DISTANCE PARTITIONING IN ATTENUATION STUDIES

John A. Blume[I]

SUMMARY

An earthquake data set consisting of 816 station-component accelerations, distances, and magnitudes is studied by a new method, which treats the data separately in ten distance bands; later the points obtained may be connected by various procedures. The object is to test the effects of regression leverage resulting from normal multiple regression analyses in three or more variables with unbalanced data sets wherein the mass of the data points may be at great distances from the distance of interest for a particular purpose. It is found that at short distances this band, or bivariate, analysis procedure provides equations that predict lower peak accelerations than other methods at short distances, and essentially the same accelerations at distances of 40 km or more. For sites close to the moving fault, the accelerations may not be as great as often estimated. The wide spread in recorded data requires careful interpretation of these results and further study.

INTRODUCTION

The attenuation of acceleration, velocity, or displacement with distance from the source has been given a great deal of attention in the literature, starting with Gutenberg and Richter [1] and followed on a somewhat different basis by Blume [2]. There have been many studies since, most of them using a form as proposed by Esteva [3], as in:

$$a = b_1 e^{b_2 M} (R + k)^{-b_3} \quad (1)$$

wherein a = peak instrumental acceleration, M = magnitude of the type specified, R = hypocentral distance, and $b_{1,2,3}$ and k are constants. In addition, the writer's site-acceleration-magnitude (SAM) equations, which have been simplified and improved over the years with increasing data bases [4], include site characteristics and probabilistic considerations, as well as magnitude and distance relationships.

A problem in deriving essentially all empirical attenuation equations has been the sparsity of data points at the short distances, especially for large magnitudes, and also at long distances. In addition, there is a rather limited dispersion of data points relative to magnitude. One result of this is wide variations in predicted ground motion, especially close to the source. The 1971 San Fernando earthquake produced a mass of data, most of which falls within a fairly narrow distance band from the source. Unless weighted in some manner, this one event tends to dominate regression analyses and to force the resulting curves up or down at the extreme distances relative to the mass of data points in the central region. This

[I]President, URS/John A. Blume & Associates, Engineers, San Francisco, California, U.S.A.

effect, here termed *regression leverage*, can result in a situation in which a best fit to such unbalanced data becomes a poor fit to reality at extreme distances. This factor explains much of the discrepancy between various attenuation curves at short distances from the source [5].

To explore this problem and to obtain more realistic attenuation relationships at various distances, a different procedure, here termed *distance partitioning*, has been employed and constitutes the subject of this paper. In this method, large populations of earthquake strong motion data are partitioned according to distance bands. Each band is treated separately in bivariate analysis with M and a, and subsequently the discrete band points are used directly or else smoothed or fitted mathematically to obtain an attenuation curve. This eliminates undesirable regression leverage and excessive extrapolation from data. Ten distance bands have been used herein on a trial basis with 816 total real earthquake data points.

THE DISTANCE PARAMETER

There are several possible expressions of distances that can be used in attenuation studies. The important point is that the empirical data should be the same as that to be used when the resulting attenuation equations are employed in practice, or else adjustments need to be made. This has rarely been done in the past. All too often, attenuation curves have been obtained by regression analysis using epicentral distances, while later applications of the resulting equations have employed the normal distances from the fault. This practice overestimates the estimated motion at short distances from the fault, as shown in Fig. 1. Other questionable practices have been to use epicentral distances and subsequently hypocentral distances, or vice versa. Another conservatism has been to assume overly large focal depths (when unknown) for regression analysis and later, in analysis, to use much shallower depths. Although consistency is desirable in these matters, it cannot always be achieved because of sparsity of data. The traditional assumptions that energy is released at the focal point rather than along the moving fault or that all energy is released opposite the site of interest in analysis or design also need to be reevaluated and made more realistic.

