Radiocarbon dating with temporal order constraints

Geoff Nicholls
Department of Mathematics
nicholls@math.auckland.ac.nz

Martin Jones
Department of Anthropology
mdj@antnov1.auckland.ac.nz

Auckland University, Private Bag 92019,
Auckland, New Zealand

April 14, 1998

A Bayesian method has been proposed for analysing radiocarbon dates. The method takes into account stratigraphic constraints on recovered calendar dates. We find that the non-informative priors in use in the literature apply a bias towards wider date ranges which is not in general supported by substantial prior knowledge. We recommend using a prior which has a uniform marginal date range. We show how such priors are derived from a model of the deposition and observation process. We apply the method to relatively large data sets, examining the effect that various priors have on the reconstructed dates.

KEYWORDS: radiocarbon dating, stratigraphy, order constraint, Bayesian inference, MCMC

1 Introduction

Buck, Kenworthy, Litton and Smith have described a Bayesian method for estimating calendar dates from radiocarbon measurements (see Buck *et al.* 1991 and Buck *et al.* 1992). Stratigraphic constraints may be incorporated in the analysis. The authors give a straightforward inference scheme, based on MCMC simulation from the posterior. Christen 1994 gives an outlier analysis, within the same Bayesian/MCMC framework. A statistical package, OxCal, described in Ramsey 1995, implements some of the methods presented in Buck *et al.* 1992. See Buck *et al.* 1996 and Litton and Buck 1996 for reviews of the field. Note that Broffit 1984 gives a Bayes estimator for ordered parameters, using conjugate priors and a class of likelihoods excluding those treated here.

It was our intention to report a simple application of these methods to a particular data set, using non-informative priors from the families described in Buck *et al.* 1996 (and below in Section 2.1). However, when dates cover an interval of time which is not large compared to the error in the radiocarbon dating process, we find that such priors push the earliest and latest reconstructed dates towards unrepresentatively early and late values respectively. We derive, in Sections 2.2 and 2.3, a family of priors from a model of the deposition and aging processes. Our models turn out to have the property that the marginal prior distribution of the date range is uniform. In Section 3 we illustrate the posterior inference in two case studies. Data sets have been chosen which are sensitive (Case I) and insensitive (Case II) to the information supplied by the priors under consideration.

We suppose a total of K uncalibrated radiocarbon ages, for K distinct objects, are obtained from M strata. Let N_m denote the number of dates obtained in the m'th stratum, running from the topmost stratum where $m = 1$ to the deepest, with $m = M$. Let $\Theta_{m,n}$ be the unknown true date (in years before the present (*years BP*), where the present is the year 1950 - like all *dates* in this paper, $\Theta_{m,n}$ is in fact an *age*) on which object (m, n) ceased exchanging carbon dioxide with its environment. This *event date* must be distinguished from the *context date*, where object (m, n)'s context date is the unknown true date on which object (m, n) was deposited. Clearly, the event date will predate the context date by an interval of time. However, we will further assume that the context and event dates are identical. We assume that all ages are known to lie in some fixed interval $[L, U]$, with $L < U$, of years BP. This interval can be chosen to be arbitrarily large, possibly equal to $[0, \infty)$, if such limits are not a part of the prior information. Thus if Θ is the K component vector $(\Theta_{1,1}, \ldots \Theta_{1,N_1}, \Theta_{2,1}, \ldots \Theta_{2,N_2}, \ldots \Theta_{1,M}, \ldots \Theta_{M,N_M})$, then $\Theta \in [L, U]^K$.

The dating laboratory returns measurements of radiocarbon ages with units "radiocarbon years BP". The mapping $c = \mu(x)$ from an event date x in calendar years BP to a radiocarbon age c is given by a pair of calibration curves $c = \mu_{\mathrm{mrn}}(x)$ and $c = \mu_{\mathrm{ttl}}(x)$, calibrating specimens of marine and terrestrial origin respectively. Sections of these two curves are shown in Figure 1. Notice that the terrestrial curve is not monotone. The uncertainties in the calibration curves are given by standard errors $\sigma^c_{\mathrm{mrn}}(\Theta), \sigma^c_{\mathrm{ttl}}(\Theta)$ depending on the age of the sample. We have interpolated the bidecadal calibrations of Stuiver and Pearson 1993 and Stuiver and Braziunas 1993 using cubic splines, storing the results in a lookup table, constant in a one year interval. Let $y_{m,n}$ denote the n'th observed (uncalibrated) radiocarbon age obtained in layer m. Each datum $y_{m,n}$ is reported from the lab with an estimated standard error $\sigma_{m,n}$. There is an uncertainty associated with calibrating radiocarbon ages measured in the southern hemisphere using curves estimated from northern hemisphere reference material. We return to this issue in Section 3. Let

$$\mu(x, m, n),\ \tilde{\sigma}_{m,n}(x)^2 = \begin{cases} \mu_{\mathrm{ttl}}(x),\ {\sigma_{m,n}}^2 + \sigma^c_{\mathrm{ttl}}(x)^2 & \text{if specimen } (m,n) \text{ is of terrestrial origin} \\ \mu_{\mathrm{mrn}}(x),\ {\sigma_{m,n}}^2 + \sigma^c_{\mathrm{mrn}}(x)^2 & \text{if specimen } (m,n) \text{ is of marine origin} \end{cases}$$

Let $\theta_{m,n}$ represent a trial value for $\Theta_{m,n}$ and let y, σ, $\tilde{\sigma}(\theta)$ and θ respectively denote the K component vectors $(y_{1,1}, \ldots y_{1,N_1}, \ldots y_{1,M}, \ldots y_{M,N_M})$ *etc*, so that $y \in \Re^K$ and $\theta \in [L, U]^K$. The likelihood density function is

$$g(y|\theta) = \mathcal{Z}_L(\theta)^{-1} \exp(-|y - \mu(\theta)|^2 / 2\tilde{\sigma}(\theta)^2),$$

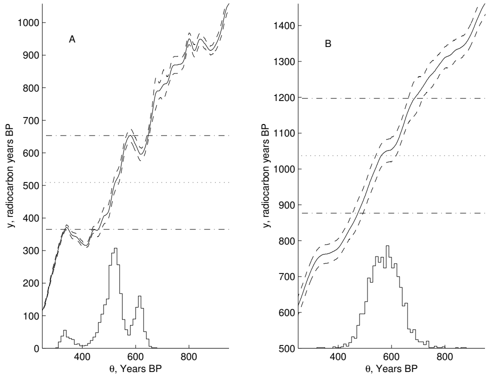

Figure 1: Calibration curves: (A) terrestrial curve, $\mu_{\mathrm{ttl}}(\Theta)$ and (B) marine curve, $\mu_{\mathrm{mrn}}(\Theta)$, from Stuiver and Pearson 1993 and Stuiver and Braziunas 1993. In each graph $\mu(\Theta)$ and $\mu(\Theta) \pm 2\sigma^c(\Theta)$ are represented by solid and dashed lines respectively. Horizontal lines indicate a radiocarbon age $y_{m,n}$ (dotted) with 2σ limits $y_{m,n} \pm 2\sigma_{m,n}$ (dot-dashed). The histogram below is proportional to the likelihood, $g(y_{m,n}|\Theta_{m,n})$.

where

$$-|y-\mu(\theta)|^2/2\tilde{\sigma}(\theta)^2 \equiv -\sum_{m=1}^{M}\sum_{n=1}^{N}(y_{m,n}-\mu(\theta_{m,n},m,n))^2/2\tilde{\sigma}_{m,n}(\theta_{m,n})^2 \tag{1}$$

and $\mathcal{Z}_L(\theta)$ is a straightforward normalising function. We must allow the K measurements $y_{m,n}$ to date up to K distinct context dates $\theta_{m,n}$. If it is known that radiocarbon ages $y_{m,n}$ and $y_{m,n'}$ measure a single context date, there is a straightforward modification of the likelihood.

