Bulletin of the Seismological Society of America, Vol. 72, No. 6, pp. S43-S60, December 1982

THE EMPIRICAL PREDICTION OF GROUND MOTION

By David M. Boore and William B. Joyner

ABSTRACT

Recent additions to the strong-motion data set, primarily from earthquakes in California and Italy, are responsible for a large number of papers examining the prediction of ground-motion measures using regression methods. Peak acceleration is still the most common measure being considered, but increasing attention is being given to peak velocity and spectral amplitudes. Although direct comparisons among the studies are hampered by differing definitions of distance and magnitude, in general the various studies give similar answers for peak acceleration in the region of distance and magnitude space in which most of the data are concentrated. As might be expected, the differences are most pronounced for large magnitudes and distances close to the fault, where data are few. Even so, widely differing assumptions about the form of the regression equation and differences in the composition and weighting of the data set can give similar answers. This was true in recent studies by Campbell (1981b) and Joyner and Boore (1981), where the predicted accelerations for large earthquakes at close distances differed by less than 40 per cent. This seemingly large uncertainty is small compared to the scatter in the data about the regression lines. A Monte Carlo study shows that the question of whether the shape of the attenuation curves is magnitude-dependent cannot be resolved by existing data.

Introduction

The old problem of predicting ground motions from earthquakes has been given new life by the demands of engineers and the increasing amount of data from within a few tens of kilometers of faults. This paper is a systematic review of the considerations needed in using the improved data set to derive curves or equations that can be used to make empirical predictions of strong ground motion. It is organized following the questions that are usually asked in deriving predictions: What will be the dependent and independent variables? What set of data will be used? What will be the form of the prediction equation (if any)? What analysis procedure will be used? Is the equation adequate? These questions may be obvious, but their answers are vital, especially when comparing various studies. For convenience, our review is illustrated by results and examples in a number of papers appearing in the December 1981 issue of the *Bulletin.* In so doing, it is unavoidable that comparisons be made of the predictions of the various studies. Our intention, however, is to illuminate the issues involved and the logic used in making predictions rather than to pass judgment on the predictions. Detailed reviews of the many previous attempts to predict ground motion are not given here; a comprehensive summary is given by Idriss (1978). Because it is outside our area of expertise, we also give only brief attention to the important problem of predicting ground motions in areas such as the Central and Eastern United States; for this, see recent papers by Battis (1981), Campbell (1981a), and Nuttli and Herrmann (1981).

Studies Used for Comparisons

Of the group of papers dealing with strong ground motion in the *Bulletin's* December 1981 issue, five are of primary concern to us because they presented empirical predictions of ground motion. Hasegawa *et al.* (1981) presented attenuation relations for strong motion in Canada. Herrmann and Goertz (1981) were primarily concerned with a numerical study of ground motion scaling, but they did present prediction equations for the Central and Western United States. The attenuation equations in these two studies are based primarily on Murphy and O'Brien (1977). Joyner and Boore (1983) and the similar study of Joyner *et al.* (1981a) derived prediction equations using data from earthquakes in western North America. Campbell's (1981b) study is similar in scope, but he included data from significant earthquakes in other parts of the world. Both Joyner and Boore and Campbell included data from the 1979 Imperial Valley earthquake in their studies. Finally, Chiaruttini and Siro (1981) gave prediction equations from European strong-motion data.

VARIABLES

Dependent variables. The first thing to be decided is what variable is to be predicted. Peak ground acceleration is the usual choice, but a number of other descriptors of ground motion have been used recently. Among these are rms acceleration (Hanks and McGuire, 1981; McCann and Boore, 1982), peak ground

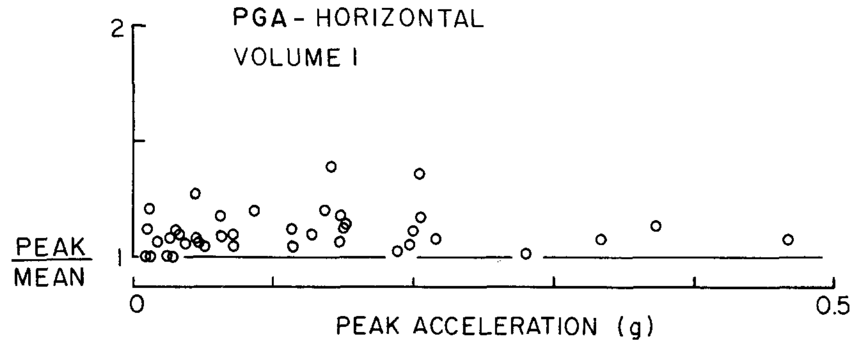


FIG. 1. Ratio of larger peak ground acceleration (PGA) on either horizontal component to mean of the two peaks. Points correspond to those recordings in the data set of Joyner and Boore (1981, Table 2) that are contained in volume 1 of the series "Strong Motion Earthquake Accelerograms," published under the direction of D. E. Hudson by the Earthquake Engineering Research Laboratory of the California Institute of Technology.

acceleration (Hanks and McGuire, 1981; McCann and Boore, 1982), peak ground
velocity (Joyner and Boore, 1981b), response spectra of various sorts (McGuire, 1974; Trifunac and Anderson, 1978; Joyner and Boore, 1982), and Fourier spectra (Trifunac, 1976; McGuire, 1978a). Because ground motions usually are recorded on three orthogonal components, a further needed decision is how to treat these three components. The motions in the horizontal plane are of greater engineering significance than those in the vertical direction, and therefore most studies have dealt with horizontal motions only; three approaches are common: (1) use the larger of either horizontal component; (2) use both components; or (3) use the mean of the estimates from both components. For peak horizontal acceleration, the first approach gives numbers that are systematically larger than the third (and presumably the second) by about 10 per cent (Figure 1).

Independent variables. Ground-motion predictions are almost always a function of the independent variables earthquake size and distance to the source. Geologic conditions are sometimes considered as well.

Although the measure of earthquake size is universally expressed by earthquake magnitude, the diversity of magnitude scales (Figure 2) can lead to confusion in comparing various predictions. There is a clear tendency for all scales except moment magnitude to reach a limiting value (saturate) as the size of the earthquake increases. Because most magnitudes are based on the peak amplitude of an instrumental recording, one might expect a good correlation of the ground-motion variable of interest with the instrumental recording having similar frequency content. For example, Boore (1980) found a strong correlation between peak velocity and peak amplitude of a Wood-Anderson instrument. This correlation, however, might not hold for earthquakes with moment magnitudes less than about 5 or greater than 7. A detailed discussion of this and other correlations is beyond the scope of this paper; we point out, however, that complications arise because of the broadband character of ground motion compared with the sometimes narrow-band instrument output, and because of the stochastic nature of large, extended ruptures (McGuire and Hanks, 1980; Hanks and McGuire, 1981). Whatever scale is used, it is important to state the choice and be consistent in its use. We prefer moment magnitude (Hanks

