

The Geological Society of America
 Special Paper 425
 2007

Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments?

Alan L. Kafka

Weston Observatory, Department of Geology and Geophysics, Boston College, Weston, Massachusetts 02493, USA

ABSTRACT

The spatial distribution of seismicity is often used as one of the indicators of zones where future large earthquakes are likely to occur. This is particularly true for intraplate regions such as the central and eastern United States, where geology is markedly enigmatic for delineating seismically active areas. Although using past seismicity for this purpose may be intuitively appealing, it is only scientifically justified if the tendency for past seismicity to delineate potential locations of future large earthquakes is well-established as a real, measurable, physical phenomenon as opposed to an untested conceptual model. This paper attempts to cast this problem in the form of scientifically testable hypotheses and to test those hypotheses. Ideally, thousands (or even millions) of years of data would be necessary to solve this problem. Lacking such a long-term record of seismicity, I make the “logical leap” of using data from other regions as a proxy for repeated samples of seismicity in intraplate regions. Three decades of global data from the National Earthquake Information Center are used to explore how the tendency for past seismicity to delineate locations of future large earthquakes varies for regions with different tectonic environments. This exploration helps to elucidate this phenomenon for intraplate environments. Applying the results of this exercise to the central and eastern United States, I estimate that future earthquakes in the central and eastern United States (including large and damaging earthquakes) have ~86% probability of occurring within 36 km of past earthquakes, and ~60% probability of occurring within 14 km of past earthquakes.

Keywords: seismicity, earthquakes, intraplate, statistics.

INTRODUCTION

Many seismic hazard studies rely heavily on seismicity as a presumed indicator of zones where future large earthquakes are likely to occur. This is particularly true for intraplate regions, where the cause of earthquakes is largely unknown, and seismically active geological and geophysical features are difficult to delineate. The most recent U.S. National Seismic Hazard Maps, for example, rely heavily on the observed record of seismicity for mapping the hazard in the central and eastern United States (e.g., Frankel, 1995; Frankel et al., 1996; Wheeler and Frankel, 2000).

While this approach may be intuitively appealing, it is only scientifically justified if the tendency for past seismicity to delineate zones where future large earthquakes are likely to occur is well-established as a real, measurable, physical phenomenon as opposed to an untested conceptual model. However, the scientific basis for measuring this tendency has yet to be fully explored, particularly for intraplate regions, and as Lord Kelvin put it: “I often say that when you can measure what you are speaking about, and express it in numbers, you know something about it; but when you cannot measure it, when you cannot express it in numbers, your knowledge is of a meagre and unsatisfac-

Kafka, A.L., 2007, Does seismicity delineate zones where future large earthquakes are likely to occur in intraplate environments?, *in* Stein, S., and Mazzotti, S., ed., Continental Intraplate Earthquakes: Science, Hazard, and Policy Issues: Geological Society of America Special Paper 425, p. 35–48, doi: 10.1130/2007.2425(03). For permission to copy, contact editing@geosociety.org. ©2007 The Geological Society of America. All rights reserved.

tory kind ...” (Lord Kelvin, Sir William Thompson, “Electrical Units of Measurement,” in *Popular Lectures and Addresses*, v. 1, p. 72–73, 1883). Furthermore, even if we were to establish that this tendency could be measured in principle, there would still remain the questions of: (1) how to actually go about measuring it, (2) whether or not currently available earthquake catalogs provide representative samples of this tendency, and (3) how it varies from one tectonic environment to the next.

The purpose of this paper is to explore these issues, with particular emphasis on intraplate environments, and with the objective of shedding some light on the extent to which seismicity delineates zones where future large earthquakes are likely to occur in the central and eastern United States. The challenge in this type of investigation is to find a way to cast the problem in the form of testable hypotheses and then to use observed earthquake catalogs to test them. Figure 1 illustrates my approach for casting this problem in the form of testable hypotheses: H1 is the hypothesis that future large earthquakes occur only where past earthquakes have occurred; H2 is the hypothesis that future large earthquakes occur only where past earthquakes have not occurred; and H3 represents the hypothesis that future large earthquakes are equally likely anywhere. This study is essentially a test of hypothesis H1 for various tectonic environments, with emphasis on intraplate regions and with a specific focus on the central and eastern United States. I find that, while H1 is not unequivocally verified for the central and eastern United States, it seems as reasonable a conceptual model for the central and eastern United States as it is for many other parts of the world, including some plate-boundary regions.

I began investigating these issues shortly after a workshop on seismic hazard mapping in the northeastern United States

Relative to the Spatial Distribution of Past Earthquakes, Future Large Earthquakes Occur:

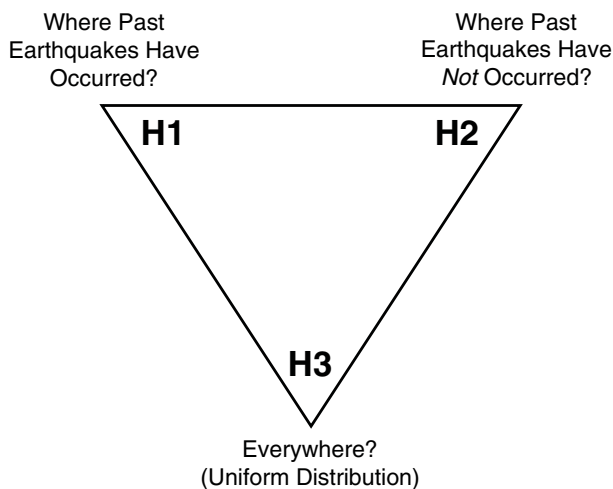


Figure 1. Diagram representing the hypotheses tested in this study (adapted from Kafka and Levin, 2000).

held in 1994 to obtain input for the next generation of the U.S. National Seismic Hazard Maps. The concept of using seismicity to indicate zones where future large earthquakes are likely to occur in the central and eastern United States was presented at the workshop, and this motivated my first investigation of this concept, an attempt to test hypothesis H1 for earthquakes in the northeastern United States. Kafka and Walcott (1998) found that, on average, “future” (i.e., later-occurring) earthquakes in the northeastern United States tended to occur in the vicinity of past earthquakes more frequently than would be expected for a random distribution of future earthquakes, and we were curious to see if the same pattern would be found in other regions. This led us to conduct similar tests of H1 in various regions, including the entire central and eastern United States (Kafka and Levin, 2000). For a variety of regions and tectonic environments, we found that future earthquakes tended to occur in the vicinity of past earthquakes more frequently than would be expected for a random distribution of future earthquakes (Kafka and Levin, 2000; Kafka, 2002). While these results are not particularly surprising, we consider them a preliminary step toward investigating the scientific basis for relying on past seismicity as an indicator of zones where future large earthquakes are likely to occur. A surprising result was that, for the regions and time periods studied, we did not detect any statistically significant difference in the percentage of future earthquakes occurring near past earthquakes for intraplate versus plate-boundary environments. In this paper, I first summarize these previous studies, and then extend these investigations to other parts of the world in an effort to discern how the tendency for past seismicity to delineate zones where future large earthquakes are likely to occur varies with tectonic environment. This provides a foundation for estimating the probability of future large earthquakes occurring near past earthquakes in the central and eastern United States.

METHODS AND REVIEW OF PREVIOUS STUDIES

In addition to the studies associated with the development of the U.S. National Seismic Hazard Maps (e.g., Frankel, 1995; Frankel et al., 1996), other studies have explored the tendency for past seismicity to delineate zones of future large earthquakes. For example, Cao et al. (1996) estimated the seismic hazard in Southern California from background seismicity, using the assumption that future large earthquakes cluster spatially near locations of historical earthquakes of magnitude ≥ 4.0 . Jackson and Kagan (1999) tested forecasts of future earthquakes in the Northwest and Southwest Pacific based on smoothed versions of past seismicity (magnitude ≥ 5.8) using the Harvard Centroid Moment Tensor (CMT) catalog of earthquakes (e.g., Dziewonski et al., 1999). They found that the actual catalogs for both regions were quite consistent with their forecast model.

In our approach, we use a method that was developed over the course of our past studies of this phenomenon (Kafka and Walcott, 1998; Kafka and Levin, 2000; Kafka, 2002). This

method is “purely statistical” in the sense that no attempt is made here to explain the physical cause of the earthquakes or the physical reasons why the earthquakes occur in some places and don’t occur in other places. On a global scale, the locations of future earthquakes will, of course, be dominated by the process of plate tectonics, but the very occurrence of intraplate earthquakes means that delineation of plate boundaries alone is not the sole indicator of where earthquakes occur. While a physical understanding of why intraplate earthquakes occur where they do is an ultimate, fundamental goal, the goal of this study is more modest: to systematically investigate the pattern of the relationship between where intraplate earthquakes occurred in the past versus where they will occur in the future.

Because our method is analogous to the configuration of a cellular phone system, we affectionately refer to it as the “cellular seismology” method. We construct circles of a given radius around each epicenter in an earthquake catalog (the “before” catalog), and investigate the percentage of later-occurring earthquakes (the “after” catalog) that were located within that radius of at least one previous earthquake (Fig. 2). The shaded zones in Figure 2 show the area surrounding the “before” earthquakes, and the filled circles are the “after” earthquakes.¹ The radius is varied so that the shaded circles fill a given percentage of the map area. If a filled circle falls within a shaded zone, we call that a “hit,” and the observed proportion of hits is called \hat{p} in this paper (see notation below under Statistical Analysis of Percentages of Hits). In the hypothetical case shown in Figure 2, six of the eight filled circles fall within the shaded zones, so \hat{p} is 75%. Although this is a rather simple method of characterizing the relationship between past earthquake seismicity and locations of future earthquakes, we tried more complex approaches (including Gaussian smoothing, following the method of Frankel, 1995) and found the results to be quite similar to what we obtained using this simpler cellular method (Kafka and Levin, 2000). Thus, we have adopted the cellular seismology method as a simple and straightforward way of measuring the tendency for past seismicity to delineate zones where future earthquakes are likely to occur.

