Bulletin of the Seismological Society of America, Vol. 77, No. 4, pp. 1382–1399, August 1987

A GENERIC RECURRENCE INTERVAL DISTRIBUTION FOR EARTHQUAKE FORECASTING

By S. P. Nishenko and R. Buland

Abstract

Earthquake recurrence intervals for characteristic events from a number of plate boundaries are analyzed using a normalizing function, T/T_{ave}, where T_{ave} is the observed average recurrence interval for a specific fault segment, and T is an individual recurrence interval. The lognormal distribution is found to provide a better fit to the T/T_{ave} data than the more commonly used Gaussian and Weibull distributions, and it has an appealing physical interpretation. The observation of a small and stable coefficient of variation (the ratio of the standard deviation to the mean) of normalized data covering a wide range of recurrence intervals, seismic moments, and tectonic environments indicates that the standard deviation of the recurrence intervals for each fault segment is a fixed fraction of the corresponding average recurrence interval. Given that the distribution of T/T_{ave} data is approximately lognormal, the analysis is refined using the median recurrence interval, $\bar{T}$, rather than T_{ave}. An approximately optimal algorithm is derived for making stable estimates of $\bar{T}$, its standard deviation $\bar{\sigma}$, and the reliability of each $T/\bar{T}$ datum. By accounting for possible errors in the data, this algorithm treats both historical and geological data properly. All information is then combined to make an optimal estimate of the distribution of the $T/\bar{T}$ data. This refined distribution is finally used to estimate the probability of a recurrence in a future forecast time interval and the reliability of the forecast. It is found that the forecast interval must be short compared with $\bar{T}$ for the forecast to be statistically meaningful. In addition, the distribution of $T/\bar{T}$ data can be used to estimate an expected time and prediction time window for a future earthquake recurrence.

Introduction

The use of probabilistic, time-dependent descriptions of earthquake hazards along simple plate boundaries has grown during the last decade. These descriptions reflect an increase in the number of historic and geologic investigations of earthquake occurrence and the documentation of a relatively stable distribution of earthquake recurrence times along individual fault segments. The increase in the number and quality of these recurrence time data have, in turn, led to the development of a number of models (i.e., seismic gaps, characteristic earthquakes, time- and slip-predictable processes) to describe the physics and temporal behavior of the earthquake recurrence cycle. By a characteristic event, we mean an earthquake which repeatedly ruptures the same fault segment and whose source dimensions define that fault segment. For fault segments which are approximately decoupled from adjacent segments, such an event will be the largest earthquake to occur in the segment and will dominate strain release in the segment. While characteristic earthquakes are not strictly periodic, it appears that their individual recurrence times are distributed in a particular way about a mean or average recurrence time for each fault segment. This observed behavior has already been exploited in the successful forecast of the 1985 Valparaiso, Chile, earthquake (Nishenko, 1985) and is presently being used as the basis for the Parkfield, California, earthquake prediction experiment (Bakun and Lindh, 1985).

One of the basic problems which confronts the application of a probabilistic approach to a specific fault segment is the relatively small number of data from

which the recurrence interval distribution and hence, the forecast probability can be estimated. Identifying the statistical properties of a population based on a small data sample is a classic problem in statistical analysis. In general, distinguishing between a number of possible statistical distribution functions requires more data than are available for individual fault segments. Previous attempts at applying probabilistic methods to forecasting earthquake hazards have either used *ad hoc* choices of probability distributions [e.g., the normal or Gaussian distribution (see Lindh, 1983; Sykes and Nishenko, 1984)] or analogs to industrial failure time statistics [e.g., Weibull distributions (see Rikitake, 1976; Nishenko, 1985)]. While both approaches represent reasonable first attempts, the resolution possible with small data sets necessarily limited the reliability of the resulting forecasts.

In this study, we investigate the utility of a simple function, T/T_{ave}, to help define the overall distribution properties of earthquake recurrence. For a series of characteristic earthquakes occurring along a specific fault or plate boundary segment, T_{ave} is defined as the average recurrence interval observed for that segment where T is an individual recurrence interval. Previous investigations in Chile (Nishenko, 1985) and California (Sykes and Nishenko, 1984) have indicated, for a limited number of examples, that the standard deviation of T/T_{ave} appears to be roughly constant. This suggests that the T/T_{ave} data from all fault segments may follow the same statistical distribution. If this is a general property of characteristic earthquake recurrence intervals, then it allows the simultaneous analysis of all available recurrence interval data to determine this distribution. The greatly increased size of the data set will provide a more significant test for distinguishing between various distribution functions. In particular, a larger data set will allow a more thorough sampling of the tail areas of the distribution which have a large influence on the estimation of the statistical parameters of the function and hence on the resulting forecast estimates. For our purposes, the tails are the portions of the distribution which are far from mean and therefore have small relative probabilities.

Data Set and Initial Analysis

The data set for this analysis consist of recurrence intervals drawn from segments of plate boundaries with histories of two or more recurrences (i.e., three or more earthquakes which are documented to have ruptured the same fault segment). The 53 recurrence intervals listed in Table 1 include both historically recorded and geologically determined dates. The data set includes events originating from both convergent and transform plate boundaries, recurrence intervals varying from 12 to 1300 yr, and seismic moments varying from 10^{24} to 10^{30} dyne-cm (10^{17} to 10^{23} N-m). In other words, a range of two orders of magnitude in recurrence time and six orders of magnitude in seismic moment are available for our analysis. The primary assumption underlying the compilation of this diverse suite of recurrence intervals is that these data are all samples from a common population (i.e., a single common process). It is assumed that the recurrence history of each fault segment is complete for the time interval used and that each recurrence recorded is actually a characteristic earthquake.

While many more data are available (even for major plate boundaries in the circum-Pacific region) than are shown in Table 1, we have carefully selected only time intervals for specific fault segments which provide as complete and unambiguous a description of characteristic earthquake activity as possible. Data were rejected due to suspected *a priori* problems only. For great earthquakes, it is difficult to guarantee completeness due to the length of the required historical record or due

TABLE 1

RECURRENCE DATA FOR CHARACTERISTIC EARTHQUAKES

Region	Events	T	σ	T_{ave}	σ_{ave}/T_{ave}	$\bar{T}$	$\bar{\sigma}/\bar{T}$	Notes*
		Mexico						
Central Oaxaca	1928–1870	58		54.0	0.104	53.9	0.152	1
	1978–1928	50						
West Oaxaca	1894–1854	40		38.0	0.092	37.9	0.124	
	1928–1894	34						
	1968–1928	40						
San Marcos	1907–1845	62		56.0	0.152	55.7	0.152	
	1957–1907	50						
Petatlan	1943–1908	35		35.5	0.020	35.5	0.152	
	1979–1943	36						
		Chile						
Southern Chile	1737–1575	162		128.3	0.244	125.8	0.124	2
	1837–1737	100						
	1960–1837	123						
Concepcion	1657–1570	87		92.3	0.097	91.9	0.108	
	1751–1657	94						
	1835–1751	84						
	1939–1835	104						
Valparaiso	1730–1647	83		84.5	0.064	84.4	0.108	
	1822–1730	92						
	1906–1822	84						
	1985–1906	79						
		California						
Parkfield	1881–1857	24		21.8	0.330	20.8	0.096	3
	1901–1881	20						
	1922–1901	21						
	1934–1922	12						
	1966–1934	32						
Pallett Creek	1350–1080	270	44	194.3	0.292	192.5	0.151	4
	1550–1350	200	46					
	1720–1550	170	46					
	1857–1720	137	27					
		Japan						
Tokachi-Oki	1763–1677	86		97.0	0.138	96.4	0.124	5
	1856–1763	93						
	1968–1856	112						
Sanriku-Oki	1793–1611	182		143.0	0.385	137.6	0.152	5
	1897–1793	104						
Miyagi-Oki	1646–1616	30		40.2	0.301	38.8	0.072	6
	1678–1646	32						
	1731–1678	53						
	1770–1731	39						
	1835–1770	65						
	1861–1835	26						
	1897–1861	36						
	1936–1897	39						
	1978–1936	42						
Nankai Trough (AB)	1707–1605	102		113.7	0.258	111.3	0.124	7
	1854–1707	147						
	1946–1854	92						
Nankai Trough (CD)	1854–1707	147		118.5	0.340	115.0	0.152	7
	1944–1854	90						