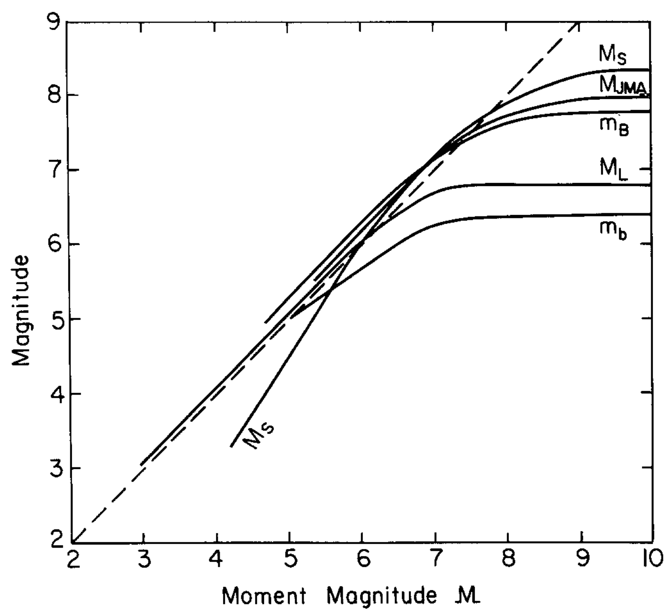


FIG. 2. Relation between moment magnitude and various magnitude scales: M_S (surface wave); m_b (short-period body wave); M_B (long-period body wave); M_L (local); and M_{JMA} (Japan Meteorological Agency). Dashed line shows a 1:1 relation for reference. Figure adapted from Heaton *et al.* (unpublished data, 1982).

and Kanamori, 1979) because it corresponds to a well-defined physical property of the source. Furthermore, the rate of occurrence of earthquakes with different moment magnitudes can be related directly to the slip rates on faults (Brune, 1968; Anderson, 1979; Molnar, 1979).

After the size is specified, the next independent variable is the distance from the source to the station. Because rupture surfaces for earthquakes can extend over tens of kilometers, a number of distance measures are in use (Figure 3). The measure used should depend on the application. The most common case, and the one of concern to us, is the use of previous data to predict motions from future earthquakes. In many applications, the fault can be identified but the hypocenter or sources of particularly energetic radiation cannot. In these cases, a measure based on the closest distance to the fault ($M4$ or $M5$ in Figure 3) seems reasonable. Some have argued that using a closest distance measure can lead to biased predictions, especially if the strong-motion stations have a nonrandom distribution around the fault and if the peak motions come from one small source on the fault (Shakal and Bernreuter, 1981; M. D. Trifunac, oral communication, 1982). These conditions are not usually met; even if they were, however, the randomization introduced by considering an ensemble of earthquakes would help eliminate any bias. (In our view, the placement of recording instruments and structures are comparable sampling processes from the statistical point of view.) Of course, if the peak motions are radiated by a small area of the overall fault, using a shortest distance measure might give a distorted view of the attenuation of the motions from a particular event. An example of such distortion is shown in Figure 4. The decrease in attenuation near the fault shown in

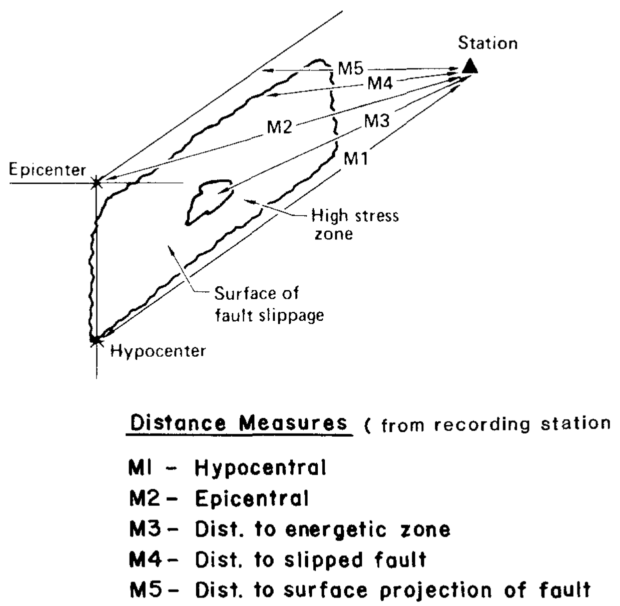


FIG. 3. Distance measures used in prediction equations (from Shakal and Bernreuter, 1981).

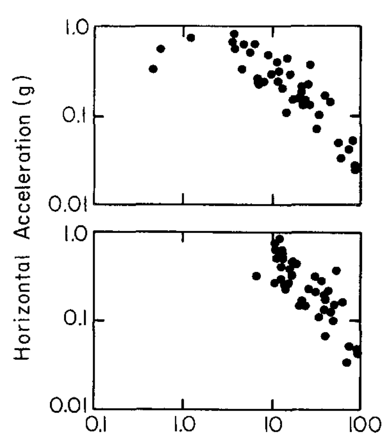


FIG. 4. Horizontal acceleration data from the 1979 Imperial Valley earthquake plotted against shortest distance to fault (*top*) and distance to a point along the Imperial fault about 16 km northwest of the epicenter (*bottom*, both from Shakal and Bernreuter, 1981).

the *top box* of Figure 4 has been predicted by some (e.g., Schnabel and Seed, 1973) as an effect of a change in geometric spreading from $1/R$ at large distances to 1 when close to the extended rupture surface. On the other hand, Shakal and Bernreuter (1981) showed some evidence that the peak motions were not radiated uniformly over the whole fault plane. When the data are plotted against the distance to their suggested source (Figure 4, *bottom box*), the decrease in attenuation rate disappears. We believe that attenuation curves based on the shortest distance to the fault will show flattening at close distances; this flattening is not due to a fundamental change in geometric spreading, however, but is a result of the largest motions coming from stress release at depth, even if the fault breaks the surface. In effect, the source of motions is always at some distance from the recording stations.

Another independent variable sometimes used in prediction studies is a measure of geologic material at the site. The usual measure is a binary classification into rock or soil. More refined methods for characterizing the site, such as using local shear-wave velocities and thickness of sediments for soil sites, might reduce the variance in the predicted motions (Joyner *et al.*, 1981a; Rogers and Tinsley, 1982).

Data Selection

Once the dependent and independent variables have been selected, a subset of the several thousand strong-motion records must be chosen for analysis. The choices usually center around avoiding biases and decreasing the scatter in the prediction equations. Biases and scatter can arise for many reasons. Among them are: including data from various tectonic provinces that may differ in rate of attenuation or source properties; differences in record processing (Figure 5); instrument depth and soil-structure interaction at the recording site (Figure 6); geologic effects not accounted for explicitly (Figure 7); azimuthal effects due to radiation pattern and source directivity; and poor knowledge of the location of the rupture surface.

The set of data finally chosen can be conveniently displayed in magnitude-distance space (Figure 8). In general, the distribution of data is not uniform; small earthquakes do not trigger distant recorders, and, more important, few recordings have been obtained close to large earthquakes.

The Regression Model

Having chosen the data, the usual next step is to fit a model to the data. Why choose a model at all? Clearly, there would be no reason to if the requirement were simply to predict motions in a region of magnitude-distance space for which data were abundant. This is not usually the case, however. Predictions are often required where data are not available, e.g., close to large earthquakes. One strategy is to fit a functional form to the data, then use the resulting equation to predict the motions. The success of this strategy depends on the functional form's adequacy in describing the physics of the problem. There are other reasons for fitting a mathematical relation to the data even if magnitude-distance space is well represented by data. The resulting equations provide a convenient summary of the data for use in computer programs (e.g., when dealing with probabilistic ground-motion assessment) and are also useful in understanding the physical processes controlling the variations of ground motion (e.g., scaling with size or comparing attenuation rates in various tectonic provinces).

