Archaeometry 37, 1 (1995), 1–24. Printed in Great Britain

REVIEW ARTICLE
THE BAYESIAN APPROACH TO THE INTERPRETATION OF ARCHAEOLOGICAL DATA*

C. D. LITTON and C. E. BUCK

Department of Mathematics, University of Nottingham, Nottingham, NG7 2RD, U.K.

Archaeologists are increasingly becoming aware of an approach to data investigation known as Bayesian statistics. In this paper we outline both the philosophical and statistical background to the approach. We show that it provides a logical and coherent framework in which to make inferences on the basis of both data and a priori expert knowledge. We note that adoption of the Bayesian framework is particularly timely since there have been recent dramatic developments in numerical methods which mean that a number of previous implementation problems have now been solved. As a result, many questions of archaeological interest, which require the use of complex statistical models, are being investigated using this methodology. We use a variety of recently published examples from a range of archaeological areas to illustrate the type of questions that can be answered and the nature of the methodologies used, and we make comparisons with the results obtained using more traditional statistical techniques.

KEYWORDS: BAYESIAN STATISTICS, MARKOV CHAIN MONTE CARLO, RADIOCARBON DATING, PROVENANCE STUDIES, FIELD SURVEY, CORBELLED TOMBS, MEGALITHIC YARD, PALAEOETHNOBOTANY

INTRODUCTION

All modern archaeological investigations include the interpretation of data, ranging from counts of objects to chemical compositional data and complex interrelationships between groups of objects. In order to aid them in their interpretation, archaeologists and other specialists will often use statistical techniques. These techniques are employed to describe data by displaying plots such as pie charts or histograms and computing summary statistics such as means, standard deviations and the like. The last twenty years has seen the development of a wide range of computer packages and the publication of good basic books which mean that descriptive statistical techniques are now widely and, generally speaking, well used in archaeology and related disciplines.

However, in order to answer specific questions, like 'how many different provenances contributed to the pottery assemblage found on this site?', more thought and more sophisticated statistical tools are required. This, almost of necessity, involves the use of a 'model' which, although not always explicitly stated, can be expressed in a statistical form. In the introductions to the well-known and often referenced books by Doran and Hodson (1975) and Orton (1980), we find that this very point is emphasized and the advantages of undertaking such modelling are highlighted. In fact, Orton (1980) selects a layout to his

* Received 24 January 1994, accepted 20 April 1994.

book which means that we can readily identify the types of questions to which archaeologists most commonly seek answers; these are:

What is it?	How old is it?
Where does it come from?	What was it for?

He discusses the contemporary statistical and mathematical methods for answering these questions and clearly indicates that greater interpretability can be obtained when model-based approaches are adopted rather than simple descriptive ones. In his final chapter, however, Orton (1980) includes a section entitled 'The Future' in which he identifies mathematical and statistical tools as yet not adopted by the archaeological community but which he feels might prove particularly useful. Amongst other techniques, he identifies the Bayesian statistical approach as having much to offer. This theme is taken up again by Ruggles (1986) (who links the need for more model-based statistical work to the increased use of information technology in archaeology) and is also stressed by other authors such as Doran and Hodson (1975, 308) and Cowgill (1989, 79).

More recently, in his *Distinguished lecture in archaeology: beyond criticizing New Archaeology*, Cowgill (1993) provides a wide-ranging review of changes in philosophical approaches to archaeological data interpretation. He highlights a number of drawbacks to some of the previously adopted approaches and states that

> I will not pursue philosophical issues further here, except to urge that one very important aid to improved thinking is to learn more about the Bayesian approach to statistical inference. The Classical approach goes to great lengths in the attempt to extirpate all subjective elements, but it only succeeds in sweeping them under the rug (Urbach 1992). It also succeeds in baffling most students in statistics courses, and often leads them to reason incoherently when they try to apply formal statistics (Cowgill 1977). In contrast, in the Bayesian approach, one's prior knowledge and beliefs can often be built explicitly into the equations and can thus be made overt rather than covert. One can formalize the interaction between prior beliefs and new data. The approach is not without technical problems and it by no means resolves all the important philosophical issues, but it is a great improvement over Classical statistical models for reasoning.

What stronger commendation could one hope for?

The major drawback noted by Cowgill (1993) was that technical problems exist. Recently, however, dramatic advances have been made in numerical techniques based on Markov chain Monte Carlo (MCMC) methods. In simple terms MCMC methods provide a means of carrying out the calculations necessary for Bayesian analysis that up to now have been difficult or impossible. See Smith and Roberts (1993) for a review of the theory of the methods. A more practical guide to their use is provided by Gilks *et al.* (forthcoming) who give several case studies. As a consequence, applications of Bayesian methods are now beginning to be more widespread in a variety of subjects such as medicine and economics. In the light of these developments, our intention in this paper is to present the Bayesian approach to data analysis and describe its past, present and potential future applications to archaeological problems.

Some readers may be unaware of the nature of the considerable philosophical differences between the Bayesian approach to statistical inference and the more traditional (or classical) one. As a result, in the next section, we will describe these differences and highlight the practical and theoretical advantages of adopting the Bayesian approach.

A PHILOSOPHICAL AND THEORETICAL PERSPECTIVE

When faced with a new problem, archaeologists will (by their training) seek to make use of any available a priori knowledge that may help them. They will use their own experience, the experiences of their colleagues and the written reports of other archaeologists. Then, when their own work is under way (be that survey, excavation, laboratory work and so forth), they will collect new data and interpret those in the light of the a priori evidence.

For example, initial interpretation of the Bronze Age site at Flag Fen, Cambridgeshire, was that it represented a habitation site built on a platform (Pryor *et al.* 1986). Further excavation over a number of years led to the archaeologists discovering additional information which suggested that in fact the posts (originally interpreted as building supports) are more likely to have been part of a palisade. As a consequence, the site of the initial excavation is now acknowledged to be just a small part of a much larger area of archaeological activity. Thus the interpretation was updated in the light of the new evidence (Pryor 1992).

Bede House Farm, North Luffenham, Leicestershire, provides another example. This building was first visited by an architectural historian whose remit was to make an assessment with a view to adding it to the national list of buildings of special interest. On the basis of external evidence alone, the structure was identified as being of late seventeenth- or early eighteenth-century construction. Later, internal inspection was possible which revealed a medieval open hall with smoke-blackened rafters, and a cruck cross-wing which clearly indicates a much earlier period of construction. Subsequently, investigation of timbers by dendrochronological dating established that the open hall was constructed soon after 1443 whereas the cross-wing was somewhat earlier at about 1400 (Litton forthcoming).

These two schematic examples are for the purpose of illustration only, and we are sure that all readers will know of other examples of updating interpretations from their own field of specialization. Basically, the Bayesian approach to statistics allows for such updating, and provides a formal framework in which to do it. Consequently, by adopting the Bayesian approach, we explicitly express our prior beliefs, collect our data and then combine these in a logical and coherent way to obtain our posterior beliefs. That is, the Bayesian framework provides a simple, but formal way of dealing with life's ubiquitous problem of learning from experience.

Statisticians, both Bayesian and classical, formalize problems by developing statistical models that represent, in a mathematical manner, the problem at hand. Within these models, the relationship between data (denoted by $\boldsymbol{x}$) and unknowns called parameters (denoted by $\boldsymbol{\theta}$) are posited. Bayesian statisticians use probabilities to reflect the uncertainty about the parameters, both a priori and a posteriori. The Bayesian framework offers a simple way of computing the posterior from the prior and the data via the model. Both the prior and the posterior are expressed as probabilities. That is, subject matter experts are asked to make probabilistic statements relating to their opinion about the values of $\boldsymbol{\theta}$ before seeing the data. Bayesian analysis is then used, in the light of both the prior and the data, to give posterior probabilities of the same values of $\boldsymbol{\theta}$.

Briefly, the Bayesian method requires the user to be able to specify the following components.

The likelihood—a function, denoted by statisticians as $p(\boldsymbol{x}|\boldsymbol{\theta})$. This is read as 'how likely are the values of the data we observed, given some specific values of the unknown parameters?'.

The prior—also a function, denoted by statisticians as $p(\boldsymbol{\theta})$. This is read as 'how much belief do I attach to possible values of the unknown parameters before (i.e., prior to) observing the data?'.

The posterior—is what we wish to obtain from the Bayesian investigation and is represented by statisticians as $p(\boldsymbol{\theta}|\boldsymbol{x})$. This is read as 'how much belief do I attach to possible values of the unknown parameters after (i.e., posterior to) observing the data?'.

We clarify ideas by the use of a simple archaeological example. Suppose we are interested in dating an organic object using radiocarbon. Let the unknown date be represented by $\boldsymbol{\theta}$. Prior to sending the sample for radiocarbon analysis, the archaeologist has beliefs about the unknown date, $\boldsymbol{\theta}$, which can be expressed (perhaps with the help of a statistician) in terms of a prior probability denoted by $p(\boldsymbol{\theta})$. The radiocarbon laboratory returns an estimate of the radiocarbon age of the object, this is the date which we denote by $\boldsymbol{x}$. The likelihood is how $\boldsymbol{x}$, the estimated radiocarbon age, is related to the true, but unknown, calendar age. For each value of $\boldsymbol{\theta}$ it measures how likely that value of $\boldsymbol{x}$ is. After receiving the laboratory results, the posterior beliefs about $\boldsymbol{\theta}$ are expressed in terms of the posterior probability, $p(\boldsymbol{\theta}|\boldsymbol{x})$. A large value of $p(\boldsymbol{\theta}|\boldsymbol{x})$ indicates that the corresponding date is, a posteriori, likely and a small value that it is unlikely.

