4 June 1965, Volume 148, Number 3675 SCIENCE

Data Analysis and the Frontiers of Geophysics

More can be learned from data by wise use of spectrum analysis, choice of expression, and straggling values.

John W. Tukey

In the last decade and a half, at least in those areas of geophysics with which I have had the most contact, relatively new techniques of analyzing data have played an increasingly important part in revealing interesting geophysical facts. It was only about two years from the time we began to obtain experimental knowledge about the natural modes of vibration of the earth until we were studying their multiplet structure. It was only a few years from the time workers in this laboratory first detected the very long waves which reach this coast from 15,000 kilometers away (*1*) until it was possible to use multistation arrays to determine the direction of arrival of such waves and to obtain a detailed and reliable correspondence between wave records on the coast of California and the known behavior of storms in the Indian Ocean and beyond.

More recently, studies by Gordon MacDonald (*2*) have converted the mystery of the source of driving force for the continual but somewhat irregular 10-monthly Chandler wobble of the earth's poles into a new, deeper mystery. For the bispectrum has shown that the many-year wobbles of the poles interact (in some way that is still a mystery) with some phenomenon of 12-month period—perhaps the 12-monthly wobble of the poles—to provide the energy for the Chandler wobble.

I have asked myself what geophysicists should ask, at such a time as this, of one who takes part in developing techniques for the analysis of data. They might, I believe, most reasonably seek help with questions such as these: What will happen over the next not-too-many years? Do we expect wholly new kinds of analysis of data to have important impacts on geophysics? Do we expect modifications of presently used methods of analysis, either of the types which have been with us for a long time or of the types which have grown up during the last decade and a half, to play an important role? Or is the impact on geophysics of the new methods of analyzing data to be short-lived; has it now come close to its end?

Surely I am not here to say that this impact is close to its end. No one can foresee what wholly new kinds of analysis of data are going to become important in the years to come; almost by definition each one will prove to be something of a surprise. The most I can hope to do is to forecast newer uses and natural developments of today's methods.

There are three areas about which I feel I must speak.

1) Spectrum analysis, in the broadest sense—a field in which one can see a variety of the newer techniques just beginning to grow up.

2) Wise expression of data for analysis—a matter which too often seems too simple to be thought as important as it really is.

3) Modification of our techniques of summarization and analysis to deal effectively with the case (almost universal, especially in geophysics, even after the observations have been expressed wisely) where fluctuations and errors have a distribution whose shape is far more straggling than that of the magic bell-shaped curve of Gauss and Laplace.

Spectrum Analysis

We rarely emphasize, or even mention, the essential reasons why the techniques of spectrum analysis are so useful.

Time series are often subjected, naturally or artificially, to many processes, physical or computational, which have two simple properties. First, when applied to a time-by-time sum of two time series, these processes yield the time-by-time sum of the results of their application to the separate series. Second, no zero of time enters into their behavior. [As a consequence of this, the observations must either be (essentially) continuous or they must be taken at equal intervals of time.] These properties are usually described as "linearity" or "superposability," on the one hand, and as "constancy," "stationarity," "time-origin-shift invariance," or merely "invariance," on the other.

The passage of ocean waves over the ocean, of pressure waves through the atmosphere, of electromagnetic waves through the ionosphere, of seis-

The author is professor of mathematics and director of the Section of Mathematical Statistics, Princeton University, Princeton, New Jersey, and associate executive director of research, Bell Telephone Laboratories, Murray Hill, New Jersey. This article is based on an address presented 26 February 1964 at the dedication of the La Jolla Laboratories of the Institute of Geophysics and Planetary Physics of the University of California.

mic waves through the earth (*3*) are all instances of situations where these two properties hold to good, but but perfect, approximations. In these instances, as in many others, both the good approximation and the nature and causes of the deviations from perfect stationarity and origin-shift invariance are of scientific importance. Both must concern those who are to analyze the data.

There may be little in a name, but there are times when much can be said for having at least one. Let us call anything that exhibits, *to an approximation adequate for immediate purposes*, the two properties of invariance and superposability an *IS-box*. Those mathematical idealizations that exhibit both properties precisely and exactly can then be called *perfect IS boxes*.