The general form chosen by the various authors in the December 1981 *Bulletin* is

$$Y = b_1 e^{b_2 M}[e^{b_3 D}/D^{b_4}]e^{b_5 S}e^{b_6 P} \tag{1}$$

in which Y is the dependent variable, M a magnitude, D a function of the distance measure, S a binary variable representing local site geology (0 if rock, 1 if soil), and P the uncertainty in the prediction (0 and 1 for 50 and 84 percentile values, respectively). The b's are parameters that must be determined from regression analyses. The physical motivation for the chosen form is as follows: the exponential dependence on magnitude stems from the basic definition of magnitude as a

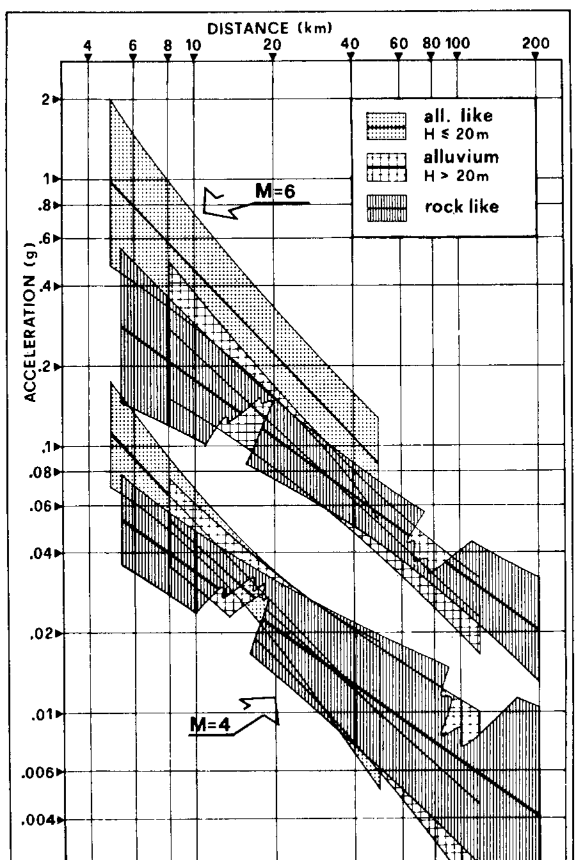


FIG. 7. Attenuation relations for data from Friuli, Italy, showing the tendency for motions on shallow soil sites to be larger than on either rock or deep soil sites (from Chiaruttini and Siro, 1981). *Cutaways* expose patterns on obscured fields.

logarithm of a measure of ground motion; the distance dependence in brackets accounts for anelastic attenuation (b_3) and geometrical spreading (b_4); the soil term is arbitrary, but agrees with the notion that site effects should be multiplicative; finally, the uncertainty follows from the assumption of a log-normal distribution of the observations about the regression line. Although the most common, it must be emphasized that equation (1) is not the only form possible. For example, Idriss (oral communication, 1982) proposes a different relation that may provide a better fit to data from a wider range of magnitudes than is usually considered (specifically,

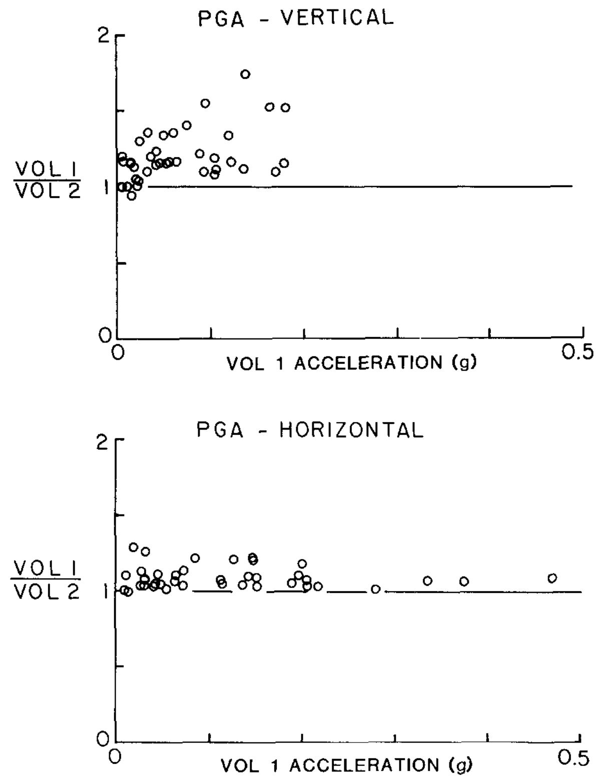


FIG. 5. Ratio of peak ground accelerations (PGA) from volumes 1 and 2 of the series referenced in Figure 1 for vertical and horizontal components. Volume 1 data are unprocessed and digitized at unequal spacing (including peaks and troughs); volume 2 data have been interpolated to 0.02-sec spacing and corrected for instrument response. The bias is probably unimportant for frequencies less than about 10 Hz but is increasingly important as frequency increases (compare the higher frequency vertical data to the horizontal data).

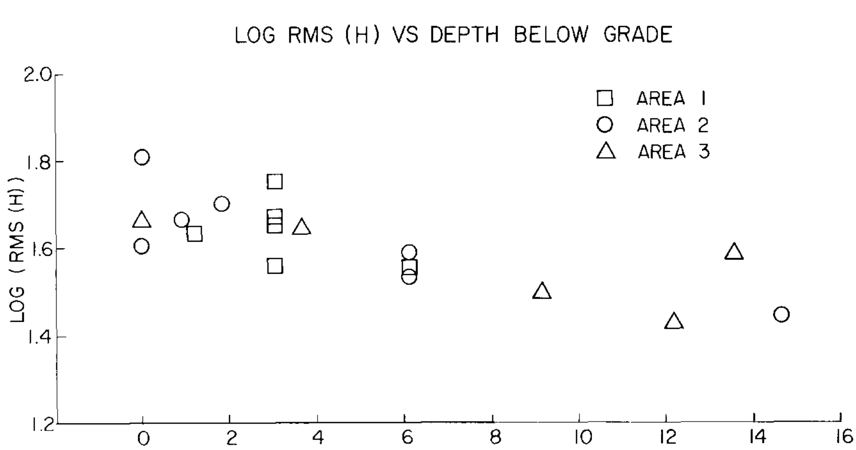


FIG. 6. The dependence rms acceleration on embedment depth of instrument for data recorded in three small areas in Los Angeles during the 1971 San Fernando earthquake (from McCann and Boore, 1983).