Figure 3 shows examples of the application of the cellular seismology method to the northeastern United States and Southern California, with radii of circles around the “before” epicenters chosen to fill 33% of the map area. I use the notation “M” in Figure 3, and throughout this paper, to represent magnitude as it was reported in the various earthquake catalogs used in this study. As in Figure 2, the shaded areas in Figure 3 indicate the portions of the map that are near the “before” epicenters, and filled circles are the “after” epicenters. I varied the magnitude cutoffs for “before” and “after” earthquake catalogs so that I could analyze forecasts of earthquakes in the various regions for comparable numbers of events and comparable peri-

¹In this paper, the terms “before” and “after” catalogs refer to the *samples* analyzed, and “past” and “future” earthquakes refer to the *populations* of past and future earthquakes that those samples were selected from.

“Cellular Seismology” Method

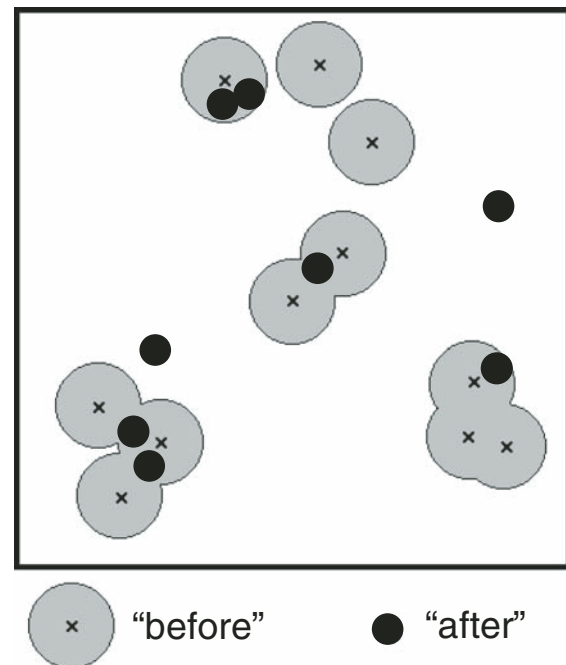


Figure 2. Illustration of the methodology used in this study to measure the extent to which future earthquakes tend to occur near past earthquakes.

ods of time. The values of \hat{p} for the northeastern United States and for Southern California are 78% and 79%, respectively.

As another example of the cellular method, Figure 4 shows the application of this method to the entire central and eastern United States for “before” earthquakes, chosen to be events between 1924 and 1987 ($M \geq 3.0$), and “after” earthquakes, chosen to be events between 1988 and 2003 ($M \geq 4.5$). For this case, with radii of the circles around the “before” epicenters again chosen to fill 33% of the map area, there are 90% hits.

The eventual goal of this type of study is to forecast the locations (if not the times) of future damaging earthquakes. For the U.S. National Seismic Hazard Maps, a minimum magnitude of 5.0 was used in the hazard calculations for the central and eastern United States (Frankel, et al., 1996) because that was considered to be the threshold for an earthquake to cause significant damage. While it would be ideal to limit this study to analyzing “after” earthquakes of about magnitude 5.0 and greater, unfortunately such an approach would mean that I would only be able to analyze very small samples (therefore making any statistical analysis very difficult to carry out). The approach taken in this study is to choose magnitude thresholds for seismicity catalogs based on the data available and the completeness of the catalogs at lower magnitudes. For the “before” catalogs, I chose magnitude thresholds such that I was reasonably confident that the catalogs were complete. For the “after” catalogs, I chose magnitude thresholds such that

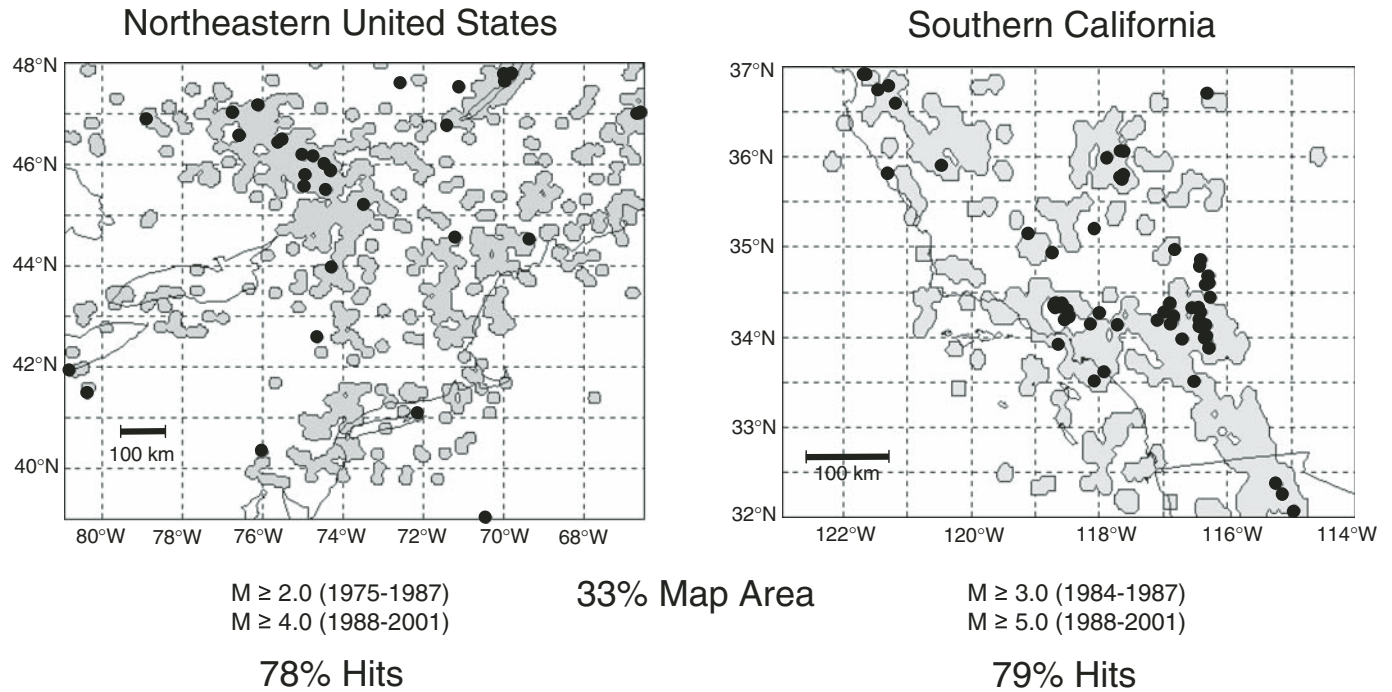


Figure 3. Examples of the application of the “cellular seismology” method (described in the text) to the northeastern United States (left) and Southern California (right), adapted from Kafka (2002). Shaded areas are zones surrounding the “before” catalog, and filled circles indicate epicenters of the “after” catalog. For the northeastern United States: “before” corresponds to $M \geq 2.0$, 1975–1987, and a radius of 15.5 km, and “after” corresponds to $M \geq 4.0$, 1988–2001. For Southern California: “before” corresponds to $M \geq 3.0$, 1984–1987, and a radius of 13.2 km, and “after” corresponds to $M \geq 5.0$, 1988–2001.

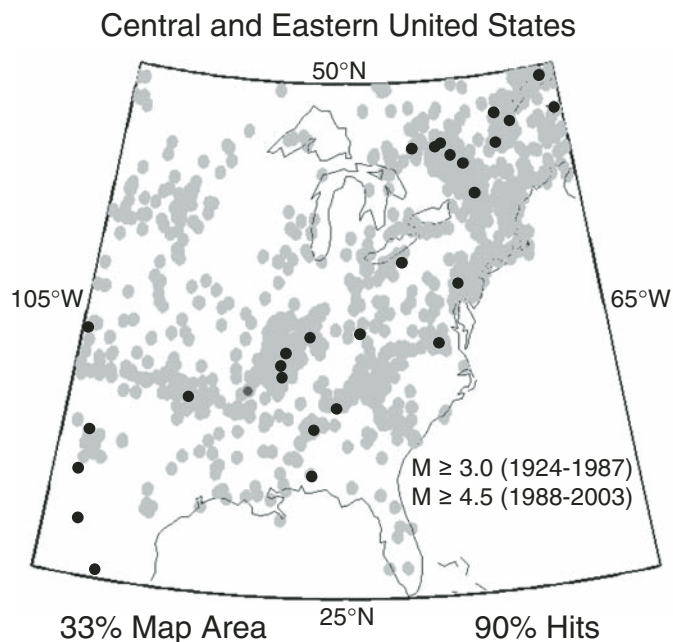


Figure 4. Example of the application of the cellular seismology method to the entire central and eastern United States. Shaded areas are zones surrounding the “before” catalog, and filled circles indicate epicenters of the “after” catalog. For this case, “before” corresponds to $M \geq 3.0$, 1924–1987, and a radius of 36.0 km, and “after” corresponds to $M \geq 4.5$, 1988–2003.

the future earthquakes were as large as possible, while still having enough of them to study the values of $\hat{\rho}$ statistically. These magnitude thresholds were then different for different regions analyzed. While this is not an ideal way to investigate the problem, it is an attempt to use the maximum amount of data available. In order to obtain at least a preliminary assessment of the extent to which these results apply to forecasting locations of truly large and damaging earthquakes, I include here an analysis of the effect of magnitude threshold used for the “after” catalogs, and, as will be seen, I found no evidence of an effect of the choice of this magnitude threshold on the results of this study.