2 Analysis

2.1 Non-informative priors

Context dates are constrained by their stratigraphic relations. In our observation model no information is recorded on intra-stratum ordering, whilst inter-strata ages are known to be ordered as $\Theta_{m,n} > \Theta_{m',n}$ if $m > m'$ (greater depth indicates greater age). The space of possible true calendar dates is Ω_Θ where

$$\Omega_\Theta = \{\theta : L \leq \theta_{1,1}, \ldots \theta_{1,N_1} < \theta_{2,1}, \ldots \theta_{2,N_2} < \ldots < \theta_{1,M}, \ldots \theta_{M,N_M} \leq U\}. \tag{2}$$

Let $f(\theta)$ be some unnormalised prior density on Ω_Θ. We give the posterior probability distribution $\mathrm{H}^\Theta(d\theta|y,f) \equiv \Pr\{\Theta \in d\theta|y,f\}$ for ages in terms of a density, $\mathrm{H}^\Theta(d\theta|y,f) \equiv h^\Theta(\theta|y,f)d\theta$ where $d\theta \equiv d\theta_{1,1} \times \ldots d\theta_{M,N_M}$,

$$h^\Theta(\theta|y,f) = \mathcal{Z}_P^\Theta(y,f)^{-1} g(y|\theta) f(\theta), \tag{3}$$

and $\mathcal{Z}_P^\Theta(y,f)$ is an intractable normalising function. We wish to recover estimates of the unknown true ages Θ given uncalibrated radiocarbon data (y,σ), and to quantify the range of Θ-values which the data will admit.

Several authors, following Buck *et al.* 1991, use the non-informative prior density $f = f^\Theta(\theta) = 1$ in Equation (3). Since this puts equal weight on all states, the mode of the posterior is equal to the maximum likelihood estimator for θ. Let $\Delta(\Theta) \equiv \max(\Theta) - \min(\Theta)$ denote the unknown time interval between the maximum and minimum of all the true ages and let δ be some trial value for Δ. We have computed the marginal prior density for Δ when $K \geq 2$ and find

$$f_\Delta^\Theta(\delta) \propto (R-\delta)\delta^{K-2}, \tag{4}$$

where $R \equiv U - L$. This result is independent of the way dates are alloted to layers (via N_m values). As the number of carbon dates is increased, the number of random variables in the model (*ie* K) goes up, driving reconstructed dates to wider intervals. Equation (4) is obtained by a straightforward summation of the marginal integrals over all complete orderings of $\Theta_{1,1} \ldots \Theta_{M,N_M}$ consistent with the partial ordering given in Equation (2). Buck *et al.* 1991 compute numerically the marginal prior distributions of the $\theta_{m,n}$ parameters. If K is small as in that paper the resulting marginals are perhaps acceptable, in the sense that they represent a plausible prior state of knowledge. However, it is clear from Equation (4) that, when K is large, this cannot be the case.

Perhaps the most popular family of prior models in the literature, particularly for K large, are the *multi-phase models* of Buck *et al.* 1992. Dated objects are here associated with phases, which may overlap, rather than strata, which do not in general overlap. We introduce variables $\alpha_m, \beta_m \in [L,U], \alpha_m > \beta_m$, with units *years BP*, parameterising the start (α_m) and end (β_m) of phase m. Thus $\alpha_m \geq \Theta_{m,n}$ and $\Theta_{m,n} \geq \beta_m$ for each $n = 1 \ldots N_m$. If it is known that the start of phase m coincides with the end of phase $m+1$, then we set $\alpha_m = \beta_{m+1}$. Let $\mathrm{C}(\alpha,\beta,L,U)$ denote an event, defined as a set of equalities and inequalities between α's, β's and L and U only. The relations $\mathrm{C}(\alpha,\beta,L,U)$ will be those constraints available to us as part of the prior knowledge. $\mathrm{C}(\alpha,\beta,L,U)$ may be represented as a digraph, with a node for each component of α and β, and nodes for L and U. The relations $a > b$ and $a \geq b$ are represented by a directed edge from a's node to b's node. The relation $a = b$ is represented by an undirected edge between nodes a and b. We assume this graph contains no cycles. For example, when phases correspond to stratigraphic layers, the bottom of layer m at α_m *years BP* naturally occurs after the top of layer $m+1$ at β_{m+1} *years BP*, so that $\beta_{m+1} \geq \alpha_m$. Thus, when phases correspond to strata,

$$\mathrm{C}(\alpha,\beta,L,U) = \{L \leq \beta_1 < \alpha_1 \leq \beta_2 < \alpha_2 \leq \ldots \beta_M < \alpha_M \leq U\}, \tag{5}$$

and the corresponding constraint graph is given in Figure 2. Returning to the general case, if A_m

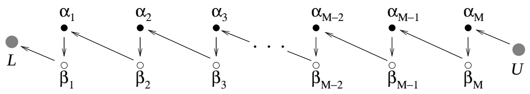

Figure 2: Graph of the constraint relations $\mathrm{C}(\alpha,\beta,L,U)$ for the simple phase model of strata, Equation (5). Nodes correspond to parameters α_m and β_m, and limits, L and U, if they are present. Directed edges $\longrightarrow$ run from a node a to node b if either of the constraints $a > b$ or $a \geq b$ are present.

and B_m denote the unknown true values of α_m and β_m, and $\alpha = (\alpha_1 \ldots \alpha_M)$ *etc*, the full state space of this second model is

$$\Omega_{A,B,\Theta} = \{(\alpha,\beta,\theta) : \mathrm{C}(\alpha,\beta,L,U);\ \theta_{1,1}, \ldots \theta_{1,N_1} \in [\beta_1,\alpha_1], \ldots \theta_{M,1}, \ldots \theta_{M,N_M} \in [\beta_M,\alpha_M]\}.$$

Each variable $\theta_{m,n}$ is taken by Buck *et al.* 1996 to be *a priori* uniform over the interval $[\beta_m,\alpha_m]$, with normalising constant $1/(\alpha_m - \beta_m)$ so the unnormalised prior density is

$$f^{A,B,\Theta}(\alpha,\beta,\theta) = \prod_{m=1}^{M}(\alpha_m - \beta_m)^{-N_m}. \tag{6}$$

Let E denote the number of equalities present in the constraint set $\mathrm{C}(\alpha, \beta, L, U)$. The number of distinct phase variables is then $2M - E$. Now the marginal prior distribution of the variable $\Delta(A, B) = \max(A) - \min(B)$ is

$$f_{\Delta}^{A,B,\Theta}(\delta) \propto (R - \delta)\delta^{2M-E-2}.$$

This result is derived in the same way as Equation (4). Again, the obvious non-informative prior favours widely spread dates. If $\mathrm{H}^{A,B,\Theta}(d\alpha\, d\beta\, d\theta | y, f) \equiv \Pr\{A \in d\alpha,\ B \in d\beta,\ \Theta \in d\theta | y, f\}$ for some density $f(\alpha, \beta, \theta)$, and $\mathrm{H}^{A,B,\Theta}(d\alpha\, d\beta\, d\theta | y, f) \equiv h^{A,B,\Theta}(\alpha, \beta, \theta | y, f) d\alpha\, d\beta\, d\theta$, then the posterior density on $\Omega_{A,B,\Theta}$ is

$$h^{A,B,\Theta}(\alpha, \beta, \theta | y, f) = \mathcal{Z}_P^{A,B,\Theta}(y, f)^{-1} g(y|\theta) f(\alpha, \beta, \theta),$$

with $\mathcal{Z}_P^{A,B,\Theta}(y, f)$ an intractable normalising function.