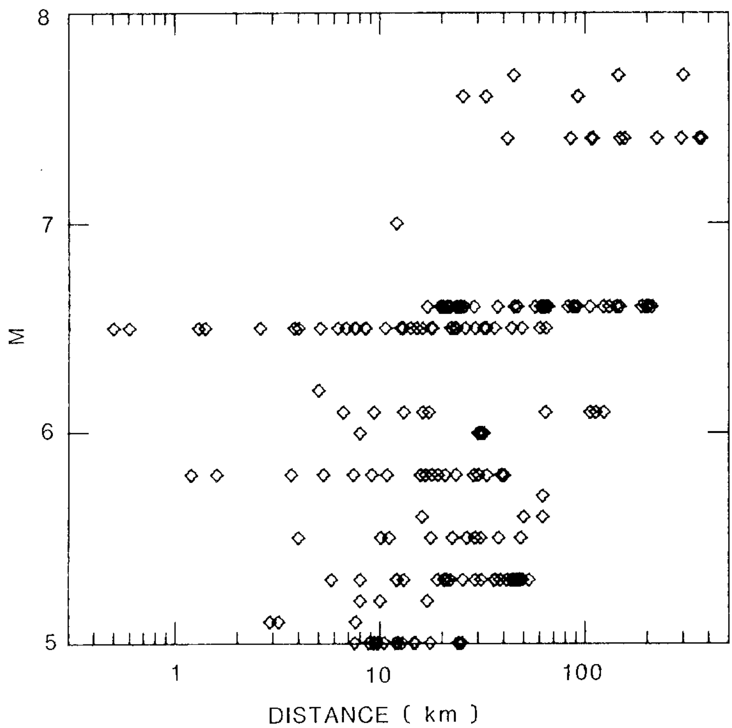


FIG. 8. Distribution in magnitude and distance of the peak horizontal acceleration data used by Joyner and Boore (1981). Each point represents one recording.

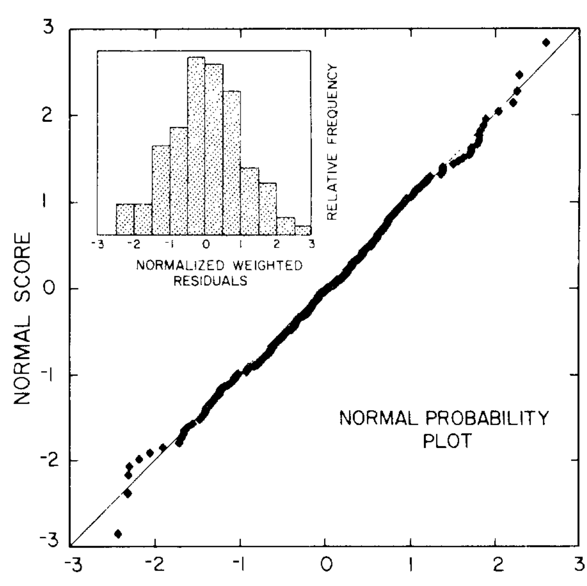


FIG. 9. Normal probability plot of residuals of the logarithms of peak acceleration around the prediction equation of Campbell (1981b); this evidence for a log-normal distribution of peak ground acceleration is confirmed by a Kolmogorov-Smirnov test at the 90 per cent confidence limit (after Campbell, 1981b).

earthquakes for magnitudes less than 5.0). The actual model used in the analysis is given by taking the logarithm of equation (1)

$$\log Y = c_0 + c_M M - c_D D(M) - c_{LD} \log D(M) + c_S S + c_P P \qquad (2)$$

where the c's are the coefficients to be determined. (Not all of these coefficients are used by all of the authors of the papers in the *Bulletin*.) The distance function D may contain coefficients, some of which may depend on magnitude. If it does not,

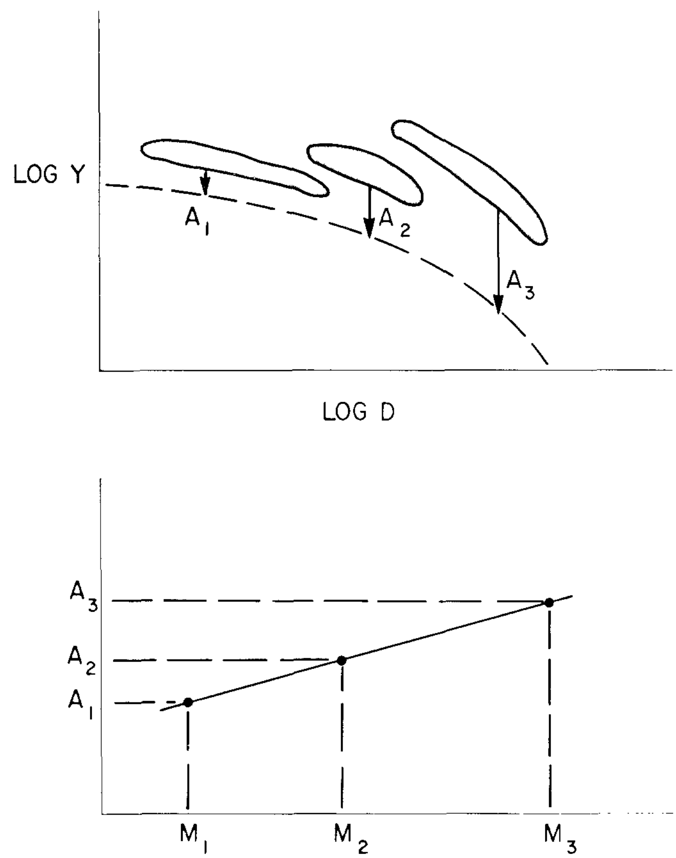


FIG. 10. Schematic showing the two-step regression procedure of Joyner and Boore (1981). The *top* figure shows the first regression, in which the shape of the dashed curve is varied, and all the data points from the ith earthquake (contained within each enclosed area) are shifted by a distance A_i so that the sum of the squares of the residuals is minimized. The *bottom* figure shows the second regression of the offset factors versus magnitude. An effect of the two-step procedure is to let each recording have equal weight in determining the shape, and each earthquake equal weight in determining the magnitude scaling.

equation (2) provides a linear equation for the unknowns. This, of course, is one reason for the form chosen. A more important reason is that, in a general way, the form satisfies physical ideas regarding the scaling of ground motion with magnitude and distance. Some *a posteriori* justification for fitting the equation to the logarithm of the dependent parameter comes from the log-normal distribution of the residuals of the data about the predictions (Figure 9). Although a normal distribution is not required for determining the coefficients through least-squares procedures, it has the convenience of allowing confidence limits to be assigned.

Another choice in deriving prediction equations involves the details of how the data are to be used in determining the unknown coefficients in equation (2). A potential for bias exists for two reasons: first, the data are not uniformly distributed in magnitude-distance space; and second, they may be dominated by many recordings from a few earthquakes. It is possible to restrict the data sample further so that no more than a certain number of data points come from a given earthquake and a given recording site (McGuire, 1978b). Campbell (1981b) used a weighting scheme to reduce the bias, and Joyner and Boore (1981) used a two-step regression in an attempt to separate the attenuation and magnitude scaling (Figure 10). Other types of bias—e.g., that due to using a lower threshold acceleration in making up the data set—are discussed by Toro (1981) and Cornell (1982).

TABLE 1

DEFINITION OF DEPENDENT AND INDEPENDENT VARIABLES

Study	Peak Acceleration* (Y')	Magnitude (M)	Distance (D)
Hasegawa *et al.* (1981)	Not specified	Not specified	Hypocentral
Herrmann and Goertz (1981)	Not specified	m_b	Epicentral
Chiaruttini and Siro (1981)	Larger	M_L	Hypocentral
Joyner and Boore (1981)	Larger	M	$[d^2 + (h_1 e^{h_2(m-6)})^2]^{1/2}$, d = shortest distance to surface projection of rupture surface ($M5$ in Figure 3)
Campbell (1981b)	Mean	M_L, M_S†	$R + h_1 e^{h_2 M}$, R = shortest distance to rupture surface ($M4$ in Figure 3)

* In reference to the peaks on each of the two horizontal accelerograms.

