MATHEMATICS, STATISTICS AND ARCHAEOMETRY – THE LAST 50 YEARS OR SO

M.J. Baxter

Division of Physics and Mathematical Sciences
School of Science and Technology
Nottingham Trent University
Clifton Campus
NOTTINGHAM NG 11 8NS
U.K.

ABSTRACT

This review of developments in the use of mathematics and statistics in archaeometry over the last 50 years is partial, personal and ‘broad-brush’. The view is expressed that it is in the last 30 years or so that the major developments have taken place. The view is also expressed that, with the exception of methods for analysing radiocarbon dates and increased computational power, mathematical and statistical methods that are currently used, and found to be useful in widespread areas of application such as provenance studies, don’t differ fundamentally from what was being done 30 years ago.

KEYWORDS: BAYESIAN, COMPUTATION, DATA TRANSFORMATION, DATING, LEAD ISOTOPE ANALYSIS, MATHEMATICS, MULTIVARIATE DATA, PROVENANCE, STATISTICS,

INTRODUCTION

The brief for this essay was to write an account of mathematics and statistics in archaeology and, more specifically archaeometry, over the last 50 years or so since the journal *Archaeometry* was founded. Having relatively recently published a book (Baxter 2003) that says most of what I wanted to say on the subject of statistics in archaeology, including archaeometry, from a technical perspective, I want here to explore some other aspects from a fairly non-technical point of view. What follows eschews some of the standard conventions of academic publication; avoids, for the most part, technical detail; and presents a partial and personal view of the way mathematics, and more particularly statistics, has been used in archaeometry. I am missing out a lot, but want to concentrate on generalities.

As this will be a personal view, some background is appropriate. Whilst not covering the full 50 years or so that this edition of *Archaeometry* celebrates, my interest in statistics and archaeology dates back some decades. The first serious statistical seminar I attended (Birmingham University 1968-1971) was by the late Professor David Kendall on a seriation problem in archaeology. Then, and when doing my PhD at Edinburgh University (1971-1974), I spent vacations working on archaeological excavations, and was sufficiently enchanted that I worked as an archaeologist in various capacities for two years after completing my PhD, before returning to statistics.

Edinburgh in the 1970s was a good place for a statistician with an interest in archaeology to be. The use of mathematics and statistics in archaeology has a history that some date to the late 19th century. Many commentators, me included, would cite the early 1950s as the important developmental phase, with lots of influential work taking place in the 1960s, and an explosion of interest in the 1970s when books dealing explicitly with the use of mathematics and statistics began to be published, several by Edinburgh University Press. Clarke's (1968) seminal text, *Analytical Archaeology*, which involves statistical ideas but is much wider-ranging, appeared slightly earlier, but it is the 1970s where I think things 'take-off'.

Owing to the vicissitudes of career development, I did not start publishing on statistical applications in archaeology until the late 1980s, since when it has been my main research interest. Baxter (1994, 2003) contains material on my views of the development of statistics in archaeology generally, and what follows involves a degree of self-plagiarism.

After a brief historical survey, a number of topics that interest me have been selected to illustrate particular points. Those involving provenance and related studies of various kinds are areas in which I have research experience, but I have strayed into important areas such as dating and Bayesian methods that I have not done work in. Technical details are dealt with largely by reference.

EARLY DAYS IN QUANTITATIVE METHODOLOGY AND ARCHAEOLOGY

Kendall, as well as publishing on seriation, worked on the statistics of the megalithic yard. Most archaeologists did not believe in the ubiquitous existence of this unit, but it was not until the work of Kendall (1974) and Freeman (1976), which involved the development of new statistical methodology, that convincing non-archaeological evidence against its existence was adduced.

Kendall (1963) made a case for Flinders Petrie, in the late 19^{th} century, as one of the first to apply essentially statistical methods to archaeological problems, but as already noted, many would date the serious impact of statistics and mathematics on archaeology from the early 1950s in the work of Robinson (1951) (on seriation), Spaulding (1953) (on typology) and others, much of it published in *American Antiquity*.

American scholars took the early lead in applying quantitative methods to archaeological problems, Lewis Binford in the 1960s being particularly influential (Binford 1964; Binford and Binford 1966). Apart from Clarke, the early seminal text about applying mathematics and statistics (and computing) to archaeological problems was Doran and Hodson's (1975) *Mathematics and Computers in Archaeology*.

Most of the work to which I have made reference so far involved few or no archaeometric applications. 'Archaeometry' has been defined as the 'application of physical sciences to archaeological problems and issues' (Dunnell 2000). Mathematics and statistics (and I distinguish between them) involve the application of the methods of these disciplines to the ideas and data generated by archaeometric analysis. This can involve the mathematical modelling of the processes used to generate the data, or the archaeological processes involved, or data analysis based on the information generated. The use of mathematics and statistics in archaeometry developed in parallel with uses in the wider world of archaeology and was not, for the most part, driven by archaeologists as far as I know.

The use of statistics in archaeometry has a 'prehistory', but I think it was only from the mid-1970s that statistics began to make a serious impact in terms of accessible publications. I have in mind Bieber *et al*. (1976) which I frequently cite. This is in part based on Sayre (1975). The latter is unpublished in a journal but I have lost count of the number of times I have been handed a copy by co-researchers anxious I should know of the work. These papers, among other things, discussed mathematical and statistical models accounting for the structure of analytically generated data, and introduced statistical techniques such as the use of Mahalanobis distance.

These papers and others emanating from the Brookhaven Laboratory established a particular approach to archaeometric data analysis that has been refined in later years (with the aid of increased computer power) that does not – I think – differ fundamentally in the essentials (e.g., Glascock 1992; Neff 2002).

