peratures.

Once T_* had been found, it was no longer necessary to cool slowly from a high initial temperature. By starting at $T = 1.0$ and cooling to $T = T_*$ in 100 sweeps, then remaining at T_* for another 50 sweeps, we obtained profiles whose energies differed from the energy of the true profile by less than 5%. Below T_*, very slow cooling for several hundred sweeps gave profiles with energies less than 1% above the energy of the true profile. We found many low-E profiles while constructing Fig. 4.

In problems of the type considered here, the most important question is not what single profile fits the data best but rather what set of profiles fits the data well. The latter question is addressed by estimating the a posteriori probability distribution (PPD) for the soundspeed at each depth. Mathematically, the PPD for the c_i, the soundspeed at the ith depth, is given by

$$p(c_i|U_1, U_2) = \sum_{c_1}\sum_{c_2}\dots\sum_{c_{i-1}}\sum_{c_{i+1}}\dots$$
$$\dots\sum_{c_M} p(\mathbf{c}|U_1, U_2) \qquad (8)$$

Numerically, we construct this PPD by the following recipe, which we call Gibbs-weighted graph-binning. Let A be the set of profiles found at $T = T_*$ during construction of Fig. 4 and let $B \subset A$ be those profiles in A for which $c_i = \gamma$, where γ is one of the allowed values for c_i. First, we estimate the partition function at $T = T_*$ by

$$Z_* \approx \sum_{\mathbf{c} \in A} \exp[-E(\mathbf{c})/T_*] \qquad (9)$$

Then the PPD for $\mathbf{c}_i$ is estimated by

$$p(\mathbf{c}_i = \gamma|U_1, U_2) \approx \frac{1}{Z_*}\sum_{\mathbf{c} \in B} \exp[-E(\mathbf{c})/T_*] \qquad (10)$$

The PPDs for our example are shown in Fig. 3. Note that Fig. 3 was constructed using only the 125 profiles found at $T = T_*$ during construction of Fig. 4, and that we constructed Fig. 4 by using only 10 (temperatures) $\times$ 25 (sweeps per temperature) $\times$ 5 = 1250 sweeps. The PPDs in Fig. 3 look the same for any T within our error bounds for T_*.

Another useful quantity is the mean profile at $T = T_*$, estimated by

$$\langle \mathbf{c} \rangle \approx \frac{1}{Z_*}\sum_{\mathbf{c} \in A} \mathbf{c} \exp[-E(\mathbf{c})/T_*] \qquad (11)$$

The ith component of $\langle \mathbf{c} \rangle$ is the centroid of the PPD for $\mathbf{c}_i$. Figure 3 shows that $\langle \mathbf{c} \rangle$ is not as good an estimator of the true profile as is the profile whose ith component is the peak value of the ith PPD. For example, in Fig. 3 the PPDs between 300 m deep and 350 m deep have their peaks at the true profile, but their centroids are to the left of the true profile indicated by the thin line.

Our numerical experiments indicate that for SA optimization and inversion problems one should spend most of the sweep budget determining T_*. Then for optimization problems, the conventional slow cooling schedule should be replaced by a schedule with rapid cooling to T_*. The process of finding T_* gives all the information necessary to estimate the PPD, a quantity much more useful in inversion than the profile with lowest energy.

REFERENCES AND NOTES

1. S. Kirkpatrick *et al.*, *IBM Res. Rep. RC 9355* (1982); S. Kirkpatrick, C. D. Gelatt, Jr., M. P. Vecchi, *Science* **220**, 671 (1983).
2. V. Černy, *J. Optimization Theory Appl.* **45**, 41 (1985).
3. C. Rebbi, in *Applications of the Monte Carlo Method*, K. Binder, Ed. (Springer-Verlag, New York, 1984), pp. 277–298; M. Creutz, *Phys. Rev. D* **21**, 2308 (1980).
4. N. Metropolis *et al.*, *J. Chem. Phys.* **21**, 1087 (1953).
5. P. J. M. van Laarhoven and E. H. L. Aarts, *Simulated Annealing* (Reidel, Boston, 1987).
6. D. H. Rothman, *Geophysics* **50**, 2784 (1985); *ibid.* **51**, 332 (1986).
7. S. Geman and D. Geman, *IEEE Trans. Pattern Anal. Mach. Intell.* **6**, 721 (1984).
8. H. Szu and R. Hartley, *Phys. Lett. A* **122**, 157 (1987).
9. J. D. Nulton and P. Salamon, *Phys. Rev. A* **37**, 1351 (1988); I. Matsuba, *ibid.* **39**, 2635 (1989).
10. W. Menke, *Geophysical Data Analysis: Discrete Inverse Theory* (Academic Press, New York, 1984); A. Tarantola, *Inverse Problem Theory* (Elsevier, New York, 1987).
11. S. Mallick and L. N. Frazer, *Geophysics* **53**, 479 (1988).
12. J. A. Carter and L. N. Frazer, *ibid.* **49**, 46 (1984).

13. Supported by the U.S. Office of Naval Research and by American Chemical Society grant PRF-20029-AC2. Hawaii Institute of Geophysics contribution 2328.