With the likelihood, the prior and the posterior defined as above, the key to all Bayesian statistical methods is Bayes's theorem which takes the form

$$p(\boldsymbol{\theta}|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|\boldsymbol{\theta})}{p(\boldsymbol{x})} \times p(\boldsymbol{\theta}) \tag{1}$$

where

$$p(\boldsymbol{x}) = \text{Average of } \{p(\boldsymbol{x}|\boldsymbol{\theta}) \times p(\boldsymbol{\theta})\}.$$

Commonly, the first term of the right-hand side of this equation (the likelihood divided by $p(\boldsymbol{x})$) is called the *standardized likelihood* and we can then rewrite the theorem in words as

$$\text{Posterior belief} = \text{Standardized likelihood} \times \text{Prior belief.}$$

Formulated in this way, we see that values of the unknown parameters which give rise to large values of the standardized likelihood will lead to higher posterior beliefs than values of the unknown parameters which give rise to small values of the standardized likelihood. And what does the standardized likelihood measure? Well, high values correspond precisely to parameter values which are 'likely' to have led to the data we actually observed. So, we increase our beliefs in parameter values which are highly compatible with the data we have seen; we correspondingly decrease our beliefs in parameter values which are unlikely to have given rise to the data actually observed. We note that equation (1) can be expressed as

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}) \tag{2}$$

where $\propto$ means 'proportional to'. Thus the posterior can be easily obtained just by multiplying the likelihood by the prior and then the posterior is normalized so that the probabilities add up to one.

What happens next? That is, if and when further related data are obtained? Again, it's very simple. Basically, today's posterior becomes tomorrow's prior. In other words, our new 'current' beliefs about $\boldsymbol{\theta}$ are $p(\boldsymbol{\theta}|\text{previous data})$; these will be revised via Bayes's theorem to:

$$p(\boldsymbol{\theta}|\text{previous and new data}):$$

and so it goes on. Bayes's theorem is a perfectly natural tool for successive revision of beliefs as each piece of new data comes in. We simply 'learn from experience' so that

$$p(\boldsymbol{\theta}|\text{previous and new data}) \propto p(\text{new data}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\text{previous data}).$$

AN HISTORICAL PERSPECTIVE

We commence this section by stressing that Bayesian ideas are not new. Indeed, they can be traced back to the Revd Thomas Bayes FRS who was born about 1702 and died in 1761. His most famous work in which he formulated the concept of 'inverse' probability (i.e., posterior probability), was published posthumously (Bayes 1763). Similar ideas, but without the detailed mathematical foundation, are to be found in the earlier work of Bernoulli (1713).

We should stress too, however, that it is not until very recently that Bayesian ideas can be seen to be well established in statistical practice. During the middle of this century, traditional (or classical) approaches were almost uncontested within the statistical community and it was not until about 1960 that a steady revival in the Bayesian approach occurred. Development of the Bayesian statistical framework can be followed in the works of Jeffreys (1961), Lindley (1965), DeGroot (1970) and Box and Tiao (1973) with the current state of Bayesian theory to be found in Bernardo and Smith (1994) and O'Hagan (1994). A review of the development of computational aspects of Bayesian statistics will be given in Gelfand and Smith (forthcoming) while Gilks *et al.* (forthcoming) will concentrate on recent advances in the use of MCMC techniques. O'Hagan (1988) and Smith (1988) both provide useful introductions to the Bayesian view of probability, de Finetti (1974–5) offers stimulating background reading and Howson and Urbach (1989) offer a Bayesian perspective on scientific reasoning. A less mathematical exposition can be found in Lindley (1985). Barnett (1982) and Stuart and Ord (1991, chap. 31) give comparisons of the conventional or classical approach to statistics with the Bayesian one.

We do not intend to develop here all the mathematics underlying the Bayesian approach: for this, the reader is referred to the above works. For discussions of the general applicability of a Bayesian approach to the analysis of archaeological data the reader is referred to Doran and Hodson (1975, chap. 3), Orton (1980, 220) and the other works cited above which all clearly indicate that a Bayesian approach has much to offer archaeologists.

A PRACTICAL PERSPECTIVE

On the face of it, Bayesian statistics and archaeological thinking have much in common. Why is it then that Bayesian methods have not been applied to archaeology more extensively? Firstly, archaeologists are largely unaware of the existence of Bayesian methods and those that are aware are hesitant about using a novel approach which has not been widely addressed in standard statistical texts. Secondly, few Bayesian statistical computer packages exist and those that do are designed for users with sophisticated understanding of statistical methods.

So, to carry out a full Bayesian analysis, one needs to (i) develop a model, (ii) specify a prior, (iii) evaluate the posterior and (iv) interpret the results. These constituent parts are important both in their own right and for their relationship to the others. Consequently, we will look briefly at each in turn.

(i) Modelling Many of the data analysis problems posed by archaeologists are non-standard to the statistical community. As a result, close collaboration is often required between a number of specialists, including a statistician, in order that the situation be realistically modelled. Off-the-peg statistical solutions are often inadequate for modelling archaeological problems and, as a result, a model needs to be specifically developed for the application in mind (see in particular Fieller 1993).
(ii) Specifying the prior One of the great stumbling blocks to more widespread use of the Bayesian approach is the difficulty some people have in specifying prior information, or even acknowledging that such information exists at all. The latter is largely a philosophical point. Regarding the former, as with model building, prior elicitation requires clear communication between experts from different fields and such interaction can realistically only be developed as part of close collaborative projects.
(iii) Evaluating the posterior As mentioned above, the posterior is obtained by evaluating (standardized likelihood$\times$prior). This deceptively simple relationship, however, disguises the true mathematical complexities involved. In the remainder of this section, we consider methods for summarizing the posterior information for each individual parameter that we need for Bayesian statistical analysis and provide a detailed outline for the use of one of these. Apart from a few, relatively simple, special cases, evaluating posteriors requires the use of advanced techniques of numerical analysis in order to carry out the necessary multidimensional integration involved. Such methods are mathematically complex and even when incorporated into computer packages (e.g., BAYES4 by Naylor and Shaw 1985) they can really only be used by statistical experts (Smith 1991). This problem is now being overcome, however, with advances in numerical techniques, based upon the Gibbs sampler and related MCMC methods (reviewed by Smith and Roberts 1993), which make the calculations relatively easy to implement if on occasions still rather time-consuming.
(iv) Interpretation At the end of the statistical analysis, the posterior represents our current beliefs. It may, however, take a complex form which does not provide us with a suitable vehicle for communicating the information it contains to others. Summarization of the major features of the posterior may be required and reported in terms of means, variances, modes and highest posterior density regions. Furthermore, our confidence in the posterior interpretation will depend upon its sensitivity to changes in the data or prior. As a result, some form of sensitivity or robustness exercise will be needed.

As mentioned in (iii) above, evaluation of the posterior probabilities for individual parameters has, until recently, been difficult if not impossible in many commonly occurring situations. The advent of MCMC methods, however, has meant that previously intractable problems can not be realistically addressed. Of the MCMC methods, one particular approach (Gibbs sampling) has proved particularly useful for archaeological applications. The Gibbs sampling algorithm is based on stochastic simulation. Its main features are given in Smith and Roberts (1993) while Buck *et al.* (1992b) give details of its application to problems in the interpretation of radiocarbon determinations.

A CLASSICAL PERSPECTIVE

It is necessary to appreciate that there are many statisticians who feel unable to adopt the Bayesian approach. There are a multitude of reasons for this, but often expressed is a dissatisfaction with having to specify prior probabilities; thus the Bayesian approach is judged not 'objective' and, in the eyes of some, therefore 'unscientific'. Of course, it is difficult to quantify one's beliefs prior to the analysis of data; this is not disputed, nor is it denied that personal judgement is part of the process of defining a prior. However, classical methods of analysis cannot usually escape an element of judgement either, but the judgement enters at the stage of interpreting the results, after processing the data. There can be a danger in interpreting the results in this manner, of *ad hoc* argument and special pleading. Such dangers are greatly increased when analysis is carried out by the inexperienced using statistical packages and off-the-peg methods. From a Bayesian stance, we would argue that it is more rigorous to state any judgements concerning the nature of the data and model explicitly from the start. For instance, in undertaking statistical analysis of chemical compositions with a view to determining provenance, it has commonly been the practice to carry out the analysis using standard algorithms but with no prior definition of what is meant by a cluster. In such a case a decision is still made, but usually after the statistical results have emerged from the computer. Moreover the decision can be open to question; namely, where one analyst might judge there to be three clusters another might recognize four. Clearly we need to distinguish between 'exploratory' methods of statistical analysis, and model-based techniques which can realistically be used to make inferences. In his recent book, Baxter (1994) highlights the continuing role of exploratory multivariate data analysis in archaeology. He comments that model-based methods are likely to attract increasing interest in the future, but he expects 'exploratory and accessible methods . . . to remain the bread-and-butter of multivariate applications'. Nonetheless, when cluster identification is to be linked to a particular type of archaeological interpretation, we would suggest that it is better and more scientific to define what is meant by a cluster before the statistical analysis is carried out and that a statistical algorithm be designed which exactly suits this need. Otherwise we are in danger of finding patterns in tea-leaves or similar random data sets!

