Statistical Analysis of the Hypothesis That Seismicity Delineates Areas Where Future Large Earthquakes Are Likely to Occur in the Central and Eastern United States

Alan L. Kafka
Weston Observatory, Boston College

ABSTRACT

Although large and damaging earthquakes occur in the central and eastern United States (CEUS), no comprehensive and scientifically sound physical models have proven to be reliable indicators of where future large earthquakes are likely to occur in this region. This situation forces seismologists who are attempting to estimate the seismic hazard in CEUS to rely heavily on the observed record of seismicity as an indicator of where future large earthquakes are likely to occur. In this study, the hypothesis that seismicity delineates areas where large earthquakes are likely to occur in CEUS (as well as in other regions) is tested and statistically analyzed. These analyses are then used as a basis for quantifying and giving statistical bounds for the percentage of large earthquakes in CEUS that can be expected to occur in areas where previous earthquakes have occurred. Based on the data analyzed in this study, I estimate that at least two thirds to three fourths of the future large earthquakes in CEUS will occur in zones delineated by historical seismicity.

INTRODUCTION

Intraplate regions such as the central and eastern United States (CEUS) present a challenge to seismologists working on the problem of characterizing the earthquake hazard. On the one hand, large and damaging earthquakes occur in CEUS, but on the other hand there are few (if any) cases in which well documented active faults were identified prior to these earthquakes. Furthermore, no comprehensive and scientifically sound physical models have proven to be reliable indicators of where future large earthquakes are likely to occur in this region. This situation forces seismologists who are attempting to estimate the hazard to rely heavily on the observed record of seismicity as an indicator of where future large earthquakes are likely to occur in CEUS (*e.g.*, Frankel, 1994; Frankel *et al.*, 1996; Wheeler and Frankel, 2000). Thus, it behooves us to test the underlying hypothesis that future earthquakes tend to occur where past earthquakes have occurred.

The purpose of this study is to use earthquakes that have already occurred, both in CEUS and elsewhere, to statistically test the hypothesis that seismicity delineates areas where future large earthquakes are likely to occur. The primary objective of this study is to investigate this hypothesis for CEUS earthquakes. I also address this issue for other regions, however, both for comparison with CEUS and to gain more general insight into the extent to which seismicity can be used as an indicator of where future large earthquakes are likely to occur.

BACKGROUND

In spite of efforts over many years to understand the cause of CEUS earthquakes, the models that have been proposed to explain these earthquakes tend to be in the category of conjecture or interesting speculation awaiting scientific hypothesis testing, rather than scientifically sound theories. Most models claiming to explain CEUS earthquakes are some variation on a "pre-existing zones of weakness" model (*e.g.*, Sykes, 1978). In this model, it is conjectured that preexisting faults and/or other geological features which formed during ancient geological episodes persist in the intraplate crust. By way of analogy with plate-boundary seismicity, earthquakes then supposedly occur when the present-day stress is released along these zones of weakness. Much of the research on CEUS earthquakes has, therefore, involved attempts to identify preexisting faults and other geological or geophysical features that might be "reactivated" by the present-day stress field (*e.g.*, Aggarwal and Sykes, 1978; Talwani, 1982; Seeber and Dawers, 1989; Kafka and Miller, 1996; Marshak and Paulson, 1997; Wheeler and Frankel, 2000).

While this concept of reactivation of old zones of weakness is commonly assumed to be valid, in reality the identification of individual active geological features has proven to be quite difficult (*e.g.*, Kafka, 2000). Thus, the relationship between locations of earthquakes and faults or other geological/geophysical features in CEUS is, for all practical purposes, still unknown.

The CEUS portion of the 1996 National Seismic Hazard Maps (developed by the U.S. Geological Survey) does not rely very heavily on presumed correlations between earthquakes and geological structures. Rather, it is heavily weighted toward using the observed record of seismicity to characterize the seismic hazard (Frankel, 1995; Frankel *et al.*, 1996; Wheeler and Frankel, 2000). Following Wheeler and Frankel (2000), I use the term "geology-based hypotheses" as shorthand for hypotheses based on information about geologic structures (including faults and fault zones) and tectonics, and "seismicity hypothesis" as shorthand for the hypothesis that seismicity delineates where future large earthquakes are likely to occur.

These two types of hypotheses are, of course, not necessarily mutually exclusive. The actual occurrence of earthquakes in CEUS and elsewhere can probably be explained by some combination of the two (plus perhaps some other as yet unarticulated hypotheses). Thus, an important question regarding earthquake hazard analysis in CEUS is, how do we decide how much weight to assign to the various hypotheses? The primary objective of this study is to shed light on the question of how much weight to assign to the seismicity hypothesis.

METHOD

Statistical Hypothesis Testing in Earthquake Studies

Statistical hypothesis testing is difficult in earthquake studies. While hypotheses should be tested on multiple independent data sets, in earthquake studies 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.) Despite this constraint that is inherent to studies of this type, I use multiple data sets that are to a large extent independent of each other to test the seismicity hypothesis for CEUS. I do this in three ways: First, I evaluate how well seismicity "retrodicts" CEUS earthquakes that have already occurred. Second, I compare how well seismicity retrodicts earthquakes across the entire CEUS with how well it retrodicts earthquakes on smaller scales, specifically regional network scales in subregions of CEUS. Finally, I evaluate how well seismicity retrodicts earthquakes that have occurred in other regions (including a variety of tectonic environments) and compare the CEUS results with results from other regions.

There is a conceptual difference between testing the seismicity hypothesis versus testing geology-based hypotheses. In the latter case, the data set used to test the hypothesis is often the same data set that was used to *formulate* the hypothesis (or is at least a data set that is not independent of that used to formulate the hypothesis). Even if a later-occurring catalog of seismicity is used to test a geology-based hypothesis, that data set is likely to be dependent on the data used to formulate the hypothesis because (1) the data set may contain aftershocks of the earthquakes used to formulate the hypothesis, and/or (2) the stress released by the earthquakes used to formulate the hypothesis has changed the potential for earthquake occurrence in the study region. In the case of the seismicity hypothesis, however, the development of the hypothesis is not based on any particular distribution of seismicity; rather, it is based on an intuitive sense that it is reasonable to imagine that future earthquakes (whether they are aftershocks or future mainshocks) will occur in areas where earthquakes have occurred in the past. Such an hypothesis is, therefore, formulated independently of the data used to test the hypothesis.

