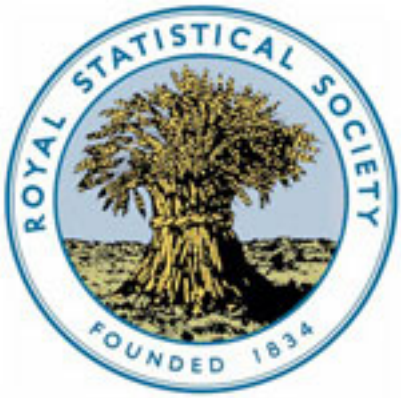

Appl. Statist. (2000)
49, *Part* 4, *pp.* 423–440

Bayesian models for relative archaeological chronology building

Caitlin E. Buck

Cardiff University, UK

and Sujit K. Sahu

University of Southampton, UK

[Received April 1999. Final revision February 2000]

Summary. For many years, archaeologists have postulated that the numbers of various artefact types found within excavated features should give insight about their relative dates of deposition even when stratigraphic information is not present. A typical data set used in such studies can be reported as a cross-classification table (often called an abundance matrix or, equivalently, a contingency table) of excavated features against artefact types. Each entry of the table represents the number of a particular artefact type found in a particular archaeological feature. Methodologies for attempting to identify temporal sequences on the basis of such data are commonly referred to as seriation techniques. Several different procedures for seriation including both parametric and non-parametric statistics have been used in an attempt to reconstruct relative chronological orders on the basis of such contingency tables. We develop some possible model-based approaches that might be used to aid in relative, archaeological chronology building. We use the recently developed Markov chain Monte Carlo method based on Langevin diffusions to fit some of the models proposed. Predictive Bayesian model choice techniques are then employed to ascertain which of the models that we develop are most plausible. We analyse two data sets taken from the literature on archaeological seriation.

Keywords: Archaeological seriation; Correspondence analysis; Hybrid Monte Carlo methods; Langevin algorithms; Markov chain Monte Carlo methods; Model choice; Predictive distribution

1. Introduction

Modern archaeological chronology building involves the use of a range of techniques to provide insights into both absolute and relative timescales. Typically, absolute dates are obtained via the use of scientific dating methods such as radiocarbon (or ^{14}C) dating and dendrochronology. Relative chronologies, however, are commonly built on the basis of information gained during excavation. Such information includes stratigraphic sequences and the nature and quantity of the artefacts recovered. Archaeologists would, on the whole, prefer to build their relative chronologies on the basis of well-excavated vertical stratigraphies, but this is not always possible. In some situations this is because the site(s) did not form in a strictly sequential manner and so vertically stratified deposits did not form; in others it is because the material was partially mixed before or after deposition and there is no longer a reliable relationship between the vertical location in the ground and the relative date of deposition; in others it is because there is stratification in parts of the site(s), but it is

Address for correspondence: Caitlin E. Buck, School of History and Archaeology, Cardiff University, PO Box 909, Cardiff, CF1 3XU, UK.
E-mail: BuckCE@cardiff.ac.uk

 0035–9254/00/49423

not possible to link layers in one area with those in another. In situations where stratification is not present, or is incomplete, as in the following examples, archaeologists commonly turn to the artefacts excavated in an attempt to derive relative chronological information.

1.1. Example 1: the refuse mounds at Awatovi, Arizona

Ever since the Spanish expeditions to the Americas in the 16th century, there has been interest in the Hopi pueblos of north-eastern Arizona. One such pueblo, at Awatovi, is known to have been visited and recorded as early as 1540. Some time around 1700, however, the occupants of the site were attacked and large numbers of the male inhabitants were killed. Although some people may have continued to live at the site after this date it was largely abandoned, leaving behind a considerable archaeological record relating to periods both before and after the first Spanish contact.

In the late 1930s, archaeologists from the Peabody Museum led an excavation of parts of the archaeological record at Awatovi. The excavation gave rise to considerable quantities of decorated ceramic shards deposited in refuse mounds at the site. Burgh (1959) reported on the ceramics found and on their interpretation in association with the considerable stratigraphic information that was also obtained during the excavation. He described several stratified profiles that contained huge numbers of ceramic shards from a long period of human occupation.

Burgh (1959) concluded that there are five general categories of decoration on the ceramics and that these relate to the chronology of the layers at the site. The data that we use here (Table 1) relate to the five ceramic decoration types identified by Burgh: black on white (BW), black on orange (BO), black on yellow (BY), orange paste polychrome (OP) and yellow paste polychrome (YP). They arise from an amalgamation, made by Burgh, of ceramics from 13 excavated profiles (see Fig. 11 of Burgh (1959) for details).

1.2. Example 2: mesolithic stone tools from southern England

Jacobi *et al.* (1980) published the results of collaborative work between archaeologists and mathematicians. They worked with data relating to the numbers of seven types of mesolithic flint tools (known as microliths) from six different sites in southern England; see Table 2. The objective was to identify the relative chronological order of the sites (rows) by studying the changes in the numbers of the seven types of microliths that were available at the various sites.

Table 1. Proportions of five different types of painted pottery in seven layers (from 13 ceramic profiles) in the refuse mounds at Awatovi, Arizona (Burgh, 1959)

Layer	*Proportions of the following types:*				
	BW	*BO*	*BY*	*OP*	*YP*
d	4	6	86	0	4
e	6	14	76	0	4
f	8	19	70	1	2
g	18	49	30	3	0
h	23	54	20	3	0
i	32	49	5	14	0
j	39	43	0	18	0

Table 2. Proportions of seven different types of microliths from six mesolithic sites in southern England (Jacobi *et al.*, 1980)

Site	*Numbers of each stone tool type*						
1	20	3	4	42	18	0	13
2	85	3	12	0	0	0	0
3	26	40	8	0	0	26	0
4	20	1	4	13	58	0	4
5	67	10	23	0	0	0	0
6	26	29	8	3	0	33	1

There are several competing methods for building relative chronologies using such data sets. When applied to the stone tools data two common approaches, discussed in detail below, produce the orders 2, 5, 3, 6, 1, 4 and 3, 6, 5, 2, 1, 4. In Fig. 1 we plot these orders against the relative abundance of the seven microliths. Since these two orders are likely to give rise to different archaeological conclusions, the primary objective of this paper is to develop methodologies for selecting between the possible chronological orders by using Bayesian model-based approaches.

1.3. Overview

The development of relative chronologies on the basis of numbers and types of artefacts found (or their presence or absence) is commonly referred to as *seriation*. Chronological seriation has a long history in archaeological research which is not discussed in detail here. It is, however, worth noting that the most famous of the early work (though non-mathematical) was formal and provides the intuitive basis for much of the seriation that is undertaken today (Petrie, 1899).