18 April 1990; accepted 29 June 1990

Forecasting Damaging Earthquakes in the Central and Eastern United States

S. P. NISHENKO AND G. A. BOLLINGER

Analysis of seismograph network data, earthquake catalogs from 1727 to 1982, and paleoseismic data for the central and eastern United States indicate that the Poisson probability of a damaging earthquake (magnitude ≥ 6.0) occurring during the next 30 years is at a moderate to high level (0.4 to 0.6). When differences in seismic wave attenuation are taken into account, the central and eastern United States has approximately two-thirds the likelihood of California to produce an earthquake with comparable damage area and societal impact within the next 30 years.

SINCE 1727 THERE HAVE BEEN SEVEN earthquakes with magnitudes greater than m_b (body-wave magnitude) 6.0 in the central and eastern United States. The largest five of these events [1811 to 1812 New Madrid, Missouri; m_b 7.0, 7.1, 7.2, 7.3 (*1*); and 1886 Charleston, South Carolina; m_b 6.7 (*2*)] occurred during the 19th century. In the 95 years that have elapsed since the last damaging $m_b \geq 6.0$ earthquake (*3*), the central and eastern United States has undergone an era of rapid urban growth along with the development of nuclear power plants, energy distribution systems, large reservoirs, and transportation and communications networks. The observations that earthquakes have occurred in this region in the past and that they are capable of producing structural damage over larger areas than their counterparts of similar size in California raise the obvious question: What are the chances for the occurrence of another $m_b \geq 6.0$ earthquake in the region east of the Rocky Mountains during the next few decades? We define a damaging earthquake as having a $m_b > 6.0$ (seismic moment magnitude, $\mathbf{M} \geq 6.1$), and consider exposure windows for the next 10, 30, 50, and 100 years.

Earthquake hazard assessments are primarily based on average rates of earthquake occurrence. Although information about earthquake recurrence times is lacking for specific fault zones in the central and eastern United States, regional rates of seismic activity are commonly estimated for hazards and engineering purposes by application of the Gutenberg-Richter frequency-magnitude or B value relationship, $\log N_C = A - BM$, where N_C is the cumulative number of earthquakes greater than or equal to a particular magnitude, M. The constants, A and B, are determined primarily from the rates of occurrence of smaller magnitude earthquakes and are used to estimate the rates of occurrence of infrequent larger magnitude

S. P. Nishenko, National Earthquake Information Center, U.S. Geological Survey, Denver, CO 80225.
G. A. Bollinger, U.S. Geological Survey and Seismological Observatory, Virginia Polytechnic Institute and State University, Blacksburg, VA 24061.

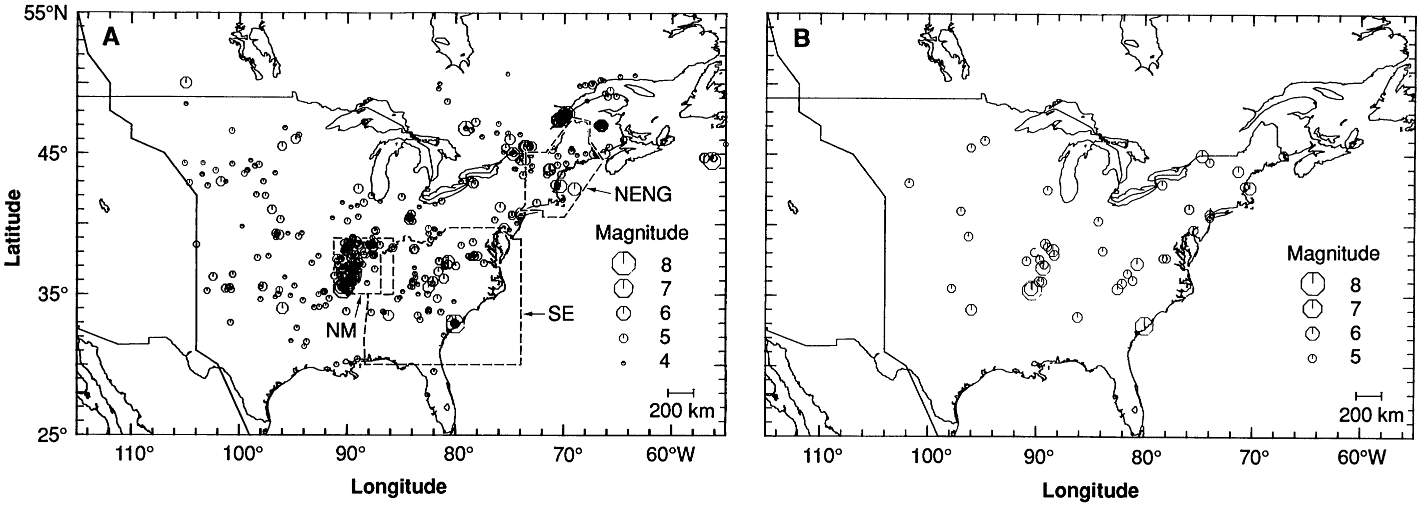

Fig. 1. Seismicity of eastern North America. (**A**) Epicenters of earthquakes with $m_b \geq 4$ from 1638 to 1982 in the United States and Canada (*10*). Sizes of symbols are proportional to the magnitude (see scale). Western limit of catalog area is shown by solid line. Dashed lines show the spatial limits of regional seismograph network data; NENG, New England; SE, Southeast United States; NM, New Madrid. Earthquakes with $m_b \geq$ **5.0** constitute the ENA (eastern North America) catalog of Table **1**. (**B**) Epicenters of earthquakes with $m_b \geq$ **5.0** from 1727 to **1980** in the central and eastern United States. These events constitute the CEUS catalog of Table 1.

events.