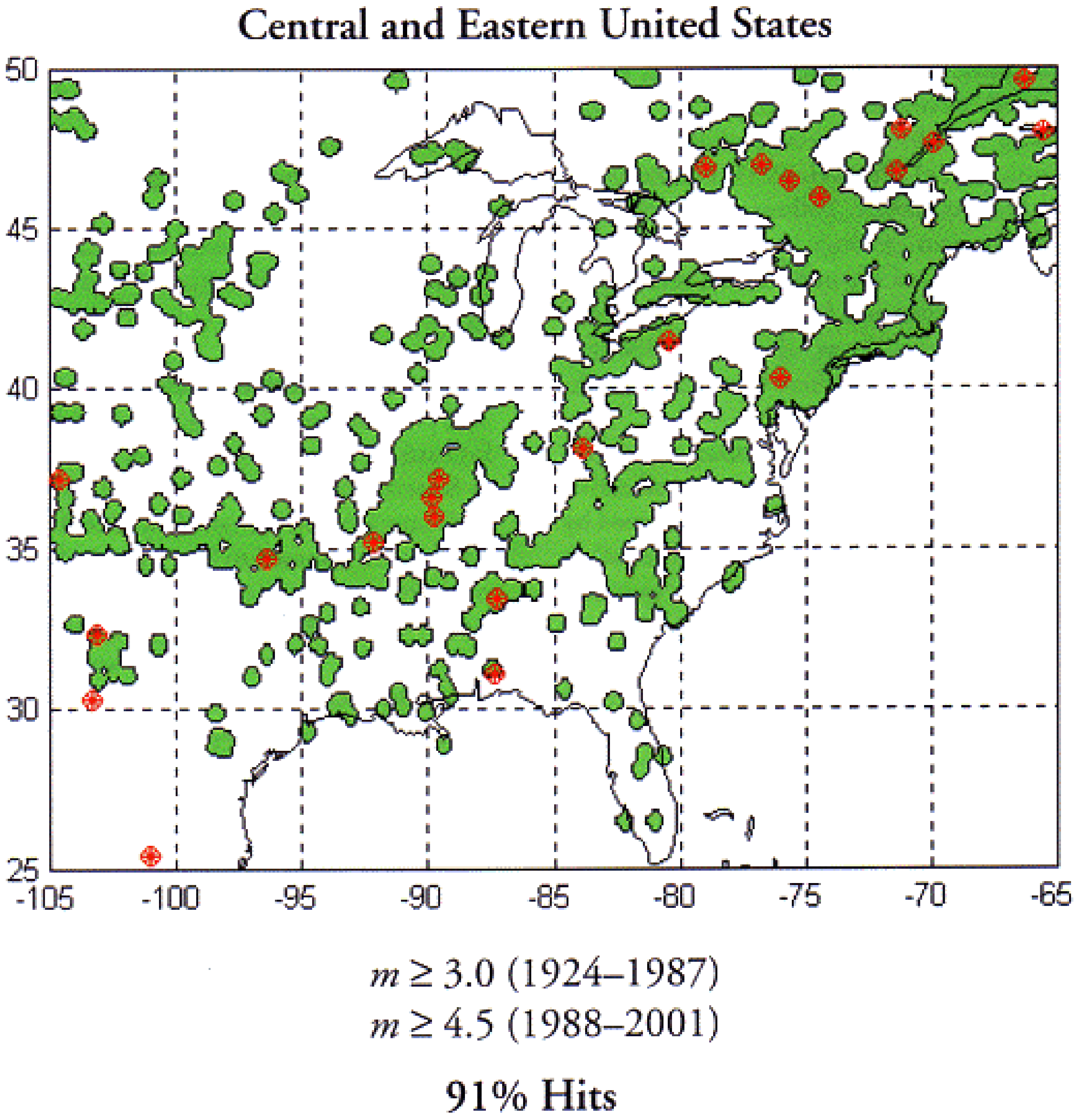
Testing the Seismicity Hypothesis

The physical model being tested here is simple and straightforward: Within a given region, some areas are hypothesized to be seismically active for a wide range of magnitude levels, and some are not. The smaller, more frequent, earthquakes in a region are assumed to be a statistical sample of areas that are prone to experience earthquakes. For each region analyzed, I define a "small-earthquake catalog" as the catalog that extends from the earliest time and smallest magnitude for which the catalog is complete up to the year before the beginning of a "large-earthquake catalog" for that region. The objective, then, is to evaluate how well the areas delineated by the small-earthquake catalog forecast the locations of the "future" large earthquakes. The specific magnitude cutoffs for "small" and "large" earthquakes depend on the level of monitoring and the level of earthquake activity in the region.

The method used for this study is analogous to the configuration of a cellular phone system. Circles of a given radius are constructed around each epicenter in a small-earthquake catalog, and the percentage of large earthquakes that occurred within the given radius of at least one previous small earthquake is systematically investigated. This is a rather simple method of characterizing the small-earthquake seismicity, but we tried more complex approaches and found that the results were not significantly different from what we obtained using this "cellular" method (*e.g.*, Kafka and Levin, 2000).

This method is illustrated in Figure 1 for the case of CEUS. Circles of a given radius are constructed around each epicenter in the small-earthquake catalog ($m \geq 3.0$, 1924–1987), and the radius is varied so that the interiors of the circles fill up a given percentage of the map area. Then the earthquakes in the "future" large-earthquake catalog ($m \geq 4.5$, 1988–2001) are analyzed to see what percentage of them occurred within that given radius of at least one of the earthquakes in the small-earthquake catalog. In the case illustrated in Figure 1, I have chosen the radius of the circles to be large enough so that their interiors fill 33% of the map area (offshore parts of the map are not included in the area calculation). If at least one previously occurring small earthquake is within the specified distance of a given large earthquake, I call that a "hit." In the case illustrated in Figure 1, the radius that is sufficient to fill 33% of the area is 36.0 km, and for that radius there are 91% hits.

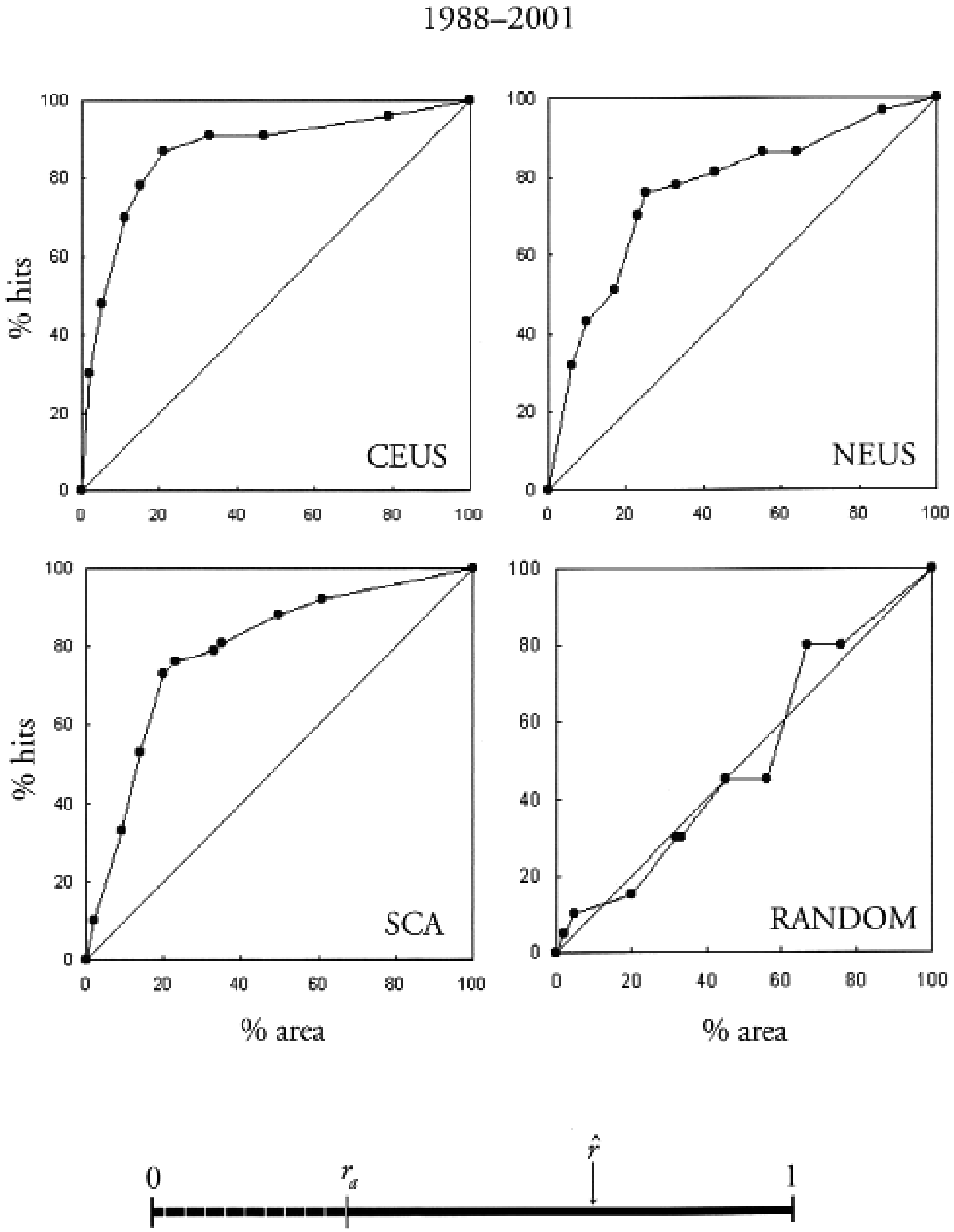
The zones created by using a radius large enough to fill 33% of the map area were chosen as an operational definition