Our instruments, and our computations, provide many other examples of IS-boxes. Many electric circuits, both those used in science and those used in stereo high-fidelity, are carefully designed to be good IS-boxes, as are both (i) the hydraulic circuits in the tsunami recorder on the pier of the Scripps Institution of Oceanography (*4*), which gives warning of so-called tidal waves, and (ii) many simple computational procedures, like differencing and most smoothing procedures.

Every IS-box must have a very simple property, one that relates to its treatment of frequencies: if a signal consisting of a single frequency enters the box, what comes out is a signal of the *same single* frequency! Once this holds for single frequencies, the superposability condition ensures that similar relationships will hold for pairs of frequencies, for triples, for quadruples, and so on. If, for example, a signal made up of only four particular frequencies enters an IS-box, the output signal must be made up of the same four frequencies and no others! And, more than this, the output at each of these frequencies must be determined solely by the input at the same frequency. This is a property of nonintertwining—of not getting certain constituents of the signal entangled with one another. This property of frequencies and IS-boxes is of extremely great importance: it is the key property of frequencies and the reason for spectrum analysis.

Just so long as the information we need about some phenomenon is expressed, at any one place and time, in a distribution of activity or energy or power over frequency, we have a hope of going from the there-and-then to the here-and-now. For if propagation, and instruments, and preliminary computations are all IS-boxes, the needed information will come through, possibly in modified form, into our analysis. We must beware the distractions of superposed noise and the deformations of wildly varying frequency responses, but we have a fighting chance to learn *here* what happened *there*.

This is the main reason why we are able to learn from data by simple spectrum analysis. There are, indeed, situations where it is interesting and useful to check detailed theories in terms of questions such as: Is the intensity of the spectrum near this frequency quantitatively what we would expect it be be? We dare not neglect these opportunities to make quantitative comparisons, but we should bear in mind that, on balance, such comparisons are probably less important than what we can learn about some phenomenon from qualitative aspects of an appropriate frequency spectrum.

So far I have been talking about the most elementary and naive form of spectrum analysis, "looking at the spectrum." Here we are really asking only one question: How much activity is present at various frequencies? This question has, on occasion, proved very illuminating, as in Munk and Snodgrass's detection (*1*) of the very-long-period waves coming 15,000 to 20,000 kilometers, from the Indian Ocean and beyond to the coast of San Clemente Island. We may both hope and expect that this approach will be equally helpful many times in the future. We should realize, however, that this is, by and large, a technique which is most effective in detecting phenomena that were really clamoring for our attention—even though the "clamor" may have been such a feeble whisper that we needed much care and amplification to hear it.

It is a commonplace of science that, where one can, one learns faster by deliberately reaching in and changing something—by seeing what happens when something is varied in a controlled way. Reaching in can be very much better than just sitting idly by. Astronomers have little choice; they *must* sit. Planetary physicists are in nearly the same situation. Geophysicists can reach in and vary little things, but not big ones. All need to ask: How can at least some of the advantages of reaching in be had when one can only sit and look?

The answer is simple—and well known: look in two places and try to assess the relationship of the two things observed. If one "causes" the other, you can learn about the process by which this occurs. If things are not this simple you may learn much about why they are not simple. If two places are not enough, try three, or four—or more. Where both what is varied and what is observed have steady-state values, one of which can be plotted against the other, this advice is both good and easy to take. It is equally good, but less easy to take, when each thing observed is a time series.

I emphasized the separateness of frequencies in an IS-box; by contrast I must emphasize the horrible unseparateness of values at specific times. The mathematician, when told that y causes x, automatically writes $y = f(x)$. This is occasionally reasonable. But when y is $y(t)$ and x is $x(t)$ it is too easy to act, or think, as if $y(t) = f[x(t)]$—as if the box representing the connection from one to another were an *instantaneous* box—so that the output *now* does not depend on the input in the past. Some simple things we do in computation are instantaneous boxes, as are some simple physical phenomena—indeed, ideal wave propagation, without dispersion, reflections, absorptions, or measuring devices (without, that is, the grim realities), differs from an instantaneous box by only a shift of time origin—but the real world is rarely even approximately instantaneous. The IS-box is of much wider application among time series than is $y = f(x)$.