For many fault systems along simple plate boundaries, current research on probabilistic earthquake hazard assessment has focused on time-dependent models, where the probability of a future event is a function of both the time since the last earthquake and the average recurrence time (*4*). Typically, hazard estimates are presented in terms of conditional probability, which is the probability that an earthquake will occur within some time interval, t to $t + \Delta t$, conditional on the event not having occurred before time t (*5*).

In contrast to the major exposed fault systems in California, much less is known about the location and behavior of the buried seismogenic faults in the central and eastern United States. Although the general lack of information about the occurrence of large earthquakes in this region currently argues against the development and widespread application of time-dependent models, such models have been proposed for the New Madrid seismic zone (*6*). Conditional probabilities based on the Poisson distribution are functions of the length of the exposure window, Δt, and the average return time, T, and not the time since the last earthquake. These Poisson-based estimates are thus time-independent (*7*). For individual fault zones, the basic Poisson assumptions of stationary behavior and temporal independence between events are inconsistent with recent concepts of strain accumulation and release (*8*). For cases where the return time between earthquakes is significantly longer than the time elapsed since the last event or the length of the historic catalog, however, the Poisson model provides a sufficient approximation of the hazard level (*9*) and has been widely used for regional seismic hazard assessments (*10*). On a regional or continental scale, however, choice of the Poisson distribution is motivated by the observation that combining the interoccurrence times of earthquakes from a number of independent faults into a single catalog does approximate a Poisson process (*6, 11–13*). Also, both the frequency-magnitude and Poisson relations are negative exponentials and exhibit a region of equivalence when the absolute value of the exponent is large. For the central and eastern United States this equivalence occurs for earthquakes with $m_b \geq 5$ (*11*). It is this correspondence that allows the use of Poisson statistics to describe regional earthquake occurrence, even though departures from stationary behavior have been documented (*14*). Specific examples illustrating the differences and similarities between time-dependent and Poisson-based earthquake forecasts will be discussed following the presentation of the data.

We compared frequency-magnitude data from two separate data bases to estimate the rate of seismic activity in the central and eastern United States. The first is based on data collected by regional seismograph networks in the New England, Southeast, and Central or New Madrid regions of the United States (*6, 11, 12*) (Fig. 1). The second is based on catalogs of eastern North American earthquakes of $m_b \geq 5.0$ from 1638 to 1982 (*15*). A further, independent constraint comes from areas where geologically derived recurrence time information is available.

The annual cumulative frequency-magnitude relations from the three regional seismograph networks are:

$$\text{New England } \log N_c = 3.49 - 0.93M \quad (1)$$

$$\text{New Madrid } \log N_c = 3.43 - 0.88m_b \quad (2)$$

$$\text{Southeast } \log N_c = 3.13 - 0.84m_b(\text{Lg}) \quad (3)$$

We have assumed that $m_b = m_b(\text{Lg}) = M$ (*6, 11, 12*). The data used to determine these regional frequency-magnitude relations span various time intervals and magnitudes. The New England equation is based on instrumental data (m_b 2 to 4.5) collected from 1975 to 1986 (*12*). The New Madrid rate is based on a combination of historic data from 1816 to 1974 (m_b 3.6 to 6.2) and instrumental data from 1975 to 1983 (m_b 1.7 to 5.0) (*6*). Southeastern rates are based on historic data from 1772 to 1977 (m_b 3.0 to 6.7) and instrumental data from 1978 to 1986 (m_b 2.0 to 5.2) (*11*). All of the above instrumental catalogs are assumed to be complete at m_b 2.0 level (that is, all earthquakes above that magnitude level have been recorded instrumentally). For historic earthquakes the degree of catalog completeness varies with location, time, and magnitude. In general, the rates of activity inferred from the historic record are in reasonable agreement with rates observed in the instrumental record (Fig. 2). For most areas, the rates of activity for earthquakes greater than m_b 6 have been extrapolated from the rates of occurrence of smaller events.

Using Eqs. 1 to 3, we calculated conditional probabilities for earthquakes with magnitudes greater than m_b 6.0 during exposure windows of 10, 30, 50, and 100 years (Table 1). For this and all subsequent calculations of probability, we have fit the

Table 1. Conditional probability estimates for earthquakes in the central and eastern United States. Probabilities are given for an earthquake of equal or greater magnitude during the next 10, 30, 50, and 100 years. New England, Southeast U.S., and New Madrid refer to the regional seismograph network data. Combined row gives the probabilities of at least one event within the above three regions. CEUS and ENA are the central and eastern United States and eastern North America (south of 55°N). See Fig. 1 for network and catalog locations.

