Seventeen—
Some Examples of the Statistical Analysis of Seismological Data

David R. Brillinger

> "Data! data! data!" he cried impatiently, "I can't make bricks without clay."
> Sherlock Holmes
> —A. Conan Doyle, The Adventure of the Copper Beeches (1892)

> "Mr. . . . has joined the society, and, like many engineers, is interested in the possible effects of earthquakes. . . . These men want to know the seismicity of given places. The Lord help them!"
> If the engineers of the county will cooperate with the Seismological Society of America in the effort to gather and publish data regarding earthquakes, the Seismological Society of America will gladly undertake to get them some help here on this earth.
> —Seismological Notes (1911, p. 185)

Introduction

A subject that has been called statistical seismology has too few researchers but a number of success stories to its credit. Vere-Jones and Smith (1981) reviewed much of the work in the subject up to 1980. This presentation concentrates on some themes of contemporary statistics that seem of some relevance to the seismological circumstance. The examples of their use are based principally on the work of my students and myself.

That statistics is important in seismology seems self-evident. This was recognized very early on. Rothé (1981) recorded that part of the program of the 1891 Tokyo Earthquake Investigation Committee was

> To draw up a list of shocks with dates and times for each phase; to study the distribution of earthquakes in space and time; to study possible relations with the seasons, the phases of the moon, meteorological conditions, etc.

These are all data sets ripe for statistical analysis. It may be mentioned generally that there are massive seismological data sets, that uncertainty abounds, and that there are floods of hypotheses and inferences. Earthquake prediction is in the public mind. Seismology is also important to statistics. This results in part from the field's remarkable generosity in making data sets available and from the intriguing formal problems it raises.

The foremost researcher in statistical seismology has to be Harold Jeffreys. His research altered the field of both seismology and statistics in major

---

fashions. His working attitude is illustrated by the remarks: ". . . I have been insisting for about twenty years that the claim of finality for any scientific inference is absurd" (Jeffreys, 1939) and "The uncertainty is as important a part of the result as the estimate itself. . . . An estimate without a standard error is practically meaningless" (Jeffreys, 1967).

Of Jeffreys's work, Hudson (1981) has written: "The success of the Jeffreys-Bullen travel time tables was due in large part to Jeffreys's consistent use of sound statistical methods." The part of Jeffreys's work that has perhaps affected statistics the most is his development of robust/resistant techniques for handling nonnormal and bad data. Other scientists whose work has had major impact on seismological statistics include: Keiiti Aki, Bruce Bolt, Allin Cornell, Yan Kagan, Vladimir Keilis-Borok, Leon Knopoff, Bob Shumway, John Tukey, and David Vere-Jones. More recent contributors include Daniele Veneziano and Yosihiko Ogata.

Likelihood-Based Procedures

In the statistical approach to data analysis it is usual to view observations as realizations of random variables. Important to that approach is the notion of likelihood. If the (multivariate) observation (Y_1 , . . ., Y_n) is assumed to come from a random variable with probability function p (y_1 , . . . ,y_n | q), depending on the unknown parameter q , then the likelihood function of q given the observation is defined to be

$$L(\theta) = p(Y_1, \ldots, Y_n | \theta)$$

Employing likelihood-based inference procedures handles and unifies a variety of problems. The procedures are often highly efficient. There are corresponding estimation, testing, and confidence procedures, (referring back to the second Jeffreys's quote). Results derived from different data sets may be combined routinely

In applications, the approach is to set down a likelihood based on a conceptual model of the situation at hand. As an example of employing a likelihood procedure, consider the problem of estimating the seismic moment and stress drop of a particular event given a particular seismogram. For a variety of source models, researchers have related the seismic moment and stress drop to characteristics of the amplitude spectrum, |W (w)|, (that is, the modulus of the Fourier transform of the signal). Suppose that the seismogram is written

$$Y(t) = u(t;\theta) + \epsilon(t) \qquad (1)$$

where u is the signal, q is an unknown parameter and $\hat{I}$ is the "noise." If W (w ;q) denotes the Fourier transform of u (t ;q), then what is given, from the

source model, is the functional form of |W($w;q$)|. Following Brune (1970), common forms (for displacement measurements) include