TABLE 1—*Continued*

Region	Events	T	σ	T_{ave}	σ_{ave}/T_{ave}	$\bar{T}$	$\bar{\sigma}/\bar{T}$	Notes*
			Alaska					
Middleton Island	1941 B.C.–3141 B.C.	1200	283	1020.8	0.355	1131.8	0.155	8
	1551 B.C.–1941 B.C.	390	283					
	471 B.C.–1551 B.C.	1080	283					
	660 A.D.–471 B.C.	1130	283					
	1964 A.D.–660 A.D.	1304	250					

* Notes: 1 = Singh *et al.* (1981); 2 = Nishenko (1985); 3 = Bakun and Lindh (1985); 4 = Sieh (1984); 5 = Hatori (1976); 6 = Wesnousky *et al.* (1984); 7 = Ishibashi (1981); 8 = Plafker (1986); M. Rubin, personal communication, 1986.

to problems in interpreting the geological record. For smaller events, it is difficult to guarantee that all events are in fact in the same place. For example, the historic record of great earthquakes along the Nankai Trough, in southwestern Japan, extends to the 7th century, but is judged spatially and temporally complete only since the 17th century (Ishibashi, 1981). The segmented rupture characteristic of the Nankai Trough region is well documented (Ando, 1975). For the purposes of our analysis, we have treated the Nankaido (AB) and Tonankai (CD) segments as separate regions, even though historically they have either broken simultaneously or within a few days to years of one another since 1700. Rupture zone identification for the Tokachi-Oki and Sanriku-Oki segments are based on a comparison of tsunami source zones (Hatori, 1976). Other historic data for Chile, Mexico, Japan, and California are based, with the exception of the most recent events, on intensity reports and comparisons with recent shocks (see references listed in Table 1).

For geologically derived data, the problems are compounded by uncertainties in the dating and identification of all possible events at a single site. For example, the poor exposure of uplifted marine terraces in the Yakataga, Alaska, region suggested that the recurrence history might be incomplete (Plafker, 1986), and so these data have not been included in Table 1. At the Pallett Creek site along the San Andreas fault in California, 12 prior earthquakes have been recognized. However, the earthquake record is complicated by major erosional events, apparent changes in the average recurrence interval as a function of time, and uncertainties that all earthquakes are similar in size to the historic 1857 shock (Sieh, 1984). In this study, we have used only the latest five events, which all appear to have offsets comparable to the 1857 event, as a conservative sample of earthquake recurrence in this region. Uncertainties in the dating of prior events represent an additional source of error not encountered in the historic data. Times of geologically dated recurrences are shown in Table 1, after correction from ^{14}C years B.P. (i.e., before 1950) to calendar years, along with estimated errors associated with dating these events.

Figure 1a shows a histogram of the T/T_{ave} values listed in Table 1. Note that while the recurrence data cover a wide temporal range, these normalized values exhibit relatively little scatter, with the majority of the data located within ± 0.20 of $T/T_{ave} = 1$. Even using all available recurrences, the data in Figure 1a are still too sparse to draw useful conclusions about the probability density function, designated $f(T/T_{ave})$, directly. In other words, the apparent detail in Figure 1a is almost certainly not resolvable, in an inverse theory sense, and is obscuring the gross structure of the distribution. We can emphasize the gross structure in Figure 1a by integrating the data to provide an estimate of the cumulative density function (CDF), designated $F(T/T_{ave})$. The theory of order statistics provides an even better

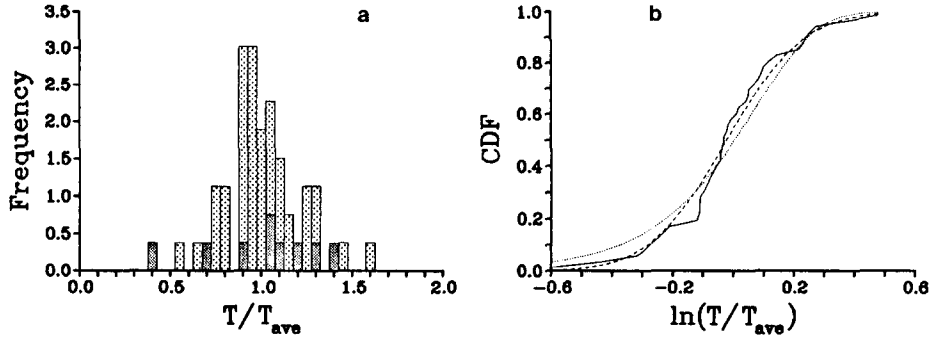

FIG. 1. Frequency histogram and cumulative distribution function for T/T_{ave}. (a) The relative frequency of T/T_{ave} values from Table 1 are plotted with light stippling denoting historic recurrence data and darker stippling denoting geologically estimated recurrence intervals. (b) Best fits to the cumulative distribution of historic $\ln(T/T_{ave})$ values. The solid line is the estimated distribution, the dashed and dotted lines show the best-fitting lognormal and Weibull functions, respectively.

method of mapping the observed data into an estimate of the CDF. In this case, the observed data provide a sampling of the independent variable, while a semi-empirical "plotting rule" specifies the associated probability level.

In the following, estimates of the T/T_{ave} CDF will be compared with the CDFs of various model distributions. We begin with the raw recurrence data in Table 1, which we will designate as $T_{ik} \pm \sigma_{ik}$; $i = 1, 2, \cdots, N_k$; $k = 1, 2, \cdots, K$, where K is the number of fault segments and N_k is the number of recurrence intervals for the kth segment. $T_{ik} \pm \sigma_{ik}$ means T_{ik} with an observed standard deviation of σ_{ik}. For historical data, $\sigma_{ik} = 0$. For geological data, σ_{ik} represents the uncertainty in the dating process. The total number of data is

$$N = \sum_{k=1}^{K} N_k. \tag{1}$$

For the kth segment, the arithmetic mean of the recurrence intervals is

$$T_{ave,k} = \frac{1}{N_k} \sum_{i=1}^{N_k} T_{ik} \tag{2}$$

and the standard deviation of the mean is

$$\sigma_{ave,k} = \left\{ \frac{1}{N_k - 1} \sum_{i=1}^{N_k} (T_{ik} - T_{ave,k})^2 \right\}^{1/2}. \tag{3}$$

Note that neither the estimate of $T_{ave,k}$ nor of $\sigma_{ave,k}$ account for the uncertainty in geologically dated recurrence intervals. T_{ave} and σ_{ave}/T_{ave} are shown in Table 1 for each fault segment. Notice that the coefficient of variation, σ_{ave}/T_{ave}, shows some scatter rather than being constant, as was initially assumed. However, due to the small number of recurrence data available from each fault segment, this scatter does not require the hypothesis to be rejected. The CDF estimate is constructed by defining $\tau_j \equiv T_{ik}/T_{ave,k}$ such that $\tau_1 \leqq \tau_2 \leqq \cdots \leqq \tau_N$. The jth datum is then a sample of the jth order statistic and is assigned a probability level $F_{j,N} = \dfrac{j - \frac{1}{2}}{N}$ (see Mann *et al.*, 1974, pp. 214–217). This is just one of a number of these so-called plotting rules. Their relative merits have been discussed at some length, in the context of a

related problem, by Knopoff and Kagan (1977). Numerical experiments show that differences across the whole spectrum of plotting rules contribute insignificantly to our conclusions. This is partly because our models are constructed by maximum-likelihood techniques rather than graphically using probability paper and partly because the problem we are addressing is much better posed than that of Knopoff and Kagan (1977).