Region	1990 to 2000		1990 to 2020		1990 to 2040		1990 to 2090	
	$m_b \geq 6.0$ $\mathbf{M} \geq 6.1$	$m_b \geq 7.0$ $\mathbf{M} \geq 7.4$	$m_b \geq 6.0$ $\mathbf{M} \geq 6.1$	$m_b \geq 7.0$ $\mathbf{M} \geq 7.4$	$m_b \geq 6.0$ $\mathbf{M} \geq 6.1$	$m_b \geq 7.0$ $\mathbf{M} \geq 7.4$	$m_b \geq 6.0$ $\mathbf{M} \geq 6.1$	$m_b \geq 7.0$ $\mathbf{M} \geq 7.4$
New England	0.08	0.01	0.21	0.03	0.33	0.04	0.55	0.08
Southeast	0.11	0.02	0.30	0.04	0.45	0.07	0.70	0.14
New Madrid	0.13	0.02	0.34	0.05	0.50	0.08	0.75	0.15
Combined	0.29	0.04	0.64	0.11	0.81	0.18	0.97	0.33
CEUS catalog	0.16 to 0.22	0.01 to 0.02	0.41 to 0.53	0.03 to 0.07	0.58 to 0.72	0.05 to 0.11	0.83 to 0.92	0.10 to 0.20
ENA catalog	0.31 to 0.39	0.03 to 0.05	0.67 to 0.78	0.10 to 0.14	0.85 to 0.92	0.16 to 0.23	0.98 to 0.99	0.29 to 0.40

interval form of the b value equation, $\log N_i = a - bM_i$, rather than the cumulative form, N_C (16). The probability that at least one event will occur within the entire area covered by all three networks (in Table 1, Combined) is given by

$$P_j = 1 - (1 - P_1)(1 - P_2)(1 - P_3) \quad (4)$$

where P_i, $i = 1, 2, 3$ are the probabilities of one or more events for the individual regions, and P_j is the combined probability.

These regional results can be compared with those based on the general catalog of earthquakes of $m_b \geq 5.0$ from 1727 to 1980 for the central and eastern United States and a larger catalog of events with $m_b \geq 5.0$ that includes both the United States and Canada from 1638 to 1982 (15). Our primary intent is to estimate earthquake probabilities for the central and eastern United States; hence, we have restricted our analysis to events that occurred south of 55°N (that is, those seismic events capable of affecting the central and eastern United States). Although these catalogs cover larger geographic regions than the combined regional networks (see Fig. 1), the degree of completeness is known to vary as a function of magnitude, time, and location (17). We have thus restricted our analysis to those seismic events occurring during time periods of relatively complete coverage (that is, earthquakes with $m_b \geq 7$ since 1727, of $m_b \geq 6$ since 1770, and of $m_b \geq 5$ since 1800). Additional variability in earthquake catalogs stems from the fact that the same historic earthquake can have various assigned magnitudes as a result of the different methods used to estimate magnitude (18).

Instrumental estimates of body-wave (m_b) magnitude saturate at approximately m_b 6.5; the magnitude estimates in our catalog for events larger than 6.5 are from the preinstrumental period and are based on special intensity studies rather than being directly derived from instrumental data (1). Hence, these magnitudes are not saturated (in an instrumental sense) and give reliable estimates of earthquake size that can be analyzed with the other, smaller events in our catalog (19). For central and eastern United States earthquakes, m_b 6.5 is equivalent to $\mathbf{M}$ (seismic moment magnitude) 6.8; m_b 7.0 is equivalent to $\mathbf{M}$ 7.4 (2, 20).

For consistency, the frequency-magnitude analysis of the central and eastern U.S. catalog was similar to that used for the network data. The interval form of the frequency-magnitude relationship was fit with the use of maximum-likelihood techniques (21) after foreshocks and aftershocks were removed (22). The maximum magnitude assumed for our analysis was varied between 7.5 and 8.0. For comparison with the network results, the interval form of the frequency-magnitude relation ($\log N_i = a - bM_i$, where N_i is the number of earthquakes with magnitude between $M_i \pm \Delta M$, where ΔM is 0.25) was converted to the cumulative form of the frequency-magnitude equation, $\log N_C = A - BM$ with the relation $B = b$ and $A = a - \log_{10}(10^{B\Delta M} - 10^{-B\Delta M})$ (23). For the entire U.S. catalog ($m_b \geq 5$, 40 to 46 events), the cumulative frequency-magnitude equation (on an annual basis) varies between $\log N_C = (4.29 \pm 0.85) - (0.98 \pm 0.15)M$ and $\log N_C = (4.80 \pm 0.97) - (1.09 \pm 0.18)M$ (Fig. 2). Analysis of the combined U.S. and Canadian catalog was similar to the above and covers the same time intervals. The frequency-magnitude equation (annual basis) for all data ($m_b \geq 5$, 56 to 79 events) varies between $\log N_C = (3.86 \pm 0.71) - (0.88 \pm 0.13)M$ and $\log N_C = (4.28 \pm 0.59) - (0.93 \pm 0.11)M$ (Fig. 2).