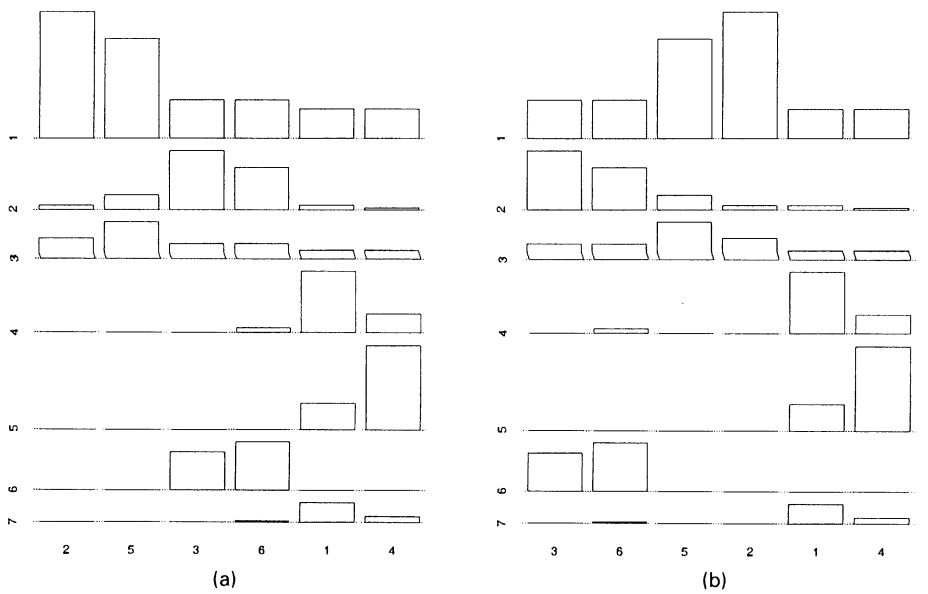

Fig. 1. Plots of the most likely orders of the six mesolithic sites from southern England using (a) the Robinson–Kendall modelling approach and (b) correspondence analysis

More recent uses of chronological seriation all have the same basic deterministic model at their core. This model was first formally made explicit by Robinson (1951) who attempted to develop a numerical method for seriating pottery assemblages. Robinson (1951), page 293, stated that 'over the course of time pottery types come into and go out of general use by a given group of people'. In other words, he assumed that artefact types come into use, go out again, but never come back again into general usage. The work of Kendall (1970, 1971) was of primary importance in formalizing Robinson's hypothesis (as a consequence we henceforth refer to this as the Robinson–Kendall (RK) model). More recently both deterministic and stochastic versions of this model have been considered by many researchers; see for example Laxton (1987), Laxton and Restorick (1989) and Buck and Litton (1991) and references therein.

Very few real data sets fit the deterministic RK model *exactly*, however. There are many good reasons for arguing that this is not to be expected. In particular, consider the following.

(a) The RK model assumes that an object type will *never* appear in the archaeological record, disappear and then reappear again sometime later. This assumption does not allow for some usage and discard behaviour that seems likely to have occurred in the past. Two examples of this are technological innovation and the passing of heirlooms from one generation of a community to the next. In the case of technical innovation a new tool type, for example, might be used and discarded sporadically. This could easily result in the appearance and disappearance of a tool type, at the beginning of its use, before it is adopted more widely. In the case of heirlooms, which might be passed through several hands before being discarded, we could expect a reappearance of objects in the archaeological record long after they have ceased being in widespread use.

(b) When used deterministically, as has commonly been the case, the RK model also fails to make any allowance for noise in the data. Unfortunately, the archaeological recovery of artefacts is very error prone. Many artefact types decompose in the ground and may not reliably be available for recovery, others are small and are simply not found and others still are not found at all because the contexts in which they lie have not yet been excavated.

For these reasons (among others) the RK model can be seen as rather prescriptive and it is desirable to seek methodologies which allow some relaxation of the constraints. In this paper we consider two approaches to achieving this. The first is to extend the RK model to accommodate the issues described above. The second is to adopt a model-based implementation of an exploratory statistical tool (correspondence analysis (CA)) which has been used to obtain seriations of real data for a considerable time. Our suggestions for extending the RK model are intuitive and relatively simple and will be described in detail in Section 2. The second approach is somewhat less intuitive, has quite an extensive history of use by archaeologists and thus requires a rather more elaborate introduction.

Archaeologists seeking a less prescriptive approach to seriation than those suggested by Kendall (1971) or Laxton (1987) have adopted a number of exploratory statistical tools. Those based on principal components analysis have proved particularly popular. A principal component analysis method, tailored for discrete data, known as CA is one such approach; see for example Madsen (1988), Baxter (1994), pages 100–139, Goodman (1986) and references therein. CA is generally viewed as an exploratory tool for recovering the dominant order of the rows of the data matrix. Total variation (according to a suitable χ^2-metric) in the data is split into different principal component axes.The seriated order is typically established via a visual inspection of a plot of the first two components of such an analysis or on the basis of the first component alone. Although archaeologists have traditionally used CA as an

exploratory tool, we note that statistical models can be developed from it. Those seeking model-based approaches to CA have usually constructed a saturated canonical correlation model in which the number of parameters is the same as the number of data points; see for example Goodman (1986). However, saturated models are not useful for statistical inference because they do not have non-zero error degrees of freedom. Consequently non-saturated versions of these models are used for statistical inference and these are the models that are adopted here.

An unsaturated CA model, when fitted using classical techniques such as the maximum likelihood method, produces only one candidate order for seriation. As a result, these methods fail to identify other possible candidate orders. Consequently, it is not possible to obtain relative odds for different chronological orders. Bayesian versions of these CA models, developed in Section 3, are attractive in this regard and, as far as we are aware, such implementations have not been considered for archaeological inference before. In addition to allowing us to obtain relative odds of the orders obtained, Bayesian statistical methods are particularly suitable for archaeological inference because they allow the possibility of incorporating prior information. Since archaeological excavations involve the identification of relative chronological relationships between features and contexts, prior information about temporal order is sometimes available and should, if possible, be interpreted in a coherent fashion along with the evidence from the artefacts found.

Conventional Markov chain Monte Carlo (MCMC) simulation methods, such as the single-component updating Gibbs sampler, tend to do very poorly for fitting the unsaturated CA models. These methods fail because the associated posterior densities are generally not log-concave and there is a high correlation between the parameters (induced by many constraints placed on them). It is also difficult to tune the proposal scaling of the global Metropolis–Hastings algorithms which do not use structural information of the target posterior distribution. In this paper we adopt MCMC methods based on Langevin diffusions which use gradient information of the posterior distribution to simulate samples from it.

At this point, it is appropriate to note that the CA approach to seriation does not incorporate the assumptions of the RK model. As a consequence, methods based on the RK model and those based on CA can produce conflicting chronological orders for the same data set. We illustrate this issue using the mesolithic stone tools data from southern England outlined in Section 1.2. In Figs 1(a) and 1(b) we plot the most likely orders of the sites using the RK modelling approach and CA respectively. The horizontal axis gives the most likely order of the six sites and the vertical axis plots the proportion of each artefact type. Note that in Fig. 1(b) artefact types 4 and 7 violate the deterministic RK model as the strict unimodal behaviour (in seriated archaeological time) is violated. Clearly, the two interpretive strategies produce results with potentially quite different archaeological conclusions.

Models based on CA have the independence model (independence between the artefacts and the sites) as their starting-point (this is not the case for the RK model). Archaeological data used in seriation studies often contain many zero counts, with the consequence that the independence model should not be expected to provide a good fit. None-the-less, since CA is popular as a descriptive tool for archaeological seriation we feel that a model-based implementation of this approach is appropriate if we are to compare previously popular methodologies. At present there is, as far as we are aware, no formal statistical methodology for comparing the CA and RK models for a given data set. As a consequence, we feel that this paper would not be complete without Section 4 in which we develop a Bayesian predictive model choice approach to comparing the two sets of models outlined above.

The remainder of this paper is organized as follows. In Section 2 we detail our extensions to the Bayesian RK model. In Section 3 we develop hierarchical Bayesian models for CA and

describe appropriate MCMC implementation methods. Section 4 is devoted to addressing Bayesian model choice methods for selecting between the models presented in the preceding two sections. Section 5 is used to report the results obtained for our example data sets and Section 6 provides some concluding remarks.

