

Characteristic and Uncharacteristic Earthquakes as Possible Artifacts: Applications to the New Madrid and Wabash Seismic Zones

Seth Stein

Northwestern University

Andrew Newman

Los Alamos National Laboratory

INTRODUCTION

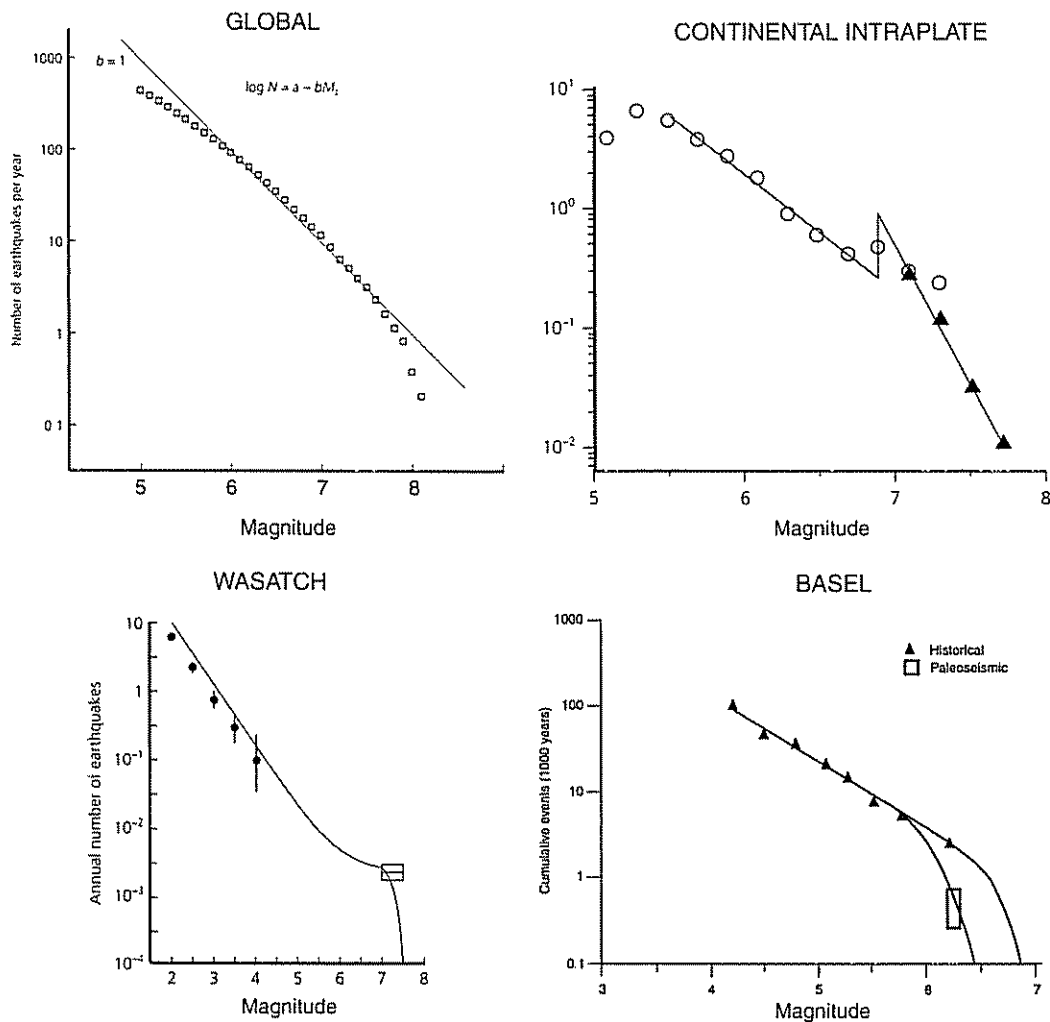
Understanding the relation between the recurrence of large and small earthquakes is a formidable challenge, because multiple data sets with poorly known uncertainties are combined to explore phenomena that may vary from place to place and whose underlying physics are unknown. This problem often arises in attempts to assess the recurrence intervals of large earthquakes, which are important for earthquake physics, tectonics, and seismic hazard analysis. For example, we often seek to infer the sizes and rates of future large earthquakes from an earthquake record containing seismological, geological, and cultural records. This process, however, is far from straightforward and is prone to a variety of biases that can make the apparent rates and sizes of the largest earthquakes appear different from their true long-term values.

In general, earthquake recurrences approximately follow a log-linear, b -value, or Gutenberg-Richter relation, $\log N = a - bM$, with $b > 1$, such that the logarithm of the annual number (N) of earthquakes above a given magnitude (M) decreases linearly with magnitude (Gutenberg and Richter, 1944). The rates of large earthquakes, however, whose size is presumably limited, show interesting deviations, as reviewed by Main (1996). Large data sets derived from seismological data, such as the global or intracontinental data shown in Figure 1 (upper panels), show that the largest earthquakes occur less frequently than expected from the rates of smaller ones. This is generally attributed to the effects of finite fault dimensions (e.g., Okal and Romanowicz, 1994), although other interpretations have been made (Main, 2000). In contrast, studies of specific areas, which commonly address the short history of seismological observations by combining seismological data for smaller earthquakes with paleoseismic data or geologic inferences for larger earthquakes, sometimes infer that large earthquakes occur more commonly than expected from the log-linear frequency-magnitude relation observed for smaller earthquakes (Figure 1, lower left).

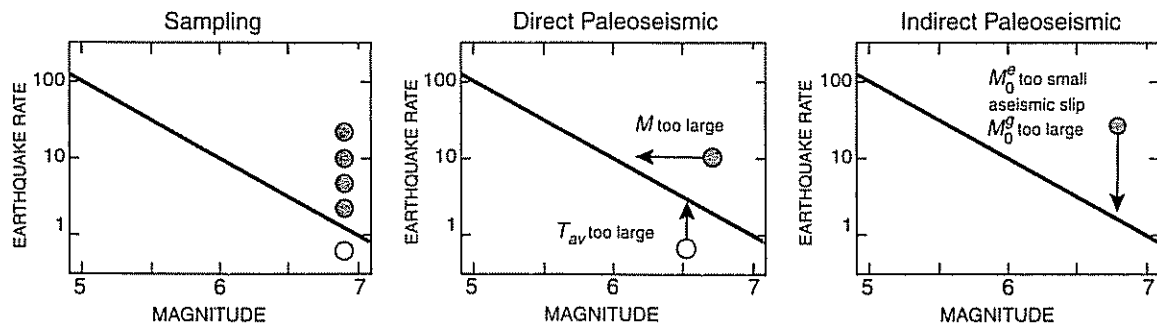
Whether this effect is real or apparent in any given region is an interesting question (Kagan, 1993). In one view, the larg-

est earthquakes, termed characteristic earthquakes in this interpretation, are in fact more frequent than expected from the small earthquakes (Schwartz and Coppersmith, 1984). As shown in Figure 1 (lower left), paleoseismic studies for the Wasatch Fault infer that earthquakes with $M \geq 7$ recur every 400–600 years, whereas a log-linear frequency-magnitude relation derived from the short instrumental history predicts a recurrence interval greater than 1,000 years. (In discussing such combinations, we use the term “history” for the period of instrumental data, and “record” for the total period of instrumental, historical (cultural), and paleoseismic data. These are the same for synthetic earthquake histories [or records].) Such behavior can be formulated as an alternative to the Gutenberg-Richter relation (Youngs and Coppersmith, 1985; Main, 2000). According to this interpretation, the characteristic earthquake magnitudes are controlled by the lengths of the faults on which they occur (Wesnowsky *et al.*, 1983; Stirling *et al.*, 1996). This assumption is often made in seismic hazard studies, typically leading to a predicted hazard higher than predicted assuming a Gutenberg-Richter distribution.

Alternatively, the difference in the inferred recurrence times of the largest earthquakes might be an artifact. Apparent differences might arise in several possible ways, or a combination thereof. One situation (Figure 2, left) might arise as a result of a time-sampling bias in areas where the length of the earthquake record under consideration is comparable to the mean recurrence time of large earthquakes predicted by a Gutenberg-Richter distribution. Apparent characteristic earthquakes can occur if seismicity follows a log-linear frequency-magnitude relation with earthquake recurrence intervals distributed about the mean for that magnitude range, because sampling bias makes those with shorter intervals more likely to be observed than those with longer ones (fractions of earthquakes cannot be observed). The short record can also cause large earthquakes (which we term “uncharacteristic”) to appear less frequently than expected from the small earthquakes. Such an effect has been inferred for the Upper Rhine graben by comparison of the recurrence times inferred from historic and paleoseismic data (Figure 1, lower



▲ **Figure 1.** Frequency-magnitude plots for various sets of earthquake data *Upper left* Global data for 1968–1997 (Stein and Wysession, 2003) *Upper right* Continental intraplate data for 1978–1994 (open circles) and 1900–1994 (solid triangles) (Triep and Sykes, 1997) The largest earthquakes are less frequent than expected from the smaller ones *Lower left*. Seismological (dots) and paleoseismic (box) data for the Wasatch Fault (Youngs and Coppersmith, 1985), showing large earthquakes to be more common than expected from the small ones *Lower right* Historical and paleoseismic data for the greater Basel (Switzerland) area (Meghraoui *et al.*, 2001) The paleoseismic data imply that large earthquakes occur at a lower rate than predicted from smaller ones



▲ **Figure 2.** Possible effects causing apparent deviations from a log-linear frequency-magnitude relation *Left* Sampling bias due to a short earthquake record makes the largest earthquakes seem more common (characteristic, dark circles) or less common (uncharacteristic, light circles) than their long-term averages *Center*. Apparent characteristic earthquakes occur if paleoseismic data yield overestimates of magnitudes Apparent uncharacteristic earthquakes occur if paleoseismic data yield overestimates of recurrence intervals *Right* If recurrence intervals are estimated indirectly by dividing the expected moment in large earthquakes by that assumed to occur annually over a long time, apparent characteristic earthquakes occur if the size of the largest earthquakes is underestimated, the long-term slip rate and hence moment release rate is overestimated, or some release occurs aseismically.

right). Similarly, large earthquakes can be absent from an earthquake record composed of small earthquakes, simply because none occurred in the short time sampled. In these situations the magnitudes of the largest observed earthquakes are correctly known, but the deviation of their recurrence interval from that predicted by a Gutenberg-Richter distribution is due to sampling bias.

