RADIOCARBON, Vol 51, Nr 1, 2009, p 337–360 © 2009 by the Arizona Board of Regents on behalf of the University of Arizona

BAYESIAN ANALYSIS OF RADIOCARBON DATES

Christopher Bronk Ramsey

Research Laboratory for Archaeology, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford, OX1 3QY.
Email: christopher.ramsey@rlaha.ox.ac.uk.

ABSTRACT. If radiocarbon measurements are to be used at all for chronological purposes, we have to use statistical methods for calibration. The most widely used method of calibration can be seen as a simple application of Bayesian statistics, which uses both the information from the new measurement and information from the ^{14}C calibration curve. In most dating applications, however, we have larger numbers of ^{14}C measurements and we wish to relate those to events in the past. Bayesian statistics provides a coherent framework in which such analysis can be performed and is becoming a core element in many ^{14}C dating projects. This article gives an overview of the main model components used in chronological analysis, their mathematical formulation, and examples of how such analyses can be performed using the latest version of the OxCal software (v4). Many such models can be put together, in a modular fashion, from simple elements, with defined constraints and groupings. In other cases, the commonly used "uniform phase" models might not be appropriate, and ramped, exponential, or normal distributions of events might be more useful. When considering analyses of these kinds, it is useful to be able run simulations on synthetic data. Methods for performing such tests are discussed here along with other methods of diagnosing possible problems with statistical models of this kind.

INTRODUCTION

Much that will be said here would be equally applicable to any dating technique. However, 3 aspects of radiocarbon dating have led to the extensive development of statistical analysis methods and tools. The first is the very wide availability of ^{14}C dates, which means that, in many applications, large number of measurements have to be considered together. Another, more fundamental reason is that ^{14}C dates are not dates at all but measurements of an isotope ratio. To interpret them as dates at all requires some form of statistical analysis using a calibration curve. Furthermore, once calibrated, ^{14}C dates have probability density functions that are not normally distributed and, therefore, many of the standard methods of classical statistics cannot be applied.

The earliest consequence of this was the development of probabilistic methods for calibration in various calibration programs (Stuiver and Reimer 1993; van der Plicht 1993; Bronk Ramsey 1994) and the very important early development of a new statistical framework for dealing with multiple ^{14}C dates undertaken by the Nottingham University statistics group quickly followed (Buck et al. 1991, 1992, 1994, 1996; Christen 1994; Christen and Litton 1995; Christen et al. 1995). At that time, the underlying statistics for the most commonly used models were developed. Indeed, there has been a general consensus over the most appropriate mathematical models to be applied to single and multiple phases. Other statisticians (Nicholls and Jones 2001) have suggested refinements, but on the whole the general approach has remained the same.

When these methods were first developed, the analysis of even quite small models took considerable computing power. However, with the development of computer hardware over the last decade and a half, most desktop and laptop computers can now handle large and sophisticated models. Although the mathematical equations underlying the models in common use are relatively simple, few have the desire to implement them by hand. For this reason, the development of widely available software, such as OxCal (Bronk Ramsey 1995), BCal (Buck et al. 1999), and DateLab (Jones and Nicholls 2002), to perform such analyses has been important in promoting their adoption.

The purpose of this paper is to provide an overview of the methods underlying much of the Bayesian analysis undertaken today. The underlying themes are relevant to analyses performed using any soft-

© 2009 by the Arizona Board of Regents on behalf of the University of Arizona

ware. However, for simplicity the illustrations of the models will be taken from the latest version of OxCal, partly because of the familiarity of the author with the package and partly because the code for the models can be succinctly expressed.

BAYES' THEOREM: THE UNDERLYING METHODOLOGY

We need to start with the mathematical basis for all of the analyses discussed here. Bayes' theorem tells us that:

$$p(\mathbf{t}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{t})p(\mathbf{t}) \quad (1)$$

where $\mathbf{t}$ is the set of parameters and $\mathbf{y}$ the observations or measurements made. $p(\mathbf{t})$ is the *prior* or the information about the parameters (mostly events) we have apart from the measurements. $p(\mathbf{y}|\mathbf{t})$ is the *likelihood* for the measurements given a set of parameters. $p(\mathbf{t}|\mathbf{y})$ is the *posterior* probability, or the probability of a particular parameter set given the measurements and the prior. In a Bayesian analysis, we have to express the information we have about the chronology in these terms.

For any statistical analysis, we need to formulate a mathematical model for the underlying chronology. This immediately forces us to make certain assumptions that need to be borne in mind when considering the results. The first assumption made is that the chronology can be broken down into events—that is, single points on a continuous timescale. This does not mean that more continuous processes cannot be modeled, but it does mean that those processes need to be described in terms of events. For example, a phase of activity in an archaeological site, or a period of deposition in a sedimentary sequence, is bounded by a start event and an end event. In other cases, the events might be more randomly sampled from a more continuous process like deposition in a sedimentary sequence. We will look later at how this works in practice and how some of the apparent limitations of this approach can be overcome.