† Campbell uses M_L if both M_L and M_S are less than 6.0 and M_S if both are greater than 6.0. Procedure is unspecified when the two measures are not both above or both below 6.

TABLE 2

COEFFICIENTS OF REGRESSION EQUATION FOR PEAK GROUND ACCELERATION (IN g)

Study	c_0	c_M	c_D	C_{LD}	h_1 (km)	h_2	c_P
Hasegawa *et al.* (1981)	−1.99	0.55	—*	1.5	—	—	—
Herrmann and Goertz (1981)†	−3.52	0.53		0.0			
	−2.32	0.53	—	1.02	—	—	—
Chiaruttini and Siro (1981) (Friuli—all sites)	−2.0	0.39	—	0.9	—	—	0.24
Joyner and Boore (1981)	−1.02	0.249	+0.00255	1.0 (fixed)	7.3	0.0 (fixed)	0.26
Campbell (1981b)	−1.80	0.377	—	1.09	0.0606	0.7	0.16

* —, coefficients not used or not given.

† First set of coefficients for $R \leqq 20$ km, second set for greater distances.

COMPARISON OF RESULTS

The aforementioned comments apply to any definition of dependent variable, but for the sake of comparison, the rest of the paper will deal with peak ground acceleration. The authors of the papers in the December 1981 *Bulletin* used a number of different definitions for the independent variables. These are summarized in Table 1; Table 2 contains the coefficient values for the various prediction equations. Three of the five studies used the same model for the regression equation (Table 2). Campbell's model was similar to that used in the first three papers, but his definition of distance contained two more coefficients to be determined. Joyner and Boore's model was the most dissimilar, but even it bore some strong resemblances to the other four.

Because the authors of the papers did not use the same definitions for the various variables, a graphical comparison requires certain assumptions in order to reconcile these differences. This has been done in the comparisons shown in Figure 11. (Not included are predictions from Herrmann and Goertz, which are not meant for western North America.) The hypocenter is assumed to occur at 7-km depth below the closest point of the surface projection of the rupture surface to the station; with one exception, magnitudes have been converted to moment magnitude using the relations in Figure 2. The exception is the paper by Hasegawa *et al.,* in which the type of magnitude was not specified; in that case no conversions were made. The

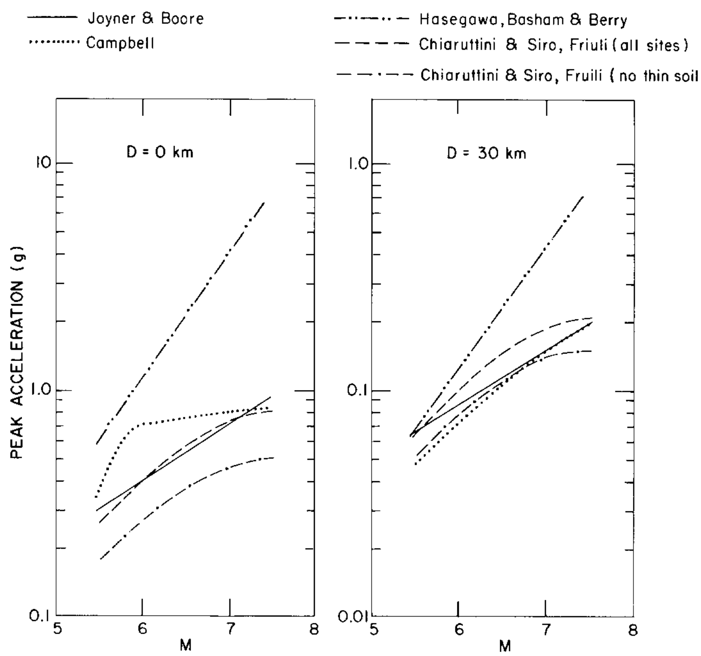


FIG. 11. Comparison of peak horizontal acceleration predictions for various papers in the December 1981 issue of the *Bulletin of the Seismological Society of America.* For the Campbell prediction, the *top edge* of the fault was assumed to be at 2.5-km depth for $M = 5.5$ and to break the surface for $M = 6.0$. See the text for other assumptions made (see Swanger *et al.*, 1980 for a similar figure comparing earlier predictions).

predictions based on using the peak of both horizontal components simultaneously (or the mean of both peaks) have been increased to account for the difference shown in Figure 1. The results show that the predictions agree with one another better at 30 km than they do at 0 km. This reflects the relative lack of data at close distances; at these distances, the predictions are quite dependent on the model, especially for the larger magnitudes. The reason for the systematically high predictions of Hasegawa *et al.* is not clear. They relied heavily on Murphy and O'Brien (1977), who included intensity in their regressions, but Hasegawa *et al.* converted intensities to accelerations. Cornell *et al.* (1979) showed that because of the poor correlations of intensity with acceleration, the use of an intermediate regression can bias the final prediction equation.

Another comparison of various predictions, shown in Figure 12, puts into per-

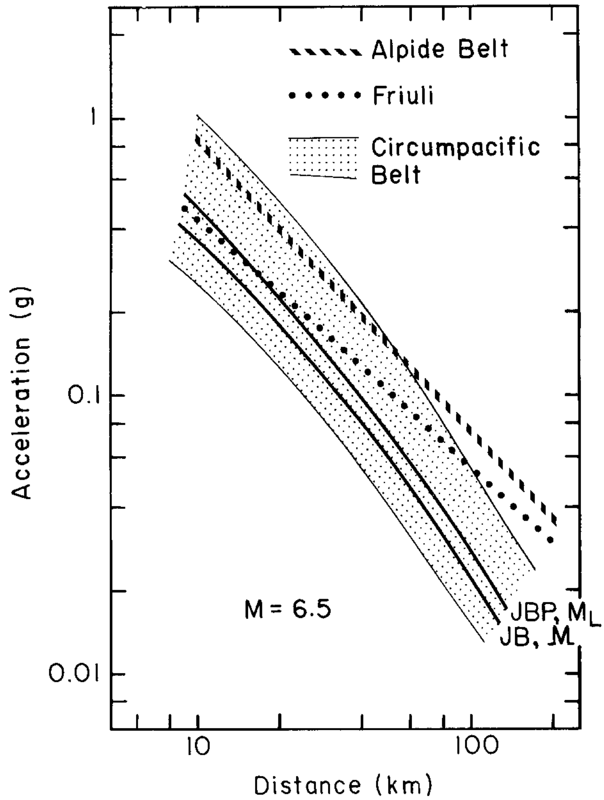


FIG. 12. Predictions of peak horizontal acceleration for the Friuli and Alpide Belt areas (Chiaruttini and Siro, 1981), western North America (JB, Joyner and Boore, 1981, using moment magnitude and JBP, Joyner *et al.*, 1981a, using local magnitude), and an envelope covering predictions for the Circumpacific Belt made by a number of authors (as given in Idriss, 1978). This figure was adapted from Chiaruttini and Siro (1981). As used by Chiaruttini and Siro, "Alpide Belt" seems to be loosely defined as the mountainous region extending from Italy (excluding Friuli) to Kashmir.