Because there is no theory for treating recurrence interval data with errors, we begin by using historical data only and assuming that the T_{ave} estimates are exact. The CDF estimate for historical data, constructed using the procedure just described, is shown by the solid line in Figure 1b. Theoretical models have been constructed by maximum-likelihood estimation of relevant parameters. The Weibull distribution (dotted line) was fit by the procedure described by Mann *et al.* (1974, pp. 184–191). The lognormal distribution (dashed line) was fit as described by Mann *et al.* (1974, pp. 265–266). Note that these estimates depend only on the T/T_{ave} data (i.e., they are completely independent of any plotting rule).

Figure 1b shows that the Weibull function does not fit the empirical distribution very well. In fact, numerical experiments indicate that no linear combination of two and three parameter Weibull functions will fit significantly better. The difficulty is that the empirical distribution is light-tailed (has lower than normal relative probabilities) for recurrence intervals which are short compared with the average and heavy-tailed (has larger than normal relative probabilities) for recurrence intervals which are long compared with the average. The tails of the Weibull function have the opposite behavior. That is, they are heavy for short intervals and light for long intervals. The lognormal function, on the other hand, mimics the asymmetry of the empirical distribution quite well, as can be seen in Figure 1b. This fit implies that

$$f(T/T_{ave}) = \frac{T_{ave}}{T\sqrt{2\pi}}\, e^{-(\ln(T/T_{ave})-\mu_D)^2/2\sigma_D^2} \tag{4}$$

or that $\ln(T/T_{ave})$ is normally distributed with a mean of $\mu_D = -0.020$ and a standard deviation of $\sigma_D = 0.205$. The expected value of this distribution is $E(T/T_{ave}) = e^{\mu_D+\sigma_D^2/2} = 1.00$. In other words, T_{ave} is an unbiased estimator of $E(T)$ as it should be. σ_D describes the intrinsic variation in the length of recurrence intervals from cycle to cycle, which presumably reflects complexities in the accumulation and release of strain. In addition, it can be shown that $f(T) = f(T/T_{ave})/T_{ave}$ (e.g., Bury, 1975, pp. 58–60). Thus, the distribution of recurrence intervals for each fault segment is also lognormal and $\ln(T)$ is normally distributed with a mean of $\ln(T_{ave}) + \mu_D \equiv \ln(\bar{T})$ and a standard deviation of σ_D. Similarly, $T/\bar{T}$ is lognormal and $\ln(T/\bar{T})$ is normally distributed with a zero mean and the same standard deviation of σ_D. $\bar{T}$ turns out to be an estimate of both the median and the mode (which are coincident for the lognormal) of $f(T)$, but is different from (and smaller than) the expected value.

A similar standard deviation was found by Nishenko (1985) using a smaller data set from Chile and assuming a normal distribution. Note that for our analysis a Gaussian was not even considered as it has an inappropriate domain ($-\infty < T/T_{ave} < \infty$). The Weibull and lognormal functions, of course, have domains ($0 \leqq T/T_{ave} < \infty$) which are appropriate for the recurrence problem. Clearly, the requirement for a semi-infinite domain constrains all realizable failure time distributions (which are

nonzero everywhere in their domains) to be asymmetric. All curves in Figure 1b are plotted as functions of $\ln(T/T_{ave})$ for comparison with Figure 2a (see below).

While the initial results of our analysis appear promising, there are a number of points that must be addressed to develop a rigorous solution to the problem. First, although the model fits are based on maximum-likelihood methods, they do not incorporate all available information. In particular, they neglect the fact that not all data are equally reliable. This is true even for historical data as the T_{ave} values are determined by different numbers of observations for each fault segment. This difficulty is compounded by the fact that σ_{ave}, the estimate of reliability of T_{ave}, is itself poorly determined due to the small number of data constraining each T_{ave}. Further, without additional assumptions, it is not possible to derive the variance of each T/T_{ave} datum even if the variance of T_{ave} were known. Note that if the reliability of each T/T_{ave} datum were accounted for in the model fitting, then there would be no hinderance to combining both historical and geologically determined recurrence intervals. All of these caveats may be addressed by first assuming that the distribution sought is lognormal and then repeating the analysis.

Further Statistical Analysis

That fact that the T/T_{ave} data are lognormally distributed to a good approximation is not only interesting in itself, but provides valuable *a priori* information which should greatly improve the reliability of forecast probabilities. We proceed as follows. First, recognizing that $\bar{T}$ is a more natural parameter to estimate for the lognormal distribution than T_{ave}, what is known about the distribution of $T/\bar{T}$ data will be used to make optimal estimates of $\bar{T}$ and to make stable estimates of its standard deviation, $\tilde{\sigma}$. Second, the estimates of $\tilde{\sigma}$ will be used to derive standard deviations for each individual $T/\bar{T}$ datum. These procedures will be generalized so that historical and geological data are treated in an identical manner. Third, all data, with their respective standard deviations will be combined to make an improved estimate of the distribution of the $T/\bar{T}$ data. Finally, the best estimate of the distribution and of the $\bar{T}$'s will be used to compute conditional probabilities of a recurrence for each segment and any forecast interval. In fact, having good estimates of the $\tilde{\sigma}$'s will allow ranges of forecast probabilities (i.e., the reliability of each forecast) to be estimated.

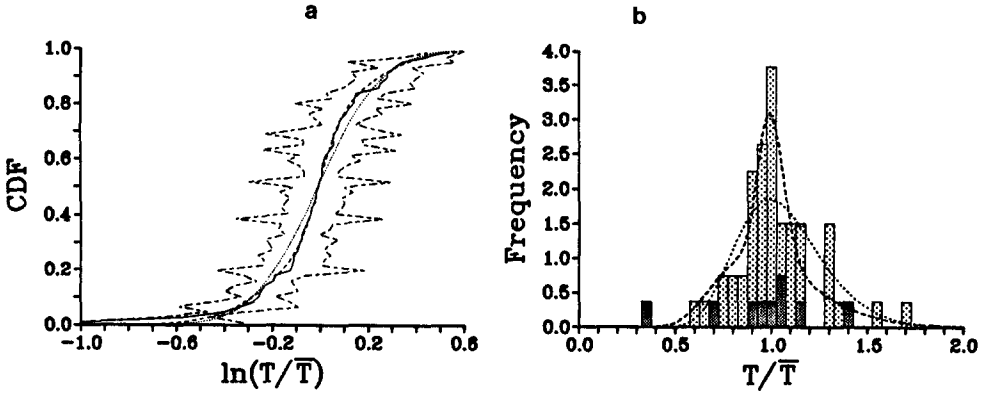

FIG. 2. Final fits to the recurrence data. (a) The estimated cumulative distribution of all of the $\ln(T/\bar{T})$ data is shown by the solid line. The long dash-short dash lines show the $\pm 1\tilde{\sigma}$ error bars on the data. The dotted and dashed lines show the fits of one and a linear combination of two lognormals, respectively. (b) The final frequency histogram of the $T/\bar{T}$ data is shown using the same conventions as in Figure 1a. The dotted and dashed lines show the probability density functions given by the one and two lognormal models, respectively.

