

EFFECTS OF UNCERTAINTY IN SEISMICITY ON ESTIMATES OF SEISMIC HAZARD FOR THE EAST COAST OF THE UNITED STATES

BY ROBIN K. MCGUIRE

ABSTRACT

The lack of knowledge of the cause of earthquakes in the Eastern United States and the short length of seismic history lead to decisional and statistical uncertainties concerning the seismicity of the East. The effect of these uncertainties on the calculated seismic hazard for the east coast is assessed by deriving probability distributions on the important parameters.

The available earthquake catalog for the Eastern United States is adequate to establish activity rates for seismic sources but not maximum possible sizes of events. Thus it is not justifiable on a statistical basis to assume that events larger than those observed historically in areas of the East will not occur in the future. The Modified Mercalli intensities associated with an annual probability of 10^{-4} of being equaled or exceeded range from VIII to IX for selected sites from Florida to Maine; these intensities are generally insensitive to the manner in which seismic sources are drawn to represent seismicity, except for those sites which lie within an important seismic source under one hypothesis and outside all seismic sources under a different hypothesis.

Several disadvantages of using a strictly deterministic procedure to determine design intensities are disclosed by this study. Specifically, design intensities established deterministically are sensitive to seismic source geometry and to the largest event observed historically in each source. The risks associated with these design intensities vary by more than a factor of ten for the sites on the east coast examined here.

INTRODUCTION

Statistical uncertainty in parameters can have an important effect on the calculated hazard for engineering problems, especially when high reliability (low risk) design levels are sought with a small amount of data available on which to base the analysis (Veneziano, 1974). This is the case for nuclear power plants in the Eastern United States, where high reliability is required for seismic design but where the available history and low levels of seismic activity give few data on which to base seismic design decisions. The incorporation of statistical parameter uncertainties in seismic hazard analysis has been proposed and illustrated by several authors (Esteve, 1969, 1974; Veneziano, 1975). However, no comprehensive study of seismic hazard for many sites along the east coast of the United States has been reported which incorporates uncertainties in seismic source zone, activity rate, b value, and maximum possible size of events. Such a study is the purpose of the research reported here.

Throughout this study, Modified Mercalli (MM) intensity is used as the measure of ground shaking and of earthquake size. MM intensity is taken to be a continuous, rather than a discrete, random variable; results for the discrete variable are easily drawn from those of the continuous variable. Conclusions regarding the effect of parameter uncertainties on design ground motions derived in this study will apply for any parameter of ground motion which behaves similarly to MM intensity with respect to event size and source to site distance.

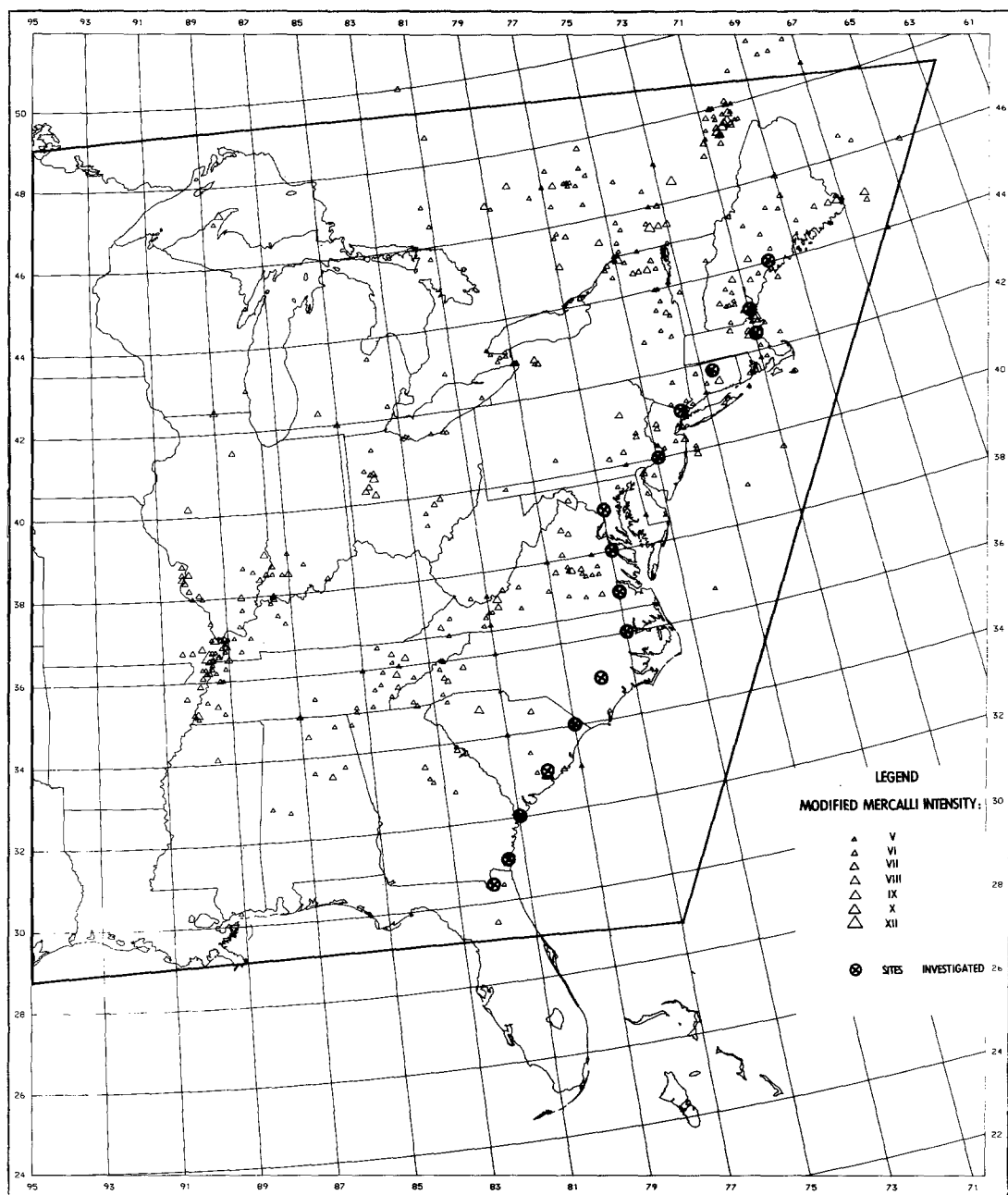


FIG. 1. Epicenters, single seismic source, and sites investigated.

To perform a probabilistic seismic hazard analysis for a site, the seismicity of the region must first be defined. In typical seismic hazard analyses (e.g., Cornell and Merz, 1975; Shah *et al.*, 1975; Algermissen and Perkins, 1976), the definition of seismicity is made as follows. Using geologic and tectonic data as well as observed earthquake locations, the region of study is divided into seismic sources, within which future events are considered equally likely to occur at any location. For each seismic source, the rate of occurrence of events above a chosen threshold level is estimated, using the observed frequency of historical events. The sizes of successive events in

each source are assumed to be independent and exponentially distributed; the slope of the log-number versus frequency relationship (which is proportional to the parameter of the exponential distribution) is estimated from the relative frequency of different sizes of events observed in the historical data. This slope, often termed the b value (Richter, 1958), is determined either for each seismic source individually or for all sources in the region. Finally, the maximum possible size of events for each source is determined, using judgment and the historical record.

In this study, judgmental and statistical uncertainties in the definition of seismicity for probabilistic hazard analysis in the Eastern United States are examined, and the effect of these uncertainties on the hazard calculated for building sites on the east coast is determined. Uncertainties in seismic sources, activity rates, b values, and maximum possible sizes of events are described, using the earthquake catalog illustrated in Figure 1 as a source of data, and the influence of these uncertainties on seismic hazard is illustrated where appropriate.

Data for the study consisted of a catalog of 467 events occurring between 1534 and 1974 in the United States and Canada east of longitude 95°W, for which maximum MM intensities between V and XII had been assigned. This catalog was obtained by removing obvious aftershocks from a larger data set. The data are plotted on a map of the Eastern United States in Figure 1; at locations where several epicenters are listed, only the largest MM intensity has been plotted.

SEISMIC SOURCES

Several alternate hypotheses were examined for specification of seismic sources within which the earthquake recurrence process is considered spatially and temporally homogeneous. The effect of these different hypotheses on the calculated seismic hazard is described below.

The first hypothesis is based on the assumption that seismicity within the entire Eastern United States is uniformly distributed, i.e. that areas of high or low rate of occurrence cannot be distinguished, and that events of MM intensity up to XII can occur anywhere. The area used to calculate the occurrence rate under this hypothesis is shown in Figure 1.

A second hypothesis for the specification of seismic sources examined is that used in a recent seismic risk study of the United States (Algermissen and Perkins, 1976). These sources were determined using historical seismicity and available geologic and tectonic evidence, and are shown in Figure 2.