$$|\Omega(\omega;\theta)| = a / \sqrt{1 + (\omega/\omega_0)^b} \text{ and } a/[1 + (\omega/\omega_0)^2]$$

where q = {a, b,w_o }, are the parameters to be estimated. Estimates of the seismic moment and stress drop may be determined once estimates of a and w_0 are available. The practice has been to estimate the unknowns graphically from a plot of the modulus of the empirical Fourier transform, |$d^T{}_Y$ (w)|, where

$$d_Y^T(\omega) = \sum_{t=0}^{T-1} Y(t)\, e^{-i\omega}$$

0 £ w£p . The following likelihood-based procedure was suggested in Brillinger and Ihaka (1982) and developed in detail in Ihaka (1985).

When the asymptotic distribution of |$d^T{}_Y$ (w)| is evaluated for the case of stationary mixing $\hat{I}$ (t), it is found to depend on |W ($w;q$)| and $f_{\hat{I}\hat{I}}$ (w) alone, *where* $f_{\hat{I}\hat{I}}$ (w) is the power spectrum of the noise. Hence, given an expression only for the modulus of W , one can proceed to estimate q . For the model (1), and small noise, one has

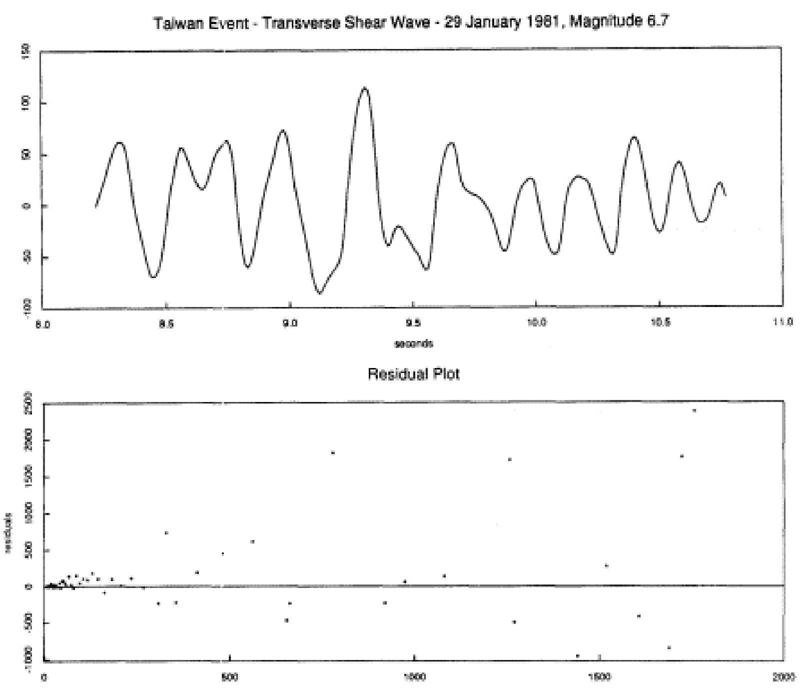
$$|d_Y^T(\omega)| = |\Omega(\omega;\theta)| + (d_\epsilon^T(\omega) + d_\epsilon^T(-\omega))/2 + \ldots$$

showing variation around $|W|$ independent of $|W|$. However, when deviations of $|d^T{}_Y|$ from a fitted version of itself are plotted versus the fitted values, dependence of the error on $|W|$ is apparent. An example is provided in figure 1. This is the result of computations for an earthquake of magnitude 6.7 that occurred in Taiwan on 29 January 1981. The data were recorded by one of the instruments of the SMART 1 array (Bolt et al., 1982). The top graph of the figure provides the transverse S-wave portion of the recorded accelerations. The lower graph provides the deviations plot just referred to. This plot suggests that the noise is in part "signal generated."