The rates of earthquake occurrence based on the frequency-magnitude analysis of seismograph network and historic earthquake catalogs indicate that the Poisson probability for an event of $m_b \geq 6.0$ ($\mathbf{M} \geq 6.1$) over the next 30 years is at a moderate to high (0.4 to 0.6) level for the central and eastern United States. When a larger area is considered (that is, eastern North America) the probability is even higher (0.7 to 0.8) (Table 1). These results can also be derived from an alternate, and simpler, analysis by noting that in both the central and eastern United States and eastern North America there have been between four and nine independent earthquakes greater than m_b 6 during the last 215 years. The average return times are between 54 and 24 years, respectively, and the corresponding Poisson probabilities for a 30-year exposure window range from 0.4 to 0.7. During this same 30-year interval, the probability for an earthquake with a magnitude greater than m_b 7.0

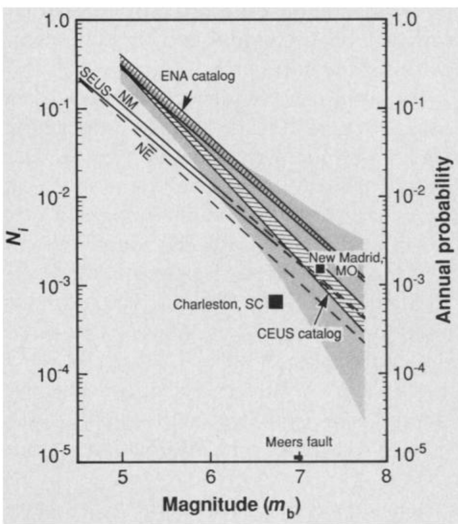

Fig. 2. Comparison of frequency-magnitude relations and annual probabilities based on regional seismograph data (NE, New England; SEUS, Southeast United States; and NM, New Madrid; solid lines are constrained by data, dashed lines are extrapolations of frequency-magnitude data); earthquake catalog data (CEUS, central and eastern United States and ENA, eastern North America); and paleoseismic data (solid squares show recurrence times from sites in Charleston, South Carolina, New Madrid, Missouri, and the Meers fault, Oklahoma; height of squares is proportional to the uncertainty in recurrence times). Hachured areas connecting the pairs of curves for the CEUS and ENA catalogs reflect the variation in the frequency-magnitude relationship due to differences in magnitude assignments of historic events and of the maximum magnitudes assumed in the analysis. Shaded area represents the 95% confidence interval for the CEUS catalog data.

(M 7.4) is approximately 0.1 (Table 1).

Paleoseismic investigations can provide earthquake recurrence information that is independent of the above catalog-based frequency-magnitude analysis. Such studies, however, have been made at only a few sites in the central and eastern United States. For the New Madrid seismic zone, ^{14}C dating of shells in disturbed sediments near the Reelfoot Lake area in northwestern Tennessee suggests that there were at least two episodes of faulting during the 2000 years before the 1811 to 1812 earthquakes (*24*). If these two events are of similar size to the 1811 to 1812 earthquakes, the average return interval for an m_b 7 (**M** 7.4) earthquake is about 600 to 700 years. These results are in reasonable agreement with the frequency-magnitude estimates for the New Madrid region [550 years (*6*)]. In contrast, excavations of 1811 to 1812 sandblow sites in northeastern Arkansas reveal no evidence of significant ground disturbance for at least 1300 years before 1811 (*25*). Hence, the 600- to 700-year intervals may represent a lower bound estimate on return times and give an upper bound on probabilities for $m_b \geq 7$ earthquakes in this region. Conditional probabilities are 0.02 for a 10-year interval and 0.05 for a 30-year interval on the basis of the Poisson model. Time-dependent models also indicate similar probabilities ($\leq$0.01 by A.D. 2000 and 0.03 to 0.04 by A.D. 2035) (*6*).

For the Charleston, South Carolina, area, ^{14}C dating of sandblows and paleoliquefaction events indicate that the average return time for earthquakes capable of producing these features ($m_b \geq 5.5$) is about 1600 years (*26*). The distribution of sandblows and paleoliquefaction events along the coastal plain of North and South Carolina indicate that the Charleston, South Carolina, area has been a persistent site of strong Holocene earthquakes (*27*). The above recurrence interval agrees with those intervals based on seismicity in the immediate Charleston area (*11*) and the corresponding Poisson probability is 0.01 for 10-year and 0.02 for 30-year exposure windows, and is less than 0.01 for the next 30 years on the basis of time-dependent models. The above probability estimates for the New Madrid and Charleston, South Carolina, earthquakes are in general agreement, regardless of whether time-dependent or time-independent models are used to portray the hazard. This agreement primarily reflects the effects of relatively long recurrence intervals and recent faulting on the calculations.

While Charleston, South Carolina, has been the site of repeated strong Holocene earthquakes, inspection of Fig. 2 indicates that the geologic rate of occurrence of Charleston events alone does not satisfy the extrapolated frequency-magnitude curve for m_b 6.5 to 7 events in the central and eastern United States. In contrast, the current paleoseismic data for $m_b \geq 7$ earthquakes in the New Madrid region are in agreement with the regional frequency-magnitude extrapolation for the central and eastern United States. At this point, the application of local paleoseismic results to regional earthquake frequency data is equivocal and we cannot use the agreement or disagreement of these data to support the existence (or nonexistence) of additional source zones to balance the b value budget.