▲ **Figure 1.** Illustration of method used in this study. Green shading shows areas surrounding the 1924–1987 catalog of small earthquakes ($m \geq 3.0$) in CEUS, calculated using the cellular method described in the text. Radii of circles surrounding the epicenters are chosen such that 33% of the map area is filled. Red symbols indicate epicenters of large earthquakes ($m \geq 4.5$) that occurred in CEUS between 1988 and 2001.

of areas where the small-earthquake seismicity is concentrated. This arbitrary choice of one third of the map area is used below only for the purpose of illustration. Figure 2 shows that choosing some other percentage of area for the operational definition does not change the essential conclusions of this aspect of the study. The graphs shown in Figure 2 are for CEUS, the northeastern United States (NEUS), and southern California (SCA), as well as for a synthetic random distribution of small and large earthquakes. For each case, the graphs show the percentage of hits as a function of the percentage of map area. For the three cases of real data, the percentage of hits consistently exceeds the percentage of map area; for the random case, the percentage of hits scatters around the percentage of map area. Based on these types of results, it seems reasonable to conclude that the large earthquakes do indeed seem to be preferentially occurring in areas of previous seismicity (Kafka and Walcott, 1998; Kafka and Levin, 2000). While this result is hardly surprising, Figure 2 suggests that is may be possible to quantify and thus statistically analyze the extent to which previous seismicity delineates locations of future large earthquakes.

The idea behind using percentage of map area for the statistical analysis presented here is to base that analysis on a dimensionless variable that captures the concept of zones delineated by seismicity within a given region. We want this variable to be dimensionless so that we can compare regions of different sizes, and it should also capture a combination of the characteristics of proximity to previous earthquakes and distribution of seismicity. We also want to choose a variable that is not overly dependent on the particular shape of the active zones of seismicity. The percentage of map area covered by circles of a given radius surrounding the smaller earthquake seismicity is chosen as a variable that, at least to some extent, satisfies these criteria. One might envision that the radius of the circles surrounding the epicenters might be a better (and more fundamental) variable to use for this purpose because it is more directly related to the physics of the earthquake process. Later in this paper I present an empirical



▲ **Figure 2.** (Above) Percentage of hits as a function of percentage of map area surrounding the small-earthquake catalog for CEUS, NEUS, and SCA. Also shown is the same calculation for a test of randomly distributed seismicity catalogs of large and small earthquakes for a hypothetical region. In this case, both the small- and large-earthquake catalog events are randomly distributed spatially over an arbitrary 10° × 10° region. (Below) Illustration of the relationship between r_a and $\hat{r}$ for the statistical analyses discussed in this paper.

analysis demonstrating that percentage of map area is a better variable than radius in terms of satisfying the criteria described above.

Statistical Analysis

The next question I address is whether the concept of observed percentage of hits for a given percentage of map area has any physical, or at least statistical, meaning. We begin by imagining that we are investigating a region in which earthquakes occur but for which we are ignorant of the cause of those earthquakes (this is quite a realistic assessment of the situation for CEUS). In Figure 1, which shows the application of the "cellular" method to the CEUS data, the green shading shows areas surrounding the 1924–1987 small-earthquake catalog ($m \geq 3.0$), and the red symbols indicate epicenters of large earthquakes ($m \geq 4.5$) that occurred between 1988 and 2001. Thus, the question is, to what extent does the green shading delineate areas where future large earthquakes (red symbols) are likely to occur? Let the logic of the problem be defined as shown at the bottom of Figure 2.

In Figure 2, the ratio r_a is the percentage of map area covered by circles of a given radius surrounding the smaller-earthquake seismicity, and $\hat{r}$ is the percentage of future large earthquakes that are observed to occur within that given radius of at least one of the smaller earthquakes. Consider the range of all possible values of $\hat{r}$ for a given value of r_a. On one end of the spectrum, we could imagine that $\hat{r}$ could be absolutely any value between 0 and 1, regardless of the information content in the spatial distribution of smaller earthquake seismicity. On the other end of the spectrum, we could imagine that the nature of earthquake processes (either specifically for CEUS or for earthquake processes in general) is such that there is some more systematic shape to the distribution of possible values of $\hat{r}$.

To gain additional insight into the nature of that distribution, let us next consider the sources of variation for $\hat{r}$. There are a number of reasons why we would expect variation in the observed percentage of hits. First, variation may be due to differences in the earthquake processes in the regions as a result of different seismotectonic environments. Second, there may be differences in the extent to which the specific small-earthquake catalogs are representative "snapshots" of the longer-term seismicity in the regions (*e.g.*, different small-magnitude cutoffs, different levels of catalog completeness, different lengths of records of small-earthquake seismicity). Finally, of course, there is random variation due to having taken samples of the population of all possible earthquakes that could occur in the region.

The approach that I take in the statistical analysis that follows is to treat the observed percentages of hits ($\hat{r}$) as samples of a random variable. Specifically, I let the interior of the 33% area contour be an operational definition of areas "near" previous small earthquakes, and I assume that a random variable (r) exists corresponding to the percentage of large earthquakes that tend to occur "near" previous small earthquakes. The process of examining whether or not a given large earthquake occurs within the 33% area contour is treated as a binomial experiment in which the binomial random variable, r, corresponds to the probability of "success" (*i.e.*, within the green shaded area) and $1-r$ corresponds to the probability of "failure." The statistical analyses presented below are based on this formulation of the problem.

ESTIMATING THE DISTRIBUTION OF *r* VALUES FOR CEUS

Having observed 91% hits for CEUS (with $r_a = 0.33$) is encouraging in terms of supporting the idea of using seismicity as an indicator of locations of future large earthquakes. However, this one observed value of $\hat{r}$ does not provide a great deal of insight into the more general problem being investigated here. Since it is only one observation, it is difficult to discern whether it is just a coincidence that we obtained such a high value for $\hat{r}$. What we really want to know is what the *distribution* of $\hat{r}$ is, so that we can decide on an appropriate estimate of r for CEUS (as well as statistical bounds on that estimate). The only way to estimate what that distribution might look like is to conduct the same type of investigation for other realizations of the same process.

Strictly speaking, we would have to wait several hundred years to obtain the data necessary to estimate the distribution of $\hat{r}$ for CEUS. We can, however, make a first step toward estimating that distribution by applying the same method to earthquake catalogs in other regions and at other scales. We can then ask, is this tendency for a high percentage of large earthquakes in CEUS to occur near previous known epicenters a coincidence for this particular data set, period of time, tectonic environment, and scale? Alternatively, there may be some general, fundamental measure of the tendency for future large earthquakes to occur in areas where earthquakes have occurred in the past, regardless of tectonic environment and scale.