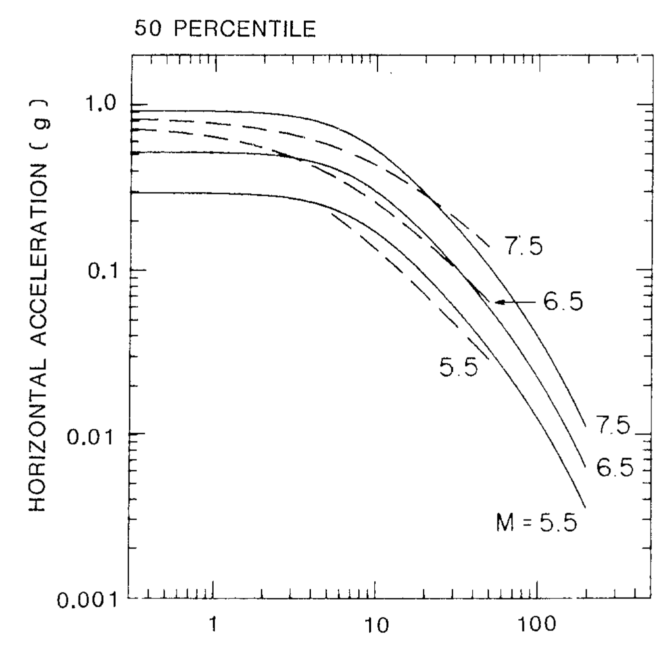
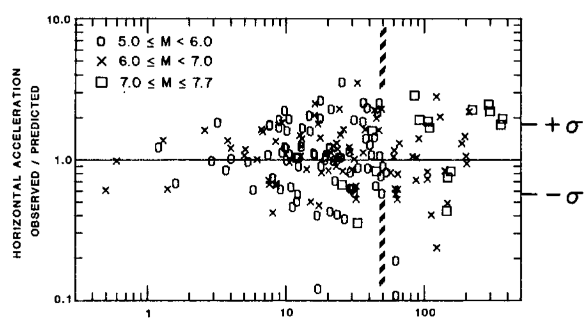


FIG. 13. Comparison of attenuation curves for peak horizontal acceleration by Campbell (1981b) (dashed lines) with the 50 percentile curves by Joyner and Boore (1981) (solid lines). Campbell's curves are raised by 13 per cent to compensate for the fact that he took horizontal acceleration as the mean of the peaks on the two horizontal components. Magnitudes at *right ends* of curves.

spective the recent predictions with respect to those based on older studies (which would be contained within the band labeled "Circumpacific Belt"). It also indicates that motions in some tectonic provinces may be systematically higher than in others ("Alpide Belt" versus Friuli). Unfortunately, the validity of the Alpide Belt results is clouded by including distances and magnitudes of uncertain accuracy in the analysis [although Chiaruttini and Siro (1981) did make a special effort to use a consistent magnitude scale]. Furthermore, at least two of the accelerations in the Alpide data set came from vertical components. These also happened to be the largest accelerations in the data set; the proper values should have been: Karakyr, 0.8 *g* and Naghan, 0.95 *g* (rather than 1.3 and 1.08 *g*, respectively; Ambraseys, 1978).

Finally, a comparison of Campbell's and Joyner and Boore's predictions show



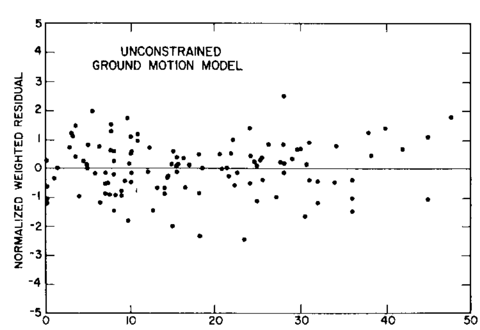


FIG. 14. Residuals of data about predictions: (*top*) from Joyner and Boore (1981); (*bottom*) from Campbell (1981b). Heavy dashed line in *top graph* marks 50-km distance from fault for ease in comparing with *bottom* residuals, which extend out to 50 km only. Campbell's residuals have been normalized by the standard deviation (0.16 log units).

them to be quite similar, in spite of many differences in procedures and data set (Figure 13). The largest deviation between the curves corresponds to a factor of 1.38, less than one standard deviation of the data about the regression line. In spite of the similarity of the absolute predictions, there are fundamental differences in the shapes of the curves; Campbell's have less curvature, and they depend on magnitude in such a way that the offset between the curves decreases as distance decreases. Joyner and Boore, on the other hand, used a model in which the shape is constrained to be independent of magnitude. We will focus on these differences to illustrate procedures for examining a model's adequacy.

Adequacy of the Model

No empirical equation can be condemned on the basis of the steps used in its derivation. It does not matter whether the steps leading to the equation are logically defensible or not; lacking the means to subject it to the ultimate test (does it predict the future?), the model is best assessed by studying the residuals of the observations about the regression curve. Such residuals are shown as a function of distance in Figure 14. There is a hint of a systematic trend in Campbell's residuals (high at both ends, low in the middle), but the trend probably has no statistical significance. From this figure, we conclude that neither model has any gross inadequacies; to address the question of a magnitude-dependent shape we must resort to other comparisons. This requires studies not contained in the published papers, and therefore our discussion will now focus on the Joyner and Boore predictions. Similar studies have been made by Campbell and Niazi (1982).

Before continuing, it is worthwhile pointing out the large scatter in Figure 14. For Joyner and Boore's data, the standard deviation corresponds to a factor of almost

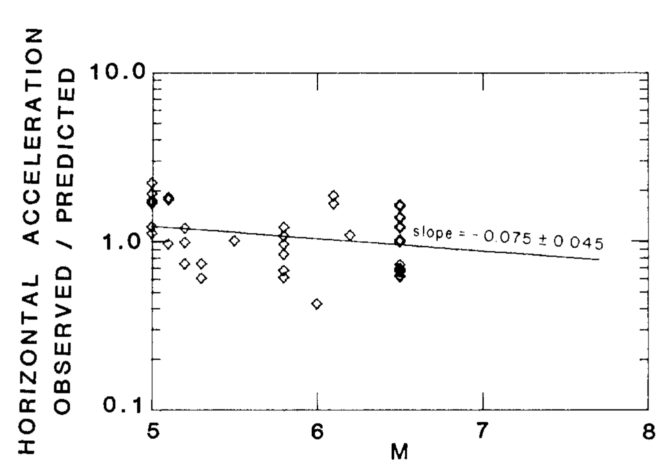


FIG. 15. Residuals of peak horizontal acceleration with respect to the predictions of Joyner and Boore (1981) for recordings within 10 km of the fault.

