This is a repository copy of *Bayesian radiocarbon modelling for beginners*.

White Rose Research Online URL for this paper:
https://eprints.whiterose.ac.uk/212885/

Version: Published Version

Article:
Buck, C.E. orcid.org/0000-0002-0872-9504 and Juarez, M.A. (2024) Bayesian radiocarbon modelling for beginners. Archaeometry. ISSN 0003-813X

https://doi.org/10.1111/arcm.12998

Reuse
This article is distributed under the terms of the Creative Commons Attribution (CC BY) licence. This licence allows you to distribute, remix, tweak, and build upon the work, even commercially, as long as you credit the authors for the original work. More information and the full terms of the licence here: https://creativecommons.org/licenses/

Takedown
If you consider content in White Rose Research Online to be in breach of UK law, please notify us by emailing eprints@whiterose.ac.uk including the URL of the record and the reason for the withdrawal request.

eprints@whiterose.ac.uk
https://eprints.whiterose.ac.uk/

Check for updates

Received: 27 September 2023 | Accepted: 17 May 2024

DOI: 10.1111/arcm.12998

ORIGINAL ARTICLE

Bayesian radiocarbon modelling for beginners

Caitlin E. Buck | **Miguel A. Juárez**

School of Mathematics and Statistics, University of Sheffield, Sheffield, UK

Correspondence
C. E. Buck, School of Mathematics and Statistics, University of Sheffield, Hicks Building, Hounsfield Road, Sheffield S3 7RH, UK.
Email: c.e.buck@sheffield.ac.uk

Abstract

Due to freely available, tailored software, Bayesian statistics is now the dominant paradigm for archaeological chronology construction in the UK and much of Europe and is increasing in popularity in the Americas. Such software provides users with powerful tools for Bayesian inference for chronological models with little need to undertake formal study of statistical modelling or computer programming. This runs the risk that it is reduced to the status of a black box, which is not sensible given the power and complexity of the modelling tools it implements. In this paper we seek to offer intuitive insight to ensure that readers from the archaeological research community who use Bayesian chronological modelling software will be better able to make well educated choices about the tools and techniques they adopt. Our hope is that they will then be both better informed about their own research designs and better prepared to offer constructively critical assessments of the modelling undertaken by others.

DECLARATION

Much of the content of this paper was originally published in Spanish as Buck and Juárez (2020). The content (including examples) has been revised, however, including changes made in response to helpful comments and suggestions made by two referees, appointed by *Archaeometry*. The resulting update is published here with the generous consent of the Editors of *Archaeometry* and the original publishers, Dextra Editorial, Madrid.

BACKGROUND

Bayesian chronological models are statistical models that allow us to represent, manage and interpret both relative and absolute chronological information from one or more archaeological

This is an open access article under the terms of the Creative Commons Attribution License, which permits use, distribution and reproduction in any medium, provided the original work is properly cited.

© 2024 The Author(s). *Archaeometry* published by John Wiley & Sons Ltd on behalf of University of Oxford.

or palaeoenvironmental research projects. They were developed over the last thirty years, specifically for the archaeological and palaeoenvironmental research communities, by statisticians and software developers who took advantage of a revolution in our ability to implement such models using a simulation-based (as opposed to an exact calculation) approach.

Users of the resulting software need not know the details of the underlying mathematics and statistics nor of the computational techniques used to implement them. They must, however, understand the concept of a model, appreciate the choices they are making when they select a particular model to represent their own project, and understand enough of the key decisions made by the statistical modellers and software developers to know which model and software is appropriate for their needs.

Much has been written in the past about the motivation for and mechanics of using Bayesian methods for chronology construction (e.g., Bayliss, 2007; Bayliss & Marshall, 2022; Buck et al., 1996). Most of these do not offer much detail, however, to help users to get started in their first forays into formal chronological modelling, which is the precursor to implementation. This paper seeks to fill that gap in the literature and to do so using simple, toy datasets so that we can focus on the ideas that are common to all projects of this sort and not become distracted by the specifics of particular case studies.

In Section 2.1 we look at some key concepts and decisions involved in modelling in general. Section 3 focuses on the basics of formal chronology construction, highlighting the fact that both section drawings and Harris matrices are types of chronological model. Formal statistical notation is then introduced in Section 4, and used to define the statistical models now in routine use for Bayesian radiocarbon dating. These are then applied to specific examples and illustrative software output is discussed. Section 5 focuses on some of the practicalities involved in using Bayesian radiocarbon calibration software, and Section 6 looks to the future.

What makes a good model?

For the purposes of this paper, we define a model as

> a representation of a person, organism, structure or concept typically smaller, simpler and/or more abstract than the original.

This definition highlights the very many different ways in which the term 'model' is used in modern parlance, but it tells us nothing about what makes a good model for any specific purpose. To move towards this, we start by thinking about a simple and perhaps rather trivial modelling problem: what makes a good model of an elephant?

Consider the model elephants pictured in Figure 1; clearly neither is anatomically accurate, but is one model definitively better than the other? Our contention is that which one is best depends on what the model is to be used for. If the model is for entertaining a 3-year-old child on a wet afternoon, then the one on the left is probably the best. It is safe for them to play with on their own, has no sharp or fragile pieces and will allow them to recognise an elephant just from its long nose. If, on the other hand, the model is to be used to help a 10-year-old child learn the key features of real elephants, then the model on the right is surely more suitable. It has more realistic legs and head and also has a tail, ears and tusks, which are all missing from the one on the left. Were we to want to move beyond these basics, however, to teach an older child or adult about, say, the differences between African and Indian elephants or about the physiology of elephants relative to other large mammals, then neither of the models pictured would be suitable and we would need to look elsewhere for something more anatomically detailed.

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

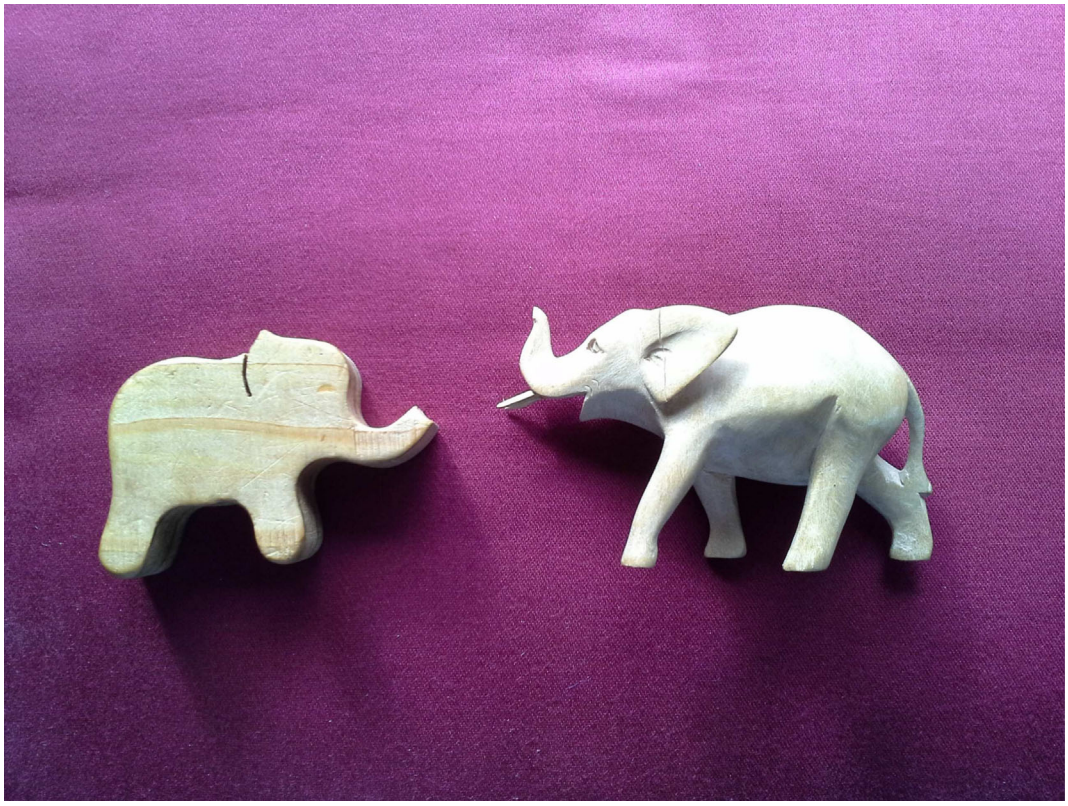


FIGURE 1 Two models of elephants: which is better?