In the next two sections, I use data from regional networks within CEUS as well as data from regions in other tectonic environments to obtain a rough estimate of the distribution of $\hat{r}$ for $r_a = 0.33$. The statistical analysis presented below is only for $r_a = 0.33$, but we have conducted similar analyses for other values of r_a and have found that the essence of the results does not depend on the specific choice of the value for r_a.

TESTING THE SEISMICITY HYPOTHESIS FOR REGIONAL SCALES WITHIN THE CEUS

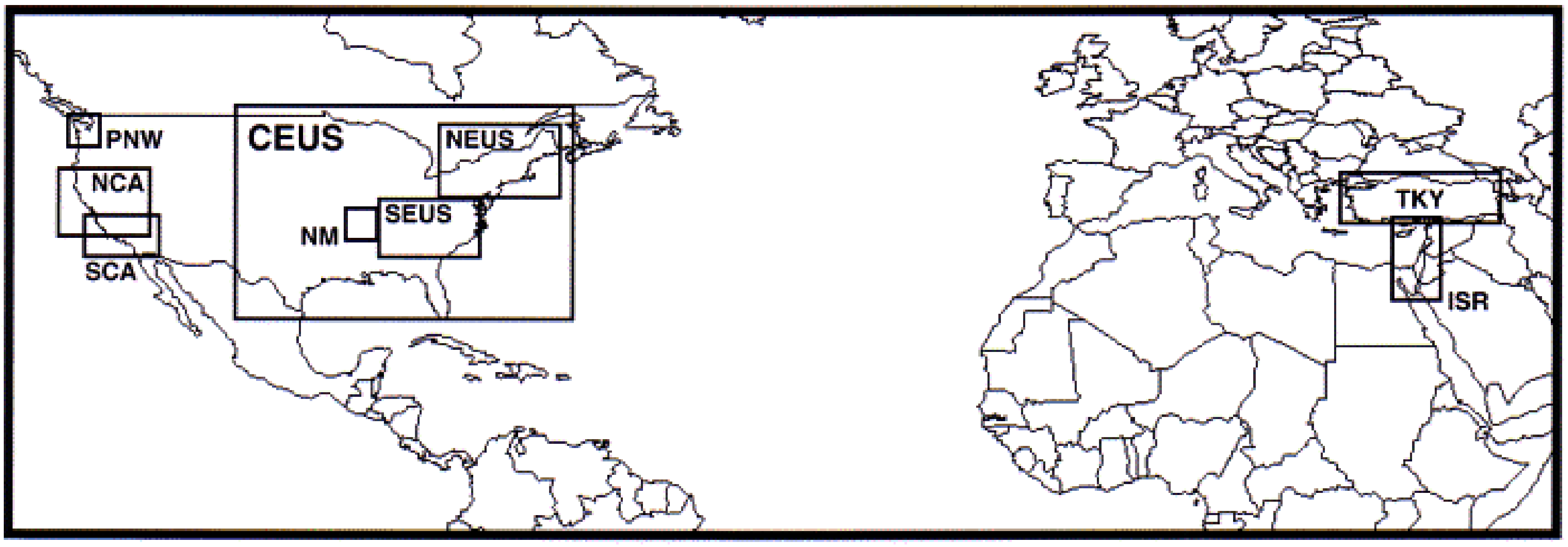
Within CEUS there are three regional catalogs that were analyzed in this study (Figure 3): NEUS, SEUS, and New Madrid (NM). Figure 4 shows an example of the application of the cellular method to one of these regions (NEUS). For that case, the small-earthquake catalog consists of all earthquakes of magnitude 2.0 and greater during the time period 1975–1987, and the large-earthquake catalog is for magnitude 4.0 and greater for 1988–2001.

The results for all three of these regions are shown in Figure 5 and Table 1. Also shown in Table 1 are the magnitude cutoffs for the small- and large-earthquake catalogs for each region, as well as the radii that correspond to 33% of the map area for each case. The values of $\hat{r}$ for these regions range from 0.60 to 0.89. Based on these results and the results for the entire CEUS, we would conclude that an estimate of the range for the distribution of $\hat{r}$ would be 0.60 to 0.91.

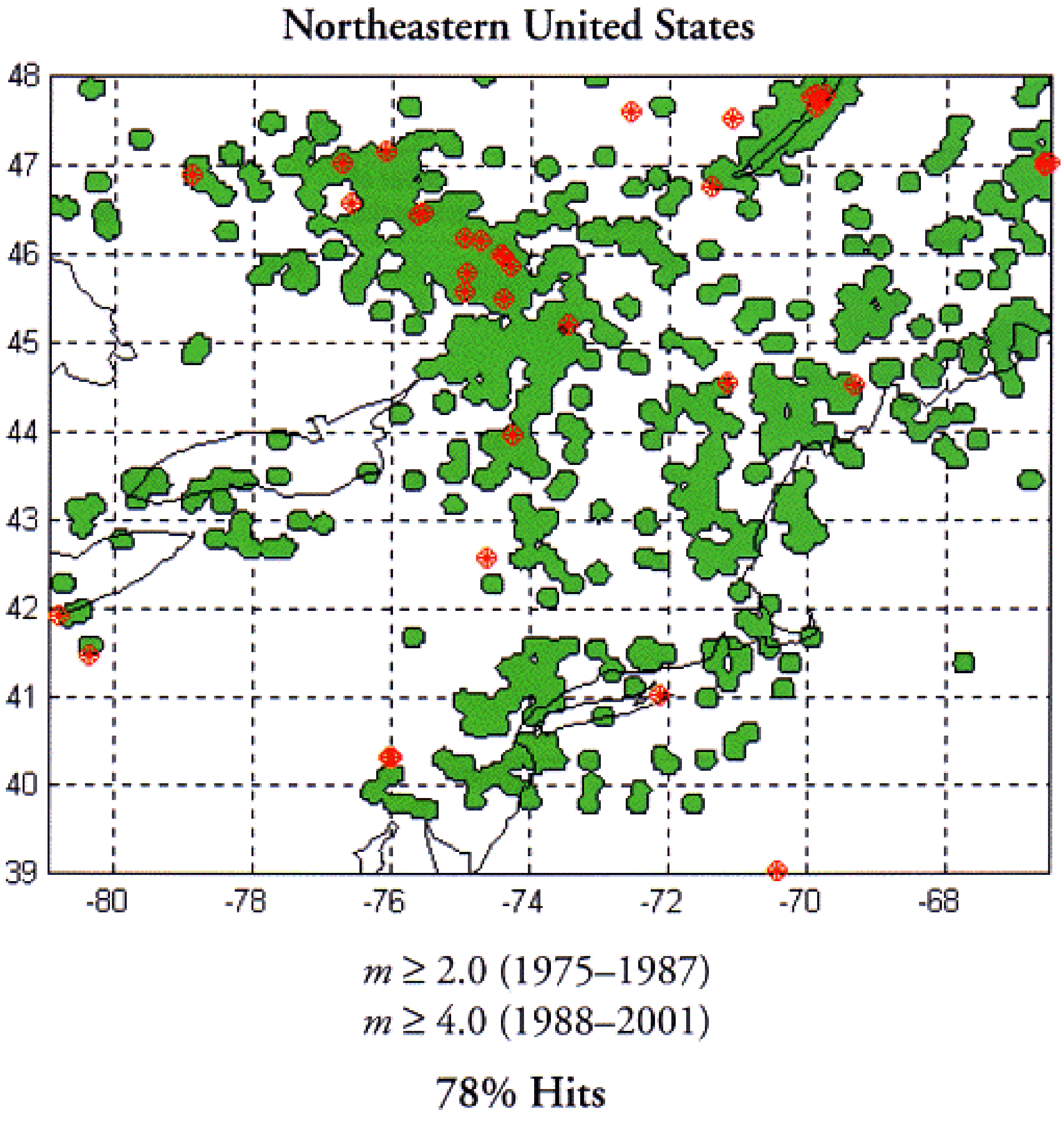
The next question I address is whether there is any statistically significant difference in the observed values of $\hat{r}$ when the cellular method is applied to regions in different tectonic environments.