It has always intrigued me that different laboratories/institutions developed their own statistical software because of the perceived inadequacies of commercial software for dealing with what were perceived as specifically archaeometric problems. Different researchers were influenced by different traditions, so a variety of different approaches grew up and everyone did their 'own thing'. I've never been entirely convinced that archaeometric data analysis is inherently different from and/or more difficult than other forms of statistical analysis.

Things have changed. The advent of powerful and programmable open source statistical software such as R makes the need for special-purpose lab-based software redundant (although some scholars will remain wedded to the methods they 'married' at an earlier age). I've argued (Baxter, 2001) that over time there has been a convergence in statistical methodologies employed by scholars emerging from different intellectual traditions. For example (and they may not agree), and in the context of provenance studies, I see – from a statisticians viewpoint – more similarities than differences in the approaches of the Brookhaven group, Pollard (1986), Beier and Mommsen (1994), Glascock (1992) and Neff (2002). There are, I think, differences in 'philosophy' but that is not the subject of this paper.

The papers cited above, and applications of the methodologies, represent what I regard as 'best practice' (at the time they were published) – even if I don't always agree with the detail. My early view, which I see no reason to change, was that many applications of statistics to archaeometric problems were fairly dire. The general mantra was 'I've got an analytical technique; I've got some artefacts that happen to be lying around; so lets publish the numerical results of the analyses'. The idea of a research design or an archaeological problem to be addressed was often conspicuously absent. If I'm honest I'd have to say that publications in *Archaeometry* in its earlier years were often guilty of this. The statistical aspects were often also wanting; frequently, what software happened to be available was used – whether commercial or designed locally, and whether appropriate or not – with little or no explanation of why. I am ignoring here instances where the statistical treatment was simply wrong.

And I have a problem here (and I am not alone), already alluded to. It is that vast amounts of archaeometric publications and the subsequent data analysis do not seem to address any particular 'archaeological problems and issues'. To put this another way, there is much past 'archaeometric' literature that treats the 'technology' (including the mathematics and statistics) as subjects worthy of study in their own right. This can be sterile. I think that archaeometry as a subject can only be justified if it is eventually harnessed to its usefulness in addressing interesting archaeological problems (I've heard the opposite view expressed more than once). I am not suggesting that technical/methodological developments are not worth addressing in their own right and have done so myself, but – if they are worthy of being called 'archaeometric' – they should eventually be used to address real archaeological problems.

The publication of *Mathematics in the Archaeological and Historical Sciences* (Hodson *et al*. 1971), a collection of conference papers, was a landmark. Most of the papers – and this is not an original observation – have a more mathematical than archaeological flavour. It depends on how you define 'archaeometry' – is it the subject matter or the methodology (in which I include the statistical methodology) – but there is little, if anything, in the volume that I suspect most readers of *Archaeometry* would regard as 'archaeometric'. Some of the mathematical and statistical methodology used remains useful for archaeometric data analysis, but is not applied in such a way in the book. *Mathematics and Computing in Archaeology*

(Doran and Hodson 1975), already noted, has little on archaeometric matters (a short section on radiocarbon dating).

Fast-forward to 1988 and the publication of *Quantifying Archaeology* (Shennan 1988), one of the first and best undergraduate texts on the subject, which has subsequently been revised (Shennan 1997). This is not to disparage the, almost invariably interesting and good, texts I know to have been published elsewhere; I want to make the point, in the context of this article, that there is (on a quick revisit) little if anything about archaeometric matters.

We have a disjunction here. Mathematics and statistics have been widely used for archaeometric analyses for some while, but the subject (archaeometry) does not feature much in the major quantitative archaeology texts that I know of before the mid-1990s.

This is not to say that there were no publications that involved mathematics and statistics in archaeometry – simply, they may have been taking place in collections, conference proceedings and so on that were unknown to, or simply irrelevant to, archaeologists with real archaeological problems.

The points suggested so far are:

(a) A lot of early archaeometric publication (including any data analysis) had little relevance to real archaeological problems (and it is these that mainly justify the existence of archaeometry as a discipline). You need the technological development, but this should not take place in an intellectual vacuum.
(b) The mathematics/statistics is used to interpret the data generated by the technology, but you need to have archaeological (as well as technological) processes to model and questions to address (hypotheses if you wish) or data to display in an (archaeologically) informative way.

I think the use of mathematics and statistics in archaeology developed almost independently of the same in archaeometry, and that this was not a good thing. Mathematics can be a thing of beauty, and worthy of study in its own right; but in the context of archaeology it needs to serve a purpose. The development of statistical ideas and thinking in the 20th century is possibly one of that century's more important intellectual achievements; the mathematics of statistics can be ugly and sometimes pointless but I think the justification for statistics ultimately lies in applicability to data analysis, including that in archaeometry. Not everyone would agree with this.

A reviewer of an early draft of this paper has suggested – and I paraphrase – that a lot of archaeometry, and the use of mathematics and statistics within it, was living within its own 'little self-contained bubble'. I did not coin this phrase, but think I agree. The more specific idea that informs this view is that the archaeometric literature, including its mathematical/statistical aspects, is largely innocent of, or uninformed by, theoretical debates about archaeology that have been raging since at least the 1970s, particularly the ideas of 'post-processual' archaeology. Whether this is because of ignorance of what has been going on, or a deliberate choice to ignore it I don't know.

Shanks and Tilley (1992 - the second edition of their 1987 text) was particularly influential. I find much of it unreadable and disagree with those bits I think I understand. It launches a fairly savage attack on the use of ‘science’ and more specifically mathematics and statistics to address archaeological problems. In the general archaeological literature Orton (1999) and Ringrose (1993) have responded to this attack, but I find it difficult to think of papers in archaeometry that have engaged with this debate that has been taking place in the wider archaeological world for some considerable time.

