American Archaeology
Past and Future

A Celebration of the
Society for American Archaeology
1935–1985

David J. Meltzer, Don D. Fowler, and
Jeremy A. Sabloff, editors

Published for the
Society for American Archaeology
by the
Smithsonian Institution Press
Washington and London
1986
Archaeological Applications of Mathematical and Formal Methods

Introduction

This paper is about archaeological uses of mathematical or formal logical techniques and concepts that go beyond the simple tabulation and inspection of measurements, counts, and proportions. Actually, the point at which even these simple operations become more than routine is a bit vague, and I will not hesitate to discuss some aspects of their use and misuse. I include statistics as a subset of mathematics. The connection with computer applications is more tenuous. Computers present us with a body of extremely useful technology, partly for implementing mathematical and logical techniques, and partly for quite different purposes.

It is useful to think rather distinctly of three broad categories: archaeological observations, analytical methods, and sociocultural theory. Analytical methods provide the connections between observations and theory. Some theory is expressed directly in mathematical terms, but, as I will argue below, at present the vast majority of archaeological uses of mathematical and formal techniques pertain to the domain of analytical methods or to the design of data collection.

There is room for improvement in many of the mathematical techniques themselves, both in general terms and in developing techniques based on models that more closely approximate realistic archaeological situations. I will not, however, say much about the mathematics. One reason is that only a few of us will participate directly in this. Most of the developments will continue to be made by statisticians and applied mathematicians who have little or no contact with archaeological problems. Even the exceptions, although they will be very important, will be produced by the few persons who combine a practical sense of archaeo-
logical conditions with the capacity for creative mathematics, or who are at least able to communicate well with mathematicians. A second reason is that the realms of data and theory are far more in need of improvement. At present it is all too easy to find examples where the power and sophistication of the mathematics far outstrip both the relevance and reliability of the data and the depth or interestingness of the theory. It can be something like using a Rolls Royce to pull a wooden plow to cultivate nettles and thistles.

Improvement of mathematical techniques will remain a specialized task, but most of us can contribute to better theory, and all of us can, and indeed must, take pains to improve our data. It is a serious mistake to think that mathematical analysis, by and large, is theory-building. It is a disastrous mistake to think that sophisticated analysis obviates the need for data of very much higher quality than archaeologists have customarily obtained.

I will begin with a very brief discussion of the bearing of computers on formal techniques. After that, I will elaborate on the relations between mathematical analysis and sociocultural theory. Then I will discuss some of the most serious problems about data quality. Finally, I will offer an assessment of the current situation for some major topics in mathematical and formal approaches.

Computers

The impact of the affordable personal computer has been genuinely revolutionary. For decades computers were really used by only a minority of archaeologists. Quite suddenly this has changed. Most archaeologists are now accustomed to using computers for word processing, many are using database management systems, and interest in computer graphics is growing very rapidly. The continuously plummeting cost of computing has reached the point where, for a one-time investment of a few thousand dollars, one can acquire the hardware and software to do quite interesting things, with no further major expense except one's own time. Also, except perhaps in the climatic extremes of arctic cold or tropical heat and humidity, it is feasible to do very substantial computing in the field, without having to worry much about links to a mainframe at some home base.

We hear a lot about "user-friendly" software. It is at least equally important to note the great increase in "computer-friendly" archaeologists. The relevance of this for mathematical and formal methods is that many (though by no means all) of the useful techniques are not feasible without the aid of computer technology, and that awe of computers should no longer either impede their use nor distract us from down-to-earth pragmatic evaluation of results.

Relations Between Analysis of Data and Sociocultural Theory

Most applications of statistics and other mathematical concepts and techniques in archaeology fall into one of two quite distinct categories. One is designing research, summarizing important aspects of data, and generally manipulating data in ways that will bring out the aspects most relevant for specific purposes, especially for making well-warranted culture-historical inferences and for evaluating and improving sociocultural theory. The other is the direct use of mathematics for expressing sociocultural theory.

A surprisingly high proportion of the archaeological literature about mathematical and formal methods concentrates on the methods themselves and has very little or nothing to say about explicit sociocultural theory (except insofar as some general theoretical propositions may be included in the assumptions that justify certain models or procedures). Other publications use results of mathematical and/or formal methods to test, support, challenge, or suggest theoretical propositions, but the theory itself is rarely couched in mathematical terms. Finally, only a few publications use mathematics directly for expressing sociocultural theory. In fact, outside of simulation and modeling and applications of ideas from mathematical geography, I cannot think of good examples.

The split is not simply due to different workers with different interests. I find the distinction very evident in my own work. There have been a number of papers (e.g., Cowgill 1968, 1970, 1972, 1974, 1977, 1982) that examined aspects of various formal procedures in some detail. There was explicit concern that the results of these procedures should be useful for some important archaeological purposes, but almost nothing about specific topics in theory. In other papers (e.g., Cowgill 1975, 1979, 1983) I have addressed various issues in sociocultural theory, but have made almost no use of explicit mathematics. In my 1975 paper on population, the equation \[ p_t = p_0 \left(1 + \frac{r}{100}\right)^t \] was central to some of the points that I made, but it was relegated to a footnote. In the other papers a few inequalities and ordinal relationships were important, but I did not see any
need to go beyond the words of ordinary English to express them. The only paper that combines substantial use of formal techniques and substantial discussion of theory is Cowgill et al. (1984) and in that paper we emphasized difficulties in connecting theory with results of formal analyses.