A second possibility (Figure 2, center) is suggested by the fact that characteristic earthquakes are often inferred because the recurrence times inferred from paleoseismic data are discordant with those expected from instrumental or historical data. This situation suggests that the apparent difference may result from uncertainties in estimating the magnitude and recurrence intervals of the largest earthquakes directly from paleoseismic data. If the uncertainties are greater than assumed, then the largest earthquakes might not in fact be more frequent than predicted by the b -value line, especially given the uncertainties in estimating the b -value line from the seismological record of seismicity. Apparent characteristic earthquakes could occur due to miscalibration of paleoseismic data, such that the inferred paleoearthquake magnitudes were overestimated. Conversely, apparent uncharacteristic earthquakes could occur if some paleoearthquakes in a series were not identified, causing the mean recurrence time T_{av} to be overestimated.

A third possibility arises when the mean recurrence times of the largest earthquakes are inferred indirectly by dividing M_0^c , the expected moment release in characteristic earthquakes, by M_0^s , the expected annual moment release rate on a fault averaged over a long time interval (Wesnousky, 1994). This approach is used primarily in plate-boundary zones, when long-term slip rates on faults can be inferred from geological data or plate-motion arguments. Apparent characteristic earthquakes might result if the sizes of the largest events are underestimated, as proposed by Kagan (1996) but discounted by Wesnousky (1996). A similar situation might arise if the long-term fault slip rate and hence moment-release rate were overestimated or some occurred aseismically, implying T_{av} too short (Figure 2, right).

Our interest in these issues derives from the New Madrid seismic zone, where the length of the earthquake record is comparable to the expected recurrence times of the largest earthquakes. Hence we focus on the first possibility, effects of a short earthquake record. However, because multiple effects may be involved, we also explore the second possibility by considering some potential biases due to uncertainties in directly estimating magnitudes and recurrence times from paleoseismic data. We do not address the third possibility, which does not arise in the intraplate New Madrid situation.

We explore these issues via simple simulations to see under what conditions characteristic, uncharacteristic, or missing large earthquakes might result from short earthquake records. These simulations extend a previous analysis of deviations of the largest earthquakes in an area from a log-linear frequency-magnitude relation. Howell (1985) illustrated this effect by dividing an earthquake population that followed a

Gutenberg-Richter distribution into subsets. Because the numbers of large earthquakes in individual subsets were small, frequency-magnitude relations for the subsets had considerable scatter, with the largest earthquakes appearing in some cases more and in other cases less frequent than for the total population. Thus the frequency-magnitude slope (b value) was reasonably well estimated from the smaller earthquakes but not the largest ones. Our simulations explore not only the scatter of recurrence estimates, but also the tendency of these estimates to be biased toward recurrence times longer or shorter than their mean value.

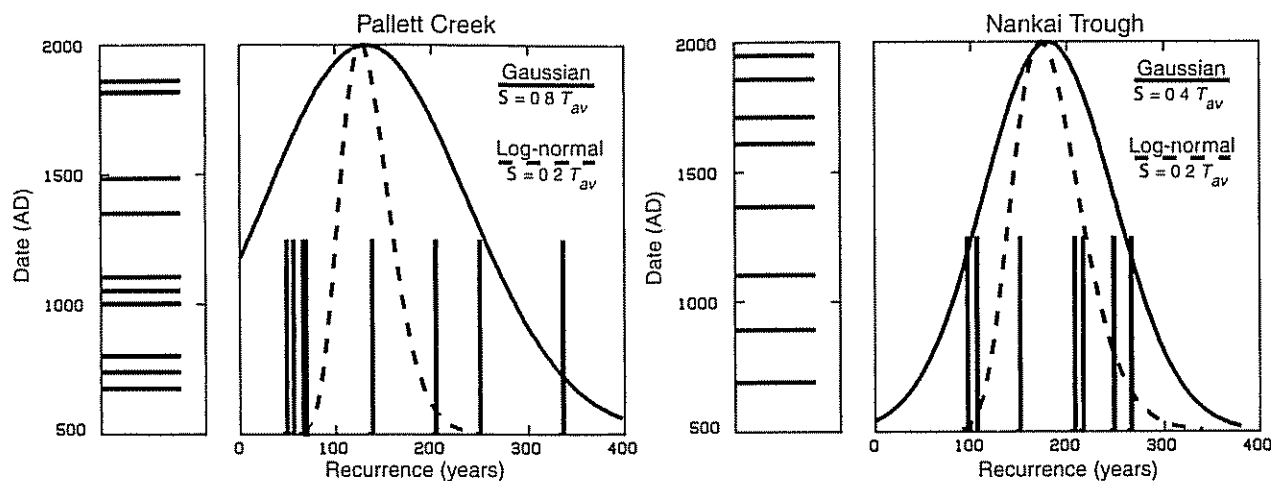
RECURRENCE VARIABILITY

The challenge in understanding earthquake recurrence is to interpret an observed earthquake record. Following a common approach in earthquake recurrence and probability studies, we view the observed record as a series of samples from a parent distribution of earthquake recurrence intervals. Ideally, the longer the record of observations, the more insight we would have into the parent distribution. Unfortunately, given the relatively short earthquake records available, we know surprisingly little about earthquake recurrence. In particular, as discussed later, we do not know whether variability in recurrence intervals, which we treat as random, in fact reflects underlying deterministic factors.

In this probabilistic view, how likely characteristic, uncharacteristic, and missing earthquakes are to occur as a result of short earthquake records depends on the frequency-magnitude relations that describe the mean recurrence intervals for large earthquakes and the distribution of recurrence times that describes the variability about the mean. Because we know little about either of these factors, we do not know whether such earthquakes reflect real or apparent deviations from a Gutenberg-Richter distribution.

Our numerical simulations used a Gutenberg-Richter distribution rather than other possible frequency-magnitude relations (Main, 1996). This choice was both for simplicity and because our goal was to see how apparent deviations from the log-linear frequency-magnitude relation, which characterizes the recurrence of small earthquakes, might result from short sampling. Thus we did not incorporate a possible decrease in earthquake frequency for the largest earthquakes (Figure 1).

To describe the variability in recurrence times, we used Gaussian distributions about the mean value predicted by a Gutenberg-Richter distribution, which were truncated to avoid negative recurrence times. A compilation of recurrence times for the largest earthquakes on different faults shows a range of the ratio of the standard deviation to the mean (σ/T_{av}) of approximately 0.1–0.4 (Nishenko and Buland, 1987). This study finds that the variability can be characterized better by assuming that the ratio of recurrence times to their mean is log-normally distributed. Most earthquake sequences used in that study are short, however, typically two to four recurrences. It appears that the variability in short sequences may not fully reflect the true variability.



▲ **Figure 3.** Observed variability in recurrence intervals for two long sequences of large earthquakes

Figure 3 illustrates this effect with two of the longest observed earthquake sequences. The upper panels show the paleoearthquake record at Pallett Creek, on the segment of the San Andreas that broke in the 1857 Fort Tejon earthquake (Sieh *et al.*, 1989). These dates yield nine recurrence intervals with $T_{av} = 132$ years and $\sigma = 105$ years, which implies $\sigma = 0.8 T_{av}$. This large variability results from the presence of several clusters of large earthquakes, which together with the observational uncertainties makes it difficult to characterize the sequence and estimate earthquake probabilities. Hence Sieh *et al.* (1989)'s estimate of the probability of a similar earthquake before 2019 ranged from 7–51%. As shown, overall variability is larger than would be characterized by a log-normal distribution with the parameters used by Nishenko and Buland (1987), who divided the Pallett Creek data into clusters and found that the less-variable individual clusters were reasonably well described by the log-normal distribution. A second example, shown in the lower panels, is the record of large subduction earthquakes in the A-B segment of the Nankai Trough (Ishibashi, 1981). These seven intervals have $T_{av} = 180$ years and $\sigma = 0.4 T_{av}$. Hence $\sigma = 0.4 T_{av}$ or greater seems like a reasonable characterization of the variability in recurrence times for long earthquake sequences.