As anecdotal evidence, I have sat over breakfast/lunch/dinner at conferences and listened to ‘scientists’ being rather scathing about archaeologists understanding of the ‘science’ and, in effect, saying ‘leave it to us’. I’ve also heard archaeologists being equally scathing about the ‘scientists’ knowledge of archaeological problems and theory. Archaeometric practice ought to, but does not always, provide a bridge between the ‘two cultures’, in the form of fruitful collaboration. And I think this has much improved in recent years. I have come close to making some horrendous mistakes in my time by undertaking statistical analyses of (archaeometric) data in advance of discussion with the archaeometrician and archaeologist involved, and was saved from humiliation by their comments on my initial efforts (Baxter 1994:219)

MATHEMATICS AND STATISTICS IN ARCHAEOLOGY

Some of what follows is about the use of multivariate statistics in archaeometry. It is what I know about and multivariate analysis is fairly ubiquitous. Simpler statistical devices (the use of means, standard deviations, standard errors, assessment of error etc.) are also ubiquitous and well-understood, so readers need to be aware that this has a partial focus on what has been an important component of mathematical and statistical practice in archaeometry.

Cluster analysis is the most widely applied multivariate technique in archaeometry. I am on record as describing (Baxter 1994:179-180) many archaeometric applications as ‘boring’, lacking in ‘quality’, ‘voluminous’, and ‘uncritical’, It would be invidious to cite specific articles, but a lot of papers in the journal *Archaeometry*, and publications in Archaeometry conference proceedings would not be exempt from this criticism (and indeed largely informed my views on this subject).

Have things changed much since my original survey in the early 1990s. The answer is ‘yes’ and ‘no’. A lot of current applications in archaeometry of multivariate analysis (not just cluster analysis) do not differ much from what was being done in the 1970s. To use a modern phrase – which I dislike – if the method is ‘fit for purpose’ then I don’t have a problem with it. If the results produced are obvious anyway they are at least not wrong. It is the archaeological interpretation that is paramount, unless a study is a purely methodological one.

Where I think there has been a major improvement in recent years is that methods of multivariate statistical analysis are treated as the start- rather than end-point of data analysis. To make this clear, there were far too many papers around that presented the results of a single cluster analysis (and its interpretation) as the end-point of the statistical analysis. More intelligent approaches now take the initial cluster analysis (or whatever other method is chosen) as a start-point and then modify/re-interpret the results using other statistical methods and/or specific archaeological knowledge about the problem at hand. I have specifically in mind approaches similar to those initiated at the Brookhaven Laboratory and paralleled (from a slightly different tradition) by Hans Mommsen and his colleagues (e.g., Beier and Mommsen, 1994). This is not an exhaustive list!

LEAD ISOTOPE RATIO ANALYSIS (LIA)

For an account of lead isotope ratio analysis (LIA) in archaeology, including some of the mathematics, see the relevant sections in Pollard and Heron (1996) and Pollard *et al.* 2007). I want to comment on some statistical aspects of LIA at a general level.

The following account of the problem is simplified. Specimens from presumed sources of metal used in the manufacture of ancient artefacts are obtained, giving rise to a sample *n* specimens (or cases).

For the source this defines a three-dimensional cloud of points which is a sample of the space, in three dimensions, occupied by the lead isotope field for the source. The actual field, the *population* in statistical terminology, needs to be estimated from the sample. The general hope is that fields for separate sources are reasonably distinct. If so, this allows for the possibility that isotope measurements on an ancient artefact can allow the artefact to be matched to a single source or none at all.

This account conceals a variety of archaeological and scientific problems, not discussed here. The statistical aspects seem straightforward. Far from it. When I became involved in this I was astonished by the amount of debate taking place over the issues involved. Statistical aspects are covered in Chapter 18 of Baxter (2003) which includes references up to 2001. For a general flavour of the debate about various issues, which I recall being described as ‘robust’ but which have also heard described as ‘vituperative’, dating from the early 1990s *Archaeometry* **34** and **35** make interesting reading (there are too many papers to note individually and plenty of other papers in similar vein published elsewhere).

A major problem was that sample sizes, *n*, were usually far too small to justify subsequent statistical treatments based on the assumption that the (population) isotope field had a three-dimensional multivariate normal distribution. Independent work by Scaife (1998) and Baxter (1999), using data sets where *n* was larger than had usually been the case, showed that the normality assumption could not often be sustained (see, also Pollard *et al.* 2007:192-4, and the references there).

The normality assumption was used to justify the estimation of the extent of isotope fields using confidence ellipsoids of the kind used in other forms of provenance studies, assignment of artefacts to fields, and the use of multivariate statistical techniques such as discriminant analysis. Analysis was often done after the removal of inconvenient ‘outliers’ that appeared to contradict the normality assumption.

Given that the data are only three-dimensional, some scholars questioned the need for ‘sophisticated’ statistical methods to examine differences between fields, and associate artefacts with them (even if this was considered a valid exercise). The argument was that, given the low-dimensionality of the data, simple graphical methods sufficed to address the questions being asked. This is possibly now the general consensus (that is, I haven't seen any recent papers in which the 'correct' form of statistical analysis is an issue). It possibly took statistical analyses using more complex methodology (e.g., Baxter and Gale 1998; Scaife *et al.* 1999) to persuade some researchers that simple methods were adequate.

This is another instance of a point made elsewhere in this review, that however sophisticated the analytical technology, you don't always need complex statistical methods to analyse the data generated. Some analysts when confronted with multivariate data reach for their favourite statistical technique without looking to see if simple methods (e.g., bivariate graphs) will answer the research question. The opposite problem can exist, in that simple methods are used when more complex ones may be needed.