The classical approach to statistics disregards any prior information, whether it is relevant or not, and regards as the only quantifiable source of information that provided by the sample data. It adopts a frequency view of probability which assumes an infinite series of essentially identical repeated circumstances. Thus, when reporting a 95% confidence interval, the correct interpretation is that 95% of intervals constructed in that manner will include the true value of the unknown parameter. However, it tells us nothing about the actual situation under consideration, only what will happen in the long term.

This long-term approach may be applicable to some situations such as the monitoring of the diameters of ball bearings made by the thousands in a factory, but does it apply to archaeology? Usually, archaeologists get one, and only one, chance to excavate a site and take samples for radiocarbon dating. How does this fit into the classical framework? The point is that most archaeological 'experiments' are unrepeatable. In contrast, Bayesians produce 95% posterior credible intervals which clearly express the posterior opinion that the unknown parameter is included in that interval *with probability 0.95* (or in other words that the unknown parameter lies in the interval with odds of 19 to 1).

Another point is that, in hypothesis testing, classical statisticians report the probability of observing the data assuming that the null hypothesis is true, whereas what is actually required is the probability that the null hypothesis is true *given* the observed data. In other words, if we let the value of a parameter under the null hypothesis be θ_0, then classical statistical hypothesis testing involves calculation of $p(x|\theta = \theta_0)$. This clearly tells us how likely the data are if the value of θ were θ_0. However, we really need $p(\theta = \theta_0|x)$, the probability of θ being θ_0 given that we have observed the data. Of course these two are easily linked via Bayes's theorem since

$$p(\theta = \theta_0|x) \propto p(x|\theta = \theta_0)p(\theta_0)$$

where $p(\theta_0)$ is the prior probability that $\theta = \theta_0$.

All these points add up to fundamental differences between the classical and Bayesian perspectives. The debate has been fiercely contested over the last two decades and continues, if somewhat less vehemently, to the present day. For a balanced and wide-ranging discussion of the points raised here and other more subtle ones relating to implementation, we refer the reader to Barnett (1982) and Stuart and Ord (1991, chap. 31).

A REVIEW OF BAYESIAN STATISTICS IN ARCHAEOLOGY

In this section, we will review many of the published applications of the Bayesian methodology to the analysis of archaeological data. We will do so by giving first a brief description of each archaeological problem and then an outline of the mathematical and/or statistical model used. The nature of the prior information will be explained and the archaeological interpretation highlighted. We will not attempt to give the mathematical details of how the results were obtained, but simply refer the interested reader to the original papers in which the work was reported.

In each of the articles discussed below, the models developed were based upon individual interpretation of the archaeological problem. Obviously, these models are not definitive and other workers, with different insights, may well wish to develop alternatives. This is just part of good scientific practice and is not unique to Bayesian investigation. What is unique is the ability to incorporate into the analysis, in a coherent fashion, prior information about the parameters of the model. The nature of the prior information and the way in which it is converted into a statistical form may vary depending upon the particular experts involved. However, in all cases, this prior information must be explicitly stated so that others can clearly see what has and has not been included.

Investigating the megalithic yard

The first published application of Bayesian statistics to archaeological data was Freeman (1976). This work concerns the analysis of measurements of the radii of prehistoric stone circles with a view to identifying the length of the 'megalithic yard'. Much previous work (Thom 1955, 1962, 1964 and 1967; Broadbent 1956; Kendall 1974; and Porteous 1973 and 1977) had concentrated on attempts to obtain a point estimate for the length of the 'megalithic yard'. In contrast, Freeman (1976) saw the need to produce a posterior distribution for the quantum, assuming it exists.

The mathematical formulation of the problem chosen by Freeman (1976) was as follows. Let the unknown quantum be denoted by θ and suppose that the data set comprises measurements from n stone circles. For the ith stone circle, let X_i be its radius and m_i be a positive but unknown integer constant which reflects the size of the circle. Then the model is

$$X_i = m_i\theta + e_i \qquad i = 1, \ldots, n$$

where e_i represents an error component which is assumed to have a normal distribution with mean zero and variance σ^2. Thus, the likelihood function for the observations $x_1, \ldots, x_n$ is given by

$$p(x_1, \ldots, x_n | m_1, \ldots, m_n, \theta, \sigma) = \left(\frac{1}{\sqrt{2\pi}\sigma}\right)^n \exp\left\{-\sum_{i=1}^{n}(x_i - m_i\theta)^2/2\sigma^2\right\}.$$

The prior information was modelled by Freeman (1976) as follows. Since θ was believed to lie between two and 11 feet, but with no value of θ more likely than any other, the prior for θ was taken as having a uniform distribution on the range two to 11. To specify the prior for m_i given the quantum, θ, Freeman postulated that there existed a maximum circle radius, denoted by k (where k is an integer multiple of the quantum size θ). In addition, it was assumed that the builders would choose to construct a circle of radius $m_i\theta$ with probability θ/k (for $1 \leqslant m_i \leqslant k/\theta$). Because of the difficulty of specifying a prior for σ, and the computational difficulties involved even if that were possible, the posterior density for θ, given by

$$p(\theta | x_1, \ldots, x_n) \propto \theta^n \prod_{i=1}^{n} \sum_{m_i=1}^{k/\theta} \exp\{-(x_i - m_i\theta)^2/2\sigma^2\},$$

was only evaluated for a number of fixed values of σ.

When investigating Thom's data (1955, 1962, 1964 and 1967) for stone circles from all over the British Isles, Freeman (1976) was unable to identify a single quantum used in the construction of all the monuments. However, in some subsets of the data (in particular, western Scotland) he felt more confident about identifying a particular value for the 'megalithic yard'.

Palaeoethnobotany

Kadane and Hastorf (1988) discuss the investigation of palaeoethnobotanical remains with a view to identifying the usage of the activity areas from which they were excavated. For the purposes of their study, it was assumed that there were m different types of palaeoethnobotanical remains and n different types of activity. Let Y_{ij} represent the number of palaeoethnobotanical remains of type i at location of activity type j (at abandonment of the site). Further, Y_{ij} was modelled as a Poisson random variable with parameter λ_{ij}. In addition, it was assumed that a particular palaeoethnobotanical object (of type i) survives from the abandonment to the date of excavation with probability p_i which is independent of the survival of other objects and the type of activity area from which it was recovered. Let X_i represent the number of palaeoethnobotanical remains of type i found during the excavation. Conditional on the area having been used for activity type j, we observe data values $x_1, \ldots, x_n$ with probability

$$\prod_{i=1}^{m} \frac{(\lambda_{ij}p_i)^{x_i}}{x_i!} e^{-\lambda_{ij}p_i}.$$

Denoting the probability that a particular area was used for activity type j by θ_j, the likelihood for $x_1, \ldots, x_n$ is given by

$$p(x_1, \ldots, x_n | \lambda_{ij}, \theta_j, i = 1, \ldots, m, j = 1, \ldots, n) = \sum_{j=1}^{n} \left\{\prod_{i=1}^{m} \lambda_{ij}^{x_i} p_i^{x_i} e^{-\lambda_{ij}p_i}\right\} \theta_j.$$

Strictly, priors for λ_{ij}, p_i and θ_j need to be specified. However, recall that this work was undertaken before the advent of MCMC methods and so, for ease of computation, Kadane and Hastorf (1988) used point estimates for λ_{ij} and p_i. $\theta_1, \ldots, \theta_n$ were assumed to have a Dirichlet distribution with parameters estimated by reference to the notebooks kept by the archaeologists during excavation. On the basis of this information, a posterior probability was calculated for the usage of each activity area on the site. As a consequence, Kadane and Hastorf (1988) were able to provide clear statements about the probability of usage of each part of the site from which palaeoethnobotanical remains were recovered. It was then left to the archaeologists, on the basis of these probabilities, to make the final interpretations as to the usage of each activity area.

The works of both Freeman (1976) and Kadane and Hastorf (1988) clearly demonstrate the constraints placed upon Bayesian statisticians by the computational methods available to them. In particular we note that, although the models were carefully formulated to reflect the natural variability in archaeological data, both papers report the use of (at the time) necessary but undesirable estimates for some of the model parameters.

Field survey

Cavanagh *et al.* (1988b) developed a Bayesian method for identifying changes in soil phosphate level recorded as part of field survey investigations. Commonly, the natural underlying soil phosphate content is enhanced when organic matter decomposes. Thus on archaeological sites, where occupation was in some sense permanent, we usually find that the habitation and midden areas are associated with a soil phosphate concentration that is appreciably higher than the underlying background level. This enhancement persists through time and can be detected using a fairly quick, simple chemical test that can be used on site.

The survey work reported in Cavanagh *et al.* (1988a) was undertaken in the Laconia region of Greece. One of the aims of this fieldwork was to establish the limit of occupation, as defined by enhanced soil phosphate concentration, on previously identified habitation sites. Since similar limits for each site had already been defined by pottery and tile distributions, the samples for phosphate analysis were collected at two metre intervals along transects from an estimated site centre. Four transects were taken, usually the four cardinal points.