The examples shown in Figures 3 and 4 illustrate how this method provides a measure of the tendency for past seismicity to delineate epicenters of future earthquakes. However, there is a fundamental problem in statistical hypothesis testing of these types of results (as well as for seismicity studies in general). Whereas hypotheses should be tested on multiple independent data sets, in earthquake studies of this type, we often have only one observed data set: the observed record of seismicity. (For additional discussion of issues related to hypothesis testing in earthquake studies, see Rhoades and Evison, 1989.) Lacking such multiple independent data sets, I make a “logical leap” of using earthquake catalogs in different regions as a proxy for repeated samples from the central and eastern United States. I then compare how well seismicity “retrodicts” earthquakes in

the central and eastern United States with how well seismicity retrodicts earthquakes in other regions (including a variety of tectonic environments).

To make these proxy data sets in various regions as comparable as possible, the essence of the distribution of seismicity must be captured in such a way that the measure of the distribution of seismicity is as independent as possible of the size and shape of the region being investigated. For this purpose, I use the percentage of map area surrounding the past earthquakes (P) as a parameter to characterize the distribution of seismicity for a given region. One might envision that the radius of the circles surrounding the epicenters might be a better (and more fundamental) variable for this purpose because it is more directly related to the physics of the earthquake process. I have found, however, that the value of $\hat{\rho}$ for a given radius is more affected by the characteristics of seismicity specific to a given region than is the value of $\hat{\rho}$ for a given percentage of map area. Thus, I consider the percentage of map area to be a more useful parameter than radius for the purpose of this study, i.e., for comparing the extent to which seismicity delineates zones of future earthquakes from one region to the next. As an illustration, Figure 5 shows the value of $\hat{\rho}$ as a function of radius and P for all regions analyzed by Kafka (2002). Notice that choosing percentage of area rather than radius results in a smaller spread of values of $\hat{\rho}$ for a given mean value of $\hat{\rho}$. Thus, it appears that $\hat{\rho}(P)$ captures the extent to which seismicity delineates zones of future earthquakes in such a way that it minimizes the effect

of the size and shape of the specific region investigated and emphasizes the phenomenon of interest itself, thus justifying the logical leap of using many regions as a proxy for many realizations of earthquake catalogs for the same region.

Figure 6 shows results for the various regions analyzed by Kafka (2002). When the radius is chosen such that 33% of the map area is filled, the average value of $\hat{\rho}$ for all of these data lumped together is 74%. The observed values of $\hat{\rho}$ range from 60% for the southeastern United States to 91% for the entire central and eastern United States. These results show no striking pattern of systematic differences in values of $\hat{\rho}$ for intraplate versus plate-boundary regions. Given the more spatially concentrated seismicity in plate-boundary regions, I had expected (intuitively) to observe a greater tendency for past seismicity to delineate zones of future earthquakes in interplate regions than in plate interiors, but no such pattern was observed.

A simple, straightforward observation from Figure 6 is that the percentage of hits exceeds the percentage of map area in all cases. Although not a surprising result, this provides a measure of the fact that, for these cases, future earthquakes are likely to be more highly concentrated in the vicinity of past earthquakes than would be expected for a random distribution of future earthquakes (i.e., values of $\hat{\rho}$ are greater than P). Without such an explicit comparison of the observations with that expected for a random distribution, there is no empirical basis for arguing that future earthquakes tend to occur in the vicinity of past earthquakes. One could imagine that for some regions $\hat{\rho}$ would be below the line

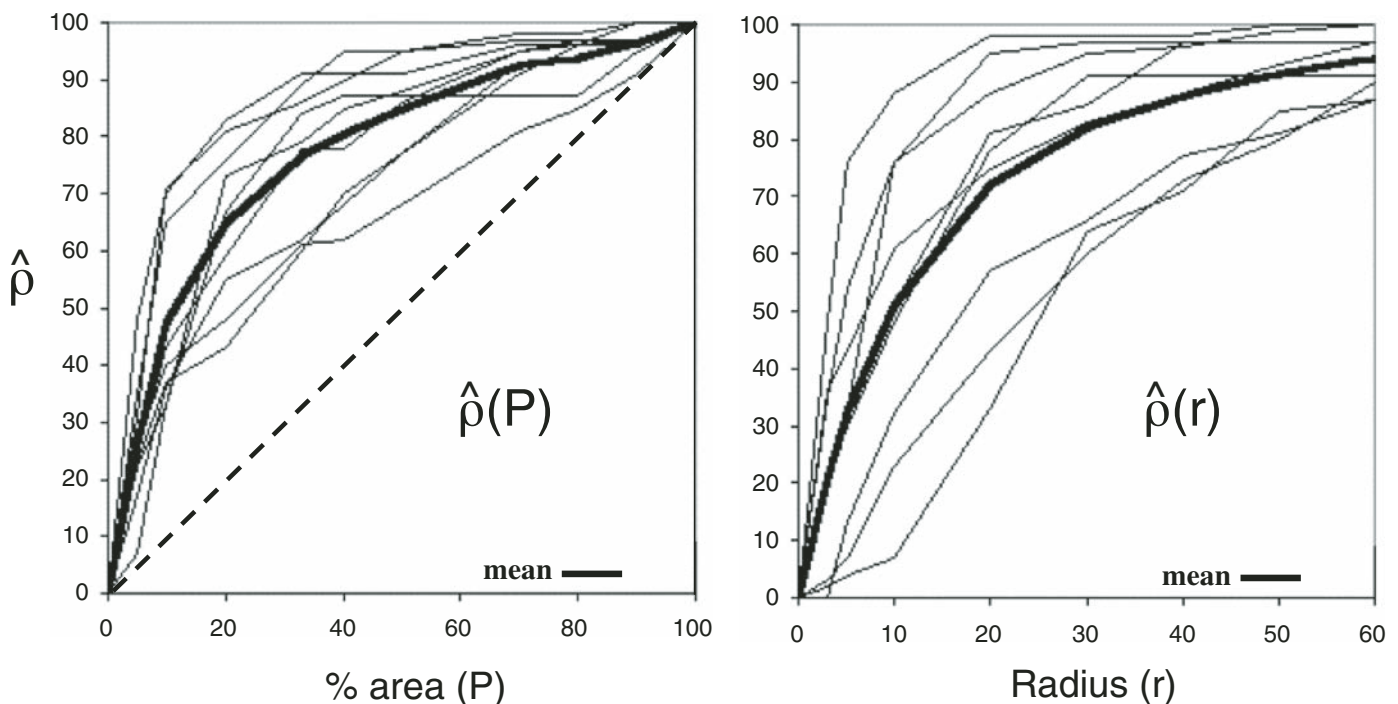


Figure 5. Percentage of hits as a function of percentage of map area and radius (km) for the regions studied by (and adapted from) Kafka (2002). Thin lines denote the individual regions, and thick lines indicate the mean for all regions. Dashed line indicates the locus of points where $\hat{\rho} = P$, corresponding to hypothesis H3.