If I tried, perhaps I could come up with theory that was more mathematical, but the fact is that I have not been motivated to do so. The concepts and issues in theory that interest me do not come from reading or reflecting on mathematical concepts. Many of them come, of course, from the writing and talk of other archaeologists. Most of the rest of them come from the work of other social scientists, to some extent social anthropologists, but especially social and economic historians.

There will always be some people who will fix on any newly fashionable mathematical idea that comes along, treating it as a hammer for which there must be some archaeological nails somewhere. These efforts should not be altogether discouraged; some of them may prove useful. Most of them, however, result only in gimmicks and the occasional bandwagon, and we do well to regard the enthusiasts with skepticism.

The most promising efforts to express aspects of theory in mathematical or quasi-mathematical terms have been in the fields of mathematical and economic geography, and the simulation of sociocultural systems (e.g., Clarke 1977; Hodder 1978a, 1978b; Hodder and Orton 1976; Renfrew and Cooke 1979; Sabloff 1981; Smith 1976). However, many of the former have been highly controversial (such as attempts to apply central place models), and the more thoughtful simulation efforts have ended by stressing how very much more we need to know to make the models good enough to be very enlightening (e.g., Aldenderfer 1981).

Paleodemography makes highly effective use of a substantial body of applied mathematics. However, the mathematics are concerned with describing and estimating vital rates and population structures, and the central problem of all demography, contemporary as well as paleo, is devising sociocultural theory that can explain, rather than merely describe, population changes. So, in this case as well, the mathematics pertain to data analysis rather than to sociocultural theory.

Thus, formal methods are extremely useful for the ordering and analysis of archaeologically recovered traces of ancient activities, but the problems of connecting the results with sociocultural theory remain formidable. I see two reasons for this, one of which is peculiar to archaeology. Sociocultural theory is about people, societies, and cultural traditions, but the objects and deposits we can actually observe have very complex relations to the things theory is about. Archaeologists have always more or less known this, and some have felt that the difficulties were overwhelming and retreated from trying to do serious sociocultural interpretation. Others have dealt with the problem by greatly underestimating it. Some of the "new archaeology" efforts of the 1960s seem to have been covertly guided by a "Hansel and Gretel" model: like the trail of breadcrumbs in the woods, people of the past thoughtfully left trails of objects to mark out just where they did. One of the great merits of the work of Michael Schiffer (1972, 1976, 1983) is that he has made us far more aware than we were that observable structure in the archaeological record is not just an incomplete, biased, and partially disordered reflection of structure in ancient sociocultural phenomena. There are many things, natural processes as well as ancient ways of dealing with discards, that can introduce new and different patterning. This means that there is usually a long and complex road between the identification and description of structure in the archaeological record and its accepted interpretation in terms of ancient sociocultural phenomena. Formal techniques can help us to move along this road, but the central insights must come from ethnoarchaeology, experimental archaeology, and taphonomy.

The second problem in connecting results of formal analyses with theory is one that we probably share with all the sciences; certainly with all the social sciences. It is, I hope, widely understood that valid objections to an "inductive" or "empiricist" approach do not hinge on the possibility of making archaeology purely deductive—of somehow sidestepping or transcending the need for inductive steps. They are ineluctable parts of the theory-building process in any science that claims to be truly about anything outside our minds. The controversy, instead, is between those who assume that the kinds of inductive steps that are most useful for theory are easy and obvious and those who argue instead that the steps are hard and not obvious. The former believe that usually "the facts speak for themselves." The difficulties with this idea have been discussed at length by other archaeologists, notably by Lewis Binford (1985 and earlier), and I will not further belabor the point. Of course, we should not be too dogmatic; many highly relevant and useful inductions, especially in everyday life, really are simple and obvious. But many others are not. This is especially the case as one approaches theoretical frontiers.

I will make one further observation on this. I do not think that the tendency for empiricists to hesitate to generalize "before all the facts are
in" is necessarily due to intellectual timidity; of being, as Kent Flannery (1967) once put it, "deathly afraid of being wrong." If the implications of data are simple and obvious, then faulty conclusions will be due mainly to faulty data, rather than to faulty reasoning. Furthermore, there is no great problem about what kinds of data are needed. The only large task is to go out and get them. If, however, the relevant inductions are not obvious and not simple, then heuristics become a major problem. It is important to begin reasoning while the data are imperfect and highly incomplete, partly because the reasoning itself is problematic and should be a subject of discussion within the archaeological community, and partly because not all the kinds of relevant data are obvious. Our early inductive efforts are immensely valuable in leading us to recognize new aspects of data that are highly relevant for theory.

Granted, then, that some kind of hard reasoning has to mediate between "raw" data (more properly, observations interpreted in terms of very widely shared and accepted concepts) and the construction and testing of explicit theory, we can ask where mathematics fits in. A view that was implicit in much of the "new archaeology" literature of the 1960s, and that probably still persists, is that statistical and/or mathematical methods can provide most of the hard reasoning. In other words, although "the facts" do not speak for themselves, there are statistical and mathematical techniques which, when applied to the data, will generate results whose relevance for theory is obvious and unproblematic.