It is important to realize that the archaeologists were not attempting to locate individual phosphate features, but simply to locate, in very broad terms, site boundaries. Therefore, the problem posed to the statistician was how to use the numerical data relating to soil phosphate concentration from a transect to identify two zones, one of 'on-site' concentration and one of 'off-site' concentration. A fairly simple statistical model of the situation is as follows. In the 'on-site' zone we might expect a mean phosphate concentration of μ_1 with a variability (due to inherent noise) of σ_1 and in the 'off-site' zone we might expect a mean of μ_2 with a variability of σ_2.

In this situation, our prior information consists of knowledge about the likely values of μ_1, σ_1, μ_2 and σ_2. Such information might be provided by the soil phosphate analyst, the archaeologist or possibly from data from other sites. Sometimes we will have very good prior information, for example, if there has previously been similar work undertaken in the same area, whereas, if the area has not previously been investigated in this way, we will have only vague prior information.

In the light of the success of the above methodology, the archaeologists undertook further seasons of field survey in Laconia and invested greater effort in obtaining soil samples for phosphate analysis. They extended their sampling scheme to one which involved collection of soil samples over a grid rather than just simple transects. With data in this form, the statistical methodology was extended to effectively become a Bayesian approach to image processing (Buck *et al.* 1988; Cavanagh *et al.* 1988a). Consequently, the statisticians presented the archaeologists with two-tone images of the 'on-site' and 'off-site' phosphate regions on their sites plotted on the survey grid. These images could then readily be compared with similar plots of data from other sources, such as surface counts of pottery and tile, to arrive at an integrated interpretation of a range of archaeological information.

In further work, the same authors worked with a number of different archaeologists whose data required a multilevel approach to Bayesian image processing and in so doing adopted the methodology of Besag (1986). Buck *et al.* (1992a) consider data from a site (Manor Farm, Borwick, North Lancashire) at which the phosphate data were modelled as multilevel, allowing the identification of possible burials that had previously been obscured by the very noisy nature of the data. In other work (Buck *et al.* 1990) a similar methodology was applied to the analysis of resistivity data from Etton, Cambridgeshire, at which they were able to identify two ditches, one of probable Neolithic date and one thought to be Iron Age or Romano-British. In the same paper, the authors also describe the investigation of magnetic susceptibility data from Barnhouse, Orkney, where it was established that increased magnetic susceptibility, associated with Neolithic occupation, covered a considerable area outside the previously excavated habitation site. In addition, the Bayesian analysis highlighted a linear feature not previously identified in magnetic susceptibility (or other) data, but whose presence was later confirmed using resistivity survey.

Structural analysis

The work of Cavanagh and Laxton (1981) relates to investigation of the stability of corbelled vaults (or domes) of the type associated with the architecture of a number of prehistoric cultures worldwide. They derived a model (based upon structural mechanics)

$$r(x) = cx^d$$

which relates the depth below the apex of the dome, x, to the radius $r(x)$. Here c and d are constants which may vary from dome to dome. The first constant, c, reflects the *size* of the dome while the second, d, is of special archaeological interest, as it determines the *shape* of the dome.

Using data from a variety of Mediterranean corbelled domes, Cavanagh and Laxton (1982 and 1987) demonstrated that this model more than adequately described the shape of the majority of structures sampled. That is, the above model was a good one. However, for a small minority of structures, the model was obviously not adequate. By using the model given above, the curvature is assumed to be 'smooth' and indeed, in a simple case, the builders could have built a stable vault from floor to apex in this 'smooth' form. Alternatively, they could have built it in two stages with a different curvature for the uppermost part than for the lower part. Mathematically this is expressed in terms of a new refined model as

$$r(x) = \begin{cases} c_1 x^{d_1} & 0 < x < \gamma \\ c_2 x^{d_2} & \gamma < x \end{cases}$$

with a change in exponents at depth γ such that $d_1 > d_2 > 0$.

This, however, leads to a serious statistical problem of how to estimate the depth, γ, in vaults where the change occurs. Moreover, this is compounded by the fact that the location of the apex is only imprecisely known. Consequently, Cavanagh *et al.* (1985) suggest that we let a be the distance of the true apex above the point from which the x measurements were made, then the model becomes

$$r(x) = \begin{cases} c_1 (x+a)^{d_1} & 0 < x < \gamma \\ c_2 (x+a)^{d_2} & \gamma < x \end{cases}$$

Adapting this formulation, Buck *et al.* (1993) showed how the model can be analysed from a Bayesian viewpoint using Gibbs sampling. Including prior information about a, posteriors of a, d_1, d_2, γ and the other parameters of interest were calculated. The methodology was illustrated on data from the Minoan tholos at Stylos, Greece. In a later paper. Laxton *et al.* (1994) apply the same methodology to Sardinian nuraghi which had previously been analysed using classical statistical techniques which did not permit inclusion of any uncertainty about a. For three of the nuraghi, the results from the Bayesian and classical approaches were in broad agreement. In the case of the other two structures, however, the results were significantly different. In Figure 1, we show plots of (a) the prior and (b) the posterior distribution for the change-point, γ, from one of these nuraghi, at Santa Barbara, Sardinia. In (b), the broken line marks the maximum likelihood estimate of γ given by Cavanagh *et al.* (1985). Clearly, the inferences based upon the Bayesian posterior and the classical maximum likelihood will be different. Since the Bayesian analysis included the uncertainty about a explicitly in the analysis, our confidence in the Bayesian result is greater than that in the classical one. In terms of archaeological interpretation, it is now clear that the change-point in these structures does not always occur at the height of the door lintel as had previously been conjectured.

Radiocarbon dating

It is in the calibration of radiocarbon determinations that Bayesian statistics has made the

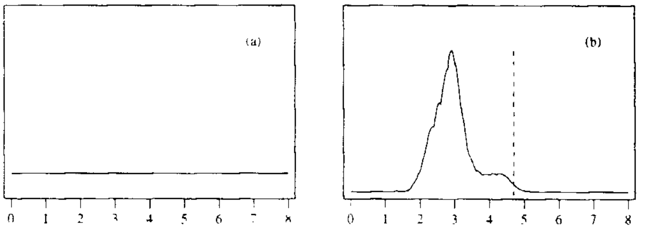

Figure 1 (a) *Prior and* (b) *posterior distributions for the location of the change-point in the Santa Barbara nuraghi, Sardinia. The broken line on* (b) *shows the maximum likelihood estimate of the change-point location.*

greatest impact on archaeological data analysis and interpretation. Naylor and Smith (1988) reported the first fully Bayesian investigation of archaeological data in which they demonstrated a new approach to the calibration of radiocarbon determinations from Danebury Iron Age hillfort, Hampshire. This initial work has led to a gradual development of a wide range of supplementary Bayesian techniques for interpreting radiocarbon determinations (see later), but from the initial work of Naylor and Smith (1988) through to the more recent work of Litton and Leese (1991) and Buck *et al.* (1991, 1992b and 1994) the basic model is essentially the same. In this section, we will first consider this model in some detail and then examine how to incorporate any archaeological prior knowledge that is available into the analysis.

Consider one particular archaeological event of interest and suppose that we have obtained organic material whose date of death coincides with that event. We denote this calendar date by θ. Now, associated with this calendar date there is a unique 'radiocarbon age' which relates to the amount of ^{14}C actually present in the sample. Statisticians have chosen to represent this 'radiocarbon age' by $\mu(\theta)$. However, due to the nature of the archaeological samples and the physical techniques used to measure the amount of ^{14}C in the sample, the experimentally derived value for $\mu(\theta)$ is not totally precise. In fact, what we obtain is an observation, x, which is an estimate of $\mu(\theta)$. Thus, x is defined as a realization of a random variable X where

$$X = \mu(\theta) + \text{noise}.$$

The noise, or error, term is assumed to be unbiased and to have a normal distribution with mean zero and standard deviation σ. This σ is taken to be the standard error quoted by the radiocarbon laboratories (see Bowman 1991 for details of how σ is obtained). That is, we have

$$X \sim N(\mu(\theta), \sigma^2).$$

The exact form of $\mu(\theta)$ is unknown but the internationally agreed calibration curve (Stuiver and Pearson 1986; Pearson and Stuiver 1986; Pearson *et al.* 1986) is (at the time of writing) used to estimate it. This consists of bi-decadal measurements of the ^{14}C content of tree-rings of known age and is usually expressed in its piecewise linear form so that

$$\mu(\theta) = \begin{cases} a_1 + b_1\theta & (\theta \leqslant t_o) \\ a_k + b_k\theta & (t_{k-1} \leqslant \theta < t_k, k = 1, 2, \ldots, K) \\ a_K + b_K\theta & (\theta > t_K) \end{cases}$$

where $K + 1$ is the number of bi-decadal measurements used to define the calibration curve and a_ks and b_ks are known constants. Consequently, the likelihood function which relates a given radiocarbon determination, x, to the calendar date, θ, is expressed as

$$p(x|\theta) = \begin{cases} \dfrac{1}{\sqrt{2\pi}\sigma}\exp - \dfrac{(x - a_1 = b_1\theta)^2}{2\sigma^2} (\theta \leqslant t_0) \\ \dfrac{1}{\sqrt{2\pi}\sigma}\exp - \dfrac{(x - a_k - b_k\theta)^2}{2\sigma^2} (t_{k-1} \leqslant \theta < t_k, k = 1, 2, \ldots, K) \\ \dfrac{1}{\sqrt{2\pi}\sigma}\exp - \dfrac{(x - a_K - b_K\theta)^2}{2\sigma^2} (\theta > t_K) \end{cases} \tag{3}$$

Using Bayes's theorem the posterior probability density of θ is given by

$$p(\theta|x) \propto p(x|\theta)p(\theta)$$

where $p(\theta)$ represents our prior knowledge about θ. For example, if we have no a priori archaeological knowledge about the value of θ, we can adopt a *vague prior* for θ so that $p(\theta) = 1$ over the whole range of the calibration curve. This is the approach adopted by most of the widely available computer calibration packages (such as CALIB; Stuiver and Reimer 1993). In practice, $p(x|\theta)$ is evaluated at different values of θ, say at yearly intervals, and $p(\theta|x)$ found by normalizing so that the sum of $p(\theta|x)$ at these different values of θ is 1.