We wish to simplify the model given above. When we date layered strata, we have no need to distinguish the end of one phase (*ie* layer) from the start of the next. We set $\alpha_m = \beta_{m+1}$ and use a restricted set of parameters ψ_m to represent these boundary transition ages. Let $\Psi = (\Psi_0, \Psi_1, \ldots \Psi_M)$ be a vector containing the unknown true layer transition ages. The state space of (Ψ, Θ) is

$$\begin{aligned} \Omega_{\Psi,\Theta} = \{(\psi, \theta) \quad &: \quad L \le \psi_0 < \psi_1 < \ldots < \psi_M \le U; \\ & \quad \theta_{1,1}, \ldots \theta_{1,N_1} \in [\psi_0, \psi_1], \ldots \theta_{M,1}, \ldots \theta_{M,N_M} \in [\psi_{M-1}, \psi_M]\}. \end{aligned} \tag{7}$$

If ψ is a vector of trial values for the components of Ψ, the prior density is given by Equation (6),

$$f^{\Psi,\Theta}(\theta, \psi) = \prod_{m=1}^{M} (\psi_m - \psi_{m-1})^{-N_m}.$$

If $\mathrm{H}^{\Psi,\Theta}(d\psi\, d\theta | y, f) \equiv \Pr\{\Psi \in d\psi,\ \Theta \in d\theta | y, f\}$ for some density $f(\psi, \theta)$, and $\mathrm{H}^{\Psi,\Theta}(d\psi\, d\theta | y, f) \equiv h^{\Psi,\Theta}(\psi, \theta | y, f) d\psi\, d\theta$, then the posterior density on $\Omega_{\Psi,\Theta}$ is

$$h^{\Psi,\Theta}(\psi, \theta | y, f) = \mathcal{Z}_P^{\Psi,\Theta}(y, f)^{-1} g(y|\theta) f(\psi, \theta),$$

with $\mathcal{Z}_P^{\Psi,\Theta}(y, f)$ an intractable normalising function.

2.2 A model of the deposition process

Up to this point we have been describing models already in use in the literature. We prefer to work with a distinct family of priors, which we derive from a simple model of the deposition process.

We work with the variables Ψ, Θ, taking values in the state space $\Omega_{\Psi,\Theta}$ constrained according to Equation (7). We wish to model the process by which the unknown true Ψ and Θ were formed. Let $\Lambda(t)$ be the intensity of a Poisson point process in time t, $t \in [L, U]$, which we will use to model the layer formation process. We suppose

$$\Lambda(t) = \begin{cases} \Lambda_0 & t \in [\Psi_0, \Psi_M], \\ 0 & t \text{ otherwise}, \end{cases}$$

where Λ_0 is a constant and $[\Psi_0, \Psi_M]$ is the period for which the layer formation process was active. Realisations of this layer process include all distinguishable point sets $(\Psi_1 \ldots \Psi_{M-1})$ with $\Psi_m \in [\Psi_0, \Psi_M]$ and $\Psi_m < \Psi_{m+1}$ for $m = 1 \ldots M-1$.

We model the deposition process as a second Poisson point process, with intensity $\lambda(t)$. Its realisations are sets of context dates, $\tilde{\Theta}$ *years BP*. The setup is illustrated in Figure 3. $\lambda(t)$ represents the

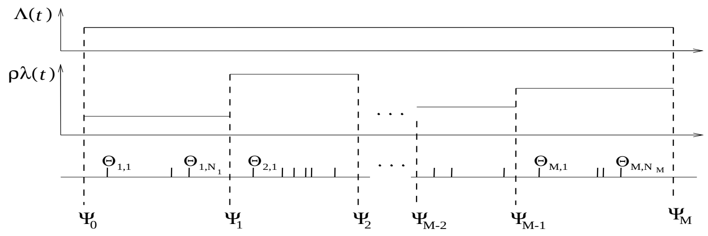

Figure 3: The deposition model is a doubly stochastic process built from a pair of Poisson point processes. Dated material is deposited at a rate $\rho\lambda(t)$. The context dates associated with these dated events are labelled $\Theta_{m,n}$. Change points Ψ_m in the deposition rate $\lambda(t)$ occur at a rate $\Lambda(t)$ which is constant through the active period, and otherwise zero. The process runs backwards in calendar time, however, it is time reversible, so the forward process is equivalent.

rate at which datable items were deposited through the active period. We allow this rate to vary from layer to layer. We suppose

$$\lambda(t) = \begin{cases} \lambda_m & t \in [\Psi_{m-1}, \Psi_m], \text{ for each } m = 1 \ldots M \\ 0 & t \text{ otherwise}, \end{cases}$$

for some constants $\lambda_1, \ldots \lambda_M$. Now, suppose each item of potentially datable deposited material is in fact found and dated with constant probability ρ, and missed with probability $1 - \rho$. Let $\Theta \subset \tilde{\Theta}$ be the unknown true ages belonging to the subset of deposited materials which were in fact dated. The set Θ is a realisation of a Poisson point process with intensity $\rho\lambda(t)$.

We assume that the *number* of layers M is counted exactly, without missing any of the change points ψ_m in the intensity $\lambda(t)$. This allows us to condition on knowledge of M and thereby set aside the unknown intensity $\Lambda(t)$. No such assumption is necessary in respect of the process generating Θ, since its realisations are just the unknown ages of the dated objects. The number N_m of Θ variables in a layer is counted exactly, and so we can condition on knowledge of N_m, and thereby avoid inference about $\rho\lambda(t)$ itself.

Our model leads to the following prior distribution for $(\Psi, \Theta) \in \Omega_{\Psi,\Theta}$,

$$\begin{aligned} \Pr\{\Psi \in d\psi,\ \Theta \in d\theta\} \quad &= \quad \prod_{m=1}^{M} \frac{d\theta}{(\psi_m - \psi_{m-1})^{N_m}} \quad \times \\ & \qquad \frac{(M-1)!}{(\psi_M - \psi_0)^{M-1}} d\psi_1 \ldots d\psi_{M-1} \ \times \ \Pr\{\Psi_0 \in d\psi_0, \Psi_M \in d\psi_M\}. \end{aligned}$$

Let $\mathcal{I}(X)$ denote the indicator function for the event X. We take a uniform prior

$$\Pr\{d\Psi_0 \in d\psi_0, d\Psi_M \in d\psi_M\} \propto \mathcal{I}(L \le \psi_0 < \psi_M \le U) d\psi_0 d\psi_M$$

on the only two remaining unmodeled variables. Now if $\Pr\{\Theta \in d\theta,\ \Psi \in d\psi\} \equiv \hat{f}^{\Psi,\Theta}(\psi, \theta) d\psi\, d\theta$, with $d\psi \equiv d\psi_0 \times \ldots d\psi_M$, our modelling has lead us to the prior density

$$\hat{f}^{\Psi,\Theta}(\psi, \theta) = \frac{1}{(\psi_M - \psi_0)^{M-1}} \times \prod_{m=1}^{M} (\psi_m - \psi_{m-1})^{-N_m}, \tag{8}$$

differing by a factor $(\psi_M - \psi_0)^{-M+1}$ from the non-informative prior density $f^{\Psi,\Theta}$ defined above.

If we wish to impose a constant deposition rate $\lambda(t)$ in $[\Psi_0, \Psi_1]$ then we set $M = 1$ in Equation (8). Integrating out variables ψ_M and ψ_0 leads to a prior density $\hat{f}^{\Theta}(\theta)$ for the variables $\Theta \in \Omega_\Theta$ given approximately by

$$\hat{f}^{\Theta}(\theta) \simeq \frac{1}{(\max(\theta) - \min(\theta))^{K-2}}.$$

The approximation treats terms $(U - \min(\theta))^{-K+2}$ and $(\max(\theta) - L)^{-K+2}$ as small.