To apply the maxim "Look here, look there, compare, and interrelate" to time series, then, we very often need to work in terms of frequency descriptions of these series, rather than in terms of time descriptions, and to introduce ideas and techniques that are closely related to some corresponding sort of spectrum analysis. To do this we must have concepts describing the interrelation of frequencies in two or more time series, and we must have ways of assessing what seems, in terms of these concepts, to be going on.

Such concepts appeared first in the

theory of polarized light, where different polarizations at different frequencies are, for example, provided by passing white light through almost any polarizer, and then through a solution of cane sugar. In terms of the cross-spectrum of two time series, a function of frequency that describes a distribution of joint activity over frequency —a description that has to use complex numbers in order to cope, for instance, with the difference between plane and circular polarization—we can develop a wide variety of tools and techniques. It was by using the cross-spectrum, for example, that the direction of arrival was fixed for the waves arriving on the California coast from 15,000 kilometers away (*5*).

Given two time series, and wishing to learn about the existence and nature of their interrelationship, it is all too easy to place a plot of y against t above a plot of x against t and then sit and stare—easy and misleading. Dispersion, or reflections, or preferential absorption, or the vagaries of instrumentation—any or all of these can conceal real relationships. In any situation, chance can cause the appearance of relationship to arise by accident, but chance's behavior is here much more threatening than usual. For the ways of chance are rarely simple, and are often unexpected, when they affect a plot of $y = y(t)$ or a paired plot of $y = y(t)$ and $x = x(t)$.

If, on the other hand, we study the cross-spectrum of the two series, giving due attention to the individual spectra, we deal with nonintertwined quantities and can reach conclusions, or suspicions, that are both more reasonable and more easily compared with the effects of chance. We cannot ask more from such a procedure than, in the non-time-series case, we could ask from a plot of y against x; if y and x are both caused by z, the plot is likely to look as if either caused the other. But we can ask as much—and we may discover more than we expected. Thus, if, for example, the relation of y to x is different for one range of frequencies than for another, we are likely to discover this fact if we look at the cross-spectrum and its relatives.

The easy way to summarize what has just been said about cross-spectra is to say that measuring relationship is almost always more rewarding than measuring relative contribution to variability. We can learn a fair amount from the second, but, whenever we can use the first, we will learn much more from it. And the cross-spectrum will enable us to do this in many more areas of geophysics.

So much for the spectrum and the cross-spectrum, which are important because of frequencies, which can be defined, if we so wish, in terms of averages of expressions involving squares and simple products of observations, and which provide contributions that will add up to give any one of these average values. Let us ask "What next?" Can we define analogous concepts in terms of third-degree expressions, such as cubes and triple products of observed values? Can we go beyond this? Will such definitions prove useful? *We can, we did, and they have.*

I spoke earlier about the preservation and nonintertwining of frequencies as signals or phenomena pass through IS-boxes, be these of rock or of water or instruments or computations. There are similar laws for the more complicated "bifrequencies," which describe the interrelatedness of what goes on at different frequencies. Bifrequencies can, if you insist, be indexed in terms of pairs of frequencies, but our thinking, our analogies, and our algebra all become much clearer if we index each of them in terms of a triple of frequencies, positive and negative, whose sum is exactly zero. Each such triple contributes its share to each third moment—to the average value, that is, of every homogeneous expression of degree 3 in the observations. Each triple does this as a whole; no part of the contribution can be validly assigned to any one frequency or to any pair of frequencies.

The key points are these. Each bifrequency passes through any perfect IS-box just as if no other bifrequency were present. Each bifrequency comes out of any perfect IS-box at exactly the same bifrequency at which it went in. What an IS-box does to a bifrequency is fixed, in a simple way, by what that IS-box does to the three frequencies that index the bifrequency.

It is this preservation and nonintertwining of bifrequencies that allows us to listen, here and now, to the labels that some nonlinear interaction has applied, far away and at some other time, to our time series. So long as the media, the processes, the devices, and the preliminary computations are all good approximations to perfect IS-boxes, the original label from that nonlinearity will remain attached, even though other, hopefully minor, nonlinearities between "there" and "here" may add other labels. (We may have to give some attention to the fact that we can observe only the arithmetic sum of these labels, so that there may be some concealment by over-labeling. But this would not be the first time we have had to look under arithmetic sums in order to find the things that were really of concern to us.) In those cases where relatively large amounts of relatively good data can be obtained, there seems to be no possible conclusion but that bispectral analysis will enable us to learn much about nonlinearities, even rather weak ones, even when data are obtained at places and times rather remote from those where the nonlinear interaction actually took place.