I do not think this is so. I am especially skeptical of the variant that suggests that the harder and more abstruse the math, the easier all the rest of the reasoning will be. Often, the formal part of developing implications that are highly relevant for theory can be very simple, perhaps no more than computing and displaying some well-chosen percentages.

Have I said anything very surprising? Doubtless many who never cared for math anyhow will feel that they already understood the limits of formal approaches. But I am not suggesting that mathematical and statistical methods are of limited value; I am saying that they only complement other kinds of hard work; they do not substitute for other things. In particular, richness of mathematical concepts cannot remedy poverty of sociocultural concepts. The remedy for the latter has to come from a much greater acquaintance with the other social disciplines, most obviously sociocultural anthropology, but also and especially the work of historians.

To help us think better about the links between data and theory and the problems of induction, I propose that we distinguish the terms "uniformity" and "regularity." By a "uniformity" I mean a statistically warranted statement about a relationship or pattern between or among the values of two or more variables. The statement refers to the values of these variables that are exhibited by the elements of some population. However, usually we have only observed the values of the variables for the elements of a subset or sample of the population. By "statistically warranted" I mean that there is strong or at least satisfactory reason to think that the relationship observed in the sample is neither accidental nor an artifact of measurement errors or bias in sample selection; in other words, the observed relationship in the sample is difficult to explain unless a rather similar relationship occurs in the population.

In contrast, I suggest that we limit the term "regularity" to a logically deduced consequence of a coherent body of theory. Examples of uniformities include the observation that bodies that are free to move without constraint or friction, regardless of composition, move toward the center of the earth at a constant rate of acceleration; that tides rise and fall in very predictable patterns that are related to the positions of the Moon and Sun; and formulas that enable one to predict quite accurately the future positions of the planets. All of these become regularities only when they are deduced from a theory, such as the Newtonian laws of motion and gravitation.

Many writers (e.g., Tilly 1984:33) use the terms "regularity" and "uniformity" as synonyms. The idea of systematically contrasting them was suggested to me by Julian Steward's usage (Steward 1955:88). However, I should say at once that the contrast I am proposing is not the same as Steward's. For him, uniformities referred to resemblances between different societies that were explainable by the fact that the societies were linked by processes of cultural transmission, while regularities were resemblances that reflect the independent operation of causal processes in unrelated societies. Steward's contrast is important, and I have appropriated the term "uniformity" in a different sense only because I do not think that Steward's usage has been very widely adopted.

What I mean by a uniformity, then, is any empirically discovered pattern that does not seem to be accidental. Merely stating a uniformity does not get us very far toward understanding or explanation. "All ravens are black" is an example of a proposed uniformity. In contrast, a regularity is a pattern that is implied by some body of theory about the causal connections between various phenomena. If, for example, a body of biological and evolutionary theory implies that black ravens will be better adapted than ravens of other colors, and moreover implies that a
trait can be explained by showing that it has adaptive value, then the statement “according to this theory, all ravens will be black” is a proposed regularity. It may be helpful to remember that the word “regularity” is derived from a Latin word meaning “rule,” and thus it at least connotes, if it does not denote, the idea of something lawful. “Uniformity” seems to me a much more passive term, with no such connotation. (Note that “statistical laws” are not necessarily synonymous with uniformities. It is perfectly possible for a regularity to be a probabilistic statement.)

Now the point of all this is that the vast majority of statistical methods are concerned with (1) identifying uniformities in samples, (2) describing uniformities in samples, and (3) making the fullest and most effective use of sample evidence as it bears on what is reasonable to think about the extent to which similar uniformities may be present in populations represented by the samples. The issue is seriously confused because we are used to talking about statistical “hypothesis testing.” In most cases, statistical hypotheses are about uniformities, and these should be thought of as quite distinct from scientific hypotheses, which are about regularities. Statistical hypothesis tests are usually concerned with how good the evidence is that some proposed uniformity actually exists in some sampled population. It may be fairly easy to see that simple statistics mostly involve the search for evidence of uniformities, but this is also true of most applications of more advanced techniques, such as principal components analysis, the general linear model, and discrete multivariate analysis. Nearly all social scientists, not just archaeologists, are too prone to treat the results of multivariate analysis as statements about regularities, when in fact in most cases they should be regarded as indications of uniformities. The results do not have any simple direct bearing on regularities unless the investigator steps in strongly and knowingly to direct and constrain the analysis in the light of explicit theory.

To come at the same thing from a slightly different angle, everyone knows that “correlation is not causality.” Introductory statistics texts commonly give examples of bivariate correlations that are not accidental but are nevertheless only explainable as the outcome of more complex multivariate causal links, such as the positive correlation between numbers of sandflies and sales of ice cream at the beach. Unfortunately, we too easily forget this when the statistics become more multivariate and complex. To be sure, many of the techniques can be used to get at regularities, provided all the variables likely to be relevant have been included, all have been well measured and appropriately expressed, and all plausible linear and nonlinear forms of functional relations between variables have been adequately imagined by the researcher and tested for by the mathematical model or models. But what if not all of the above has been well done? The result is likely to be something that sort of fits, that accounts for (“explains” in the quite peculiar sense of that term used by statisticians) perhaps 40% to 70% of the data variance—far more than could often be achieved by accident, but far less than we would like. It is a dead wrong strategy to take the outcome of such procedures as an imperfect theory; that is, as the best foundation upon which to build better theory. Unless theory was strongly used to guide the analyses from the beginning, statistical results should be taken as evidence for uniformities, and uniformities should be taken as symptoms of possible regularities. The statistical analysis is only one part of the inductive process.