Analogies of this sort are useful only up to a point and this one could certainly be taken too far, but before we leave it and move to think about chronological models it is worth noting a couple of similarities between them and model elephants. Both chronological models and model elephants can be off-the-shelf or tailor-made and descriptive or mechanistic. The elephants in Figure 1 are both off-the-shelf and descriptive, but were we to seek anatomically correct models for more sophisticated purposes, then they may well need to be tailor-made and mechanistic.

In a similar way, for many purposes, chronological models that are descriptive and available off-the-shelf might be all that is required to complete our archaeological or palaeoenvironmental inference. In other situations, we might need off-the-shelf but mechanistic chronological models which capture (in part) the mechanisms that led to the chronological observations we have made. In some situations, however, none of the models on offer in existing off-the-shelf software is suitable for our needs, and then we must ask a statistician and/or software developer to construct a tailor-made, probably mechanistic, model for us.

FORMAL CHRONOLOGY CONSTRUCTION

All modelling clearly involves making choices. Generally, we start by selecting a medium, framework or paradigm in which to construct the model, and other choices then follow. In statistical modelling, our first choice is between the Frequentist (or Classical) and the Bayesian paradigm. Frequentist statistics is based on an interpretation of probability that defines an event's probability as the limit of its relative frequency in a very large number of trials. Bayesian

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

statistics—named after Thomas Bayes (1701–1761)—is a paradigm in which evidence about the fixed but unknown state of the world is expressed in terms of *degrees of belief*, formalised as personal probability statements (see, e.g., Kyburg & Smokler, 1980; Lindley, 1965; Press & Tanur, 2001; Robert, 2007). Because of this formalisation, the Bayesian paradigm is ideally suited to the representation and management of expert opinion and prior knowledge, as well as scientific data, making it particularly appealing to archaeologists and palaeoenvironmental scientists looking for a coherent way to draw together information from several disparate sources.

Buck et al. (1996) make the case for the use of the Bayesian paradigm in archaeology and we do not recap those arguments here. Instead, we consider the circumstances we find ourselves in when constructing chronological models and propose that, for all but the simplest problems, the Bayesian paradigm seems most natural.

Appropriate chronological models

Just as with the elephant modelling problem, in order to choose the appropriate nature and scale of model for our archaeological or palaeoenvironmental research project, we need to know precisely what we are modelling and why. For example, if we are seeking to date a single event in the archaeological record, like the death of an individual human whose articulated skeleton has been found in a well-sealed grave, then a Frequentist approach might be suitable. We could (in theory) repeatedly date the same event by sending multiple samples from the skeleton to the laboratory for dating and then summarise the results within the Frequentist statistical paradigm. However, if we want to date a sequence of events (e.g., stratigraphy from an archaeological site or a sediment core) then the relative chronological knowledge needs modelling and, probably, some expert chronological knowledge/opinion too. In that case we will surely need the Bayesian paradigm. If we want to construct a chronology for a whole site or landscape then multiple sequences need comparing and combining: several experts might be involved, working at different times and/or in different places, and so the model must be modular and will almost certainly need to account for personal probability statements from multiple experts. In such circumstances, the Bayesian paradigm is the only one that allows a mathematical representation of the causal flow, enabling a coherent propagation of uncertainty and a robust interpretation of all of the interrelated sources of information simultaneously.

Pictures as chronological models

Almost all chronologists who are drawn towards the Bayesian paradigm are motivated in part by the desire to combine scientific dating and relative or absolute prior knowledge. Stratigraphic information is the most common such knowledge from archaeological excavations and is usually held in the research archive in the form of sketches or plots: section drawings, phase diagrams, Harris matrices or age/depth plots. These are all pictorial models of relative chronology and are a key starting point in chronological modelling. Figure 2 offers an example section drawing from an imagined archaeological site, which we use to illustrate how such information can be utilised as the starting point for chronological modelling.

The first step in creating a formal chronological model is to simplify the stratigraphic drawing and to focus purely on the temporal information it contains. The sketch in Figure 3a shows just such a simplification for the illustrative section in Figure 2. Figure 3b shows a simplification, with a focus only on contexts that contain samples that could be submitted for chronometric dating.

By drawing the two sketches in Figure 3 side by side, we highlight a key choice that all chronological modellers must make—that is, which contexts and samples to include. Most modellers

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

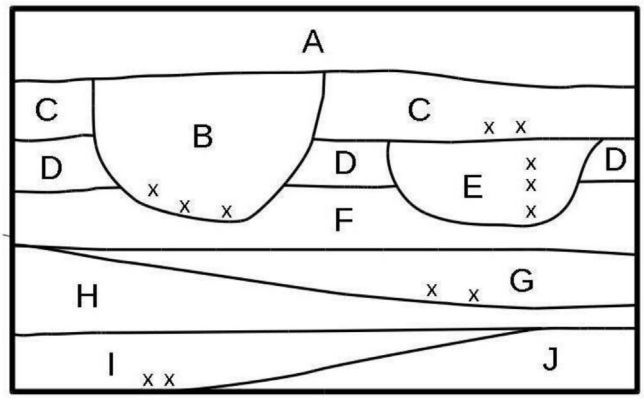
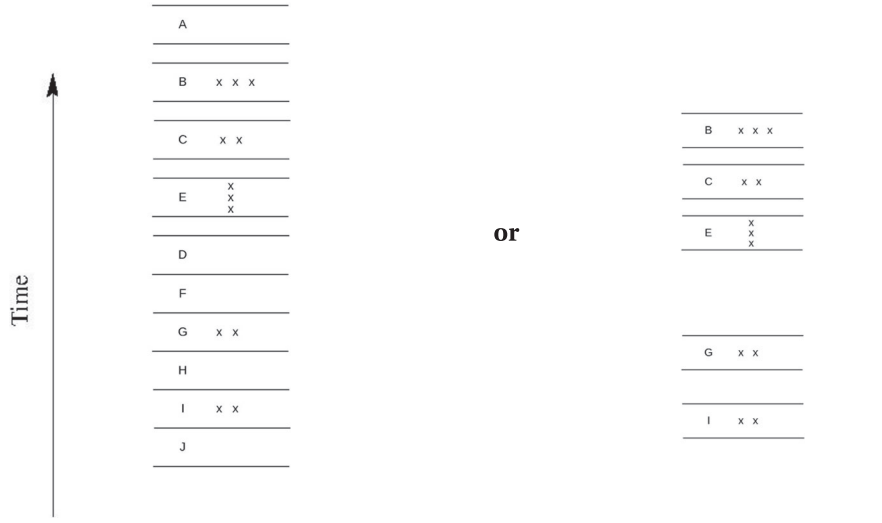


FIGURE 2 An illustrative section drawing from part of an (imagined) archaeological site. The locations of samples suitable for chronometric dating are indicated by crosses.