And the cross-bispectrum offers us opportunities which go at least as far beyond the bispectrum as the cross-spectrum goes beyond the spectrum. For those cases where we can get the rather long series required, and can afford rather substantial amounts of calculation, such types of data analysis have an important and useful future.

Bispectrum and cross-bispectrum analysis are far from being the only higher-order types of spectrum analysis which offer promise. The study of higher-order nonlinearities, or interactions involving more than two series at a time, may be investigated with the aid of the trispectrum, or of other higher analogs of the bispectrum and the cross-bispectrum. These will require still more data and still more computing time; we will hope not to be forced to use them too often, but we should be glad to know that they will be there when we need them most.

Moreover, while it is true that all fourth moments can be represented in terms of the trispectrum, all fifth moments in terms of the tetraspectrum, and so on, this is not the only possibility; there are other approaches, sometimes more useful. If one is interested in some aspects of the noise made by a symphony orchestra, it is natural to say that the orchestra plays a certain note, now loudly, now softly, now not at all, or that a certain note recurs every so often. One way of studying data from such a point of view is to find an apparent spectrum

for each of many short intervals of time and then study the variations from one spectrum to another. If one goes so far as to calculate the spectrum of the spectrum, as has been done, for instance, at Rocketdyne Division of North American Aviation, then one is again decomposing a fourth moment, but not in terms of specific trifrequencies. As time goes on, we may expect a slowly increasing variety of higher-order techniques of this sort to develop.

Moreover, there are situations in which one may want to use nonlinear techniques to answer questions which are about linear phenomena and which, at least naively, would seem to be most naturally answered by linear procedures of analysis. In one situation—the analysis of time series for echoes—there is at least a moderate amount of evidence that this is a good thing to do. The first work in this connection was directed toward solution of a geophysical problem: Can one assess the echoes on a seismograph record well enough to be able to localize the depth of the source of the signal?

The work that Bruce Bogert, Mike Healy, and I did on this (*6*) was not sufficient to produce any great advance in the geophysical problem. We did learn, I think, enough about what can be done with linear and nonlinear techniques to realize that there may be very great advantages in using nonlinear techniques in such studies. On the one hand, there are procedures involving a quadratic analysis of a logarithm of the result of a quadratic analysis—a combination which, as you see, already escapes all polynomial description. The *cepstrum* so obtained seems to have very real possibilities for certain sorts of echo-like problems. The sequence of a quadratic procedure, a logarithm, a linear procedure, an antilogarithm, and another linear procedure leads, among many other possibilities, to what we called the pseudoautocorrelation function, which seems to be a better indicator of echoes than the autocorrelation function itself and to offer interesting possibilities of its own.

It would be fair, I think, to say that the science and art of frequency analysis—of a frequency analysis combining the statistician's ideas, frequency ideas, the power of digital computers, and good data—have now advanced to a crucial stage: If one can be really specific as to what aspect (expressed clearly in frequency terms) of a time series or space function one wishes to study, one can at least make reasonable, and probably moderately effective, suggestions as to how this can be done. We have come a long way in the last decade and a half; we can expect to go further in the future.

Expression of Data

The growing power and flexibility of frequency methods are going, willy-nilly, to lead the geophysicists—along with many others—into more careful and effective ways of proceeding through an anlysis of data into ways that they will inevitably then begin to carry over into other problems, some of them classical, where time series do not appear. The logical consequences of Gordon MacDonald's analysis of the bispectra of very long pressure and temperature records offer one example of how this may begin. In these bispectra he has found the traces of several kinds of nonlinearity in the process by which the daily influence of the sun on the rotating earth has been transmitted to the pressure, or temperature, that was actually measured. And it is rather clear that different nonlinearities occur at different stages of this process. As it becomes more and more interesting and important to unravel and understand successive nonlinearities, geophysicists will undoubtedly begin, in many areas, to reprocess their observational records in such a way as to remove one nonlinearity after another. By doing this, beginning perhaps with the largest nonlinearity, perhaps with the latest one, they will reveal, more and more clearly, the traces of earlier and smaller nonlinearities.