Various physical phenomena can lead to signal-generated noise. These include multipath transmission, reflection, and scattering. The following is an example of a model that includes signal-generated noise.

$$Y(t) = u(t) + \sum_k [\gamma_k u(t-\tau_k) + \delta_k u^H(t-\tau_k)] + \epsilon(t)$$

where t_k are time delays, u^H is the Hilbert transform of u , g_k , d_k like a and b above, are parameters to be estimated reflecting the vagaries of the transmission process, and $\hat{I}(t)$ is unrelated noise. The inclusion of the Hilbert transform allows the possibility of phase shifts. Assuming g_k , $d_k t_k$ are random, and evaluating the large sample variance, one is led to approximate the distribution of the discrete Fourier transform values, $Y_j = d^T{}_Y(w_j)$ by a complex normal with mean $W(w_j;q)$ and variance $\Gamma_j = 2\pi T[\rho^2|\Omega(\omega_j;\theta)|^2 + \sigma^2]$, $\omega_j = 2pj/T$. Here it has also been assumed that $\hat{I}$ is white noise (of variance s^2),



[Full Size]

Figure 1

The top graph provides the computed transverse shear wave component derived from data recorded by the SMART 1 array. The bottom graph provides residuals, that is, the difference between the absolute values of the empirical Fourier transform values and their mean values determined from the final fitted values. These are plotted against the fitted values. Wedging is apparent.

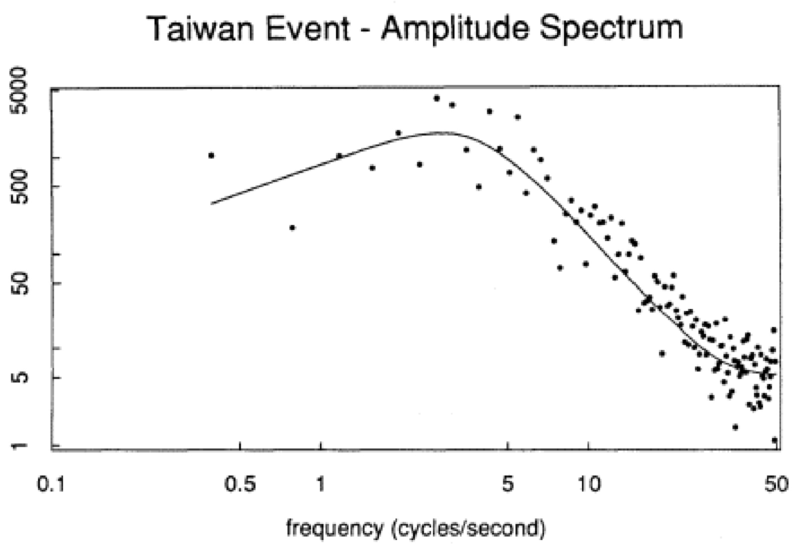
that the expectations of g_k and d_k are zero, and that the process t_k is Poisson. The ratio r^2 / s^2 measures the relative importance of signal-generated noise. In the likelihood approach one proceeds to estimate q by deriving the marginal distribution of the $|Y_j|$ and then setting down the likelihood. This likelihood when evaluated is found to be approximately

$$\prod_j \left\{ \exp\left[-\frac{|Y_j|^2+|\Omega|^2}{\Gamma_j}\right] I_0\left[\frac{2|Y_j|\,|\Omega_j|}{\Gamma_j}\right]\frac{1}{\Gamma_j}\right\}$$

[Full Size]

where I_0 denotes a modified Bessel function. Figure 2 shows a fit of the model $|\Omega(\omega)| = a|\omega| / [1 + (\omega/\omega_0)^4]$ to the data of figure . The fit is good.

Once estimates of a , w_0 are at hand, they may be converted to estimates of



[Full Size]

Figure 2

The plotted points are the absolute values of the discrete Fourier transform of the data of figure 1. The smooth curve is the result of fitting the Brune-type model $|w| / [1 + (w/w_0)^4]$

the seismic moment and stress drop via the theoretical relationships that have been developed. Uncertainty measures are directly available for the estimates. Details of this technique and a study of its theoretical properties may be found in the thesis of

Ihaka (1985).

Borrowing Strength

"Borrowing strength" is the colorful term John Tukey has introduced for the class of statistical procedures that seek to improve on naive estimates by incorporating data from parallel but formally distinct circumstances. These procedures also go under other names, such as pooling, random effects, James-Stein, shrinkage, empirical Bayes, and Bayes. The technique of damped regression provides an example most known to seismologists. Of the notion generally, Mallows and Tukey (1982) have remarked: "Knowing when to borrow and when not to borrow is one of the key aspects of statistical practice." A popular account of "improved" estimates is given in Efron and Morris (1977). The case of the linear model is developed, with examples, in Dempster et al. (1981).