The third hypothesis used to specify seismic sources was derived from a seismotectonic map of the Eastern United States (Hadley and Devine, 1974). Again, this seismotectonic map was compiled by interpretation of historic seismicity and tectonic data. The sources used in this study, derived from the Hadley and Devine map, are shown in Figure 3. Some slight modifications of the areas as originally published by Hadley and Devine were required for computer input, such as the use of straight line segments rather than curves to define area boundaries. These changes in no way affect the results or conclusions reached in this study.

The differences between the seismic sources in Figures 1, 2, and 3 are worth emphasizing. The source in Figure 1 assumes no knowledge of the causative mechanism underlying the earthquake process, whereas the sources of Figures 2 and 3 presume that the pattern of future activity can be estimated, in a gross, statistical sense, by historical seismicity and available geologic and tectonic evidence. The differences between the sources of Figures 2 and 3 are important; Figure 2 shows a rather broad geographical interpretation of seismic, geologic, and tectonic data, whereas Figure 3 shows a more

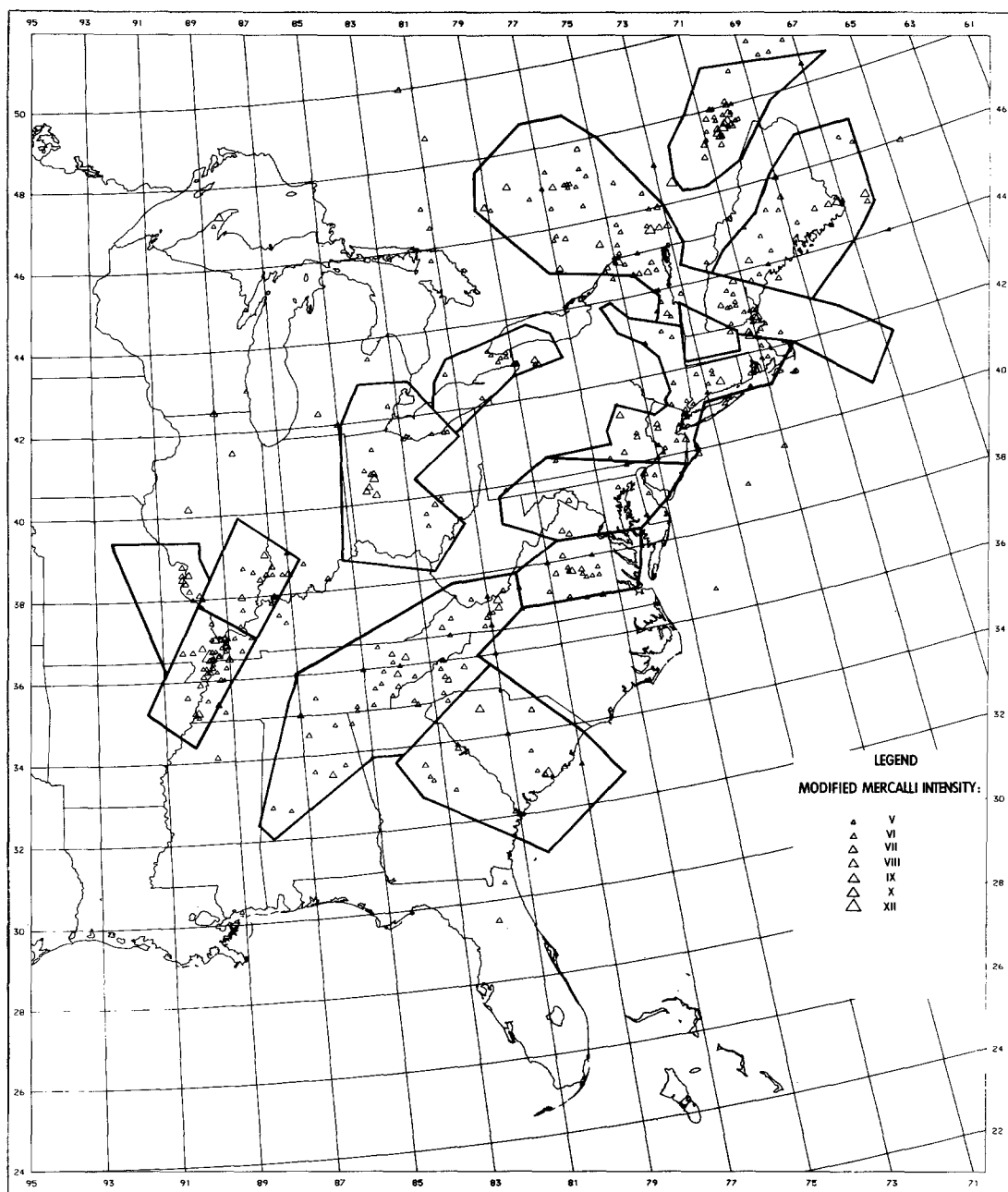


FIG. 2. Seismic sources after Algermissen and Perkins (1976).

restrictive interpretation (compare, for instance, the sizes of the sources around Charleston, South Carolina, in the two figures). Other differences between Figures 2 and 3 are evident: Figure 2 shows a source corresponding to the so-called "Boston-Ottawa trend," whereas Figure 3 assumes no continuity between the seismicity near Boston and that of the St. Lawrence valley. Figure 3 shows a continuous zone of seismicity along the Appalachian mountains, whereas no such continuous trend is depicted by the sources of Figure 2. Thus, whether or not one agrees that any of the sources of Figures 1, 2, and 3 represent the "real" seismic sources, these hypotheses

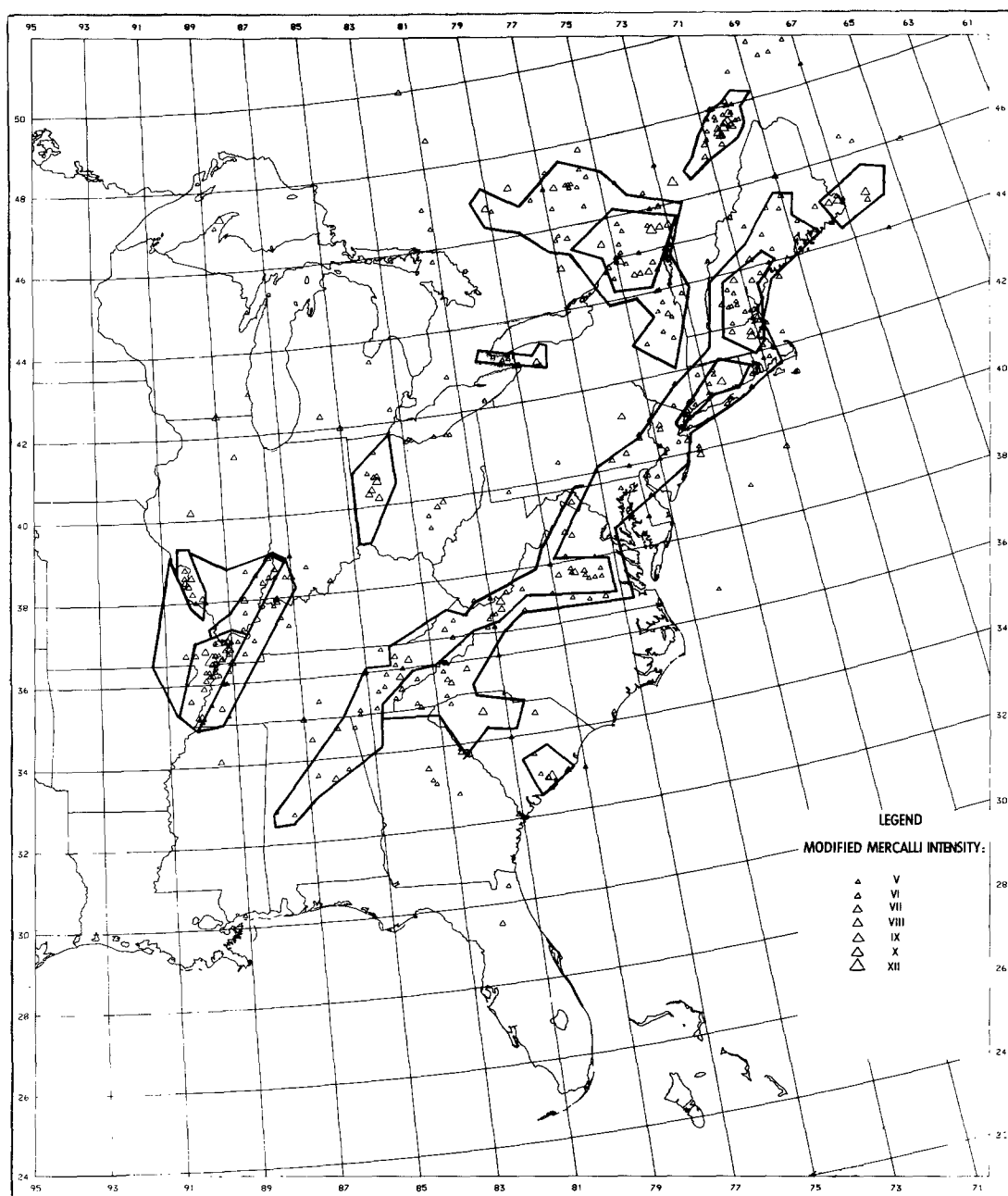


FIG. 3. Seismic sources after seismotectonic map of Hadley and Devine (1974).

certainly represent a wide range of interpretation of available data for the Eastern United States. The influence of these different interpretations will be illustrated later in this paper.