As a further example, suppose we know that a certain event should (on the basis of archaeological or historic evidence) have occurred between two known calendar dates ψ_1 and ψ_2 (note that all dates are quoted BP). We assume that $\psi_1 > \psi_2$ and that no date in the range is more likely than any of the others. This information can easily be incorporated in to the calibration process since the prior for θ is now

$$p(\theta) = \begin{cases} (\psi_1 - \psi_2)^{-1} & \psi_1 > \theta > \psi_2 \\ 0 & \text{otherwise} \end{cases}$$

In this case we proceed as for the vague prior example given above, but now only evaluate $p(\theta|x)$ for values of θ within the interval ψ_1 to ψ_2 BP.

Another common type of archaeological information is that a priori one event must be later than another. For example, let θ_1 and θ_2 be the unknown calendar dates of two events. On the basis of stratigraphic information, the event with calendar date θ_1 occurred after that with calendar date θ_2. Thus, if the dates are expressed in years BP, we have $\theta_1 < \theta_2$. Now, let x_1 and x_2 be the corresponding observed radiocarbon determinations and let σ_1 and σ_2 be the associated standard deviations. Assuming that the two radiocarbon determinations are independent of each other, the likelihood relating x_1 and x_2 to θ_2 to θ_1 and θ_2 is given by

$$p(x_1, x_2|\theta_1, \theta_2) = p(x_1|\theta_1)p(x_2|\theta_2)$$

where $p(x_i|\theta_i)$ is given by equation (3) above.

If the archaeological evidence suggests that $\theta_1 < \theta_2$ but nothing else is known about θ_1 and θ_2 then we take the prior for θ_1 and θ_2 as

$$p(\theta_1, \theta_2) = \begin{cases} 1 & \theta_1 < \theta_2 \\ 0 & \text{otherwise} \end{cases}$$

By Bayes's theorem, the posterior distribution of θ_1 and θ_2 is given by

$$p(\theta_1, \theta_2|x_1, x_2) \propto p(x_1, x_2|\theta_1, \theta_2)p(\theta_1, \theta_2).$$

In this situation we evaluate $p(\theta_1, \theta_2|x_1, x_2)$ over a grid of values of θ_1 and θ_2, noting that it is zero if $\theta_1 > \theta_2$. Inferences about θ_1 and θ_2 can be made from this posterior or probabilities of θ_1 can be evaluated using the relationship

$$p(\theta_1|x_1, x_2) = \sum_{\theta_1 < \theta_2} p(\theta_1, \theta_2|x_1, x_2)$$

(and similarly for the posterior probability of θ_2).

Of course, the three situations just considered are all very simple because they involve no more than two events. More realistic and commonly occurring scenarios will comprise many events with possibly fairly complex interrelationships. However, in our experience, they may be broadly subdivided into two types of problem relating to (i) *events* and (ii) *phases*. The two types of problem are different both archaeologically and in terms of the models that statisticians use to investigate them.

Events Event-based investigations are relatively straightforward in as much as the archaeologist can usually obtain organic material which directly dates the events in question. The types of archaeological questions posed when dating events are, however, many and varied. Examples include, 'what is the chronological sequence of these events?', 'given that stratigraphy provides a chronological sequence for some of the events, can other events be placed within it?' and 'if we have chronologies of events from two different sites, can we relate the dates of events on one site to those on the other?'.

Buck *et al.* (forthcoming) describe the application of Bayesian methods to the analysis of 15 radiocarbon determinations from the Early Bronze Age settlement of St. Veit-Klinglberg, Austria. The primary objective in excavating this site was to investigate the organization of a community believed to be associated with the production and distribution of copper. During the excavation archaeologists collected samples for radiocarbon dating in the hope that they would provide (a) an estimate of the absolute date of the development of copper production in the region and (b) a foundation for the dating of Early Bronze Age domestic pottery. Because of its sloping location on a hill spur the site showed complex depositional patterns. In some places erosion had removed all archaeological deposits; in others deposits had accumulated to produce stratigraphic sequences. However, such localized sequences could not be related to one another stratigraphically because of the intervening erosion. A similar problem was posed by the defensive wall around the site which had been built at some point during the occupation; again it had no stratigraphic connection with other deposits.

Nevertheless, using stratigraphic evidence, it was possible to partially order the calendar dates of ten events. Let θ_i denote the calendar date of context i (where i is the context number allocated at the time of excavation) then the archaeological information can be expressed in the form of the following inequalities.

$$\theta_{758} > \theta_{814} > \theta_{1235} > \theta_{358} > \theta_{813} > \theta_{1210},$$

$$\theta_{493} > \theta_{358},$$

$$\theta_{925} > \theta_{923} > \theta_{358}$$

and

$$\theta_{1168} > \theta_{358}$$

With no other archaeological knowledge, the θ_is are assumed to have a uniform prior subject to these constraints. Thus the posterior of $\boldsymbol{\theta}$ is given by

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \prod_i \frac{1}{\sigma_i} \exp\left\{ -\frac{(x_i - \mu(\theta_i))^2}{2\sigma_i^2} \right\}$$

for values of $\boldsymbol{\theta}$ satisfying the above constraints.

In Figure 2 we illustrate the nature of the results obtained in investigations of this sort by providing plots of the posterior probability of θ_{1235}. In Figure 2 (a) the stratigraphic information is included in the calibration process and in 2 (b) it is ignored. We see clearly here the marked effect of the stratigraphic information and note that when it is included the calibrated date range is 200 years shorter than when it is ignored. Even in this example, where the archaeological information is relatively complex, the application of Bayes's theorem to produce the posterior of θ is deceptively simple. However, when attempting to

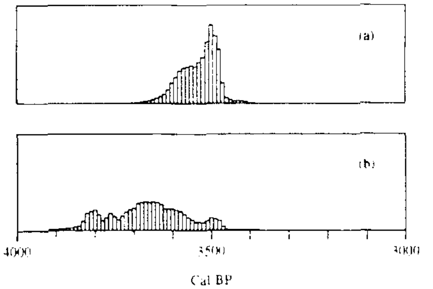

Figure 2 *Calibrated date distributions for context 1235 at St. Veit-Klinglberg* (a) *when the stratigraphic information is accounted for and* (b) *when it is ignored.*

obtain the posterior probabilities for each of the θ_is, computational difficulties arise if we attempt to proceed as before. This is due to the large number of parameters and the complex inequalities involved. Consequently, in obtaining the information represented in Figure 2, recent developments in numerical methods (especially the Gibbs sampler) proved particularly powerful.

Buck *et al.* (forthcoming) address questions of particular archaeological interest such as (a) the dates of the ten events and (b) the time period between certain specific events. In addition, they are able to use radiocarbon dating to place into the chronology several deposits which cannot be stratigraphically related to the main sequence. As a result, archaeologists now have a firm indication of the occupation periods of the site and how localized stratigraphic chronologies relate to each other. On broader questions, they now have dates for the development of copper production in the area and for the pottery assemblages found on the site.

Phases A phase is usually seen as a period of time during which a particular type of archaeological activity took place. Phases can, therefore, cover broad time spans (such as the Bronze Age) or rather shorter periods of time (such as the occupation of a particular habitation structure). In either case archaeologists will obtain radiocarbon samples that date events within the phase and wish to use these to allow them to make inferences about, for example, the dates of the start and end of the phase, its duration and its relation to other phases on the same or different sites. In these situations we are not particularly interested in the dates of the organic samples, in their own right, but the chronological information they provide about the phases in which they were deposited. Thus, with phase-based investigations we need to use a model to relate the available radiocarbon evidence to the phases we actually wish to date. The most famous study in which radiocarbon determinations were used to fix the calendar dates of phase boundaries is that of the Danebury Iron Age hillfort ceramic phases (see Buck and Litton forthcoming for the latest statistical analysis).

Zeidler *et al.* (forthcoming) report the analysis of radiocarbon determinations from 16 sites in the Jama River Valley in Ecuador. As a result of recent archaeological survey in the Jama River Valley, seven major archaeological phases have thus far been defined spanning about 4000 years. Several lines of investigation have been utilized in establishing this phasing including archaeological stratigraphy (interspersed with layers of volcanic ash) and detailed study of ceramic typologies. Initially, one major site was excavated (San Isidro) in the upper valley which, through its deep stratigraphy, led to a 'master' ceramic sequence.

Over subsequent seasons of study, the archaeologists undertook further excavation, correlated the information between the various sites studied and selected suitable samples from archaeological contexts on these sites to be submitted for radiocarbon dating. Some 37 radiocarbon determinations were made, with each being clearly and unambiguously assigned to one of the seven phases.