2. The Robinson–Kendall model

Kendall (1971) and other more recent researchers have worked with deterministic models for seriation. Stochastic Bayesian versions of this model have also been considered; see for example Buck and Litton (1991), Halekoh and Vach (1999) and references therein. Here we present extensions to the Buck and Litton (1991) modelling approach.

Let n_{ij} denote the observed number of artefacts (e.g. pottery or tools types) of type j ($j = 1, \ldots, J$) found in archaeological site, feature or context i ($i = 1, \ldots, I$). Also, let θ_{ij} denote the underlying proportion of artefact j available for deposition at i and let Θ denote the matrix with elements θ_{ij}. Since the θ_{ij} are proportions it is implicitly assumed that $\theta_{ij} \geqslant 0$ and

$$\sum_i \sum_j \theta_{ij} = 1$$

(for more discussion of this assumption see expression (3) and the preceding paragraph). The problem, then, is to estimate the true temporal order of the I rows which is a permutation of the indices $1, \ldots, I$. We represent this permutation using $p(1), p(2), \ldots, p(I)$. In theory, there are $I!$ temporal orders. In practice, however, since the models used for seriation do not contain information to allow a distinction to be made between the start of a sequence and its end (we rely on the skill of archaeologists to make this assessment), only $I!/2$ of the possible orders need to be considered.

Assume that the true chronological order is the given natural order, i.e. $p(1) = 1, p(2) = 2, \ldots, p(I) = I$. Then, for each j, the RK model assumes that there are integers $1 \leqslant a_j \leqslant I$ such that

$$\begin{aligned} \theta_{ij} &\leqslant \theta_{i+1j} \qquad \text{for } i = 1, \ldots, a_j - 1, \\ \theta_{i+1j} &\leqslant \theta_{ij} \qquad \text{for } i = a_j, \ldots, I - 1. \end{aligned} \tag{1}$$

When a_j is either 1 or I only one set of inequalities in the above equations is required and the other set is redundant. In the archaeology literature a matrix Θ satisfying assumption (1) is called a Q-matrix (for theoretical work on such matrices see, for example, Kendall (1971) and Laxton (1976)). In practice the true chronological order is unknown and we attempt to find an order $p(1), p(2), \ldots, p(I)$ such that Θ is a Q-matrix for a set of unknown integers $a_j, j = 1, \ldots, J$.

As mentioned previously, model (1) is overly prescriptive for most real archaeological data. The first issue that we address in tackling this is that the strict, temporal, unimodal sequence assumed in the RK model may be violated because of the nature of use and discard of objects in the past. To account for this type of violation we develop the following extension. Suppose that the matrix Φ is a Q-matrix in the natural order and let $\|\cdot\|$ denote a suitable distance measure between two matrices Θ and Φ. For example, we may consider the Kullback–Leibler distance

$$\|\Theta - \Phi\| = \sum_{ij} \theta_{ij} \log(\theta_{ij}/\phi_{ij}) \tag{2}$$

or the Euclidean distance

$$\|\Theta - \Phi\| = \sqrt{\left\{\sum_{ij} (\theta_{ij} - \phi_{ij})^2\right\}}.$$

In the remainder of this paper we illustrate our approach by adopting distance (2), although it should be clear that any suitable measure could be used. Our extended model is then that, for prespecified $\epsilon > 0$, we have a matrix Θ which also satisfies the extended RK model in the natural order if $\|\Theta - \Phi\| \leqslant \epsilon$. It is clear that when ϵ is chosen to be 0 the extended model reduces to model (1). In this sense the parameter ϵ dictates how much relaxation we want to allow our models to have over the basic RK model. A large value of ϵ will produce all the $I!/2$ possible permutations for plausible seriation of the data. In contrast smaller values will typically produce only a few of the possible permutations of the rows for seriation. See our discussion of the results in Section 5 for a further explanation of the role played by ϵ.

The above models are deterministic. In other words, for a given ϵ a matrix Θ can be reordered to give a Q-matrix with probability 0 or 1. So, the second modelling issue that must be addressed is to account for random variation in the data, for example due to poor recovery of some or all artefact types. Hence, stochastic models relating the data matrix (n_{ij}) and the parameters θ_{ij} must be formulated.

Let $n_{i+} = \Sigma_{j=1}^{J} n_{ij}$ and similarly, in general, a subscript is replaced by a plus sign to denote the sum over that subscript. Let $N = n_{++}$ be the total number of artefacts in the present study and let $\mathbf{n}_i = (n_{i1}, \ldots, n_{iJ})$ and $\mathbf{n} = (\mathbf{n}_1, \mathbf{n}_2, \ldots, \mathbf{n}_I)$. Similarly let $\boldsymbol{\theta}_i = (\theta_{i1}, \ldots, \theta_{iJ})$ and $\boldsymbol{\theta} = (\boldsymbol{\theta}_1, \ldots, \boldsymbol{\theta}_I)$, i.e. $\boldsymbol{\theta}$ is the vector representation of the matrix Θ.

When the rows represent I sites that are geographically far apart, it is reasonable to assume that $\mathbf{n}_i$ has an independent multinomial distribution with parameters n_{i+} and $\boldsymbol{\theta}_i$; this is the approach adopted by Buck and Litton (1991). (Note that this assumes $\Sigma_j \theta_{ij} = 1$ for each i.) However, there are many situations in which the independence assumption may not be appropriate. For example, if the features to be ordered are graves in the same graveyard, or the sites to be ordered are close with known cultural interactions between inhabitants.

Henceforth, we develop and work with the following alternative multinomial model to address these issues. We assume that $\mathbf{n}$ has a multinomial distribution with parameters N and $\boldsymbol{\theta}$, i.e. the probability of obtaining the observed configuration, n_{ij}, is given by

$$N! \prod_{ij} \theta_{ij}^{n_{ij}} / n_{ij}!. \tag{3}$$

(Here we assume that $\Sigma_i \Sigma_j \theta_{ij} = 1$.) This probability model does not incorporate any underlying assumptions except that N is given. As a consequence this could be viewed as a nonparametric statistical model because we are not assuming any probability distribution for the prevalence and/or characteristics of each artefact type. This is in contrast with parametric approaches in which such assumptions are made (see for example Scott (1993)).

A suitable way to represent prior information about θ_{ij} is to use the Dirichlet distribution; thus

$$\pi(\boldsymbol{\theta}) \propto \prod_{ij} \theta_{ij}^{\alpha_{ij}-1}, \tag{4}$$

where $\alpha_{ij} > 0$ for all i and j. We can use this to incorporate both informative and non-informative prior information quite successfully. For example, setting $\alpha_{11} = 5$ can be thought

of as having four artefacts of type 1 in row 1. To specify non-informative prior information we simply set $\alpha_{ij} = 0.5$ for all i and j.