PARAMETERS OTHER THAN DISTANCE

In addition to distance, as discussed above, common parameters are magnitude, M, peak instrumental motion (acceleration, velocity, or displacement), site conditions, transmission-path characteristics, and probability of occurrence. The SAM equations [4] consider acceleration, distance, and all the above except transmission path characteristics. Other attenuation procedures ignore transmission paths, and most also ignore site conditions and probabilistic considerations. The SAM equations include the concept that the relationship of magnitude and acceleration changes at large M values. When more data become available, all these parameters can be considered more effectively. Magnitude may be local, body wave, surface wave, etc., and it often varies numerically, depending upon the recording station. In this study, the magnitude given in *United States Earthquakes* [6] has been used. Peak instrumental acceleration is defined as the peak absolute value during the entire time history (record); it is not to be implied that this value should be used in design analysis for the zero-period spectral anchor-point value without adjustment, especially at short distances. Acceleration is sensitive in the high-frequency range and is considered an effective parameter for all except long natural periods. Anchoring at periods other than zero -- whether with acceleration, velocity, or displacement -- can lead to errors in the important high-frequency range.

There have been many variations in the use of acceleration for attenuation studies. One question is whether to use the single peak value or both horizontal peak values at a recording station. The writer considers both components to be valid samples of the population and has used both when the values were known. Some extremists would argue for the single maximum, or even the greatest vector. Another important point is the lower cutoff value. Should all recorded data be used (in which case the instrument contributes the cutoff), or all data above some arbitrary acceleration, such as 0.01g, or 0.05g, or 0.10g? The greater the cutoff value the greater the mean and, often, the lesser the dispersion from the mean. Because peak acceleration is, by itself, a poor index of damage in most cases, it seems preferable at this stage to use all available valid data without arbitrary filtering. This has been done in this study; *United States Earthquakes* data have been used from the start in 1933 through 1970. This excludes the San Fernando data of 1971.

THE DATA SET

The data set as a whole consists of 816 station-component records. Most of the earthquakes were in California and western Nevada, and a few were in the Northwest. Magnitudes range from 2.1 to 7.6, with a mean value of 5.4; the standard deviation of the magnitude set is 0.91, so the coefficient of variation is 0.17. The hypocentral distances of the total data set range from essentially zero to 449 km, with a mean value of 84.4 km and a standard deviation of 81.0 km. The peak recorded accelerations range from 1 to 490 Gal, or cm/sec^2, with a mean value of 26.1 Gal and a standard deviation of 51.2 Gal; thus the coefficient of variation of peak acceleration is 1.96 and indicates wide dispersion. The data set has often been treated as a regression problem in three variables with a as the dependent variable. Various analyses show a rather wide variation of results at short distances depending upon the data set employed and the form of the equation selected.

For the partitioning or bivariate analysis that is the subject of this paper, the set of 816 data points was partitioned into 10 distance bands as shown in Table 1. The data on M and a for each band are provided in Table 2. The magnitudes in Table 2 cover a wide range, especially at the short distances. The mean values increase with distance, while the standard deviations are in the narrow range of $0.6M$ to $0.8M$. The measured accelerations cover a very wide range, especially at the shorter distances. This is reflected as well in the wide dispersion from the mean; the acceleration coefficients of variation are generally well in excess of unity.

PROCEDURE AND RESULTS

If enough empirical data were available, all important parameters and their interrelationships could be treated with confidence by multiple regression analysis. Such data, however, are not available and may not be for many decades. Theory can and should be introduced and be reconciled with the data whenever possible. However, there are great variations between theoretical models and measured data from real earthquakes. For example, although alluvial response should decrease at a greater rate with distance than rock outcrop response, empirical data clearly show the opposite effect with few exceptions [4]. This is not anomalous, but the result of overcompensating local effects such as natural mode excitation in deep and/or soft soil deposits. Multivariate analyses produce results in the distant ranges of sparse data that are highly subject to the nature of the mass of the data elsewhere and also to assumptions made about the shape of the curve or the character and the weight of the variables. Bivariate analysis is used herein to investigate the response for each distance range of interest without it being affected by data from other directions. This was a trial investigation. The results are compared with those from other methods.