Another recently described example of an active, prehistoric earthquake source zone is the Meers fault in southwestern Oklahoma. The ^{14}C dating of soil samples from trenches and ponded alluvium sites indicates that the last surface-faulting event ($\sim m_b$ 6.7) occurred 1,200 to 1,300 years ago, and that recurrence intervals may be on the order of 100,000 years or more (*28*). In this case, both the Poisson and time-dependent models indicate probabilities of less than 0.01 for the time interval considered in this study.

More precise definition of recurrence times and corresponding probabilities for specific fault zones or segments is difficult beyond our regional analysis. Along the New Madrid seismic zone, Johnston and Nava (*6*) applied a suite of time-dependent models (Gauss, Weibull, lognormal) to estimate a 0.40 to 0.63 probability of an event with $m_b \geq 6$ by the year 2000, and a 0.86 to 0.97 chance by the year 2035. These estimates are a factor of 2 to 3 larger than the Poisson model estimates (see Table 1, New Madrid). The two areas that produced $m_b \geq 6$ earthquakes in the 19th century (Marked Tree, Arkansas, 1843, and Charleston, Missouri, 1895) occupy different geographic locations with respect to the New Madrid seismic zone, and have no history of earlier events in the same magnitude range. The current uncertainties in understanding of the tectonic regime of the New Madrid seismic zone are clearly illustrated by the above difference in probability estimates for $m_b \geq 6$ earthquakes and highlight the need for further research (*29*).

Central and eastern U.S. earthquakes are capable of producing damage over larger areas than their counterparts in California, and the higher occurrence rates in the West are balanced by this lower rate of attenuation of damaging vibrations in the East. For a hypothetical m_b 7.2 earthquake, the areas of peak horizontal ground acceleration in the East are estimated to be up to ten times as large as those in the West, and areas of peak ground velocity are five to ten times as large (*30*). If we compare the probability for a single earthquake with equivalent damage areas [for example, modified Mercalli (MM) intensity greater than VIII] east and west of the Rockies, **M** 6 in the central and eastern United States, and **M** $\geq$ 7 in either southern California or the San Francisco Bay area (*31*), the probabilities are about equal for a 30-year time interval (*32*). On a larger scale, the Poisson probability for a **M** $\geq$ 6 earthquake in the central and eastern United States is two-thirds that of a **M** $\geq$ 7 in all of California. Hence, while California is more hazardous than the central and eastern United States on the basis of frequency of occurrence, both regions have comparable risk [that is, similar probabilities for producing earthquakes with comparable damage areas (*33*)]. Because the last earthquake in the central or eastern United States greater than m_b 6 occurred in 1895, the response of buildings and critical structures to a damaging earthquake has not been tested in more than 90 years.

REFERENCES AND NOTES

1. R. Street and O. W. Nuttli, *U.S. Geol. Survey Open-File Rep. 84-770* (1984), p. 33.
2. O. W. Nuttli, G. A. Bollinger, R. B. Herrmann, *U.S. Geol. Surv. Circ. 985* (1986).
3. 1895 Charleston, Missouri, m_b 6.2.
4. S. P. Nishenko and R. Buland, *Bull. Seismol. Soc. Am.* 77, 1382 (1987).
5. In a recent example, the Working Group on California Earthquake Probabilities [*U.S. Geol. Surv. Open-File Rep. 88-398* (1988)] had identified the site of the 17 October 1989 Loma Prieta, California, earthquake as having a probability of 0.3 for the occurrence of a M_S 6.5 earthquake during the time interval 1988 to 2018. See also U.S. Geological Survey, *Science* **247**, 286 (1990).
6. A. C. Johnston and S. J. Nava, *J. Geophys. Res.* **90**, 6737 (1985). m_b is the 1-s body-wave magnitude scale.
7. If the number of events are Poisson-distributed, the time between events follows an exponential distribution, and the probability distribution function, $f(t)$, for the occurrence of a single earthquake is $f(t) = (1/T)\exp(-t/T)$ where t is the time elapsed since the last event, and T is the average interevent time, or $1/N$, where N is the number of earthquakes per unit time. The cumulative distribution, $F(t)$ is $F(t) = 1 - \exp(-t/T)$ and the conditional probability, $P_C[t < x < t + \Delta t | x > t]$, is equal to $[F(t + \Delta t) - F(t)]/[1 - F(t)]$ or $P_C = 1 - \exp(-\Delta t/T)$. The conditional probability, P_C, is independent of t, the time since the last event. This relation illustrates the memoryless property of the Poisson distribution and how conditional probabilities are only a function of Δt and T, the length of the time window and the average interevent time.
8. K. Shimazaki and T. Nakata, *Geophys. Res. Lett.* **86**, 279 (1980).
9. C. A. Cornell and S. R. Winterstein, *Bull. Seismol. Soc. Am.* **78**, 1522 (1988).
10. S. T. Algermissen, D. M. Perkins, P. C. Thenhaus, S. L. Hanson, B. L. Bender, *U.S. Geol. Surv. Open-File Rep. 82-1033* (1982); P. C. Thenhaus *et al.*, *Earthquake Spectra* 1, 285 (1985).
11. G. A. Bollinger, F. C. Davison, Jr., M. S. Sibol, J. B. Birch, *J. Geophys. Res.* **94**, 2857 (1989). m_b(Lg) is the 0.8- to 1.3-s Lg magnitude scale.
12. J. E. Ebel, *Seismic Hazards, Ground Motions, Soil-Liquifaction, and Engineering Practice in Eastern North America*, K. H. Jacob, Ed. (*NCEER-87-0025*, National Center for Earthquake Engineering Research, Buffalo, NY, 1987), pp. 178–188. *M* is a higher frequency (that is, less than 0.8 s) Lg magnitude scale.