Given the assumptions above, the posterior distribution of $\boldsymbol{\theta}$ follows a Dirichlet distribution; thus

$$\pi(\boldsymbol{\theta}|\mathbf{n}) \propto \prod_{ij} \theta_{ij}^{n_{ij}+\alpha_{ij}-1}. \tag{5}$$

However, this Bayesian model does not directly estimate the unknown order $p(1)$, $p(2)$, . . ., $p(I)$; nor does it incorporate model (1)–(2). We adopt the following *non-Markovian* sampling-based approach towards solving this problem. We simulate $\boldsymbol{\theta}$ from the Dirichlet distribution (5). For this simulated $\boldsymbol{\theta}$ we first identify the a_js in model (1). Then for each artefact type j we arrange the θ_{ij}, $i = 1, \ldots, a_j - 1$, in ascending order and the θ_{ij}, $i = a_j + 1, \ldots, I$, in descending order. The resulting matrix is taken as the matrix Φ for the extended model (2). The distance (2) is then calculated and a conclusion is reached about whether or not the simulated $\boldsymbol{\theta}$ satisfies the extended model for the given order. This process is repeated for all possible permutations of the rows ($I!/2$) of the simulated $\boldsymbol{\theta}$. Finally, the simulations are repeated a large number of times (5000 for the results reported in Section 5) to produce all possible seriations of the data on the basis of the extended RK model.

3. Correspondence analysis

3.1. Hierarchical Bayesian models for seriation

As stated earlier, CA is viewed as an alternative to adopting the RK model for seriation (see for example Baxter (1994), chapter 5, and Goodman (1986)) but it has usually been used only in an exploratory fashion in archaeology. Here we adopt a model-based approach using hierarchical Bayesian models.

In the first stage of model building we assume that $\mathbf{n}$ has a multinomial distribution with parameters N and $\boldsymbol{\theta}$, i.e. the likelihood of the data is as given in expression (3). Let $1 \leqslant M \leqslant \min(I-1, J-1)$ be a positive integer. We then assume the following model for θ_{ij}:

$$\theta_{ij} = \theta_{i+}\theta_{+j}\left(1 + \sum_{k=1}^{M} \lambda_k x_{ik} y_{jk}\right), \tag{6}$$

where $0 \leqslant \lambda_M \leqslant \ldots \leqslant \lambda_1 \leqslant 1$ and

$$\sum_{i=1}^{I} x_{ik}\theta_{i+} = 0, \qquad \sum_{i=1}^{I} x_{ik}^2\theta_{i+} = 1, \qquad \sum_{i=1}^{I} x_{ik}x_{ik'}\theta_{i+} = 0, \qquad k \neq k' = 1, \ldots, M, \tag{7}$$

and

$$\sum_{j=1}^{J} y_{jk}\theta_{+j} = 0, \qquad \sum_{j=1}^{J} y_{jk}^2\theta_{+j} = 1, \qquad \sum_{j=1}^{J} y_{jk}y_{jk'}\theta_{+j} = 0, \qquad k \neq k' = 1, \ldots, M. \tag{8}$$

The parametric constraints in equations (7) and (8) orthogonalize and normalize the row and column scores x_{ik} and y_{jk}. The λ-parameters are called the canonical correlations and these correspond to the eigenvalues (with the score vectors as the eigenvectors) for the χ^2 distance matrix between the observed and the fitted cell counts in the contingency table.

The independence model ($\theta_{ij} = \theta_{i+}\theta_{+j}$) is obtained if these scores are all assumed to be 0. The chronological order produced by the CA is usually taken as the ordering of the score

vector $x_{11}, x_{21}, \ldots, x_{I1}$. It is often argued that row scores that are close together represent rows in which conditional distributions across the columns are similar. In addition, a two-dimensional plot of the vector $\mathbf{x}_{i1}$ against $\mathbf{x}_{i2}$ gives a visual representation of the relative closeness of the rows. To avoid misinterpretation, we note here that this 'closeness' is only an indication of relative (and not absolute) chronology and as such the relative distance between two seriated x_{ik}s does not relate to the length of time elapsed between archaeological depositions.

It is clear that in equation (6) a saturated model is obtained by taking $M = \min(I - 1, J - 1)$, since in this case the number of parameters is equal to the number of IJ-observations n_{ij}. For positive values of M, less than the above maximum, unsaturated models result; see for example Goodman (1986). Our model choice methodology, described in the next section, can be used to decide which model to adopt.

In practice, the first canonical correlation λ_1 can explain most of the variation (according to the Pearson χ^2-norm) between the fitted and the observed cell counts. In addition, the CA model (6) with $M = 1$ is more easily interpreted both in terms of the model parameters and its utility for seriation. With $M = 1$ model (6) reduces to the *canonical correlation model*

$$\theta_{ij} = \theta_{i+}\theta_{+j}(1 + \lambda x_i y_j), \tag{9}$$

where $0 < \lambda < 1$ is an unknown parameter, and

$$\sum_{i=1}^{I} x_i\theta_{i+} = 0, \qquad \sum_{i=1}^{I} x_i^2\theta_{i+} = 1, \qquad \sum_{j=1}^{J} y_j\theta_{+j} = 0, \qquad \sum_{j=1}^{J} y_j^2\theta_{+j} = 1.$$

Consider the situation in which the θ_{i+} are all equal. This represents equal proportions of all artefacts in each row. Now, consider model (9). For a particular artefact j, when $y_j > 0$ the proportion of artefact j in each row (θ_{ij}) follows the same pattern as the x_i (i.e. the proportions of artefacts increase and decrease with the row scores). Conversely, if $y_j < 0$, the proportions in the jth column have a pattern that is the opposite of the x_i (i.e. the proportions increase as the row scores decrease and decrease as the row scores increase). This is why the order of the x_i is held to indicate the seriated order for the rows of the data under study.

In the second stage of our model building, we take a hierarchical Bayesian approach by specifying prior distributions for all the parameters in model (6). The two marginal vectors of parameters $\boldsymbol{\theta}_I = (\theta_{1+}, \ldots, \theta_{I+})$ and $\boldsymbol{\theta}_J = (\theta_{+1}, \ldots, \theta_{+J})$ are assigned independent Dirichlet prior distributions

$$\pi(\boldsymbol{\theta}_I) \propto \prod_i \theta_i^{c_i - 1}, \qquad c_i > 0,$$

$$\pi(\boldsymbol{\theta}_J) \propto \prod_j \theta_j^{d_j - 1}, \qquad d_j > 0.$$

The constants c_i and d_j have the same interpretations as the α_{ij} in equation (4). The score parameters x_{ik} and y_{jk} are given normal prior distributions

$$\pi(x_{ik}) \propto \exp\left(-\frac{1}{2\sigma^2}x_{ik}^2\right),$$

$$\pi(y_{jk}) \propto \exp\left(-\frac{1}{2\sigma^2}y_{jk}^2\right),$$

for all i, j and k, where σ^2 is the assumed *a priori* variance of the scores. The correlation parameter λ_k is assigned a vague prior in the unit interval. Note that the constraints for different parameters (given in equations (7) and (8)) are not imposed *a priori*. These are imposed in the model through the likelihood contributions.

An advantage of the above hierarchical Bayesian models is the ability to incorporate informative prior distributions should we so wish. The most likely type of prior information would relate to the relative chronological relationships between two (or more) sites, features or contexts in the study. For example, if some stratigraphic information is available we might be certain that graves 1 and 2 were deposited in this chronological order. In such a situation, the constraint $x_{1k} < x_{2k}$ for each $k = 1, \ldots, M$ can be imposed as prior information.