I cannot resist commenting on the 'sociology', or even 'sociopathy', that attended the debate about LIA interpretation in the 1990s. As far as I could tell the primary data was produced in a limited number of centres, and as well as owning the facilities used to produce the data, the producers – as it seemed to me – also felt they owned the data, *and* the interpretation of the data, *and* the statistical methods used to aid interpretation.

Scientists, of course, publish their data (eventually) at which point it is in the public domain, and anyone can comment on it. In other areas of archaeological endeavour (e.g., the publication of site reports) authors are often grateful for their work to be noticed, even if others disagree with aspects of the interpretation. In the case of the LIA debate such gratitude was not apparent. Questioning of the data quality, interpretation and statistical methods led to what could be charitably described as a 'resentful' response by the produces that gained in bitterness as the 1990s progressed. I became involved on the periphery, and as a non-partisan, at a fairly late stage. It is the only time in my career when I have been urged to withhold publication of papers – the academic quality of which was not in doubt – simply because it was perceived by some scientists that the findings might give succour to other scientists whose views on LIA interpretation differed.

DATA TRANSFORMATION AND CONSTANT SUM CONSTRAINTS

This is essentially a mini-essay on aspects of the subject of data transformation in archaeometry. This has been dealt with at some length in recent issues of *Archaeometry* (e.g., Aitchison *et al.* 2002; Baxter and Freestone 2006).

The application area in mind is provenance and related studies. A sample of n cases is available and the elemental/oxide composition of a case is measured with respect to p variables that are a subset of the periodic table. Measurements can be percentages and/or ppm (for trace elements) but either can be converted to the other if needed. Let the variables be denoted by $x_1, x_2, \ldots, x_p$.

For simple applications, such as bivariate plotting, the data in their raw form may be sufficient. In some instances a logarithmic transformation is useful for making plots more readable, and some researchers do this as a matter of course because they think it renders the data more normally distributed. For more complex multivariate analyses some form of data transformation is usually desirable. This is needed to avoid the predictable dominance of a subset of the variables in widely used multivariate methods such as principal component (PCA) and cluster analysis.

The data envisaged here are compositional. If for each case the sum of the variables is (approximately) 100% the data are fully-compositional, and subject to a constant sum constraint (typically 100% or 1 – we shall assume 100%). A sub-set of such a set of variables is sub-compositional.

Since multivariate compositional data began to be analysed in archaeometry there has been debate about appropriate forms of data transformation (e.g. Pollard 1986). Aitchison *et al.* (2002) make a case, grounded in mathematical arguments going back to at least the 1970s, for using log-ratio analysis (LRA). For fully-compositional data, variable values for a case are scaled either by dividing by the geometric mean of values for the case, or by the value for a selected variable, and then taking logarithms. For sub-compositional data, values are rescaled to be fully compositional before doing this.

A main idea is that compositions give information on relative values and that some form of ratio transformation is needed to respect this. This idea has been used in archaeometry since at least the 1970s but for pragmatic rather than theoretical reasons (see the review in Baxter and Freestone 2006).

The knowing use of log-ratios for the reasons advocated by Aitchison *et al*. (2002) is not common in archaeometric publications. Buxeda i Garrigós, with collaborators, (e.g., Buxeda i Garrigós *et al.* 2001, 2002, 2003) is possibly the most enthusiastic exponent of the method, the references cited being to recent publications in *Archaeometry*.

By happy chance it turns out that analyses of sub-compositional data of trace elements based on a logarithmic transformation are mathematically more or less the same as a log-ratio analysis. This approach to data analysis has a long and honourable tradition in archaeometry dating back to at least Sayre (1975) (see the section on 'Early days in quantitative methods in archaeology'). Thus, a lot of what has been done in practice is in the spirit of log-ratio analysis or approximates to it, even if not knowingly so.

However, the need for logarithmic transformation (possibly of ratios), even when mainly using trace element data, has been disputed (e.g., Beier and Mommsen, 1994), Some researchers, myself included, who are familiar with LRA and do not dispute its theoretical underpinning, either prefer to use some form of standardized data (incorrect according to LRA advocates) instead of or along with LRA. Why?

The basic answer is simple – it is that LRA can provide poor substantive results when applied to typical fully-compositional archaeometric data sets. Examples are given in Baxter and Freestone (2006) and papers they refer to. Often the use of standardized data leads to valid substantive conclusions much more rapidly. I am mainly thinking of situations when exploratory pattern-seeking methods such as PCA or cluster analysis are used, and these are the most common multivariate methods used in archaeometry.

Debate about the appropriate choice of transform will continue to surface in archaeometric publications. I would certainly not wish to discourage anyone from experimenting with LRA. Typically LRA is applied without subsequent standardization and this is often where the problem lies. On the log-ratio scale with methods such as PCA too much importance is given to variables with a low absolute presence, leading to results with no obviously useful substantive interpretation.

Within an LRA framework some form of variable selection may be needed to get sensible results. Another possibility is to weight variables differentially, and I am aware of research along these lines. Omitting variables is an extreme form of weighting. Standardization of log-ratios is also possible but, based on unpublished research, I have found that results are then very similar to those obtained with standardized data.

Much statistical analysis in archaeometry is exploratory in nature. Many successful applications of LRA to other areas involve well-specified models for the data. I'd distinguish between model-based and exploratory methods, and have seen the view expressed more than once that the choice of methods to use in an exploratory application is a domain-specific one. In other words you use what appears to produce the most interpretable substantive results on a reasonably consistent basis. As far as archaeometric applications go there is as yet no body of evidence that LRA does this (and several examples where it doesn't), and plenty of instances where the use of standardized data 'works'. It would be useful to have more case studies where LRA is used to get a better picture.