We seek estimates of $\bar{T}_k$ and $\bar{\sigma}_k$ for each of the K segments. If all of the recurrence intervals for a fault segment are measured with equal accuracy, the maximum-likelihood estimate of $\bar{T}$ is given by

$$\ln(\bar{T}) = \frac{1}{N_k} \sum_{i=1}^{N_k} \ln(T_{ik}) \tag{5}$$

(Mann *et al.*, 1974, pp. 265–266). To account for the differences in the quality of the recurrence interval measurements, the var[ln(T_{ik})] must be estimated for each datum. In making these estimates, two sources of variability must be accounted for. The first source is the error in estimating the length of the recurrence interval T_{ik} (i.e., σ_{ik}). It measures the reliability of each datum. The second source is the natural variation among the recurrence intervals (i.e., σ_D). It measures how close an individual observation is likely to come to $\bar{T}$. The variance due to the first source can be derived from var(T_{ik}) = σ_{ik}^2 by applying the theorem

$$\text{var}[Z(z_1, z_2, \cdots, z_m)] = \sum_{j=1}^{m} \left(\frac{\partial Z}{\partial z_j}\right)^2 \text{var}(z_j) \tag{6}$$

(e.g., Eshbach, 1963, pp. 2–37), which yields var[ln(T_{ik})] = σ_{ik}^2/T_{ik}^2. The variance due to the second source can be derived by remembering that ln($T_{ik}/\bar{T}_k$) is normally distributed. Therefore, var[ln($T_{ik}/\bar{T}_k$)] = var[ln(T_{ik}) − ln($\bar{T}$)] = var[ln(T_{ik})] = $\sigma_D{}^2$. As the two sources of variability are independent, the total variance of ln(T_{ik}) for estimating ln($\bar{T}_k$) is just, var[ln(T_{ik})] = $\sigma_{ik}^2/T_{ik}^2 + \sigma_D{}^2$. The use of equation (6) requires that the random variables involved be normally distributed. This is only approximately true for the first source of variability. Even if the z_i's are normally distributed, equation (6) is still only an approximation. Below, we will show that these approximations have apparently not had a significant impact on our conclusions.

Ln($\bar{T}_k$) can now be estimated by means of weighted least squares. In the usual manner, the logarithm of each recurrence observation in the kth segment is considered to be an estimate of ln($\bar{T}_k$). However, because not all data are equally good estimates of ln($\bar{T}_k$), they should each be weighted by the inverse of their variance. This is done by solving

$$\min \sum_{i=1}^{N_k} \frac{[\ln(T_{ik}) - \ln(\bar{T}_k)]^2}{\sigma_{ik}^2/T_{ik}^2 + \sigma_D{}^2}. \tag{7}$$

Note that the form of the variance is such that poorly measured recurrence intervals are down-weighted relative to better measured ones. Further, for equally well (but not perfectly) measured intervals, longer intervals are weighted less heavily than shorter ones. This turns out to compensate for the nonlinearity of the logarithm.

Differentiating equation (7) with respect to ln($\bar{T}_k$),

$$\sum_{i=1}^{N_k} \frac{\ln(T_{ik}) - \ln(\bar{T}_k)}{\sigma_{ik}^2/T_{ik}^2 + \sigma_D{}^2} = 0, \tag{8}$$

or

$$\ln(\bar{T}_k) = \sum_{i=1}^{N_k} \frac{\ln(T_{ik})}{\sigma_{ik}^2/T_{ik}^2 + \sigma_D{}^2} \Bigg/ \sum_{i=1}^{N_k} \frac{1}{\sigma_{ik}^2/T_{ik}^2 + \sigma_D{}^2} \tag{9}$$

and

$$\bar{\sigma}_k^2 = \bar{T}^2 \Big/ \sum_{i=1}^{N_k} \frac{1}{\sigma_{ik}^2/T_{ik}^2 + {\sigma_D}^2}. \tag{10}$$

$\bar{T}$ and $\bar{\sigma}/\bar{T}$ are shown in Table 1 for each fault segment. Note that as expected, $\bar{T}$ is almost always smaller than T_{ave}. The one exception, Middleton Island, has such large errors that $\bar{T}$ and T_{ave} are not, in fact, significantly different. Also, due to the *a priori* information in the lognormal distribution, $\bar{\sigma}$ is much more stable than σ_{ave}. This can be seen in Table 1 by comparing the relative standard deviations (i.e., the coefficients of variation). σ_{ave}/T_{ave} scatters because the estimates depend on whether the small data set available for each segment happens to be drawn from the center of the distribution or from its tails. $\bar{\sigma}/\bar{T}$, on the other hand, depends primarily on the number of data drawn from the lognormal distribution rather than the consistency of the data.

In order to refine the estimate of the distribution, we wish to use all data (geological as well as historical) and to account for differences from segment to segment in our ability to estimate $\bar{T}$. As we now have sufficient information to estimate the var$[\ln(T_{ik}/\bar{T}_k)]$, which we will call $\tilde{\sigma}_{ik}^2$, for each datum, weighted least squares again presents itself as the method of choice. In this case, var(T_{ik}) must be considered to be σ_{ik}^2. The variability of T_{ik} as an estimator of $\bar{T}_k$ is not relevant for this problem, as we now need a measure of the reliability of each $T_{ik}/\bar{T}_k$ datum. Repeated applications of equation (6) yields,

$$\tilde{\sigma}_{ik}^2 = \frac{\sigma_{ik}^2}{T_{ik}^2} + \frac{\bar{\sigma}_k^2}{\bar{T}_k^2}. \tag{11}$$

Following the procedure outlined above, define $\tau_j \equiv T_{ik}/\bar{T}_k$ such that the τ_j are sorted into nondecreasing order and assign a probability level F_j to each τ_j. The least-squares problem is now formulated as

$$\min \sum_{j=1}^{N} \frac{(\ln(\tau_j) - \ln(F^{-1}(F_j))^2}{\tilde{\sigma}_j^2}. \tag{12}$$

Knopoff and Kagan (1977) point out that the use of least squares in conjunction with a plotting rule can be dangerous because order statistics are not, in general, normally distributed and are, in general, correlated with each other. Notice, however, that in equation (12), in order to take advantage of our knowledge of $\tilde{\sigma}_{ik}$, F^{-1} is being fit rather than F. In other words, the F_j estimated from the plotting rule are being taken as the independent variables, whose errors are, for the purpose of this analysis, negligible. This is a reasonable approximation because the aggregate sample size from all fault segments (which is used to estimate the order statistics) is large compared with the sample size from any one fault segment (which is used to estimate $\bar{T}$). Thus, errors due to the plotting rule will be small compared with $\tilde{\sigma}_{ik}$. This argument is supported by the insensitivity of our conclusions to the plotting rule chosen.

The fit shown in equation (12) has been done in the logarithmic domain primarily for numerical convenience. Had it been done in the linear domain, the result would necessarily have been the same, but $\tilde{\sigma}_{ik}$ would have had to have been suitably redefined. The determination of F by this procedure is implicit and requires

iteration. The solution of equation (12) defines a new value for σ_D which then changes $\bar{T}_k$, $\tilde{\sigma}_k$, and $\tilde{\sigma}_{ik}$ as defined in equations (9) to (11). This cycle is found to converge in several iterations. The estimate of the CDF is shown by the solid line in Figure 2a. The error bars (long dash-short dash lines) show $\pm 1\tilde{\sigma}$. The dotted line shows the fit to a lognormal with $\mu_D = -0.010$ and $\sigma_D = 0.215$. The major effect of using all data and accounting for the dominant errors in the data has been to increase σ_D by about 5 per cent. The change in μ_D is dominated by the redefinition of the data set from T/T_{ave} to $T/\bar{T}$.