(a) *First simplification* **(b)** *Second simplification*

FIGURE 3 Sketches of the relative chronological information contained within the stratigraphic profile shown in Figure 2. Horizontal lines represent archaeological context boundaries and crosses indicate the location within the stratigraphic sequence of samples suitable for chronometric dating. The profile on the left shows all of the contexts in Figure 2; the one on the right shows only the contexts that contain samples suitable for absolute dating.

are agreed that, all other things being equal, we should follow Occam's razor and keep the model as simple as possible. This seems like good, straightforward advice, but in practice all other things are seldom equal—by which we mean that excluding or including contexts or samples from our chronological model, and precisely how we represent the ones we do include, will almost always have at least some impact on the results we get.

Given this, and the fact that there are many more such choices we need to make as we undertake the implementation process, in the later sections of this paper we offer some general guidance for those seeking to make responsible use of Bayesian chronological modelling software. Before we can do that, however, we need to (a) clearly identify the key chronological components that we wish to manage or interpret and (b) think and write rather more formally,

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

thus constructing statistical models. We address (a) in the next section and move to (b) in Section 4. In doing so we draw on a large body of existing literature, but, in particular, Naylor and Smith (1988), Buck et al. (1996), Blackwell and Buck (2008), and citations therein.

Key components of a statistical chronological model

There are broadly two types of information to be represented in a statistical model for chronology construction: relative and absolute. Relative chronological information typically relates to the (prior) ordering of events, whereas absolute chronological information usually arises from historical records or from scientific dating methods.

In the remainder of this paper, we focus on absolute dates that arise from radiocarbon dating, but Bayesian methods have also been developed for dendrochronology (Hassan et al., 2019; Jones, 2013; Litton & Zainodin, 1991; Millard, 2002), luminescence dating (Li et al., 2023; Zink, 2015) and electron spin resonance dating (Millard, 2006). We will also centre what we say around seeking to manage and interpret the chronological information represented in Figures 2 and 3. Bayesian models now exist to represent a considerably wider range of chronological features than those needed for this purpose, but our goal here is to be introductory rather than comprehensive, and we encourage the interested reader to study further work by the authors cited herein.

There are two key types of chronological event in Figures 2 and 3: those that relate to directly datable objects (like the deposition of the samples indicated by crosses) and those, like the creation of context boundaries, that do not. Absolute date estimates for the context boundaries can only be obtained by modelling their relationship with the datable objects, via the stratigraphic sequence. So, in summary, we need model components to represent the following:

- the underlying dates we wish to learn about, only some of which relate directly to datable objects;
- stratigraphic relationships between the underlying dates of all components of the stratigraphic record;
- the relationship between the underlying dates and radiocarbon determinations, including laboratory uncertainties and the necessary calibration.

In the next section we look at all three of these, starting with a simple model which includes only directly datable objects and their associated stratigraphic relationships, and then moving to include context boundaries and their stratigraphic relationships to the datable objects. In so doing, we suppose that the datable samples in Figures 2 and 3 (identified by crosses) give rise to the radiocarbon determinations indicated in Table 1.

MODELS FOR BAYESIAN RADIOCARBON DATING

Since the focus in this section is on formal statistical modelling, some readers may find it daunting. For those who do, we suggest that you focus on appreciating the notation used and on the general structure of the equations provided. It is not essential to understand the details of the equations to gain insight into the structure of the models and the general nature of the way in which they are constructed, and it is these that are the most important. Given this, we offer Figure 4, which we hope readers will use to follow the structure of the model as it is described. The top part of Figure 4 relates to the ideas in Sections 4.1 and the lower part to those in Section 4.2. We would like to draw your attention to some key aspects of this Bayesian model:

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

TABLE 1 Radiocarbon determinations assumed to be associated with the illustrative stratigraphic sequence in Figure 2, along with an indication as to whether or not we are assuming each sample to be stratified within the relevant context. Samples that are known a priori to be stratigraphically older are presented below those that are younger.

Context	Sample label	Stratified within context	Mean ^{14}C age	Lab error
B	θ_1	No	9,700	30
	θ_2	No	9,670	30
	θ_3	No	9,650	30
C	θ_4	No	9,520	30
	θ_5	No	9,530	30
E	θ_6	Yes	9,730	30
	θ_7	Yes	9,700	30
	θ_8	Yes	9,670	30
G	θ_9	No	9,800	30
	θ_{10}	No	9,800	30
I	θ_{11}	No	9,960	30
	θ_{12}	No	10,000	30

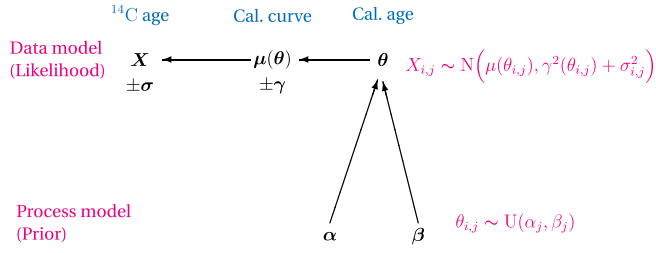


FIGURE 4 Pictorial representation of the (hierarchical) statistical model developed by Naylor and Smith (1988). Here we represent the relationship between chronological parameters, $\boldsymbol{\theta}$ and $\boldsymbol{\sigma}$, that relate directly to data, $\boldsymbol{X}$, and those used to represent the underlying archaeological processes, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$. To emphasise the process relationships, the arrows represent the causal direction, which is the opposite direction from the one in which we make inferences.

- **Parameters**—Models are parameterised. Some of the parameters represent quantities of interest while some others are included only for modelling convenience. An example of a parameter is θ, which represents the true unknown calendar age.
- **Levels**—Any hierarchical Bayesian model consists of a number of levels or hierarchies. There is a top and a bottom level, and each is related only to its immediate neighbours. These relations are directed and go from bottom to top. Note that in Figure 4 there are just two levels, but hierarchical Bayesian models can (and in real applications usually) have many levels. Of course, model levels should not be confused with the layers in a stratigraphic archaeological sequence.
- **Data model**—The top layer in the model represents the data generation mechanism—that is, the parametric model we have decided to use to represent the available data.
- **Uncertainty**—Parameters in a statistical model are unknown. Each layer in a Bayesian hierarchical model identifies the sources of uncertainty for its parameters and measures them using probability distributions.
- **Propagation**—Systematic use of Bayes theorem and the hierarchical structure enables the coherent propagation of uncertainty through a Bayesian model.

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

It is easiest to understand Figure 4 if one starts from the top of the diagram and considers a single organic sample with true but unknown calendar age, $X_{i,j}$ BP. The sample (and the parameters associated with it) are identified by the index i, which is just one of n_j samples found in a particular archaeological context, j, which in turn is one of many such contexts (a total of J, say) found on a particular archaeological site. So, given that, $i = 1,2,\ldots,n_j$ and $j = 1,2,\ldots,J$, the total number of samples available across the whole site are $n = \sum_{j=1}^{J} n_j$. Suppose that our single sample derives from an organism that ceased metabolising on a true (but unknown) calendar date $\theta_{i,j}$ cal BP. Given that the proportion of radioactive carbon atoms in the Earths atmosphere has not been constant over time, we must use the calibration curve, $\mu(\theta_{i,j})$, to map between radiocarbon and calendar ages. The current estimate of the calibration curve at $\theta_{i,j}$ has an associated uncertainty denoted by $\gamma(\theta_{i,j})$. For much more about the radiocarbon calibration curve and how it is estimated, see Reimer et al. (2020) and other papers in the same volume.

When an estimate of $X_{i,j}$ is made by a radiocarbon laboratory, we write that estimate $x_{i,j}$, and the associated laboratory standard error we represent by $\sigma_{i,j}$. Taken together, $x_{i,j}$ and $\sigma_{i,j}$ are often referred to as a radiocarbon determination. We can then use prior knowledge from the stratigraphic evidence at our excavation to further enhance our chronological model. Towards the bottom right of Figure 4, in the lower level of the hierarchical model, we assume that calendar date $\theta_{i,j}$ lies in the interval α_j cal BP to β_j cal BP with every date in that interval taken as equally likely (i.e., $\theta_{i,j}$ is assumed to be Uniformly distributed between α_j and β_j).