2.3 Modified priors

Since the prior is a vehicle for archaeological fore-knowledge, we suggest taking a non-informative (and unnormalised) density with known properties, and using this as a base, incorporating knowledge via further multiplied densities and conditioning.

We define such *base priors* by dividing the original densities f^{Θ}, $f^{A,B,\Theta}$ and $f^{\Psi,\Theta}$ by the associated marginal density for Δ. Recall that $R \equiv U - L$. For the simplest model, with $\theta \in \Omega_\Theta$,

$$\tilde{f}^{\Theta}(\theta) \equiv \frac{1}{R - (\max(\theta) - \min(\theta))} \times \frac{1}{(\max(\theta) - \min(\theta))^{K-2}}, \tag{9}$$

while for the general phase model, with $(\alpha, \beta, \theta) \in \Omega_{A,B,\Theta}$,

$$\tilde{f}^{A,B,\Theta}(\alpha, \beta, \theta) \equiv \frac{1}{R - (\max(\alpha) - \min(\beta))} \times \frac{1}{(\max(\alpha) - \min(\beta))^{2M-2}} \times \prod_{m=1}^{M} (\alpha_m - \beta_m)^{-N_m},$$

and finally for the simplified layer model, with $(\psi, \theta) \in \Omega_{\Psi,\Theta}$,

$$\tilde{f}^{\Psi,\Theta}(\psi, \theta) \equiv \frac{1}{R - (\psi_M - \psi_0)} \times \frac{1}{(\psi_M - \psi_0)^{M-1}} \times \prod_{m=1}^{M} (\psi_m - \psi_{m-1})^{-N_m}.$$

These prior densities are attractive to us for two reasons. Firstly, they have the property that the marginal density for the difference between the top and bottom dates, or phases, is *a priori* uniformly distributed over its allowed range $[0, R]$, that is

$$\Theta \sim \tilde{f}^{\Theta}(\Theta) \quad \Rightarrow \quad \max(\Theta) - \min(\Theta) \sim \mathsf{U}(0, R),$$

$$(A, B, \Theta) \sim \tilde{f}^{A,B,\Theta}(A, B, \Theta) \quad \Rightarrow \quad \max(A) - \min(B) \sim \mathsf{U}(0, R),$$

$$(\Psi, \Theta) \sim \tilde{f}^{\Psi,\Theta}(\Psi, \Theta) \quad \Rightarrow \quad \Psi_M - \Psi_0 \sim \mathsf{U}(0, R).$$

Secondly, $\tilde{f}^{\Theta}$ and $\tilde{f}^{\Psi,\Theta}$ are equal to the corresponding deposition model prior densities, $\hat{f}^{\Theta}$ and $\hat{f}^{\Psi,\Theta}$, up to factors $R - (\max(\theta) - \min(\theta))$ and $R - (\psi_M - \psi_0)$ respectively. These extra state space factors are needed to get exactly uniform marginal priors for the range parameter, Δ. They will be near to constant in analyses where a conservative ($L << \Psi_0$ and $\Psi_M << U$) range is chosen, as the likelihood will then penalise states sensitive to $R - (\max(\theta) - \min(\theta))$ or $R - (\psi_M - \psi_0)$. When a tight range can be used, so that L and U are close to the true start and end of the dated process, we regard the uniform marginal range as too valuable to give up, and retain the factors $R - (\max(\theta) - \min(\theta))$ and $R - (\psi_M - \psi_0)$.

3 Case studies

The marginal prior distributions of the range variable Δ under the priors f^{Θ}, $f^{A,B,\Theta}$ and $f^{\Psi,\Theta}$ show strong power-law dependence. Thus if δ and 2δ are two candidate date ranges in a model with M phases and K data points, the range 2δ is favoured over the range δ by a factor of 2^K under prior f^{Θ}, 2^{2M-E-2} under prior $f^{A,B,\Theta}$ and 2^{M-1} under prior $f^{\Psi,\Theta}$. However, the likelihood has a Δ dependence which is approximately Gaussian (at least over large date ranges) and so any spreading effect of the prior will be visible in the posterior only if the likelihood admits both δ and 2δ. Under priors $\tilde{f}^{\Theta}$, $\tilde{f}^{A,B,\Theta}$ and $\tilde{f}^{\Psi,\Theta}$ candidate ranges δ and 2δ are, *a priori*, equally probable.

In simulation studies from synthetic data we have found that context dates reconstructed under priors $\tilde{f}^{\Theta}$, $\tilde{f}^{A,B,\Theta}$ or $\tilde{f}^{\Psi,\Theta}$ are typically distributed across their true values while context dates reconstructed under priors f^{Θ}, $f^{A,B,\Theta}$ or $f^{\Psi,\Theta}$ are not. When the synthetic dates Θ are spread across a range which is not large compared to the observation uncertainties $\tilde{\sigma}$ then priors f^{Θ}, $f^{A,B,\Theta}$ or $f^{\Psi,\Theta}$ give results which are often quite wrong. We do not present these studies, as the conclusion is unsurprising, given the marginal densities of the range variable Δ in these priors.

We present two case studies comparing all three pairs of models f^{Θ} and $\tilde{f}^{\Theta}$, $f^{\Psi,\Theta}$ and $\tilde{f}^{\Psi,\Theta}$, and $f^{A,B,\Theta}$ and $\tilde{f}^{A,B,\Theta}$. In Case I, the likelihood functions $g(y_{m,n}|\theta_{m,n})$ show strong overlap in $\theta_{m,n}$, from the earliest to the latest data (see Figure 4). When the data has this property, our conclusions will be sensitive to the choice of prior. The data in Case II is taken from Buck *et al.* 1996. We repeat their analysis using the prior $f^{A,B,\Theta}$, and place it alongside an analysis using the prior $\tilde{f}^{A,B,\Theta}$. In this case the data is scattered over range which is large in comparison to the range of support of the likelihood functions $g(y_{m,n}|\theta_{m,n})$ (see Figure 7), and our conclusions are consequently insensitive to the choice of prior. The models used in these case studies are set out in Table 1.

Case, Data	*Model label*	*Number of parameters*	*State space*	*Prior density*	$[L, U]$	*Prior density of range, δ*
I Shag mouth	$\mathcal{M}_1$	$K = 34$, $M = 0$	Ω_Θ	$f^{\Theta}(\theta)$	$[250, 950]$	$(R - \delta)\delta^{32}$
I Shag mouth	$\mathcal{M}_2$	$K = 34$, $M = 10$	$\Omega_{\Psi,\Theta}$	$f^{\Psi,\Theta}(\psi, \theta)$	$[250, 950]$	$(R - \delta)\delta^{9}$
I Shag mouth	$\mathcal{M}_3$	$K = 34$, $M = 0$	Ω_Θ	$\tilde{f}^{\Theta}(\theta)$	$[250, 950]$	1
I Shag mouth	$\mathcal{M}_4$	$K = 34$, $M = 10$	$\Omega_{\Psi,\Theta}$	$\tilde{f}^{\Psi,\Theta}(\psi, \theta)$	$[250, 950]$	1
II Jama river	$\mathcal{M}_5$	$K = 37$, $M = 7$	$\Omega_{A,B,\Theta}$	$f^{A,B,\Theta}(\alpha, \beta, \theta)$	$[0, 5500]$	$(R - \delta)\delta^{11}$
II Jama river	$\mathcal{M}_6$	$K = 37$, $M = 7$	$\Omega_{A,B,\Theta}$	$\tilde{f}^{A,B,\Theta}(\alpha, \beta, \theta)$	$[0, 5500]$	1

Table 1: Models trialed in case studies I and II. Minimum and maximum age limits L and U were chosen to be computationally equivalent to $[0, \infty)$ and otherwise as tight as possible.