3.2. Computations

The full posterior distribution for the hierarchical Bayesian model is given by

$$\pi(\boldsymbol{\theta}_I, \boldsymbol{\theta}_J, \mathbf{x}, \mathbf{y}, \boldsymbol{\lambda}|\mathbf{n}) \propto \prod_{ij} \theta_{ij}^{n_{ij}} \prod_i \theta_i^{c_i - 1} \prod_j \theta_j^{d_j - 1} \prod_{ik} \exp\left(-\frac{x_{ik}^2}{2\sigma^2} \right) \prod_{jk} \exp\left(-\frac{y_{jk}^2}{2\sigma^2} \right),$$

where θ_{ij} is as in model (6) and subject to the constraints in equations (7) and (8). The Gibbs sampling implementation for this model is rather challenging. The primary difficulty lies in having to impose the constraints of the model. The single-component updating Gibbs sampler is problematic because (owing to the constraints) one-dimensional conditional densities of the score parameters are degenerate. We have also used Metropolis algorithms without much success. The difficulty lies in tuning the scaling parameters of the algorithms. The joint distribution of any score vector ($\mathbf{x}$ or $\mathbf{y}$) given other parameters lies in a very constrained and narrow region and it is extremely difficult for the Metropolis algorithm to reach such constrained regions of high probability without using further information about the target distribution.

As a result, we adopt the hybrid Monte Carlo updating method based on Langevin diffusion. This algorithm (see for example Roberts and Rosenthal (1998) and references therein) incorporates information about the d-dimensional target density, $\pi(u)$ say, by using the gradient vector $\nabla(u) = \partial[\log\{\pi(u)\}]/\partial u$ for $\log\{\pi(u)\}$. The algorithm works by augmenting the target vector u with a d-dimensional standard normal random variable z (often called the 'momentum' vector). The current Markov chain iterate, $U^{(t)}$ for $t \geqslant 0$, is updated to $U^{(t+1)}$ by using a Metropolis step as follows. Let $\delta > 0$ be a known constant and I_d denote the identity matrix of order d.

(a) Simulate $Z \sim N\{(\delta/2)\nabla(U^{(t)}), I_d\}$.
(b) Let $U_* = U^{(t)} + \delta Z$ and $Z_* = Z + (\delta/2)\nabla(U_*)$.
(c) Move from the current value of $(U^{(t)}, Z)$ to $(U^{(t+1)}, Z_{\text{new}})$ where

$$(U^{(t+1)}, Z_{\text{new}}) = \begin{cases} (U_*, Z_*) & \text{with probability } \alpha, \\ (U^{(t)}, Z) & \text{with probability } 1 - \alpha \end{cases}$$

where

$$\alpha = \min\left[1, \frac{\pi(U_*)}{\pi(U^{(t)})} \exp\left\{ -\frac{1}{2}(Z_*'Z_* - Z'Z) \right\}\right].$$

This algorithm has been shown, by Roberts and Rosenthal (1998), to be more efficient than

the Metropolis algorithm. They also showed that, under certain conditions, 0.574 is the optimal acceptance rate which maximizes the efficiency of the algorithm. Note that the acceptance rate depends on the scale δ, the only parameter which is to be supplied by the user.

For our problem, we run a Gibbs sampler on the full parameter vector. Since it is difficult to update the normalized score vectors we decided to work with non-normalized versions of those. The Gibbs sampler is run on the non-normalized scores and other remaining parameters. Hence the usual geometric convergence is guaranteed since the joint posterior density is constructed to be proper. At each iteration we also obtain the normalized scores by a Gram–Schmidt orthogonalization scheme; see for example Rao (1973), page 9. One full Gibbs cycle is as follows.

Step I—update the non-normalized x-score parameters $(x'_{1k}, \ldots, x'_{Ik})$ for each $k = 1, \ldots, M$, by using the above Langevin step.
Step II—simulate $\boldsymbol{\theta}_I$ from its conditional posterior distribution which is a Dirichlet distribution with appropriate parameters ignoring the constraints in equations (7).
Step III—perform Gram–Schmidt orthogonalization of the non-normalized scores x'_{ik} to obtain x_{ik} which satisfy constraints (7) with the simulated value of $\boldsymbol{\theta}_I$.
Step IV—repeat the above three steps for the y-scores and $\boldsymbol{\theta}_J$.
Step V—simulate the λs from their conditional distributions one after another by using the adaptive rejection Metropolis sampling of Gilks *et al.* (1995).
Step VI—relabel the scores and the λs so that the constraint $0 \leqslant \lambda_M \leqslant \ldots \leqslant \lambda_1 \leqslant 1$ is observed as a result.

The proposal variances for updating the score parameters are tuned to have an acceptance rate of approximately 0.574. In each case Gibbs sampling was run for 12000 iterations with quite arbitrary starting values. Block updating of Θ_I, Θ_J, $\mathbf{x}$ and $\mathbf{y}$ helps convergence of the Gibbs sampler. The first 2000 iterations were discarded before forming the ergodic averages for estimation and model choice. In all our implementations we took $\sigma^2 = 10^4$.

To address the problem of assessing MCMC convergence we have monitored and analysed many interesting characteristics of the posterior distributions. In particular we obtained autocorrelation plots of the typical sample paths of various parameters and those did not show any problem in convergence. For illustrative purposes we provide plots of the sample paths of θ_{11} and θ_{33} from our first example (the Awatovi data set) for the CA model with $M = 1$ (Fig. 2).

4. Model selection

In the previous two sections we have attempted to model the parameter θ_{ij} on the basis of the contingency table n_{ij}. The independence model $(\theta_{ij} = \theta_{i+} \times \theta_{+j})$ is often viewed as the starting model for analysing contingency tables. Clearly, the independence model is not a nested submodel of the RK model (1). In contrast, the hierarchical CA models of Section 3 do admit the independence model as one of their special cases. This in itself does not, necessarily, show that one model is superior to the other; rather it indicates that it is not easy to make intuitive comparisons between the two sets of models. For these reasons, among others, it is not a surprise that there is as yet no satisfactory method for deciding which approach to archaeological seriation is most appropriate for a given data set. Hence we feel that further statistical investigation is necessary to allow a formal comparison between the two methods.

Predictive Bayesian model choice techniques can be used to facilitate model comparisons. A pure Bayesian solution for model comparison is to consider the Bayes factor; see for

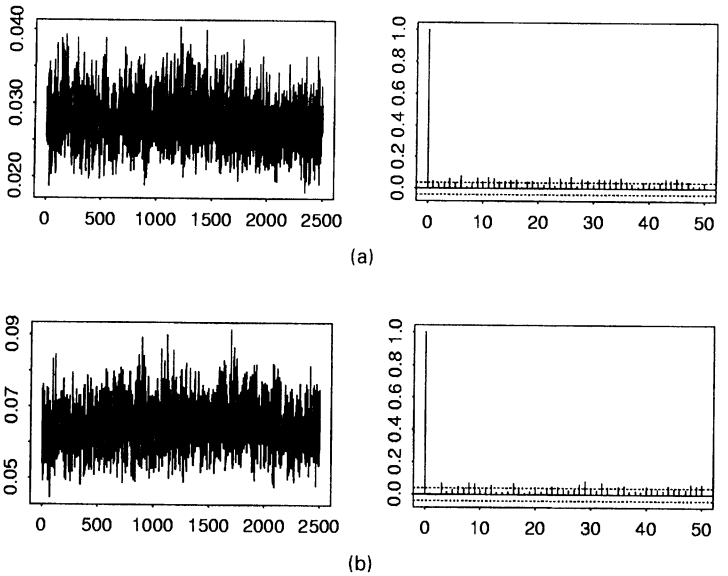

Fig. 2. Time series and autocorrelation function plots of (a) θ_{11} and (b) θ_{33} from our analysis of the Awatovi mound data discussed in Section 5 (for the CA model with $M = 1$)

example Key *et al.* (1999) and Bernardo and Smith (1994). Although there have been recent advances in computing Bayes factors (see for example Raftery (1996) for a review), there are still problems in calculating them for high dimensional models such as those advocated here. Also for improper priors the Bayes factor is not meaningful since it cannot be calibrated.