The equation at the top right of Figure 4 (in the top level of the model) is the data generation mechanism, commonly referred to as the likelihood. Following the widely adopted convention, we are assuming that the true radiocarbon date, $X_{i,j}$, is located around $\mu(\theta_{i,j})$, with some variability described by a Normal distribution with a standard deviation that derives both from the error on the current estimate of the radiocarbon calibration curve and on the laboratory standard error on the radiocarbon determination. Given the form of Normal probability distribution function, this assumption leads to the following (probability) distribution for $X_{i,j}$:

$$\frac{1}{\sqrt{2\pi\left(\sigma_{i,j}^2 + \gamma(\theta_{i,j})^2\right)}} \times \exp\left[-\frac{\left(x_{i,j} - \mu(\theta_{i,j})\right)^2}{2\sigma_{i,j}^2 + 2\gamma(\theta_{i,j})^2}\right].$$

As we will see below, this equation can be combined with the hierarchical information represented in Figure 4, using Bayes theorem and, in so doing, we are then able to manage coherently all of the information in a single, joint representation of the complete collection of the chronological information from a given archaeological site. Once we have done that, we can then provide probabilistic statements not only about one calendar age, but all n calendar ages. In the central part of Figure 4 and elsewhere we represent that full set of calendar ages using the vector $\boldsymbol{\theta}$. In a similar way, the full set of all of the J context boundary dates is represented as $\{\boldsymbol{\alpha},\boldsymbol{\beta}\}$.

We are now almost ready to get into mathematical details relating to the use of Bayes theorem but, before we do that, it is worth highlighting a few notational conventions. We are using Roman letters to represent observable or known quantities and Greek letters to represent unknown components of statistical models, which are parameters. Some parameters, like $\theta_{i,j}$ and α_j, take single numerical values (known as scalars), and we indicate these using standard fonts. Parameters defined using a bold font, like $\boldsymbol{\theta}$ and $\boldsymbol{\alpha}$, indicate a fixed-length sequence of scalars, known as a vector. Since vectors contain a set of values, we often undertake calculations systematically for each entry in the vector. For example, if $\boldsymbol{y}$ is a vector of length u (i.e., contains a sequence of u scalars), then to multiply the elements in $\boldsymbol{y}$ together, we would write $\prod_{m=1}^{u} y_m$, where Π indicates multiply and y_m is the m-th element of $\boldsymbol{y}$—that is, a scalar.

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

We will also need to use the term 'probability distribution function' (or 'probability density'), which refers to a mathematical function whose value at any given point provides a probabilistic statement about the parameter represented by the function. The familiar bell-shaped curve of the Normal probability distribution function is one that readers will probably be able to call to mind. Such distributions can be used one-at-a-time to represent individual, independently varying parameters. More commonly in statistical models, however, we work with collections of such functions that are interrelated or covarying. When we do this, we typically undertake calculations on several variables simultaneously (or jointly) so that we can keep track of their interrelationships. Then, when we create plots of the final results, we usually focus on variables one at a time, since visualising high-dimensional distributions is notoriously difficult. Such single-variable representations are referred to as marginal (as opposed to joint) probability distribution functions.

The basic Bayesian radiocarbon model

We start by considering just the top level of our hierarchical model. Recall that a radiocarbon determination has two parts: a radiocarbon date estimate before present (BP) and an associated laboratory error. We want to use such determinations to learn about the (true underlying) calendar date(s), on which one or more samples sent for dating ceased metabolising. Sets of such radiocarbon ages, laboratory errors and calendar dates are represented by the vectors $\boldsymbol{x}$, $\boldsymbol{\sigma}$ and $\boldsymbol{\theta}$, respectively. To help us we often have prior information (historical, stratigraphic, etc.) about the calendar date, which we represent with the (prior) probability density $p(\boldsymbol{\theta})$.

In order to use the data and the prior information to learn about the calendar date, we need to formalise the link between the information we have—that is, $\boldsymbol{x} \pm \boldsymbol{\sigma}$ BP and $p(\boldsymbol{\theta})$, and what we want to learn about—that is, $p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\sigma})$ (note that the symbol | can be read as 'given', so this is the probability density of the unknown parameter $\boldsymbol{\theta}$ given that we know the values $\boldsymbol{x}$ and $\boldsymbol{\sigma}$). Since the proportion of radioactive carbon atoms in the Earths atmosphere has not been constant over time, we need a calibration curve to map between radiocarbon and calendar ages. It is this curve that we refer as $\mu(\boldsymbol{\theta})$. Hence, strictly, we want to learn about $p(\boldsymbol{\theta}|\boldsymbol{x},\boldsymbol{\sigma},\mu(\boldsymbol{\theta}))$.

Focusing on context E from Figure 2, for example, we have three radiocarbon determinations from a stratigraphic sequence (see Table 1), and so $\boldsymbol{x} = (9670, 9700, 9730)$ BP and $\boldsymbol{\sigma} = (30, 30, 30)$ and the true underlying dates associated with them are $\boldsymbol{\theta} = (\theta_6, \theta_7, \theta_8)$ cal BP. The stratigraphic information allows us to be sure (a priori) that the three calendar dates are strictly ordered, $\theta_8 > \theta_7 > \theta_6$ cal BP. Note that here and in what follows (where it does not introduce ambiguity), we have simplified the notation by dropping the second subscript on $\theta_{i,j}$, since we are focusing on data and results only from context E.

Then, using the standard statistical model for radiocarbon (motivated above, but described in deal more detail in Buck et al., 1996, chap. 9):

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{i=a}^{b} \exp\left[-\frac{1}{2}\frac{(x_i - \mu(\theta_i))^2}{\sigma_i^2 + \gamma(\theta_i)^2}\right],$$

where $\gamma(\theta_i)$ is the standard deviation on the current internationally agreed estimate of the calibration curve at θ_i and, in our example $a = 6$ and $b = 8$. Note that in the formulation in Naylor and Smith (1988) and Buck et al. (1996) the parameter $\gamma(\theta_i)$ is not present, since those authors assumed this to be very small relative to σ_i, which was reasonable at the time, but no longer holds because laboratory errors have reduced considerably.

The (prior) stratigraphic information can then be formalised as

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$p(\boldsymbol{\theta}) \propto \mathrm{I}(\theta_i) = \begin{cases} 1 & \theta_{i+1} > \theta_i > \theta_{i-1} \\ 0 & \text{otherwise.} \end{cases}$$

and we thus finally arrive at a statement of Bayes' theorem that is useful for basic chronology construction:

$$\begin{aligned} p(\boldsymbol{\theta}|x) &\propto p(x|\boldsymbol{\theta}) \times p(\boldsymbol{\theta}) \\ &\propto \prod_{i=a}^{b} \exp\left[-\frac{1}{2}\frac{(x_i - \mu(\theta_i))^2}{\sigma_i^2 + \gamma_{\mu(\theta_i)}^2}\right] \times \mathrm{I}(\boldsymbol{\theta}). \end{aligned} \quad (1)$$

If carefully computed, this gives us a posterior distribution for our calendar age vector, $\boldsymbol{\theta}$, given the radiocarbon data, the prior stratigraphic information and the modelling assumptions outlined above. From this we could obtain any relevant estimates. Such derivations are, however, not straightforward, since the mathematical integration calculations needed cannot be made exactly, even with very powerful computers.