ACTIVITY RATE

The occurrence rate of events in each seismic source with epicentral MM intensity greater than or equal to V was treated as a random variable. Its distribution was obtained through the likelihood function (Raiffa and Schlaifer, 1961; Benjamin and

Cornell, 1970). If the seismic history indicates n observations of intensity $i_1, i_2, \dots, i_n$ in time τ in the seismic source area of interest, and the sizes of these events are assumed to be independent of their number, the joint likelihood function of the activity rate ν , the maximum possible intensity I_m , and the parameter β of the intensity distribution (to be discussed below) is

$$L(\nu, I_m, \beta | i_1, i_2, \dots, i_n \text{ in } \tau) = P[i_1, i_2, \dots, i_n \text{ in } \tau | \nu, I_m, \beta]. \quad (1)$$

Because the number of events in time τ depends only on ν and not on I_m or β , Equation (1) can be written

$$L(\nu, I_m, \beta | i_1, i_2, \dots, i_n \text{ in } \tau) = p[n \text{ in } \tau | \nu]P[i_1, i_2, \dots, i_n | I_m, \beta]. \quad (2)$$

From the separable form of equation (2), it is apparent that given the n observations in time τ , ν is independent of I_m and β . Thus the likelihood function for ν , and its probability density function, may be considered separately from those of I_m and β .

It is assumed that earthquakes occur as Poisson arrivals; this process has been found to describe adequately the occurrence of large events when aftershocks are disregarded (Lomnitz, 1966, 1973; Gardner and Knopoff, 1974). With this assumption, the likelihood function for ν is equal to the Poisson distribution

$$\begin{aligned} L(\nu | n \text{ in } \tau) &= P[n \text{ in } \tau | \nu] \\ &= \frac{e^{-\nu\tau}(\nu\tau)^n}{n!}. \end{aligned} \quad (3)$$

A density function for ν can be obtained from the likelihood function by normalizing so that the integral over ν from 0 to ∞ is unity. This gives

$$f_\nu(\nu) = \frac{\tau}{n!} e^{-\nu\tau}(\nu\tau)^n. \quad \nu \geq 0 \quad (4)$$

In the context of Bayesian probability this is the posterior distribution of ν given the likelihood function and a uniform or noninformative prior distribution.

With this result, the effect on calculated risks of treating the activity rate ν as a random variable, rather than simply choosing a fixed value, can be demonstrated as follows. Let $G_1(i_s)$ be the complementary-cumulative probability for intensity i_s (i.e., the risk that i_s will be equaled or exceeded) at a site from a single event of random intensity and location in a seismic source. Under the Poisson process assumption, if ν is fixed (or is known exactly), the complementary-cumulative distribution of the *maximum* site intensity in time t is (Benjamin and Cornell, 1970)

$$G_{\max}(i_s) = 1 - \exp(-\nu t G_1(i_s)). \quad (5)$$

Uncertainty in the rate ν can be accounted for by multiplying the right side of equation (5) by the density function on ν and integrating over all values, yielding

$$G_{\max}(i_s) = 1 - \left[\frac{\tau}{\tau + t G_1(i_s)} \right]^{n+1} \quad (6)$$

(Benjamin, 1968). For the values of n , τ , t , and G_1 of interest in this study it can be shown that the ratio of risk given by equation (6) to that given by (5) is approximately $1 + 1/n$. The term $1/n$ thus indicates the increase in risk due to uncertainty in the rate of occurrence of events. For n as small as 5, there is only a 20 per cent increase in the calculated risk using equation (6) rather than (5). As shall be shown, an increase of 20 per cent in risk results in a very small change in the design intensity associated with that risk. From this point of view there certainly have been enough earthquakes recorded in the Eastern United States to define rates of occurrence in seismic sources, for the purpose of calculating design intensities through risk analysis.

For the practical case when the catalog is considered complete for varying lengths of time for different intensity levels, a similar analysis is straightforward. Equation (6) may be used to account for statistical variability in ν , replacing n by $n_1 + n_2 + n_3 + \dots$ and τ by $\tau_1 + p_2\tau_2 + p_3\tau_3 + \dots$, where n_1 is the number of events in time τ_1 during which the lowest intensity level is considered complete, n_2 is the number of events during the *previous* time τ_2 during which the next higher intensity is considered complete, and p_2 is the probability of observing an event of the second intensity level or greater given that an event occurs, and so on. The effect of accounting for statistical variation in ν thus is not sensitive to the length of time during which the upper intensity levels are assumed complete.

RELATIVE FREQUENCY OF EVENTS OF DIFFERENT SIZES

The logarithm of the number of events occurring in a region at each intensity level is assumed to be linearly proportional to epicentral intensity, i.e.

$$\log N(I_e) = a - bI_e \quad (7)$$

where N is the number of events of epicentral intensity I_e in a given time period, and a and b are parameters applicable to the region of study. This assumption is consistent with a linear log-number versus magnitude relation (Richter, 1958) and a linear relationship between magnitude and epicentral intensity. Equation (7) has been found to be useful by several investigators in describing seismicity in the Eastern United States (e.g. Algermissen, 1969; Chinnery and Rogers, 1973; Cornell and Merz, 1975). From Equation (7) it follows that the distribution of event sizes is exponential

$$f_{I_e}(i) = \frac{\beta \exp(-\beta(i - i_o))}{(1 - \exp(-\beta(I_m - i_o)))} \quad i_o \leq i \leq I_m \quad (8)$$

where I_m is an upper bound at which the distribution is truncated, i_o is a chosen lower bound intensity for which the activity rate is estimated, and β equals $b \ln 10$.

To determine the observed frequency of events with different epicentral intensities, it is enlightening to plot the number of events per decade for each intensity level. Such a plot is shown in Figure 4 for the entire catalog. (The sequence of events in 1811–1812 near New Madrid, Missouri, has been considered one event of intensity XII. The exact treatment of this sequence is not critical to the results which follow.)

It is evident from Figure 4 that the record of events for the higher intensities is complete for a longer period of time than that for the lower intensities. This is reasonable because earthquakes with higher intensities are felt over large areas and are more likely to be reported in early news accounts than small shocks. To determine an average number of events per year, it is necessary to estimate the length of time that

the catalog is complete at each intensity level. Such estimates are subjective; two sets of estimates for the catalog of Figure 4 are shown in Figure 5.

Once the frequency of events at each intensity level has been chosen, the value of parameter b in equation (7) can be calculated by one of several methods. Figure 5 shows the estimates of event frequency corresponding to two values of b . The lower

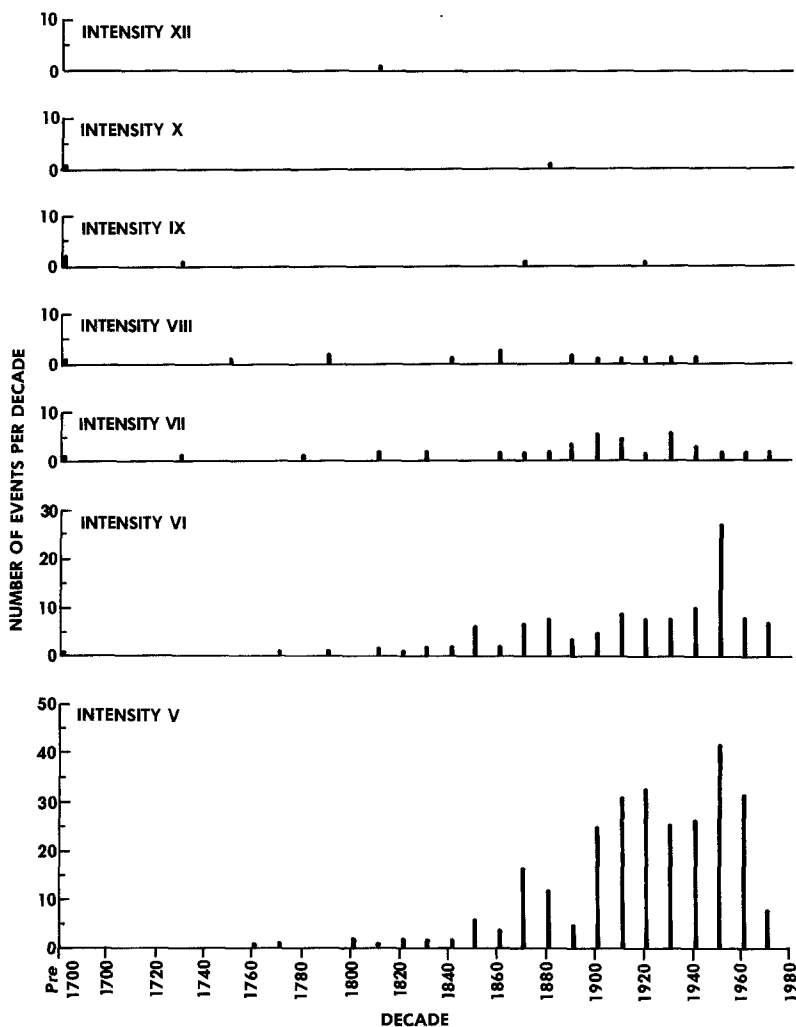


FIG. 4. Number of events per decade in Eastern United States, for different epicentral MM intensities.

line (b equal 0.57) was obtained by applying a least-squares procedure to the open circles shown on the figure, neglecting the intensity XII observation. A least-squares procedure is not logical because it weights all points used in the analysis equally, whereas the frequencies for the lower intensities are better determined (and therefore should carry more weight) because they are estimated using a larger number of shocks. However, the least-squares procedure has gained acceptance among some investigators because the lower intensities are sometimes not completely reported, and the least-squares method does not emphasize the possibly inaccurate rates observed for

these lower intensities. The upper line in Figure 5 (b equal 0.45) was obtained using a maximum likelihood criterion applied to the closed squares. Such a criterion weighs the frequencies observed at lower intensities very heavily and those at high intensities very lightly.