To begin with a simple example, suppose that one wishes to estimate the mean μ_i of a population i , and one has available the mean $\bar{Y}_i$ of a sample of

values from that population. Then the naive estimate of μ_i is $\bar{Y}_i$. Suppose, however, that other populations beyond the ith, and corresponding sample means, are available. Suppose that these populations are all somewhat similar. Let $\bar{Y}$ denote the mean of all the sample means of the populations. Consider borrowing strength, in the estimation of μ_i , from the other populations; specifically consider forming an estimate

$$q\bar{Y}_i + (1 - q)\bar{Y} \qquad (2)$$

for some q lying between 0 and 1. One would like to choose q to be near 1 if $\bar{Y}_i$ can almost stand on its own, but q to be near 0 if the $\bar{Y}_i$ are highly variable. This problem may be formalized via a random effects model, specifically by setting down a model

$$Y_{ij} = \mu + \epsilon_i + \epsilon_{ij}$$

with the $\hat{I}_i$, say, independent variates with mean 1 and variance t^2 , and the $\hat{I}_{ij}$ independent variates with means o and variance s^2 . Then, for the case of samples all the same size, J , the "best" linear unbiased estimate of $\mu_i = \mu + \hat{I}_i$ is given by expression (2) with

$$q = J\tau^2/(J\tau^2 + \sigma^2)$$

In the case that t is zero, q is 0, and the estimate is $\bar{Y}$. In the case that t is infinity, q is 1, and the estimate is $\bar{Y}_i$.

As an example of what is involved here, consider the problem of developing attenuation relationships. Quite a variety of specific functional forms, involving a finite number of real-valued parameters, have been set down. For example, Joyner and Boore (1981) develop the relationship

$$\log A = -1.02 + 0.249M - \log \sqrt{d^2 + 7.3^2} - 0.00255\sqrt{d^2 + 7.3^2} \qquad (3A)$$

for (mainly) western United States earthquakes with A peak horizontal acceleration, with M moment magnitude, and with d closest distance to the surface fault rupture in kilometers. To prevent earthquakes with many recordings from dominating the estimates, Joyner and Boore carried out the fitting in two stages. First magnitude was not included in the model, but an event constant was. Then the event constant estimates were regressed on magnitude to obtain the term – 1.02 + 0.249M. There were 23 events and 182 records in all.

One may obtain "improved" estimates as follows. The Joyner-Boore functional form will be retained. Let the subscript i index the event, and j index the record within the event. Consider the (random effects) model

$$\log A_{ij} = \alpha_i + \beta_i M_i - \log\sqrt{d_{ij}^2 + \delta_i^2} - \gamma_i\sqrt{d_{ij}^2 + \delta_i^2} + \epsilon_{ij} \qquad (3B)$$

where a_i, $b_i g_i d_i I = 1, \ldots,$, are independent realizations of random variables with means $m_a, m_b m_g m_d$ and variance $\sigma_\alpha^2, \sigma_\beta^2, \sigma_\gamma^2, \sigma_\delta^2$, respectively.

TABLE 1		
Parameter	Estimate	Standard Error
μ_a	–0.969	0.210
μ_b	0.239	0.034
μ_g	0.00187	0.00091
μ_d	6.99	2.29
s_a	0.0617	0.0700
s_b	0.148	0.066
s_g	0.00193	0.00127
s_d	0.0294	132.
s	0.213	0.014

The $\hat{I}_{ij}$ are independent noises with mean 0 and variance s^2 . This model ties together the events, but each event has its own a , b , g , d . (The usual nonlinear regression model corresponds to s_a, s_b, s_g, s_d. identically 0.) Implications of this model are that records for the same event are correlated and that the disparate numbers of records for the events are handled automatically. Assuming that the random variables involved are normal, the model can be fit by maximum likelihood (employing numerical quadrature as needed). The results are provided in table 1. In some cases, for example s_a, s_d there is a clear suggestion that the corresponding population parameter may be 0.