My discussion bears on the Salmons’ “statistical relevance” approach to scientific explanation (Salmon and Salmon 1979). I am enthusiastic about this approach, but, as the Salmons point out, just statistical relevance is not enough. It has to be coupled with postulates about causality. To use statistical relevance as the only criterion would lead to endlessly piling up well warranted uniformities, without ever getting to any regularities.

Please make no mistake about this. I am not denouncing statistical methods. Most emphatically, I am not, repeat not, suggesting that we should bypass or ignore them in trying to connect data and theory. There are plenty of examples, to be sure, where intelligent use of very simple statistics would have been far better than some ill-conceived attempt to use abstruse techniques. In general, however, statistics can be extremely useful to us. I am particularly attracted, as are many other archaeologists, to the “robust” techniques of “exploratory data analysis” (Tukey 1977; Mosteller and Tukey 1977; Mosteller et al. 1983), techniques that are often quite simple and well adapted to the nonnormality and general untidiness of much archaeological data. I simply urge that we not mistake the results of most statistical analyses for theory. We must understand that even when the results are expressed in law-like form, they are usually really uniformities that are useful clues about regularities, rather than themselves regularities. It is no mean task to identify and characterize uniformities; the fallacy lies in thinking that that is the only hard job that needs to be done to connect data and theory.

The distinction between uniformities and regularities also offers clarification of Flannery’s (1973) well-known contrast between the “law
and order" and the "serutan" (or systems) approaches. Spaulding (1973) was absolutely correct in insisting that exactly the same logic of verification applies to systems approaches as to any other body of theory. The difference is that the "law and order" approach, as described by Flannery, seems too narrowly inductive and is likely to mistake the accumulation of well-warranted uniformities for the building of theory.

What are the implications of my arguments? On the inductive side, we simply have to recognize that there is a gap between uniformities and theory that has to be bridged (or sometimes leapt) by creative imagination, by processes that are not likely to be reduced to algorithmic forms. On the deductive side, we should spend less time testing null hypotheses (whose rejection usually implies that an observed sample property is probably also a property of the relevant population) and more time assessing the fit between data and implications logically deduced from explicit theory. In doing so, it is not very useful to merely establish that the data fit the implications of theory better than would be expected by chance. It is far more useful to look at the discrepancies between data and theory. These discrepancies do not simply suggest that a theory urgently needs to be superseded by a better theory; they are apt to be among the best heuristic devices, as we re-enter the inductive phase of the scientific cycle, for suggesting how theory might be improved. Richard Gould (1980) calls this (rather unhappily I think) "argument by anomaly." It is probably better expressed as simply paying close attention to the residuals or discrepancies. One nice example is Robert Zeitlin's (1982) interpretation in sociopolitical terms of the ways in which proportions of obsidian from various sources fail to fit the implications of a simple gravity distance decay model.

Data and Data Analysis

So far I have talked about the relation between data analysis and theory. It is time to turn to the relation between observations and data analysis. Stephen Dyson has recently referred to "the criticism made by some Northern European archaeologists that archaeologists working in the Americas have overrefined their post-excavation analytical skills while not making comparable advances in field methods" (Dyson 1985:456). I disagree because I do not think our analytical skills are the least bit over-refined in relation to the demands made by the topics in culture history and theory with which we would like to deal. However, I strongly agree that there is far too often a disparity between the sophistication of the analyses and the quality of the data on which they are based. Sophisticated formal and mathematical techniques simply cannot remedy the problems caused by poor data. Mathematical concepts can, however, aid in a discussion of some of the problems.

First, there is the myth of "total survey coverage." Plog et al. (1978) show that although all the large and conspicuous sites may be reliably found in open landscapes, no surveys have yet reached the level of intensity at which a still more intensive survey fails to reveal additional inconspicuous but significant occurrences of archaeological data. In order to be able to even begin to compare the results of one survey with another, we must routinely describe the exact procedures used, and recognize that more intensive survey would always modify the picture.

Of course, a great deal of survey work does not even aspire toward total coverage; frequently some kind of sampling is used. If one must sample, I have only a few general guidelines to suggest. It would be good to do at least a quick preliminary coverage of the entire region, in order to locate all the large conspicuous occurrences; this will ensure that we do not miss Teotihuacan in a survey of the Teotihuacan Valley. Next, both statistical theory and practical expediency suggest that one should make the fullest possible use of sociocultural theory and prior knowledge of the region in order to stratify the survey area according to theoretically relevant criteria. Probability methods can then be used to select quadrats or transects that provide good representation of all strata, and more intensive survey done within these tracts. Such a scheme should provide fairly good estimates of the numbers of inconspicuous but common types of occurrences to be found in the different survey strata (Flannery 1976:159-160). There are at least two things this procedure cannot do well. First, it is a poor way to find out about occurrences that are inconspicuous and scarce (either throughout the region or within a specific stratum) but highly significant, such as Paleoindian remains. Second, it is a very poor way to find out about the spatial organization of sites relative to one another, unless the individual survey tracts are substantially larger than the largest meaningful spatial patterns. In fact, it is worse than poor, because attempts to treat fragments of patterns as whole patterns are very apt to result in outright misinformation.