The purpose of these comparisons is to show that, for a given earthquake catalog, it is possible to derive different b values depending on subjective interpretation of the length of completeness of the catalog and on the method of analysis used. The values of b shown in Figure 5 (0.57 and 0.45) represent what are thought to be reasonable

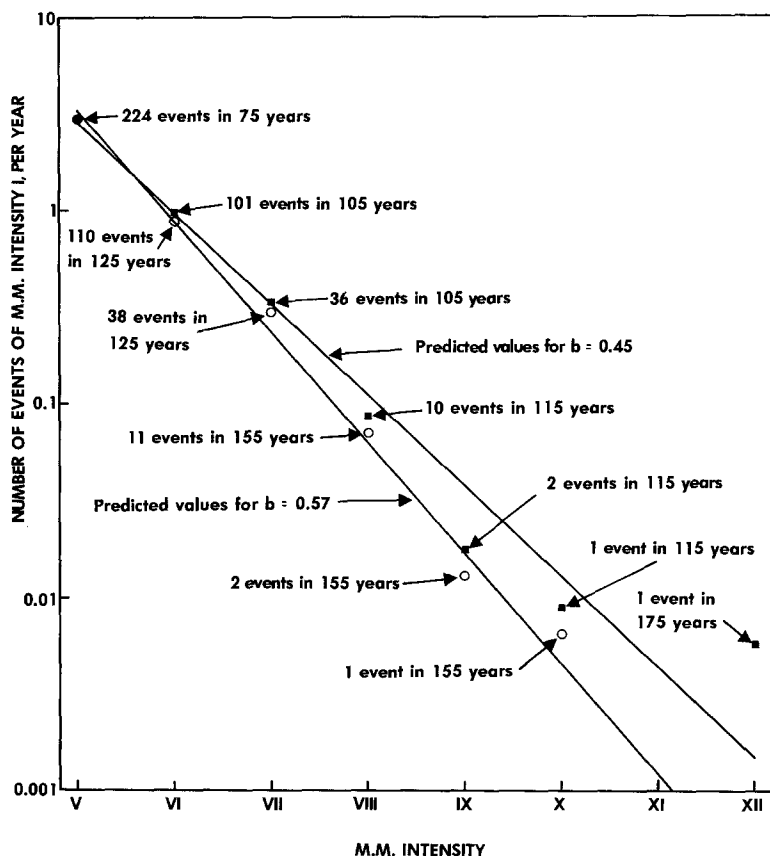


FIG. 5. Two estimates of event frequency for different M.M. intensities, and predictions based on two b values.

extremes for the catalog of Figure 4. For computations in this study, a b value of 0.50 was used; this corresponds to a minimum chi-squared solution for the closed squares of Figure 5, and also to a weighted least-squares solution with weights proportional to the number of events used to determine each frequency. The sensitivity of results to the value of b used will be demonstrated.

Having selected a b value for the entire catalog, it is reasonable to question whether the b values determined for the seismic sources individually will correspond to that for the entire catalog. In general they will not; b values computed (by the method of maximum likelihood) for the sources of Figure 2 ranged from 0.35 to 0.58, whereas that for the entire catalog (as described above) was 0.45. The values obtained, shown in Figure 6, are plotted as a function of the number of events used in the determina-

tion. It is evident that the largest variation in b is observed for seismic sources which have a small number of events upon which the calculation can be based. One conclusion, then, which is not contradicted by the data, is that variations in b value from source to source are the result of statistical variations due to the small sample sizes, rather than to underlying differences in the earthquake process itself. For this reason it is reasonable to use a single b value for all seismic sources.

Also shown for comparison in Figure 6 are the one-sigma bounds on b computed from the likelihood function, using the mean value of 0.45 as determined from the entire catalog. The data follow the shape of these curves well, but demonstrate less variation for a given sample size than that indicated by the likelihood function. In fact, if the likelihood function were modified by a reasonable non-uniform prior distribution, the Bayesian posterior distribution would have one-sigma bounds which agreed closely with the observed data.

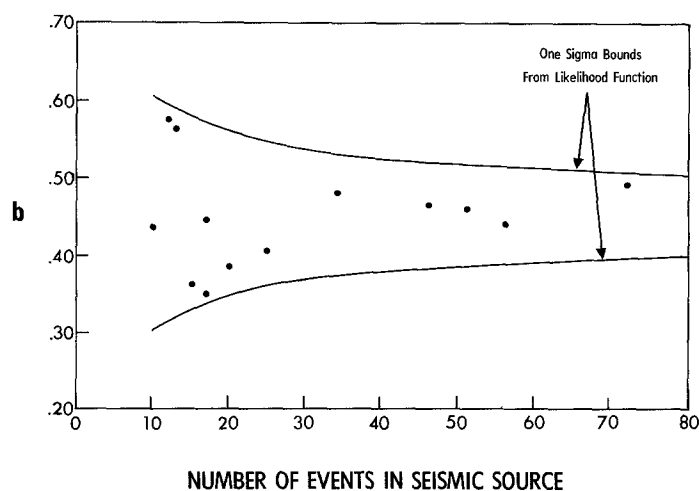


FIG. 6. Maximum likelihood b values for seismic sources of Figure 2, as a function of number of events in the source.

The exponential distribution on epicentral intensity is justified by the observation (made also by other investigators, e.g. Algermissen, 1969; Chinnery and Rogers, 1973; Cornell and Merz, 1975) that the number of events of each MM intensity level plots linearly on a semi-logarithmic scale. It is also consistent with the observed exponential distribution of earthquake magnitudes (Richter, 1958) and an assumed linear relationship between magnitude and the expected maximum intensity. However, some objection has been raised in the past to the use of exponential distributions and b values to describe the occurrence of events with different maximum intensities, on the grounds that MM intensities were developed as a qualitative description of damage, rather than as a quantitative measure of earthquake size. The risk analysis could be done without the exponential distribution and without a b -value, simply by integrating numerically over all maximum intensities using the observed rate of occurrence of each intensity level. The objection to this procedure is that it is highly sensitive to the observed numbers of high-intensity earthquakes and to the assumed length of completeness of the record for these events. If no shocks with intensity greater than, say, VII had occurred in a region, this procedure might predict larger events with equal probability regardless of size. Observation indicates, however, that

events of maximum intensity VIII are more common than those of intensity XII. On balance, the assumption of an exponential distribution on maximum intensities of events is considered far more useful and realistic than alternate procedures, given the present state of knowledge of seismicity in the Eastern United States.