Case study I: The Shag River Mouth data

We will investigate the extent to which our choice of prior effects our summary of a large data set gathered at the mouth of Shag river, in southern New Zealand. Quantities of interest to archaeologists include the length of time for which the site was occupied, and the actual dates at which occupation began and ended. We analyse a large subset of the data, a set of 34 radiocarbon ages from a single series of 10 distinct strata. The data is published in Anderson *et al.* 1996. We are analysing the SM/C:Dune series. Relevant information is summarised in Table 2.

The SM/C:Dune series is up to 2.5 metres deep, comprising many distinct layers. For this and other

Layer	*Terrestrial radiocarbon ages*	*Marine radiocarbon ages*
L1	$580_{1,1}(47)$ $630_{1,2}(82)$	
L3	$600_{3,1}(50)$ $1201_{3,2}(38)$	$747_{3,3}(38)$ $960_{3,4}(45)$ $1020_{3,5}(50)$ $990_{3,6}(45)$ $980_{3,7}(45)$ $1010_{3,8}(50)$ $1160_{3,9}(50)$ $965_{3,10}(26)$ $1060_{3,11}(45)$ $980_{3,12}(40)$ $1070_{3,13}(80)$ $1040_{3,14}(45)$ $980_{3,15}(40)$ $950_{3,16}(45)$ $1050_{3,17}(50)$ $1070_{3,18}(45)$
L4	$537_{4,1}(44)$ $1170_{4,2}(70)$ $600_{4,3}(50)$	
L5	$670_{5,1}(47)$ $624_{5,2}(58)$ $560_{5,3}(45)$	
L6	$646_{6,1}(47)$ $630_{6,2}(35)$ $509_{6,3}(72)$	$1022_{6,4}(29)$
L7	$570_{7,1}(45)$	
L10	$660_{10,1}(46)$ $787_{10,2}(72)$	$974_{10,3}(49)$

Table 2: The SM/C:Dune series radiocarbon data of Anderson *et al.* 1996. 34 radiocarbon ages are arranged in 10 layers. We do not model an empty level above L1. As a consequence our level labels L1 to L10 differ from those in Anderson *et al.* 1996 (our L1 is their L2). There were no dated materials in layers 2, 8 or 9. The notation is $y_{m,n}(\sigma_{m,n})$, with the layer index m and within-layer index n in subscript (m,n).

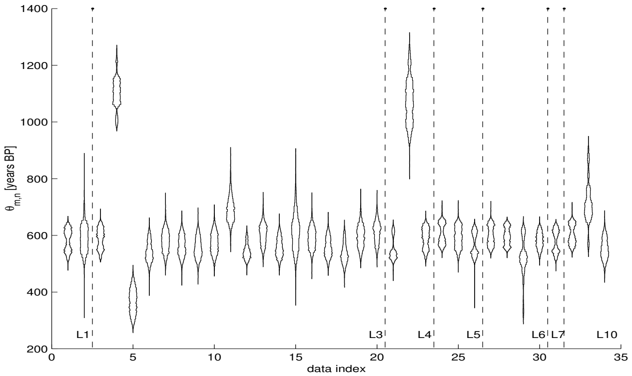

Figure 4: Likelihood factors showing $g(y_{m,n}|\theta_{m,n})$ for each data point as a histogram against $\theta_{m,n}$. The width of the strip is proportional to the likelihood at that value of $\theta_{m,n}$. Indices 1 through 34 correspond to labels $(1,1)$ through $(10,3)$ of text. Indices 4, 5 and 22 of this graph (*ie* $(3,1)$, $(3,2)$ and $(4,2)$ of text) were identified as outliers.

reasons, it was expected that the occupation span of the site would be some significant number of years, possibly centuries. Anderson *et al.* 1996 suggest 20 to 50 years, however their analysis is based on grouping context dates according to the chi-squared statistics of the associated radiocarbon ages. Such a methodology is unreliable when the terrestrial radiocarbon ages correspond to a non-monotonic section of the calibration curve, as is the case here. However, we find that a short occupation span, consistent with the results of Anderson *et al.* 1996, is not ruled out.

Archaeological and chronometric considerations led Anderson *et al.* 1996 to reject several dates as outliers. We applied the outlier analysis of Christen 1994 to this set of data, as well as exploring a scheme like that of George and McCulloch 1996. We identified data $(3,1)$, $(3,2)$ and $(4,2)$ (NZ-7743, NZ-7804 and NZ-7737 of Anderson *et al.* 1996) as outliers (see Figure 4). Since the outliers did not lie in the sensitive top or bottom layers, we choose to carry out the analysis with the outliers in place, and recover their radiocarbon ages in simulation. The outliers were assigned $\tilde{\sigma}_{m,n}(\theta_{m,n}) = 200$, so that their individual likelihood was effectively uniform over the allowed range. Thus all our models have $K = 34$ distinct context date parameters $\theta_{m,n}$ corresponding to the radiocarbon ages of Table 2.

In Section 1 we referred to the southern hemisphere offset. It is believed this can be modelled as an offset d to be subtracted from the $y_{m,n}$ (since they are ages, with units *years BP*). However, measurements reported in Sparks *et al.* 1995 and Barbetti *et al.* 1995 indicate that, for calibration of terrestrial material from New Zealand, this offset is consistent with zero. We do not attempt to account for the uncertainty introduced by this aspect of the measurement. For calibration of marine material from New Zealand, McFadgen and Manning 1990 give $d_{\mathrm{mrn}} = -30$ with standard deviation $\sigma^d_{\mathrm{mrn}} = 13$. We substitute $y_{m,n} \leftarrow y_{m,n} - d_{\mathrm{mrn}}$ and $\sigma^2_{m,n} \leftarrow \sigma^2_{m,n} + (\sigma^d_{\mathrm{mrn}})^2$ so that our expression for the likelihood $g(y|\theta)$ is unchanged.

We focus on the posterior distribution of the minimum occupation span, $\max(\Theta) - \min(\Theta)$ (rather than $\Psi_M - \Psi_0$, to allow comparison between f^{Θ} and $f^{\Psi,\Theta}$ *etc*). We trial 4 models, with properties given in Table 1. We have 34 data points, and no *a priori* justification for treating groups of data points as measurements of a single context date. We therefore have $K = 34$ of the $\theta_{m,n}$ parameters in models $\mathcal{M}_1$ through $\mathcal{M}_4$. Our phase models $f^{\Psi,\Theta}$ and $\tilde{f}^{\Psi,\Theta}$ have 11 layer-age parameters, ψ_m $m = 0 \ldots 10$.

Samples $\{\theta^{(j)}\}_{j=0}^{J}$ and $\{\psi^{(j)}, \theta^{(j)}\}_{j=0}^{J}$, distributed with densities $h^{\Theta}(\theta|y,f)$ and $h^{\Psi,\Theta}(\psi,\theta|y,f)$, were simulated using MCMC for models $\mathcal{M}_1$ - $\mathcal{M}_4$. See Appendix A for details of the sampling algorithm. We have checked our MCMC simulations for convergence according to the conditions of Geyer 1992. In particular the integrated auto-covariance time, per sample, computed from the likelihood, was close to 1 in all our simulations of the Shag mouth data. Each of the 10000 realisations sampled from each of the four posterior distributions are therefore very nearly independent. This approach is efficient when computer memory for storing samples is in short supply.