To conclude this section on a slightly more technical note, PCA is a technique that involves mathematical operations on the correlation or covariance matrix of a set of data – assume a correlation matrix for present purposes. If you take a fully-compositional data set, select a sub-composition and rescale it to sum to 100%, then this changes the correlations between variables and can lead to the PCAs of the fully- and rescaled sub-compositional data giving different results. Advocates of LRA regard this as incoherent. Examples abound in the statistical literature that I don't disagree with. The main point though, is that archaeometricians working with sub-compositions rarely rescale the data, so that analyses of standardized data in PCA are based on exactly the same correlations as apply to the fully-compositional data set. Differences in results may occur (often they are similar), but this is a function of the variables used in the sub-composition rather than 'incoherence'.

I'm suggesting here that some of the theoretical concerns raised by advocates of LRA don't necessarily apply to the way archaeometric data analysis is often carried out. It is not that I dispute the theory; it is that I think it may not always be relevant to the practical concerns of exploratory data analysis.

I recall having a discussion with Jaume Buxeda i Garrigós some years ago about the relative merits of LRA and analysis of standardized data in exploratory archaeometric data analysis; my view was that whatever the theoretical objections the latter often 'worked well'. Jaume's point of view, if I recall correctly, was that just because an 'incorrect' method sometimes worked well was not a reason for using it, particularly if more 'correct' methods such as LRA were available.

While I understand this position, in terms of practical archaeometric data analysis more subtle considerations may apply. Would you rather work with an 'incorrect' method that, on the whole produces useful results, or a 'correct' method that is less consistently useful? It can be argued that the 'correct' method is often being applied incorrectly, but I don't think so for the applications to exploratory data analysis I have in mind. An analogy I have used before is with single-linkage cluster analysis. According to some accounts this is one of the few theoretically sound methods of cluster analysis available (it satisfies theoretically prescribed desiderata). It is used very little in archaeometry because the results produced are often un-interpretable, unless the data structure is so obvious that any sensible method would reveal it.

I know how I would currently answer my own question posed above, but would not wish to be dogmatic. Sometimes different approaches to data transformation can yield different results, equally valid, that shed different light upon the data being used. In the next 50 years of *Archaeometry* – I hope sooner rather than later – it might be useful for more comparative studies to be undertaken to establish if any one approach is uniformly 'best'. I rather doubt if there is such a 'best' approach .

DATING METHODS

My main expertise, as noted, is in the application of multivariate statistics to archaeological and archaeometric data. If I had to pick out the single most important contribution of mathematics and statistics to archaeology in the last 50 years, however, I'd single out the development of dating methods and, in particular, Bayesian methods that are used to recalibrate 'raw' radiocarbon dates.

I've written on this (Baxter 2003:187-199) but am not an expert so what follows is both simplified and derivative. Radiocarbon dating led to a revolution in archaeology. It produced 'absolute' dates that in some cases radicalised previous interpretations of the data. The technique can be explained mathematically (although not here).

Unfortunately the assumptions that underpin the mathematics can be wrong so the dates are wrong. Recalibration aims to correct for this. As is often the case – in order to get a usable mathematical model for some phenomenon – you make simplifying assumptions that you either don't believe in, or which turn out to be wrong, do the mathematics, discover that there are problems, and take it from there.

Most archaeologists who use radiocarbon dates know, I hope, what the problems are. Some are summarised in Baxter (2003:189). What you want to do is associate a radiocarbon date with a calendar date; this is not easy (one reason being that the isotope ratios used in the dating process are not as well-behaved as you would like them to be). It means, for example that the radiocarbon date can be biased and a particular radiocarbon date can correspond to more than one calendar data as the curve you use for conversion is 'wiggly'.

How is chronological sense to be made of the information available? To repeat an earlier comment, archaeological research does not take place in a vacuum. Often multiple radiocarbon dates are available; contextual information is available (i.e. you know the relative dating of the material sampled, or may know that it comes from the same context; and with luck you may also have absolute limits for the date(s)). The major achievement that has been achieved by (Bayesian) statisticians is to provide a way of integrating this kind of information with the raw radiocarbon dates to come up with realistic calendar dates and the sequencing (you are being spared technical details here). The practically important point is that all this is available to archaeologists in a fairly painless way through packages such as OxCal (Bronk-Ramsey 1995) and others that hide the – in some cases formidable – mathematics from the less numerate archaeologist.

As ever, this sort of research has its prehistory. I would pick out as fundamental a paper published in the *Journal of the American Statistical Association* by my late colleague Dr John Naylor and Professor Adrian Smith, at Nottingham University at the time (Naylor and Smith 1988). The paper is flawed – it gets the archaeology wrong (subsequently corrected) – and was written mainly to demonstrate a particular mathematical/statistical approach to data modelling that could not be handled by conventional methods. In this case, the combination of radiocarbon dates where prior, archaeological and contextual, information was available.

The subsequent importance of this work cannot be overstated. The baton was taken up and ran with by numerate and archaeologically informed researchers at Nottingham University and elsewhere (see Baxter 2003:178 for more details). Buck *et al.* (1996), although wider-ranging than dating matters contains much on the subject that they and others developed on the basis of Naylor and Smith's paper. Things have obviously developed since 1996 (e.g., Buck and Millard 2003) but the seeds were sown in the 1980s.

In interpreting a single radiocarbon date the Bayesian approach takes account of measurement uncertainty and the wiggly nature of the radiocarbon calibration curve and results in a distribution of possible calendar dates which may be multi-modal. If several radiocarbon dates are available and their *relative* calendar dates are known (possibly partially) because of stratigraphic or artefactual evidence then the Bayesian approach allows such prior information to be incorporated into a statistical analysis in a manner not easily achieved by non-Bayesian methods. This allows for a much more nuanced interpretation of the radiocarbon dating evidence, potentially leading to fresh archaeological interpretations.