MAXIMUM POSSIBLE INTENSITY

The final parameter necessary to define seismicity is the maximum possible intensity in each seismic source, that is, the upper-bound intensity I_m in the exponential distribution, equation (8). This upper bound is often taken to be fixed, and, for each seismic source, is often set equal to the maximum intensity observed historically

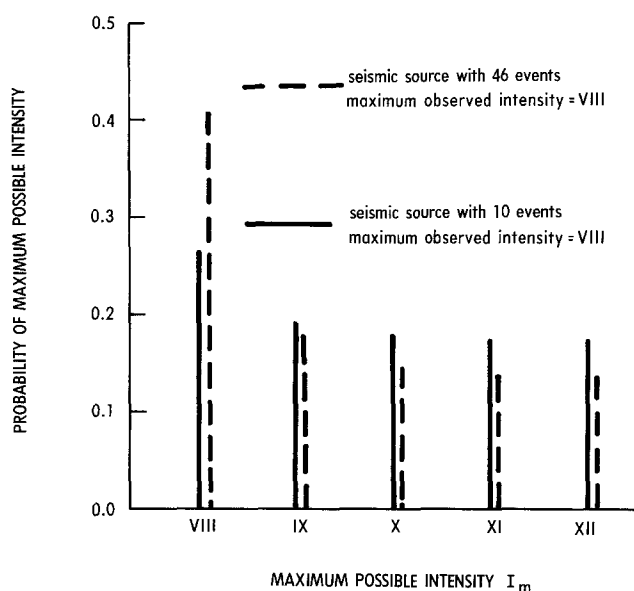


FIG. 7. Distribution of maximum possible intensity for two seismic sources of Figure 2, each with maximum observed MM intensity of VIII.

in that source; in this study the upper bound is treated as a random variable. Its distribution is obtained through the likelihood function, which, for observations $i_1, i_2, \dots, i_n$ in time τ , is equal to the denominator of the intensity function, equation (8), raised to the n th power

$$L(I_m | i_1, i_2, \dots, i_n \text{ in } \tau) = (1 - \exp(-\beta(I_m - i_o)))^{-n} \quad (9)$$

where $I_m \geq i_1, i_2, \dots, i_n$. Corresponding results are easily derived for the practical case when the catalog is considered complete for different lengths of time at different intensity levels. From the likelihood function, the distribution of maximum possible intensity can be calculated by computing likelihood values for discrete MM intensities and normalizing the likelihood function so that the sum of all possibilities, from the maximum observed intensity up to XII, is unity. Figure 7 illustrates this distribution for two seismic sources in Figure 2 for which the largest event in the catalog was MM intensity VIII. In one source 46 events were observed during the complete record, in the other, only 10 events were observed. The influence of the number of shocks in the record is evident: for a larger number of events, all MM in-

tensity VIII or less, it is more probable that VIII is, in fact, the true maximum possible intensity. However, the probability associated with this hypothesis may be quite low, and in fact may be almost equal to the probabilities for higher intensities. In general, for the seismic sources examined in this study, the probability that the maximum possible intensity in a source was equal to the maximum intensity observed in historical time was 0.4 or less; for many sources this probability was about 0.25 (the only notable exceptions were certain sources taken from the Hadley-Devine seismotectonic map whose boundaries were drawn specifically to exclude events of MM intensity greater than VI). Thus, although the maximum observed intensity is always *most* likely to be the maximum possible intensity (e.g., the probability for intensity VIII in Figure 7 is larger than that for any other single intensity), the probability that this is the case is generally low.

Accordingly, a single value for the maximum possible intensity is not assigned to the seismic sources examined in this study. Rather, a uniform distribution for I_m is used, with intensities between the maximum observed in each seismic source and XII assigned equal probability. Thus, for instance, the intensities shown in Figure 7 would be assigned probabilities of 0.2 each. This procedure has the advantage of being straightforward to apply; it also avoids the sensitivity of the calculated distributions (e.g., those in Figure 7) to the assumed length of completeness of the catalog at different intensity levels. Because it assigns larger probabilities to the higher intensities, this assumed distribution is in all cases conservative (that is, it results in higher risks than would be obtained using the calculated distributions). For the majority of seismic sources it is also a very accurate approximation, since the probabilities associated with the hypotheses of higher maximum intensities, which contribute most to the calculated risk, are not changed greatly.

The distributions on maximum possible intensity described here do not deny the possibility of eventually excluding events larger than those observed, for geologic or tectonic reasons. When accepted theories are available which explain the earthquake process in the Eastern United States, they can be used to modify the distributions adopted here. The important point to be emphasized is that the maximum size of events in the Eastern United States cannot be limited on a *statistical* basis, using the available earthquake record. From a purely statistical standpoint, larger events than those yet observed in regions of the Eastern United States are in fact quite likely.

The comment is sometimes made that differences between MM intensities X, XI, and XII are really differences of ground effects, rather than indicators of different amounts of energy release. This observation, while perhaps valid, is of little consequence if one admits that energy releases larger than those associated with MM intensity IX events are possible in the Eastern United States. If this is accepted, then the concept of MM intensities up to XII is useful as a description of these larger energy releases, whether or not one agrees that a given historical shock of MM intensity X, XI, or XII was one of these larger energy events. In case one does *not* believe that earthquakes can release more energy than generally associated with a MM intensity IX event in the Eastern United States, results are presented later in this study using IX as the largest event possible.

ATTENUATION FUNCTION

The attenuation function used in this study was derived from 783 intensity reports from the 1886 Charleston, South Carolina, earthquake (Bollinger, 1976). The result-

ing equation for mean site intensity I_s is

$$\begin{aligned} I_s &= I_e + 3.08 - 1.34 \ln \Delta & \Delta \geq 10 \text{ km} \\ I_s &= I_e & \Delta < 10 \text{ km} \end{aligned} \quad (10)$$

where Δ is epicentral distance in kilometers and I_e is epicentral intensity. In Figure 8, equation (10) is compared to several other attenuation relations reported for the Central and Eastern United States. For epicentral distances less than 500 km, the slopes of the various equations are approximately the same; equation (10) lies 1 to 1.5 intensity units below the other relations because these others were derived using distances to isoseismals rather than individual intensity reports. Use of the individual reports is the appropriate method to predict intensities at a site. From examination of

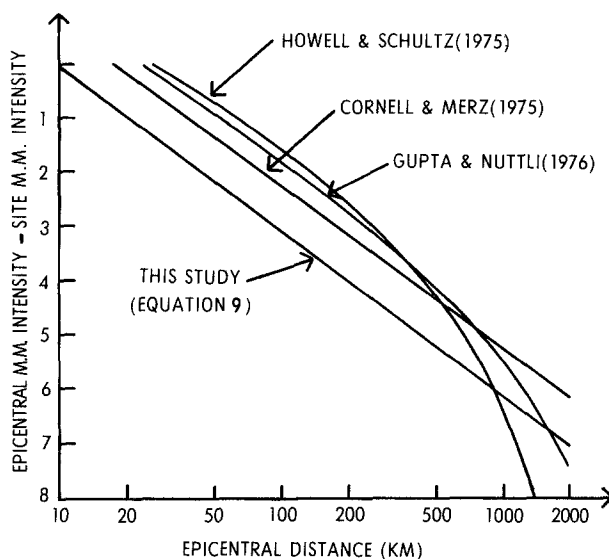


FIG. 8. Attenuation of MM intensity with distance: Relation used in this study (derived from intensity reports) and three relations reported by others (derived from isoseismal maps).

residuals about equation (10), it is found that scatter in the attenuation relation can be modeled using the normal distribution with a standard deviation of 1.19 intensity units. In the data there is no apparent upper truncation in the residual distribution, at least up to 3.5 standard deviations above the predicted value; in the risk analysis, however, the distribution of residuals is truncated in such a manner that site intensities cannot be greater than the epicentral intensity. This implies the assumption (which is not critical) that the epicentral intensity is the maximum intensity observed.

HAZARD ANALYSIS

A seismic hazard analysis was performed for a series of sites on the east coast; the sites chosen are shown in Figure 1. The method used is described elsewhere in detail (Cornell, 1968, 1971; McGuire, 1974); calculations were made with a standard computer program (McGuire, 1976).

Briefly, the total annual probability that a chosen MM intensity is equaled or exceeded at each site was calculated by summing the probabilities of this occurrence

due to all events affecting the site. The exponential distribution, equation (8), was used to describe the relative frequency of different size events, and the attenuation equation (10) was used to quantify the decrease with distance of expected intensity level from source to site. A separate analysis was performed for each of the seismic source hypotheses shown in Figures 1, 2, and 3. Under each hypothesis, the seismicity was assigned to each source according to the events in the catalog within that source. For the sources shown in Figures 2 and 3, the events which did not lie in any source were treated as "background seismicity"; the observed maximum intensity and rate per unit area were obtained from these background events and were used to define the seismicity not associated with discrete sources.

For each source and for background seismicity, the activity rate and maximum possible intensity were treated probabilistically, as previously described. Thus the

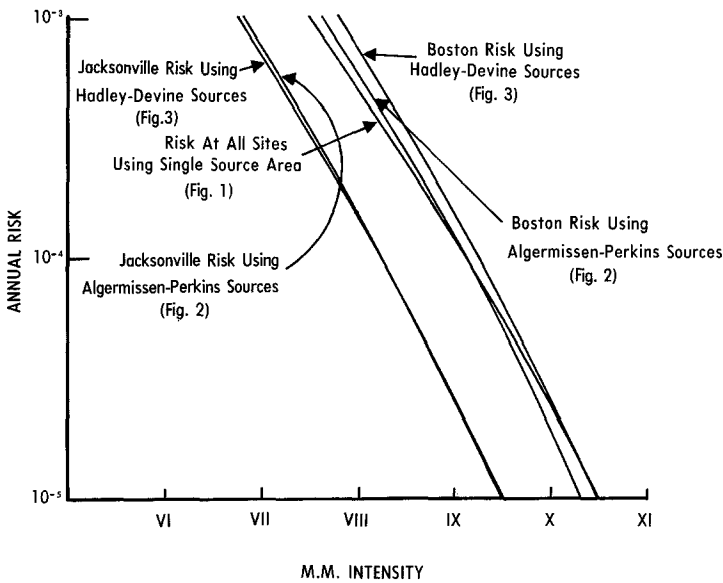


FIG. 9. Annual risk versus MM intensity for several sites and seismic source hypotheses.

occurrence of an event with MM intensity XII was considered possible at all locations, for all seismic source hypotheses. (For the sources of Figures 2 and 3, the probability of occurrence of this event varied greatly from one location to another, due to differences in the occurrence rate and maximum observed intensity.) For all sources, a b value of 0.5 was used.