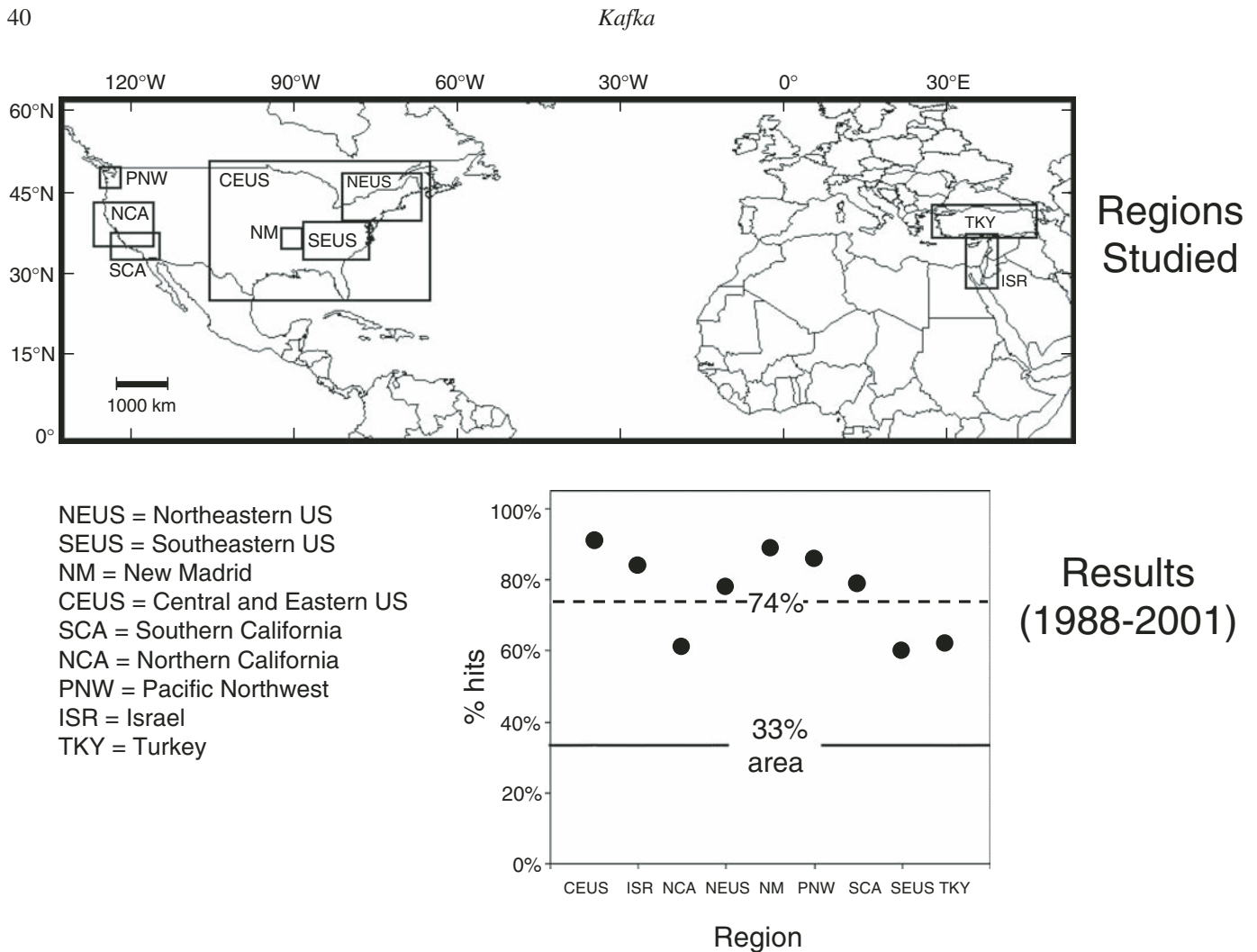


Figure 6. Summary of results for the nine regions analyzed by (and adapted from) Kafka (2002). Radii around the “before” catalog epicenters were chosen such that 33% of the map area was covered. The radii corresponding to 33% map area are: northeastern United States (15.5 km), southeastern United States (31.0 km), New Madrid (13.5 km), central and eastern United States (36.0 km), Southern California (13.2 km), Northern California (10.0 km), Pacific Northwest (9.2 km), Israel (48.0 km), and Turkey (26.0).

representing the 33% map area, which would support hypothesis H2 of Figure 1 for that area. The fact that no such case has been found is encouraging for those who would like to use seismicity as a basis for seismic hazard mapping.

Beyond this simple, straightforward conclusion, however, is the question of how much we can glean from such analyses. To what extent, for example, does the value of $\hat{\rho}$ (for a given P) vary from one region to the next? To what extent does it vary with different magnitude cutoffs for “before” and “after” earthquake catalogs? To what extent does it vary with time after the end of the “before” earthquake catalog? In statistical terms: we can envision an underlying distribution of values of $\hat{\rho}$ for a given value of P , and then attempt to discern how that distribution varies for different regions, different magnitude cutoffs, and different time periods after the end of the “before” earthquake catalog. In the next two sections of this paper, I cast these questions in statistical terms and use that formulation to investigate these variations.

STATISTICAL ANALYSIS OF PERCENTAGES OF HITS

The following statistical formulation of the problem is used for this study. P is the percentage of map area covered by circles of a given radius surrounding “before” earthquake epicenters, $\hat{\rho}$ is the observed percentage of “after” earthquakes that occur “near” (i.e., within that given radius of) at least one of the “before” earthquakes, and ρ is the underlying probability that a future earthquake will occur “near” at least one of the past earthquakes. Our objective, then, is to estimate ρ based on observations of $\hat{\rho}$.

Consider the range of possible distributions of $\hat{\rho}(P)$. At one extreme, we could imagine that $\hat{\rho}$ has equal probability of having any value between 0 and 1, implying that there is no information content in the distribution of past seismicity (or the choice of P) that is relevant to where future earthquakes are likely to occur. At the other extreme, we could imagine that the nature of earthquake processes (and the choice of P) is such that (regardless of the cho-

sen region) there is a very narrow distribution of possible values of $\hat{\rho}$, once the value of P is chosen. Between these two extremes, we can imagine that there is some characteristic (presumably region-dependent) shape to the distribution of $\hat{\rho}$, which provides information about the probability that future earthquakes will be located near past earthquakes for a given value of P .

Figure 7 shows the distribution of values of $\hat{\rho}$ from data analyzed by Kafka (2002), with P chosen to be 33%. The fact that the observed distribution of $\hat{\rho}$ lies consistently to the right of P for all cases studied is evidence that future earthquakes are more likely to occur near past earthquakes than where previous earthquakes have not occurred (regardless of the specific region investigated). If the distribution lay far to the right, it would be evidence that future earthquakes are highly concentrated in zones of past seismicity. A high variance in the distribution would indicate great variation from one region to the next (and presumably one tectonic environment to the next) in the probability of future earthquakes occurring near past ones.

Based on this type of analysis, I conducted systematic hypothesis tests to obtain estimates of $\rho(P)$ based on a given data set of values of $\hat{\rho}$. I treated $\hat{\rho}$ as a random variable, and estimated the probability distribution of $\hat{\rho}$. The process of examining whether a future large earthquake occurs within a given radius of past ones is thus modeled as a binomial experiment with success defined as the event that the future earthquake (filled circle in Fig. 2) occurs within that radius of at least one past earthquake (shaded zones in Fig. 2).

Using this formulation, Kafka (2002) estimated that, at the 95% level of statistical significance, more than 71% of the earthquakes in a region will tend to occur near previous epicenters

(defined by the 33% map area contours). Furthermore, the results of that study suggested that there was no statistically significant difference in the values of $\hat{\rho}$ for intraplate versus plate-boundary regions. While these results are a good first step toward providing statistical support for the notion that past seismicity delineates where future earthquakes are likely to occur, an obvious problem with these studies to date is that they have been based on “convenience sampling” (i.e., choosing data samples because they happen to be available or relatively easy to obtain). The regions studied were chosen primarily because they are areas where high-quality seismic monitoring resulted in earthquake catalogs that were readily available for analysis. Thus, the data were not random samples of seismicity from various tectonic environments, so what seems like statistical significance might be highly biased and merely a coincidence of the particular data sets analyzed. One would instead want to apply these methods to data sets that are more randomly and objectively chosen from different regions and different periods of time.

Another problem with our previous studies was that the sizes and shapes of the regions varied greatly (Fig. 6), and thus the characteristic of “nearness” to previous earthquakes was different for the different regions. Furthermore, in our previous studies the number of earthquakes in the catalogs was not large enough to, in any systematic way, have more than one statistical sample of “future” large earthquakes for a given region. Thus, we didn’t really have a basis from which to effectively investigate what the distribution of $\hat{\rho}$ might look like for a given region (i.e., for a given tectonic environment).

The ideal data sets to test these hypotheses are impossible to obtain without another century or more of monitoring. However, the three decades of global earthquake data available from the U.S. Geological Survey (USGS) National Earthquake Information Center (NEIC, Fig. 8) provide an opportunity to test such hypotheses in a more systematic and objective manner than in our previous studies.

IN SEARCH OF A MORE SYSTEMATIC TEST OF H1

Using this NEIC database, the statistical variation of the values of $\hat{\rho}$ was explored by dividing the world into eight subregions (labeled R11 through R24), all with the same shape and same surface area (Fig. 8). The cellular seismology method was then applied uniformly to all of these subregions. In an effort to simulate a series of realizations of hypothetical cases like the situation we face in the central and eastern United States, I envision this as a “thought experiment” in which we have eight regions where earthquakes are occurring, and the earthquakes are caused by some tectonic process (or processes) which I will treat here as unknown. Although we actually are aware of the tectonic processes in these regions, we can still think of them in this manner because the boundaries were chosen independently of tectonic environment. Continuing with this thought experiment, we envision trying to discern from a catalog of earthquakes alone the extent to which past seismicity can be used to delineate zones

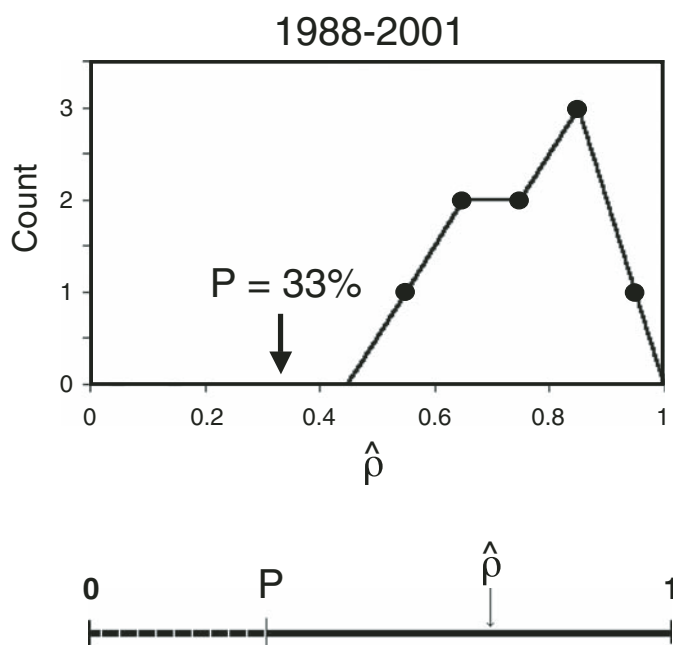


Figure 7. Distribution of $\hat{\rho}$ based on the data analyzed by (and adapted from) Kafka (2002). The $\hat{\rho}$ values shown here are calculated for $P = 33\%$.