The procedure followed involved first partitioning the data in distance bands for analysis, then making the band analyses with only M and a as variables, smoothing the results, and, if desired, returning to the multivariate use of all the data by best fit. Each of the R-partitioned elements was treated as a bivariate data set using the mean values of R from Table 1 as constants for the ten distance bands. This is then a simple calculation procedure that eliminates most of the possible numerical problems of multivariable data-set analysis. The equation applied to each band is:

$$\log_{10} a = \mathrm{b}M - \mathrm{c} + y\sigma \tag{2}$$

wherein b and c are constants from analysis of the data, σ is the standard deviation of log a, and y is the desired factor for standard deviations from the mean; $y = 0$ provides the mean value. Table 3 is a tabulation of the values obtained for b, c, and σ.

Although the values of b and c for the various bands seem to be somewhat erratic, their use in Eq. 2 produces accelerations that follow a more logical pattern. Fig. 2 shows with large, solid dots the mean peak acceleration for 7.5M at the mean R-value for each distance band. Horizontal lines are drawn through these points for the width of each band, which is defined by the vertical lines. The letters represent the band designations. The open circle directly above each point represents that mean-plus-one-standard-deviation acceleration value for the same band. Curve I in Fig. 2 is an exponential best fit to the set of ten mean points, each assigned equal weight. The equation for Curve I is:

$$\bar{a} = 311\, e^{-0.0171R} \tag{3}$$

wherein $\bar{a}$ is the mean peak acceleration in Gal. Curve II in Figure 2 is the Blume SAM V equation [4] used here with $M = 7.5$, $\rho V_s = 2{,}000$ fps, and $y = 0$ for the mean value. Curve II provides an excellent fit to the partitioned data.

A separate study (results not shown in the figure) was made wherein 52 close-in data points were added to the original set of 816. These new data sets were for strong motion recorded in Italy (1972) and India (1967) and for several post-1970 earthquakes in the United States. The magnitudes ranged from 3.2 to 6.5 and the hypocentral distances from 5 to 15 km [7]. The peak accelerations were from 60 to 700 Gal, with a mean value of 216 Gal. This additional input slightly reduced the derived accelerations close in but had no significant effect elsewhere.

There are various ways in which the bivariate data derived band by band can be returned to three-variable form in M, R, and a. One method used was first to develop exponential curves similar to Curve I of Fig. 2 for other assumed magnitudes using b and c constants from Table 3 in Eq. 2. Multiple regression analysis was performed in the format of Eq. 1 using k = 25 km and selected point sets in M, R, and a from the curves. The resulting three-variable equation for the mean acceleration (in Gal) was:

$$\bar{a} = 18.4\, e^{0.941M}(R + 25)^{-1.27} \qquad (4)$$

Fig. 3 shows plots of Eq. 4 for M = 7.5, 6.5, and 5.5 in the solid lines. Also shown in Fig. 3, in dashed lines, are curves for M = 7.5, 6.5, and 5.5, derived directly from the points calculated from the raw data. A sample of such points (for M = 7.5) is provided in Fig. 2 as the solid dots. The use of the raw data provided the three-variable Eq. 5:

$$\bar{a} = 102\, e^{0.970M}(R + 25)^{-1.68} \qquad (5)$$

which is plotted in Fig. 3 as the three dashed lines for M = 7.5, 6.5, and 5.5. Eq. 5 is probably more valid than Eq. 4, which suffers to a mild degree from regression leverage, as can be seen in Fig. 2.

The SAM V curve for an assumed site impedance ρV_s of 2,000 fps (Curve II as in Fig. 2) is also shown in Fig. 3 for comparison at M = 7.5. The SAM V equation agrees quite well with Eq. 5, which is the one with essentially no regression leverage. Any of the curves in Fig. 3 seem more reasonable than those derived from a large proportion of data from one earthquake (such as from the San Fernando earthquake of 1971) and subject to considerable leverage and distance extrapolation.

The figures have been plotted in logarithmic form for scale and convenience. Although all the data are shown, it is not easy to visualize the relative accelerations. Table 4 provides a direct comparison of mean acceleration for specific R values and for M = 7.5. Of course, all these values are subject to adjustment for values other than the mean, and care needs to be taken in all comparisons as to the probability of exceedance. For example, if a curve is drawn as an envelope of all points, it obviously is well above the mean expected value.