TESTING THE SEISMICITY HYPOTHESIS IN OTHER TECTONIC ENVIRONMENTS

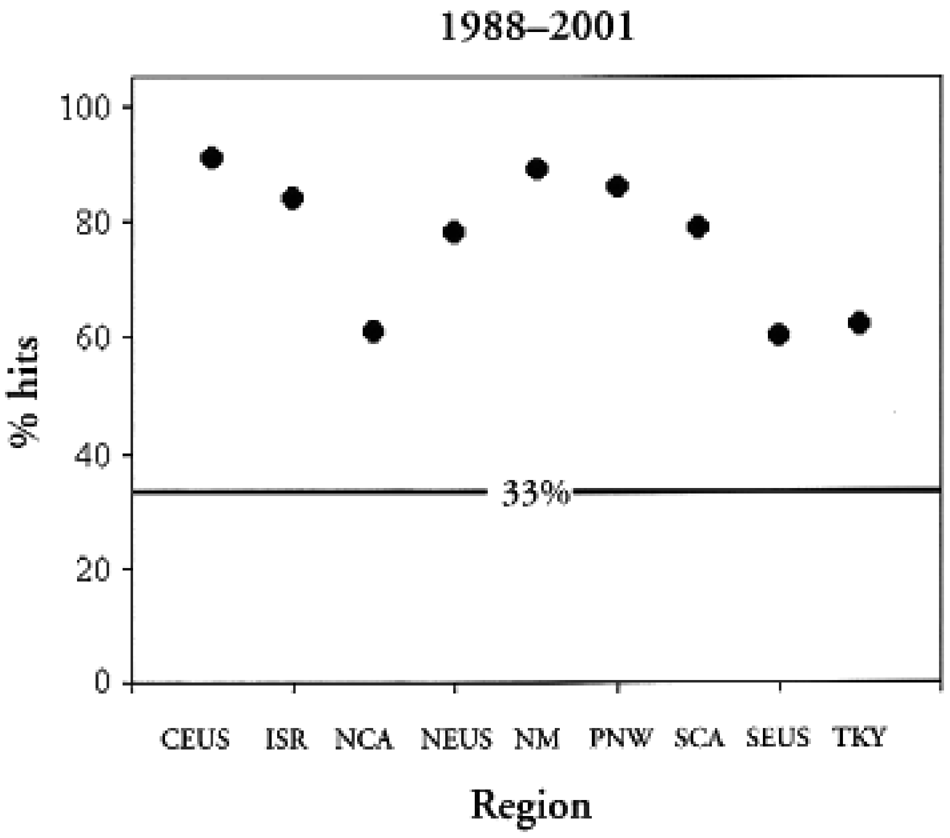
As was done in the previous section for smaller regional scales within CEUS, here the method described above is applied to earthquake catalogs in regions within different tectonic environments (Figure 3). The specific regions analyzed are SCA, northern California (NCA), Pacific Northwest (PNW), Turkey (TKY), and Israel (ISR). Figure 6 shows, as an example, the details of the results for SCA. In the case of SCA, the small-earthquake catalog consists of all earthquakes of magnitude 3.0 and greater during the time period 1984–1987, and



▲ **Figure 3.** Map indicating locations of all regions analyzed in this study. Regional network data are from northeastern United States (NEUS), southeastern United States (SEUS), southern California (SCA), northern California (NCA), New Madrid (NM), Turkey (TKY), Israel (ISR), and the Pacific Northwest (PNW). For CEUS, I used the catalog that was developed for the 1996 National Seismic Hazard Maps.



▲ **Figure 4.** Green shading shows areas surrounding the 1975–1987 catalog of small earthquakes ($m \geq 2.0$) in NEUS, calculated using the cellular method described in the text. Radii of circles surrounding the epicenters are chosen such that 33% of the map area is filled. Red symbols indicate epicenters of large earthquakes ($m \geq 4.0$) that occurred in NEUS between 1988 and 2001.



▲ **Figure 5.** Percentage of hits for r_a equal to 33% of the map area for all of the regions investigated in this study.

TABLE 1
Percent Hits for 33% Area, 1988–2001

Region*	Type	Radius (km)**	Number of Earthquakes	Number of Hits	%Hits
CEUS (3.0, 4.5)	IP	36.0	23	21	91%
ISR (4.0, 5.0)	PB	48.0	55	46	84%
NCA (3.0, 5.0)	PB	10.0	112	68	61%
NEUS (2.0, 4.0)	IP	15.5	37	29	78%
NM (2.2, 3.2)	IP	13.5	37	33	89%
SCA (3.0, 5.0)	PB	13.2	78	62	79%
SEUS (2.0, 3.5)	IP	31.0	30	18	60%
TKY (4.3, 5.3)	PB	26.0	47	29	62%
PNW (2.5, 3.5)	PB	9.2	42	36	86%
All PB			334	241	72%
All IP			127	101	80%

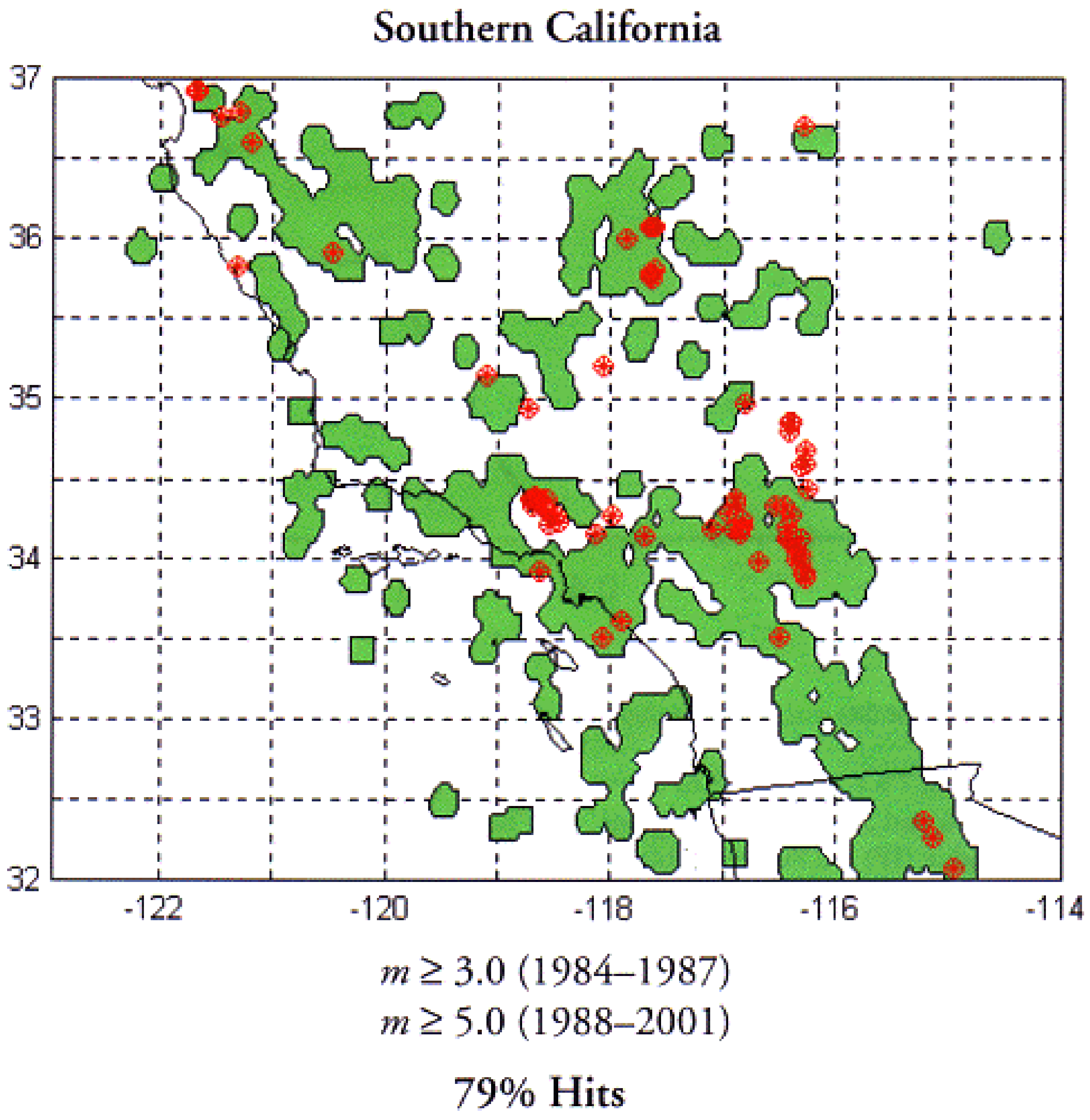
$\hat{r}$ (PB) = 72%
$\hat{r}$ (IP) = 80%
$\hat{r}$ = 74%