Posterior distributions of $\max(\theta) - \min(\theta)$ and of $\max(\theta)$ and $\min(\theta)$ are plotted for models $\mathcal{M}_1$ to $\mathcal{M}_4$ on axes with equal age ranges and scales in Figure 5. Models $\mathcal{M}_1$ and $\mathcal{M}_2$ lead to far higher estimates for the minimum occupation period than do models $\mathcal{M}_3$ and $\mathcal{M}_4$, as the marginal prior range distributions of Table 1 lead us to expect. We found it difficult to estimate accurate Bayes factors for these models and data. In lieu of Bayes factors we plot in Figure 6 distributions of reconstructed dates alongside their original likelihood distributions. We have chosen a subset of seven dates spanning the strata. Posterior distributions from models $\mathcal{M}_3$ and $\mathcal{M}_4$ lie within the range suggested by the likelihood; distributions from models $\mathcal{M}_1$ and $\mathcal{M}_2$ favour dates close to the boundary of the region where the likelihood becomes small. Priors $\tilde{f}^{\Theta}$ and $\tilde{f}^{\Psi,\Theta}$ give a better summary than do f^{Θ} and $f^{\Psi,\Theta}$ of the prior knowledge available in this analysis. Figures 5-M1 and 5-M2 will very likely overestimate the occupation span.

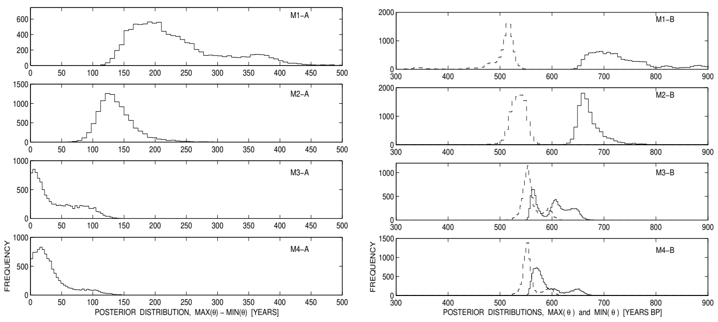

Figure 5: Labels *M1-M4* indicate the prior used, as in Table 1. $M1$-A to $M4$-A: histograms quantifying the range of plausible posterior values for the minimum occupation time, $\max(\theta) - \min(\theta)$, at the Shag mouth site. $M1$-B to $M4$-B: posterior distributions of $\max(\theta)$ (solid histograms) and $\min(\theta)$ (dashed histogram).

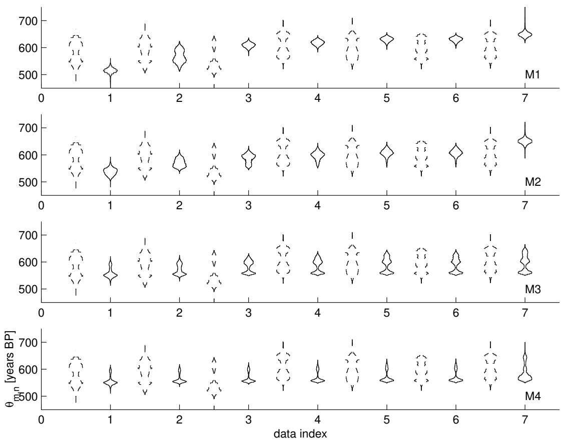

Figure 6: Labels *M1-M4* indicate the prior used, as in Table 1. Posterior distributions of $\theta_{m,n}$ for selected data, $1 = (1,1), 2 = (3,1), 3 = (4,1), 4 = (5,1), 5 = (6,1), 6 = (6,2), 7 = (10,1)$ (solid histograms); likelihood of $\theta_{m,n}$ as in Figure 4 (dashed histogram left of solid). The prior densities of models $\mathcal{M}_1$ and $\mathcal{M}_2$ favour dates near the limits of the allowed range, while those of models $\mathcal{M}_3$ and $\mathcal{M}_4$ allow a range of dates from the centre of the likelihood.

Case study II: The Jama river valley, Ecuador

The data treated in this section is taken from Buck *et al.* 1996 pages 226 to 238 where it is analysed using the multi-phase prior $f^{A,B,\Theta}$ as in model $\mathcal{M}_5$. All data is terrestrial. There are $M = 7$ seven phases, and $K = 37$ data points. The likelihood functions $g(y_{m,n}|\theta_{m,n})$ of all 37 dates are plotted

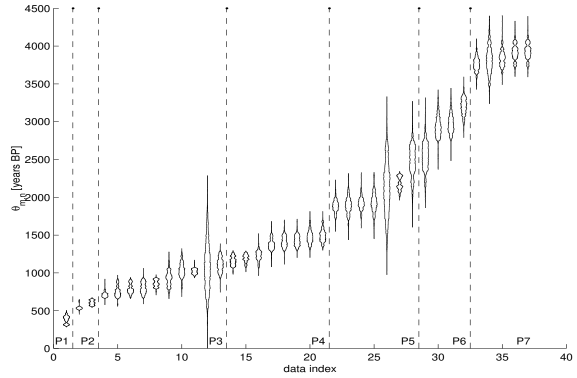

Figure 7: Likelihood factors $g(y_{m,n}|\theta_{m,n})$ as for Figure 4. Indices 1 through 37 correspond to labels $(1,1)$ through $(7,5)$ of the Jama river valley data of Buck *et al.* 1996 pages 226 to 238. Phases Muchique 5 through 1, Tabuchila and Piquigua of Buck *et al.* 1996 are here labelled phases P1 through P7 respectively.

as histograms in Figure 7. Prior knowledge restricts the phases of this model in the following way

$$\begin{aligned} \mathrm{C}(\alpha,\beta,L,U) = \{\ & \alpha_m > \beta_m, m = 1\ldots M;\ \beta_1 > L;\ U > \alpha_7; \\ & \beta_7 > \alpha_6;\ \beta_6 > \alpha_5;\ \beta_5 > \alpha_4;\ \beta_4 > \beta_3 > \beta_2;\ \alpha_4 > \alpha_3 > \alpha_2;\ \beta_2 = \alpha_1\ \}. \end{aligned} \quad (10)$$

A graph representing the constraint set $\mathrm{C}(\alpha,\beta,L,U)$ is given in Figure 8. We examine the posterior distribution $\mathrm{H}^{A,B,\Theta}(d\alpha\, d\beta\, d\theta|y,f)$ under priors $f = f^{A,B,\Theta}(\alpha,\beta,\theta)$ and $\tilde{f}^{A,B,\Theta}(\alpha,\beta,\theta)$, *ie*, models $\mathcal{M}_5$ and $\mathcal{M}_6$ of Table 1.

Simulation was carried out using a Metropolis-Hastings algorithm similar to that described in Appendix A. As in Case I, the integrated auto-correlation time of the likelihood per sample was close to 1 in both MCMC runs (under models $\mathcal{M}_5$ and $\mathcal{M}_6$). The posterior distributions of the phase

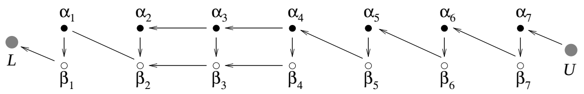

Figure 8: Graph of the constraint relations $\mathrm{C}(\alpha,\beta,L,U)$ for the Jama multi-phase model of Equation (10). An undirected edge between the nodes of variables a and b represents the constraint $a = b$. Refer to Figure 2 for label conventions.

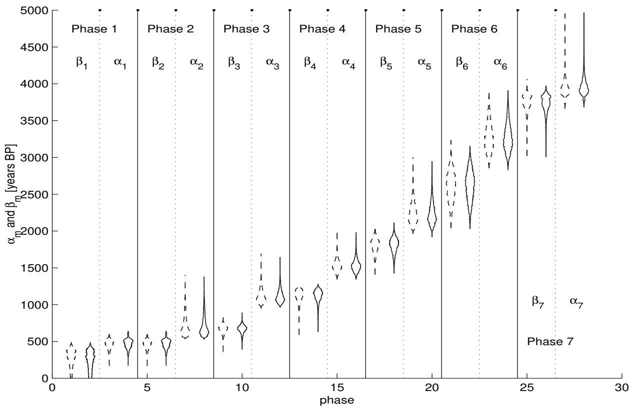

Figure 9: A comparison of the posterior distributions of the phase variables α and β under models $\mathcal{M}_5$ and $\mathcal{M}_6$. Solid histograms $\mathcal{M}_5$, dashed histograms, $\mathcal{M}_6$.

variables α and β, under models $\mathcal{M}_5$ and $\mathcal{M}_6$, are plotted in Figure 9. As anticipated for Case II data, there is little difference between the posterior distributions under priors with uniform and spreading marginal range distributions.