Instead, simulation-based methods (known as Markov chain Monte Carlo, MCMC, methods) are used, within which we generate estimates of groups of related parameters (e.g., θ_6, θ_7 and θ_8 in our toy example) all together, ensuring as we do so that all of the modelling constraints are obeyed simultaneously (details of such methods are given in Buck et al., 1996, and discussed below). Applying such methods, even to very large stratigraphic sequences, is now straightforward using off-the-shelf Bayesian radiocarbon calibration packages such as OxCal (Bronk Ramsey, 2009), ChronoModel (Lanos & Philippe, 2017, 2018) or BCal (Buck et al., 1999). For illustration, we used BCal with the IntCal20, internationally agreed, estimate of the radiocarbon calibration curve (Reimer et al., 2020) to calibrate the three stratified determinations from context E in Figure 2 and obtained the results shown in Figure 5, where black indicates the probability distribution function obtained when each determination is calibrated individually (without inclusion of the prior knowledge about the dates derived from the stratigraphy) and ochre indicates the results from computing the Bayesian posterior density function in Equation (1) (which does include the stratigraphic evidence). We discuss below the extensions to the model needed to provide the green and pink lines in Figure 5.

This simple Bayesian model is clearly powerful, allowing inclusion of basic stratigraphic information that would otherwise have been ignored or handled in an ad hoc manner. In so doing, we typically obtain more precise date estimates than those that can be obtained by calibrating determinations individually. For context E, the gain in precision is fairly modest because we have only three determinations that are temporally rather widely spaced. With more data in a similar time interval, however, the gains can be considerably more substantial. Despite these benefits, the model in Equation (1) does not have sufficient complexity to allow us to include all of the features of the stratigraphy in Figures 2 or 3, since there is no way to represent events for which we do not have direct dating evidence. In particular, we have no way to represent the concept of depositional history, as recorded by the relationships between archaeological contexts.

Modelling deposition

To represent boundaries between different depositional contexts or phases, Naylor and Smith (1988) introduced two further parameters to the model, $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ (both cal BP) with $\alpha_j > \beta_j$ (i.e., α_j and β_j are, respectively, the early and late boundary dates for context j). Typically, we have some relative chronological information, which relates such parameters to datable

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

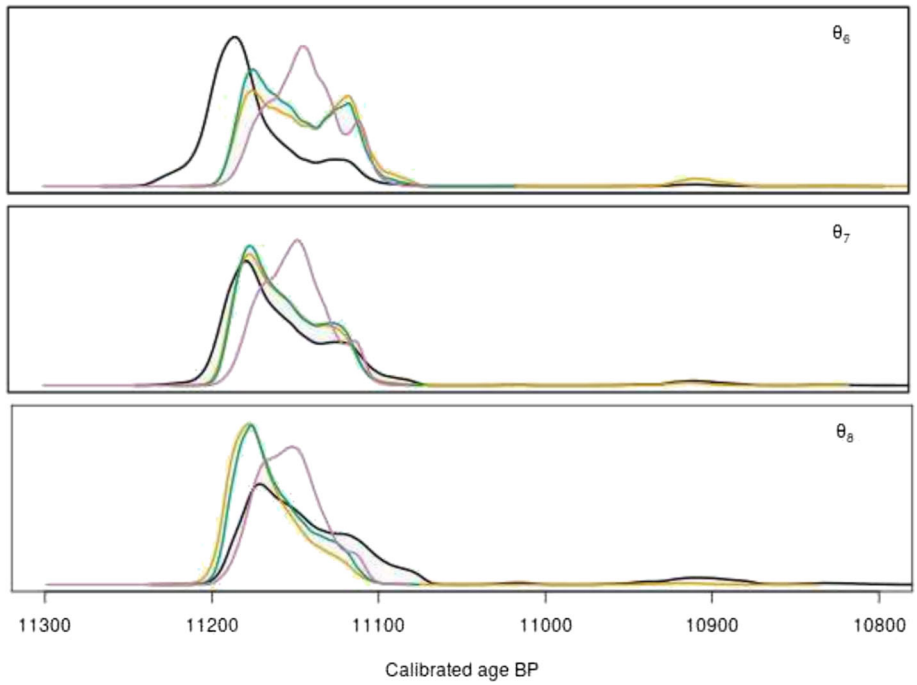


FIGURE 5 Results of calibrating the three radiocarbon determinations represented in context E: black when all stratigraphic info is ignored; ochre when ordering only within this context is included (using equation 1); green when ordering within this context is included *and* all of the other contextual relationships shown in the right-hand panel of Figure 3 are modelled (using equation 2); pink when ordering within this context is included *and* all of the other contextual relationships shown in the left-hand panel of Figure 3 are modelled (using equation 2). Note, first, that we are showing only the outline of the marginal probability density in each case, so that all four distributions can clearly be seen, and, second, that as we add more stratigraphic detail to our model the probability densities become increasingly peaked (i.e., the date estimates are more precise).

material, but there is no direct scientific dating evidence associated with them. This situation is very common, of course, since in archaeology and palaeoenvironmental research we seldom find datable material directly associated with all of the key locations in our stratigraphic sequences.

Naylor and Smith (1988) thus amended the model in Equation (1) so that, for contexts with no internal stratigraphy (and hence no a priori ordering of the dates of the samples),

$$p(\boldsymbol{x}|\boldsymbol{\theta}) \propto \prod_{j=1}^{J} \left[\left(\alpha_j - \beta_j\right)^{-n_j} \prod_{i=1}^{n_j} z_{i,j} \mathrm{I}_B\left(\theta_{i,j}\right) \right],$$

where J is the number of contexts or phases in the model, n_j is the number of datable samples in context or phase j,

$$z_{i,j} = \exp\left[-\frac{\left(x_{i,j} - \mu\left(\theta_{i,j}\right)\right)^2}{2\sigma_{i,j}^2 + 2\gamma\left(\theta_{i,j}\right)^2} \right]$$

and

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

$$\mathrm{I}_B(\theta_{i,j}) = \begin{cases} 1 & \beta_j \geq \theta_{i,j} \geq \alpha_j \\ 0 & \text{otherwise.} \end{cases}$$

Assuming that the deposition rate for material in each context is constant, but allowing varying deposition rates between contexts, they then modelled the prior knowledge as

$$p(\boldsymbol{\theta}) \propto I_A(\boldsymbol{\alpha},\boldsymbol{\beta}) = \begin{cases} 1 & \boldsymbol{\alpha},\boldsymbol{\beta} \in A \\ 0 & \text{otherwise,} \end{cases}$$

where A is the set of values of $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ that satisfy the prior chronological (e.g., stratigraphic) information. Thus obtaining calibrated dates via

$$p(\boldsymbol{\theta}|\boldsymbol{x}) \propto \mathrm{I}_A(\boldsymbol{\alpha},\boldsymbol{\beta}) \times \prod_{j=1}^{J} \left[(\alpha_j - \beta_j)^{-n_j} \prod_{i=1}^{n_j} z_{i,j} I_B(\theta_{i,j}) \right]. \quad (2)$$

In situations where we have stratigraphic ordering of samples within a context (as in context E), it is computationally straightforward to add the relevant parts of Equation 1 when we implement Equation 2 in software. It is also trivial to allow for contexts or phases for which there is stratigraphic information, but no direct dating evidence at all (i.e., pairs of α_j,β_j that have a direct relation only to other boundary parameters and not to any samples that can be directly dated). It is also relatively straightforward to amend the details of the equations we present above to allow users greater flexibility about how they model the relationships between the boundary dates and the dates for individual contexts. One widely adopted example is provided by Nicholls and Jones (2001), who pointed out some weaknesses in the models we outline above when large numbers of radiocarbon dates are available.