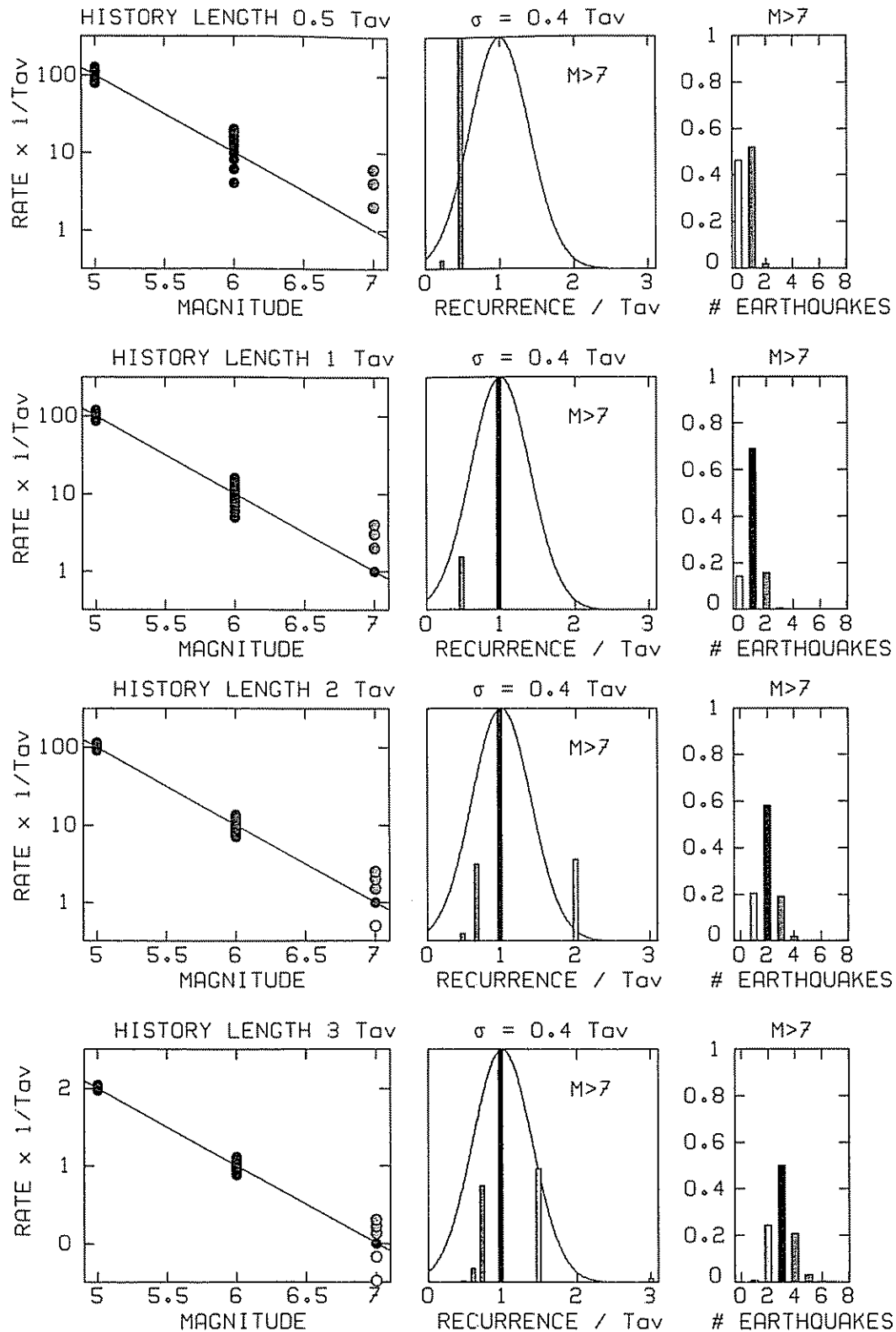
SIMULATIONS

We generated synthetic earthquake histories assuming that the seismicity followed a log-linear frequency-magnitude relation with $a = 4$, $b = 1$. Recurrence times of earthquakes with $M \geq 5$, 6, and 7 were assumed to be samples of a Gaussian (normal) parent distribution with a standard deviation of 0.4 times the mean recurrence for each of the three magnitudes. These sequences start 1,000 years into the simulations to ensure that “now” has no special significance. All times were then normalized by T_{av} , the mean recurrence time in the Gaussian distribution for earthquakes with $M \geq 7$, making the results independent of the b value.

Figure 4 shows the results for 10,000 synthetic earthquake sequences for each of four lengths: 0.5, 1, 2, and 3 times T_{av} . For each sequence, we find the mean recurrence for earthquakes with $M \geq 5$, 6, and 7 by dividing the sequence length by the number of earthquakes. Hence no recurrence time is computed if no earthquakes of a given size occur, and the nominal recurrence time equals the sequence length if only one occurs.

For each history length (row), the right panel shows the fraction of sequences in which a given number of $M \geq 7$ earthquakes occurred. We assume that the frequencies of different simulation outcomes reflect the probabilities that these would actually occur. The left panel shows the log-linear frequency-magnitude relation and dots marking the “observed” mean recurrence rates for $M \geq 5$, 6, and 7 events in each sequence. The center panel shows the parent distribution of recurrence times for $M \geq 7$ that was sampled, and a histogram of the observed mean recurrence times for the sequences. Apparent characteristic (more frequent than expected) earthquakes, for which the observed recurrence time is less than T_{av} , plot above the log-linear frequency-magnitude relation in the left panels, and to the left of 1 in the center panels. Conversely, uncharacteristic (less frequent than expected) earthquakes plot below the log-linear relation in the left panels, and right of 1 in the center panels.

For example, the first row shows the results for earthquake sequences of length $T_{av}/2$. Due to their short length, 46% of the sequences contain no earthquakes with $M \geq 7$, 52% have only one, all but one of the remaining 2% have two earthquakes, and one has three. The mean inferred recurrence times for sequences with one, two, and three earthquakes are $T_{av}/2$, $T_{av}/4$, and $T_{av}/6$. Earthquakes with mean recurrence intervals greater than or equal to T_{av} are not observed in a sequence half that length, because the recurrence interval is inferred by dividing the history length by the number of earthquakes. Hence, although the parent distribution of recurrence times contains earthquakes with recurrence intervals greater than $T_{av}/2$, these are not observed. As a result, all



▲ **Figure 4.** Results of numerical simulations of earthquake sequences. Rows show results for sequences of different lengths. Left panels show the log-linear frequency-magnitude relation sampled, with dots showing the resulting mean recurrence times. Center panels show the parent distribution of recurrence times for $M \geq 7$ earthquakes (smooth curve) and the observed mean recurrence times (bars). Right panels show the fraction of sequences in which a given number of $M \geq 7$ earthquakes occurred. In each panel, darkly and lightly shaded circles and bars represent characteristic and uncharacteristic earthquakes, respectively. Open bars in the right panels are instances where no $M \geq 7$ events occurred, thus not appearing in the left and center panels.

observed $M \geq 7$ earthquakes are characteristic. Hence, due to the short sampling interval, in about half the cases we observe characteristic earthquakes, whereas in the other half no large earthquakes are observed. These missing earthquakes are the limiting case of uncharacteristic earthquakes.

In contrast, for $M \geq 5$ and 6, the dots in the left panel are symmetric about the b -value line, showing that the observed recurrences have no bias to shorter or longer intervals. (The visual bias due to plotting the logarithms is not apparent.) Any earthquake sequence is equally likely to contain characteristic and uncharacteristic earthquakes. The scatter increases with earthquake magnitude, because the recurrence of the more common smaller earthquakes is better determined.

The second row shows results for histories with length equal to the mean recurrence time for $M \geq 7$ earthquakes. Because of the longer series, only 14% contain no earthquakes, 69% have one, and 16% have two, so fewer than 1% have three or more. The most common recurrence times are the nominal value of T_{av} inferred for only one earthquake, and $T_{av}/2$ inferred from two earthquakes. Hence 17% of the series have characteristic earthquakes, a similar fraction have no large earthquakes, and uncharacteristic earthquakes less common than the mean recurrence still cannot be detected.

The situation changes for a history two times T_{av} . Very few sequences have no $M \geq 7$ earthquakes, and the most common (58%) number of earthquakes is two, as expected from the mean recurrence time. Moreover, earthquakes with recurrence times greater than T_{av} can be detected because sequences with one earthquake give an inferred recurrence time of $2 T_{av}$. Hence the histogram of observed recurrence times becomes more symmetrical and closer to the parent distribution. About 20% of the sequences have characteristic earthquakes, and a similar fraction have uncharacteristic ones, so the bias toward characteristic earthquakes is gone. For an even longer sequence, $3 T_{av}$ (bottom row), the observed distribution of recurrence times looks much like a discrete version of the parent distribution. Characteristic and uncharacteristic earthquakes each occur for about 25% of the sequences.

For even longer histories (Figure 5) the mean recurrence interval for $M \geq 7$ earthquakes continues to approach the mean of the parent distribution. Although characteristic and uncharacteristic earthquake behavior occurs, the mean recurrence times are grouped more closely about the parent mean than is the parent distribution, because the sequence mean is a random variable whose standard deviation decreases approximately inversely as the square root of the number of earthquake recurrences in the history.

Comparing the simulations for different history lengths illustrates a systematic trend. Histories shorter than or similar to the mean recurrence time of large earthquakes are biased toward either observations of apparent characteristic earthquakes or an absence of large earthquakes. Longer histories are equally likely to yield characteristic and uncharacteristic earthquakes, and the deviations from the mean recurrence time decrease.

These simulations illustrate the difficulty in inferring the true recurrence times of large earthquakes, especially for a region that has a short earthquake record relative to the recurrence time of large earthquakes. The largest observed earthquakes are likely to appear characteristic, although their true recurrence interval is much longer. This effect is likely to be significant, because we are drawn to examine fault systems that have records of large and destructive earthquakes. Conversely, we may underestimate the size and/or frequency of the largest earthquakes in a region, because some or all of these are missing from the short earthquake record. These effects become less likely for longer earthquake records, but even in these cases there is a reasonable probability of observing apparent characteristic or uncharacteristic earthquakes purely by chance.