$z\left[\hat{r}(\mathrm{PB}) - \hat{r}(\mathrm{IP})\right] = -1.62 \Rightarrow p$ value = 0.11
For C = 0.71:
$z(\hat{r} - C) = 1.64 \Rightarrow p$ value = 0.05

*Numbers in parentheses are small- and large-magnitude cutoffs, respectively.
**Radius necessary to fill 33% of the map area.

PB = Plate Boundary
IP = Intraplate

Earthquake catalog data for these analyses have been obtained from Center for Earthquake Research and Information, Geophysical Institute of Israel, Kandilli Observatory and Earthquake Research Institute, Lamont-Doherty Earth Observatory, Massachusetts Institute of Technology, Northern California Earthquake Data Center, Pacific Northwest Seismograph Network, Southern California Earthquake Center, U.S. Geological Survey, Virginia Polytechnic Institute, and Weston Observatory.



▲ **Figure 6.** Green shading shows areas surrounding the 1984–1987 catalog of small earthquakes ($m \geq 3.0$) in SCA, calculated using the cellular method described in the text. Radii of circles surrounding the epicenters are chosen such that 33% of the map area is filled. Red symbols indicate epicenters of large earthquakes ($m \geq 5.0$) that occurred in SCA between 1988 and 2001.

the large-earthquake catalog is for magnitude 5.0 and greater for 1988–2001. The results for those regions are shown in Figure 5 and Table 1. The values of $\hat{r}$ for these regions range from 0.61 to 0.86 (a range that is quite similar to that obtained for the CEUS intraplate examples described above).

These results from other tectonic environments provide an opportunity to address the question of whether the percentage of large earthquakes that tend to occur near previous small earthquakes depends on the tectonic environment of the region studied. I explored this question by testing the hypothesis that the proportion (r) of large earthquakes occurring near previous small earthquakes is the same for plate-boundary (PB) regions versus intraplate (IP) regions. For this hypothesis test, the null hypothesis is that the percentage of large earthquakes that tend to occur near previous small earthquakes is the same for IP versus PB regions. The alternative hypothesis is that the percentages are different for the two types of regions.

Let r(PB) and r(IP) be the proportions of large earthquakes that tend to occur "near" previous small earthquakes (*i.e.*, within the 33% map area contours) in plate-boundary and intraplate regions, respectively, and let the observed percentages be $\hat{r}$ (PB) and $\hat{r}$ (IP). Also, let the population proportion and observed proportion for all of the data lumped together be r and $\hat{r}$, respectively. The hypotheses to be tested are

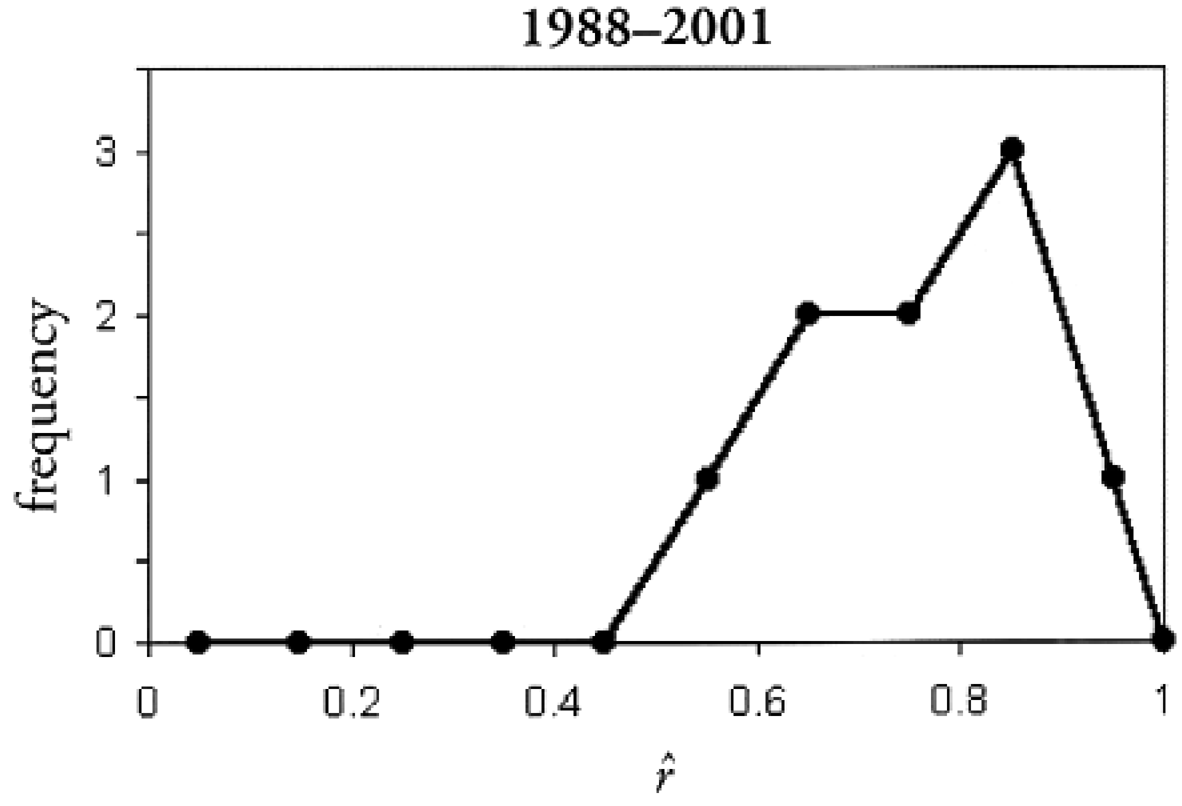
$$H_0\text{: } r(\mathrm{PB}) = r(\mathrm{IP}) \quad \text{and}$$

$$H_a\text{: } r(\mathrm{PB}) \neq r(\mathrm{IP}), \tag{1}$$

where H_0 is the null hypothesis and H_a is the alternative hypothesis. The appropriate test statistic for this binomial experiment is (*e.g.*, Weiss and Hassett, 1982):

$$z = \frac{\hat{r}(\mathrm{PB}) - \hat{r}(\mathrm{IP})}{\sqrt{\hat{r}(1-\hat{r})\left(\dfrac{1}{n(\mathrm{PB})} + \dfrac{1}{n(\mathrm{IP})}\right)}} \tag{2}$$

where n(PB) and n(IP) are the numbers of large earthquakes in the PB and IP regions, respectively.