OxCal (Bronk Ramsey, 2009), ChronoModel (Lanos & Philippe, 2017, 2018) and BCal (Buck et al., 1999) between them offer tools to implement a wide range of such models, and we adopt the latter to formalise the information in Figure 3. When we do this we are, of course, embedding context E within a larger model for the whole of the stratigraphic sequence shown in Figure 2. Nonetheless, for illustration, we can focus just on the (marginal) results for the dates of the samples within that context, thus producing the green and pink lines in Figure 5. Taken together, all four lines in Figure 5 suggest that as we add more stratigraphic detail (and hence more parameters) to our model we (appear to) learn increasingly precisely about the dates for the samples in context E.

Superficially, this seems attractive since greater precision is almost always the goal of chronologists; however, we need to be cautious here. Surely there must be a point at which adding more contexts to our model, without adding any more absolute dating evidence, leads to a false sense of extra information. Does knowing that there are an extra two contexts between contexts G and E really provide very much extra chronological information (given that neither is directly associated with any absolute dating evidence) and, if so, is our statistical modelling of the information in Figure 4 capturing it well? This and other implementation questions are discussed in the next section but, before we move to that, we look first at some of the estimates obtained for the dates of the context boundaries in our example.

Given standard model selection advice (summarised above and discussed further below), all other things being equal, we prefer simple over more complex models and so Figure 6 shows the posterior estimates of the boundary dates for the contexts in the right-hand sequence of Figure 3. Similar results could be shown for the left-hand sequence, should there be a strong archaeological reason for doing so. What we see in Figure 6 is the clear impact of the temporal ordering imposed by the prior distribution on the boundary parameters, which were derived

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

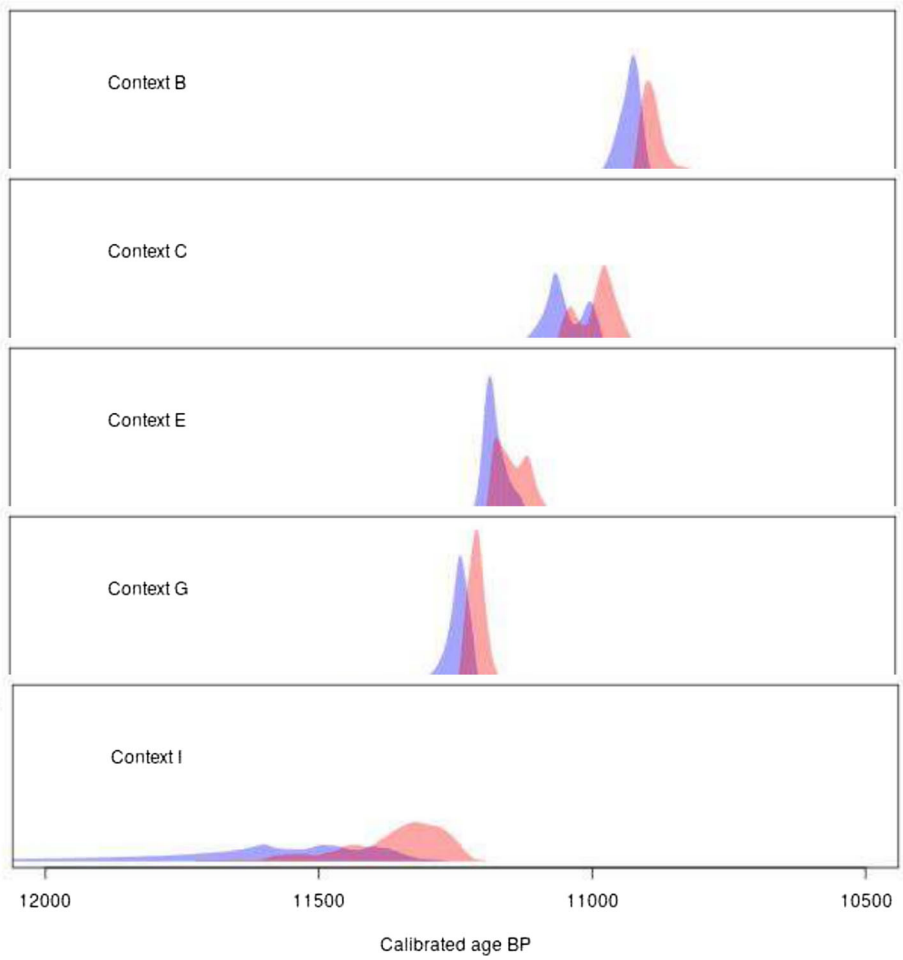


FIGURE 6 Estimated calibrated context boundary dates for the contexts shown in the right-hand panel of Figure 3. Early boundary dates for each context are in blue and late ones in red.

from the context relationships shown in Figure 3. We also find that we can date each of the context boundaries to within at most 500 years and, in the case of context B, to within 100 years, despite the fact that we have no direct absolute dating evidence for any of these parameters. Clearly, this is a very powerful archaeological inference tool, even for this very simple stratigraphic sequence. Before using such methods to aid in the interpretation of real archaeological projects, however, we must return to some of the inferential and modelling decisions related to their use, particularly the modelling dilemma outlined in the previous paragraph.

MODELLING AND IMPLEMENTATION DECISIONS

Model selection

When Naylor and Smith (1988) first proposed the models in Equations (1) and (2) they implemented them using a technique known as numerical integration, which is time consuming to program, and the resulting code is computationally demanding. Taken along with the fact

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

that, at the time, computational power was extremely limited, this meant that only the simplest of archaeological models could be implemented. The rapid increase in adoption of Bayesian chronological modelling did not come, therefore, until several years later when coders started to adopt the MCMC-simulation-based methods, outlined above. These are relatively easy to code initially and also to modify when new components are needed, thus making them more attractive for modular modelling problems of the sort discussed here, and highly amenable to the implementation of generalisable algorithms like those used in OxCal, ChronoModel and BCal.

With modern computational power and these faster, more flexible implementation tools, a range of software options are now available to chronological modellers. This has led to an ever-increasing choice of modelling options, and selecting precisely which model to implement is thus a routine dilemma. Currently, the only widely used formal model choice tools for chronologists are those offered by OxCal (Bronk Ramsey, 2009). These compare the likelihoods for the individual radiocarbon determinations (when their stratigraphic relations with one another are ignored) with the posterior probability densities obtained from a selection of more complex models (in which the stratigraphic information is included). Of the more complex models those with posterior densities closest to the likelihoods for the individual radiocarbon determinations are preferred.

Such methods are somewhat arbitrary, however, and best suited to comparing models of similar structure and size, and so are not ideal for addressing the dilemma we faced at the end of Section 4.2. For that we really need a formal model selection technique, and these are notoriously difficult to construct, especially when the numbers of parameters vary greatly between the various representations, as can easily be the case with large archaeological sequences. For this reason, model selection is an art that Bayesian chronological modellers need to acquire and which, like most arts, can only really be learnt by experience.

The present authors have some experience of such modelling and, in general, with Occam's razor in mind, we prefer models with fewer parameters over those with more. So, in the case of the dilemma in the previous section, we preferred the model represented by the right-hand panel of Figure 2, over that on the left. We do this because the simpler model allows representation of the key depositional history for the datable samples, but does not explicitly include those contexts that contain no absolute dating evidence and so add relatively little extra information to the inference process.

We might quite readily make a different choice, however, if the archaeological event that we most needed to date related to one of the contexts that contain no datable samples. For illustration, suppose that context F were a destruction layer between two periods of human activity—then we may well decide to include context F in the model, accepting that the total amount of chronological information we have available has not increased very much, but allowing us to do our best to date the context about which we are most eager to learn.

Responsible use of software

Given what we have said about implementation of chronological modelling software, coding it is clearly not a task for a novice and so we assume that readers of this paper will be using one of the off-the-shelf packages like OxCal, ChronoModel or BCal. These are very powerful tools when used carefully, but they do come with some responsibilities. These relate to gaining at least an intuitive insight into what models the software is implementing and how. This paper provides a basic introduction to some aspects, but users should also read the specific background literature cited by the software providers and the explanations offered in the user manuals.