There is no statistical scheme that can make regional survey by means of spatial sampling very good. It is always a very inferior alternative, to be resorted to only if imminent destruction of sites and limited resources leaves no possibility of complete or nearly complete coverage.
If spatial sampling must be done at all, then of course it should be done as effectively as possible. One should be clear about what kinds of information are to be maximized (or better, which losses are to be minimized). One family of strategies, often involving predictive modeling, is adapted to identifying the settings where sites are most likely to occur, in order to maximize the number of sites identified per person-day or per dollar of survey. Predictive modeling is thoroughly legitimate when it is used to identify high-risk localities that should be avoided by construction projects or to test and improve theory about reasons for site location choices. However, it cannot be legitimately used to “write off” unsurveyed areas as archaeologically insignificant. Furthermore, when it is used simply to maximize the number of sites located, it implies that we should look hardest where we already expect the most sites to occur, and this will lead to self-fulfilling prophecies and a very distorted picture of overall regional patterns. Predictive modeling is usually not good for designing regional surveys because, for most purposes, the sheer number of sites discovered is not the thing we want to maximize or optimize. Often, we want instead to optimize our knowledge of what is typical of each of the strata in the survey area, and to do this we must devote a substantial part of the survey effort to settings that we think are not very likely to contain many sites. Another purpose of survey is to find occurrences that are neither very typical nor very conspicuous, but are exceptionally valuable for culture history or exceptionally relevant for theory. Luck and intuition can help in this quest, but I think “luck” is mostly hard work, and “intuition” means being alert to the right hints. To find what is not common and not obvious, but very important, it helps to have good hunches, but there is just no substitute for lots of intensive survey.

One reason that I put so much emphasis on spatial sampling as an inferior option is the feeling that sampling has been much misused in cultural resource management work (Berry 1984). I suspect that there are powerful, even if sometimes subtle, pressures in favor of doing a survey quickly and not finding very much. Obviously it would never do, however, to produce a report that said “Well, we went out and poked around for a while in the survey area, never set foot on quite a bit of it, and didn’t find much of anything.” Therefore, some stuff about “stratified two-stage systematic unaligned 10% cluster sampling” should be thrown in, to bemuse bureaucrats and make the fact that it was a real quick and dirty job sound scientifically respectable. I know that a great deal of work has been done more responsibly than this, and that often the people involved have sincerely believed that explicit sampling designs would effectively provide all the relevant data for research questions and that fuller regional coverage would quickly reach a point of diminishing returns. Nevertheless, I do not think my caricature is too far from some of the work that has been done. The final word about spatial sampling is simply this: at least until we know very much more about regions and have very much stronger accepted theory than we now have, there is no possible sampling scheme for regional survey that could make the point of diminishing returns fall at much less than 100% coverage.

I have been referring to surveys intended to discover occurrences of archaeological data. How should one deal with an occurrence once it has been found? This brings me to a second pernicious myth: the “total collection.” Excavation reports frequently state that deposits were screened, and specify the mesh size. Even statements of this kind leave many unanswered questions, such as how much time was allowed to screen a liter of material, how skilled were the screeners, and exactly what criteria were used to decide which of the material that did not pass through the screen was discardable junk. But at least such statements are better than nothing. Amazingly often, descriptions of intensive surface survey begin and end with the flat assertion that “everything” was collected. Such a statement could only be sensible for sites that bear little resemblance to any site I have ever worked on. Unless the surface is actually devoid of small objects altogether, there is simply no lower limit to the size of fragments to be found. A literal interpretation of the directive to collect everything would lead to days or weeks of crawling around on one square meter with magnifying glass and tweezers. I do not, of course, advocate such a procedure. The point is simply that there has to be some cutoff; some level below which fragments are too small and too insignificant to be collected. This is not a matter of choice. The only choice is whether to make the cutoff criteria explicit or to leave them inexplicit and unreported. The latter choice greatly increases the difficulty of trying to compare one survey report with another. Furthermore, absence of any discussion of this point creates the suspicion that the matter was never even considered by the field workers, and that cutoff choices may have been made very differently by different workers, or even by the same worker at different times. In other words, this is a major source of unreliability, in the technical sense, that is usually simply ignored.

One’s reaction may be that any responsible worker, told to collect “everything” from a specified tract, will spot and collect virtually all the important stuff, and that I am only talking about variations in how many additional unimportant little bits get collected. Suppose, however, that
that term. This is, at best, obfuscation. Some of the energy spent on re-
how they were controlled. But writers often simply assert that collections
were "controlled," as if we all understand and agree on the meaning of
good, but what do they mean? They can, of course, mean something,
fully exactly which aspects of collection procedures were controlled, and
strated and largely unconsidered.

The solution is simple, at least in principle. First, do not just tell
workers to "collect everything." Establish explicit guidelines about how
far to go in collecting the tiny and insignificant. Second, spend enough
time on replications so that the workers learn how to get relatively sim-
lar results from the same surfaces, and so that one has a good statistical
basis for determining the actual reliability of the work.