▲ **Figure 7.** Distribution of $\hat{r}$ based on the data analyzed in this study. The $\hat{r}$ values shown here are calculated for $r_a = 0.33$.

Considering ISR, NCA, SCA, TKY, and PNW to be PB regions, and CEUS, NEUS, NM, and SEUS to be IP regions, we find that $\hat{r}$ (PB) = 0.72 and $\hat{r}$ (IP) = 0.80 (see Table 1). This difference in observed percentages yields a z value of −1.62, which corresponds to a statistical p value of 0.11, implying that there is not sufficient evidence to reject the null hypothesis at the 90% level of statistical significance. Thus, although there may be differences in r values between plate-boundary and intraplate regions, these observations do not provide strong evidence of such differences. For the remainder of this paper, therefore, it will be assumed that the r values for all the regions shown in Figure 3 can be combined together to estimate the distribution of $\hat{r}$ (Figure 7).

PERCENTAGE OF MAP AREA AS A VARIABLE THAT CHARACTERIZES PROXIMITY TO PREVIOUS EARTHQUAKES AND DISTRIBUTION OF SEISMICITY

In this study, percentage of map area (r_a) is used as a dimensionless variable that is intended to capture the characteristics of proximity to previous earthquakes and distribution of seismicity. The idea behind basing the statistical analysis on this variable is to find a way to compare regions of different sizes and different distributions of seismicity. One might envision that the radius of the circles surrounding the epicenters might be a better (and more fundamental) variable to use for this purpose because it is more directly related to the physics of the earthquake process. I have found, however, that the percentage of hits for a given radius is more affected by the characteristics of the specific region investigated than is the percentage of hits for a given percentage of map area. Thus, percentage of map area is a more useful variable than radius for comparing the characteristics of seismicity for different regions.

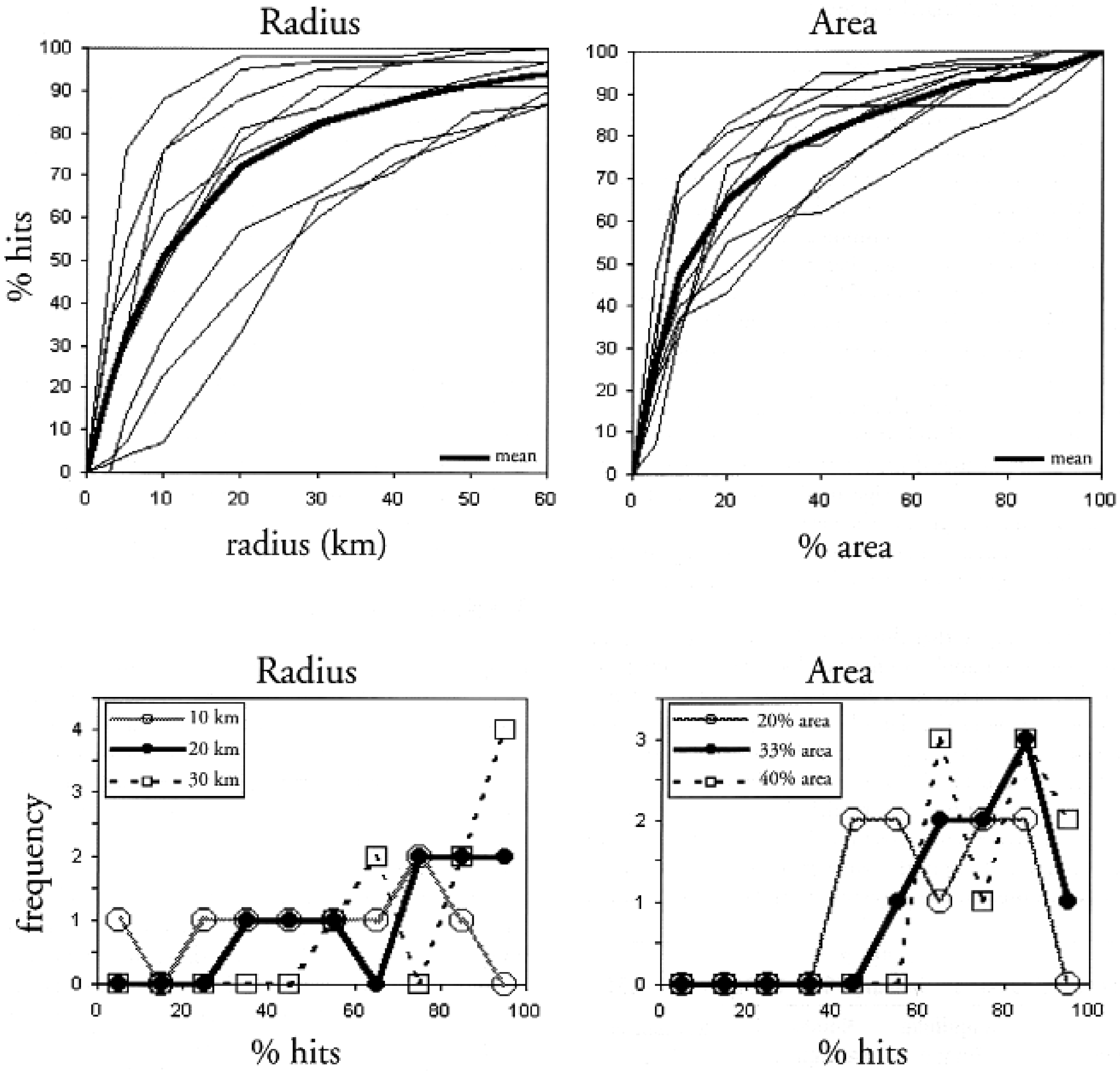
Figure 8 shows the percentage of hits as a function of radius and as a function of the percentage of map area for all regions analyzed in this study. Also shown are histograms of percentage of hits for selected examples of radius and percentage of map area. Choosing percentage of area rather than radius results in a lower variance and also yields histograms that are closer to being normally distributed. For a given mean level of percentage of hits, the standard deviation of the radius histograms is almost twice that of the area histograms. In general, the percentage of map area is seen to have more desired characteristics as a variable for comparing regions and for making statistical inferences regarding estimates and statistical bounds for r values.

WHAT IS AN APPROPRIATE ESTIMATE OF THE *r* VALUE FOR THE CEUS?

Assuming that the distribution shown in Figure 7 can be applied to CEUS, I next use that distribution to address the question of what might be an appropriate r-value for CEUS (given $r_a = 0.33$). Since a reasonable objective would be to choose the value of r that forecasts the locations of as many earthquakes as possible, an appropriate r value would be the largest value of r that is statistically acceptable given the observed percentage of hits. Thus, the next hypothesis to be tested is that r is greater than some critical value (C) for higher and higher values of C. The observed value of r ($\hat{r}$) for all of the data lumped together is 0.74 (see Table 1), and the hypotheses to be tested are

$$H_0\colon r = C \quad \text{and}$$

$$H_a\colon r > C. \tag{3}$$



▲ **Figure 8.** (Above) Percentage of hits as a function of radius and percentage of map area for all regions analyzed in this study. Percentage of map area is the percentage covered by circles of a given radius surrounding the smaller-earthquake seismicity, as discussed in the text. Thin solid lines denote the individual regions, and thick solid lines indicate the mean for all regions. (Below) Histograms of percentage of hits for radius and for percentage of map area. Note that, for a given mean percentage of hits, the variance is greater for radius than it is for percentage of map area.