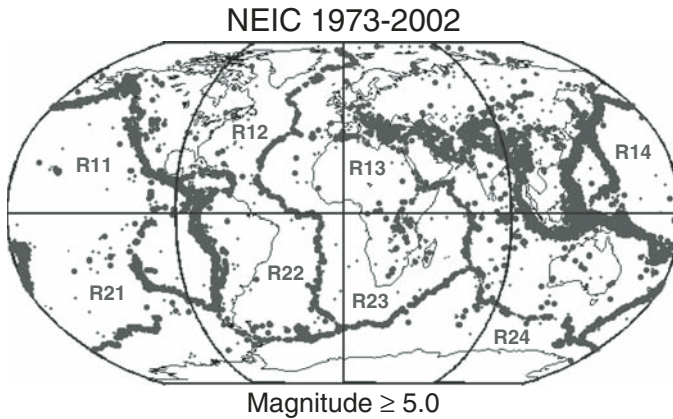


Figure 8. Filled circles indicate shallow (≤ 70 km) earthquakes $M \geq 5.0$ recorded by the National Earthquake Information Center (NEIC) from 1973 to 2002. Size of circles is proportional to magnitude (ranging from 5.0 to 8.8). Also shown are the eight subregions (R11 to R24) of equal surface area and shape for which the cellular seismology method was applied as described in the text.

where future large earthquakes are likely to occur. (After this “blind” statistical analysis, I will later speculate on how tectonic environment within these regions actually does contribute to differences in the distributions of $\hat{\rho}$.)

Using the NEIC data shown in Figure 8 (with a lower magnitude cutoff of 5.0), there is a sufficiently large number of events (from a complete earthquake catalog) that the cellular seismology method can be applied to many independent “two-year-before catalogs” forecasting “two-year-after catalogs” (Figs. 9 and 10). Figure 9 shows three examples (regions R14, R22, and R24) where two years of $M \geq 5.0$ earthquakes represent past seismicity, two years of $M \geq 5.5$ earthquakes represent future earthquakes, and $P = 10\%$ of map area. For these cases, the values of $\hat{\rho}$ are: 92% for R14, 84% for R22, and 70% for R24. Given the length of the NEIC catalog, I was able to analyze the data in the manner illustrated in Figure 9 for 14 pairs of two-year-before/two-year-after catalogs for each of the eight regions (Fig. 10).

For simplicity, I use the notation “ M_x+ ” to represent earthquakes of $M \geq x$. The top graph in Figure 10 shows results for 14 cases of two years of $M5.0+$ earthquakes representing the past, and two years of $M6.0+$ earthquakes representing the future for the eight subregions shown in Figure 8. The bottom graph shows the same $M5.0+$ “before” earthquake catalogs forecasting locations of $M5.5+$ earthquakes. For the top graph, the number of “after” earthquakes for the 14 cases ranged from 1 to 109, with a mean of 26 and a standard deviation of 22. For the bottom graph the number of “after” earthquakes ranged from 11 to 312, with a mean of 87 and a standard deviation of 66. Thus, the values of $\hat{\rho}$ shown for each two-year sample for the $M5.5+$ plot were based on large enough samples to be considered as statistical measures of the underlying value of ρ expected for that region. The results for the $M5.5+$ case are quite similar to those of the $M6.0+$ case, except that the variances for the $M5.5+$ case tended to be lower

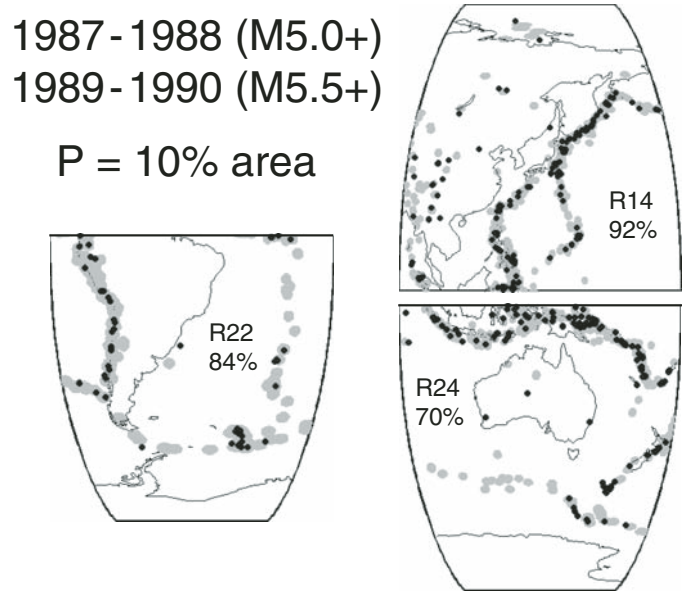


Figure 9. Examples of the application of the cellular seismology method to the subregions shown in Figure 8. Gray shading shows zones surrounding the “before” catalog, and filled black circles indicate epicenters of the “after” catalog. For this case, “before” corresponds to $M \geq 5.0$, 1987–1988, and “after” corresponds to $M \geq 5.5$, 1989–1990. Radii around the “before” catalog epicenters were chosen such that 10% of the map area was covered. The radii corresponding to 10% map area are: R22 (168 km), R14 (109 km), and R24 (117 km).

than for the $M6.0+$ case. This observation suggests that the $M5.5+$ case is measuring the same phenomenon as the $M6.0+$ case, but is just based on larger sample sizes. Figure 11 shows the mean values for the $M6.0+$ case for each region.

We can see in Figures 10 and 11 that there is a characteristic mean value of $\hat{\rho}$ for a given subregion, as well as a characteristic variance. For example, region R11 has a high mean (93% for $M6.0+$ forecasts) and a low variance, while region R23 has a low mean (49%) and a high variance. Below, in the Discussion and Conclusions section, I speculate on possible relationships between the mean and variance of the values of $\hat{\rho}$ and the tectonic environments in the subregions analyzed.

As in the cases analyzed by Kafka (2002), for these eight subregions, the values of $\hat{\rho}$ for a given radius are more affected by the characteristics of the specific region than are the values of $\hat{\rho}$ for a given percentage of map area (Fig. 12). Again, choosing percentage of area rather than radius results in a lower range of values of $\hat{\rho}$ for a given mean value of $\hat{\rho}$. This effect is not as pronounced here as in the cases analyzed by Kafka (2002), and I suspect that this is due to the fact that the eight subregions are all the same size and shape, so that radius and percentage of area are more directly related to each other in this situation. In additional cases investigated next, where I analyze regions with specific tectonic environments and with significant differences in the shape and size of the regions, this effect is again more clearly observed (Figs. 13 and 14).

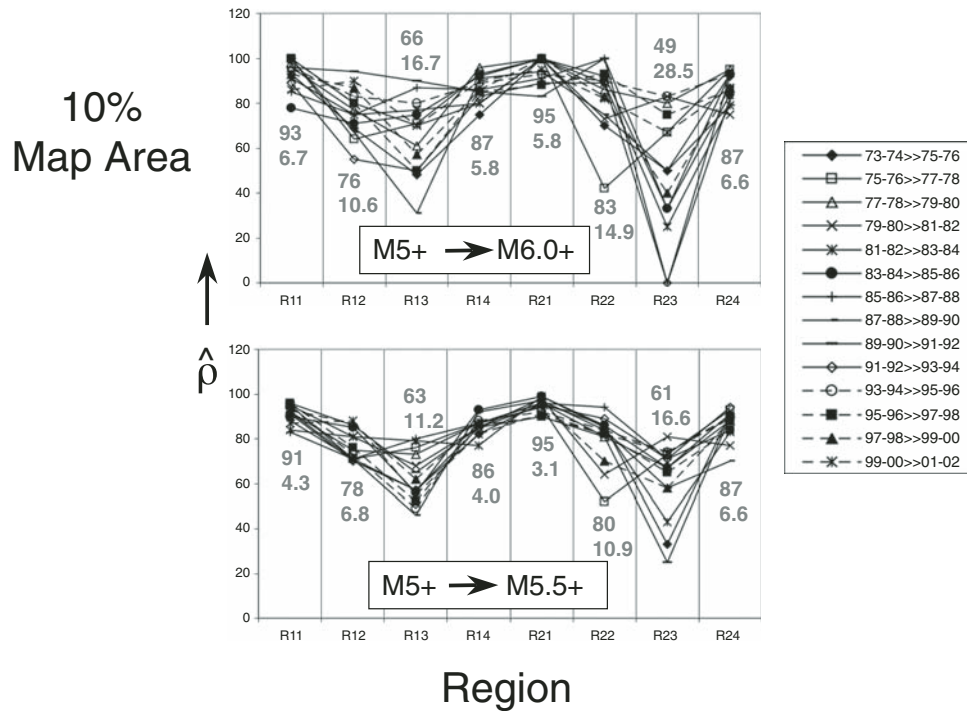


Figure 10. Results of the application of the cellular seismology method to the eight subregions shown in Figure 8. Numbers shown in gray are the mean (above) and the standard deviation (below) of $\hat{\rho}$ for each region.

While the analysis of the eight global subregions described here has the advantages of being objective and being based on regions of the same shape and size, a disadvantage of this analysis is that each region arbitrarily includes a variety of tectonic environments. In an effort to discern more specific effects of tectonic environment on the distribution of $\hat{\rho}$, I applied the cellular seismology method to four additional subregions of Earth, each representing a different tectonic environment (Figs. 13 and 14). The tectonic environments chosen for these analyses were: a subduction zone (labeled Subduction in Fig. 13), an oceanic spread-

ing center (labeled Ridge), a region of active continental collision (labeled Continent), and the interior of the North American plate (labeled INAP).