RESULTS

Figure 9 shows the annual risks (probabilities of equaling or exceeding a chosen MM intensity) for several sites and the various seismic source hypotheses. The risk-versus-intensity curve for the single seismic source of Figure 1 is the same for all sites; this results from the uniform seismicity assumed throughout the region.

The curves in Figure 9 have approximately the same slope. This results from the fact that the same b value has been used for all seismic sources. Slight differences in slopes in Figure 9 are due to different maximum observed intensities for the different seismic source hypotheses.

Defining a "design intensity" as the MM intensity corresponding to a chosen an-

nual risk level, the slope of the curves in Figure 9 provides some insight into the sensitivity of design intensities to the associated levels of risk. Changing the risk by a factor of 2 will change the corresponding design intensity by only about 0.4 intensity units. Thus, the effect of statistical variation in the occurrence rate for, say, seismicity defined by only five observations is to increase the risk by 20 per cent over that computed for a rate without accounting for statistical variation; the corresponding increase in the design intensity (at whatever risk level is chosen) is about 0.1 intensity units. For the source areas examined in this study, each area encompassed more than five earthquakes in the catalog. Thus statistical uncertainty in the recurrence rate has a very minor effect on the design intensity level. Similarly any other

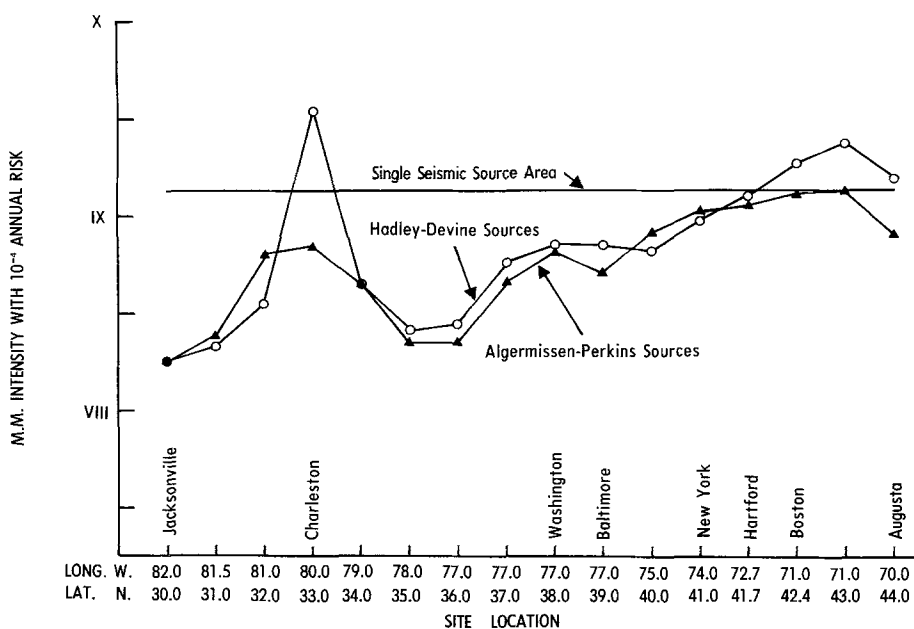


FIG. 10. MM intensity with 10^{-4} annual risk at the sites examined, under the three seismic source hypotheses.

variation of hypothesis which changes the calculated risk at a site by 20 per cent will change the corresponding design intensity by 0.1 units.

Figure 10 shows the MM intensities associated with a risk of 10^{-4} per year for the sites shown in Figure 1 and for the three seismic source hypotheses of Figures 1, 2, and 3. The most striking observation about this comparison is the closeness of intensities at most sites for the different hypotheses used. The range in intensities under the three hypotheses is less than 0.35 intensity units for the majority of sites, and in no case is it greater than 0.8 units. It is evident from Figure 9 that the range of intensities at a site for different seismic source hypotheses will not change for risk levels of 10^{-3} or 10^{-5} per year; the curves of Figure 10 simply shift up or down for smaller or larger risks, respectively.

It is useful to compare the results obtained from this hazard analysis with intensities derived using a deterministic procedure. Such a comparison is shown in Figure 11. The solid lines are the intensities with 10^{-4} annual risk for the two sets of discrete seismic sources, and are the same curves shown in Figure 10. The dashed lines are the

intensities derived from a deterministic procedure using the same sets of seismic sources. The procedure followed is that currently used to derive the safe shutdown earthquake for nuclear power plant design (Nuclear Regulatory Commission, 1973). This is, to calculate the maximum intensity for a site, use the largest intensity observed historically in the source in which the site lies, or use the largest intensity in nearby sources and attenuate the motion to the site using the shortest source-to-site

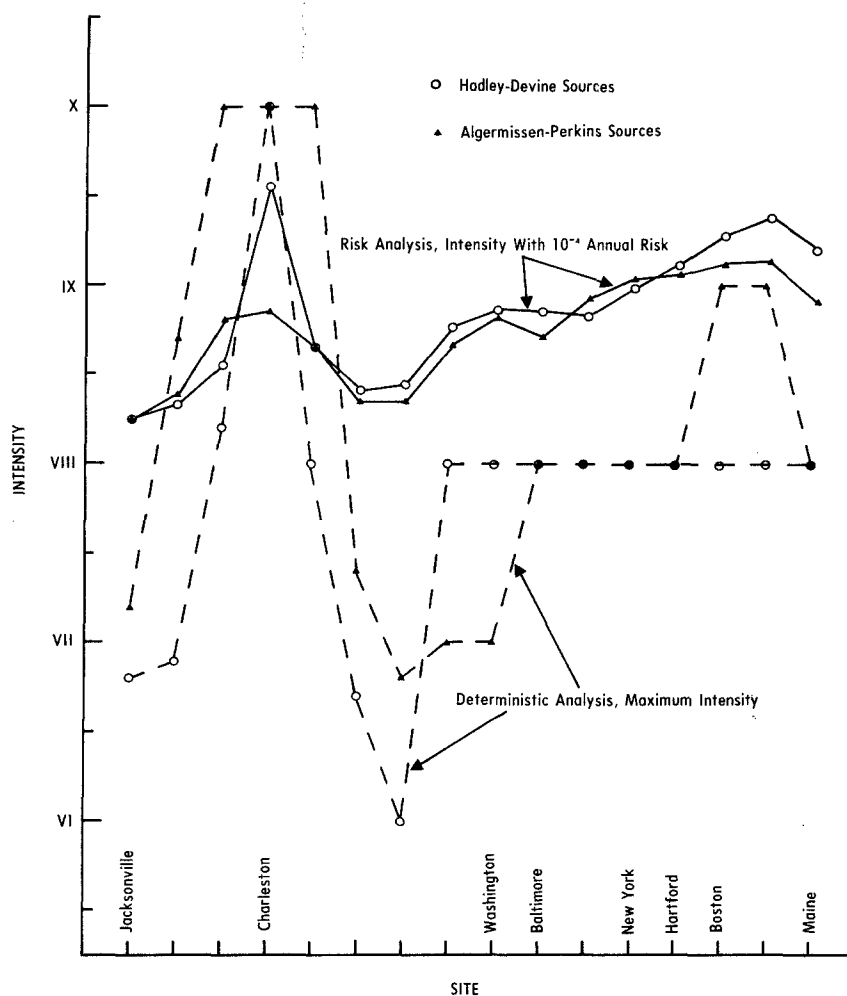


FIG. 11. Comparison of MM intensities from risk analysis with those from deterministic analysis, for the two seismic source hypotheses of Figures 2 and 3.

distance. (The attenuation used for this analysis was the Cornell and Merz curve of Figure 8.)

Several conclusions are immediately apparent from Figure 11. First, the risk associated with intensities from the deterministic procedure vary widely from site to site, from less than 10^{-4} per year to greater than 10^{-3} per year. Second, the intensities calculated deterministically are highly sensitive to the manner in which the seismic sources are drawn; differences of up to two MM intensity units are obtained for the two sets of sources considered. There is no such sensitivity in the results from the probabilistic procedure.