DISCUSSION

Which is the correct procedure? With limited data no one knows with absolute certainty. It is believed that some equations in the literature are highly leveraged by the data set and also affected by the form of equation selected. The SAM equation is less leveraged than many because it is from a large data set and yet does not contain the massive San Fernando data. The band analysis or partitioned results in this paper will no doubt be considered low at short distances by some. More data and time will tell.

In the meantime, this writer believes the SAM equations constitute a reasonable basis for estimation. However, if it is assumed that the only available data were concentrated in one distance band, confidence in analysis for that band would be gained. If that assumption were then repeated for other bands, more confidence would be gained. The manner in which these band points are connected is not too important if done reasonably. Partitioning analysis is useful, simple, reasonable, and instructive.

The general tendency is to consider only the largest recorded values. Although these are indeed important, they alone do not answer questions as to what shaking is most likely given certain conditions, such as M and R, and what shaking level has what chance of being exceeded. Distance partitioning is a useful device in attempting to answer such questions.

REFERENCES

1. Gutenberg, B., and C. Richter, 1956, Earthquake magnitude, intensity, energy, and acceleration, second paper, Bull. Seis. Soc. Am., 46:2.
2. Blume, J. A., 1965, Earthquake ground motion and engineering procedures for important installations near active faults, Proceedings, Third World Conference on Earthquake Engineering.
3. Esteva, L., 1970, Seismic risk and seismic design input for nuclear power plants, Seismic Design for Nuclear Power Plants, MIT Press.
4. Blume, J. A., 1977, The SAM procedure for site-acceleration-magnitude relationships, Proceedings, Sixth World Conference on Earthquake Engineering.
5. Blume, J. A., and A. S. Kiremidjian, 1978, Near-field data effects and further treatment of attenuation relationships, Seismic Evaluation for Postulated 7.5M Hosgri Earthquake, Diablo Canyon Site, U.S. Nuclear Regulatory Commission.
6. United States Earthquakes, 1933-1970, U.S. Department of Commerce, Washington, D.C.
7. Hanks, T. C., and D. A. Johnson, 1976, Geophysical assessment of peak accelerations, Bull. Seis. Soc. Am., 66:3.
8. Trifunac, M. D., 1976, Preliminary analysis of the peaks of strong earthquake ground motion, Bull. Seis. Soc. Am., 66:1.

TABLE 1

DISTANCE PARTITIONING DATA

Distance Band	Hypocentral Distance (Range), km	Number of Data Points	Distance, R, km			
			Minimum	Maximum	Mean	Standard Deviation
A	0 - 9.9	58	0	8	5.2	2.2
B	10 - 19.9	99	10	19	14.2	3.0
C	20 - 29.9	91	20	29	23.7	3.0
D	30 - 39.9	80	30	39	33.3	2.8
E	40 - 49.9	52	40	48	44.1	3.1
F	50 - 59.9	37	50	59	52.3	3.1
G	60 - 99.9	101	60	97	74.9	10.3
H	100 - 139.9	129	100	138	116.0	10.6
I	140 - 199.9	96	140	190	168.6	15.4
J	≥200	73	200	449	269.8	78.8
		816				

TABLE 2

DATA ON M AND a, BY DISTANCE BAND

Distance Band	Magnitude, M				Acceleration, a, Gal			
	Minimum	Maximum	Mean	Standard Deviation	Minimum	Maximum	Mean	Standard Deviation
A	2.1	6.0	4.6	0.68	2.0	490	61.7	98.3
B	3.7	7.1	4.7	0.69	1.6	389	50.0	76.0
C	3.8	6.6	5.0	0.72	1.3	363	40.8	55.7
D	3.8	5.8	4.8	0.53	1.0	275	19.7	35.9
E	3.5	7.6	5.3	0.81	2.0	162	29.2	38.3
F	4.0	7.1	5.3	0.83	1.0	141	29.6	36.1
G	4.3	7.6	5.5	0.61	1.3	195	19.6	27.2
H	4.5	7.6	5.8	0.72	1.9	32	9.6	8.7
I	5.0	7.6	6.2	0.58	1.3	316	13.7	44.7
J	5.3	7.6	6.5	0.69	1.0	17	4.5	3.7

