more detail.



in the south Hayward fault paleoseismic series, but we have no knowledge of it other than that. To avoid the start time having influence, any number of events are allowed to happen in a 500-yr period prior to the first observed event window. Conversely, simulations that include earthquakes within the open interval between the 1868 earthquake and present time (or any other interval when no earthquakes occurred) are discarded.

[9] To summarize the technique, exponential distributions with mean recurrence intervals from 10 to 5000 yrs are generated in 10-yr increments, and a group of Brownian Passage Time distributions are generated with mean recurrence intervals from 10 to 5000 yrs in 10-yr increments. The time dependent distributions also range through coefficients of variation between 0.01 and 0.99 (0.1 increments) for every mean recurrence interval. Each individual distribution is given 15 million chances to reproduce the entire south Hayward fault paleoseismic series within 95% confidence intervals on radio carbon dating uncertainties (Figure 2b). The number of successful matches is tallied for every possible distribution. The most likely distributions to represent south Hayward fault recurrence are taken as the ones with significantly more matches to the record.

### 3. Results: Earthquake Recurrence on the South Hayward Fault

[10] The comparison of time dependent and time independent earthquake recurrence distributions described above was run to determine which is most consistent with the south Hayward fault paleoseismic series of *Lienkaemper and Williams* [2007]. A comparison between one- and two-parameter distributions is valid because each individual histogram is treated independently, and given the same number of chances to duplicate the observed series. The result is that individual time dependent distributions produced >5 times more matches to the observed record than the most common time independent distributions (Figure 4). Within the framework of the test, the most likely recurrence distribution is time dependent, with mean recurrence of  $\mu = 210$  yrs, and coefficient of variation of  $\alpha = 0.6$ . This result is somewhat different than the mean of 170 yrs reported by *Lienkaemper and Williams* [2007], probably for the reasons outlined in the introduction, wherein the mean of a small sampling of intervals is likely to be shifted towards the mode of a skewed recurrence distribution (Figure 3).

[11] The significance of the difference between the result for time dependent and time independent recurrence distributions is tested with a two-tailed Z test for relative proportions [Sachs, 1984]. The significance of the difference in the number of matches from the most frequent value ( $\mu = 210$  yrs,  $\alpha = 0.6$ ) is noted on the contours of Figure 3; for example, the difference between 60 and 100 matches is significant at 99.9% confidence. Thus the difference in the result between time dependent and time independent recurrence (100 vs < 20) is highly significant.

[12] If there were unrecognized short-interval events missing from the paleoseismic record, the result presented here might be biased away from the Poisson model. However, it is estimated that fewer than 10%–20% of events go unrecognized [Weldon et al., 2004], and that it

is equally likely that too many indicators are interpreted as paleoearthquakes [Scharer et al., 2007].

### 4. Conclusions

[13] Results described here show that the south Hayward fault paleoseismic series is more consistent with predefined time dependent recurrence distributions than time independent exponential distributions. The actual earthquake recurrence distribution for the Hayward fault remains unknown. However the key difference between Brownian Passage Time and exponential distributions lies in the much higher frequency of short-interval recurrences evident in the time independent model (Figure 3). Thus the results from Monte Carlo analysis suggest that, while recurrence is moderately variable ( $\alpha \sim 0.6$ ), there is a relative absence of short intervals between large Hayward fault earthquakes that is inconsistent with time independent recurrence. If this result is viewed from stress renewal perspective, it is suggestive of a relatively lowered stress state on the Hayward fault such that two large earthquakes are unlikely to occur in rapid succession. Rather, on average it appears necessary for some time to pass for stress regeneration before another large earthquake can occur.

[14] **Acknowledgment.** This paper is dedicated to Allin Cornell (1938–2007), whose generous interest greatly enhanced my ability to understand and explain many of the concepts presented here. Rodolfo Console and an anonymous reviewer made constructive comments that improved this paper.

### References

- Console, R., M. Murru, G. Falcone, and F. Catalli (2008), Stress interaction effect on the occurrence probability of characteristic earthquakes in Central Apennines, *J. Geophys. Res.*, *113*, B08313, doi:10.1029/2007JB005418.
- Davis, P. M., D. D. Jackson, and Y. Y. Kagan (1989), The longer it has been since the last earthquake, the longer the expected time till the next?, *Bull. Seismol. Soc. Am.*, *79*, 1439–1456.
- Frankel, A. D., et al. (2002), Documentation for the 2002 update of the national seismic hazard maps, *U.S. Geol. Surv. Open File Rep.*, *02-420*, 33 pp.
- Hagiwara, Y. (1974), Probability of earthquake occurrence as obtained from a Weibull distribution analysis of crustal strain, *Tectonophysics*, *23*, 318–323.
- Kagan, Y. Y., and D. D. Jackson (1999), Worldwide doublets of large shallow earthquakes, *Bull. Seismol. Soc. Am.*, *89*, 1147–1155.
- Kagan, Y. Y., and L. Knopoff (1987), Random stress and earthquake statistics: Time dependence, *Geophys. J. R. Astron. Soc.*, *88*, 723–731.
- Lienkaemper, J. J., and P. L. Williams (2007), A record of large earthquakes on the southern Hayward fault for the past 1800 years, *Bull. Seismol. Soc. Am.*, *97*, 1803–1819, doi:10.1785/0120060258.
- Matthews, M. V., W. L. Ellsworth, and P. A. Reasenberg (2002), A Brownian model for recurrent earthquakes, *Bull. Seismol. Soc. Am.*, *92*, 2233–2250.
- Nishenko, S. P., and R. Buland (1987), A generic recurrence interval distribution for earthquake forecasting, *Bull. Seismol. Soc. Am.*, *77*, 1382–1399.
- Parsons, T. (2008), Monte Carlo method for determining earthquake recurrence parameters from short paleoseismic catalogs: Example calculations for California, *J. Geophys. Res.*, *113*, B03302, doi:10.1029/2007JB004998.
- Petersen, M. D., et al. (2008), Documentation for the 2008 update of the United States national seismic hazard maps, *U.S. Geol. Surv. Open File Rep.*, *2008-1128*, 61 pp.
- Sachs, L. (1984), *Applied Statistics*, 707 pp., Springer, New York.
- Scharer, K. M., R. J. Weldon II, T. E. Fumal, and G. P. Biasi (2007), Paleoseismicity on the southern San Andreas fault, Wrightwood, California, 3000 to 1500 B.C.: A new method for evaluating paleoseismic evidence and earthquake horizons, *Bull. Seismol. Soc. Am.*, *97*, 1054–1093, doi:10.1785/0120060137.

Weldon, R. J., K. M. Scharer, T. E. Fumal, and G. P. Biasi (2004), Wrightwood and the earthquake cycle: What a long recurrence record tells us about how faults work, *GSA Today*, 14, 4–10.

Working Group on California Earthquake Probabilities (WGCEP) (2003), Earthquake probabilities in the San Francisco Bay region: 2002 to 2031, *U.S. Geol. Surv. Open File Rep.*, 03-214.

Working Group on California Earthquake Probabilities (WGCEP) (2008), The uniform California earthquake rupture forecast, version 2 (UCERF 2), *U.S. Geol. Surv. Open File Rep.*, 07-1437.

Yu, E., and P. Segall (1996), Slip in the 1868 Hayward earthquake from the analysis of historical triangulation data, *J. Geophys. Res.*, 101, 16,101–16,118.

---

T. Parsons, U.S. Geological Survey, Mail Stop 999, 345 Middlefield Road, Menlo Park, CA 94025, USA. (tparsons@usgs.gov)