In the Jama River Valley study, the primary use of the radiocarbon determinations was in estimating the beginning and ending dates for each of the seven phases. Consequently let ψ_{2j-1} and ψ_{2j} represent the beginning and ending dates (measured BP) of phase $j(j = 1, 2, \ldots, 7)$. Let n_j be the number of samples assigned to the jth phase. Let θ_{ij} represent the calendar date of the ith radiocarbon sample in the jth phase but, since nothing is known about its date within the phase, it is assumed a priori that θ_{ij} takes any value in the range ψ_{2j-1} to ψ_{2j} with equal probability. Prior information about the ψ_js consists of constraints since information is available which partially orders the phases. In fact the ordering information may be expressed as

$$\psi_1 > \psi_2 \geqslant \psi_3 > \psi_4 \geqslant \psi_5 > \psi_6 \geqslant \psi_7 > \psi_8$$
$$\psi_7 > \psi_9 > \psi_{11} \text{ and } \psi_8 > \psi_{10} > \psi_{12}$$

and, since phase 6 and phase 7 are known to abut, we have

$$\psi_{12} = \psi_{13}$$

The posterior density of $\boldsymbol{\theta}$ and $\boldsymbol{\psi}$ is given by

$$p(\boldsymbol{\theta}, \boldsymbol{\psi}|\boldsymbol{x}) \propto \prod_{j=1}^{7} \left\{ (\psi_{2j-1} - \psi_{2j})^{-n_j} \prod_{i=1}^{n_j} \frac{1}{\sigma_{ij}} \exp - \left\{ \frac{(x_{ij} - \mu(\theta_{ij}))^2}{2\sigma_{ij}^2} \right\} \right\}.$$

for values of the ψ_{2j}s, ψ_{2j-1}s and θ_{ij}s satisfying the above constraints.

Of particular archaeological interest in this example are (a) the dates of the phase boundaries, (b) the lengths of the phases and (c) the sensitivity of the calendar dates obtained to the archaeological information used to build the model. In their paper Zeidler *et al.* (forthcoming) address all three questions and place particular emphasis on the sensitivity analysis which established that the archaeological prior knowledge and the radiocarbon evidence were in broad agreement. The calendar dates resulting from the Bayesian analysis of the Jama River Valley radiocarbon determinations provide a regionally specific cultural sequence which is firmly grounded in detailed prior knowledge. As such it can be profitably contrasted with the generalized absolute chronology established for the pan-Ecuadorian periodification scheme as a means of highlighting the parallels and the discrepancies between the two. Thus it enables the archaeologists to test whether ideas and theories based on better-studied areas of Ecuador can be carried across to the Jama River Valley study area which, until now, have not been adequately tested through sustained field research.

Extensions Clearly there will be situations in which the archaeology requires us to combine both phases and events within the same model. An example of the statistical modelling involved in such investigations is given by Christen (1994a). Moreover, in the light of research by Baillie (1990), it is almost essential that archaeological investigations using radiocarbon dating involve the identification of possible aberrant determinations (outliers) within groups of samples from both single events and from phases. A Bayesian method for the detection of outliers in radiocarbon dating is given by Christen (1994a and 1994b) and Buck *et al.* (1994).

Another application of radiocarbon dating is the widely discussed field of 'wiggle matching' (e.g., Pearson 1986; Baillie and Pilcher 1988; Baillie 1990; Clymo *et al.* 1990; Manning and Weninger 1992). The important feature of 'wiggle matching' is that a priori evidence exists about the time interval between a group of related radiocarbon samples. This evidence may take the form of exact time intervals (e.g., from floating tree-ring sequence) or postulated time intervals (e.g., from stratigraphic evidence). The Bayesian approach to 'wiggle matching' is given by Christen and Litton (forthcoming) and is illustrated by two examples of different complexity.

A somewhat more subtle, but nonetheless important, consideration for the calibration of radiocarbon determinations is that the evaluations that comprise the calibration data are themselves 'uncertain'. Like radiocarbon determinations from archaeological samples, the bi-decadal evaluations which make up the curve each have associated with them their own quoted laboratory error. These errors are generally small in comparison to the size of errors quoted on archaeological samples since the calibration data are based on multiple analyses of material with the same calendar date. However, with recent decreases in standard deviations reported by the laboratories, the uncertainty associated with the calibration curve will need to be modelled into the calibration process. A Bayesian approach for doing this is given by Christen (1994a).

Implications Clearly, archaeologists who elect to use a Bayesian approach to the calibration of their radiocarbon determinations now have available to them the means for answering questions not widely addressed before. There are, however, some important concepts that need to be understood by the archaeological community (such as model building, prior elicitation, sample selection and interpretation of statistical distributions) before the full benefits of the Bayesian approach can be realized. From a practical viewpoint, since modern Bayesian approaches have only recently begun to be applied to radiocarbon dating, there is as yet no generally available computer software. However, the potential interpretational benefits are great and, assuming that the Bayesian philosophy gains support, the demand for such software will surely increase. Fortunately, the algorithms are not difficult to implement and radiocarbon laboratories will find it in their own best interests to support this work.

BAYESIAN CLUSTER ANALYSIS

One particularly common application of statistics to archaeological data interpretation arises when archaeologists wish to subdivide multivariate data into groups, within which the samples have similar properties. The desire to make such subdivisions occurs, for example, in the development of artefact typologies, the provenancing of ceramics, the sourcing of obsidian and the assigning of handmade objects to the manufacturer who made them on the basis of their form. Investigations of this type give rise to two distinct statistical problems: those where the nature and number of clusters is known a priori and those where they are to be defined as part of the statistical analysis. Statistical methods used for tackling problems of the first type are known as discriminant analysis techniques and those for approaching problems of the second type are known as cluster analysis.

A number of classical statistical techniques are available for undertaking both discriminant and cluster analysis. Many of these are implemented as part of well-known computer packages and so are widely adopted by archaeologists. For many archaeological applications, however, these algorithms are of limited use since they do not permit the inclusion of any a priori knowledge that might be available about links between samples or about the nature of the distributions from which the samples arise. As with other fields of archaeological study considered above, the fact that such information exists and is not included in the statistical analysis may lead to results with little, or no, archaeological interpretability. This point was noted by Craddock *et al.* (1983) who used discriminant analysis of chemical compositional data to aid in the sourcing of polished flint axes. In this particular case, knowledge based upon radiocarbon determinations and upon production styles at several flint mines suggested, a priori, that some of the potential sources were unlikely to have given rise to certain of the flint axes. To allow for this in their discriminant analysis, the authors adopted weighting factors which reduced the prior probability of assigning those particular axes to the 'unlikely' flint mines, while increasing their probability of assignment to the more 'likely' ones. Although not a full Bayesian approach, this is the only investigation we are aware of in which such prior archaeological knowledge is explicitly stated in this way.

The theoretical details necessary to allow full Bayesian approaches to both discriminant and cluster analysis were first suggested some time ago (Binder 1978) but practical difficulties have always hampered their implication and general use. However, the recent advances in MCMC methods using iterative resampling schemes (in particular the Gibbs sampler) have made it feasible to tackle such problems from a Bayesian perspective. Cluster analysis does require rather more complex modelling than does discriminant analysis, but the extension involved is not theoretically intractable, it simply needs greater sophistication of software and larger amounts of computer time. Naturally, some specific applications will require additions to any basic model and some unusual problems will require significant alterations. Nonetheless, we feel that it is worthwhile outlining a generalized framework from which such developments can be made.

For the Bayesian approach to discriminant analysis, we view the problem in the following way. Suppose that there are k underlying populations (e.g., artefact types, ceramic provenances, obsidian sources) from which the samples under investigation are drawn. An observation, $\boldsymbol{x}_j$, is drawn from population, i, with probability θ_i. Conditional on being from the ith population, $\boldsymbol{x}_j$ is modelled as having a distribution which depends on $\boldsymbol{\lambda}_i$ where $\boldsymbol{\lambda}_i$ is a vector of parameters. Thus the overall distribution of $\boldsymbol{x}_j$ is viewed as a discrete mixture of k distributions where the mixing probabilities, θ_js, and the population parameters are in general unknown. To carry out an analysis, the form of the distribution needs to be specified. In many situations, the data (or if necessary a transformation of the data) are modelled as having a multivariate normal distribution with mean vector $\boldsymbol{\mu}$ and covariance matrix Σ, denoted by $N(\boldsymbol{\mu}, \Sigma)$. In this situation, $\boldsymbol{\lambda}_i = (\boldsymbol{\mu}_i, \Sigma_i)$ and so, $\boldsymbol{x}_j \sim N(\boldsymbol{\mu}_i, \Sigma_i)$ with probability θ_i.

Of course, the objective of our analysis is to allocate individual observations to groups. To do so, let z_j be a classification variable such that $z_j = i$ implies that the jth observation is assigned to the ith population. In practice, most if not all the z_js will be unknown and have to be estimated within the Bayesian procedure. It is possible, however, that some of the classifications will be known and these can then be used as *training data* to provide prior information about some of the populations (see Lavine and West 1992).

The above discriminant methodology assumes that k, the number of underlying populations, is known. Erkanli *et al.* (1992) have, however, also developed a Bayesian methodology for the much harder problem of grouping data when the number of populations in the mixture is unknown. Nonetheless, when first adopting such a methodology for investigating archaeological data we can expect some difficulties. In our experience, these include problems associated with the high dimensionality of many of the problems (e.g., chemical compositional analyses of ceramics) which result in a need for large amounts of data and computer power, and the lack of experience of chemists and archaeologists in expressing their prior information in a suitable form (Buck 1994).