Although this curve is quite adequate in the sense that it is entirely within $1\tilde{\sigma}$ of the estimated CDF, the systematics of the misfit encouraged us to try one more experiment. The dashed line shows a fit to a linear combination of two lognormals. This procedure is nonlinear, and the solution of equation (12) requires iteration [surprisingly, equation (12) yields linear equations when the model is a single lognormal]. A partial F-test analysis (Montgomery and Peck, 1982, pp. 132–137) indicates that the reduction of variance justifies the addition of the three new parameters at well above the 99th percentile. The probability density function for this model is given by

$$f(T/\bar{T}) = \frac{\bar{T}}{T\sqrt{2\pi}}\left\{\frac{\alpha}{\sigma_1} e^{-(\ln(T/\bar{T})-\mu_1)^2/2\sigma_1^2} + \frac{1-\alpha}{\sigma_2} e^{-(\ln(T/\bar{T})-\mu_2)^2/2\sigma_2^2}\right\}, \tag{13}$$

where $\mu_1 = -0.001$, $\sigma_1 = 0.066$, $\mu_2 = -0.019$, $\sigma_2 = 0.260$, and $\alpha = 0.341$. For comparison, it is still possible to define equivalents to μ_D and σ_D as $\mu_D = \alpha\mu_1 + (1 - \alpha)\mu_2 = -0.013$ and $\sigma_D = \sqrt{\alpha\sigma_1^2 + (1-\alpha)\sigma_2^2} = 0.215$. Figure 2b shows the final histogram of $T/\bar{T}$ data with the curves for the one and two lognormal fits superimposed as dotted and dashed lines, respectively.

The improvement gained by the use of a linear combination of lognormal functions is encouraging. Although the introduction of additional parameters to a regression problem never degrades the overall fit, unless the function added is well chosen, the improvement to the fit is usually found to be negligible. The significant improvement of the fit in this case provides intuitive support for the lognormal nature of the recurrence time distribution. This may indicate that deviations from the initial fit just happen to be lognormally distributed or it may imply that the distribution of recurrence times is controlled by several competing physical processes, of which the dominant ones lead to lognormal distributions. In any case, the parameters of the two lognormal fit are found to be much less stably determined than the linear combinations of μ_D and σ_D. Further, conclusions drawn from the two lognormal model are nearly identical to those drawn from the one lognormal model. Therefore, despite the statistical significance of the additional lognormal function, the added complexity is not considered to be worthwhile at the present.

From Mann *et al.* (1974, p. 265), the expected recurrence interval can also be written as

$$T_{exp} = E[T] = \bar{T}e^{\mu_D + \sigma_D^2/2}. \tag{14}$$

Thus, T_{exp} is about 1.0 per cent higher than $\bar{T}$. As pointed out above, μ_D should be zero if the $T/\bar{T}$ data is actually lognormally distributed. This is indeed the case if our complete algorithm is applied to historical data only. It turns out that the small, but nonzero value of μ_D for the full data set is due to the approximation var$[\ln(T_{ik})]$

$\approx \sigma_{ik}^2/T_{ik}^2$ derived from equation (6). This approximation has the effect of biasing $\bar{T}$ estimates for fault segments with geologically determined recurrence intervals which leads to a small average bias in our lognormal fit. As pointed out above, $f(T)$ is lognormal for each fault segment. It can be easily shown that the median of $f(T)$ according to our new model is approximately $\bar{T}(1 + \mu_D)$. Thus, our estimates of $\bar{T}$ are biased toward high values by about 1.3 per cent on the average. Notice that the bias in $\bar{T}$ has been explicitly accounted for in estimating T_{exp} so that T_{exp} will, on the average, be unbiased.

Conditional Probabilities

Given the date of the last event in a segment, $t_0 \pm \sigma_0$, and given $\bar{T} \pm \bar{\sigma}$ for the segment, we wish to compute P_C, the probability of a recurrence in the time interval $[t_1, t_2]$ conditional on the fact that the event has not happened yet (i.e., by the current time, t_c). The desired result is

$$P_C = \frac{\int_{t_1-t_0}^{t_2-t_0} f(T/\bar{T})\, dT}{\int_{t_c-t_0}^{\infty} f(T/\bar{T})\, dT} = \frac{F((t_2 - t_0)/\bar{T}) - F((t_1 - t_0)/\bar{T})}{1 - F((t_c - t_0)/\bar{T})}. \tag{15}$$

This is simply the usual conditional probability adapted to the generic distribution determined above. However, we can go one step further. As we have seen, $\ln(\bar{T})$ has been estimated by a linear combination of $\ln(T)$ values which are normally distributed. Therefore, by the central limit theorem, $\ln(\bar{T})$ will be normally distributed to a very good approximation. We can, therefore, estimate the range of P_C values corresponding to varying $\bar{T}$ within a predetermined confidence interval. If we know how σ_0 was determined, the same procedure may be applied to t_0. As most of the t_0 values are historical (i.e., $\sigma_0 = 0$), we will neglect σ_0 for simplicity in the following. Quoting a range of conditional probabilities in the manner of a weather forecast is a useful device for conveying information about the precision of the forecast. For example, a forecast of 0 to 60 per cent probability of a recurrence will (and should) be interpreted as saying that insufficient information is available to make a more useful statement. On the other hand, a forecast of 50 to 60 per cent will (and should) be taken quite seriously.

In our experiments, we have used a confidence interval of 90 per cent or $\ln(\bar{T}) \pm 1.645\bar{\sigma}/\bar{T}$. Because the central portion of the lognormal distribution is approximately normal, this is approximately equivalent to $\bar{T} \pm 1.645\bar{\sigma}$. Although P_C generally decreases as $\bar{T}$ increases, this is not always true locally. Therefore, an algorithm of determining P_C at five points within the confidence interval and then taking the minimum and maximum values as an estimate of the variability of the conditional probability was used. These experiments showed that even given $\bar{\sigma}/\bar{T}$ in the range of 7 to 15 per cent, the range of resulting conditional probabilities can be so large as to make the forecast essentially useless. Forecasts for all of the segments in our data set for four forecast time intervals, shown in Table 2, illustrate this point, ($t_1 = t_c = 1987$ for all forecast intervals). Clearly, the only sensible way to make the range of forecast probabilities small enough to be meaningful is to limit the size of the forecast interval. It turns out that, given $\bar{\sigma}/\bar{T}$, the range of forecast probabilities is only a function of $\Delta t/\bar{T}$, where $\Delta t = t_2 - t_1$. This is illustrated in Figure 3 where $P_{C,\max} - P_{C,\min}$ is contoured at the 10, 20, 30, and 40 per cent levels as a function of $\bar{\sigma}/\bar{T}$ and $\Delta t/\bar{T}$ (with $t_c = t_1$). Figure 3a shows the worst case taken over all values of $t_0/\bar{T}$ in the range (0.5, 1.5) and Figure 3b shows the average under the same conditions. Of course, an individual case may fortuitously yield a small range of