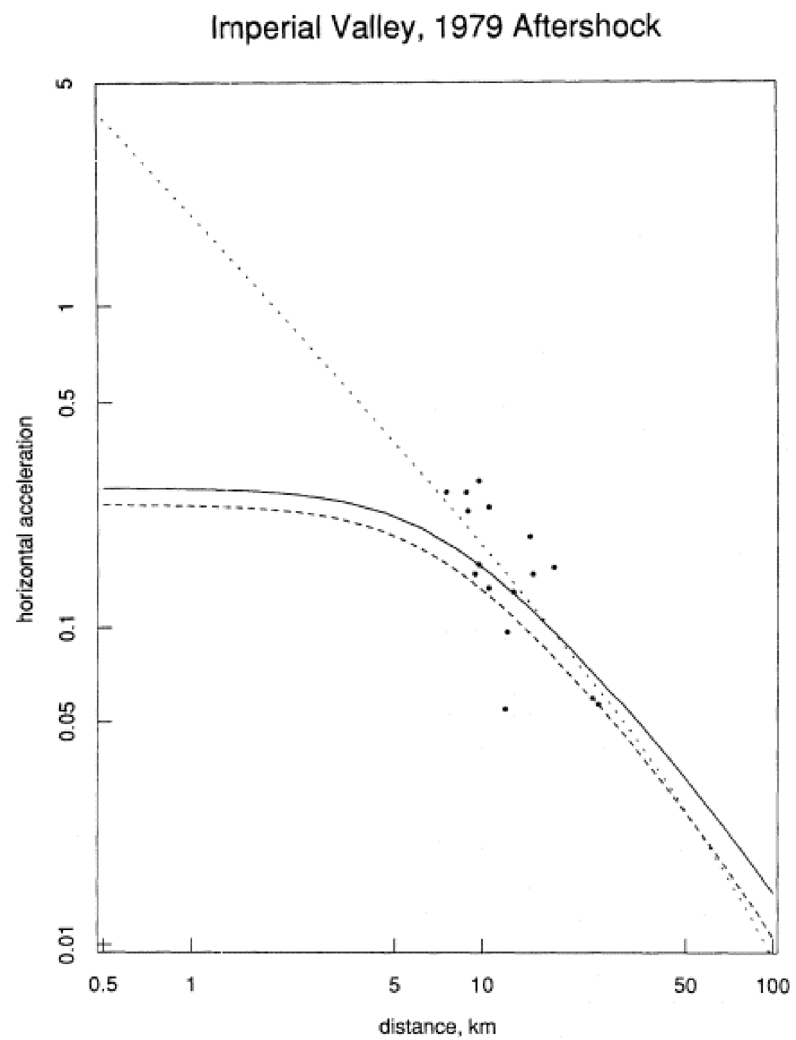
Once fit, model (3B) may be used, for example, for obtaining "improved" estimates of the attenuation behavior of the individual events. Consider for example the 1979 Imperial Valley aftershock. The data for this event are the points plotted in figure 3. Also plotted, as the curve of short dashes, is the result of fitting the Joyner-Boore functional form to the data for this event alone. Clearly, this curve is not too useful away from the cluster of observations. It has high uncertainty as well.

The solid curve graphed is the estimate of

$$E\{\alpha_0 + \beta_0 M_0 - \log(\sqrt{d^2 + \delta_0^2}) - \gamma_0 \sqrt{d_2 + \delta_0^2} + \epsilon_0 \mid \text{all the data}\} \quad (4)$$

with subscript 0 referring to this particular event. One has obtained a much more reasonable curve. This curve would be of use if one wished to estimate, *a posteriori,* an acceleration experienced in the Imperial Valley aftershock at a specified distance from the epicenter, for example to relate it to damage experienced at that distance.

The curve of long dashes in figure 3 is the Joyner-Boore curve, equation



[Full Size]

Figure 3

Points plotted are observed accelerations at the indicated distances. The curve of short dashes is the result of fitting the Joyner-Boore functional form to these data points only. The curve of long dashes is the curve developed by Joyner and Boore using the data set of twenty-three events. The solid curve is the "improved" estimate developed from expression (4) and the model (3B).

(3A). It is not inappropriate. A thing to note however is that the Joyner-Boore curve is the same for all events of the same magnitude, here $M_0 = 5.0$. It does not take special note of the actual data for the event.