We have long expressed many geophysical quantities, such as sunspot numbers and various magnetic character figures, in terms that all would admit to be somewhat arbitrary. Through our choice of how these quantities are expressed, we are probably guilty of producing the largest nonlinearity of all those each of these time series exhibits. If—often by so simple a change as taking square roots, or logarithms—we can eliminate most of these nonlinearities by a more appropriate and useful choice, perhaps guided by the bispectrum of a trial expression, we will be able to see much more deeply into the phenomena these time series describe.

Those who like to assess such useful changes at far less than their actual value often speak of their use as "making transformations," but we dare not follow their assessments. We must realize that wise choice of what is to be analyzed is a natural part of trying to let the data speak for themselves—something which each of us owes to his data as part of careful and painstaking analysis. It seems to me inevitable that, once we have learned, with the guidance of the bispectrum, to make such a choice in analyzing time series, we will carry over similar considerations, and equal care in the expression of data for analysis, to many other circumstances—to circumstances where we could, long ago, though perhaps on less objectively clear evidence, have improved our analysis of data by expressing them in a wisely chosen mode before starting our arithmetic mills, today computerized and speeded up, to grinding.

Even the simplest procedures of classical data analysis—calculating an arithmetic mean, or taking the difference of two observations—are really delicate devices intended to be as successful as they reasonably may be in screening out the unwanted while passing through the wanted. To do this with great skill and high effectiveness, such devices must depend, though their usual formulations often do not emphasize or notice this, upon very precise definitions of what is unwanted and what is wanted. It is usually far easier and far more effective to express the data to fit the analysis than to tortuously warp the analysis to fit a twisted expression of the data. Thus, for example, taking logarithms of the raw data is far easier than converting a complex calculation from arithmetic means to geometric means. It is our obligation to express our data so that the precise definitions inherent in our processes of analysis correspond as closely as possible to those definitions that will serve us best.

The choice of a mode of expression seems dull, it lacks the glitter of spectrum analysis. It seems to offer only a way to do the old things better, not a way to do the new and unsuspected. Yet its long-term promise for geophysics, like that of better methods of dealing with errors and fluctuations that do *not* follow the magic bell-shaped curve of Gauss and Laplace, may be greater, not less.

The analysis of data is not simple; accordingly, simple-seeming things may

have great importance. For the three decades from Schuster to modern spectrum analysis, geophysicists thought the examination of time series for frequencies was a simple, well-understood subject—they merely did not like the results. How many other techniques of data analysis are today in a similar period of repose and misunderstanding?

Modification of Techniques

Geophysicists have long had a substantial understanding of the most obvious problems associated with what in general might be called spotty data. No one who has associated with Sir Edward Bullard, for example, can fail to be aware of the importance which he places on the easy and automatic editing of time series for gross errors and omissions. In this sense, it is fair to say that geophysicists came early to some of the problems of spotty data and started long ago to do effective things about them. And one can say this in other senses as well, for the early work on the rejection of observations seems to involve either surveying, which we might consider applied geophysics, or actual geophysical measurement. And, among those who have devoted much effort to the analysis of data in a statistical framework, it is probably Sir Harold Jeffreys who has given most attention to what would be a relatively precise and efficient way to analyze observations which are, in fact, not distributed according to the magic bell-shaped curve of Gauss and Laplace. Geophysical data and geophysicists have been deeply involved in such questions, but the consequences of these considerations for the actual analysis of data have been far less deep and less widespread than one might have expected or hoped. We do not, yet, see many data analyzed by methods appropriate for situations in which distribution tails straggle more —that is, fall less abruptly—than those of Gaussian distributions. There are papers, it is true—for example, one in the *Monthly Notices of the Royal Astronomical Society*, by Henry Hulme and L. S. T. Symms (*7*)—which deal with the use of these methods for analyzing large bodies of data, where one can select a specific modified method very closely adapted to the data at hand.

This, however, is not the real problem of spotty data. The real problem is the devising of methods which will work well over a wide range of distributional shapes—methods which will squeeze as much from the data as it is reasonable to try to get; methods which can be applied to small and medium-sized sets of data as well as to large sets.