Methods based on features of the posterior distribution of the likelihood added to a penalty factor are also available for model selection. For example, Aitkin (1997) interpreted the p-values by using the posterior distribution of the likelihood function. Spiegelhalter *et al.* (1998) proposed a model selection criterion for arbitrarily complex models called the deviance information criterion. They estimated the effective number of parameters for such models by using quantities similar to the leverages in linear models. The penalty factor, which is the expected deviance minus the deviance evaluated at the posterior expectations, is calculated and added to the posterior expectation of the deviance to form the deviance information criterion.

Although these methods have value and are useful in practical cases, in this paper we adopt a decision theoretic approach to model selection based on loss functions. Some general notation is introduced here to develop this approach further. Let $\pi(\cdot)$ denote the density of its argument and $\mathbf{n}_{\text{obs}}$ denote the observed data with individual data points $n_{ij,\text{obs}}$, $i = 1, \ldots, I$, $j = 1, \ldots, J$. Similarly, let $\mathbf{n}_{\text{rep}}$ with components $n_{ij,\text{rep}}$ (where rep is an abbreviation for replicate) denote a future set of observables under the assumed model.

The current model is a 'good' fit to the observed data $\mathbf{n}_{\text{obs}}$ if $\mathbf{n}_{\text{rep}}$ can replicate the data well. Hence, many model choice criteria can be developed by considering different loss functions for measuring the divergence between $\mathbf{n}_{\text{obs}}$ and $\mathbf{n}_{\text{rep}}$ (see for example Rubin (1984)). In particular, we consider the following loss function between the two:

$$L(\mathbf{n}_{\text{rep}}, \mathbf{n}_{\text{obs}}) = 2 \sum_{ij} n_{ij,\text{obs}} \log\left(\frac{n_{ij,\text{obs}}}{n_{ij,\text{rep}}}\right). \tag{10}$$

Since equation (10) is an entropy-like divergence measure between $\mathbf{n}_{\text{obs}}/N$ and $\mathbf{n}_{\text{rep}}/N$, it is

likely to yield high values if the predicted data $\mathbf{n}_{\text{rep}}$ are not close to the observed data $\mathbf{n}_{\text{obs}}$. Furthermore, the ijth term in the summation is strictly convex in $n_{ij,\text{rep}}$ if $n_{ij,\text{obs}}$ is positive. We can avoid difficulties with the zero counts by removing the corresponding terms from the sum in equation (10) or by adding $\frac{1}{2}$ to every cell as is often done in practice; see for example Waller *et al.* (1997). The above loss function is sometimes called the deviance loss function.

The best model among a given set of models is the model for which the expected value of the above loss function is the minimum, where the expectation is to be taken with respect to a suitable predictive distribution of $\mathbf{n}_{\text{rep}}$. The *posterior predictive density* of $\mathbf{n}_{\text{rep}}$, given by

$$\pi(\mathbf{n}_{\text{rep}}|\mathbf{n}_{\text{obs}}) = \int \pi(\mathbf{n}_{\text{rep}}|\boldsymbol{\theta})\, \pi(\boldsymbol{\theta}|\mathbf{n}_{\text{obs}})\, \mathrm{d}\boldsymbol{\theta}, \tag{11}$$

is the predictive density of a new independent set of observables, $\mathbf{n}_{\text{rep}}$ under the model, given the actual data $\mathbf{n}_{\text{obs}}$. The posterior predictive density is easier to work with than other similar densities, because features of $\mathbf{n}_{\text{rep}}$ having density (11) can be estimated easily when MCMC samples from the posterior $\pi(\boldsymbol{\theta}|\mathbf{n}_{\text{obs}})$ are available.

The expected value of loss function (10) with respect to the predictive distribution (11) is the proposed model selection criterion. This has attractive interpretations in terms of the classical likelihood ratio test statistic for comparing two models. Let the fitted probabilities (based on the maximum likelihood estimate) for a full and a reduced model be denoted by $\hat{\boldsymbol{\theta}}$ and $\tilde{\boldsymbol{\theta}}$ respectively. The likelihood ratio statistic (often called the scaled deviance) for comparing the two models is given by

$$d(\hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) = 2 \sum_{ij} n_{ij,\text{obs}} \{\log(\hat{\theta}_{ij}) - \log(\tilde{\theta}_{ij})\}.$$

If a so-called saturated model is taken as the full model then $\hat{\theta}_{ij} = n_{ij,\text{obs}}/N$. Now the above statistic reduces to

$$d(\hat{\boldsymbol{\theta}}, \tilde{\boldsymbol{\theta}}) = 2 \sum_{ij} n_{ij,\text{obs}} \log\left(\frac{n_{ij,\text{obs}}}{N\tilde{\theta}_{ij}}\right). \tag{12}$$

Let $\theta^*_{ij} = E(n_{ij,\text{rep}}/N)$ where the expectation is taken with respect to the predictive distribution (11). Now we have

$$\begin{aligned} E\{L(\mathbf{n}_{\text{rep}}, \mathbf{n}_{\text{obs}})\} &= 2 \sum_{ij} n_{ij,\text{obs}} [\log(n_{ij,\text{obs}}/N) - E\{\log(n_{ij,\text{rep}}/N)\}] \\ &= 2 \sum_{ij} n_{ij,\text{obs}} [\log(n_{ij,\text{obs}}/N) - \log(\theta^*_{ij}) + \log(\theta^*_{ij}) - E\{\log(n_{ij,\text{rep}}/N)\}] \\ &= \text{LRS} + 2 \sum_{ij} n_{ij,\text{obs}} [\log(\theta^*_{ij}) - E\{\log(n_{ij,\text{rep}}/N)\}] \end{aligned}$$

where LRS is the likelihood ratio statistic (12) with the maximum likelihood estimate $\tilde{\boldsymbol{\theta}}$ replaced by $\boldsymbol{\theta}^*$. The LRS provides a goodness-of-fit measure as can be seen from its connection with the likelihood ratio test. Usually a more complex model provides a better fit and hence the LRS should go down when a more complex but nested model is fitted to the data set.

Using a second-order Taylor series expansion for the log-function we see that the second part of the above expression can be approximated by

$$\sum_{ij} \text{var}(n_{ij,\text{rep}}/N) = \sum_{ij} \theta^*_{ij}(1 - \theta^*_{ij})/N.$$

This is a penalty term PEN and is likely to be high if the fitted model is too large for the data

set. In other words, this takes care of uncertainty in estimation since the variability of $\mathbf{n}_{\text{rep}}$ is likely to be higher if a more complex model is fitted. This is because, intuitively, the parameters are less clearly identified and, hence, more poorly estimated (i.e. the variability is higher) under a more complex model.

In practice we do not use the above approximation to obtain the penalty term. Instead we calculate the expected loss and LRS directly and the penalty is obtained by subtraction. Henceforth, we use the following notation and decomposition for the expected loss function

$$\text{EPD} \equiv E\{L(\mathbf{n}_{\text{rep}}, \mathbf{n}_{\text{obs}})\} = \text{LRS} + \text{PEN}, \tag{13}$$

where PEN is obtained by subtraction. The notation EPD stands for the expected predictive deviance.