Figure 4 provides "improved" estimates for three other events. In each case, the improved estimates (solid curves) are plotted, as well as the Joyner-Boore (dashed) curves given by equation (3A). The general effect of borrowing strength here, and typically, has been to provide a curve lying nearer to the mass center of the points observed in the particular event of concern. Of particular note is the case of the 1957 Daly City event where but one observation was available. One could not sensibly fit a curve to that data point alone. The Joyner-Boore curve has some validity. The "improved" curve pulls the Joyner-Boore shape nearer to the available observation. In the case of the 1979 Imperial Valley event the two curves are very close to each other. This is the case with the most observations (38).

Nonparametric and Semiparametric Estimation

Traditionally, the formal theories of statistical estimation were directed at cases involving a finite dimensional parameter. Exceptions consisted mainly of the cases of histograms and power spectral density estimates. Another exception was provided by the various curve estimates developed by seismologists, particularly Jeffreys, to deal with travel-time data (which correspond to a problem of infinite dimensional regression analysis, albeit one with a multivalued regression function). Recently, statisticians have turned to the problem of curve estimation in broad general situations. Problems studied include: estimation of a nonparametric transformation of the dependent variable, transformations of variates involved in quantal models, and (semiparametric) situations involving both finite and infinite dimensional parameters. In some cases the estimates are based on likelihoods, are adaptive, and may be anticipated to be highly efficient. References, with discussion, to statistical aspects of this work, are Breiman and Friedman (1985) and Hastie and Tibshirani (1986). Wegman (1984) is a survey article on some aspects.

As an example of what is involved here, return to the problem of developing attenuation relationships. Above, the Joyner-Boore functional form

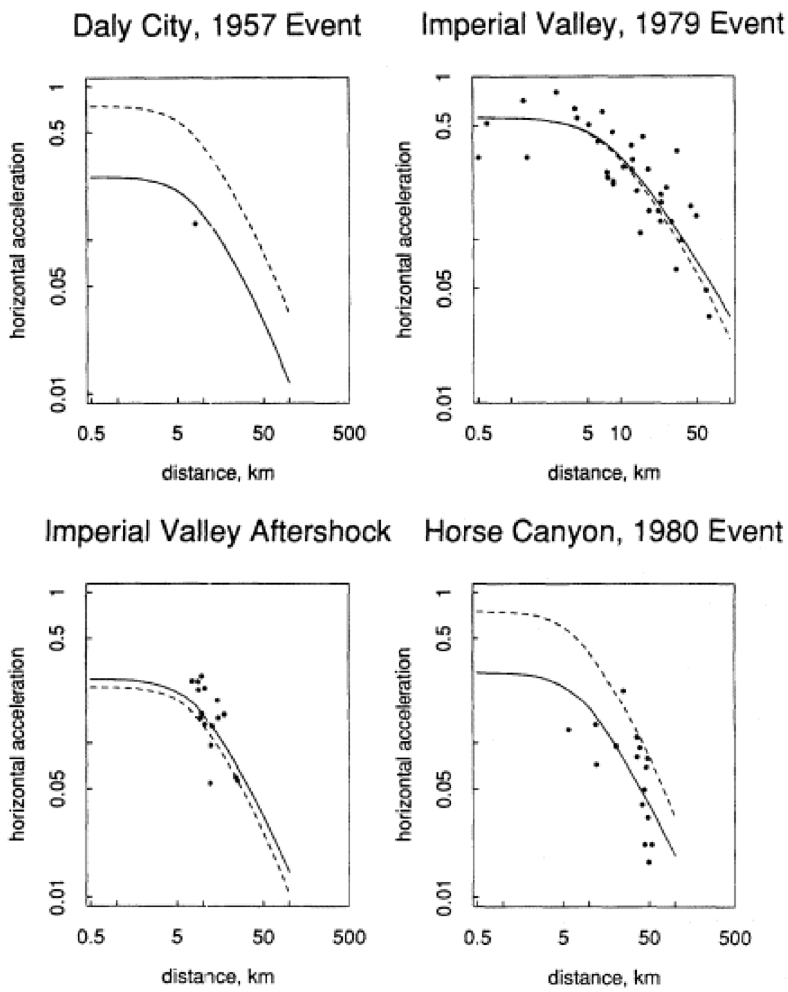
$$\log A = \alpha + \beta M - \log\sqrt{d^2 + \delta^2} - \gamma \sqrt{d^2 + \delta^2} \qquad (5)$$

was employed. Some theory suggests the use of the log and square root transformations in such a relationship; however, the theory is not definitive, and variants of equation (5) have been proposed.