Once upon a time, when we calculated by hand, fitting by least squares could be a lot of effort, but the computer has changed all this. Today, when the constants to be fitted enter linearly into the expressions for the typical values of the observations, fitting by least squares is easy to handle in practice, and for a very simple reason: minimizing a quadratic criterion leads inevitably to solutions obtained by linear processes and to assessments of the quality of these solutions obtained by processes not more complicated than quadratic.

If we must be realistic (as I assert we will have to be), if we are to get anywhere near the most out of our data, then we must face up to the possibility—nay, the near-certainty—that the distribution of our observations will often be one for which minimizing a quadratic is a very bad choice. And it requires only almost imperceptible modifications of the shape of a Gaussian distribution to make the use of the arithmetic mean a slightly inferior way to summarize the typical-value behavior of a sample, and to make its sum-of-squared-deviations an almost unbearably poor summary of its spread (*8*). The price of thinking more generally and more usefully here, of acting more realistically, will be that our procedures will become more nonlinear, and almost certainly iterative, and that, in particular, the procedures for assessing the quality of the result will have to be at least somewhat more complicated than evaluating a fixed quadratic function. There are many cases where improved methods of this sort, when we have them, will allow us to recover two or three times as much information from a given body of data as we are now able to get. Such increases are not trivial, they are not the sort of thing that can be neglected, or that can be regarded as less than important. In geophysics, as in so many other fields, there are many situations where it would be extremely difficult, if not impossible, for a worker to obtain a two- or threefold increase in the number of observed data bearing on his problem. In these circumstances, increase in the effectiveness of analysis may be the only way to a more precise result.

It is relatively easy to describe, in more than one way, the essential character of a least-squares fit. The problem is to choose the description that will best illuminate the question at hand. For our present purpose, one suitable description runs as follows, at least for the moderately realistic case where the constants to be fitted enter linearly and where the observations, rigidly uncorrelated, are allowed to have different relative precisions in known ratios—to be subject to fluctuations and errors of differing variances which are known up to a multiplicative constant. We have a certain number of parameters, we have tried various combinations of values for them, we have found a combination such that we are not urged, on balance, to modify any of them. Here the forces that urge us to increase (or decrease) a particular parameter are represented by the algebraic sum of contributions from every observation, where each contribution consists of the product of two factors, one factor reflecting the relative precision of the observation in question and the other measuring the displacement of that observation from the typical value which would be assigned to it by the constants that we have fitted (that is, the values contemplated for the parameters). This type of least squares has long been known to have, in suitable circumstances, many advantages. Among the procedures which are linear in the observations and give correct values on the average, it is known to give fitted values with minimum variance and minimum average-square deviation. [This result was due to Gauss and is usually known as the Gauss-Markov theorem (*9*).] And, at the other extreme, if we know that our individual errors and fluctuations follow the magic bell-shaped curve *exactly*, then the resulting estimates are known to have almost all the nice properties that people have been able to think of. These latter results are very pleasant and very important, but it would be easy to give them far too much weight. For, in practice, errors and fluctuations very rarely have magic properties, particularly the magic bell-shaped distribution.

We can generalize this procedure of "fitting by balancing forces" beyond this simple case. We can generalize in various ways, but it is clear that—at least in those situations where we know

the distribution of fluctuations and errors exactly—almost any of these generalizations will be better than the corresponding form of least squares. The essential issue underlying any generalization must be that the pulling power associated with the deviation of an observation from the typical value corresponding to the fitted constants is no longer directly proportional to the size of this deviation. To modify this dependence logically, we have to specify something about the nature of the distribution of fluctuations and error. In general, as I am sure almost every geophysicist knows, distributions of actual errors and fluctuations have much more straggling extreme values than would correspond to the magic bell-shaped distribution of Gauss and Laplace.