To conclude, the priors $\tilde{f}^{\Theta}$ and $\tilde{f}^{\Psi,\Theta}$ are appropriate for radiocarbon dating in the presence of hard temporal constraints, since they are derived from a model of the deposition process, albeit a simple one. Moreover their marginally uniform prior range makes them convenient non-informative base priors, to be modified as further information becomes available.

References

Anderson, A., Smith, I., and Higham, T. 1996. *Shag river mouth: the archaeology of an early Southern Maori village*. The Australian National University, Canberra: ANH Publications, RSPAS. Chap. 7, pages 61–69.

Barbetti, M., Bird, T., Dolezal, G., Taylor, G., Francey, R., Cook, E., and Petersen, M. 1995. Radiocarbon variations from tasmanian conifers: results from three holocene logs. *Radiocarbon*, **37**, 361–369.

Broffit, J.D. 1984. A Bayes estimator for ordered parameters and isotonic Bayesian graduation. *Scand. Actuarial J.*, 231–247.

Buck, C.E., Kenworthy, J.B., C.D.Litton, and Smith, A.F.M. 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity*, **65**, 808–821.

Buck, C.E., Litton, C.D., and Smith, A.F.M. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science*, **19**, 497–512.

Buck, C.E., Cavanagh, W.G., and C.D.Litton. 1996. *Bayesian approach to interpreting archaeological data*. Chichester: John Wiley.

Christen, J.A. 1994. Summarizing a set of radiocarbon determinations: a robust approach. *Appl. Statist.*, **43**, 489–503.

George, E.I., and McCulloch, R.E. 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall. Chap. 12, pages 203–214.

Geyer, C.J. 1992. Practical Markov chain Monte Carlo. *Stat. Sci.*, **7**, 473–511.

Litton, C., and Buck, C. 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall. Chap. 25, pages 466–486.

McFadgen, B.G., and Manning, M. 1990. Calibrating new zealand radiocarbon dates of marine shells. *Radiocarbon*, **32**, 229–232.

Raftery, A.E. 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall. Chap. 10, pages 163–187.

Ramsey, C.B. 1995. Radiocarbon calibration and analysis of stratigraphy: The OxCal program. *Radiocarbon*, **37**, 425–430.

Sparks, R., Melhuish, W., McKee, J., Ogden, J., Palmer, J., and Molloy, B. 1995. 14C calibration in the southern hemisphere and the date of the last Taupo eruption: evidence from tree ring sequences. *Radiocarbon*, **37**, 155–163.

Stuiver, M., and Braziunas, T.F. 1993. Modeling atmospheric 14C influences and 14C ages of marine samples to 10000 BC. *Radiocarbon*, **35**, 137–191.

Stuiver, M., and Pearson, G.W. 1993. High-precision bidecadal calibration of the radiocarbon time scale AD 1950-500 BC and 2500-6000 BC. *Radiocarbon*, **35**, 1–25.

Tierney, L. 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall. Chap. 4, pages 59–74.

Appendix A: Sampling algorithm

We implemented a Metropolis Hastings MCMC algorithm to generate samples ψ, θ distributed with density $h^{\Psi,\Theta}(\psi, \theta|y, f)$. Earlier authors note that the Gibbs sampler functions quite efficiently for sampling $h^{\Psi,\Theta}(\psi, \theta|y, f)$ when the prior density is $f^{\Theta,\Psi}$, and that in this case the necessary conditional distributions are fairly straightforward to write down and compute. However, prior densities like $f = \tilde{f}^{\Theta}(\theta)$ lead to more complicated expressions for conditionals, depending on $\max(\theta)$ and $\min(\theta)$. This is not a significant obstacle to simulation: a small set of Metropolis Hastings updates give ergodic behaviour on useful time scales.

We define an algorithm generating a realisation $\{\psi^{(j)}, \theta^{(j)}\}_{j=0}^{J}$ of a Markov chain of random variables $\{\Psi^{(j)}, \Theta^{(j)}\}_{j=0}^{J}$, with equilibrium distribution $\mathrm{H}^{\Psi,\Theta}(d\psi\, d\theta|y, f)$. Once this is done we will be able to form empirical estimates $\bar{u}$ for the posterior expectation $\mathsf{E}\{u(\Psi, \Theta)|y, f\}$ of any function $u(\psi, \theta)$ on $\Omega_{\Psi,\Theta}$ we wish. If C is a suitable *burn-in* period, chosen to allow for the sequence of samples to reach equilibrium, we estimate

$$\mathsf{E}\{u(\Psi, \Theta)|y, f\} \simeq \frac{1}{J - C} \sum_{j=C+1}^{J} u(\psi^{(j)}, \theta^{(j)}),$$

for data y and prior density f. We define the Markov chain via a stochastic update rule. This update is chosen so that it determines a transition kernel

$$\Pr\{\Psi^{(j+1)} \in d\psi',\ \Theta^{(j+1)} \in d\theta' | \Psi^{(j)} = \psi, \Theta^{(j)} = \theta\}$$

from ψ, θ into $d\psi' d\theta'$ which preserves $\mathrm{H}^{\Psi,\Theta}(d\theta\, d\psi | y, f)$. In order to get an efficient sampler, it is most effective to use several transition rules, labelled $v = 1, 2 \ldots$, each with its own transition kernel. Each rule is made up of a generation step, in which a candidate state (ψ', θ') is generated according to a distribution Q_v for rule v,

$$Q_v(d\psi' d\theta' | \psi, \theta) \equiv q_v(\psi', \theta' | \psi, \theta) d\psi' d\theta',$$

and an acceptance step, in which the candidate state is accepted with probability $\alpha_v(\psi', \theta' | \psi, \theta)$, or rejected. The algorithm is as follows.

Let $(\Psi^{(j)}, \Theta^{(j)}) = (\psi, \theta)$. $(\Psi^{(j+1)}, \Theta^{(j+1)})$ is determined in the following way.

1. Pick a new state $(\psi', \theta') \sim Q_v(d\psi' d\theta' | \psi, \theta)$.
2. Accept (ψ', θ') (*ie*, set $(\Psi^{(j+1)}, \Theta^{(j+1)}) = (\psi', \theta')$) with probability

$$\alpha_v(\psi', \theta' | \psi, \theta, f) \equiv \min\left\{1, \frac{\mathrm{H}^{\Psi,\Theta}(d\psi' d\theta' | y, f) Q_v(d\psi\, d\theta | \psi', \theta')}{\mathrm{H}^{\Psi,\Theta}(d\psi\, d\theta | y, f) Q_v(d\psi' d\theta' | \psi, \theta)}\right\}.$$

If (ψ', θ') is not accepted, set $(\Psi^{(j+1)}, \Theta^{(j+1)}) = (\psi, \theta)$ (*ie*, no change).

Sufficient conditions for (Ψ, Θ) to have equilibrium distribution $\mathrm{H}^{\Psi,\Theta}$ are given, for example, in Tierney 1996. We have four update types, with distinct generation distributions Q_v, so the above algorithm is repeated for each update type, $v = 1, 2, 3, 4$ in sequence, and the sequence repeated until enough samples have been generated.

Our update types are as follows.