1.9; although not shown in the figure, the comparable number in Campbell's studies is about 1.5. McCann and Boore (1983) found factors of about 1.3 for motions recorded close to one another (within circles of 0.5 km radius) during the 1971 San Fernando earthquake. This latter factor shows that the scatter due to local geologic (and perhaps building) effects may be considerable. An important task for research is to separate the scatter due to geologic effects, which is potentially predictable, from that due to source effects. Unfortunately, there are indications that a significant amount of the scatter may be due to nonpredictable source effects. A scatter corresponding to a factor of 1.35 was found by Joyner and Boore (1981) in their second regression (see Figure 10); this is presumably due to variations in dynamic stress release in earthquakes. Somerville and Nelson (1982) attributed a large amount of the scatter in data from the 1971 San Fernando earthquake to source directivity effects.

To return to the question of a magnitude-dependent shape, we note that the mean square residual is dominated by the data at distances beyond 20 km or so and therefore is not sensitive to magnitude-dependent differences in shape, which are most pronounced at short distances (see Figure 13). A better way to see if such magnitude dependence is demanded by the data is to plot the residuals at close distances as a function of magnitude; a correlation of the residuals with magnitude would indicate that the basic assumption in the Joyner and Boore model (magnitude independence) is invalidated by the data. Such a plot (Figure 15) does show a trend (and in the sense predicted by Campbell's study), but the slope is controlled by a single earthquake and is marginally significant.

A more powerful test involves a Monte Carlo simulation. This is especially useful in nonlinear regression analysis, for it makes possible the derivation of confidence limits on the regression coefficients (Gallant, 1975); the confidence limits derived from standard linearized approximations are useless if the problem is strongly

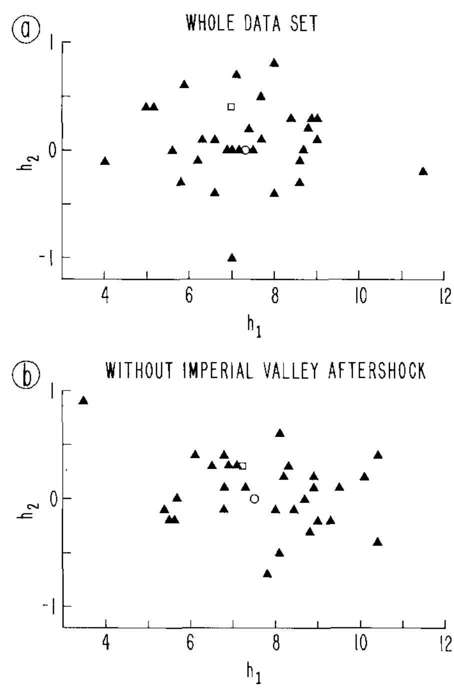


FIG. 16. (h_1, h_2) pairs obtained from the Monte Carlo study described in the text. Solid triangles are derived from the simulated data, which used the (h_1, h_2) value shown by the open circle, and the open square shows the (h_1, h_2) pair from analysis of the observed data. (a) Whole data set. (b) Data set excluding the 1979 Imperial Valley aftershock.

nonlinear, as is the case for the regression performed by Joyner and Boore. The Monte Carlo procedure is straightforward. We assume that the derived prediction equation (given in Table 2) is correct and use it to generate a set of artificial data at magnitude and distance values corresponding to the data set used in deriving the prediction equation. Random noise with the appropriate variance is added to each value. The resulting artificial data set is then subjected to the same regression analysis as was the observed data, with one crucial difference: the h_2 coefficient in the distance definition (Table 1) is not constrained to be zero; this allows for a magnitude-dependent shape. This procedure is repeated many times, and an (h_1, h_2) pair is derived from the observed data. If the h_2 value for the observed data were nonzero but within the range of values derived from the artificial data, then we could say that at a certain significance level the hypothesis of magnitude-independent shape ($h_2 = 0$) could not be rejected. That is, it could well be by chance alone that the derived value of h_2 differed from zero. The results of the Monte Carlo simulation are shown in Figure 16 for data with and without the aftershock of the 1979 Imperial Valley earthquake.

The value of h_2 derived from the data is 0.3, implying a magnitude dependence similar to that found by Campbell. The simulations show that at least 20 per cent of the time, the derived h_2 can be greater than or equal to 0.3, even if the real h_2 equals zero. Therefore, the data do not require a magnitude-dependent shape ($h_2 \neq 0$); neither, however do they reject it.

The question of magnitude-dependent shape is affected by the definition of distance since large earthquakes break through to the surface and small ones may not. Restricting our remarks to the definition used by Joyner and Boore (1981), we see that the data are consistent with either magnitude-dependent or magnitude-independent shape. This leaves only general philosophical considerations (e.g., choose the model with the fewer free parameters) or theoretical considerations as a basis for deciding between the models. Theoretical considerations, however, may be of little help at present, for they depend on unknown aspects of the earthquake source process. For example, if one assumes that h [defined as $h_1 \exp(h_2[M - 6.0])$] should scale as the fault width, then the attenuation curve will have a magnitude-dependent shape up to the magnitude at which the fault width equals the width of the seismogenic zone and will have a magnitude-independent shape for higher magnitudes. If one assumes that h should scale as the fault length, then the attenuation curve will have a magnitude-dependent shape at all magnitudes. If, however, one assumes that h should scale as some characteristic dimension of the seismogenic zone or some parameter of the stress distribution in depth, then the attenuation curve will have a magnitude-independent shape. We know of no firm basis for deciding among these possibilities, but as Figure 13 shows, the corresponding differences in predicted acceleration are small.

Concluding Remarks

As we have shown, there are a number of options available in making empirical predictions of strong ground motion. To decrease confusion and the chance for misapplication, it is important that the choice of dependent and independent variables, as well as the data selection and analysis procedures, be clearly stated. Although these factors contribute to the scatter in predictions from various methods, we should not forget that a significant amount of variance in the residuals about the regression equations is undoubtedly due to things not accounted for in the model—such as azimuthal, propagation, and site effects (e.g., see Boatwright and Boore, 1982 and Mueller and Boore, 1982 for examples of azimuthal and site effects in the acceleration records of two recent California earthquakes). Sorting out these effects and reducing the variance is an important topic for future research. We are only now acquiring the data needed to carry out this task.

Finally, the predictions that are most often of concern are for magnitudes above about 7 and distances closer than 25 km, for which we have few or no data. In this case, the empirical prediction of strong ground motion necessarily represents an extrapolation and is model-dependent. In the future, more data (although never enough) will be obtained in this critical part of magnitude-distance space, and better theoretical models will be available to guide us in our predictions. The prediction of strong ground motions in western North America for earthquakes of magnitude less than 7 seems to be well understood; not only are the most recent empirical studies in substantial agreement with each other, but they are also in agreement with independently derived theoretical predictions (Hanks and Kanamori, 1981).

Acknowledgments

We are grateful to Ken Campbell for pointing out the application of Monte Carlo methods to testing the significance of nonlinear regression coefficients, to Allin Cornell for clarifying the use of the Monte Carlo method, and to C. S. Mueller for help in the Monte Carlo experiment and for preparing a number of the figures. T. Hanks, R. Page, and P. Spudich provided thoughtful comments on the manuscript.

References

Ambraseys, N. N. (1978). Preliminary analysis of European strong-motion data 1965–1978, *Bull. Eur. Assoc. Earthquake Eng.* **4,** 17–37.

Anderson, J. G. (1979). Estimating the seismicity from geological structure for seismic-risk studies, *Bull. Seism. Soc. Am.* **69,** 135–158.

Battis, J. (1981). Regional modification of acceleration functions, *Bull. Seism. Soc. Am.* **71,** 1309–1321.