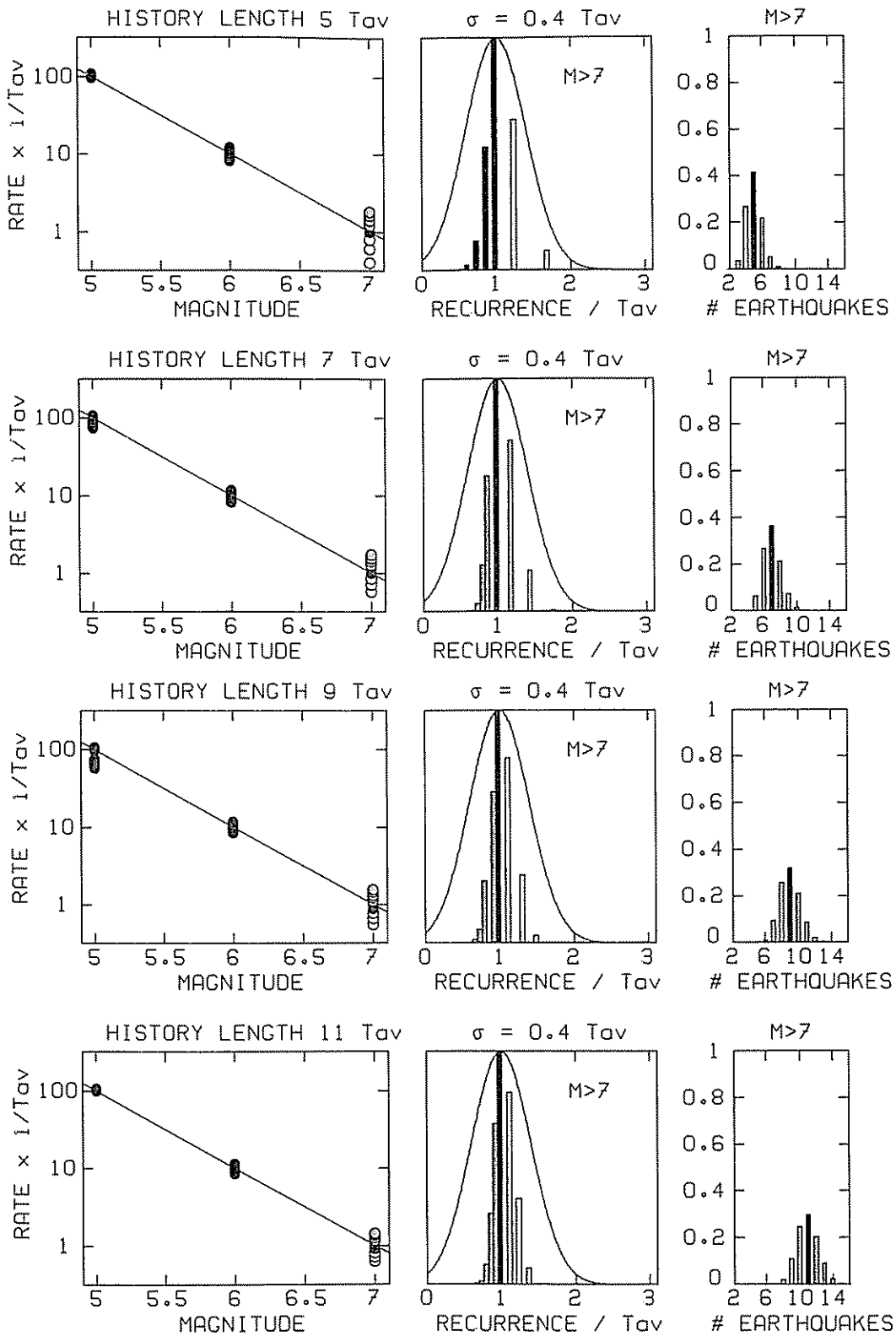
EXAMPLES: NEW MADRID AND WABASH

New Madrid

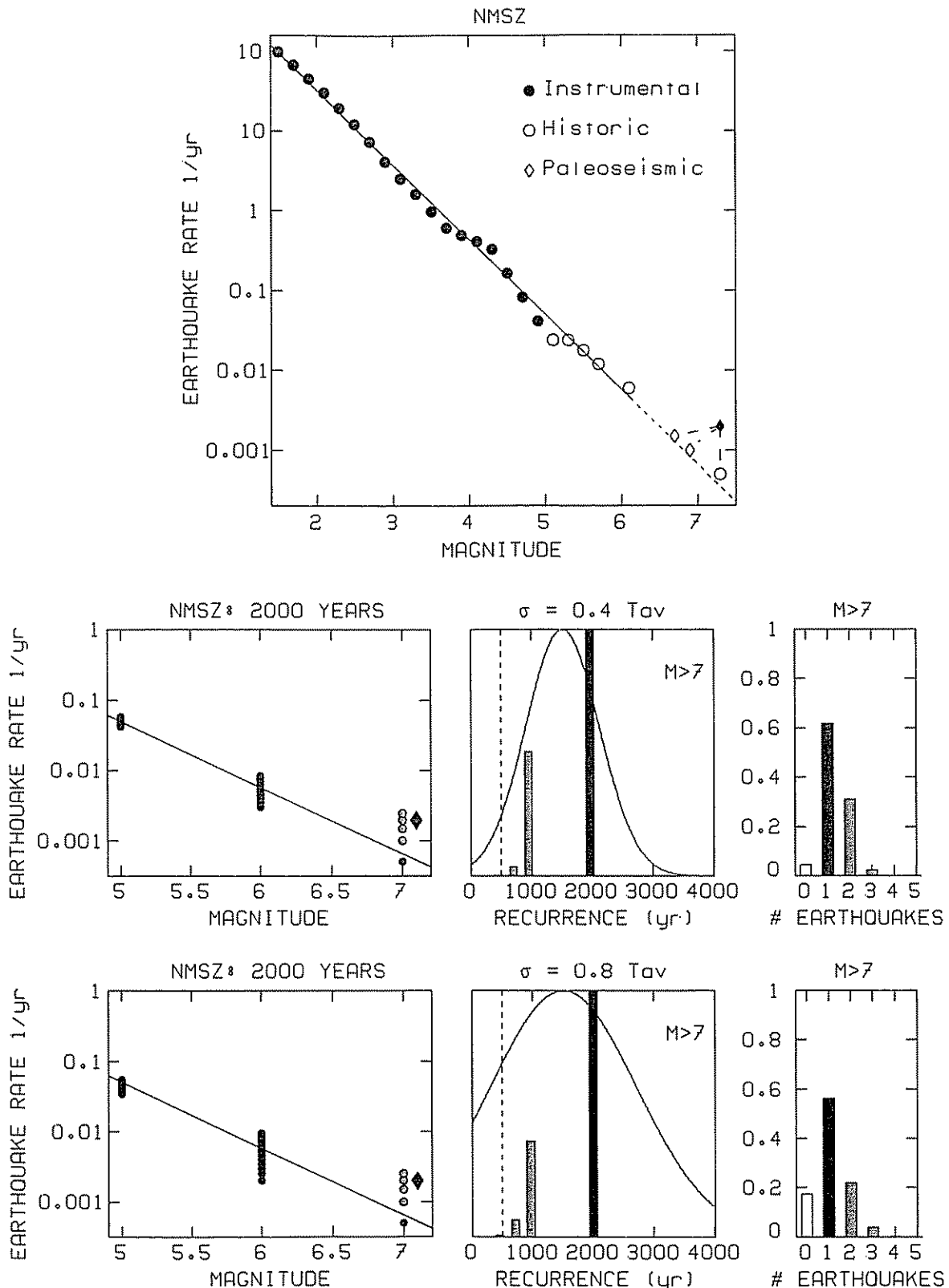
Our interest in the possible effects of short earthquake records was stimulated by the New Madrid seismic zone (NMSZ) in the central U.S. Historic and recorded earthquakes from the Nuttli and CERI catalogs for 1804–1974 and 1975–March 2003, http://www.eas.slu.edu/Earthquake_Center and <http://folkworm.ceri.memphis.edu>, define the three most active segments of the NMSZ. These are presumably the subsurface faults and show a Gutenberg-Richter distribution with a -3.2 and b -0.91. Since much current activity and possibly the paleoearthquakes occur in a broader region, we consider the seismicity within the region 35°–38°N/88°–91°W, which has similar a and b values (-3.4 and -0.94) (Figure 6; Newman *et al.*, 1999). Thus, in the NMSZ, we expect earthquakes with magnitudes greater than 5, 6, or 7 about every 20, 175, and 1,515 years, respectively.

Large earthquakes occurred in 1811 and 1812, with M_w -7.0–7.4 inferred from intensity reports (Hough *et al.*, 2000). The frequency and magnitudes of earlier such events have been inferred from the distribution of paleoliquefaction features similar to those associated with the 1811–1812 earthquakes. Wesnousky and Leffler (1992) did not find paleoliquefaction features comparable to those attributed to the 1811–1812 earthquakes and hence suggested that such large earthquakes were less common than implied by the instrumental and historic seismicity. In contrast, later studies found paleoliquefaction features that are interpreted as showing that earthquakes comparable to or perhaps somewhat smaller than those in 1811–1812 occurred in about 1450 ± 150 ($M \geq 6.7$) and A.D. 900 ± 100 ($M \geq 6.9$) (Tuttle, 2001).