Events occur on a well-defined timeline. In our analysis, we wish to find the date of the events of interest, and the intervals between them. In some periods, the precision we need is subannual (in the post-bomb era of ^{14}C dating), but for most periods we are dealing in years, decades, or centuries. So, although it is usually sufficient to work in integer years, it makes sense for any analysis to be on a floating point timescale, which can give any required precision. CaliBomb (Reimer et al. 2004) and OxCal both use a fractional year timescale where, for example, 2000.1 is early in AD 2000 and 2000.9 is late in the same year. OxCal uses the same floating point timescale internally for all periods (see Figure 1).

There are two main types of date information available to us in the study of chronology: calendar and relative. Calendar date information, which essentially puts events on an "absolute" timescale, can come from historical records or from scientific dating methods. Relative date information usually comes in the form of stratigraphy or from a broader study of the context. Ultimately, both types of information come from observations made on the material studied and there is no intrinsic reason to trust one type of information more than the other. The main functional difference between these two types of information is that the calendar date information usually relates to the age of single samples, whereas the relative date information provides a much more complex relationship between the timing of events. For this reason, Bayes' theorem is normally applied in chronological analysis so that the calendar date information is expressed as the *likelihood* and the relative date information as the *prior*. This usually makes conceptual sense given that the relative date information is normally available before any direct dating takes place. However, ultimately the distinction is somewhat arbitrary and one can simply see the statistical methods as a way of combining all of this information together.

The main question to be addressed in most analyses is what the most appropriate priors are. There is no easy answer to this. Some might argue that we should not include anything but the dating information and so the prior becomes a constant, independent of the dates of any of the events. However, as we shall see ("Groupings" section below) this null prior:

$$p_0(\mathbf{t}) \propto constant \tag{2}$$

is really quite inappropriate in most cases where we have many dates. This is essentially why most attempts to analyze ^{14}C dates without a proper formal model give misleading results and is perhaps why, when asked to look at a series of calibrated ^{14}C dates from a single phase, almost everyone will instinctively overestimate their spread.

Age models can be expressed in purely mathematical terms. However, when defining and using age models it is useful to have a convenient way of summarizing them that can be easily read, published, and shared between collaborators. For this reason, the OxCal program uses a higher-level notation (or Chronological Query Language, CQL: Bronk Ramsey 1998) to encapsulate the models. This notation has been further developed for OxCal v4 (CQL2, hereafter just referred to as CQL) and now uses a C or ECMAScript format:

$$command(parameters)\ \{further\ information\};$$

CALENDAR DATE INFORMATION

Ideally, we have direct evidence for the date of an event. This is either as an actual historical date or, more often, in terms of scientific measurements, a date with some associated uncertainty.

Some other methods provide ages in a more convenient form than ^{14}C dating. This is because ^{14}C is really not an calendar dating technique at all. We make measurements and have to make use of the the data in the calibration data sets in order to interpret them in calendar form. Many other dating techniques give either age ranges (e.g. from historical records) or means with standard uncertainties. In some cases (notably older samples in uranium-series dating), the errors are not normally distributed. In any case, we introduce such information as a probability density function, which is treated in the Bayesian analysis as a likelihood. Before we look at these likelihood functions, we first of all need to define the timescale over which they operate.

Timescale

Central to any specification of calendar age information is an agreed timescale, and this is an area of some confusion. In ^{14}C, cal BP is used to mean before AD 1950, but in other dating methods BP is also sometimes used to simply mean before the measurement was made. Where the uncertainties are large, this usually does not matter, but it is unfortunate that such confusion continues. The reason is that people prefer to think in terms of ages rather than absolute points on a calendar timescale, but unfortunately "age" is a continually shifting scale. Those who need higher precision such as historians and archaeologists working in the historical period use the calendar BC/AD or BCE/CE scale, which has no year zero and is well defined. For dates within a year, the calendar becomes very complicated before the modern era due to different local practices. Astronomers and other scientists who require very good precision over long timescales use Julian years, though confusingly not from the same starting point as the original Julian calendar with a year length of 365.25 days; for convenience they also define fraction years of the form 2000.5 to mean half-way through the year 2000. There is also a well-defined calendar, ISO-8601, which can be extended to cover all times and based on the Gregorian calendar; this is now the underlying standard for most computer applications. As with

astronomical year numbering, this does include the year zero. In ISO-8601 or the Gregorian calendar, the average year length is 365.2425 days. Where high precision is needed, such as with post-bomb ^{14}C dating, this seems the best basis for a calendrical timescale (see Figure 1).