I am in the process of complicating my life by using recent intensive
collections to get a better idea of the reliability of Teotihuacan Mapping
Project surface collections. When that work is completed I will be able
to document the arguments I have made above with actual examples, as
as having a far better basis for substantiating the degree of reliability
of our quantitative data, rather than merely asking others to have faith
in it. Meanwhile, I hope that the argument I have made will be clear and
convincing in its own terms. I am amazed by the number of projects that
in it. Meanwhile, I hope that the argument I have made will be clear and
convincing in its own terms. I am amazed by the number of projects that

The solution is simple, at least in principle. First, do not just tell
workers to "collect everything." Establish explicit guidelines about how
far to go in collecting the tiny and insignificant. Second, spend enough
time on replications so that the workers learn how to get relatively sim-
lar results from the same surfaces, and so that one has a good statistical
basis for determining the actual reliability of the work.

I am in the process of complicating my life by using recent intensive
collections to get a better idea of the reliability of Teotihuacan Mapping
Project surface collections. When that work is completed I will be able
to document the arguments I have made above with actual examples, as
as having a far better basis for substantiating the degree of reliability
of our quantitative data, rather than merely asking others to have faith
in it. Meanwhile, I hope that the argument I have made will be clear and
convincing in its own terms. I am amazed by the number of projects that

Another problem is the "controlled" collection. The words sound
good, but what do they mean? They can, of course, mean something,
and it is a legitimate expression when one explains clearly and thought-
fully exactly which aspects of collection procedures were controlled, and
how they were controlled. But writers often simply assert that collections
were "controlled," as if we all understand and agree on the meaning of
that term. This is, at best, obfuscation. Some of the energy spent on re-
fining terms such as "chiefdom" would be better spent in thinking harder
about what is meant by "controlled collection."

A fourth error is the belief that 100 or so objects will probably be
enough for most statistical purposes. To illustrate the trouble with this,
suppose we have collections of size 100 from each of a number of sites or loci. To give them the benefit of the doubt, suppose that each collection
was obtained by processes that reasonably approximate simple random
sampling, so that each can be regarded as a sample of a population consisting of all sherds in or on a site or specified part of a site; each sample representing a different population. Suppose we are interested in
differences in the proportion of redware sherds. Common sense tells us
that random samples of size 100 from a population with 10% redware
sherd do not always contain exactly 10 examples; they often have 8 or 9 or 11 or 12, and sometimes they have 7 or fewer or 13 or more. Ele-
mentary statistical knowledge (e.g., Blalock 1979:197-198) enables us
to be more exact and to calculate that about a fifth of such samples will
have 7 or fewer objects of a category whose true proportion is 10%.
Now consider an assemblage where the true proportion is 5%. A calcu-
lation shows that nearly one-fourth of the simple random samples of size
100 from such an assemblage will have 7 or more redware sherds. Thus,
even when the true assemblage proportions differ by 5%, it is easy to get
samples of size 100 in which the sample proportions do not reflect this
difference or even reverse the order of the assemblage proportions.

If one only wants to reliably distinguish assemblages with, say, 20%
of a given category from assemblages with 10% or 30%, then samples
of size 100 may be adequate. But, unless we have defined extremely few
categories, there are bound to be important categories that are never
more than a few percent of any assemblage. For these categories, it will
be critical to reliably distinguish assemblages with 10% from those with
5% or 15%. Samples of size 100 are at best marginal for this purpose.
If we hope to reliably distinguish between true proportions that differ by
only 3% or 4% in different assemblages, samples of size 100 are wholly
inadequate.

A further serious consequence of small samples is that correlations
between categories are not simply made less reliable. They are also sys-
tematically biased toward low absolute values. This effect, called "atten-
uation," is bound to be serious for collections of the sizes used for most
correlation matrices by archaeologists. Spearman (1904) showed this
many years ago. I discussed it in a not very accessible publication
Still another effect of small collection size is that assemblage diversity is underestimated. Many categories that constitute a moderate proportion of an assemblage can easily fail to occur at all in a small collection. Kintigh (1984) and Jones et al. (1983) give vivid demonstrations of this. For example, if one category constitutes 5% of a population, the probability that it will be totally absent from a sample of size 100 is about 1 in 21. If several categories are this scarce in the population, the probability is very high that one or more of them will be absent in a sample of size 100.

There will remain many cases where we have to try to make the best of collections of 100 or fewer objects, simply because they’re all we’ve got. We have to recognize, however, how desperately inadequate such collections are for most purposes, and we have to understand that the problem is not just unreliability. Small samples also cause serious distortions. We must be very much more conscious of the complications involved in comparing collections of diverse sizes, and we should try much harder to get larger collections. I realize that one’s perceptions of the tradeoffs involved are significantly affected in the field by heat, cold, rain, difficulty of terrain, distance to vehicles, number of bags already filled, and the like. Nevertheless, from the viewpoint of subsequent analyses, the point of diminishing returns in collection size is nowhere near 100, unless there is exceedingly little intrinsic diversity in the assemblages themselves, so that very few categories have been defined and each category constitutes a fairly high proportion in at least some assemblages. I will not attempt to suggest a collection size at which diminishing returns set in, but I feel sure it would not be less than 300 objects for assemblages of even moderate diversity.