We all also have responsibilities when writing up modelling work. It is vital, for example, that readers are not simply told which model (or software) was used, but that they are also

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

offered a clear explanation as to why the model in question was selected for the current application. Typically, as in the examples above, some reasons will relate to the author's interpretation of the archaeology and some to theoretical or practical constraints. Such details are important because scientific dating evidence is often collected before models are specified, and so the only way that readers can be sure that the same sources of information were not used multiple times is for all authors to offer clear justification for their choices of model and prior and that these do not rely on any aspect the scientific dating evidence.

Users should also know enough about how the software implements their model, so that they can make principled decisions and report clearly on them. This is important because MCMC-based software uses simulation and so each time it is run (slightly) different results are obtained (for more details see, e.g., Brooks et al., 2011; Gilks et al., 1995; Lambert, 2018; and references therein). Key concepts here include the following:

- Burn-in—samples discarded from the beginning of MCMC simulations, which may not be very representative of the posterior density. This is necessary because all simulations start from somewhat arbitrarily selected values that might not have high posterior probability.
- Convergence—the desired state of the MCMC sampler, from which samples can reliably be used to estimate the posterior. The burn-in phase is intended to ensure that the samples we store are from this part of the simulation.
- Thinning and effective sample size—the process of storing only a subset of MCMC samples (thinning) to leave a subset that is smaller than the total but nonetheless conveys equivalent information (and so has the same effective sample size). Subsetting the output in this way is most important when parameters of our model are highly correlated (as is often the case in stratigraphic sequences), and so the simulated values for each parameter change only very slowly between neighbouring steps of the Markov chain.

Most software providers offer automated convergence checking tools that provide guidance about the amount of burn-in and thinning required. However, users should understand intuitively the checks being conducted and record the choices made in their write-up so that others may replicate and/or build appropriately on their work. The OxCal, ChronoModel and BCal manuals both offer more advice on these issues.

Since MCMC methods are simulation-based, users must check reproducibility by making multiple runs for each model, with different start values for the sampling chains, to check that the results obtained for key parameters are the same to an *appropriate* level of precision. *Appropriate* is, of course, a relative term. We will typically require a different level of accuracy in historic periods than the palaeolithic, for example. Reproducibility experiments, and the accuracy to which we report results, should reflect this.

Since applied Bayesian statistics involves so many personal judgements, users should also explore the sensitivity of results to the choices they made. For each application users should vary models and/or priors, to explore other plausible options, and report resulting changes in the posteriors densities, just as we did here when we considered two formalisations of the stratigraphic information in Figure 2. Such checks are essential, since without them we have no idea how robust results are. If they are not robust to key decisions, which for large or complex models is not unusual, considerable further exploration may be needed as to why particular choices were made.

OTHER CURRENT AND FUTURE OPTIONS

There are a considerable number of modelling options now available to users of off-the-shelf Bayesian chronological modelling software, including several that are much more sophisticated

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

than the ones we focused on for our illustrations above. We will not attempt to discuss here all—or even very many—of these. The following are worth highlighting, however:

- model extensions to allow inclusion of a wide range of absolute prior knowledge not discussed here—for example, that derived from historical documents or classical texts and/or from scientific dating evidence other than radiocarbon;
- a range of deposition models—for example, of the gradual colonisation of a site (Jones, 2013; Lee & Bronk Ramsey, 2012) and of peat or sediment accumulation (Blaauw & Christen, 2005, 2011; Bronk Ramsey, 2008; Christen et al., 1995; Haslett & Parnell, 2008);
- detection of outliers in radiocarbon dating (Bronk Ramsey, 2008; Christen, 1994).

These are in routine use and are implemented in several freely available software packages, in particular OxCal (Bronk Ramsey, 2009).

Other desirable features not currently available in any of the standard software are as follows:

- Automated selection of samples during an incremental radiocarbon dating program in order to use the dating budget cost effectively (as first suggested in Christen & Buck, 1998; Buck & Christen, 1998). OxCal already has a feature, known as R_Simulate, which allows users to generate likely additional radiocarbon dates for potential samples within an existing model. This is the first step towards automated selection of any further samples that might be dated, but until the software can work systematically through all of the possible combinations of samples that might be dated (for a given budget) and automatically provide users with some measure of cost-effectiveness, it is unlikely to be widely taken up. Given the very large dating projects now undertaken (particular on government-funded and/or high-profile excavations) and the value-for-money requirements of many funding agencies, adding this functionality to at least one of the standard packages would seem like a good investment of time.
- Extensions of the purely temporal Bayesian chronological models to include spatial components, thus creating a spatio-temporal modelling framework. Such extensions are desirable in projects relating to the spread of animals, plants, humans or ideas in time and space. However, they are computationally very demanding to implement since they require data from entire landscapes to be analysed simultaneously. As a result, there are two commonly adopted approaches to such problems. The simplest is to spatially partition the data and then analyse the spatial groups within a purely temporal model (as, for example, in Blackwell & Buck, 2003). A more sophisticated approach is to develop a mechanistic (typically deterministic) model for the spatial process (such as demographic spread) and then to use statistical methods to compare the resulting spread patterns with the available chronological evidence using formal statistical methodology (as, for example, in Baggaley, Sarson, et al., 2012; Baggaley, Boys, et al., 2012).
- Tools to automate production of pictures like those in Figure 4, from archaeological site databases, and then to generate from them chronological models of the sort explored herein. Dye and Buck (2015) provide a proof of concept about how such a tool might be developed, using a refinement of Harris matrices and techniques adopted from graph theory. A recent PhD thesis (Moody, 2023) provides a first attempt at the kind of software that would be needed to implement such an approach, but there is considerable work to do before such software would be ready for routine use.
- Improvements to age–depth models to allow for geomorphology. At present, such models effectively assume that the cores from lake sediments were derived from cylindrical lake basins. In practice, we do not know the morphology of most of the lakes from which cores are taken, but we can be fairly sure that they are not cylindrical, and recent work by Bennett and Buck (2016) shows that basin geomorphology can have a considerable impact on the

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

age–depth relationships. Given this, more work is needed to (a) find a good, cost-effective way to establish basin morphology and (b) develop the existing Bayesian age–depth models to take account of the information obtained.

Clearly, such extensions would take chronology construction considerably beyond the relatively simple but powerful models and methods outlined above. Some may eventually be fairly readily added to the general-purpose chronological modelling software like OxCal, ChronoModel and BCal, but others—in particular, the spatio-temporal modelling options—require inclusion of parameters with a completely different structure from those used thus far. As a consequence, at least for the moment, those wishing to adopt such models will need to learn not just modelling skills but computer programming too. The chronological modelling community is short of such skills and we hope that one or two readers of this paper might already have programming skills and be keen to help, or might be willing to learn them in order to do so. We would be delighted to hear from anyone interested in such work and to encourage them to help us to make the next thirty years of chronological modelling as productive as the last.

PEER REVIEW

The peer review history for this article is available at https://www.webofscience.com/api/gateway/wos/peer-review/10.1111/arcm.12998.

DATA AVAILABILITY STATEMENT

Data sharing is not applicable to this article as no new data were created or analyzed in this study.

ORCID


Caitlin E. Buck https://orcid.org/0000-0002-0872-9504

REFERENCES

Baggaley, A. W., Boys, R. J., Golightly, A., Sarson, G. R., & Shukurov, A. (2012). Inference for population dynamics in the Neolithic period. *Annals of Applied Statistics*, *6*, 1352–1376. https://doi.org/10.1214/12-AOAS579

Baggaley, A. W., Sarson, G. R., Shukurov, A., Boys, R. J., & Golightly, A. (2012). Bayesian inference for a wavefront model of the Neolthisation of Europe. *Physical Review E*, *86*, 016105. https://doi.org/10.1103/PhysRevE.86.016105

Bayliss, A. (2007). Bayesian buildings: An introduction for the numerically challenged. *Vernacular Architecture*, *38*, 75–86. https://doi.org/10.1179/174962907X248074

Bayliss, A., & Marshall, P. (2022). *Radiocarbon dating and chronological modelling: Guidelines and best practice*. Historic England.