In the Ridge and Subduction regions, the seismicity is (not surprisingly) densely concentrated near the plate boundary. Future earthquakes in these regions have a high probability of occurring near past seismicity, with $\hat{\rho}$ (for $P = 33\%$) equal to 98% for Ridge and 100% for Subduction. The INAP and Continent regions, by contrast, have (also not surprisingly) more diffusely distributed seismicity, and future earthquakes in these regions appear to have

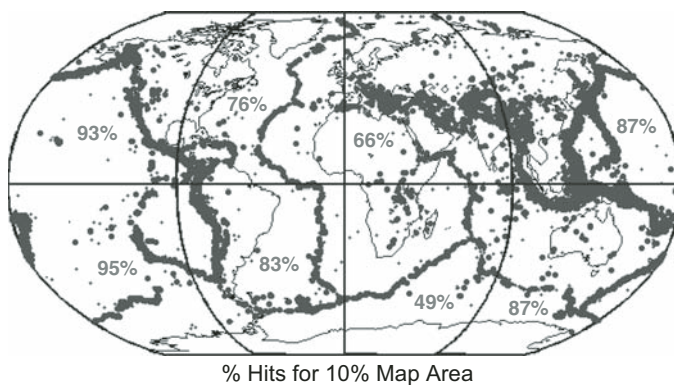


Figure 11. Mean values of $\hat{\rho}$ from the analyses shown in Figure 10 (M6.0+ case).

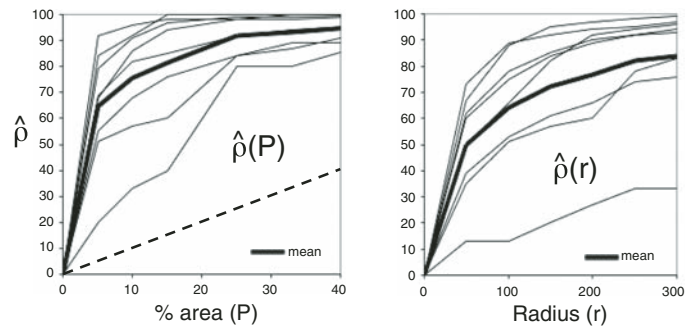


Figure 12. Values of $\hat{\rho}$ as a function of percentage of map area and radius (km) for the eight subregions shown in Figure 8. Thin lines denote the individual regions, and thick lines indicate the mean for all regions. Dashed line indicates the locus of points where $\hat{\rho} = P$, corresponding to hypothesis H3.

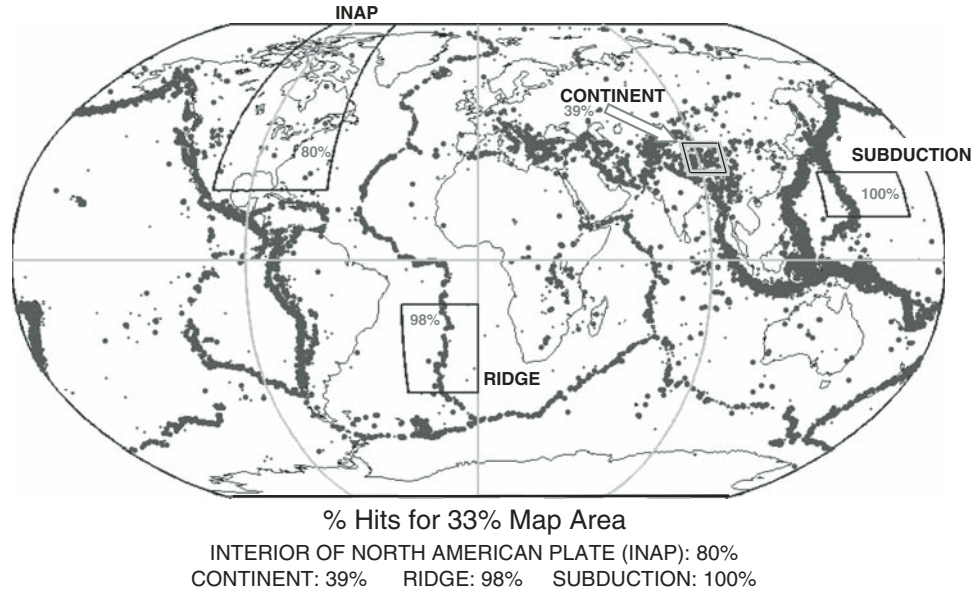


Figure 13. Percentage of hits for 33% map area calculated for regions with different tectonic environments (M5.0+, 1973–1987 forecasting M5.5+, 1988–2002). The radii corresponding to 33% map area are: interior of the North American plate (INAP; 420 km), Continent (63 km), Ridge (320 km), and Subduction (330 km). Percentages of hits for other values of percentage of map area, and for variation in radius, are shown in Figure 14.

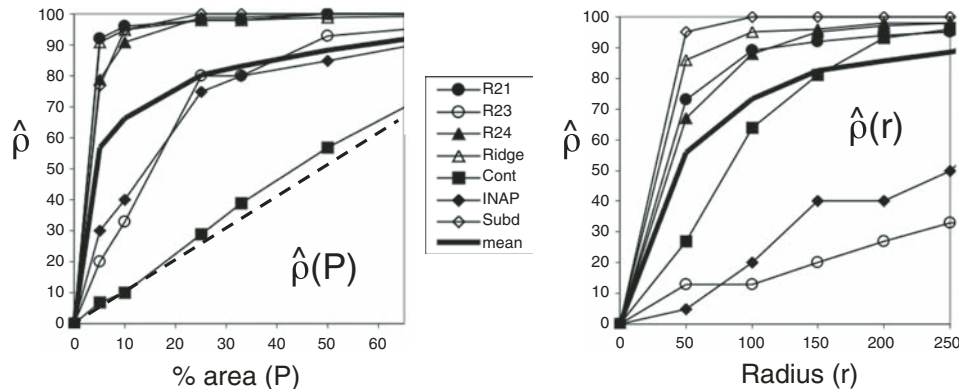


Figure 14. Percentage of hits as a function of percentage of map area and radius (km) for the four tectonic regions shown in Figure 13 (M5.0+, 1973–1987 forecasting M5.5+, 1988–2002), and for regions R21, R23, and R24 of Figure 8 (M5.0+, 1973–1974 forecasting M5.5+, 1975–1976). Thin lines denote the individual regions, and thick black lines indicate the mean for all regions. Dashed line indicates the locus of points where $\hat{p} = P$, corresponding to hypothesis H3. INAP—interior of North American plate, Cont—Continent, and Subd—Subduction.

lower probabilities of occurring near past seismicity, with \hat{p} (for $P = 33\%$) equal to 80% for INAP and 39% for Continent. More specific aspects of these results do not have obvious interpretations. For example, why is the value of \hat{p} for the Continent region so low (39%), and how might this low value of \hat{p} be related to that of R13, which has a relatively low \hat{p} for 10% map area and a high variance (see Figs. 10 and 11)? Also, when we compare these results for the Continent region with results for the central and eastern United States, we find that the central and eastern United

States has surprisingly high values of \hat{p} . For the central and eastern United States case shown in Figure 4, \hat{p} is 90% for 33% map area, and \hat{p} is 57% for 10% map area.

In an effort to investigate the extent to which these results apply to forecasting locations of truly large and damaging earthquakes, I studied the effect of the minimum magnitude threshold used for the “after” catalogs in these analyses. For nine cases, the minimum magnitude cutoff for “after” earthquakes, $M(\min)$, was varied systematically to investigate the effect on

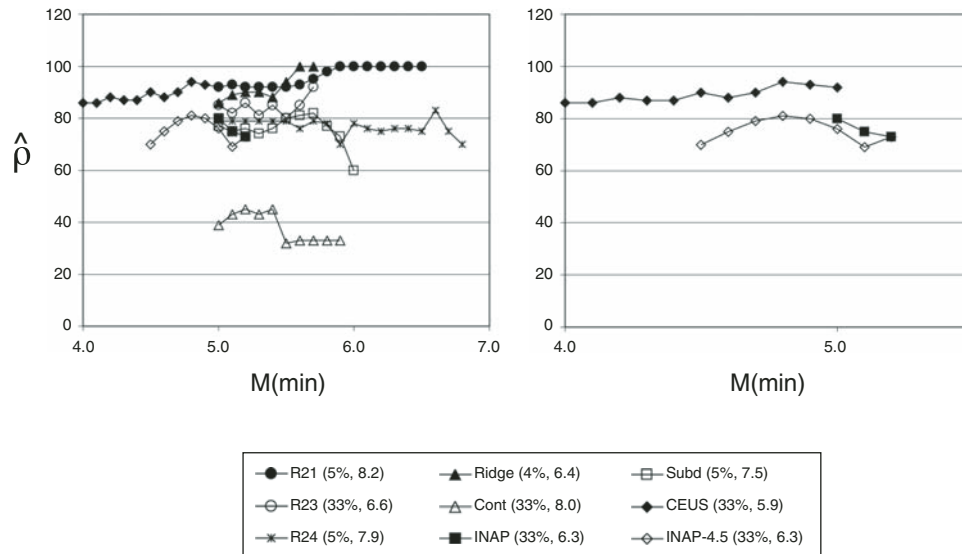


Figure 15. $\hat{\rho}$ as a function of minimum magnitude cutoff for the various regions analyzed in this study. Numbers in parentheses are percentage of map area corresponding to the “before” earthquake catalog and magnitude of largest “after” earthquake for a given region. For this analysis, it was necessary to vary the percentage of map area for the different regions so that the initial value of $\hat{\rho}$ was low enough that the variation in $\hat{\rho}$ could be seen as $M(\min)$ increased. INAP—interior of North American plate, Cont—Continent, Subd—Subduction, and CEUS—central and eastern United States.

$\hat{\rho}$ as $M(\min)$ increases (Fig. 15). For each region, $M(\min)$ was increased by 0.1 unit intervals up to the point where there were at least 10 “after” earthquakes remaining to be analyzed. The minimum magnitude for the “before” earthquake catalogs was 5.0 for all cases shown in Figure 15, except for central and eastern United States (where it was 3.0). Since the INAP region had much fewer earthquakes than the other regions, there are two cases shown for INAP in Figure 15, one in which the “before” and “after” catalogs start at magnitude 5.0, and the other in which they start at magnitude 4.5.