The final data problem I want to mention is insufficient attention to distinct depositional layers in excavation. This may seem an unexpectedly down to earth topic for a paper on mathematical and formal techniques, but one comes back over and over again to the fact that mathematics do not enable one to somehow rise above bad field techniques. If anything, the more sophisticated the analyses, the more exacting are the demands placed on the data. American archaeologists nearly all repeat the formula that “arbitrary levels were used only when no natural stratigraphy could be detected.” This stock phrase conceals enormous differences in the care and skill with which the search for natural stratigraphy is conducted. Far too many of us are far too ready to resort to arbitrary levels. Field records, including the timely drawing of profiles, also often leave a great deal to be desired.

The State of Some Major Topics

My discussion of specific technical topics is necessarily brief and selective. I will try to say something useful in a few words about matters that deservedly receive whole chapters in review volumes.

Spatial Analysis

Spatial analysis is a domain strongly affected by scale. I suggest that the intra/intersite dichotomy is too crude and that at least three overlapping levels can usefully be distinguished. The largest, generally involving whole regions, is concerned with distances on the order of 1 to upwards of 1000 kilometers, and is dominated by concepts from economic and mathematical geography; especially by "central place" and other models of hierarchical patterns of site size, importance, and location.

I have two comments. One is that doubts about the importance of retail marketing in ancient societies do not challenge the potential value of some sort of central place model, since there are many other kinds of services and activities that may be associated hierarchically with specific places. Second, I am struck by the frequency with which arguments about whether a particular effort to apply a geographical model has in fact been “successful” are carried out without even a hint of the possibility that formal goodness of fit statistics might be appropriate. I suppose the reason is that it is exceedingly difficult to quantify the vague notion that some hierarchical spatial models fit the data better than others do. One has to assess the joint significance of deviations in both the hierarchy and the spatial locations, as well as often serious measurement ambiguities. Nevertheless, formal measures of the goodness of fit of spatial models should be given more attention.

At the other end of the spatial scale are studies involving distances on the order of 0.1 to perhaps as much as 100 meters. These usually involve small sites or individual structures or living surfaces within sites.
Investigations at this scale are dominated by the Schifferian transforms. If spatial patterning in the archaeological residues has been identified and described, what does it really tell us about who did what, where? It is crucial to dig and record in ways that will give the best possible chances for distinguishing between de facto, primary, and secondary refuse; and to do our utmost to discriminate layers reflecting a sequence of occupations, activities, processes, and events. Problem-directed ethnoarchaeology and experimental archaeology are essential.

The intermediate spatial level, concerned mostly with distances on the order of 10 to 10,000 meters, is the least well defined. It involves the study of large sites, especially cities, and also clusters of small sites and site catchment areas. Problems of redeposition are less serious than for smaller scale studies, but not negligible. Relevant theory depends strongly on the type of site, and for urban sites theory comes largely from the history and sociology of more recent nonindustrialized cities.

A serious technical problem in spatial studies, regardless of scale, is that most statistical techniques deal only with the values of two or more variables that jointly occur on each of a number of independent cases, without considering the spatial proximity of the cases. If, for example, a surface is divided into a number of cells and we know the number of endscrapers and the number of backed blades found in each cell, then it is easy to do a bivariate plot of their joint occurrence, which we can inspect to decide whether it would be appropriate to compute a correlation coefficient, a regression equation, and so on. But this approach ignores two questions: what are the consequences of making the spatial cells larger or smaller, and what is going on in the cells close to a given cell? Statisticians have done a good deal of work on these questions, but the results are harder to use and less well developed than the techniques that ignore spatial relations.

One question, of course, is whether an apparent spatial pattern could easily be just accidental. Berry et al. (1983) sketch what looks like an impressive technique to answer this question. However, many spatial patterns are clearly not altogether accidental, and we are more in need of ways to describe patterns. Whallon (1984) explains a key difficulty in many approaches: they derive from models that postulate processes operating globally over the study area, whereas realistic archaeological models must postulate multiple processes that are diverse in the shape and size of the affected areas, as well as diverse in their impacts on these areas. A great deal remains to be done to develop methods that are based on realistic models and that enable us to go with confidence much beyond our considerable inbuilt pattern-recognition abilities.

Problems of Multiple Processes

Problems of multiple processes are not limited to spatial analyses. Read (1985) argues in detail that statistical models generally presuppose populations generated by single processes; a presupposition that is rarely likely to be true of archaeological databases. Perhaps it would be better to take an “information” approach for a vague guiding model: the task can be seen as that of discriminating the archaeological “signal” of a target process amidst the confusion of noise, distortion, and crosstalk from the signals of other processes. I do not think, however, that this vague guiding model can easily be implemented by adopting well-developed communications engineering methods for filtering signals, such as Fourier techniques. One problem is that although such techniques postulate multiple concurrent processes, the processes are nevertheless, in a strong sense, global. That is, each process is seen as operating over the whole study area. The approach does not really take account of the possibility that totally different things may be happening in different places. It may be more profitable to think and experiment more about the archaeological consequences of single processes, in order to deduce the probable consequences of a mix or sequence of processes. Also, the common-sense recognition that some deposits are more informative than others can be made more exact by suggesting that we try hard to identify deposits that reflect the operation of relatively few processes, possibly only one process.

The problem of multiple processes recalls my earlier discussion of uniformities and regularities. Mechanically subjecting the kinds of data generated by customary archaeological procedures to standard multivariate statistical techniques is likely to yield ill-formed and nearly meaningless uniformities reflecting the confounded operation of diverse processes, rather than approximations to regularities in sociocultural phenomena.