Figure 6 (top) shows two different interpretations of the frequency and magnitudes of the largest NMSZ earthquakes. The solid diamond assumes that the paleoearthquakes were comparable to the 1811–1812 events (treated as one earthquake, as they appear paleoseismically), as favored by Tuttle (2001). In this case, characteristic earthquakes with $M > 7$ recur about every 500 years, more frequently than expected from the smaller earthquakes. The minimum magnitude esti-



▲ **Figure 5.** Results of numerical simulations for earthquake sequences longer than in Figure 4



▲ **Figure 6.** *Top* Frequency-magnitude data for seismicity of the New Madrid seismic zone (NMSZ). Solid line is log-linear fit to instrumental and historic data; dashed line is extrapolation. Solid diamond gives recurrence of large earthquakes assuming that two recorded paleoearthquakes were comparable in magnitude to the 1811–1812 events; open diamonds are for the minimum paleoseismic magnitudes. *Center and bottom* Results of numerical simulations of earthquake sequences approximating the NMSZ, plotted using conventions of Figure 4. Solid diamond in left panels corresponds to that in top panel; vertical dashed line in center panels illustrates the corresponding 500-yr recurrence for $M \geq 7$.

mates (open diamonds) are consistent with the smaller earthquakes and do not require characteristic earthquakes. We then explore the implications of the two different interpretations.

Because the earthquake record's duration is comparable to the expected recurrence interval of large earthquakes, we ran simulations for variables approximating New Madrid seismicity. Figure 6 shows two sets of simulations, each with 10,000 earthquake sequences derived from a log-linear frequency-magnitude relation with $a = 3.4$ and $b = 0.94$. All extend over 2,000 years, the period sampled by the paleoseismic data, which is only slightly longer than the ~1,500-yr mean recurrence interval for earthquakes with $M \geq 7$ predicted by the log-linear fit to the frequency-magnitude data. For the case $\sigma = 0.4 T_{av}$ (upper panels), about 5% of the sequences have no earthquakes, and 62% have only one, yielding a nominal 2,000-yr recurrence. The remaining 33% yield apparent characteristic earthquakes. Simulations for larger variability in recurrence times, $\sigma = 0.8 T_{av}$, are shown in the lower panels. More of the sequences (17%) have no earthquakes, and 26% have two or more.

It is interesting to compare the simulations to the characteristic earthquake interpretation of the paleoseismic data (solid diamond in Figure 6). Characteristic earthquakes occur for 33% and 26% of the simulations for $\sigma = 0.4$ and $0.8 T_{av}$. However, three or more earthquakes, corresponding to mean recurrence times of 667 years or less, are rare, occurring for only 2% and 4% of the simulations. Hence, if the simulations are representative, the short recurrence times inferred paleoseismically may be due to the sampling bias, but the probability of such a large difference between the inferred recurrence time and that predicted from the b -value line is small. Thus it seems more likely that the difference is either a real effect or reflects the paleoearthquakes being smaller than the 1811–1812 earthquakes (open diamonds in Figure 6).

Paleoseismicity Issues

This analysis bears out the issue of how well magnitudes estimated from the paleoliquefaction data compare to those inferred from seismological or historical data. Ideally, we would like to compare all three data types for the same earthquakes, because each has its own uncertainties. Unfortunately, this cannot be done at New Madrid, because the 1811–1812 earthquakes predate instrumental seismology. Comparisons with intensity data are possible, however.

The paleoearthquake magnitudes are inferred primarily from the spatial extent of the liquefaction features, although other methods are used to estimate the shaking at individual sites. The uncertainty in these estimates results from various sources and is difficult to quantify. Some of the uncertainty results from the scatter in the calibration curves used to infer the magnitude from the maximum distance of liquefaction, whereas other uncertainty results from estimates of the area, which can vary between studies. Results from different techniques vary, suggesting precision of about ± 0.2 magnitude units (Obermeier *et al.*, 2001). The accuracy of these estimates is difficult to assess for several reasons. For example,

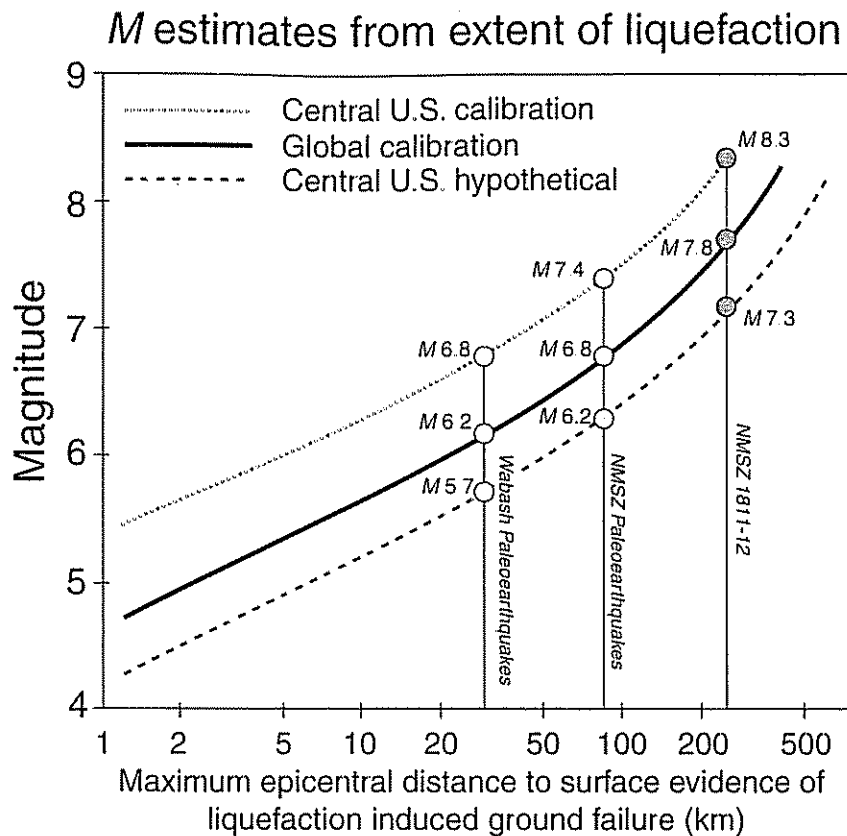
due to the limited datable markers before and after paleoliquefaction events, it is rarely possible to determine if such features resulted from one large or from numerous smaller events. Hence the 1811–1812 series of three large and many smaller earthquakes over a distributed region produced what appears as a single large liquefaction event.

For New Madrid, a further complexity results because it has been common to interpret paleoliquefaction data in the central U.S. using a different curve than that used globally (Figure 7). This curve implies that New Madrid earthquakes cause liquefaction out to a smaller distance than is typical globally. For example, liquefaction features ~250 km from the assumed epicenter are interpreted as evidence for a M -8.3 earthquake, rather than M -7.8 as would be inferred from the global curve. This practice arose because the curve was calibrated assuming the 1811–1812 to have been M 8.3 events (Johnston, 1996). Because more recent analysis finds that these earthquakes were low-to-mid M 7 (Hough *et al.*, 2000), a distinct central U.S. curve now seems unnecessary. Because the central U.S. curve biased paleoseismic magnitude estimates upward (Newman, 2000), using the global curve reduces the estimated magnitudes. Figure 6 uses magnitudes inferred from the global curve (Turtle, 2001), rather than the central U.S. curve often used for New Madrid. Even so, these estimates may still be biased upward, because the revised paleoseismic magnitude for 1811–1812 of 7.8 is about 0.5 units higher than inferred from the intensity data. One possibility is that New Madrid events cause liquefaction to a greater distance than is typical globally, much as they cause strong ground motion to larger distances than comparable California earthquakes, owing to differences in seismic wave attenuation.

Wabash

The issue of magnitudes from paleoliquefaction data is even more crucial in areas where the large earthquakes predate both instrumental and historical data. This is the case for the Wabash Valley seismic zone, a northward extension of the New Madrid zone (Bear *et al.*, 1997). Although the present seismicity is even lower than for the New Madrid zone, paleoliquefaction studies report evidence for four earthquakes with inferred magnitudes greater than 6.8 in the past 12,000 years (Pond and Martin, 1997). This magnitude estimate, however, relies on the assumption that central U.S. earthquakes differ from the global curve (Figure 7). Application of the global curve suggests that these should be viewed instead as magnitude greater than 6.2. The inferred rate of large earthquakes thus depends significantly on the inferred magnitudes.

This effect is illustrated in Figure 8 (top), which shows frequency-magnitude data for the Wabash zone, treated here as the box 85.8°–88.75°W, 37.6°–39.7°N. Combining the CERI catalog of seismologically recorded small earthquakes and the Nuttli catalog of historic earthquakes with estimated magnitudes yields a Gutenberg-Richter distribution out to about M 6 with a -2.1 and b -0.74. The four paleoearthquakes discussed by Pond and Martin (1997) thus appear



▲ **Figure 7.** Illustration of estimation of earthquake magnitude from the extent of paleoliquefaction. For a given set of observations, the inferred magnitude depends on whether the global calibration curve or one for the central U.S. is used. Moreover, if NMSZ earthquakes cause liquefaction to greater distances than the global average, the inferred magnitude would be even lower (modified from Pond and Martin, 1997).

uncharacteristic whether their magnitudes are inferred from the global (solid diamonds) or central U.S. (open diamonds) calibration curves. The log-linear fit to the small earthquakes predicts a recurrence time for $M \geq 7$ events greater than 1,700 years, whereas the paleoseismics imply a recurrence time for $M \geq 7$ events greater than 5,000 years for the central U.S. calibration, or 10,000 years for the global one.

To explore whether this uncharacteristic behavior might be a sampling artifact, we ran two sets of simulations (Figure 8). Although these show both characteristic and uncharacteristic earthquakes, neither yields such dramatic uncharacteristic behavior, because the 12,000-year paleoseismic record is about seven times the predicted recurrence time for $M \geq 7$ earthquakes. Hence either the uncharacteristic behavior is real, or the paleoseismic record captures only a small fraction of the large pre-instrumental earthquakes.