These days one can often turn to a nonparametric analysis, estimating general transformations from the data. In Brillinger and Preisler (1984), monotonic functions *q* , *ϕ* , and *y* were estimated for a relationship

$$\theta(A) \approx \phi(M) + \psi(d)$$

— 275 —



[Full Size]

Figure 4

Observed accelerations are plotted for the four indicated events. The solid curve is the "improved" estimate, while the dashed curve is that of Joyner and Boore.

— 276 —

In determining such functions, critical assumptions were that the functions were smooth and the relationship additive. The formal model fit was

$$\theta(A_{ij}) = \phi(M_i) + \psi(d_{ij}) + \epsilon_i + \epsilon_{ij}$$

with *i* indexing an event and *j* a record within an event. The model was fit by a variant of the ACE procedure of Breiman and Friedman (1985). Figure 5 presents the results, namely the estimated functional transformations, q, ϕ , y , for the Joyner-Boore data. The transformation of magnitude is essentially linear. The general transformation of amplitude found is nearer to a cube root than a logarithm. The transformation of distance decays in a steady manner, as might have been anticipated.

From these curves one can obtain broadly applicable, predicted values of acceleration corresponding to specified magnitudes and distances.

Other Topics

Had time and space allowed, other topics that would have been reviewed include: general procedures for uncertainty estimation (such as the jackknife and the bootstrap), dimensionality estimation procedures (such as Akaike's information criterion), adaptive techniques, modeling incomplete data (or biased sampling), regression diagnostics, influence measures, and techniques for analyzing quantal data.

A Concluding Remark

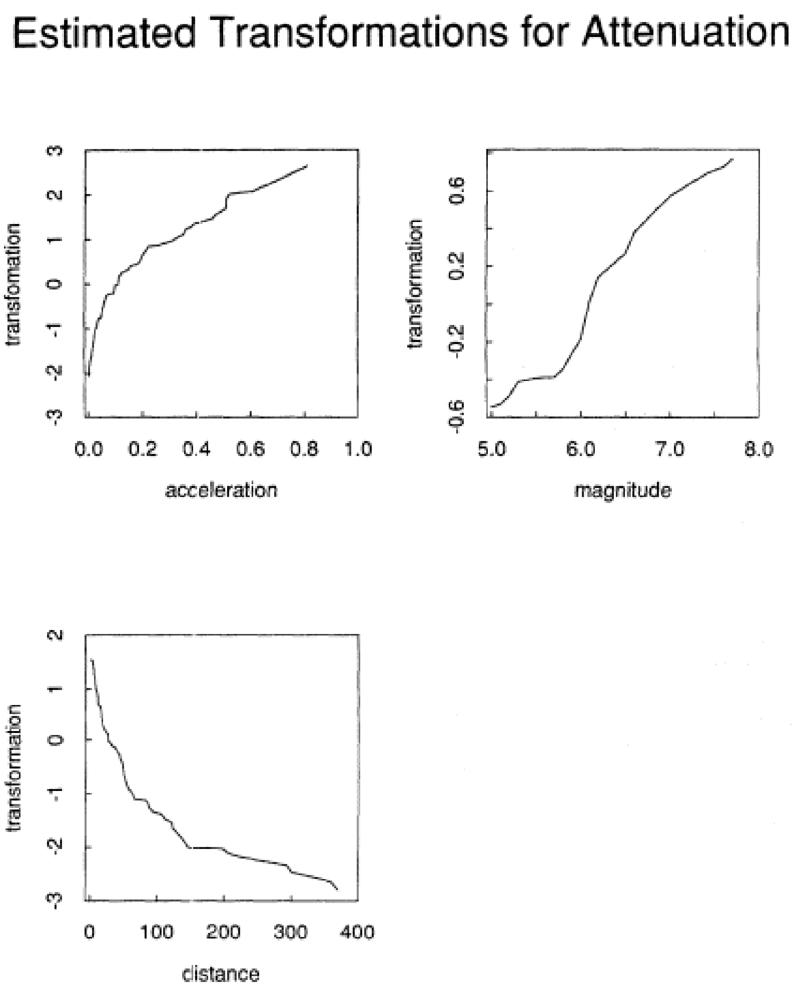
I end with a personal comment, based on a "noncollaboration" with a seismic researcher. A year or so ago, a young geologist came to see me because he had been advised that I might be able to help in computing uncertainties attached to some risk figures he had prepared. Happy to oblige was my feeling; however, as we talked, it became a highly frustrating business for both of us. As we tried to establish a common language it turned out that we really did not have an operational one. He had never taken any sort of statistics course. His problem was a hard one, so subtle techniques were called for. Sadly that is where the matter ended. Had he been at Berkeley, steady contact would have allowed a continuation, but he was not. There is no denying that there is much material that earth scientists have to be expert in. However, I would hope that statistics could be more routinely included in the list.