Representations of the time-scale

Floating point timescale (G):

```
 .    .    .    .    .    .    .
-2   -1    0    1    2    3    4
```

Division into conventional years BC/AD:

```
 .     .     .     .     .     .     .
 | 3BC | 2BC | 1BC | AD1 | AD2 | AD3 |
```

As plotted on axes:

```
 .     .     .     .     .     .     .
 3     2     1  1BC/1AD  2     3     4
```

calBP timescale:

```
    .     .     .     .     .     .
 1952  1951  1950  1949  1948  1947
```

Expressing dates

Expression	Floating point date (G)	Meaning
`2008.0`	2008.0	the start of the year AD 2008
`AD(1066)`	1066.5	the middle of AD 1066
`AD(1950)`	1950.5	the middle of AD 1950
`BC(12)`	-10.5	the middle of 12 BC
`CE(1812)`	1812.5	the middle of 1812 CE
`BCE(79)`	-77.5	the middle of 79 BCE
`CE(-78)`	-77.5	the middle of 79 BCE or ISO-8601 year -78
`calBP(100)`	1850.5	the middle of AD 1850
`calBP(0)`	1950.5	the middle of AD 1950

Figure 1 Floating point timescale and date expressions used in OxCal: Any analysis needs to be performed on a well-defined scale. OxCal uses a floating point timescale based on the ISO-8601 calendar. 1.0 is defined as 00:00:00 UT on Monday 0001-01-01 and the year length is 365.2425 days. This timescale is very close to the constantly changing Besselian fractional year scale (denoted by a prefix B) and is close to the Julian fractional year system (denoted by a prefix J) in current use by astronomers. As it is based on the Gregorian year length, which is very close to the variously defined astronomical years, it keeps phase with the seasons for the last 100,000 yr (by which point it is about a month out). Essentially, the same timescale is used in CaliBomb for the post-bomb calibration data sets. The cal BP timescale, which is really only defined in terms of integer years, is taken to be the number of Gregorian years prior to the middle of AD 1950. A prefix of G is used to distinguish it from fractional Besselian or Julian timescales.

Probability Density Functions

As soon as we make measurements, or other observations, that tell us about the timing of an event, this makes some dates much more likely than others. This is usually expressed in terms of a probability density function (see Figure 2). To take a simple example, if we make an observation y_i that implies that an event t_i took place in England within the reign of William the Conqueror, we know that it must take place at sometime between 25 December 1066 and 9 September 1087. We would give such an event a uniform probability density function:

$$p(y_i|t_i) \sim U(1066.995,1087.703) \quad (3)$$

In many scientific dating techniques, the results are given as ages (before 1950, or sometimes before measurement, expressed as BP) with normally distributed errors. For example, the Ar-Ar date for the eruption of the Laacher See Tephra 29,500 ± 560 (van den Bogaard 1995) can be represented by a normal distribution:

$$p(y_i|t_i) \sim N(1995 - 29500, 560) \tag{4}$$

When we come to ^{14}C, the measurement is made on the isotopic composition of the sample, not on the age of the sample. To convert this to an age, we need to go through the process of calibration. The most basic form of calibration is the intercept method, which just gives a range, or set of ranges. The probability method in use in most calibration programs has some advantages (Bowman and Leese 1995) and is essentially Bayesian in nature. We define the actual ^{14}C ratio of the sample for an event t_i to be ρ_i. We also have information y_i from the measurement, which is usually expressed as a mean and standard uncertainty $r_i \pm s_i$. This gives:

$$p(y_i|\rho_i) = \frac{1}{s_i\sqrt{2\pi}}\exp(-(r_i - \rho_i)^2/(2s_i^2)) \tag{5}$$

Note that we can express the ^{14}C measurement here either in terms of a ^{14}C age (BP, Stuiver and Polach 1977) or a ratio ($F^{14}C$, Reimer et al. 2004). For measurements close to background, however, the errors are approximately normal for the ratio, but are far from normal for the ^{14}C age. For this reason, calibration of older dates should always be made in reference to the ratio. In addition to the measurement, we have other (prior) information about ρ_i, which comes from the calibration curve, defined in its functional form as $r(t) \pm s(t)$. The prior is normally taken to be:

$$p(\rho_i, t_i) \propto \frac{1}{s_i(t_i)}\exp(-(\rho_i - r(t_i))^2/(2s^2(t_i))) \tag{6}$$