TABLE 2

EXPECTED RECURRENCE TIMES AND CONDITIONAL PROBABILITY ESTIMATES

Region	t_{pred}*	Conditional Probability (%)†			
		2 yr	5 yr	10 yr	20 yr
		Mexico			
Central Oaxaca	2033	0.0	0.0	0.0	0.0–6.8
West Oaxaca	2006	0.0–3.4	0.2–14.0	2.0–43.6	24.7–89.1
San Marcos	2013	0.0–5.2	0.1–16.3	0.6–39.8	6.8–79.8
Petatlan	2015	0.0	0.0–0.1	0.0–3.8	1.8–61.1
		Chile			
Southern Chile	2087	0.0	0.0	0.0	0.0
Concepcion	2032	0.0–1.1	0.1–3.5	0.2–10.0	1.7–31.3
Valparaiso	2070	0.0	0.0	0.0	0.0
		California			
Parkfield	1987	21.7–49.5	52.5–84.1	84.7–98.2	99.2–100.0
Pallett Creek	2052	0.1–3.8	0.2–9.7	0.4–19.5	1.4–38.4
		Japan			
Tokachi-Oki	2066	0.0	0.0	0.0	0.0–0.1
Sanriku-Oki	2036	0.1–4.9	0.2–12.5	0.5–25.3	2.0–49.4
Miyagi-Oki	2017	0.0	0.0	0.0–0.4	3.4–23.6
Nankai Trough (AB)	2059	0.0	0.0–0.1	0.0–0.6	0.0–4.5
Nankai Trough (CD)	2061	0.0–0.1	0.0–0.3	0.0–1.3	0.0–7.8
		Alaska			
Middleton Island	3111	0.0	0.0	0.0	0.0

* $t_{pred} = t_0 + T_{exp}$, where t_0 is the date of the last event and T_{exp} is from equation (15) and Table 1.

† Calculated from present (i.e., $t_c = t_1 = 1987$). Range in values reflect 90 per cent confidence interval for t_{pred}.

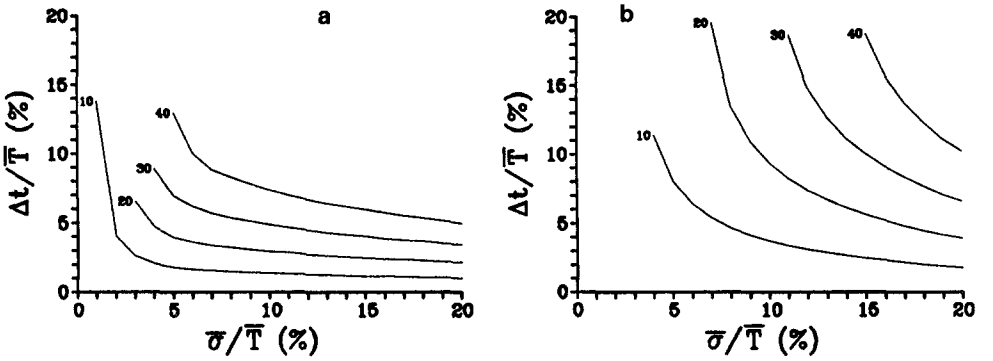

FIG. 3. Relative forecast intervals as a function of the relative standard deviation in $\bar{T}$. (a) Worst case conditional probability ranges corresponding to ±90 percentile extremes in $\bar{T}$ are contoured at the 10, 20, 30, and 40 per cent levels. The abscissa is the coefficient of variation of $\bar{T}$ in per cent and the ordinate is the ratio of the forecast time interval and $\bar{T}$ in per cent. (b) Average ranges in conditional probability estimates are displayed as in (a).

conditional probabilities, especially if the forecast interval lies in the tails of the recurrence interval distribution. However, in general, a meaningful conditional probability cannot be guaranteed unless the forecast interval is chosen to be compatible with $\bar{\sigma}$ in accordance with Figure 3.

EARTHQUAKE PREDICTION

Recently, several studies have made earthquake predictions using exactly the same information (i.e., the recurrence intervals of characteristic earthquakes) used above to derive forecast probabilities (e.g., Bakun and Lindh, 1985; McNally, 1985). This raises the question: what is the relationship between earthquake forecasting and earthquake prediction? The potential equivalence between these two methodologies is recognized and briefly discussed by Wallace *et al.* (1984). In the context of the model presented above, this equivalence can be made precise. The predicted time of the next characteristic event will just be $t_{pred} = t_0 + T_{exp}$ [the date of the last characteristic event plus the expected time of the recurrence interval as given by equation (14)]. The prediction window, W_{pred}, will be defined to be the time interval $(t_{pred} - \eta, t_{pred} + \eta)$, where η will be taken to be the 90 per cent confidence interval of T_{exp}. As for $\ln(\bar{T})$, $\ln(T_{exp})$ will be normally distributed. T_{exp} will be approximately normally distributed and $\eta \approx 1.645\ \mathrm{var}(T_{exp})^{1/2}$. As in the analysis above, two independent sources of variability must be accounted for in estimating $\mathrm{var}(T_{exp})$: (1) the ability to estimate $\bar{T}$ (i.e., $\bar{\sigma}$) and (2) the intrinsic variability of recurrence intervals (i.e., σ_D). Combining these contributions yields

$$\mathrm{var}(T_{exp}) = \{\bar{\sigma}^2 + \bar{T}^2\sigma_D{}^2\}e^{2\mu_D+\sigma_D{}^2} = \{\bar{\sigma}^2/\bar{T}^2 + \sigma_D{}^2\}T_{exp}^2. \tag{16}$$

t_{pred}, η, and W_{pred} for the next characteristic earthquake have been computed for all of the segments in our data set. These results are shown in Table 3 and are compared to previously published estimates. Note that given $\bar{\sigma}/\bar{T}$, the size of the prediction window, $\Delta W_{pred} = 2\eta$, is linear in $\bar{T}$. However, the distinction between a long-term earthquake prediction and a long-term earthquake potential according to the definition of Wallace *et al.* (1984) depends only on the magnitude of ΔW_{pred} being less than or greater than a "few" decades, respectively. Therefore, t_{pred} will represent an earthquake prediction if $\bar{T}$ is sufficiently short. This relationship is shown graphically in Figure 4 where a few has been taken to be three. Even though some of the predictions given in Table 3 qualify as long-term earthquake predictions in a formal sense, they are of a singularly uninteresting nature. This is due to the fact that ΔW_{pred} is approximately 74 to 86 per cent of T_{exp} (for $\bar{\sigma}/\bar{T}$ in the range 7 to 15 per cent). Over a large number of recurrences, the average recurrence interval will be T_{exp}. Therefore, on the average $100\eta/T_{exp}$ per cent of each cycle would be spent in a prediction window. In other words, people living in the predicted epicentral region would spend about 40 per cent of their lives in a prediction window and about 60 per cent out of one.

Equations (14) and (16) differ from other prediction results in two significant ways. The first and most obvious difference is that our estimate of T_{exp} will be more stable than T_{ave}, by virtue of the additional information available from the recurrence interval distribution. The second difference is more subtle. Both Bakun and Lindh (1985) and McNally (1985) use procedures which are parallel to standard error analysis. In essence, W_{pred} is the confidence interval of a predicted datum derived from the standard deviation of the mean of a fit to the existing data. The fallacy of this approach can be seen by considering the following thought experiment. Suppose that it was possible to obtain data for as many recurrence intervals as one desired. Then, the standard deviation of the mean of the fit and, hence, the prediction window could be made as small as desired. However, the scatter in historically recorded recurrence intervals is real. Thus, there must be some minimum below which ΔW_{pred} can never fall. From equation (16), this minimum is given by

TABLE 3

COMPARISON OF EARTHQUAKE PREDICTION TIME WINDOWS

Region	t_{pred}*	η†	W_{pred}‡	Other Estimates§	Notes¶
		Mexico			
Central Oaxaca	2033	23.6	2009–2057		
West Oaxaca	2006	15.7	1990–2022		
San Marcos	2013	24.5	1988–2037		
Petatlan	2015	15.6	1999–2031		
		Chile			
Southern Chile	2087	52.1	2035–2139	2088 ± 51.0	1
Concepcion	2032	36.9	1995–2069	2031 ± 12.3	1
Valparaiso	2070	33.8	2036–2104	2071 ± 8.2	1
		California			
Parkfield	1987	8.2	1979–1995	1988 ± 4.3	2
Pallett Creek	2052	84.3	1968–2136	1957–2057	3
		Japan			
Tokachi-Oki	2066	39.9	2026–2106	2065 ± 18.1	4
Sanriku-Oki	2036	60.4	1976–2096	2040 ± 41.0	4
Miyagi-Oki	2017	14.7	2002–2032	2018 ± 11	4
Nankai Trough (AB)	2059	46.1	2013–2105		
Nankai Trough (CD)	2061	50.5	2010–2111	2007 ± 28	4#
		Alaska			
Middleton Island	3111	500.5	2610–3611		

* $t_{pred} = t_0 + T_{exp}$, where t_0 is the date of the last event, and T_{exp} is from equation 15 and Table 1.