Note the conflicting behaviour of the two components in equation (13). As a more complex model is fitted LRS should go down, whereas the penalty should go up. Hence when fitting a sequence of more complex and nested models (for example as described in Section 3) a trade-off must arise. At some intermediate model the increase in PEN will not be offset by the decrease in LRS. The model choice criterion (13) chooses this model as the best model for the data set.

Criterion (13) is estimated by using Monte Carlo integration as follows. The criterion is the expected value of the loss function (10) under the posterior predictive distribution (11). At each MCMC iteration we obtain the fitted probabilities $\boldsymbol{\theta}$ using the assumed model. Then we obtain a new multinomial observation $\mathbf{n}_{\text{rep}}$ and evaluate loss function (10). The average of these evaluations at the end of the MCMC run is an estimate of EPD. The fitted probabilities at each iteration are also averaged after the MCMC run to obtain an estimate of $\boldsymbol{\theta}^*$. These probabilities are then put back in equation (12) to obtain LRS. PEN is obtained by subtraction.

5. Results

5.1. Example 1: the refuse mounds at Awatovi, Arizona

The orders produced by the RK model and the hierarchical canonical correlation model

Table 3. Orders obtained from the extended RK model with $\epsilon = 10^{-4}$ for the Awatovi data

Order of layers	*Percentage*
d e f g h i j	84.9
d e f h g i j	5.2
e d f g h i j	4.9
d f e g h i j	3.4
Remaining 2516 orders	1.6

Table 4. Orders obtained by using the hierarchical canonical correlation models with $M = 1$ for the Awatovi data

Order of layers	*Percentage*
d e f g h i j	70.7
e d f g h i j	11.1
d e f h g i j	10.1
d f e g h i j	2.3
Remaining 2516 orders	5.8

Table 5. Model choice for the Awatovi data

Model	*LRS*	*PEN*	*Expected loss*
RK with $\epsilon = 10^{-2}$	7.9	46.1	54.0
RK with $\epsilon = 10^{-3}$	7.0	44.7	51.7
RK with $\epsilon = 10^{-4}$	5.9	43.6	49.5
Independence	446.6	30.2	476.8
CA with $M = 1$	92.0	31.6	123.6
CA with $M = 2$	61.4	39.5	100.9

when applied to the data from the refuse mounds at Awatovi are given in Tables 3 and 4. Note that the two models produce the same dominant order and that this is known, on the basis of the stratigraphic evidence obtained during excavation, to be the true chronological order. Hence there is no conflict of orders to be resolved here.

However, we are still interested in which model most adequately represents the data and so the model choice criteria are given in Table 5. From the expected loss estimates, in the fourth column of Table 5, we see immediately that the Awatovi data set is most appropriately modelled by using the RK model rather than by the CA model. Of the results reported, it is the RK model with $\epsilon = 10^{-4}$ that best fits the data. Note, in fact, that in the order (d, e, f, g, h, i, j) the raw data matrix is already a Q-matrix and thus when $\epsilon = 0$ the same order is dominant. By taking ϵ somewhat larger than 0, however, we observe a wider range of possible orders that archaeologists might wish to consider. Although, for this particular data set, the use of non-zero values of ϵ in this way is simply for illustration, we hope that its potential utility when working with other data is clear (this will be discussed again later when we look at the results from the stone tools data).

We also wish to note that, on the basis of the expected loss, the CA models can be seen as being rather inadequate for representing the processes that gave rise to the patterns of pottery deposition observed in the refuse mounds at Awatovi. The $M = 1$ and $M = 2$ CA models provide worse fits because they are based on the basic independence model. The Awatovi data set has seven (out of 35) cell counts which are 0 and the independence model is not good for data sets with such high zero cell counts, as can be seen from the value of the model choice criteria in Table 5. Given that LRS for the independence model is 446.6, observe that even the CA model with $M = 1$ offers quite a marked improvement in model fit.

Table 6. Orders obtained from the extended RK model (with $\epsilon = 10^{-4}$) for the stone tools data

Order of sites	*Percentage*
2 5 3 6 1 4	71.5
1 4 6 3 5 2	19.6
2 5 6 3 1 4	3.7
4 1 6 3 2 5	2.9
1 4 3 6 5 2	1.0
1 4 6 3 2 5	0.8
1 4 2 5 6 3	0.2
4 1 3 6 2 5	0.2
3 6 5 2 1 4	0.1
Remaining 351 orders	<0.1

Table 7. Orders obtained by using the hierarchical canonical correlation model (with $M = 1$) for the stone tools data

Order of sites	*Percentage*
3 6 5 2 1 4	35.7
3 6 2 5 1 4	20.0
3 6 2 5 4 1	10.0
3 6 5 2 4 1	6.6
Remaining 356 orders	27.7

5.2. Example 2: stone tools from southern England

In Tables 6 and 7 we give the distribution of the orders for the six stone tool sites using the RK model and the hierarchical canonical correlation model (9) respectively. Clearly the two methods produce different candidate orders for the relative chronology of the six sites. The most likely candidate order on the basis of the extended RK model (2, 5, 3, 6, 1, 4) has an odds ratio of about 5:2. On the basis of the canonical correlation model, however, this order is not high on the list of candidate orders. In fact, the hierarchical canonical correlation model prefers the order (3, 6, 5, 2, 1, 4) with an odds ratio of about 1:2. This order is also a candidate order when the extended RK model is adopted, but under this model it has an odds ratio of just 1:1000. In other words, the order (2, 5, 3, 6, 1, 4) is more emphatically selected than (3, 6, 5, 2, 1, 4).

Before we turn to the results of the Bayesian model choice analysis, it is interesting to compare the relative chronological orders in Tables 6 and 7 with those obtained by Buck *et al.* (1996). In Table 12.4 (page 333) Buck *et al.* (1996) gave the results of an analysis of the same data by using a somewhat different modelling approach (the data in each row, rather than the whole table, are taken as multinomial) and without a parameter equivalent to our ϵ. When ϵ is 0, we obtain results that are close to those of Buck *et al.* (1996), but we do find somewhat different orderings for non-zero values of ϵ. As the value of ϵ is increased, more orders become possible seriations. In particular, note that the order (3, 6, 5, 2, 1, 4) does not appear as a possible order when $\epsilon = 0$ but appears in Table 6. A glance at Fig. 1 reveals that this order does not violate the RK model by very much and that these violations might reasonably be attributed to noise rather than to valuable features of the data. As a result, we feel that ϵ is an important feature of our extended modelling approach.

Turning now to consider the model choice criteria, reported in Table 8, exactly as expected, the independence model provides the worst fit of all to the data. Also, the likelihood ratio statistic for CA with the $M = 2$ model is smaller than with the $M = 1$ model since the $M = 1$ model is a reduced version of $M = 2$. The $M = 2$ model has more parameters than the $M = 1$ model has; hence it receives more penalty. In this way as M increases the penalty will increase

Table 8. Model choice for the stone tools data

Model	*LRS*	*PEN*	*Expected loss*
RK with $\epsilon = 10^{-2}$	14.3	45.2	59.5
RK with $\epsilon = 10^{-3}$	12.7	44.3	57.0
RK with $\epsilon = 10^{-4}$	12.0	43.2	55.2
Independence	685.4	49.3	734.7
CA with $M = 1$	376.0	51.1	427.1
CA with $M = 2$	270.5	75.2	345.7

but LRS will decrease. Finally for $M = 5$ the model becomes saturated and consequently LRS will be close to 0. We do not consider models with $M = 3$ or higher since the dominant orders produced when $M = 2$ and higher are the same as with the $M = 1$ model.