Boatwright, J. and D. M. Boore (1982). Analysis of the ground accelerations radiated by the 1980 Livermore Valley earthquakes for directivity and dynamic source characteristics, *Bull. Seism. Soc. Am.* **72,** 1843–1865.

Boore, D. M. (1980). On the attenuation of peak velocity, *Proc. Seventh World Conf. on Earthquake Engineering (Istanbul)* **II,** 577–584.

Brune, J. N. (1968). Seismic moment, seismicity, and rate of slip along major fault zones, *J. Geophys. Res.* **73,** 777–784.

Campbell, K. W. (1981a). Ground motion model for the central United States based on near-source acceleration data, *Proc. Conf. on Earthquakes and Earthquake Engineering: The Eastern United States,* vol. 1, J. E. Beavers, Editor, Ann Arbor Science Publishers, Inc., Ann Arbor, Michigan, 213–232.

Campbell, K. W. (1981b). Near-source attenuation of peak horizontal acceleration, *Bull. Seism. Soc. Am.* **71,** 2039–2070.

Campbell, K. W. and M. Niazi (1982). Evidence for saturation of peak acceleration in the near field (abstract), *Earthquake Notes* **53,** 91.

Chiaruttini, C. and L. Siro (1981). The correlation of peak ground horizontal acceleration with magnitude, distance, and seismic intensity for Friuli and Ancona, Italy, and the Alpide belt, *Bull. Seism. Soc. Am.* **71,** 1993–2009.

Cornell, C. A. (1982). Sources and uncertainty in treatment of ground motion prediction (abstract), *Earthquake Notes* **53,** 45.

Cornell, C. A., H. Banon, and A. S. Shakal (1979). Seismic motion and response prediction alternatives, *Earthquake Eng. Struct. Dyn.* **7,** 295–315.

Gallant, A. R. (1975). Nonlinear regression, *The American Statistician* **29,** 73–81.

Hanks, T. C. and H. Kanamori (1979). A moment magnitude scale, *J. Geophys. Res.* **84,** 2348–2350.

Hanks, T. C. and R. K. McGuire (1981). The character of high-frequency strong ground motion, *Bull. Seism. Soc. Am.* **71,** 2071–2095.

Hasegawa, H. S., P. W. Basham, and M. J. Berry (1981). Attenuation relations for strong seismic ground motion in Canada, *Bull. Seism. Soc. Am.* **71,** 1943–1962.

Herrmann, R. B. and M. J. Goertz (1981). A numerical study of peak ground motion scaling, *Bull. Seism. Soc. Am.* **71,** 1963–1979.

Idriss, I. M. (1978). Characteristics of earthquake ground motions, Earthquake Eng. and Soil Dynamics (Proc. ASCE Geotechnical Eng. Div. Specialty Conf., June 19–21, 1978, Pasadena, California), **III,** 1151–1265.

Joyner, W. B. and D. M. Boore (1981). Peak horizontal acceleration and velocity from strong-motion records including records from the 1979 Imperial Valley, California, earthquake, *Bull. Seism. Soc. Am.* **71,** 2011–2038.

Joyner, W. B. and D. M. Boore (1982). Estimation of response-spectral values as functions of magnitude, distance, and site conditions (abstract), *Earthquake Notes* **53,** 70.

Joyner, W. B., D. M. Boore, and R. L. Porcella (1981a). Peak horizontal acceleration and velocity from strong-motion records (abstract), *Earthquake Notes* **52,** 80–81.

Joyner, W. B., R. E. Warrick, and T. E. Fumal (1981b). The effect of quaternary alluvium on strong ground motion in the Coyote Lake, California, earthquake of 1979, *Bull. Seism. Soc. Am.* **71,** 1333–1349.

McCann, M. W. and D. M. Boore (1983). Variability in ground motions: root mean square acceleration and peak acceleration for the 1971 San Fernando, California, earthquake, *Bull. Seism. Soc. Am.* **73,** (in press).

McGuire, R. K. (1974). Seismic structural response risk analysis, incorporating peak response regression on earthquake magnitude and distance, Research Report R-A -51, Department of Civil Engineering, Massachusetts Institute of Technology, Cambridge, Massachusetts, 371 pp.

McGuire, R. K. (1978a). A simple model for estimating Fourier amplitude spectra of horizontal ground acceleration, *Bull. Seism. Soc. Am.* **68,** 803–822.

McGuire, R. K. (1978b). Seismic ground motion parameters relations, *Proc. Am. Soc. Civil Eng. J. Geotech. Eng. Div.* **104,** 481–490.

McGuire, R. K. and T. C. Hanks (1980). rms accelerations and spectral amplitudes of strong ground motion during the San Fernando, California, earthquake, *Bull. Seism. Soc. Am.* **70,** 1907–1919.

Molnar, P. (1979). Earthquake recurrence intervals and plate tectonics, *Bull. Seism. Soc. Am.* **69,** 115–133.

Mueller, C. S. and D. M. Boore (1982). Detailed study of site amplification at El Centro strong-motion array station #6, *Proc. Third Int. Conf. Earthquake Microzonation* **1,** 413–424.

Murphy, J. R. and L. J. O'Brien (1977). The correlation of peak ground acceleration amplitude with seismic intensity and other physical parameters, *Bull. Seism. Soc. Am.* **67,** 877–915.

Nuttli, O. W. and R. B. Herrmann (1981). Consequences of earthquakes in the Mississippi Valley, Am. Soc. Civil Eng., 81-519 (preprint).

Rogers, A. M. and J. C. Tinsley (1982). A study of the relation between geological and ground motion parameters in Los Angeles, California (submitted for publication).

Schnabel, P. B. and H. B. Seed (1973). Accelerations in rock for earthquakes in the western United States, *Bull. Seism. Soc. Am.* **63,** 501–516.

Shakal, A. F. and D. L. Bernreuter (1981). Empirical analyses of near-source ground motion, U.S. Nuclear Regulatory Commission NUREG/CR-2095.

Somerville, P. and G. Nelson (1982). The distribution of the RMS acceleration and duration around faults (abstract), *Earthquake Notes* **53,** 96.

Swanger, H. J., J. R. Murphy, T. J. Bennett, and R. Guzman (1980). State-of-the-art study concerning near-field earthquake ground motion, U.S. Nuclear Regulatory Commission NUREG/CR-1340.

Toro, G. R. (1981). Biases in seismic ground motion prediction, Massachusetts Institute of Technology, Department of Civil Eng. Res. Rept. R81-22, Cambridge, Massachusetts, 133 pp.

Trifunac, M. D. (1976). Preliminary empirical model for scaling Fourier amplitude spectra of strong ground acceleration in terms of earthquake magnitude, source-to-site distance, and recording site conditions, *Bull. Seism. Soc. Am.* **66,** 1343–1373.

Trifunac, M. D. and J. G. Anderson (1978). Preliminary empirical models for scaling pseudo-relative velocity spectra, Appendix A in Methods for Prediction of Strong Earthquake Ground Motion, U.S. Nuclear Regulatory Commission NUREG/ER-0689.

U.S. Geological Survey
MS 77
Office of Earthquake Studies
345 Middlefield Road
Menlo Park, California 94025

Manuscript received 24 May 1982