To illustrate, Whittle *et al*. (2008) discuss aspects of the interpretation of the Neolithic in Britain, using as a case study five long barrows in southern England dating to the fourth millennium BC. Their approach is motivated by the view that archaeologists have 'preferred a single scale of analysis, generally a long-term view of change and a large-scale perspective on society' and 'have neglected the interpretive importance of shorter time scales' (Whittle *et al* 2008:65). Using Bayesian methods it is possible to estimate calendar dates for the different monuments, including the period for which different phases were in use, expressed in probabilistic terms. Where the relative dating of different phases of construction is unknown from archaeological evidence the Bayesian also throws light on this.

The results need to be interpreted of course. It is, for example, possible to draw inferences about the extent to which different monuments and their phases were or were not in use at the same time, and the time-scale involved between differences in construction dates, leading to a temporal sequencing of the sites. To quote the authors '… instantly we move from a largely undifferentiated Early Neolithic, within which different sorts of construction and styles of activity float timelessly, to a period in which varying public architectures constitute historical sequences' (Whittle *et al*. 2008:68). The paper concludes by noting a variety of questions that the imposition of an 'historical' sequence on previously undifferentiated monuments raises, but also that it has not previously been possible to pose such questions in a manner permitting possible answers (e.g., why were some long barrows constructed at a particular times; why were some used for only relatively short periods of time; what antecedents motivated the form of construction of the long barrows).

MORE ON BAYES AND OTHER DEVELOPMENTS

As just stated, I think the advances made in radiocarbon calibration, and specifically the use of Bayesian methods, constitute one of the most important contributions that mathematics/statistics has made to archaeometry in the last 50 years or so.

Where will mathematics and statistics in archaeometry go in the next 50 years? I don't know, except that I expect increased computational power to be highly influential. I remember being at a *Computer Applications in Archaeology* conference in the 1990s where Clive Orton, in a keynote address, was addressing the more general question about applications in archaeology and made the point that if he knew the answer he would be writing grant applications rather than addressing the conference (to general mirth). Bayesian applications are as good a place to start from. I'm prepared (and even wish) to be totally wrong about what follows.

I find it difficult to convince myself that Bayesian methods, dating apart, have added much so far to archaeometric data analysis. I mention this because Bayesian methods are enjoying an increasing interest in statistics, and can undoubtedly do things, using prior knowledge, which classical methods cannot cope with.

Why do I think this? One reason is that (once-off problems apart and excluding dating) I find it difficult to think of archaeometric examples where Bayesian methods achieve what cannot be accomplished by simpler and more transparent methods. All the examples I am thinking of demonstrate that Bayesian methods work – in the sense that you can use complex methodology that no-one else is ever going to use – to reveal features of the data that can be recovered by far simpler methods.

This highlights a more general issue, already mentioned in the discussion of LIA. This is that a lot of the 'bread-and-butter' data that archaeometricians generate that is suitable for mathematical/statistical analysis is (from the purely statistical view) often quite simple. This is not a criticism. I mean by it that, in terms of what you are looking for, patterns in the data may be sufficiently simple (or non-existent), that older and tried-and-trusted methods may be more than adequate to find out what you need to know.

I have in mind a lot of applications of mathematics and statistics to provenance studies. Archaeologically it can be complicated (are you looking at the technology; the source of the raw materials; or the workshop where things were made and how things were put together, for example)? Statistical analysis of the compositional data is not necessarily straightforward, and can be done in various ways – but do useful ways of doing this differ much from what could have been done 30 years ago?

THE FUTURE

The future cannot easily be predicted. What I am capable of doing now statistically, and with the resources now available, could not have been foreseen 10 years ago.

This is a technological matter. Is archaeology/archaeometry different? As far as archaeometry goes, probably yes. Some researchers build their career on their access to a particular technique – Neutron Activation Analysis (NAA) comes to mind. Techniques improve or change. The technique is 'sold' in journal articles to establish the reputation of the author(s) and is claimed to be the 'best' available, in terms of the number and quality of the measurements available. It generates a lot of numbers which is where statistical analysis comes in.

NAA facilities have diminished in recent years (presumably because of expense) and other and cheaper technologies are now the rage (e.g., ICP methods). From a statistical point of view they generate numbers that look pretty much the same. You need to worry about the appropriateness, precision and accuracy of the variables and numbers used, but fundamentally they don't present new statistical or mathematical problems.

As far as I can tell the kind of archaeological problem that the data are designed to address has not usually changed.

So what about the mathematics and statistics? Do we need anything new? Newer methodology needs to be explored (Baxter 2006), but there should be no presumption that it will necessarily be more informative. The point of the exercise should be to address archaeologically interesting questions with appropriate methodology.

To reiterate, a lot of archaeometric data does not demand 'new' techniques to make sense of it. You need to be able to handle the techniques, of course, and not everyone does. You also ought to be asking appropriate questions of your data, and collecting and addressing it in an appropriate statistical manner – but that is a different question.

I like classification trees, partly because the presentation of the results is reasonably transparent. I've yet to see an application of neural networks to an archaeometric problem (some of which has been published in *Archaeometry*) that convinces me that the results could not have been achieved more simply. I've already said the same about non-dating Bayesian applications to archaeometric data, and the same may be true of other 'modern' statistical methods.

To return to provenance studies – a common ground for applications of mathematics and statistics in archaeology – and to simplify, suppose you have material from two sites, A and B, and p measurements on elemental/oxide concentrations and want to see if A and B differ.

Sometimes A and B are so obviously different that you don't need any complicated statistics to show this; ditto, if there is so much overlap that you cannot detect any difference at all. If there is an issue then the 'standard' techniques of multivariate analysis may help to make a decision (PCA, cluster analysis, discriminant analysis). What often underpins the analysis of such data is the assumption that the groups you are investigating have a multivariate normal distribution (possibly after logarithmic transformation). This can be questionable, but given the assumption (even if implicit) classical methods of analysis are often adequate. I could produce counter-examples to this sort of generalization, based on my own work (Baxter 2006), but as generalizations go I think it 'works'.