It is not possible to compare the RK model with the hierarchical canonical correlation model by the likelihood ratio statistic because these models are not nested. However, the model choice criterion (the fourth column in Table 8) is comparable since it is obtained as the expected loss function under the predictive distributions of the data arising from the two models. This criterion favours the RK model. Hence we may conclude that the RK model fits the data better than does the CA model with $M = 1$. In other words, the relative order produced by the RK model is more supported by the data than is the order produced by the canonical correlation model. This is also the only order in which the observed proportions form a Q-matrix.

6. Concluding remarks

Over the last 50 years, there has been much published work on the use of seriation to aid in archaeological chronology building. Among this work, one deterministic model has been predominantly used, but problems with noisy data have meant that statistical methods have also been adopted. Until now, there has been little formal assessment of the relative merits of the likely candidates for statistical models and we hope that this paper has addressed this shortcoming.

We have offered some new perspectives on the use of model-based statistical methods for archaeological seriation and provided an approach to selecting between and within groups of models that might be used for archaeological chronology building. We have shown that modern Bayesian model choice techniques can be used to compare two sets of models and to decide which gives a better explanation of the data. The CA models of Section 3 give a worse model fit than the RK models because of high numbers of zero cell counts in our example data sets. However, note that, for the maximal value of M, LRS is 0, i.e. it is possible to improve the fit by choosing a higher value of M. As mentioned in Section 1, however, such saturated models are not useful for statistical inference.

We have also demonstrated that the point estimates of the relative archaeological orders provided by erstwhile non-Bayesian analyses fail to capture other orders which could (given that the data are inherently noisy) be considered appropriate seriations. In addition, it seems that it is not possible to obtain relative odds for the non-Bayesian point estimates of the true unknown order. The Bayesian statistical approaches developed in this paper, however, give rise to posterior probabilities of the likely orders of the sites, features or contexts on which archaeologists can base their informed assessment of the most probable seriation for the data that are currently available.

Although not explicitly illustrated by our examples, we hope that it is clear that there is scope within both types of models for the inclusion of informative prior information and that this can be added with little extra work. It is our belief that, for more complex archaeological problems, prior information about the relative chronological order of at least some of the rows is often available. By utilizing such prior information we would, of course, be able to restrict the number of permutations that need to be investigated and hence to increase the size of data set that can be handled with the same computing resources. In addition, because we have adopted the Bayesian framework for formulating the models, it is possible to use today's posteriors to form priors for tomorrow. This might be seen as an appealing feature for archaeologists who undertake fieldwork over periods of many years and would like to be able to make interim interpretations as work progresses.

Acknowledgements

We are grateful to an associate editor and three referees for their helpful comments on an earlier version of the paper.

References

Aitkin, M. (1997) The calibration of P-values, posterior Bayes factors and the AIC from the posterior distribution of the likelihood. *Statist. Comput.*, **7**, 253–261.

Baxter, M. J. (1994) *Exploratory Multivariate Analysis in Archaeology*. Edinburgh: Edinburgh University Press.

Bernardo, J. M. and Smith, A. F. M. (1994) *Bayesian Theory*. Chichester: Wiley.

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

Buck, C. E. and Litton, C. D. (1991) A computational Bayes approach to some common archaeological problems. In *Computer Applications and Quantitative Methods in Archaeology 1990* (eds K. Lockyear and S. P. Q. Rahtz), pp. 93–99. Oxford: Tempus Reparatum.

Burgh, R. F. (1959) Ceramic profiles in the western mound at Awatovi, northeastern Arizona. *Am. Antiq.*, **25**, 184–202.

Gilks, W. R., Best, N. G. and Tan, K. K. C. (1995) Adaptive rejection Metropolis sampling within Gibbs sampling. *Appl. Statist.*, **44**, 455–472.

Goodman, L. A. (1986) Some useful extensions of the usual correspondence analysis approach and the usual log-linear models approach in the analysis of contingency tables. *Int. Statist. Rev.*, **54**, 243–309.

Halekoh, U. and Vach, W. (1999) Bayesian seriation as a tool in archaeology. In *Archaeology in the Age of the Internet: Proc. 25th Anniv. Conf. Computer Applications and Quantitative Methods in Archaeology, Birmingham, April 1997* (eds L. Dingwall, S. Exon, V. Gaffney, S. Laflin and M. van Leusen). Oxford: Archaeopress.

Jacobi, R. M., Laxton, R. R. and Switsur, V. R. (1980) Seriation and dating of mesolithic sites in southern England. *Rev. Archeom.*, **4**, 165–173.

Kendall, D. G. (1970) A mathematical approach to seriation. *Phil. Trans. R. Soc. Lond.* A, **269**, 125–135.

———(1971) Seriation from abundance matrices. In *Mathematics in the Archaeological and Historical Sciences* (eds D. G. Kendall, F. R. Hodson and P. Tautu), pp. 215–252. Edinburgh: Edinburgh University Press.

Key, J. T., Pericchi, L. R. and Smith, A. F. M. (1999) Bayesian model choice: what and why? In *Bayesian Statistics 6* (eds J. M. Bernardo, J. O. Berger, A. P. Dawid and A. F. M. Smith), pp. 343–370. Oxford: Oxford University Press.

Laxton, R. R. (1976) A measure of pre-Q-ness with applications to archaeology. *J. Arch. Sci.*, **3**, 43–54.

———(1987) Some mathematical problems in seriation. *Acta Applic. Math.*, **10**, 213–235.

Laxton, R. R. and Restorick, J. (1989) Seriation by similarity and consistency. In *Proc. Conf. Computer Applications and Quantitative Methods in Archaeology, York*, pp. 215–225. Oxford: British Archaeological Reports.

Madsen, T. (1988) *Multivariate Archaeology: Numerical Approaches in Scandinavian Archaeology*. Aarhus: Aarhus University Press.

Petrie, W. M. F. (1899) Sequences in prehistoric remains. *J. Anthrop. Inst.*, **29**, 295–301.

Raftery, A. E. (1996) Hypothesis testing and model selection. In *Markov Chain Monte Carlo in Practice* (eds W. R. Gilks, S. Richardson and D. J. Spiegelhalter), pp. 163–187. London: Chapman and Hall.

Rao, C. R. (1973) *Linear Statistical Inference and Its Applications*. New York: Wiley.

Roberts, G. O. and Rosenthal, J. S. (1998) Optimal scaling of discrete approximations to Langevin diffusions. *J. R. Statist. Soc.* B, **60**, 255–268.

Robinson, W. S. (1951) A method for chronologically ordering archaeological deposits. *Am. Antiq.*, **16**, 293–301.

Rubin, D. B. (1984) Bayesianly justifiable and relevant frequency calculations for the applied statistician. *Ann. Statist.*, **12**, 1151–1172.

Scott, A. (1993) A parametric approach to seriation. In *Computing the Past* (eds T. M. J. Andresen and I. Scollar), pp. 317–324. Aarhus: Aarhus University Press.

Spiegelhalter, D. J., Best, N. G. and Carlin, B. P. (1998) Bayesian deviance, the effective number of parameters, and the comparison of arbitrarily complex models. *Technical Report*. Medical Research Council Biostatistics Unit, Cambridge.

Waller, L. A., Carlin, B. P., Xia, H. and Gelfand, A. E. (1997) Hierarchical spatio-temporal mapping of disease rates. *J. Am. Statist. Ass.*, **92**, 607–617.