TABLE 3

RESULTS FOR USE IN EQUATION 2, FROM PARTITIONED ANALYSES

Distance Band, i	b_i	c_i	Standard Deviation, σ_i
A	0.381	0.276	0.46
B	0.576	1.413	0.45
C	0.334	0.372	0.49
D	0.298	0.452	0.52
E	0.449	1.208	0.40
F	0.470	1.335	0.43
G	0.460	1.490	0.34
H	0.329	1.090	0.32
I	0.433	1.955	0.38
J	0.063	-0.121	0.32

NOTE: $\log_{10} a_i = b_i M - c_i + y\sigma_i$
(Eq. 2 repeated for convenience)

TABLE 4

MEAN PEAK ACCELERATIONS FOR M = 7.5, GAL

Source	Hypocentral Distance, R, km				
	10	40	70	100	200
Equation 3	262	157	94	56	10
Equation 4	234	107	66	46	22
Equation 5	375	133	70	44	16
Sam V [4] (ρV_s = 2,000 fps)	455	131	60	35	11
Trifunac et al. [8] (Case 1, s = 0)	1,170	230	80	44	14

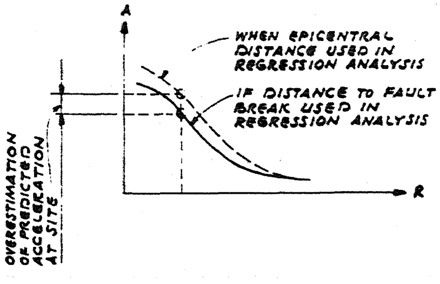

FIG. 1 EFFECT OF DISTANCE CONSIDERATION

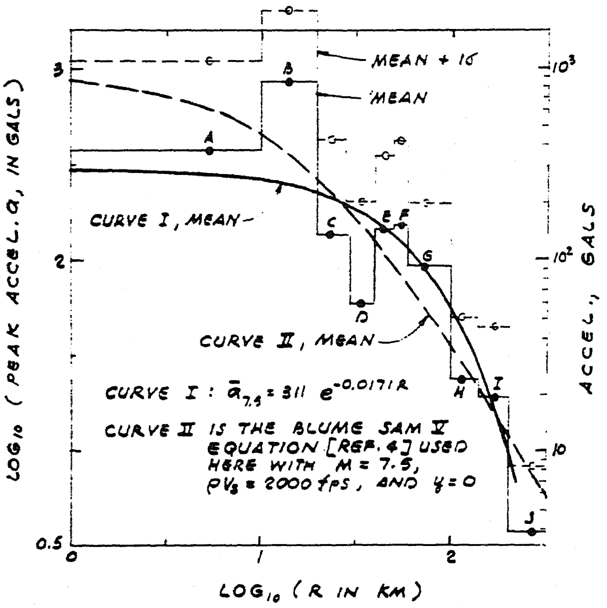

FIG. 2 ATTENUATION FOR M = 5

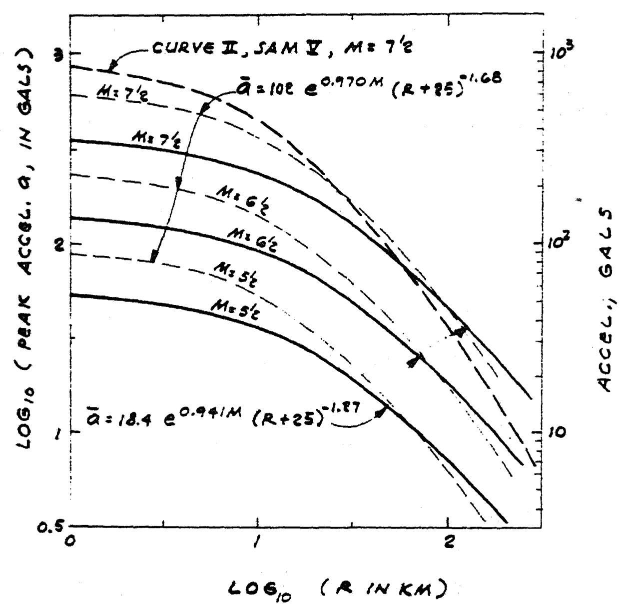

FIG. 3 COMPARISON OF MEAN ATTENUATION CURVES