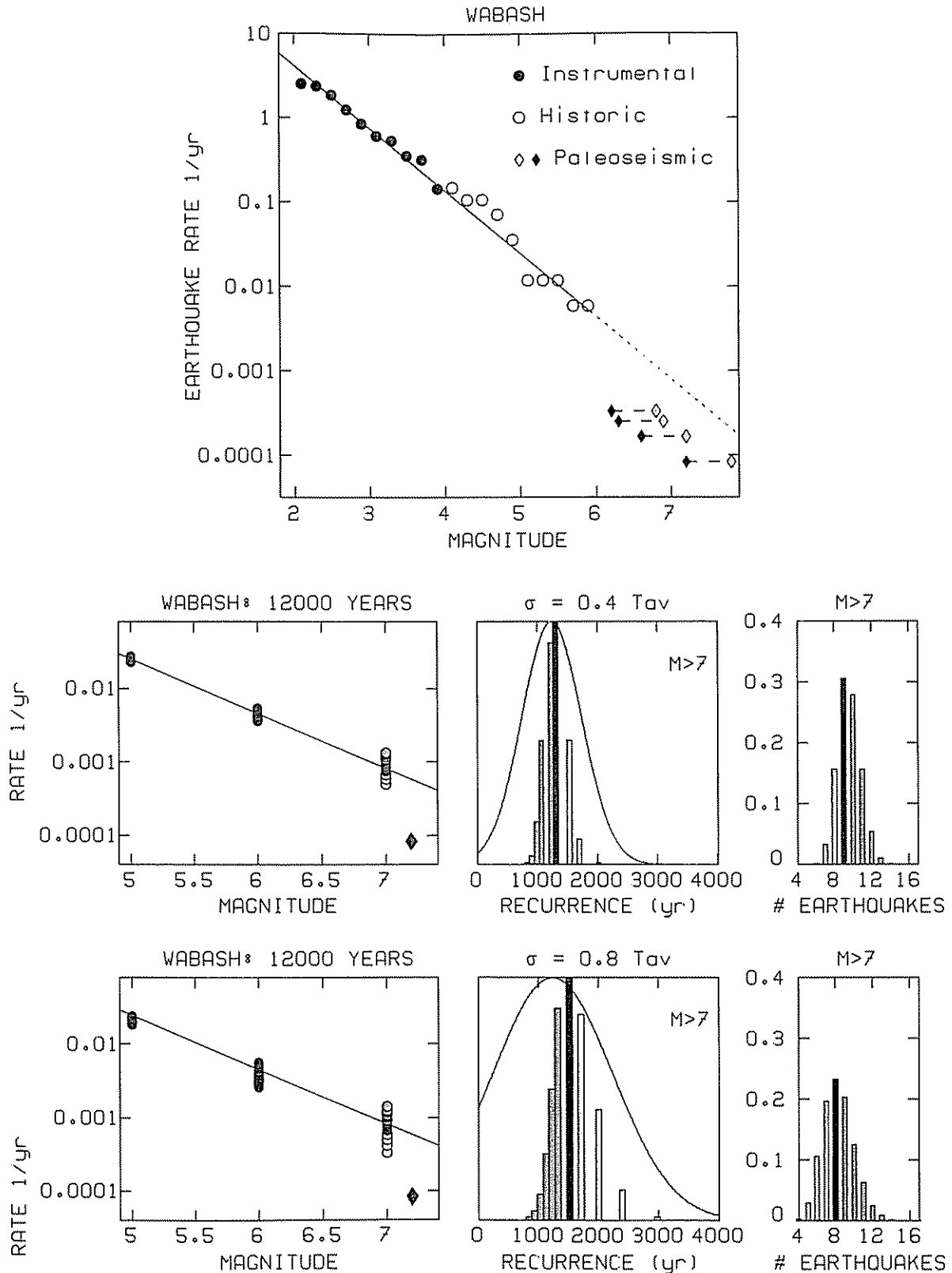
It is interesting to compare Wabash to New Madrid (Figure 9). The instrumental and historic records reveal New Madrid to be more active, with moderate earthquakes ($M \geq 5$) occurring once ~20 years as opposed to once every 40 years for Wabash. However, because the b values (slopes) of a linear fit to the frequency-magnitude data differ, the lines cross at about $M 6.5$, predicting that the Wabash seismicity would be comparable for larger magnitudes. The paleoseismic data, however, suggest lower Wabash seismicity. It is

unclear whether this discrepancy reflects limitations in estimating the frequency-magnitude relation, perhaps due to catalog incompleteness, limitations in the paleoseismic record, a real physical difference between large and small earthquakes, or a combination of the above.

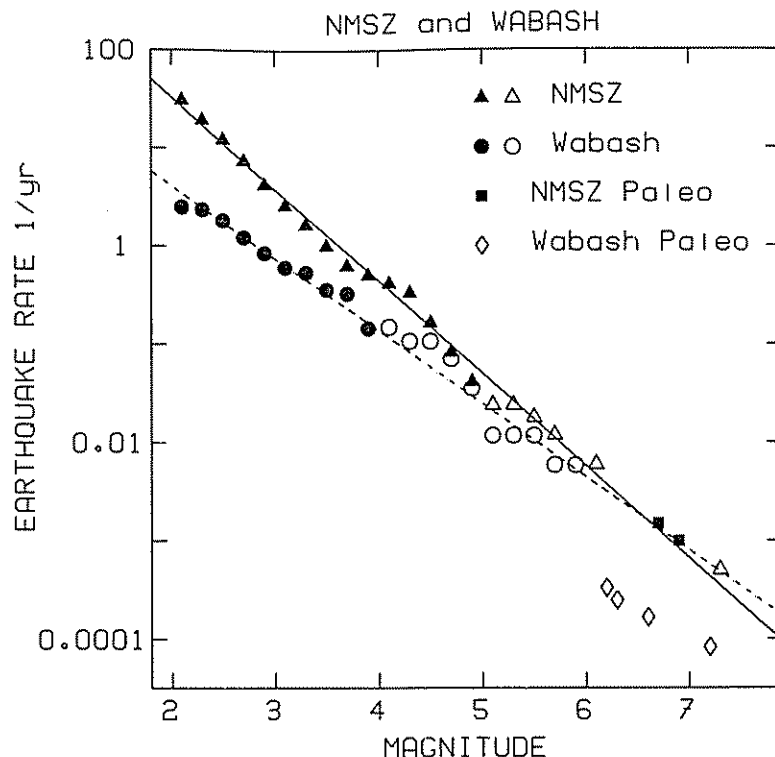
New Madrid Seismicity as Aftershocks?

A final point worth noting about New Madrid is that the rate of seismicity is sufficiently low that much or almost all of the seismicity might be aftershocks of the large 1811–1812 earthquakes. Large earthquakes are typically followed by aftershock activity that decays to a lower level of seismicity interpreted as “normal” background seismicity (e.g., Ogata and Shimazaki, 1984). However, as shown in Figure 10 (top), such a transition is not obvious in the New Madrid data. It is unclear whether this is simply because of the low seismicity rate and limitations of the catalog, or because most of the catalog reflects a long aftershock sequence. The simplest test, comparison of pre- and post-1811 data, is not possible.

Some insight can be obtained by considering a compilation of aftershock durations from Dieterich (1994). As shown in Figure 10 (bottom), aftershock durations (t_d) generally increase with the inferred recurrence time of the mainshock (t_r), such that $t_d/t_r \sim 0.05$. Hence for New Madrid, with $t_r \sim 500$ – $1,000$ years, one would expect a 25–50-year-long after-



▲ **Figure 8.** *Top* Frequency-magnitude data for seismicity of the Wabash Valley seismic zone. Solid line is log-linear fit to instrumental and historic data; dashed line is extrapolation. Diamonds give recurrence of large earthquakes for magnitudes inferred from the global (solid) or central U.S. (open) paleoliquefaction calibration curves. *Center and bottom.* Results of numerical simulations of earthquake sequences approximating the Wabash seismic zone, plotted using conventions of Figure 4.



▲ **Figure 9.** Comparison of frequency-magnitude data for the New Madrid and Wabash Valley seismic zones. Solid (NMSZ) and dashed (Wabash) lines show log-linear fit to instrumental (solid symbols) and historic (open symbols) data. Paleoseismic data are shown using minimum magnitudes for New Madrid and global magnitude calibration for Wabash.

shock sequence, which could not be resolved with the available catalog.