Classification

A volume edited by Whallon and Brown (1982) provides a range of views about typology. One purpose of classification is data management.
Computer technology is increasingly relevant for this. Other purposes come under the general heads of organizing and expressing similarities and differences between objects in a single assemblage (that is, a data set that as near as possible pertains to people in close interaction doing much the same things over a relatively short time), similarities and differences between objects in different assemblages, and similarities and differences between assemblages. The similarities and differences that can be detected will depend on how objects are observed; how they are observed will depend on what we think it important to notice; and what we think it important to notice is fundamental and will depend on our research problems. We also have a choice of different ways to express similarity/difference patterns, and each choice is better for some problems than for others. Of course, ideal techniques have not yet been developed for most problems, and there are some techniques that are inferior for any reasonable problem. But techniques and observations are not fundamental; problems are fundamental, and they should guide our selection of techniques and observations. We have been saying this for 40 years now (Brew 1946). Most work on classification is still centered on a specific technique or dominated by a few of the reasonable purposes.

It would be useful to look more systematically at the various implications for techniques of all the currently known reasonable purposes of classification.

Simulation

I have already said the main thing about simulation. It is one of the few mathematical or formal approaches that can be used not just to analyze data in ways relevant for theory but to actually express theory. Nevertheless I have been disappointed by most simulation efforts. Publications that never get beyond the verbal stage and that never quantify the proposed relationships or actually lead to a simulation also rarely get much beyond expressing some fairly simple ideas in nonsimple language (Lowe 1985 is an important exception). Many of the simulation efforts have taken ready-made computer systems whose built-in limitations and presuppositions were grossly (and sometimes grotesquely) mismatched with the phenomena the simulation was supposed to elucidate (I will not name awful examples, since I have avoided that elsewhere in this paper). I think the simulation bubble has broken now, and a more sober attitude prevails (e.g., Aldenderfer 1981; Cordell 1981).

An essential point about simulation is that it is not a body of sociocultural concepts; it is a way to embody concepts. In most simulations, the concepts have come from some variant of systems theory. Contrary to the views in many quarters, I do not think that systems theory, at least the versions used by archaeologists, has furnished us with a very rich body of highly relevant concepts. It is useful up to a point, but it offers our imaginations very thin sustenance about sociocultural phenomena. We can only be nourished by reflection on ethnographic and historical data. If we have the knowledge and imagination to formulate better models and good enough data so that none of the key input values have to be wildly guessed, it should be technically feasible to create very useful simulations.

Conclusion

Both the number and sophistication of archaeological publications that are concerned with mathematical or formal methods are greater than they were a few years ago. However, there are still very few, if any, mathematical or formal techniques that are applied by archaeologists both frequently and also in ways that are simultaneously technically correct, appropriate to data and problems, and highly useful. In that sense, the art is still in a weakly developed state. Most publications fit into one of three broad categories. First are those that make substantial use or misuse of mathematical or formal techniques in the service (or alleged service) of significant theory. A second, and I think much larger category, consists of those publications that describe and advocate a technique, method, or approach; illustrating it with “data” that are either made up or of little intrinsic interest. The third and also large category consists of publications that criticize or defend other publications belonging to one of the first two categories. Some of the debate is needlessly argumentative or petty, but much of it raises important points. There is probably also somewhat better pre-publication review and screening than there was 20 years ago. I think it is harder for incorrect or inappropriate work to gain acceptance now than it was in the 1960s. This seems to me the major advance; not a long list of widely accepted accomplishments, but advances in the conceptualization of problems, the sophistication of experimentation, and the level of self-criticism.

However, in spite of the growing number of archaeologists who have some acquaintance with mathematical and formal methods, there
remains a high proportion who either avoid these methods or who make serious errors in their attempts to apply them. It is unrealistic to expect most archaeologists to become mathematical experts, but we still sorely need better and more general teaching of some simple techniques and, above all, the basic logic of statistical inference. These are not topics that can be treated casually, nor can students be expected any longer to learn them on their own.

Finally, I will repeat what I have suggested throughout. Improvements in mathematical/formal techniques and their applications are much to be encouraged. But what we need above all are data, sociocultural theory, and understandings of the connections between archaeological evidence and past behavior that are all worthy of the techniques.

Acknowledgments

I thank David J. Meltzer and anonymous readers for a number of comments on the draft of this paper prepared for presentation in Denver. They have aided materially in the preparation of the final version.

Literature Cited

Aldenderfer, Mark S.

Berry, Kenneth J., Kenneth L. Kvanme, and Paul W. Mielke, Jr.

Berry, Michael S.

Binford, Lewis R.

Blalock, Hubert M., Jr.

Brew, John O.

Clarke, David L. (editor)

Cordell, Linda S.

Cowgill, George L.


Cowgill, George L., Jeffrey H. Atlschul, and Rebecca S. Sload

Dyson, Stephen L.

Flannery, Kent V.


Flannery, Kent V. (editor)

Gould, Richard A.

Hodder, Ian (editor)


Mosteller, Frederick, and John W. Tukey 1977 *Data Analysis and Regression*. Addison-Wesley, Reading, Massachusetts.


Tukey, John W. 1977 *Exploratory Data Analysis*. Addison-Wesley, Reading, Massachusetts.