- Update a single θ variable, $v = 1$, $(\psi, \theta \to \psi, \theta')$.
A parameter $\theta_{i,j}$ is chosen uniformly at random (UAR) from the set $\{\theta_{1,1} \ldots \theta_{N,N_M}\}$. A new value for $\theta_{i,j}$ ($\theta'_{i,j}$ say) is selected UAR on the interval $[\psi_{i-1}, \psi_{i+1}]$. Let $\theta' = (\theta_{1,1}, \ldots \theta'_{i,j}, \ldots \theta_{M,N_M})$. In this case $q_1(\theta' | \psi, \theta)$ is equal to $q_1(\theta | \psi, \theta')$ since $[\psi_{i-1}, \psi_i]$ is the same for both transitions $\theta \to \theta'$ and $\theta' \to \theta$. Now

$$\begin{aligned} \alpha_1(\theta' | \psi, \theta, f) &= \min\left\{1, \frac{g(y|\theta') f(\theta')}{g(y|\theta) f(\theta)}\right\} \\ &= \min\left\{1, \frac{e^{-(y_{i,j} - \mu(\theta'_{i,j}, i, j))^2 / 2\tilde{\sigma}_{i,j}(\theta'_{i,j})^2}}{e^{-(y_{i,j} - \mu(\theta_{i,j}, i, j))^2 / 2\tilde{\sigma}_{i,j}(\theta_{i,j})^2}} \times \frac{\tilde{\sigma}_{i,j}(\theta_{i,j})}{\tilde{\sigma}_{i,j}(\theta'_{i,j})}\right\} \end{aligned}$$

for both $f = f^{\Psi,\Theta}$ and $f = \tilde{f}^{\Psi,\Theta}$.

- Update a single ψ variable, $v = 2$, $(\psi, \theta \to \psi', \theta)$.
An index i is chosen UAR in $0, 1 \ldots M$. A new value of ψ'_i for ψ_i is selected UAR on the interval $[\max(\theta_{i-1,\cdot}), \min(\theta_{i,\cdot})]$ (the interval is $[L, \min(\theta_{1,\cdot})]$ if $i = 0$ and $[\max(\theta_{M-1,\cdot}), U]$ if $i = M$). Let $\psi' = (\psi_0, \ldots \psi'_i, \ldots \psi_M)$. Here $q_2(\psi' | \psi, \theta)$ is equal to $q_2(\psi | \psi', \theta)$ since the sample interval is the same for transitions $\psi \to \psi'$ and $\psi' \to \psi$. Now

$$\alpha_2(\psi' | \psi, \theta, f) = \min\left\{1, \frac{f(\psi', \theta')}{f(\psi, \theta)}\right\}$$

where, if $f = f^{\Psi,\Theta}$,

$$\frac{f^{\Psi,\Theta}(\psi', \theta')}{f^{\Psi,\Theta}(\psi, \theta)} = \begin{cases} (\psi_{i+1} - \psi_i)^{N_{i+1}} (\psi_i - \psi_{i-1})^{N_i} / (\psi'_{i+1} - \psi'_i)^{N_{i+1}} (\psi'_i - \psi'_{i-1})^{N_i} & \text{if } 0 < i < M \\ (\psi_1 - \psi_0)^{N_1} / (\psi'_1 - \psi'_0)^{N_1} & \text{if } i = 0 \\ (\psi_M - \psi_{M-1})^{N_M} / (\psi'_M - \psi'_{M-1})^{N_M} & \text{if } i = M \end{cases}$$

and if $f = \tilde{f}^{\Psi,\Theta}$

$$\frac{\tilde{f}^{\Psi,\Theta}(\psi', \theta')}{\tilde{f}^{\Psi,\Theta}(\psi, \theta)} = \frac{f^{\Psi,\Theta}(\psi', \theta')}{f^{\Psi,\Theta}(\psi, \theta)} \times \left[\frac{R - (\psi_M - \psi_0)}{R - (\psi'_M - \psi'_0)} \times \frac{(\psi_M - \psi_0)^{M-1}}{(\psi'_M - \psi'_0)^{M-1}}\right]$$

though the factor in square brackets is equal one unless $i = 0$ or $i = M$.

- Shift all dates, $v = 3$, $(\psi, \theta \to \psi', \theta')$.
In this update S is a positive constant with units *years* chosen to give a reasonable acceptance rate for the update. A scalar shift s is chosen uniformly at random in $[-S, S]$ and added to each variable, $\theta'_{m,n} = \theta_{m,n} + s$, and $\psi'_m = \psi_m + s$, for each $m = 0 \ldots M$ and $n = 1 \ldots N$. In this case $q_3(\psi', \theta' | \psi, \theta) = q_3(\psi, \theta | \psi', \theta')$, since the probabilities to choose the forward and reverse shifts are equal. This time

$$\alpha_3(\psi', \theta' | \psi, \theta, f) = \min\left\{1, \frac{g(y|\theta')}{g(y|\theta)}\right\}$$

with

$$\frac{g(y|\theta')}{g(y|\theta)} = \frac{e^{-|y - \mu(\theta')|^2 / 2\tilde{\sigma}(\theta')^2}}{e^{-|y - \mu(\theta)|^2 / 2\tilde{\sigma}(\theta)^2}} \times \prod_{m=0}^{M} \prod_{n=1}^{N} \frac{\tilde{\sigma}_{i,j}(\theta_{i,j})}{\tilde{\sigma}_{i,j}(\theta'_{i,j})}$$

for both priors $f^{\Psi,\Theta}$ and $\tilde{f}^{\Psi,\Theta}$ we are considering. Refer to Equation (1) for notation.

- Expand the dates about their mean, $v = 4$, $(\psi, \theta \to \psi', \theta')$.
A scalar multiplying factor ρ is chosen uniformly at random in $[2/3, 3/2]$. Let $\mathrm{av}(\psi, \theta)$ denote the arithmetic mean of the sequence ψ, θ of dates. We set $\theta'_{m,n} = \rho\theta_{m,n} - (\rho - 1)\mathrm{av}(\psi, \theta)$ and $\psi'_m = \rho\psi_m - (\rho - 1)\mathrm{av}(\psi, \theta)$, for each $m = 0 \ldots M$ and $n = 1 \ldots N$. This operation scales all the dates by ρ whilst keeping $\mathrm{av}(\psi, \theta) = \mathrm{av}(\psi', \theta')$. This time the correct acceptance rates are more difficult to calculate. We find

$$\alpha_4(\psi', \theta' | \psi, \theta, f^{\Psi,\Theta}) = \min\left\{1, \frac{g(y|\theta')}{g(y|\theta)} \times \rho^{M-2}\right\}$$

and

$$\alpha_4(\psi', \theta' | \psi, \theta, \tilde{f}^{\Psi,\Theta}) = \min\left\{1, \frac{g(y|\theta')}{g(y|\theta)} \times \frac{R - (\psi_M - \psi_0)}{R - \rho(\psi_M - \psi_0)} \times \frac{1}{\rho}\right\}$$

with $g(y|\theta')/g(y|\theta)$ as in move $v = 3$.

Our implementations were checked by varying the proportions of moves, and checking that all statistics measured on the output were unchanged. We checked also that the prior distribution of the range was uniform under models $\mathcal{M}_3$, $\mathcal{M}_4$ and $\mathcal{M}_6$. Finally, we compared the posterior distributions of phase variables α and β under $\mathcal{M}_5$ with those reported in the identical earlier analysis of Buck *et al.* 1996. No visible differences were detected.

The posterior $\mathrm{H}^{\Psi,\Theta}(d\psi\, d\theta | y, f^{\Psi,\Theta})$ may in general be simulated adequately with moves $v = 1$ and $v = 2$ alone. However, in order to simulate reliably the posterior with prior density $\tilde{f}^{\Theta}$, we require in addition move $v = 4$. Move $v = 3$ improves the efficiency of the sampling process.