So, what are the advantages of approaching cluster analysis in this complex and difficult way? Firstly, a Bayesian analysis of the type described above will give us posterior probabilities of the number of groups, their sizes and their compositions. Secondly, the same algorithms are easily adapted to take account of missing datum values. Thirdly, the methodology is extremely flexible and can be adapted to situations modelled using distributions other than the multivariate normal.

In contrast, traditional methods only offer *ad hoc* procedures for answering some important points. For instance, when using traditional approaches, samples whose data contain missing values must be discarded or those values imputed before the analysis can take place, whereas the complication of missing datum values fits neatly into the Bayesian framework. In addition, traditional algorithms are built around just one type of model which cannot readily be altered. And, finally, using traditional methods, interpretation of how many groups there are relies upon subjective assessment of graphs and other plots since they do not provide any probabilistic information about number of groups or group sizes. Surely it is preferable to provide results—in the form of probabilities—that have direct relation to the questions posed. Such results will be easier both to interpret and to report to other workers.

CONCLUSIONS

As we have shown, several applications of Bayesian statistics to archaeology have already been undertaken, most of which have lead to results of far greater interpretability than those obtained using traditional statistical approaches. As yet, however, the Bayesian approach is not routinely used and there is still some hostility to, and controversy about, the change in philosophy required. So, what can we expect for the future?

First and foremost we stress the point, raised by the examples of Flag Fen and Bede House Farm, that many archaeologists already think and work within a philosophical framework which is essentially Bayesian. Consequently the explicit adoption of Bayesian statistical methods seems a natural progression. This will mean, however, that archaeologists need to learn to couch their prior beliefs in terms of probabilities. Secondly, in adopting Bayesian methods, archaeologists will need to make explicit statements about the underlying processes that give rise to their data. In developing suitable models we must encourage archaeologists to think very carefully about the problems they wish to solve and to make clear statements about the questions they wish to answer. Thirdly, and directly related, archaeologists and other subject matter experts will need to express clearly their a priori opinions about the problem under investigation. It will no longer be sufficient to wait until after the statistical analysis is complete before deciding which pieces of prior information are relevant. All modelling assumptions and possible sources of prior information must be openly discussed and clearly stated before analysis begins. Subsequent sensitivity analysis may then be undertaken where necessary. Explicit statements of the model, prior information and the results of any sensitivity analysis are essential parts of any Bayesian investigation which will surely be welcomed by fellow workers who wish to undertake follow-up studies or make comparisons with other work in the same field.

The Bayesian philosophical perspective does have its detractors and each archaeologist will need to form his or her own opinion about its validity. This is the case, however, with any perspective on data interpretation and we feel that many workers will see that the benefits far outweigh the difficulties. There are, however, two commonly quoted reasons for the lack of adoption of Bayesian methodologies.

(a) There is a widespread belief that computation of posterior distributions for realistic, practical problems is not possible.

(b) It is widely held that prior information does not always exist and, when it does, its elicitation can appear extremely daunting.

As explained above, criticism (a) can now largely be overcome by the application of MCMC methods and criticism (b) is, we believe, also not insurmountable for the following reasons. Firstly, if there really is no useful prior information about one or more parameters of the model (as will occasionally be the case) this is not precluded by the Bayesian approach and can readily be handled. In such situations, inferences will be based solely on information from the data as with more traditional approaches. Secondly, adoption of MCMC methods has made it possible to use an enormously wide range of different distributions to represent prior beliefs. Consequently many types of information which were, for practical reasons, once hard to allow for within the Bayesian framework can now easily be incorporated. Thirdly, many people already use the information that Bayesians call *prior knowledge* in a retrospective fashion when they come to interpret the results obtained using more traditional statistical methods. All that is required now is that we raise awareness of the value of the information and encourage statisticians to seek it along with the information used to build models.

So, we encourage archaeologists to expend this extra effort, but how do they actually go about obtaining the posterior information they require? As noted above, genuinely practical Bayesian statistics is a new and rapidly growing field, and few software packages are, as yet, readily available. Over the next few years this will change, as workers in many other branches of science are appreciating the advantages of adopting such approaches and development of computer packages is now under way. For the time being at least, archaeologists will need to collaborate with statisticians, just as they do with other professional experts. It is unreasonable to expect archaeologists who usually lack sophisticated mathematical knowledge to undertake advanced statistical modelling and investigation without some guidance from the statistical community. As an analogy, most archaeologists already rely on the skills of pottery experts and radiocarbon laboratories, why not on those of statisticians too? For such a consultation process to be fruitful, however, both statisticians and archaeologists will need to be prepared to take responsibility for technology transfer between the two disciplines and to learn at least some of the professional vocabulary of the other subject area.

ACKNOWLEDGEMENTS

Without the collaboration of a number of colleagues, from a variety of disciplines, the development of Bayesian applications in archaeology would not have been possible. We wish to acknowledge the contributions of many people, in particular, Bill Cavanagh, James Kenworthy, Bob Laxton and Adrian Smith. In addition, we are grateful to a great number of other individuals who have permitted us to use their data or brought interesting problems to our attention. We also thank the referees for their encouraging and constructive comments. The second-named author is grateful to both the Mathematics and the Science-Based Archaeology Committees of the SERC who have supported her work during the last six years.

REFERENCES

Baillie, M. G. L., 1990, Checking back on an assemblage of published radiocarbon dates, *Radiocarbon*, **32** (3), 361–6.

Baillie, M. G. L., and Pilcher, J. R., 1988, Make a date with a tree, *New Scientist*, **117**, 48–51.

Barnett, V., 1982, *Comparative statistical inference*, 2 edn., Wiley, New York.

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

Bayes, T. R., 1763, An essay towards solving a problem in the doctrine of chances, *Phil. Trans. Roy. Soc. London*, **53**, 370–418.

Bernardo, J. M., and Smith, A. F. M., 1994, *Bayesian theory*, Wiley, London.

Bernoulli, J., 1713, *Ars Conjectandi*, Book Four.

Besag, J., 1986, On the statistical analysis of dirty pictures (with discussion), *J. Roy. Stat. Soc. (B)*, **48**, 259–302.

Binder, D. A., 1978, Bayesian cluster analysis, *Biometrika*, **65**, 31–8.

Bowman, S., 1990, *Radiocarbon dating*, British Museum Publications, London.

Box, G. E. P., and Tiao, G. C., 1973, *Bayesian inference in statistical analysis*, Addison-Wesley, Reading, MA.

Broadbent, S. R., 1956, Examination of a quantum hypothesis based on a single set of data, *Biometrika*, **43**, 32–44.

Buck, C. E., 1994, *Towards Bayesian archaeology*, unpubl. Ph.D. thesis, University of Nottingham.

Buck, C. E., and Litton, C. D., forthcoming, Statistical analysis of radiocarbon results from Danebury Iron Age hillfort, in *Danebury: an Iron Age hillfort in Hampshire: Vol. 6* (ed. B. Cunliffe).

Buck, C. E., Cavanagh, W. G., and Litton, C. D., 1988, The spatial analysis of site phosphate data, in *Computer applications and quantitative methods in archaeology, 1988* (ed. S. P. Q. Rahtz), 151–60, Brit. Archaeol. Rep. Internat. ser., **S446** (i), Oxford.

Buck, C. E., Cavanagh, W. G., and Litton, C. D., 1990, Image segmentation methods for archaeological field survey data, in *New tools from mathematical archaeology* (ed. J. Wiltowski), 55–68, Scientific Information Centre of the Polish Academy of Sciences, Warsaw, Poland.

Buck, C. E., Kenworthy, J. B., Litton, C. D., and Smith, A. F. M., 1991, Combining archaeological and radiocarbon information: a Bayesian approach to calibration, *Antiquity*, **65**, 808–21.

Buck, C. E., Cavanagh, W. G., Litton, C. D., 1992a, Tools for the interpretation of soil phosphate data from archaeological field surveys, in *Geoprospection in the archaeological landscape* (ed. P. Spoerry), 75–87, Oxbow monographs, **18**, Oxford.

Buck, C. E., Litton, C. D., and Smith, A. F. M., 1992b, Calibration of radiocarbon results pertaining to related archaeological events, *J. Archaeol. Sci.*, **19**, 497–512.

Buck, C. E., Litton, C. D., and Stephens, D. A., 1993, Detecting a change in the shape of a prehistoric corbelled tomb, *The Statistician*, **42**, 483–90.

Buck, C. E., Christen, J. A., Kenworthy, J. B., and Litton, C. D., 1994, Estimating the duration of archaeological activity using ^{14}C determinations, *Oxford J. Archaeol.*, **13** (2), 229–40.

Buck, C. E., Litton, C. D., and Shennan, S. J., forthcoming, A case study in combining radiocarbon and archaeological information: the Early Bronze Age settlement of St. Veit-Klinglberg, Land Salzburg, Austria, *Germania*.

Cavanagh, W. G., and Laxton, R. R., 1981, The structural mechanics of the Mycenaean tholos tomb, *Annu. Brit. Sch. Athens*, **76**, 109–40.