13. D. M. Perkins, *U.S. Geol. Surv. Open-File Rep. 87-673* (1987), p. 428.
14. A. C. Johnston, *Earthquakes and Earthquake Engineering in the Eastern United States*, J. E. Beavers, Ed. (Ann Arbor Science, Ann Arbor, MI, 1981), pp. 161–181.
15. Catalogs used in this study are from M. S. Sibol and G. A. Bollinger, *Listing of Central and Eastern North American Earthquakes* (Virginia Polytechnical Inst. Seismological Observatory, Blacksburg, VA, 1989); *Electric Power Research Institute Catalog of Central and Eastern North American Earthquakes to 1985* (Electric Power Research Institute, Palo Alto, CA, 1985).
16. In the interval form of the frequency-magnitude equation, it is assumed that there is a finite or maximum magnitude cutoff for the analysis (we used m_b 7.5 to 8.0). As the cumulative form of the frequency-magnitude equation is unbounded, direct conversion of the interval results to the cumulative form for computing return times violates the maximum magnitude assumption.
17. Statistical tests of catalog completeness in the central and eastern United States indicate that intervals of complete reporting vary from 110 to 280 years for m_b 5, 125 to 300 years for m_b 6, and 160 to 300 years for m_b 7; O. W. Nuttli and R. B. Herrmann, *Misc. Pap. S-73-1* (U.S. Army Waterways Experiment Station, Vicksburg, MS, 1978); D. Veneziano and J. Van Dyck, *Seismic Hazard Methodology for Nuclear Facilities in the Eastern United States*, vol. 2 (Electric Power Research Institute, Palo Alto, CA, 1985).
18. The variability in magnitudes assigned to earthquakes in the catalogs consulted is greatest in the $m_b < 6$ range (which includes 85 to 93% of the total data set) whereas the number of events greater than m_b 6 is generally constant from catalog to catalog. By comparing different earthquake catalogs we have attempted to bracket at least some of the above variations in the smaller magnitude band.
19. Although a frequency–seismic moment regression would have dealt with the problem of instrumental magnitude saturation, we have used body-wave magnitudes (m_b) as the independent variable for our analysis to provide continuity with earlier studies. Table 1 includes seismic moment magnitude (**M**) equivalents of m_b values.
20. O. W. Nuttli, *Bull. Seismol. Soc. Am.* **73**, 519 (1983); T. C. Hanks and H. Kanamori, *J. Geophys. Res.* **84**, 2348 (1979).
21. B. Bender, *Bull. Seismol. Soc. Am.* **73**, 831 (1983); D. H. Weichert, *ibid.* **70**, 1337 (1980).
22. These consisted of five events: a m_b 5 foreshock to the 1897 m_b 5.8 Giles County, Virginia, earthquake, a m_b 5.5 aftershock of the 1940 m_b 5.5 New Hampshire event, and three of the 1811 to 1812 New Madrid earthquakes. The New Madrid earthquakes are determined to be dependent events as the probability of two independent earthquakes of $m_b \geq 7$ occurring within 2 months of the first event is $\sim 5 \times 10^{-8}$.
23. R. B. Herrmann, *Earthquake Notes* **48**, 47 (1977).
24. D. P. Russ, *Geol. Soc. Am. Bull.* **90**, 1013 (1979).
25. R. T. Saucier, *Geology* **17**, 103 (1989).
26. S. F. Obermeier, R. E. Weems, R. B. Jacobson, *U.S. Geol. Surv. Open-File Rep.* 87-504 (1987); P. Talwani and J. Cox, *Science* **229**, 379 (1985).
27. S. F. Obermeier, R. E. Weems, R. B. Jacobson, G. S. Gohn, "Earthquake hazards and the design of constructed facilities in the eastern United States," *Ann. N.Y. Acad. Sci.* **558**, 183 (1989).
28. A. J. Crone and K. V. Luza, *Bull. Geol. Soc. Am.* **102**, 1 (1990).
29. R. M. Hamilton and A. C. Johnston, *U.S. Geol. Surv. Circ. 1066* (1990).
30. O. W. Nuttli, in *Earthquakes and Earthquake Engineering in the Eastern United States*, J. E. Beavers Ed. (Ann Arbor Science, Ann Arbor, MI, 1981), pp. 25–51.
31. Damage area–magnitude comparison based on areas of MM VIII damage associated with the 1886 Charleston, South Carolina, and 1895 Charleston, Missouri, earthquakes. The minimum estimated MM VIII area for the m_b 6.7 (**M** 7.0) 1886 event is $\sim 3.8 \times 10^4$ km^2 [G. A. Bollinger, *U.S. Geol. Surv. Prof. Pap. 1028* (1977)]; the MM VIII area for the m_b 6.2 (**M** 6.4) 1895 event is $\sim 2 \times 10^4$ km^2 [I. N. Gupta and O. W. Nuttli, *Bull. Seismol. Soc. Am.* **66**, 743 (1976)]. On the basis of magnitude-intensity area relationships for California [T. R. Toppozada, *ibid.* **65**, 1223 (1975); J. F. Evernden, W. M. Kohler, G. D. Glow, *U.S. Geological Surv. Prof. Pap. 1223* (1981)], the equivalent magnitudes for California earthquakes with comparable damage areas are **M** 8.0 to 8.3.
32. The Poisson probability for a $M \geq 7$ earthquake in California is 0.85 for 30 years (11 events in 178 years). For southern California and the San Francisco Bay area, the 30-year Poisson probabilities are 0.57 and 0.54, respectively (five events in 178 years, and four events in 154 years, respectively). The 30-year time-dependent probabilities for $\mathbf{M} \geq 7$ earthquakes along the San Andreas fault system are 0.6 (*5*) in southern California and 0.67 for the San Francisco Bay area [Working Group on California Earthquake Probabilities, *U.S. Geol. Surv. Circular 1053* (1990)]. Probabilities for $\mathbf{M} \geq 6$ earthquakes in the eastern and central United States are shown in Table 1.
33. L. Seeber and J. G. Armbruster, in *The Geology of North America*, R. E. Sheridan and J. A. Grow, Eds. (Geological Society of America, Boulder, CO, 1988), vol. I-2, pp. 565–582.