† η is the 90 per cent confidence interval of T_{exp}.

‡ $W_{pred} = t_{pred} - \eta, t_{pred} + \eta$.

§ Published σ corrected to 90 per cent confidence intervals.

¶ Notes: 1 = Nishenko (1985): 2 = Bakun and Lindh (1985); 3 = Sieh (1984); 4 = Wesnousky *et al.* (1984); 4# = Wesnousky *et al.* (1984), date based on time-predictable model.

$\Delta W_{pred} \approx 3.290\sigma_D T_{exp}$ (*i.e.*, $\bar{\sigma} \to 0$ as $N_k \to \infty$) which represents the variability of the recurrence intervals alone. That is, if one were to perform the same thought experiment for W_{pred} derived from equation (16), the minimum ΔW_{pred} is about 71 per cent of T_{exp}.

Bakun and Lindh (1985) justify their statistics by hypothesizing a physical mechanism which explains the observed pattern of "normal" recurrence intervals and paired short and long intervals. Their model implies that σ_D for Parkfield is much smaller than the generic value derived here. However, it is still an approximation to neglect it. On the other hand, if the apparent pattern of Parkfield recurrence intervals is hypothesized to be a fortuitous byproduct of the small number of observations, then the Parkfield data fit our model quite well.

DISCUSSION

We have examined the distribution of earthquake recurrence intervals along a number of simple plate boundaries using a normalizing function, T/T_{ave}, and have found the lognormal distribution to provide a significantly better fit to the data than either of the previously suggested Gaussian or Weibull distribution functions. As the natural independent variable for this analysis was $\ln(T/T_{ave})$, it was equiv-

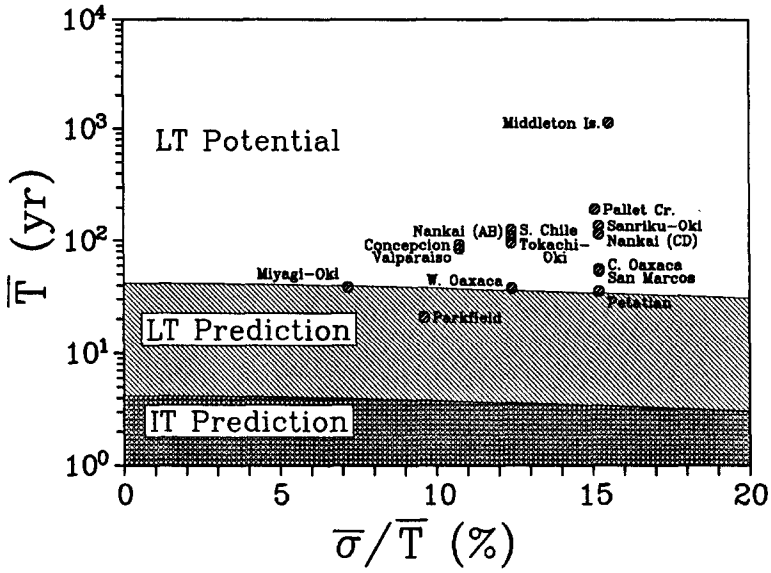

FIG. 4. Definition of intermediate term (IT), long-term (LT) predictions and long-term potentials, as a function of the coefficient of variation of $\overline{T}$. The relative location of the recurrence data used in this analysis illustrates which earthquakes may be "predicted" and which events may be "forecast" using historic and geologic data.

alent to measuring the deviation of the logarithms of individual recurrence intervals about the logarithm of the average recurrence interval. We found that the majority of the normalized data exhibit little scatter, having a coefficient of variation of 0.205 for historic data and 0.215 for a combination of both historic and geologic data. While the original data span two orders of magnitude in recurrence time, six orders of magnitude in seismic moment, and originate from both convergent and transform plate boundaries, the small and relatively stable coefficient of variation appears to confirm our basic assumption that we are sampling from a common population. More importantly, this analysis indicates that the standard deviation of the recurrence intervals is a fixed fraction of the average recurrence interval. This seems to imply a relatively uniform, well-behaved, scale-independent physical structure underlying the characteristic earthquake cycle for simple plate boundaries worldwide. Exactly which factors contribute to the observed variations are not clear at present. Changes in source size from event to event, the basis for the proposal of the time-predictable earthquake model of Shimazaki and Nakata (1980), is probably part of the explanation. Note in Table 1, that the deviations from the mean for the Nankai Trough region (where the time-predictable model was formulated) are similar to those observed elsewhere. Another possible factor, which is appealing in view of the shape of the lognormal distribution, is nonlinear strain accumulation (Thatcher, 1983, 1984) in which the strain field initially recovers rapidly following a characteristic earthquake and then progressively slower until the next characteristic event. Another model which may contribute to the variation in recurrence intervals is contagion whereby the release of strain energy on one fault segment may increase or decrease the strain on near by fault segments (J. H. Dieterich, 1986, personal communication). With regards to the analysis presented here, the heavy-tailed nature of the lognormal distribution favors later than average recurrences over earlier than average events. Hence, in an aggregate sense, it may be suggested that stress changes and strain transients tend to retard rather than accelerate earthquake recurrence.

Another interesting aspect of the observation that the standard deviation of recurrence intervals scales with the average recurrence interval impacts the definitions of earthquake prediction and earthquake forecasting. It appears that, using the statistics of recurrence intervals of characteristic earthquakes alone, only characteristic earthquakes with short average recurrence intervals (and hence small standard deviations) can be predicted, while those events which, by virtue of their size, have longer recurrence intervals can only be forecast. The physical behavior in either case appears to be the same, but because of the constraints imposed on earthquake prediction by the needs of society, these two cases will be categorized and treated quite differently. The limited prediction capability of recurrence interval data reinforces the need for other types of data to provide the reliable intermediate and short-term precursory information needed to meet societal requirements.