Cavanagh, W. G., and Laxton, R. R., 1982, Corbelled vaulting in the late Minoan tholos tombs of Crete, *Annu. Brit. Sch. Athens*, **77**, 65–78.

Cavanagh, W. G., and Laxton, R. R., 1987, Notes on building techniques in Mycenaean Greece and nuragic Sardinia, *Studies in Sardinian archaeology III* (ed. M. S. Balmuth), 39–55, Brit. Archaeol. Rep. Internat. ser., **S387**, Oxford.

Cavanagh, W. G., Laxton, R. R., and Litton, C. D., 1985, An application of change-point analysis to the shape of prehistoric corbelled domes; I. the maximum likelihood method, *PACT*, **11** (III.4), 191–203.

Cavanagh, W. G., Buck, C. E., and Litton, C. D., 1988a, The interpretation of noisy data from archaeological field survey: phosphate analysis, *Environmental Geochem. Health*, **10** (3/4), 92–5.

Cavanagh, W. G., Hirst, S., and Litton, C. D., 1988b, Soil phosphate, site boundaries and change-point analysis, *J. Field Archaeol.*, **15** (1), 67–83.

Christen, J. A., 1994a, *Bayesian interpretation of ^{14}C results*, unpubl. Ph.D. thesis, University of Nottingham.

Christen, J. A., 1994b, Summarising a set of radiocarbon determinations: a robust approach, *Applied Stat.*, **43** (3), 489–503.

Christen, J. A., and Litton, C. D., forthcoming, A Bayesian approach to wiggle-matching, *J. Archaeol. Sci.*

Clymo, R. M., Oldfield, F., Appleby, P. G., Pearson, G. W., Ratnesar, P., and Richardson, N., 1990, The record of atmospheric deposition on a rainwater-dependent peatland, *Phil. Trans. Roy. Soc. London (B)*, **327**, 331–8.

Cowgill, G. L., 1977, The trouble with significance tests and what we can do about it, *Am. Antiquity*, **42**, 350–68.

Cowgill, G. L., 1989, Formal approaches in archaeology, in *Archaeological thought in America* (ed. C. C. Lamberg-Karlovsky), 74–88, Cambridge University Press, Cambridge.

Cowgill, G. L., 1993, Distinguished lecture in archaeology: beyond criticizing New Archaeology, *Am. Anthropologist*, **95**, 551–73.

Craddock, P. T., Cowell, M. R., Leese, M. N., and Hughes, M. J., 1983, The trace element composition of polished flint axes as indicator of source, *Archaeometry*, **25** (2), 135–63.

de Finetti, B., 1974–5, *Theory of probability: a critical introductory treatment*, 2 vols, Wiley, New York.

DeGroot, M. H., 1970, *Optimal statistical decisions*, McGraw-Hill, New York.

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

Erkanli, A., Müller, P., and West, M., 1992, *Curve fitting using Dirichlet process mixtures*, Inst. Stat. & Decision Sci. Discussion Pap., **92–A09**, Duke University.

Fieller, N. R. J., 1993, Archaeostatistics: old statistics in ancient contexts, *The Statistician*, **42**, 279–95.

Freeman, P. R., 1976, A Bayesian approach to the megalithic yard, *J. Roy. Stat. Soc. (A)*, **139**, 20–55.

Gelfand, A. E., and Smith, A. F. M., forthcoming, *Bayesian practice*, Wiley, London.

Gilks, W., Richardson, S., and Spiegelhalter, D., forthcoming, *Markov chain Monte Carlo in practice*, Chapman and Hall.

Howson, C., and Urbach, P., 1989, *Scientific reasoning: the Bayesian approach*, Open Court, Illinois.

Jeffreys, H. S., 1961, *Theory of probability*, 3 edn., University Press, Oxford.

Kadane, J. B., and Hastorf, C. A., 1988, Bayesian palaeoethnobotany, in *Bayesian statistics 3* (eds J. M. Bernardo, M. H. DeGroot, D. V. Lindley and A. F. M. Smith), 243–60, Oxford University Press, Oxford.

Kendall, D. G., 1974, Hunting quanta, *Phil. Trans. Roy. Soc. (A)*, **276**, 231–66.

Lavine, M., and West, M., 1992, A Bayesian method for classification and discrimination, *Canadian J. Stat.*, **20** (4), 451–61.

Laxton, R. R., Cavanagh, W. G., Litton, C. D., Buck, C. E., and Blair, R., 1994, The Bayesian approach to archaeological data analysis: an application of change-point analysis to prehistoric domes, *Archeologia e Calcolatori*, **5**, 53–68.

Lindley, D. V., 1965, *Introduction to probability and statistics from a Bayesian viewpoint*, University Press, Cambridge.

Lindley, D. V., 1985, *Making decisions*, 2 edn., Wiley, London.

Litton, C. D., forthcoming, Methods and problems of dendrochronology as exemplified by Bede House Farm North Luffenham, Leicestershire, in *Historic buildings and dating by dendrochronology* (ed. G. I. Meirion-Jones), Oxbow, Oxford.

Litton, C. D., and Leese, M. N., 1991, Some statistical problems arising in radiocarbon calibration, in *Computer applications and quantitative methods in archaeology, 1990* (eds K. Lockyear and S. P. Q. Rahtz), 101–9, Brit. Archaeol. Rep. Internat. ser., **S565**, Oxford.

Manning, S. W., and Weninger, B., 1992, A light in the dark: archaeological wiggle matching and the 'J absolute chronology of the close of the Aegean Late Bronze Age, *Antiquity*, **66**, 636–63.

Naylor, J. C., and Shaw, J. E. H., 1985, *Bayes four user guide*, Statistics Group, University of Nottingham.

Naylor, J. C., and Smith, A. F. M., 1988, An archaeological inference problem, *J. Am. Stat. Assoc.*, **83** (403), 588–95.

O'Hagan, A., 1988, *Probability: methods and measurements*, Chapman and Hall, London.

O'Hagan, A., 1994, *Kendall's advanced theory of statistics: Vol. 2B, Bayesian inference*, Edward Arnold, London.

Orton, C., 1980, *Mathematics in archaeology*, William Collins Sons and Co. Ltd., Glasgow.

Pearson, G. W., 1986, Precise calendar dating of known growth-period samples using a 'curve fitting' technique, *Radiocarbon*, **28** (2A), 292–9.

Pearson, G. W., and Stuiver, M., 1986, High-precision calibration of the radiocarbon time scale, 500–2500 BC, *Radiocarbon*, **28** (2B), 839–62.

Pearson, G. W., Pilcher, J. R., Baillie, M. G. L., Corbett, D. M., and Qua, F., 1986, High-precision ^{14}C measurements of Irish oaks to show the natural ^{14}C variations from AD 1840 to 5210 BC, *Radiocarbon*, **28** (2B), 911–34.

Porteous, H. L., 1973, Megalithic yard or megalithic myth?, *J. Hist. Astronomy*, **4**, 22–4.

Porteous, H. L., 1977, The saga of the megalithic yard, *Bull. Applied Stat.*, **4** (1), 28–39.

Pryor, F. M. M., 1992, Current research at Flag Fen, Peterborough, *Antiquity*, **66**, 439–57.

Pryor, F. M. M., French, C. A. I., and Taylor, M., 1986, Flag Fen, Fengate, Peterborough I: discovery, reconaissance and initial excavations (1982–85), *Proc. Prehist. Soc.*, **52**, 1–24.

Ruggles, C., 1986, You can't have one without the other? I.T. and Bayesian statistics and their possible impact within archaeology, *Sci. & Archaeol.*, **28**, 8–15.

Smith, A. F. M., 1991, Bayesian computational methods, *Phil. Trans. Roy. Soc. London (A)*, **337**, 369–86.

Smith, A. F. M., and Roberts, G. O., 1993, Bayesian computation via the Gibbs sampler and related Markov chain Monte Carlo methods, *J. Roy. Stat. Soc. (B)*, **55** (1), 3–23.

Smith, J. Q., 1988, *Decision analysis: a Bayesian approach*, Chapman and Hall, London.

Stuart, A., and Ord, J. K., 1991, *Kendall's advanced theory of statistics: Vol. 2, classical inference and relationships*, 5 edn., Edward Arnold, London.

Stuiver, M., and Pearson, G. W., 1986, High-precision calibration of the radiocarbon time scale, AD 1950–500 BC, *Radiocarbon*, **28** (2B), 805–38.

Stuiver, M., and Reimer, P. J., 1993, Extended ^{14}C data base and revised CALIB 3.0 ^{14}C age calibration program, *Radiocarbon*, **35** (1), 215–30.

Thom, A., 1955, A statistical examination of the megalithic sites in Britain, *J. Roy. Stat. Soc. (A)*, **118** (III), 275–95.

Thom, A., 1962, The megalithic unit of length, *J. Roy. Stat. Soc. (A)*, **125**, 243–51.

Thom, A., 1964, The larger units of length of megalithic man, *J. Roy. Stat. Soc. (A)*, **127**, 527–33.

Thom, A., 1967, *Megalithic sites in Britain*, Clarendon Press, Oxford.

Urbach, P., 1992, Regression analysis: classical and Bayesian, *Brit. J. Phil. Sci.*, **43**, 311–42.

Zeidler, J. A., Buck, C. E., and Litton, C. D., forthcoming, *The integration of archaeological phase information and radiocarbon results from the Jama River Valley, Ecuador: a Bayesian approach.*