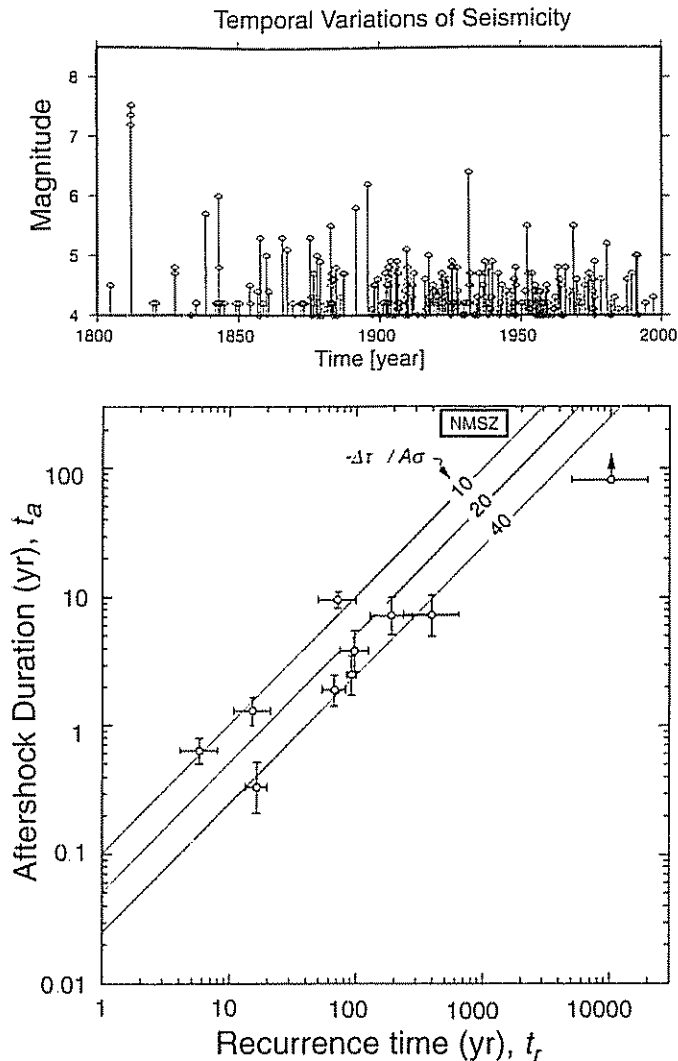
Thus if New Madrid seismicity is a long-lived aftershock sequence, it would differ from the other examples in Figure 10. This may be the case, because its continental intraplate tectonic setting differs from the others, which occur either at plate boundaries (California, Alaska, Japan) or at an oceanic hot spot (Hawaii). In Dieterich's (1994) model, aftershock duration and mainshock recurrence are related by $t_d/t_r = A \sigma / \Delta \tau$, where A is a fault constitutive parameter, σ is the normal stress on the fault, and $\Delta \tau$ is the mainshock stress drop. We can crudely estimate A for an interplate setting by assuming that $\Delta \tau \sim 30$ bars (Kanamori and Anderson, 1975) and $\sigma \sim 100$ bars (Zoback *et al.*, 1987), so $t_d/t_r \sim 0.05$ implies $A \sim 0.015$. On intraplate faults, we might expect both higher earthquake stress drops, perhaps 100 bars (Kanamori and Anderson, 1975), and larger normal stresses, perhaps 1 Kbar (Zoback, 1992; Grana and Richardson, 1996). Hence if A were the same, $t_d/t_r \sim 0.15$, which would predict a much longer aftershock sequence. This estimate would be roughly consistent with present New Madrid seismicity being an aftershock sequence, with $t_d/t_r \sim 200/500 \sim 200/1,000 = 0.4 \sim 0.2$.

From these considerations, the idea of a long-duration aftershock sequence is possible, but hard to prove or disprove. Interestingly, the largest earthquakes since 1811–1812 are at the ends of or outside the area that seems likely to have been the mainshock rupture, a pattern often observed in after-

shock studies (Das and Henry, 2003). This possibility would have interesting consequences for New Madrid, and perhaps other low-seismicity intraplate settings, if the large earthquakes are followed by long-duration aftershock sequences. If current NMSZ seismicity is dominated by long-lived aftershocks, and should not be included in the Gutenberg-Richter relation, then the $M > 7$ earthquakes every 500 years would appear considerably more characteristic than they already do in Figure 6. Alternatively, this might be consistent with the idea that New Madrid seismicity results from a transient zone of stress release that became active within the past few million or even tens of thousands of years (Schweig and Ellis, 1994) and is now "shutting" down (Newman *et al.*, 1999). If so, it would be hard to define meaningfully either the background seismicity rate or the recurrence time of large earthquakes, making a characteristic earthquake discussion moot. It is worth noting that numerical modeling by Rydelek and Pollitz (1994) found that fossil strain effects from the 1811–1812 earthquakes may still be significant, although the reported high geodetic strain rates (Liu *et al.*, 1992) that prompted this analysis are not found in subsequent studies (Newman *et al.*, 1999; Santillan *et al.*, 2002).

DISCUSSION

These simulations and examples illustrate the challenge of understanding earthquake recurrence in areas of low seismicity. They indicate that characteristic, uncharacteristic, and



▲ **Figure 10.** *Top* Time sequence of seismicity for the New Madrid seismic zone (seismicity within the region 35°–38°N/88°–91°W) *Bottom.* Relation between aftershock durations and mainshock recurrence times (Dieterich, 1994), with estimate from New Madrid assuming a 200-year aftershock sequence duration. Such a duration would be longer than expected from the other data, but might be consistent with an intraplate setting

missing earthquakes might occur purely as an artifact of short earthquake records. The significance of this effect depends on the frequency-magnitude relations and how well they can be estimated, and the distribution of recurrence times for actual faults. Unfortunately, we know little about these factors and hence do not know whether discrepancies in the rates of large and small earthquakes are a real effect or are artifacts of sampling and/or other biases.

The fact that deviations might occur from small sampling does not prove that observed deviations are due to this effect. We think it difficult to determine whether characteristic earthquakes in regions with short earthquake records are real. A possible argument against their reality is that global seismological data sets and seismological data for continental interiors (Figure 1) show the largest earthquakes to be less frequent than expected from the smaller ones. However, under Wesnousky *et al.*'s (1983) hypothesis, the global and regional

data sets would reflect the sum of many faults, each with a characteristic earthquake distribution

There is the possibility that characteristic or uncharacteristic earthquakes are real in some areas and artifacts in others. Wesnousky (1994) and Stirling *et al.* (1996) found that strike-slip faults with larger offsets are less complex, having fewer steps per unit length, and more prone to characteristic earthquakes than Gutenberg-Richter behavior. They propose that faults evolve, such that as the total slip increases the fault becomes smoother, increasing the rate of larger earthquakes relative to smaller ones. Hence long-lived large-offset faults on strike-slip plate boundaries, such as the San Andreas, would tend toward characteristic behavior. In contrast, young intraplate faults with smaller offsets, such as those in the New Madrid zone where the absence of large-scale crustal deformation suggests that the seismic zone is a young feature (Schweig and Ellis, 1994), might tend toward Gutenberg-Richter behavior. This could be a general tendency if such zones of

intraplate seismicity are transient features, the present locus of intraplate strain release that migrates with time between fossil weak zones (Newman *et al.*, 1999).

Although the simulations explored only the possibility that time-sampling artifacts could produce deviations from log-linear frequency-magnitude relations, apparent deviations might result from other limitations of the earthquake record. Often the log-linear frequency-magnitude relation is inferred from seismological data, whereas the recurrence intervals and magnitudes of the largest earthquakes are characterized from paleoseismic data. Hence uncertainties in paleoseismic data could cause a mismatch with the seismological data, as illustrated for New Madrid in Figure 6. If the paleoseismic data overestimate magnitudes or underestimate recurrence times, apparent characteristic earthquakes will occur. Conversely, underestimating magnitudes or overestimating recurrence times will yield uncharacteristic earthquakes. For example, the Wabash zone paleoseismic data might be consistent with the instrumental data if paleoliquefaction data record only a small fraction of the paleoearthquakes. Assessing these possible effects, and whether they could occur in areas where the paleoseismic data are from surface faulting observations rather than paleoliquefaction, is a challenge. This possibility is suggested by the observation that fault slip can be overestimated by focusing on those sites where surface slip was largest, perhaps due to special soft ground conditions, or due to complex displacement patterns where surface rupture occurs in a complex shear zone (Johnson, 1997).

Finally, it is worth noting that our simulations treat earthquake recurrence in a simple statistical fashion and assume that the frequency of different simulation outcomes reflects the probability that these would actually occur. Hence we do not address the physics of earthquake recurrence, because the relative roles of deterministic and random factors are not understood and seem likely to remain so for a very long time. The variability of earthquake recurrence times may reflect random failure processes (Main, 1996) and/or deterministic factors such as the amount of slip in previous earthquakes (Shimazaki and Nakata, 1980) and the stress effects of nearby earthquakes (Stein, 1999). Some of these effects could be included in more complex simulations. For example, log-normal distributions of recurrence times, or other frequency-magnitude relations, could be used. Similarly, rather than treating the sizes of successive earthquakes as independent, the slip and hence moment release rate over a number of seismic cycles could be required to remain constant, such that a larger-than-average earthquake makes the next one more likely to be smaller. We have not done so both for simplicity and because it is not clear whether this should be the case over a few earthquake cycles. On plate boundaries, it seems likely that the long-term slip rate over many seismic cycles would tend to a steady value, assuming that the aseismic slip fraction remained constant. Whether this would be the case for only a few cycles is not clear, since we do not know whether the vari-

ation in recurrence times (*e.g.*, Figure 3) reflects corresponding variation in slip (Schwartz and Coppersmith, 1984). The situation is even less clear for intraplate earthquakes. For example, we have argued that the slip rate in New Madrid earthquakes is related to the relative motion across the seismic zone and that the recurrence of large earthquakes is consistent with the Gutenberg-Richter relation (Newman *et al.*, 1999), whereas Kenner and Segall (2000) favor a non-steady-state model, in which recent relaxation of the lower crust allows transient release of accumulated stress, so that for some time large earthquakes occur more frequently than implied by relative motion across the seismic zone. These issues are likely to remain unresolved for a long time, because long earthquake records are needed to test models (Kagan and Jackson, 1991), and the few long records available sometimes show complicated behavior. Hence simulations that replicate some aspects of seismicity are one way to gain insight into these issues.