As a simple test of whether there is any systematic effect on $\hat{\rho}$, I counted the number of times that $\hat{\rho}$ increased when $M(\min)$ was changed to $M(\min) + 0.1$, the number of times it remained the same, and the number of times it decreased. For example, in the case of central and eastern United States, $\hat{\rho}$ is 86% for $M(\min) = 4.0$, remains the same (86%) for $M(\min) = 4.1$, and then increases to 88% for $M(\min) = 4.2$. If there was, for example, a systematic decrease in $\hat{\rho}$ with increasing $M(\min)$, we would expect there to be significantly more decreases than increases. The results were: 30 times for “increase,” 27 times for “decrease,” and 21 times for “remained the same.” The similarity between numbers of increases and numbers of decreases suggests that there is no observed effect of $M(\min)$ on $\hat{\rho}$. Using this result as a guide, it appears that the results of this study are independent of $M(\min)$, and therefore apply to “large and damaging” earthquakes in the central and eastern United States.

Although somewhat beyond the scope of this study, this investigation would not be complete without at least some mention of the question of the time dependence of the values of

$\hat{\rho}$ for a given region. One would intuitively imagine that the values of $\hat{\rho}$ would tend to decrease for forecasting locations of earthquakes farther into the future. Figure 16 shows the distribution of $\hat{\rho}$ for 10% map area for the eight global regions analyzed ($M5.0+$ seismicity forecasting locations of $M5.5+$ earthquakes). The “before” catalog is for 1973–1974, and Figure 16 shows the distribution of values of $\hat{\rho}$ for “after” earthquakes occurring in 1975 and 1976, as well as in 2001 and 2002. What is most striking about the results shown in this figure is the lack of evidence for systematically lower values of $\hat{\rho}$ for the 2001–2002 forecasts than for the 1975–1976 forecasts (even though the 2001–2002 catalog starts 26 yr after the end of the “before” seismicity catalog). For 1975–1976, the mean value of $\hat{\rho}$ is 75.6 and the median is 84, while for 2001–2002, the mean is 79.6 and the median is 82. Thus, at least on the time scale of a few decades, there does not seem to be any evidence of a decrease in the tendency for past seismicity to delineate zones where future earthquakes are likely to occur.

As an additional test of the time dependence of this phenomenon, there were sufficient numbers of $M4.0+$ earthquakes in the central and eastern United States catalog to divide the results for observed values of $\hat{\rho}$ in the central and eastern United States from 1988 to 2003 into six subsamples of 15 earthquakes each. Each one of the six subsamples is ordered in Figure 17 such that they consist of sequentially later events in time, and for each subsample, $\hat{\rho}$ was calculated. Again as in the analysis described in the previous paragraph, there was no evidence of any decrease in the values of $\hat{\rho}$ as time increased after the end of the past earthquake seismicity catalog. While these issues need to be analyzed

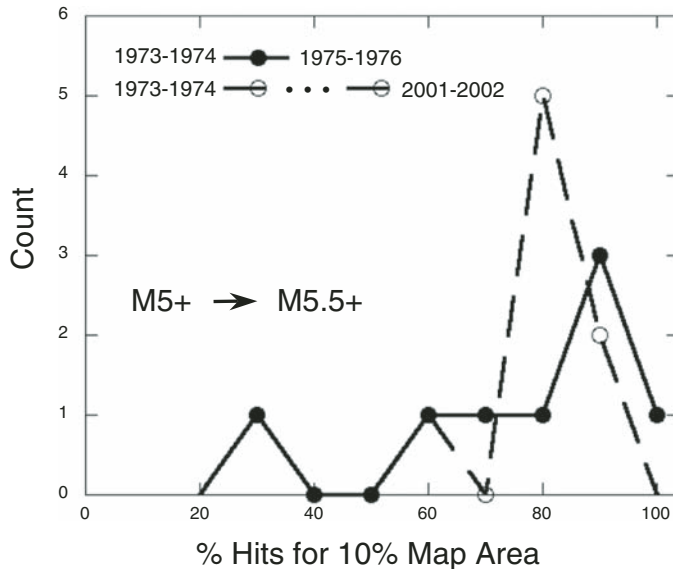


Figure 16. Distribution of percentage of hits for 10% map area for the eight regions shown in Figure 8 (M5.0+ seismicity forecasting locations of M5.5+ earthquakes). Filled circles represent results for 1973–1974 seismicity forecasting locations of earthquakes occurring between 1975 and 1976 (mean percentage of hits = 75.6, median percentage of hits = 84). Open circles are for 1973–1974 seismicity forecasting locations of earthquakes occurring between 2001 and 2002 (mean = 79.6, median = 82).

in greater detail to make any strong conclusions about the time dependence of this phenomenon, I see no evidence of a decrease in the values of $\hat{\rho}$ as time increases.

ESTIMATING ρ FOR THE CENTRAL AND EASTERN UNITED STATES

Having explored the statistical variation of the values of $\hat{\rho}$ for a variety of tectonic environments, magnitude thresholds, and sizes and shapes of regions, it seems clear that the tendency for past seismicity to delineate zones where future large earthquakes are likely to occur is a real, measurable, physical phenomenon. Furthermore, the time scale over which this phenomenon varies appears to be such that it is possible to obtain representative samples of this phenomenon from seismicity maps and to use those samples to estimate the probability of future earthquakes occurring near past earthquakes in a region of interest. Based on these results, I can thus estimate the probability of a future large earthquake occurring in zones delineated by past seismicity in the central and eastern United States, as well as a confidence interval for that estimate.

Let $\rho(\text{CEUS}, P)$ represent the probability that a future large earthquake in the central and eastern United States will occur within the zones defined by $P\%$ map area, defined as discussed already using the cellular seismology method. Based on M4.0+ earthquakes occurring between 1988 and 2003 (with $P = 33\%$),

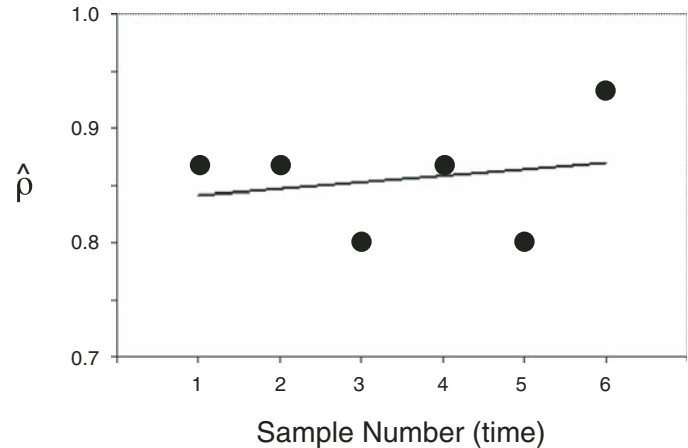


Figure 17. M4.0+ earthquakes in the central and eastern United States catalog (1988–2003) divided into six subsamples of 15 earthquakes each. The six subsamples are ordered such that they consist of sequentially later events in time, and for each subsample, $\hat{\rho}$ was calculated. The solid line is a least squares fit to the data.

I have a large enough sample to apply methods of statistical inference, and we find $\hat{\rho} = 0.86$. Thus, I form a 95% confidence interval, as follows (e.g., Weiss and Hassett, 1982):

$$\rho(\text{CEUS}, 0.33) = \hat{\rho} \pm 1.96 \sqrt{\frac{\hat{\rho}(1-\hat{\rho})}{n}}, \quad (1)$$

where $n = 91$ is the number of “after” earthquakes. The 95% confidence interval for this case is $0.79 \leq \rho(\text{CEUS}, 0.33) \leq 0.93$, and I estimate that the probability of a given future earthquake occurring in the light shaded zones in Figure 18 is 0.86 ± 0.072 . For $P = 10\%$ map area (dark shaded zones in Fig. 18), I find that $\hat{\rho} = 0.60 \pm 0.100$, and the 95% confidence interval is $0.50 \leq \rho(\text{CEUS}, 0.10) \leq 0.70$.

This estimate of ρ for the central and eastern United States is not inordinately low compared to observed values of $\hat{\rho}$ for regions around the world, but it falls within the lower end of the range of values (Figs. 10 and 13). Stated in terms of radius surrounding the epicenters of past earthquakes, these results suggest that future earthquakes in the central and eastern United States have $\sim 86\%$ probability of occurring within 36 km of past earthquakes, and $\sim 60\%$ probability of occurring within 14 km of past earthquakes.