A fairly extreme case, going beyond what one sometimes sees in practice, is the case where the probability of a deviation from a central value is inversely proportional to a quadratic function of this deviation. This type of distribution is known to the statisticians by the name of the French mathematician Cauchy and is, in particular, the distribution of the tangent of a uniformly distributed angle. In order to provide one physical interpretation for it, we might consider the simplest model for the width of a spectral line, one involving merely a simple resonance. Here the distribution of energy over frequency will follow the Cauchy distribution. In this situation, if we were able to catch individual quanta and measure their energy exactly (as by measuring their frequency) and wished to make an inference from a collection of such measurements to the center frequency of the emission line, we would be tempted to proceed merely by averaging the observations, forming the equally weighted arithmetic mean, and then looking at this. But a simple and well-known theorem states that, given n observations which independently follow the same Cauchy distribution, their arithmetic mean has a distribution exactly like that of each individual observation, in shape, in location, *and* in degree of spread. The man who takes an arithmetic mean of a sample from a Cauchy distribution has done the equivalent of throwing away all but one observation.

If we ask what the pulling function should be in a situation like this, we find that, for small deviations, the pulling power at first increases almost proportionately to the displacement, then it increases less and less rapidly, reaches a maximum, and eventually turns down. For very large deviations, indeed, the pulling power becomes arbitrarily small. It ought to surprise no geophysicist to be told that, when your distributions have very long tails, you are wise to make very little use of extreme observations. But the consequences of this for more or less linear methods of analysis are greater than you might expect.

The natural thing to do, once we have got as far as dealing with the sum of pulling powers, is to ask: Can we express the effects upon the fitted constants in terms of a weighted influence of the different observations—more particularly, a weighted influence of the values of the different observations? To do this in the most reasonable way we must ask: How much will the fit change when the value of any single specific observation is changed? The answer to this depends on: How fast does the pulling power change as the value of this observation is changed? Indeed, this latter change can quite wisely be taken as expressing the "weight" which is given to a specific observation.

In the case of the Cauchy distribution, and of many straggling-tailed distributions, the pulling power reaches a maximum and then decreases as the deviation from the typical value based on the fit becomes more and more positive. At that maximum value, the pulling power will not change with a small change in the deviation. Small changes in the value of the observation whose deviation falls at this maximum will have no effect at all on the resulting fit. The weight assigned to the value of such an observation is zero. For even larger deviations, increasing the deviation—moving the observation further away from the value that one would select by intelligent fitting—will *decrease* its pulling power. Such observations act as if their values received *negative* weights, not positive ones!

There is no escape from this type of conclusion. If the distribution of fluctuations and errors is sufficiently long-tailed, and precisely known, the only way we can make effective use of the values of very extreme observations is by giving them negative weights and letting them push where we would expect them to pull, and pull where we would expect them to push.

If we take a sample from a Cauchy distribution, for example, in a situation where we know the value of the parameter analogous to line width, so that we are merely trying to decide where the center of the distribution falls, the sample median—the value such that there are as many observations of greater value as of lesser value—conveys almost all the available information about the location of this center. There will remain some information to be recovered, but almost all of it is to be recovered by the backward process, just discussed, of saying: "The more positive the value of this observation, the more negative a typical value will I fit." This is a somewhat paradoxical conclusion, and we can be happy to learn that it follows only if we can trust, precisely and in detail, the assumed way in which the probability of occurrence of a deviation decreases as the size of that deviation increases. This extra information appears to be obtainable, but only because we pretend that we are dealing with a tight specification—that we have perfect knowledge of distribution shape. The real world is usually not amenable to tight specification.

The unreality of tight specifications, however, does not mean that we can ignore the problem of straggling tails; all our nice results about the arithmetic mean are associated, in one way or another, with the tight specification of the magic bell-shaped curve. They fail for the tight specifications corresponding to many straggling-tailed, but still bell-shaped, distributions, such as the Cauchy distribution.

What are we to do about situations of this sort? Rather clearly, the extra information to be recovered by negative weighting comes by courtesy of specific assumptions which we are unlikely to really know to be valid. The real situation is likely to be one in which we "fear" distributions like the Cauchy distribution rather than one where we know that we are, in fact, dealing with the Cauchy distribution. Accordingly, we ought to give up any thought of recovering this dubious information, and we ought to be content with what we can learn from something that really provides us with the information that it appears to provide, a property we may reasonably expect to find in a measure which does at least moderately well for tight specifications involving a rather wide range of distribution shapes. For, in this case,

things are as they should be: we do not have to worry very much about the exactness with which any particular tight specification represents the behavior of the observations.