Figure 12 indicates MM intensities with 10^{-4} annual risk for various hypotheses on the maximum possible intensity. The results shown are for the Algermissen-Perkins seismic sources (Figure 2); results for the Hadley-Devine sources (Figure 3) are similar. Results for the hypothesis used previously in Figures 10 and 11, that is, that the maximum possible intensity in each source is a random variable with uniform distribution up to intensity XII, are repeated in Figure 12 as solid circles. If one assumes *with certainty* that intensity XII events will occur eventually in all seismic sources, the MM intensities with 10^{-4} annual risk which result are shown as squares. It is thought that this hypothesis is unduly conservative in that intensity XII events are not expected to occur eventually in *all* seismic sources. Thus the random maxi-

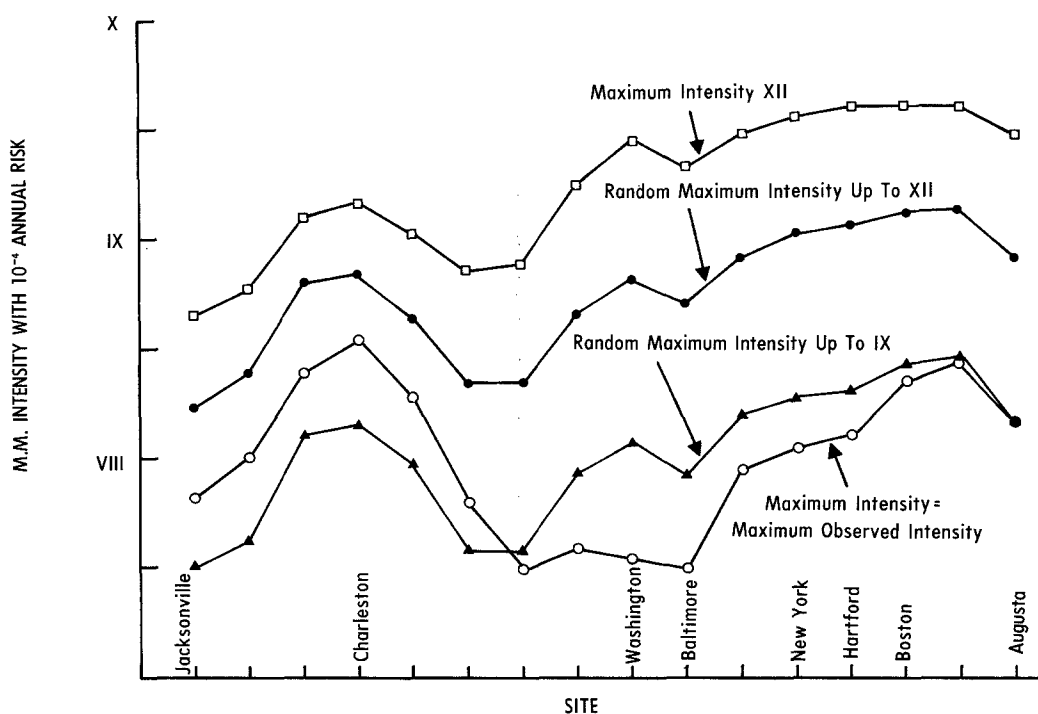


FIG. 12. MM intensities with 10^{-4} annual risk for various hypotheses on maximum possible intensity in each seismic source, using seismic sources of Figure 2.

imum intensity curve is more realistic, because it accounts for present uncertainty in determining which sources are capable of producing large events and which are not.

As discussed previously, one can assert that events with observed maximum MM intensity greater than IX are simply the results of geologic conditions near the epicenter rather than large amounts of energy release. To adopt this assertion for the hazard analysis, observed events of intensity X and above are down-graded to intensity IX, and the maximum possible intensity in each source is assumed random and uniformly distributed between the maximum observed intensity and IX. The results from this hypotheses are shown in Figure 12 as triangles. The conclusion is that, although the maximum credible intensity has been reduced three units (from XII to IX), the intensities with 10^{-4} annual risk are reduced by $\frac{3}{4}$ of a unit. (The reduction in ground acceleration associated with this change would be significant, however.)

The last set of points in Figure 12, the open circles, are intensities with 10^{-4} annual risk which result from the assumption that the maximum possible intensity in each seismic source is the largest intensity which has been observed in historical time. As discussed previously at length, such an assumption is not supported by statistical analysis of the available earthquake catalog for the Eastern United States.

The sensitivity of results to the value of b used in the hazard analysis is shown in Figure 13. The intensities at all sites vary within a range of about one unit, depending on the value of b used (as mentioned above, the b values of 0.45 and 0.57 represent reasonable limits on interpretation of the earthquake catalog). In determining MM

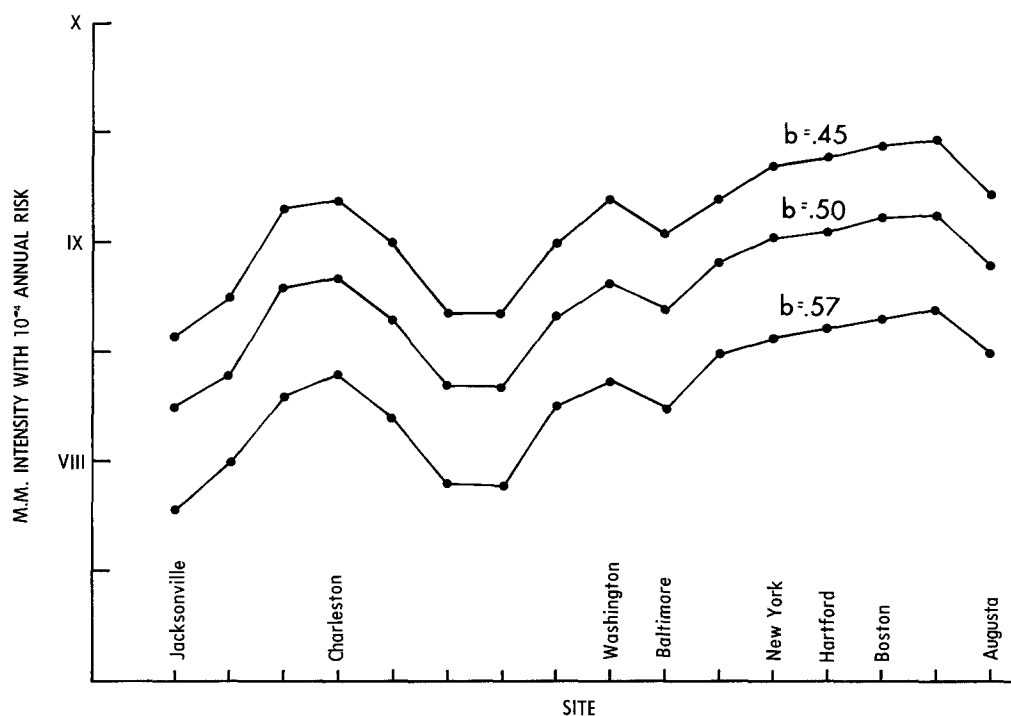


FIG. 13. MM intensities with 10^{-4} annual risk for various b values, using random maximum intensity up to XII and seismic sources of Figure 2.

intensities for design, the subjectivity in determining b could be avoided simply by specifying a single value (such as 0.5) to be used in all analyses.

DISCUSSION

The first observation from these results is that MM intensities for a chosen risk level are generally insensitive to the manner in which the region of study is divided into seismic sources. The reason for this insensitivity is that the various source hypotheses do not result in large changes in seismicity, as this seismicity affects the calculated hazard. The insensitivity of intensity levels to changes in risk is important; changing rates of occurrence by a factor of 2 (by using an alternate set of sources) will change risks at any intensity by a factor of 2, but will change design intensities (at a given risk level) by only 0.4 intensity units. (Note, however, that ground acceleration is often observed to vary in a logarithmic manner with MM intensity. Thus a change of one MM intensity unit may imply a doubling of ground acceleration.)

A compensating factor in the comparison between seismic source hypotheses is that the effect of high occurrence rates (per unit area) in the discrete seismic sources of Figures 2 and 3 (as compared to that of the single source in Figure 1) is partially offset by accounting for the possibility of maximum possible MM intensities less than XII in these discrete sources.

This insensitivity of design intensity to the sources used to define seismicity is not a conclusion of universal applicability for all sites in the Eastern United States. Design intensities will usually vary most at a site in a region which is considered highly seismic under one hypothesis and aseismic under another. In such cases, it is always possible to assign subjective probabilities to each hypothesis (e.g., Cornell and Merz, 1975) and to weigh the conditional risks to obtain the total risk at various intensity levels. This procedure will always give a design intensity which lies within the range of the intensities indicated by the individual hypotheses. Note also that the single seismic source of Figure 1 was drawn specifically with risk analysis of the east coast in mind. For analysis at other locations, different geometries for the single seismic source would likely be more appropriate. Thus the intensities and risks indicated by Figures 8 and 9 for the single seismic source are not necessarily applicable to sites other than those considered here.

An advantage of the probabilistic treatment of upper bound intensity is that the intensity associated with a chosen risk level at a site is insensitive to the maximum intensities observed in the hypothesized sources. For the sites examined here, and for the sources of Figures 2 and 3, altering the maximum observed intensity in each source by one intensity unit would change the intensity associated with an annual risk of 10^{-4} by approximately 0.1 intensity unit. Thus, arguments about the precise intensity of some historical earthquake which was the largest observed event in a source would have minor importance, since the design intensity would not be sensitive to this largest value. In contrast, under the deterministic procedure of designing for the largest observed intensity, changing this intensity by one unit will change the design intensity by one unit.