DISCUSSION AND CONCLUSIONS

Given our less-than-complete understanding of the cause of earthquakes in the central and eastern United States, seismic hazard analysis for this region (and probably most intraplate regions) will likely continue to depend, to a large extent, on seismicity for delineating locations of future large earthquakes. Thus, it is important to understand the scientific basis underlying the tendency for seismicity to delineate zones where future large earthquakes are likely to occur. This will not be an easy task, but if

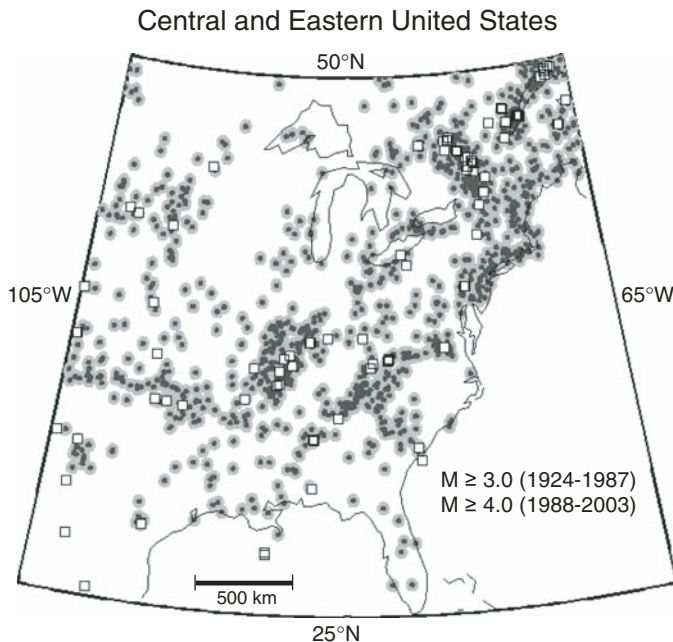


Figure 18. M3.0+ earthquakes in the central and eastern United States (1924–1987) forecasting locations of M4.0+ earthquakes (1988–2003). Lighter shading corresponds to a 36 km radius and 33% map area, and darker shading corresponds to a 14 km radius and 10% map area. M4.0+ earthquake epicenters for 1988–2003 are shown by the open squares.

we do not undertake this task, then one of the major inputs into seismic hazard analysis will be based merely on an untested conceptual model. The analysis presented here supports the notion that the tendency for future large earthquakes to occur in zones delineated by past seismicity is a real, measurable, physical phenomenon that can be investigated scientifically. While this investigation is only at a very early stage of development, it shows some systematic patterns regarding similarities and differences in this tendency for regions characterized by different tectonic processes. The results suggest that using seismicity as an indicator of zones where future large earthquakes are likely to occur is as reasonable a conceptual model for the central and eastern United States as it is for many other parts of the world, including some plate-boundary regions.

The analysis of the eight regional subdivisions of Earth, chosen objectively, shows that a given region, within which some given combination of tectonic processes is occurring, is characterized by a mean value and variance of $\hat{\rho}$. As can be seen in Figures 10 and 11, regions that contain large proportions of subduction zones (such as R11, R14, and R24) have characteristically high mean values of $\hat{\rho}$ and characteristically low variances of $\hat{\rho}$. While this same pattern might have been expected for regions containing large proportions of oceanic spreading centers (such as R23), such regions have characteristically lower mean values of $\hat{\rho}$ and higher variances of $\hat{\rho}$ compared to regions dominated by subduction zones. The variations

in $\hat{\rho}$ seem to be the result of a combination of not only tectonic differences, but also differences in the way in which the NEIC samples the different types of tectonic zones. Although we are just beginning to scratch the surface of how to measure and categorize this phenomenon, it appears that $\hat{\rho}$ is higher when a major subduction zone is included in a region, and it is lower for continental areas and mid-ocean ridges.

The analysis of specific tectonic regions (Fig. 13) supports the (not surprising) result discussed already regarding a strong tendency for future earthquakes to occur near past earthquakes in subduction zones. The very high value of $\hat{\rho}$ for the Ridge region (98%), however, suggests that the relatively low values of $\hat{\rho}$ for region R23 are not merely the result of the high percentage of ridge-type plate boundaries in region R23. Unraveling the effects of tectonic environment on values of $\hat{\rho}$ is, of course, complicated by differences in the sizes and shapes of the regions chosen for analysis (even though use of the percentage of map area does mitigate this problem to some extent).

The value of $\hat{\rho}$ for the Continent region (39% for 33% map area) is among the lowest values of $\hat{\rho}$ for any region analyzed in our studies. In fact, 39% hits for 33% map area is very close to what would be expected for a random spatial distribution of future earthquakes. The reason for this low value is not clear at this point, but it should caution us not to conclude that seismicity will always be a good indicator of where future earthquakes will occur.

The value of $\hat{\rho}$ for the INAP region (80% for 33% map area) is neither particularly high nor particularly low compared to other regions around the world. Thus, while hypothesis H1 is not unequivocally verified for intraplate regions, these results, along with our specific results for the central and eastern United States, suggest that H1 is as reasonable a conceptual model for the central and eastern United States as it is for many parts of the world, including some plate-boundary regions. Our studies, therefore, support the approach of using past seismicity as an indicator of zones where future large earthquakes are likely to occur in the central and eastern United States for the purpose of seismic hazard analysis (at least until the physical cause of earthquakes in this region is better understood).

The analysis of the time dependence of the tendency for future earthquakes to occur in zones delineated by past seismicity yielded no evidence of a decrease in this tendency as time increases after the end of the past seismicity catalog. This result is encouraging for seismic hazard analysis as it suggests that seismic hazard maps developed today may provide meaningful estimates of hazards to facilities that are expected to have very long lifetimes.

The results of these preliminary analyses of time dependence suggest that we are not (yet?) seeing evidence of the “paleoseismicity” model of Ebel et al. (2000). This model hypothesizes that the major earthquakes that have occurred in the central and eastern United States might be long-delayed aftershocks of large early historical or prehistorical earthquakes. If this were the case, we might expect to eventually see a decrease in the number of earthquakes near past historical

earthquakes as time increases, and therefore might expect our ability to forecast locations of future earthquakes based on seismicity to decrease as time increases. The time spans covered by the earthquake catalogs used in this study are probably too short, however, to detect such a pattern. Thus, the results of this study cannot rule out the paleoseismicity model.

We find no evidence to suggest that the tendency for seismicity to delineate zones where future large earthquakes are likely to occur is any “less real” for the central and eastern United States than for any other regions. Given what we have been able to discern from this and our previous studies on this topic, we estimate that the probability of a given future earthquake (including large and damaging earthquakes) occurring in the light shaded zones in Figure 18 is 0.86 ± 0.072 , and the probability of such an earthquake occurring in the dark shaded zones in that figure is 0.60 ± 0.100 .

This study provides evidence that the tendency for seismicity to delineate zones where future large earthquakes are likely to occur is a real, measurable, physical phenomenon. Furthermore, the time scale over which seismicity data are available appears to be representative of the time scale over which this phenomenon occurs. Many years of additional seismicity data will, nonetheless, be required to fully test the ideas discussed in this paper. For this (and many other reasons), there is value in continuing to monitor seismicity in intraplate regions. Continued monitoring of seismicity will provide a better basis for discerning the scientific justification for using seismicity as an indicator of zones where future large earthquakes are likely to occur in the central and eastern United States and elsewhere.

ACKNOWLEDGMENTS

I thank the editors of this volume, Seth Stein and Stéphane Mazzotti, as well as the three anonymous reviewers of this manuscript for their thoughtful critiques, which improved the paper in many ways. This study benefited from many years of valuable

discussions with John Ebel on various aspects of intraplate earthquakes in general and “cellular seismology” in particular. I thank Heather Wagner for assisting me with the data processing for the analysis of the eight subregions of Earth.

REFERENCES CITED

- Cao, T., Petersen, M.D., and Reichle, M.S., 1996, Seismic hazard estimate from background seismicity in Southern California: *Bulletin of the Seismological Society of America*, v. 86, no. 5, p. 1372–1381.
- Dziewonski, A.M., Ekström, G., and Maternovskaya, N.N., 1999, Centroid-moment tensor solutions for April–June 1998: *Physics of the Earth and Planetary Interiors*, v. 112, p. 11–19, doi: 10.1016/S0031-9201(98)00157-5.
- Ebel, J.E., Bonjer, K.P., and Oncescu, M.C., 2000, Paleoseismicity: Seismicity evidence for past large earthquakes: *Seismological Research Letters*, v. 71, no. 2, p. 283–294.
- Frankel, A., 1995, Mapping seismic hazard in the central and eastern United States: *Seismological Research Letters*, v. 66, no. 4, p. 8–21.
- Frankel, A., Mueller, C., Barnhard, T., Perkins, T.D., Leyendecker, E.V., Dickman, N., Hanson, S., and Hopper, M., 1996, *National Seismic Hazard Maps, Documentation, June 1996*: U.S. Geological Survey Open-File Report 96-532, 41 p.
- Jackson, D.D., and Kagan, Y.Y., 1999, Testable earthquake forecasts for 1999: *Seismological Research Letters*, v. 70, no. 4, p. 393–403.
- Kafka, A.L., 2002, Statistical analysis of the hypothesis that seismicity delineates areas where future large earthquakes are likely to occur in the central and eastern United States: *Seismological Research Letters*, v. 73, no. 6, p. 990–1001.
- Kafka, A.L., and Levin, S.Z., 2000, Does the spatial distribution of smaller earthquakes delineate areas where larger earthquakes are likely to occur?: *Bulletin of the Seismological Society of America*, v. 90, no. 3, p. 724–738, doi: 10.1785/0119990017.
- Kafka, A.L., and Walcott, J.R., 1998, How well does the spatial distribution of smaller earthquakes forecast the locations of larger earthquakes in the northeastern United States?: *Seismological Research Letters*, v. 69, no. 5, p. 428–439.
- Rhoades, D.A., and Evison, F.V., 1989, On the reliability of precursors: *Physics of the Earth and Planetary Interiors*, v. 58, p. 137–140, doi: 10.1016/0031-9201(89)90049-6.
- Weiss, N., and Hasset, M., 1982, *Introductory Statistics*: Reading, Massachusetts, Addison-Wesley Publishing Co., 651 p.
- Wheeler, R.L., and Frankel, A., 2000, *Geology in the 1996 USGS Seismic Hazard Maps, central and eastern United States*: *Seismological Research Letters*, v. 71, no. 2, p. 273–282.

MANUSCRIPT ACCEPTED BY THE SOCIETY 29 NOVEMBER 2006