Acknowledgments

This research was carried out with the partial support of NSF Grant MCS-8316634.

It has benefited substantially from many discussions on the statis-

— 277 —



[Full Size]

Figure 5

Estimated monotonic transformations of acceleration, magnitude, and distance providing the "best" additive relationship of acceleration in terms of magnitude and distance for the Joyner-Boore data set.

— 278 —

tical analysis of seismological data with Bruce Bolt and David Vere-Jones through the years. I thank them for all the help and encouragement they have provided. I thank Bob Darragh for preparing the Smart 1 data record for analysis.

References

Bolt, B. A., Y. B. Tsai, K. Yeh, and M. K. Hsu (1982). Earthquake strong motions recorded by a large near-source array of digital seismographs. *Earthquake Engin. Structural Dynamics,* 10: 561–573.

Breiman, L., and J. H. Friedman (1985). Estimating optimal transformations for multiple regression and correlation. *J.Am. Statist. Assoc.,* 80: 580–597.

Brillinger, D. R., and G. R. Ihaka (1982). Maximum likelihood estimation of source parameters. *Earthquake Notes,* 53: 39–40.

Brillinger, D. R., and H. K. Preisler (1984). An exploratory analysis of the Joyner-Boore attenuation data. *Bull. Seism. Soc. Amer.,* 74: 1441–1450.

Brune, J. N. (1970). Tectonic stress and the spectra of seismic shear waves from earthquakes. *J . Geophys. Res.,* 75: 4997–5009.

Dempster, A. P., D. B. Rubin, and R. K. Tsutakawa (1981). Estimation in covariance components models. *J. Am. Statist. Assoc.* 76: 341–353.

Efron, B., and C. Morris (1977). Stein's paradox in statistics. *Scientific American,* 236: 119–127.

Hastie, T., and R. Tibshirani (1986). Generalized additive models. *Statistical Sci.,* 3: 297–309.

Hudson, J. A. (1981). Mathematics for seismology. *J . Inst. Math. Appl.,* 17: 34–39.

Ihaka, G. R. (1985). Ruaumoko. Ph.D. Diss., Statistics Dept., University of California, Berkeley.

Jeffreys, H. (1939). The times of P, S and SKS, and the velocities of P and S. *Mon. Not. R. Astr. Soc. Geophys.,* Suppl. 4: 498–547.

——— (1967). Statistical methods in seismology. In K. Runcorn, ed., *International Dictionary of Geophysics.* Pergamon, London, 1398–1401.

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.

Mallows, C. M., and J. W. Tukey (1982). An overview of techniques of data analysis emphasizing its exploratory aspects. pp. 111–172 In J. Tiago de Oliveira and B. Epstein, eds., *Some Recent Advances in Statistics.* Publicacoes do II Centenario da Academia das Ciencias de Lisboa, Portugal.

Rothé, J-P. (1981). Fifty years of history of the International Association of Seismology (1901–1951). *Bull. Seism. Soc. Am.,* 71: 905–923.

Vere-Jones, D., and E. G. C. Smith (1981). Statistics in seismology. *Commun. Statist. Theor. Meth.,* A10(15): 1559–1585.

Wegman, E. J. (1984). Optimal nonparametric function estimation. *J.Statistical Planning and Inference,* 9: 375–387.