To be quite clear here, in the provenance example, if you are comparing material from several sites, it can be rather difficult to make sense of the statistical results. Some sites may differ and some may not. The potential difficulties are various; the questions posed that the data are supposed to answer may be inappropriate; the data may not be behaving in the way anticipated; the analyst may not really understand the (mathematical/statistical) techniques; there may simply be not much to say; or the answers may be obvious.

I am wondering here whether more modern/specialized methods of statistical analysis might often be needed (and it is an open question). In parenthesis, a lot of development in multivariate methods in statistics is taking place as a result of what is happening in areas such as bioinformatics. Abstracting a bit, and from a purely statistical point of view, you often have vast amounts of data with far more variables (elemental compositions if you wish) than cases (artefacts). This poses problems, which are being dealt with, but possibly do not apply to too much archaeometric data (and I am aware of exceptions).

CONCLUSION/SUMMARY

This has been an impressionistic survey of aspects of the role of mathematics and statistics in archaeometry over the last 50 years. For those who want it Baxter (2003), while not specifically about archaeometry, covers the technical detail, has a fairly comprehensive bibliography, and a lot of archaeometric applications. Some references cited in the present text draw attention to later work.

I started with the idea of concentrating on the generally important trends in the use of mathematics and statistics in archaeometry. Other than computational aspects, and developments in dating, I've found it difficult to convince myself that such trends exist. I could persuade myself that increasing sophistication in the use of Mahalanobis distance in provenance studies constitutes a trend of sorts, but the essence of the ideas involved is in Sayre (1975).

This need not be viewed as a criticism. There are papers I admire, which deal with once-off problems using interesting mathematics and statistics, but I have mainly commented on what I see as common rather than uncommon practice.

Not everyone whose work I like has been name-checked earlier, and I've refrained from citing papers that I think are unsatisfactory. The view that developments in the analysis of radiocarbon data, associated with an explicitly Bayesian approach, is one of the most important contributions that mathematics/statistics has made to archaeology is based on my perception that it has altered *archaeological* interpretations in a fundamental way – and I wish to stress *archaeological* here.

I have ambivalent views about what has gone on over the survey period. Notwithstanding earlier work, it is about the mid-1970s that things get interesting. In archaeology generally and not just archaeometry, a lot of the work published from this period into the 1980s is more interesting and innovative than much of what has appeared since. You need to explore methodologies, but they don't necessarily improve on what has been done before. Academics publish for career reasons; what they publish rarely has any lasting impact or importance and archaeometric publication is not immune from this effect.

To finish with an example to illustrate some of the problems of applying mathematics and statistics to archaeometric data, one of the more important statistical ideas to be exploited in archaeometric (provenance) studies is that of Mahalanobis distance. Given p variables and enough data (cases) it allows you to measure the 'distance' between, for example, two artefacts based on their chemical composition. This allows you to assess how similar they are and whether they could come from the same source. It is a very nice idea that has been around in the statistical literature for a long time (1930s at least); it has been exploited in archaeometric studies since at least the 1970s; it has been refined (I have in mind leave-one-out methodologies).

This is wonderful. The ideas frequently can't be used because you have too many variables, too many groups and not enough cases. In practice you may have to be pragmatic about the way you handle data (regardless of any mathematical model that might underpin a research hypothesis). The point here is that 'old' ideas are often fine; they can't, because of data limitations, always be exploited; you sometimes need to use scientific 'common-sense'; and 'new' statistical techniques may often suffer from similar limitations.

A common problem, I think, is that sometimes the quality/quantity of a set of data does not support the use of the mathematical/statistical techniques you might wish to use. As a hypothetical example, if you measure the properties of 10 artefacts with respect to 30 variables that you wish to use, forget about the use of Mahalanobis distance. Often pragmatism has to reign in doing data analysis.

Archaeometry is pointless unless it serves the purposes of archaeological research, and the connection is sometimes difficult to discern. Many commentators have noted that archaeology is difficult. Mathematics and statistics has, in the context of archaeology, a similarly subservient role to archaeometry.

REFERENCES

Aitchison, J., Barceló-Vidal, C. and Pawlowsky-Glahn, V., 2002, Some comments on compositional data analysis in archaeometry, in particular the fallacies in Tangri and Wright's dismissal of logratio analysis, *Archaeometry*, **44**, 295-304.

Baxter, M.J., 1994, *Exploratory multivariate analysis in archaeology*, Edinburgh University Press, Edinburgh.

Baxter, M.J, 1999, On the multivariate normality of data arising from lead isotope fields, *Journal of Archaeological Science*, 26, 117-124.

Baxter, M.J., 2001, Statistical modelling of artefact compositional data, *Archaeometry*, **43**, 131-147.

Baxter, M.J., 2003, *Statistics in archaeology*, Arnold, London.

Baxter, M.J., 2006, A review of supervised and unsupervised pattern recognition in archaeometry, *Archaeometry*, **48**, 671-694.

Baxter, M.J., and Gale, N.H., 1998, Testing for multivariate normality via univariate tests: a case study using lead-isotope ratio data, Journal of Applied Statistic, **25**, 671-683.

Baxter, M.J., and Freestone, I.C., Log-ratio compositional data analysis in archaeometry, *Archaeometry*, **48**, 511-531.

Beier, T., and Mommsen, H., 1994, Modified Mahalanobis filters for grouping pottery by chemical composition, *Archaeometry*, **36**, 287-306.

Bieber, A.M., Brooks, D.W., Harbottle, G. and Sayre E.V., 1976, Application of multivariate techniques to analytical data on Aegean ceramics, *Archaeometry*, **18**, 59-74.