Figure 12 shows that the calculated intensities are sensitive to the assumed largest event which can occur. Reducing this largest event from intensity XII to IX reduces design intensities by $\frac{3}{4}$ of one unit, and would reduce corresponding ground accelerations by about 40 per cent. Thus, determination of the largest event possible in the Eastern United States, and the MM intensity associated with it, would be of great value. Similarly, limiting the maximum possible intensity in a seismic source on the basis of geologic or tectonic evidence would have important consequences for sites whose seismic hazard is governed by that source. This emphasizes the importance of continued research in understanding the earthquake process in the Eastern United States.

It should be emphasized that the attenuation function, equation (10), was obtained using all available intensity data from the Charleston earthquake (this is the only major event in the Eastern United States for which a large data base is available). Hence the results presented represent intensities for average soil conditions of the sites from which intensity reports from 1886 were available. There is presently no generally accepted method of modifying intensities on average site conditions to predict those on firm or loose materials. It is expected that any such modification would affect the intensities of Figure 11 in a similar manner, so that conclusions from the comparison between probabilistic and deterministic results would not change.

The MM intensity with 10^{-4} annual risk at a site is often called the intensity with a

10,000-year return period. (In fact this illustrates the definition of return period.) These intensities (from Figure 11, for instance) are *not rare events* for the entire Eastern United States. In fact, intensity IX or larger events occur on the average about once every 30 years in the Eastern United States; this can be seen in Figure 5. Thus the probabilistic analysis described here, which indicates annual risks of 10^{-4} for intensities VIII or IX (depending on the site) on the east coast does not involve extrapolation beyond the available seismic data. What is indicated, roughly speaking, is that a common event which occurs on the average once every 30 years in a large region, has a small annual probability of occurring at a specific location in that region.

CONCLUSIONS

It is evident from this study that, using a probabilistic analysis which accounts for uncertainty in activity rates and maximum possible intensities of seismic sources, the MM intensity associated with a chosen level of risk at many sites on the east coast of the United States is not sensitive to the sources used to represent seismicity. This allows many alternate methods to be used to combine geology, tectonics, and seismic history to draw seismic sources, without greatly affecting design intensities at many sites. For other sites in the Eastern United States, the geometry of sources may be critical; this needs further investigation.

As few as five or ten earthquake observations are adequate to define the occurrence rate in an area, for the purpose of deriving risk-associated design intensities. The seismic history of the Eastern United States is *not*, however, adequate to define accurately the maximum possible intensity in chosen source areas. In the absence of geologic or tectonic evidence which would limit the size of events in the area, it is logical and conservative to assume a uniform distribution on the maximum possible MM intensity, between the maximum observed value and XII. The b value describing the relative frequency of small and large events cannot be determined unequivocally, because of different possible interpretations of seismic history and different methods of numerical analysis. It is evident, however, that variations in b from area to area in the Eastern United States can be attributed to statistical variations resulting from small numbers of events available to make the determinations. Thus it is reasonable, for the purpose of calculating design intensities, to adopt a single value of b for the Eastern United States.

In comparing the results of probabilistic analyses with those of the deterministic method of calculating intensities presently used for seismic design of nuclear power plants in the United States, several conclusions are apparent. First, the risks associated with the deterministic intensities vary by more than a factor of ten, for the sites considered on the east coast. Also, these intensities are very sensitive to the geometry of sources used to describe seismicity, and to the maximum intensity which has been observed in historical times in each source. The deterministic procedure is not sensitive to rates of activity in seismic sources, nor to the b value, because these parameters are not considered in the analysis.

This study places in perspective many of the objections to using probabilistic analysis to determine seismic design levels. In general there is sufficient information, in terms of seismic history and geology, to produce seismic design values at very small levels of risk, given the present understanding of the physical processes governing earthquake occurrences in the Eastern United States. Furthermore the results obtained here provide a means of evaluating the conservatism and consistency (in

terms of risk) of deterministic analyses which assume that the maximum epicentral intensity possible in an area is the maximum intensity which has been observed in recorded history.

ACKNOWLEDGMENTS

The author would like to acknowledge the help of David M. Perkins in assessing methods of determining b values for seismic sources, and for providing useful criticism throughout this study.

REFERENCES

- Algermissen, S. T. (1969). Seismic risk studies in the United States, *Proc. World Conf. Earthquake Eng., 4th, Santiago, Chile*, 1, 14-27.
- Algermissen, S. T. and D. M. Perkins (1976). A probabilistic estimate of maximum ground acceleration in the contiguous United States, *U.S. Geol. Surv. Open-File Rept.* 76-416.
- Benjamin, J. (1968). Probabilistic models for seismic force design, *J. Struct. Div., Am. Soc. Civil Eng.* 94, 1175-1196.
- Benjamin, J. R. and C. A. Cornell (1970). *Probability, Statistics, and Decision for Civil Engineers*, McGraw-Hill, New York.
- Bollinger, G. A. (1977). Reinterpretation of the intensity effects of the 1886 Charleston, South Carolina, earthquake in Geological, Geophysical, and Seismological Studies Related to the Charleston South Carolina Earthquake of 1886; A Preliminary Report, D. W. Rankin, Editor, *U.S. Geol. Surv. Profess. Paper 1028* (in press).
- Chinnery, M. A. and D. A. Rogers (1973). Earthquake statistics in southern New England, *Earthquake Notes*, 45, 89-103.
- Cornell, C. A. (1968). Engineering seismic risk analysis, *Bull. Seism. Soc. Am.* 58, 1583-1606.
- Cornell, C. A. (1971). Probabilistic analysis of damage to structures under seismic load, Chapter 27 in *Dynamic Waves in Civil Engineering*, D. A. Howells, I. P. Haigh, and C. Taylor, Editors, Wiley Interscience, London, 473-488.
- Cornell, C. A. and H. A. Merz (1975). Seismic risk analysis of Boston, *J. Struct. Div., Am. Soc. Civil Eng.* 10, 2027-2043.
- Esteva, L. (1969). Seismicity prediction: A Bayesian approach, *Proc. World Conf. Earthquake Eng., 4th, Santiago, Chile* 1, 172-184.
- Esteva, L. (1974). Geology and probability in the assessment of seismic risk analysis, *Proc. Intern. Congress Intern. Assoc. Eng. Geologists*, 2d, Sao Paulo, Brazil.
- Gardner, J. K. and L. Knopoff (1974). Is the sequence of earthquakes in southern California, with aftershocks removed, Poissonian, *Bull. Seism. Soc. Am.* 64, 1363-1368.
- Gupta, I. N. and O. W. Nuttli (1976). Spatial attenuation of intensities for Central U.S. earthquakes, *Bull. Seism. Soc. Am.* 66, 743-751.
- Hadley, J. F. and J. F. Devine (1974). Seismotectonic map of the Eastern United States, *U.S. Geol. Surv. Misc. Field Study MF-620*.
- Howell, B. F., Jr. and T. R. Schultz (1975). Attenuation of Modified Mercalli intensity with distance from the epicenter, *Bull. Seism. Soc. Am.* 65, 651-665.
- Lomnitz, C. (1966). Statistical prediction of earthquakes, *Rev. Geophys.* 4, 3.
- Lomnitz, C. (1973). Poisson processes in earthquake studies, *Bull. Seism. Soc. Am.* 63, 735.
- McGuire, R. K. (1974). Seismic Structural Response Risk Analysis, Incorporating Peak Response Regressions on Earthquake Magnitude and Distance, *Massachusetts Institute Technology Dept. Civil Eng. Rept. R74-51*, August.
- McGuire, R. K. (1976). Fortran computer program for seismic risk analysis, *U.S. Geol. Surv. Open-File Rept.* 76-67.
- Nuclear Regulatory Commission (1973). Part 100, Reactor Site Criteria, *Federal Register* 38 (218), 31279-31285.
- Raiffa, H. and R. Schlaifer (1961). *Applied Statistical Decision Theory*, Massachusetts Institute Technology Press, Cambridge.
- Richter, C. F. (1958). *Elementary Seismology*, Freeman, San Francisco.
- Shah, H. C., C. P. Mortgat, A. Kiremidjian, and T. C. Zsutty (1975). A Study of Seismic Risk for Nicaragua, *Rept. 11, Dept. of Civil Eng.*, Stanford Univ., Stanford, California.
- Veneziano, D. (1974). Statistical Estimation and Prediction in Probabilistic Models, with Ap-

plication to Structural Reliability, *Ph.D. Thesis*, Dept. Civil Eng., Massachusetts Institute Technology, September.

Veneziano, D. (1975). Probabilistic and Statistical Models for Seismic Risk Analysis, *Massachusetts Institute Technology Dept. Civil Eng. Rept. R75-34*, July.

U.S. GEOLOGICAL SURVEY
DENVER FEDERAL CENTER
Box 25046, Stop 978
DENVER, COLORADO 80225

Manuscript received August 6, 1976