In summary, in a variety of applications including seismic hazard studies, we think it worth bearing in mind that the rates and sizes of the largest observed earthquakes in a region may differ significantly from their true long-term values. In some cases the largest observed earthquakes may not be the largest that will occur, and in others the largest earthquakes may be more or less frequent than the available record implies. ☒

ACKNOWLEDGMENTS

We thank Arvid Johnson, Hiroo Kanamori, Susan Hough, and Mark Stirling for helpful comments. This research was supported by USGS grant 03HQGR0104.

REFERENCES

- Bear, G. W., J. A. Rapp, and A. J. Rudman (1997). Seismic interpretation of the deep structure of the Wabash Valley fault system, *Seismological Research Letters* **68**, 624–640.
- Das, S. and C. Henry (2003). Spatial relation between main earthquake slip and its aftershock distribution, *Review of Geophysics* **41**, doi:10.1029/2002RG000119.
- Dieterich, J. H. (1994). A constitutive law for rate of earthquake production and its application to earthquake clustering, *Journal of Geophysical Research* **99**, 2,601–2,618.
- Grana, J. P. and R. M. Richardson (1996). Tectonic stress within the New Madrid seismic zone, *Journal of Geophysical Research* **101**, 5,445–5,458.
- Gutenberg, B. and C. F. Richter (1944). Frequency of earthquakes in California, *Bulletin of the Seismological Society of America* **34**, 185–188.
- Hough, S., J. G. Armbruster, L. Seeber, and J. F. Hough (2000). On the Modified Mercalli Intensities and magnitudes of the 1811/1812 New Madrid, central United States, earthquakes, *Journal of Geophysical Research* **105**, 23,839–23,864.
- Howell, B. F., Jr. (1985). On the effect of too small a data base on earthquake frequency diagrams, *Bulletin of the Seismological Society of America* **75**, 1,205–1,207.
- Ishibashi, K. (1981). Specification of a soon-to-occur seismic faulting in the Tokai district, central Japan, based upon seismotectonics, in D. Simpson and P. Richards (editors), *Earthquake Prediction*, Maurice Ewing Series, volume 4, Washington, DC: AGU, 297–332.

- Johnson, A. (1997). Structures formed during the Landers-Big Bear, California earthquake sequence, *Proceedings of the 4th Workshop on Non-Linear Dynamics and Earthquake Prediction*, Trieste, Italy: International Centre for Theoretical Physics.
- Johnston, A. C. (1996). Seismic moment assessment of earthquakes in stable continental regions - II. New Madrid 1811-1812, Charleston 1886 and Lisbon 1755, *Geophysical Journal International* **126**, 314-344.
- Kagan, Y. Y. (1993). Statistics of characteristic earthquakes, *Bulletin of the Seismological Society of America* **83**, 7-24.
- Kagan, Y. Y. (1996). Comment on "The Gutenberg-Richter or characteristic earthquake distribution, which is it?" by S. Wesnousky, *Bulletin of the Seismological Society of America* **86**, 274-285.
- Kagan, Y. Y. and D. D. Jackson (1991). Seismic gap hypothesis: Ten years after, *Journal of Geophysical Research* **96**, 21,419-21,431.
- Kanamori, H. and D. L. Anderson (1975). Theoretical basis of some empirical relations in seismology, *Bulletin of the Seismological Society of America* **65**, 1,073-1,095.
- Kenner, S. J. and P. Segall (2000). A mechanical model for intraplate earthquakes: Application to the New Madrid seismic zone, *Science* **289**, 2,239-2,332.
- Main, I. G. (1996). Statistical physics, seismogenesis, and seismic hazard, *Review of Geophysics* **34**, 433-462.
- Main, I. G. (2000). Apparent breaks in scaling in the earthquake cumulative frequency-magnitude distribution: Fact or artifact, *Bulletin of the Seismological Society of America* **90**, 86-97.
- Meghraoui, M., B. Delouis, M. Ferry, D. Giardini, P. Huggenberger, I. Spotteke, and M. Graner (2001). Active normal faulting in the Upper Rhine graben and the paleoseismic identification of the 1356 Basel earthquake, *Science* **293**, 2,070-2,073.
- Newman, A. (2000). *Geodetic and Seismic Studies of the New Madrid Seismic Zone and Implications for Earthquake Recurrence and Seismic Hazard*, Ph.D. thesis, Northwestern University.
- Newman, A., S. Stein, J. Weber, J. Engeln, A. Mao, and T. Dixon (1999). Slow deformation and lower seismic hazard at the New Madrid seismic zone, *Science* **284**, 619-621.
- Nishenko, S. P. and R. Buland (1987). A generic recurrence interval distribution for earthquake forecasting, *Bulletin of the Seismological Society of America* **77**, 1,382-1,399.
- Obermeier, S., E. C. Pond, and S. M. Olson (2001). Paleoliquefaction studies in continental settings: Geologic and geotechnical factors in interpretations and back-analysis, U.S. Geological Survey Open-File Report, 1-29.
- Ogata, Y. and K. Shimazaki (1984). Transition from aftershock to normal activity: The 1965 Rat Islands earthquake aftershock sequence, *Bulletin of the Seismological Society of America* **74**, 1,757-1,765.
- Okal, E. A. and B. Romanowicz (1994). On the variation of b -values with earthquake size, *Physics of the Earth and Planetary Interiors* **87**, 55-76.
- Pond, E. C. and J. R. Martin (1997). Estimated magnitudes and accelerations associated with prehistoric earthquakes in the Wabash Valley region of the central U.S., *Seismological Research Letters* **68**, 611-623.
- Rydelek, P. A. and F. F. Pollitz (1994). Fossil strain from the 1811-1812 New Madrid earthquakes, *Geophysical Research Letters* **21**, 2,303-2,306.
- Santillan, V., R. Smalley, and M. Ellis (2002). Deformation in the New Madrid seismic zone from continuous GPS geodesy (abstract), *Eos, Transactions of the American Geophysical Union* **83**, Fall Meeting Supplement, F378.
- Schwartz, D. P. and K. J. Coppersmith (1984). Fault behavior and characteristic earthquakes: Examples from the Wasatch and San Andreas Fault zones, *Journal of Geophysical Research* **89**, 5,681-5,698.
- Schweig, E. S. and M. A. Ellis (1994). Reconciling short recurrence intervals with minor deformation in the New Madrid seismic zone, *Science* **264**, 1,308-1,311.
- Shimazaki, K. and T. Nakata (1980). Time-predictable recurrence model for large earthquakes, *Geophysical Research Letters* **7**, 279-282.
- Sieh, K., M. Stuiver, and D. Brillinger (1989). A more precise chronology of earthquakes produced by the San Andreas Fault in Southern California, *Journal of Geophysical Research* **94**, 603-624.
- Stein, R. (1999). The role of stress transfer in earthquake occurrence, *Nature* **402**, 605-609.
- Stein, S. and M. Wyssession (2003). *Introduction to Seismology, Earthquakes, and Earth Structure*, Oxford: Blackwell.
- Stirling, M. W., S. G. Wesnousky, and K. Shimazaki (1996). Fault trace complexity, cumulative slip, and the shape of the magnitude-frequency distribution for strike-slip faults, a global survey, *Geophysical Journal International* **124**, 883-868.
- Triep, E. G. and L. R. Sykes (1997). Frequency of occurrence of moderate to great earthquakes in intracontinental regions: Implications for changes in stress, earthquake prediction, and hazard assessments, *Journal of Geophysical Research* **102**, 9,923-9,948.
- Turtle, M. P. (2001). The use of liquefaction features in paleoseismology: Lessons learned in the New Madrid seismic zone, central U.S., *Journal of Seismology* **5**, 361-380.
- Wesnousky, S. G. (1994). The Gutenberg-Richter or characteristic earthquake distribution, which is it?, *Bulletin of the Seismological Society of America* **84**, 1,940-1,959.
- Wesnousky, S. G. (1996). Reply to Y. Kagan's comment on "The Gutenberg-Richter or characteristic earthquake distribution, which is it?", *Bulletin of the Seismological Society of America* **86**, 286-291.
- Wesnousky, S. G. and L. M. Leffler (1992). The repeat time of the 1811 and 1812 New Madrid earthquakes: A geological perspective, *Bulletin of the Seismological Society of America* **82**, 1,756-1,785.
- Wesnousky, S. G., C. H. Scholz, K. Shimazaki, and T. Matsuda (1983). Earthquake frequency distribution and the mechanics of faulting, *Journal of Geophysical Research* **88**, 9,331-9,340.
- Youngs, R. R. and K. J. Coppersmith (1985). Implications of fault slip rates and earthquake recurrence models to probabilistic seismic hazard estimates, *Bulletin of the Seismological Society of America* **75**, 939-964.
- Zoback, M. D., M. L. Zoback, V. S. Mount, J. Suppe, J. P. Eaton, J. H. Healy, D. Oppenheimer, P. Reasenber, L. Jones, C. B. Raleigh, I. G. Wong, O. Scotti, and C. Wentworth (1987). New evidence on the state of stress of the San Andreas Fault system, *Science* **238**, 1,105-1,111.
- Zoback, M. L. (1992). Stress field constraints on intraplate seismicity in eastern North America, *Journal of Geophysical Research* **97**, 11,761-11,782.

Department of Geological Sciences
Northwestern University
Evanston, IL 60208
Telephone: +1-847-491-5265
E-mail: seth@earth.northwestern.edu
(S.S.)

Los Alamos National Laboratory
EES-9, MS D462
Los Alamos, NM 87545
Telephone: +1-505-665-3570
E-mail: anewman@lanl.gov
(A.N.)