Binford, L.R.B., 1964, A consideration of archaeological research design, *American Antiquity*, **29**, 429-451.

Binford, L.R.B. and Binford, S.R., 1966, A preliminary analysis of functional variability in the Mousterian of Levallois facies, *American Anthropologist*, **68**, 239-295.

Bronk-Ramsey, C., 1995, Radiocarbon calibration and analysis of stratigraphy: the OxCal program, *Radiocarbon*, **37**, 425-430.

Buck, C.E., and Millard, A.R. (eds.), 2004, *Tools for Constructing Chronologies: Crossing Disciplinary Boundaries,* Springer-Verlag, London.

Buck, C.E., Cavanagh, W.G., and Litton, C.D., 1996, *Bayesian Approach to Interpreting Archaeological Data*, Wiley, Chichester.

Buxeda i Garrigós, J., Kilikoglou, V., and Day, P.M., 2001, Chemical and mineralogical alteration of ceramics from a Late Bronze Age kiln at Kommos, Crete: the effect on the formation of a reference group, *Archaeometry*, **43**, 349-371.

Buxeda i Garrigós, J., Mommsen, H., and Tsolakidou, A., 2002, Alterations of Na, K and Rb concentrations in Mycenaean pottery and a proposed explanation using X-ray diffraction, *Archaeometry*, **44**, 187-198.

Buxeda i Garrigós, J., Cau Ontiveros, M.A., and Kilikoglou, V., 2003, Chemical variability in clays and pottery from a traditional cooking pot production village: testing assumptions in Pereruela, *Archaeometry*, **45**, 1-17.

Clarke, D.L., 1968, *Analytical Archaeology*, Methuen, London.

Doran, J.E. and Hodson, F.R.,1975, *Mathematics and Computers in Archaeology*, Edinburgh University Press, Edinburgh.

Dunnell, R., 2000, Archaeometry. In *Archaeological Method and Theory*, (ed. L. Ellis), Garland Publishing, New York, 47-52.

Freeman, P.R., 1976, A Bayesian analysis of the megalithic yard (with discussion), *Journal of the Royal Statistical Society A*, **139**, 20-55.

Glascock, M.D., 1992, Characterization of archaeological ceramics at MURR by neutron activation analysis and multivariate statistics, in *Chemical characterization of ceramic pastes in archaeology* (ed. H.Neff), 11-26, Prehistory Press, Madison, Wisconsin.

Hodson, F.R., Kendall, D.G., and Tautu, P., 1971, *Mathematics in the Archaeological and Historical Sciences*, Edinburgh University Press, Edinburgh.

Kendall, D.G., 1963, A statistical approach to Flinders Petrie's sequence dating, *Bulletin of the International Statistical Institute*, **40**, 657-680.

Kendall, D.G., 1974, Hunting quanta, *Philosophical Transactions of the Royal Society of London*, **A276**, 231-266.

Naylor, J.C., and Smith, A.F.M., 1988, An archaeological inference problem, *Journal of the American Statistical Association*, **83**, 588-595.

Neff, H., 2002, Quantitative techniques for analyzing ceramic compositional data, in *Source determination by INAA and complementary mineralogical investigations* (eds. D.M. Glowacki and H. Neff), 15-36, Monograph 44, The Corsten Institute of Archaeology at UCLA, Los Angeles.

Orton, C., 1999, Plus ça change? - 25 years of statistics in archaeology, in *Archaeology in the Age of the Internet: CAA97*, (eds. Dingwall, L., Exon, S., Gaffney, V., Laflin, S and van Leusen, M), Archaeopress, Oxford, 25-34.

Pollard, A.M., 1986, Multivariate methods of data analysis, in *Greek and Cypriot Pottery: A Review of Scientific Studies* (ed. R.E. Jones), 56-83, Fitch Lab., Occasional Paper, **1**, British School at Athens, Athens.

Pollard, A.M., and Heron, C., 1996, *Archaeological chemistry*, Royal Society of Chemistry, Cambridge.

Plooard, A.M., Batt, C,M., Stern, B., and Young, S.M.M., 2007, *Analytical chemistry in archaeology*, Cambridge University Press, Cambridge.

Ringrose, T.J., 1993, Bone counts and statistics - a critique, *Journal of Archaeological Science*, **20**, 121-157.

Robinson, W.R., 1951, A method for chronologically ordering archaeological deposits, *American Antiquity*, **16**, 293-301.

Sayre, E.V., 1975, Brookhaven procedures for statistical analyses of multivariate archaeometric data. Unpublished manuscript.

Scaife, B., 1998, *Lead isotope analysis in archaeology*, Unpublished PhD thesis, University of Bradford, UK.

Scaife, B., Budd, P., McDonnell, J.G., and Pollard, A.M., 1999, Lead isotope analysis, oxhide ingots and the presentation of scientific data in archaeology, in *Metals in antiquity* (eds. S.M.M.Young, A.M. Pollard, P. Budd. and R.A. Ixer, R.A.), 122-133, BAR International Series 792, Archaeopress Oxford.

Shanks, M., and Tilley, C., 1992, *Re-constructing Archaeology: Theory and Practice: Second Edition*, Routledge, London.

Shennan, S., 1988, *Quantifying archaeology*, Edinburgh University Press, Edinburgh.

Shennan, S., 1997, *Quantifying archaeology: 2nd edition*, Edinburgh University Press, Edinburgh.

Whittle, A., Bayliss, A., and Healey, F., 2008, The timing and tempo of change: examples from the fourth millennium cal. BC in Southern England, *Cambridge Archaeological Journal*, **18**, 65-70.

Spaulding, A.C., 1953, Statistical techniques for the discovery of artefact types, *American Antiquity*, **18**, 305-313.