34. We thank B. Bender, J. Dewey, K. Jacob, D. Perkins, M. Sibol, and R. Wheeler for stimulating discussion and advice.

30 May 1990; accepted 29 July 1990

No Excess of Homozygosity at Loci Used for DNA Fingerprinting

B. DEVLIN, NEIL RISCH,* KATHRYN ROEDER

Variable number of tandem repeat (VNTR) loci are extremely valuable for the forensic technique known as DNA fingerprinting because of their hypervariability. Nevertheless, the use of these loci in forensics has been controversial. One criticism of DNA fingerprinting is that the VNTR loci used for the "fingerprints" violate the assumption of Hardy-Weinberg equilibrium (H-W), making it difficult to calculate the probability of observing a genotype in the population. If one can assume H-W, the probability of observing the pair of alleles constituting an individual's genotype can be calculated by taking the product of the alleles' frequencies in the population and multiplying by two if the alleles are different. The evidence cited against assuming H-W is homozygote excess, which is presumed to be caused by an undetected mixture of two or more populations with limited interpopulational mating and distinct allele frequencies. For most VNTR loci, measurement error makes it impossible to test these claims by standard methods. The Lifecodes database of three VNTR loci used for forensics was used to show that the claimed excess of homozygotes is not necessarily real because many heterozygotes with similar allele sizes are misclassified as homozygotes. A simple test of H-W that takes such misclassifications into account was developed to test for an overall excess or dearth of heterozygotes in the sample (the complement of homozygote dearth or excess). The application of this test to the Lifecodes database revealed that there was no consistent evidence of violation of H-W for the Caucasian, black, or Hispanic populations.

THE DISCOVERY OF HYPERVARIABLE VNTR loci in human DNA in the early 1980s (*1*) was seen as a boon to a number of areas of scientific interest, in particular forensic science (*2, 3*). The loci are called hypervariable because, in any population, there are a very large number of alleles present at each locus (*3, 4*). Each allele of a VNTR locus is composed of a distinct sequence of base pairs, which one can detect indirectly by excising that region of the DNA with a restriction enzyme and estimating the length of the fragment by gel electrophoresis. Much of the allelic variation is generated by variation in the number of short, repeated sequences of base pairs linked in tandem in the core region of the locus (hence the acronym VNTR), which leads to fragment length variation on electrophoresis. It is the presence of a large number of alleles at these loci that renders the loci valuable as "fingerprints" (*5*), because it is quite likely that different individuals will have distinct genotypes at these loci. Nevertheless, the use of VNTR loci for forensics has led to controversy (*6, 7*). We focus on one of these controversies: whether H-W can be assumed for several VNTR loci used in forensics. Some researchers have asserted that H-W cannot be assumed because there is an excess of homozygotes at these loci (*6–8*).

H-W is an attribute of large, randomly mating populations. A population is in

B. Devlin, Division of Biostatistics, Department of Epidemiology and Public Health, Yale University, Post Office Box 3333, 60 College Street, New Haven, CT 06510.
N. Risch, Division of Biostatistics, Department of Epidemiology and Public Health, and Department of Human Genetics, Yale University, Post Office Box 3333, 60 College Street, New Haven, CT 06510.
K. Roeder, Department of Statistics, Yale University, Post Office Box 2179 Yale Station, New Haven, CT 06520.

*To whom correspondence should be addressed.