The use of a lognormal function for describing the distribution of T/T_{ave} can also be motivated by a comparison with fatigue and failure time analysis in the engineering literature. Both the Weibull and lognormal distributions have been found to be applicable to a wide range of materials and testing conditions. Both provide good fits to the data in the 10 to 90 per cent probability range, and significant differences between the two are only distinguishable with large sample sizes (i.e., $N > 30$). Lognormal distributions are commonly used to describe the measured distribution of fatigue crack sizes, growth rates, and lives or failure times (Heller, 1972; Little and Ekvall, 1981). In addition to an empirical basis for adopting the lognormal distribution, it can also be derived from a physical model for fatigue crack growth. According to the proportional effect model of Kao (1965), the change in crack size at the nth step is randomly proportional to the size of the crack at the previous step. In other words, $\Delta X_{i-1} = X_i - X_{i-1} = C_i X_{i-1}$, where X_i is the crack size at the ith step and C_i, the proportional growth constant, is a random variable. For incremental changes in crack size,

$$\sum_{i=1}^{n} C_i = \sum_{i=1}^{n} \frac{\Delta X_{i-1}}{X_{i-1}}. \tag{17}$$

In the limit as $\Delta X_{i-1} \to 0$, and n becomes large,

$$C_n = \sum_{i=1}^{n} C_i \to \int_{x_0}^{x_n} \frac{dx}{X} = \ln(X_n/X_0). \tag{18}$$

Due to the central limit theorem, C_n will be normally distributed and X_n/X_0 will be lognormally distributed. Das and Scholz (1981) have applied two basic fracture mechanics concepts (stress intensity and crack velocity) to describe a number of earthquake-related phenomena (foreshocks, aftershocks, multiple events, slow earthquakes, etc.). Neglecting the effects of friction in their model, crack growth is dependent on the magnitude of the stress intensity factor k, where $k \approx \sqrt{X}$, and X is the crack size. From the previous derivation, if X is lognormally distributed, it may be expected that other crack and earthquake-related parameters will also have lognormal distributions. When frictional effects are considered, however, it is found that the magnitude of the stress change or transient dominates the logarithm of time to failure (Dieterich, 1986).

In conclusion, the analysis of a large number of earthquake recurrence intervals, in a normalized reference frame, has lead to the development of a scale invariant description of characteristic earthquake recurrence along simple plate boundaries. Whether scale invariance is a characteristic feature of recurrence for all earthquakes

or just a subset (i.e., interplate events) must await further analysis. Knowledge about the underlying distribution of earthquake recurrence intervals enables an improvement in recurrence time estimates based on small samples and quantitative descriptions of the reliability of these estimates. Finally, the development of a theory for treating recurrence interval data with errors enables the combination of geologic and historic data, thus expanding the data bases available for probabilistic forecasts of earthquake hazards.

ACKNOWLEDGMENTS

Conversations with Jim Dieterich, Klaus Jacob, Dave Perkins, George Plafker, and Meyer Rubin were invaluable during the formative stages of this study. We thank Bill Spence and Dave Perkins for critically reading the manuscript. This study was funded by the Agency for International Development Agreement BOF-000-P-IC-4051-00 and the U.S. Geological Survey.

REFERENCES

Ando, M. (1975). Source mechanisms and tectonic significance of historical earthquakes along the Nankai trough, Japan, *Tectonophysics* **27,** 119–140.

Bakun, W. H. and A. G. Lindh (1985). The Parkfield, California earthquake prediction experiment, *Science* **229,** 619–624.

Bury, K. V. (1975). *Statistical Models in Applied Science*, John Wiley and Sons, New York, 625 pp.

Das, S. and C. H. Scholz (1981). Theory of time-dependent rupture in the earth, *J. Geophys. Res.* **86,** 6039–6051.

Dieterich, J. H. (1986). A model for the nucleation of earthquake slip, in *Earthquake Source Mechanics, Maurice Ewing Series*, vol. 6, S. Das, J. Boatwright, and C. H. Scholz, Editors, American Geophysical Union, Washington, D.C., 37–47.

Eshbach, O. W. (Editor) (1963). *Handbook of Engineering Fundamentals*, John Wiley and Sons, New York, 1314 pp.

Hatori, T. (1976). Propagation of tsunamis from sources off the Pacific coast of northeastern Japan, *Bull. Earthquake Res. Inst., Tokyo Univ.* **51,** 197–207.

Heller, R. A. (1972). Probabilistic aspects of fatigue, *ASTM Special Technical Publication 511*, American Society for Testing and Materials, Philadelphia, Pennsylvania, 203 pp.

Ishibashi, K. (1981). Specification of a soon-to-occur seismic faulting in the Tokai district, central Japan, based upon seismotectonics, in *Earthquake Prediction, An International Review, Maurice Ewing Series*, vol. 4, D. W. Simpson and P. G. Richards, Editors, American Geophysical Union, Washington, D.C., 297–332.

Kao, J. H. K. (1965). Statistical models in mechanical reliability, *Proceedings of the Eleventh National Symposium on Reliability and Quality Control*, 240–247.

Knopoff, L. and Y. Kagan (1977). Analysis of the theory of extremes as applied to earthquake problems, *J. Geophys. Res.* **82,** 5647–5657.

Lindh, A. G. (1983). Preliminary assessment of long-term probabilities for large earthquakes along selected fault segments of the San Andreas fault system in California, *U.S. Geol. Surv., Open-File Rept. 83-63*, 1–15.

Little, R. E. and J. C. Ekvall (Editors) (1981). Statistical analysis of fatigue data, *ASTM Special Technical Publication 744*, American Society for Testing and Materials, Philadelphia, Pennsylvania, 143 pp.

Mann, N. R., R. E. Schafer, and N. D. Singpurwalla (1974). *Methods for Statistical Analysis of Reliability and Life Data*, John Wiley and Sons, New York, 564 pp.

McNally, K. C. (1985). San Andreas fault, central California, in *Minutes of the National Earthquake Prediction Council*, C. F. Shearer, Editor, *U.S. Geol. Surv., Open-File Rept. 85-754*, 335–349.

Montgomery, D. C. and E. A. Peck (1982). *Introduction to Linear Regression Analysis*, John Wiley and Sons, New York, 504 pp.

Nishenko, S. P. (1985). Seismic potential for large and great interplate earthquakes along the Chilean and southern Peruvian margins of South America: a quantitative reappraisal, *J. Geophys. Res.* **90,** 3589–3615.

Plafker, G. (1986). Geologic studies related to earthquake potential and recurrence in the "Yakataga seismic gap," in *Minutes of the National Earthquake Prediction Council*, C. F. Shearer, Editor, *U.S. Geol. Surv., Open-File Rept. 86-92*, 135–143.

Rikitake, T. (1976). Recurrence of great earthquakes at subduction zones, *Tectonophysics* **35,** 335–362.

Shimazaki, K. and T. Nakata (1980). Time-predictable recurrence model for large earthquakes, *Geophys. Res. Letters* **7,** 279–282.

Sieh, K. (1984). Lateral offsets and revised dates of large prehistoric earthquakes at Pallett Creek, southern California, *J. Geophys. Res.* **89,** 7641–7670.

Singh, S. K., L. Astiz, and J. Havskov (1981). Seismic gaps and recurrence periods of large earthquakes along the Mexican subduction zone: a reexamination, *Bull. Seism. Soc. Am.* **71,** 827–843.

Sykes, L. R. and S. P. Nishenko (1984). Probabilities of occurrence of large plate rupturing earthquakes for the San Andreas, San Jacinto and Imperial faults, California, 1983–2003, *J. Geophys. Res.* **89,** 5905–5927.

Thatcher, W. (1983). Nonlinear strain buildup and the earthquake cycle on the San Andreas fault, *J. Geophys. Res.* **88,** 5893–5902.

Thatcher, W. (1984). The earthquake deformation cycle, recurrence and the time-predictable model, *J. Geophys. Res.* **89,** 5674–5680.

Wallace, R. E., J. F. Davis, and K. C. McNally (1984). Terms for expressing earthquake potential, prediction, and probability, *Bull. Seism. Soc. Am.* **74,** 1819–1825.

Wesnousky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda (1984). Integration of geologic and seismologic data for the analysis of seismic hazard: a case study of Japan, *Bull. Seism. Soc. Am.* **74,** 687–708.

NATIONAL EARTHQUAKE INFORMATION CENTER
U.S. GEOLOGICAL SURVEY
BOX 25046, STOP 967
DENVER FEDERAL CENTER
DENVER, COLORADO 80225

Manuscript received 10 August 1986