There are such measures; the median, which provides about two-thirds of the information in a large sample from a magic bell-shaped curve (and a larger fraction in small samples) and a very much larger percentage in samples from a Cauchy curve, is one such. It is not the best sort of compromise, as far as one can tell at the moment; one does better—here "doing better" must be a matter of arriving at a better compromise rather than doing better everywhere—with, for example, the arithmetic mean of the central 80 or 90 percent of the observations.

Besides this particular procedure, known as trimming, in which we *appear* to neglect the outer 5 percent of the observations at each end, there are many other ways of playing down the effect of the exact values of extreme observations; trimming can be replaced by Winsorizing (*10*) and by various other procedures which should not even be named here. But it is very important to notice that in none of these procedures are we wholly eliminating the effect of the extreme observations.

We do not take explicit account of the quantitative values of these extreme observations—in fact we may go to considerable lengths to make sure that we have not done so! But, merely by existing and having values in a certain general range, such extreme observations make it clear that other observations are, in fact, not extreme, and thus determine what we do with these other observations. Thus, for example, if one were to wipe out the highest 5 percent of observations in a large sample, such a measure as the mean of the central 90 percent of the observations would shift downward, because we would have to eliminate more observations at the top before forming the arithmetic mean of the remainder.

In short, we have arranged to take qualitative, rather than quantitative, account of extreme observations. In very many circumstances this is the right thing to do. It is through procedures which do just this—even though it may require a little thought to see what is being done—that we can hope to greatly increase the effectiveness of our analysis of many kinds of data, especially in situations where the magic bell-shaped curve, though actually inappropriate, has been relied upon in the past.

Summary

I have tried to talk about three areas in which I think newer techniques of data analysis will come into, or be developed in, geophysics which will advance our knowledge about the plasma-like, gaseous, liquid, and solid earth: (i) developments along the line of the concepts of the spectrum, with emphasis both on the use of cross-spectral methods of studying relationship and on bispectrum and higher methods of studying nonlinearities (including modulations) and frequency interactions; (ii) wise choice of modes of expression for analysis; (iii) the use of methods appropriate and effective in situations where distributions of errors and fluctuations do not have exactly the magic bell-shaped distribution.

From past experience, I know which one of these you will think is of least importance: the choice of modes of expression for analysis. I would urge you to reconsider your intuitions, to give up the idea that the methods and techniques of data analysis are the only real keys to doing better, and to believe that the way you express your data for analysis may make it possible to learn much more. If geophysicists do this, individually and as a profession, then I think we will find, over the next decade or so, that all three of these areas have made major contributions to the unraveling of the problems of geophysics. Beyond this, I hope, as I am sure you do, that other methods which we cannot now anticipate will appear and have similarly large effects.

References and Notes

1. W. H. Munk and F. E. Snodgrass, *Deep-Sea Res.* **4**, 272 (1957).
2. G. J. F. MacDonald, *Space Sci. Rev.* **2**, 473 (1963).
3. Whether or not seismic *waves* are stationary may depend on whether you think in minutes, years, or tens of millennia; the *process* of transmission itself, however, is stationary: the rocks form a stationary box, at least over any seismologist's lifetime.
4. W. H. Munk, H. Iglesias, T. Folsom, *Rev. Sci. Instr.* **19**, 654 (1948).
5. W. H. Munk, *Ocean Wave Spectra* (Prentice-Hall, New York, 1963), p. 158.
6. B. P. Bogert, M. J. R. Healy, J. W. Tukey, *Time Series Analysis*, M. Rosenblatt, Ed. (Wiley, New York, 1963), p. 209.
7. H. R. Hulme and L. S. T. Symms, *Monthly Notices Roy. Astron. Soc.* **99**, 642 (1939).
8. J. W. Tukey, *Contributions to Probability and Statistics*, I. Olkin *et al.*, Eds. (Stanford Univ. Press, Stanford, Calif., 1960), p. 448.
9. R. L. Plackett, *Biometrika* **36**, 458 (1949).
10. W. J. Dixon, *Ann. Math. Statist.* **31**, 385 (1960).
11. The address on which this article is based was prepared in part in connection with research done at Princeton University, sponsored by the Army Research Office (Durham).