Bennett, K. D., & Buck, C. E. (2016). Interpretation of lake sediment accumulation rates. *Holocene*, *26*(7), 1092–1102. https://doi.org/10.1177/0959683616632880

Blaauw, M., & Christen, J. A. (2005). Radiocarbon peat chronologies and environmental change. *Journal of the Royal Statistical Society: Series C: Applied Statistics*, *54*(4), 805–816.

Blaauw, M., & Christen, J. A. (2011). Flexible paleoclimate age-depth models using an autoregressive gamma process. *Bayesian Analysis*, *6*, 457–474. https://doi.org/10.1214/ba/1339616472

Blackwell, P. G., & Buck, C. E. (2003). The late glacial human reoccupation of north western Europe: New approaches to space-time modelling. *Antiquity*, *77*(296), 232–240. https://doi.org/10.1017/S0003598X00092231

Blackwell, P. G., & Buck, C. E. (2008). Estimating radiocarbon calibration curves. *Bayesian Analysis*, *3*, 225–248.

Bronk Ramsey, C. (2008). Deposition models for chronological records. *Quaternary Science Reviews*, *27*(1–2), 42–60. https://doi.org/10.1016/j.quascirev.2007.01.019

Bronk Ramsey, C. (2009). Bayesian analysis of radiocarbon dates. *Radiocarbon*, *51*(1), 337–360. https://doi.org/10.1017/S0033822200033865

Brooks, S., Gelman, A., Jones, G., & Meng, X. L. (2011). *Handbook of Markov chain Monte Carlo*. CRC Press. https://doi.org/10.1201/b10905

Buck, C. E., Cavanagh, W. G., & Litton, C. D. (1996). *The Bayesian approach to interpreting archaeological data*. Wiley.

Buck, C. E., & Christen, J. A. (1998). A novel approach to selecting samples for radiocarbon dating. *Journal of Archaeological Science*, *25*, 303–310. https://doi.org/10.1006/jasc.1997.0207

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License

Buck, C. E., Christen, J. A., & James, G. N. (1999). BCal: An on-line Bayesian radiocarbon calibration tool. *Internet Archaeology*, *7*. http://intarch.ac.uk/journal/issue7/buck/

Buck, C. E., & Juárez, M. A. (2020). Modelización bayesiana de radiocarbono para principiantes. In *Cronométricos en Historia y Arqueología* (pp. 293–310). Dextra Editorial.

Christen, J. A. (1994). Summarizing a set of radiocarbon determinations: A robust approach. *Applied Statistics*, *43*(3), 489–503. https://doi.org/10.2307/2986273

Christen, J. A., & Buck, C. E. (1998). Sample selection in radiocarbon dating. *Applied Statistics*, *47*, 543–557.

Christen, J. A., Clymo, R. S., & Litton, C. D. (1995). A Bayesian approach to the use of ^{14}C dates in the estimation of the age of peat. *Radiocarbon*, *37*(2), 431–442. https://doi.org/10.1017/S0033822200030915

Dye, T. S., & Buck, C. E. (2015). Archaeological sequence diagrams and Bayesian chronological models. *Journal of Archaeological Science*, *83*, 84–93.

Gilks, W. R., Richardson, S., & Spiegelhalter, D. (1995). *Markov chain Monte Carlo in practice*. Chapman & Hall/CRC Interdisciplinary Statistics. CRC Press. https://doi.org/10.1201/b14835

Haslett, J., & Parnell, A. (2008). A simple monotone process with application to radiocarbon-dated depth chronologies. *Applied Statistics*, *57*, 399–418.

Hassan, M. M., Jones, E., & Buck, C. E. (2019). A simple Bayesian approach to tree-ring dating. *Archaeometry*, *61*(4), 991–1010. https://doi.org/10.1111/arcm.12466

Jones, E. (2013). *Practical Bayesian Dendrochronology*. PhD thesis, University of Sheffield.

Kyburg, E., & Smokler, H. E. (1980). *Studies in subjective probability*. Wiley.

Lambert, B. (2018). *A stundent's guide to Bayesian statistics*. SAGE.

Lanos, P., & Philippe, A. (2017). Hierarchical Bayesian modeling for combining dates in archeological context. *Journal de la Société Française de Statistique*, *158*, 72–88.

Lanos, P., & Philippe, A. (2018). Event date model: A robust Bayesian tool for chronology building. *Communications for Statistical Applications and Methods*, *25*, 131–157.

Lee, S., & Bronk Ramsey, C. (2012). Development and application of the trapezoidal model for archaeological chronologies. *Radiocarbon*, *54*(1), 107–122. https://doi.org/10.2458/azu_js_rc.v54i1.12397

Li, B., Jacobs, Z., & Roberts, R. G. (2023). A Bayesian hierarchical age model for optical dating of single grains of quartz. *Quaternary Geochronology*, *77*, 101455. https://doi.org/10.1016/j.quageo.2023.101455

Lindley, D. V. (1965). *Introduction to probability and statistics* (Vol. 2). University Press. https://doi.org/10.1017/CBO9780511662973

Litton, C. D., & Zainodin, H. J. (1991). Statistical models of dendrochronology. *Journal of Archaeological Science*, *18*, 429–440. https://doi.org/10.1016/0305-4403(91)90036-O

Millard, A. (2002). Bayesian approach to sapwood estimates and felling dates in dendrochronology. *Archaeometry*, *44*(1), 137–143. https://doi.org/10.1111/1475-4754.00048

Millard, A. (2006). Bayesian analysis of ESR dates, with application to border cave. *Quaternary Geochronology*, *1*(2), 159–166. https://doi.org/10.1016/j.quageo.2006.03.002

Moody, B. (2023). *Semi-Automation of Bayesian Chronology Construction Using a Graph-Theoretic Approach*. PhD thesis, University of Sheffield.

Naylor, J. C., & Smith, A. F. M. (1988). An archaeological inference problem. *Journal of the American Statistical Association*, *83*(403), 588–595.

Nicholls, G., & Jones, M. (2001). Radiocarbon dating with temporal order contraints. *Applied Statistics*, *50*(4), 503–521.

Press, J., & Tanur, J. M. (2001). *The subjectivity of scientists and the Bayesian approach*. Wiley. https://doi.org/10.1002/9781118150634

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kromer, B., Manning, S. W., Muscheler, R., … Talamo, S. (2020). The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kbp). *Radiocarbon*, *62*(4), 725–757. https://doi.org/10.1017/RDC.2020.41

Robert, C. P. (2007). A defense of the Bayesian choice. In *The Bayesian choice: From decision-theoretic foundations to computational implementation* (pp. 507–530). Springer.

Zink, A. (2015). Bayesian analysis of luminescence measurements. *Radiation Measurements*, *81*, 71–77. https://doi.org/10.1016/j.radmeas.2015.04.009

How to cite this article: Buck, C. E., & Juárez, M. A. (2024). Bayesian radiocarbon modelling for beginners. *Archaeometry*, 1–18. https://doi.org/10.1111/arcm.12998

14754754, 0, Downloaded from https://onlinelibrary.wiley.com/doi/10.1111/arcm.12998 by University Of Sheffield, Wiley Online Library on [20/06/2024]. See the Terms and Conditions (https://onlinelibrary.wiley.com/terms-and-conditions) on Wiley Online Library for rules of use; OA articles are governed by the applicable Creative Commons License