For this situation, the test statistic is (*e.g.*, Weiss and Hassett, 1982)

$$z = \frac{\hat{r} - C}{\sqrt{C(1-C)\left(\frac{1}{n}\right)}} \qquad (4)$$

where $n = n(\mathrm{PB}) + n(\mathrm{IP})$.

Based on this test statistic, we can raise the value of C as high as 71% and still reject the null hypothesis at the 95% level of statistical significance. Conceptually, this means that at the 95% level of statistical significance, we can expect that on average more than 71% of the large earthquakes in a region will tend to occur near previous small earthquakes. Thus, I conclude from this analysis that, to the extent that these samples are representative of seismicity in general, about two thirds or more of the large earthquakes in CEUS (or any other region?) tend to occur in zones delineated by previous seismicity.

DISCUSSION AND CONCLUSIONS

If there were no information content in the distribution of seismicity that bears upon the question of what the *r* value is likely to be, then there would be no reason to base any component of seismic hazard analysis on the distribution of historical seismicity. If that were the case, we would be left with the situation in which seismic hazard can be estimated based only on knowledge of the seismotectonic processes in the region. In that situation, given that we know so little about the seismotectonic processes in CEUS, we would be able to conclude only that large earthquakes are equally likely anywhere in the region. The results of this study suggest that the situation is not that hopeless.

The distribution of *r* values does appear to have a shape that is more complex than just being uniform between 0 and 1, and the shape of that distribution can be estimated from studies of this type. Knowledge of the shape (and how it might vary from one region to the next) would make it possible to estimate an appropriate *r* value for a given region. This study is a first step toward estimating an appropriate *r* value for CEUS.

One might argue that this type of analysis does not say anything about our ability to forecast locations of truly large earthquakes, those large enough to be of real concern. Since in some of the regions the "large" earthquakes are only in the magnitude 3.0 to 4.0 range, extrapolating these results to truly large earthquakes involves arbitrarily invoking some type of "self similarity" in the model. A preliminary way to investigate this issue is to extract the three largest earthquakes in each region and apply the same method to those earthquakes as was used in the previous examples. Table 2 shows the results of that investigation. For each large-earthquake catalog, the three largest earthquakes were extracted. Out of a total of 28 of the largest earthquakes in this time period, 18 of them occurred in the areas where small earthquakes were concentrated, *i.e.*, in the areas defined by choosing the radii of the circles such that 33% of the map area is covered. Thus, we have 64% hits in this case (*i.e.*, a value of $\hat{r}$ that is quite consistent with the other values of $\hat{r}$ found in this study). While this exercise does not resolve the entire issue of whether this type of analysis applies to truly large earthquakes, we can at least conclude that it does not yield evidence against such an extrapolation.

TABLE 2
Largest Earthquakes 1 January 1988–12 December 2001

Region	Magnitude Range	Number of Earthquakes*	Number of Hits
CEUS	5.2–5.8	4	3
ISR	6.2–6.9	3	2
NCA	7.0–7.4	3	2
NEUS	5.0–5.2	3	3
NM	4.3–4.8	3	2
SCA	6.7–7.4	3	1
SEUS	4.3–4.8	3	2
TKY	6.8–7.8	3	1
PNW	5.4–6.8	3	2
Total		**28**	**18**

Percent hits for 33% map area = 18/28 = 64%.
*For CEUS, two earthquakes were tied for the third largest magnitude. Thus, there are four events listed for that region.

It is intuitively reasonable to expect that the *r* values would vary from one region to the next, but the data analyzed in this study are not able to resolve such differences (if they exist). In the absence of evidence to the contrary, I conclude that an appropriate *r* value for CEUS is 0.71. Although this study is still "exploratory", the results do suggest that it is possible to quantify the extent to which seismicity delineates areas where future large earthquakes are likely to occur.

ACKNOWLEDGMENTS

I appreciate the efforts of the many individuals at Weston Observatory, Lamont-Doherty Earth Observatory, Massachusetts Institute of Technology, Center for Earthquake Research and Information, Virginia Polytechnic Institute, Southern California Earthquake Center, Northern California Earthquake Data Center, Geophysical Institute of Israel, Kandilli Observatory and Earthquake Research Institute, Pacific Northwest Seismograph Network, U.S. Geological Survey, and numerous other institutions responsible for operating regional seismic networks and global monitoring stations. Without the efforts of those individuals, seismic monitoring would never have survived long enough to generate the earthquake catalogs that are used as the basis for this study. I thank John Ebel, Rus Wheeler, Art Frankel, Shoshana Levin, Jessica Walcott, and Kevin Harrison for many valuable discussions about various aspects of this study.

REFERENCES

Aggarwal, Y. P. and L. R. Sykes (1978). Earthquakes, faults and nuclear power plants in southern New York and northern New Jersey, *Science* **200**, 425–429.

Frankel, A. (1995). Mapping seismic hazard in the central and eastern United States, *Seismological Research Letters* **66**(4), 8–21.

Frankel, A., C. Mueller, T. Barnhard, D. Perkins, E. V. Leyendecker, N. Dickman, S. Hanson, and M. Hopper (1996). *National Seismic Hazard Maps, Documentation—June 1996*, U.S. Geological Survey Open-File Report 96-532.

Kafka, A. L. (2000). Public misconceptions about faults and earthquakes in the eastern United States: Is it our own fault?, *Seismological Research Letters* **71**, 311–312.

Kafka, A. L. and S. Z. Levin (2000). Does the spatial distribution of smaller earthquakes delineate areas where larger earthquakes are likely to occur?, *Bulletin of the Seismological Society of America* **90**, 724–738.

Kafka, A. L. and P. E. Miller (1996). Seismicity in the area surrounding two Mesozoic rift basins in the northeastern United States, *Seismological Research Letters* **67**(3), 69–86.

Kafka, A. L. and J. R. Walcott (1998). How well does the spatial distribution of smaller earthquakes forecast the locations of larger earthquakes in the northeastern United States?, *Seismological Research Letters* **69**, 428–439.

Marshak, S. and T. Paulson (1997). Structural style, regional distribution, and seismic implications of midcontinent fault-and-fold zones, United States, *Seismological Research Letters* **68**, 511–520.

Rhoades, D. A. and F. F. Evison (1989). On the reliability of precursors, *Physics of the Earth and Planetary Interiors* **58**, 134–140.

Seeber, L. and N. Dawers (1989). Characterization of an intraplate seismogenic fault in the Manhattan prong, Westchester, Co., N.Y., *Seismological Research Letters* **60**, 71–78.

Sykes, L. R. (1978). Intraplate seismicity, reactivation of preexisting zones of weakness, alkaline magmatism, and other tectonism postdating continental fragmentation, *Reviews of Geophysics and Space Physics* **16**, 621–688.

Talwani, P. (1982). An internally consistent pattern of seismicity near Charleston, South Carolina, *Geology* **10**, 654–658.

Weiss, N. and M. Hassett (1982). *Introductory Statistics*, Reading, MA: Addison-Wesley Publishing Co.

Wheeler, R. L. and A. Frankel (2000). Geology in the 1996 USGS Seismic-Hazard Maps, Central and Eastern United States, *Seismological Research Letters* **71**, 273–282.

Weston Observatory
Department of Geology and Geophysics
Boston College
Weston, MA 02493
kafka@bc.edu