This is a suitable prior because if we integrate over ρ_i we then get a constant value, independent of t_i, so it is neutral with respect to date. This prior reflects that fact that some ^{14}C measurements (notably those on the plateaus) are more likely than others. Given this, we can now integrate out the parameter ρ_i, which we do not need:

$$p(y_i|\rho_i, t_i) \propto \frac{1}{s_i s(t_i)}\exp(-(r_i - \rho_i)^2/(2s_i^2))\exp(-(\rho_i - r(t_i))^2/(2s^2(t_i))) \tag{7}$$

$$p(y_i|t_i) \propto \int_{\rho_i=-\infty}^{\infty} p(y_i|\rho_i, t_i)d\rho_i \tag{8}$$

$$\propto \frac{\exp\left(-\frac{(r_i - r(t_i))^2}{2(s_i^2 + s^2(t_i))}\right)}{\sqrt{s_i^2 + s^2(t_i)}} \tag{9}$$

This is the standard calibration equation and is usually treated as the likelihood for the event association with the ^{14}C measurement, although it already includes the prior information about the calibration curve. This complication needs to be addressed in some cases, most particularly where several ^{14}C dates are all made on the same sample, in which case the measurements must be combined before calibration so that the same prior calibration curve information is not included more than once in the analysis (see Bronk Ramsey 2009 for further discussion of the combination of ^{14}C dates).

As far as single events go, this is all that we need. The likelihood defines the probability of obtaining a measurement given a particular date for an event. If we only have a single event, we normally take the prior for the date of the event to be uniform (but unnormalized):

$$p(t_i) \sim U(-\infty, \infty) \sim constant \tag{10}$$

and so, from Bayes' theorem, the posterior probability density is the same as that for the likelihood. So for a single ^{14}C measurement, we can say that:

$$p(t_i|y_i) \propto p(y_i|t_i)p(t_i) \propto \frac{\exp\left(-\frac{(r_i - r(t_i))^2}{2(s_i^2 + s^2(t_i))}\right)}{\sqrt{s_i^2 + s^2(t_i)}} \tag{11}$$

See Figure 2 for an explanation of how likelihoods are defined in OxCal.

RELATIVE CONSTRAINTS

One thing that we will need to consider in most models is constraints on the order of events. We frequently have information about the ordering of events either from stratigraphy or more generally from our understanding of the chronology. To introduce a single constraint $t_a < t_b$ between 2 event parameters, we introduce an element in the prior:

$$p_H(t_a, t_b) = \begin{cases} 1, \text{if } t_a < t_b \\ 0, \text{otherwise} \end{cases} \tag{12}$$

This function is labeled p_H because it is the same as the Heaviside function $H(t_b - t_a)$ except in the case where $t_a = t_b$. For multiple constraints in a sequence, we define a prior along the same lines:

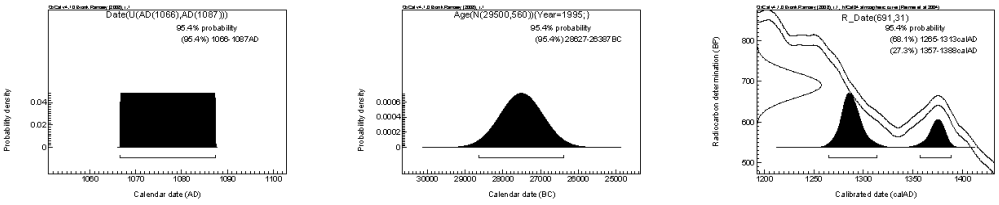
$$p_H(t_a, t_b, t_c, \ldots) = \begin{cases} 1, \text{if } t_a < t_b < t_c < \ldots \\ 0, \text{otherwise} \end{cases} \tag{13}$$

Any constraints that we wish to include in a prior model are simply further factors of this form.

Rather than dealing with constraints between individual events, groups of events are often considered together in terms of sequences and phases (see Figure 3), but such constraints can all be broken down and expressed mathematically in the same way.

GROUPINGS

Although a suitable prior for a single ^{14}C date is fairly easy to find, it is much less obvious what a suitable prior is for a larger number of events. If we have 2 events t_b and t_b, and we use a simple constant prior over all time, then it is interesting to look at the implications for the interval $s = |t_b - t_a|$. The effective prior for our parameter s is also uniform from 0 to ∞, which seems fairly satisfactory. However, if we have more events $t_1 \ldots t_n$, the span of these events has an effective prior, which is $\propto s^{n-2}$, which is very highly weighted to higher values when n is large. This is clearly very unsuitable in most cases as demonstrated by Steier and Rom (2000) and discussed in Bronk Ramsey (2000). This means that for any more than 2 events we need to find a more appropriate prior.



CQL code	Examples	Explanation
General probability density functions		
`N(`μ`,`σ`)`	`N(100,10)`	Normal distribution
`U(`a`,`b`)`	`U(90,110)`	Uniform distribution in range $a \rightarrow b$
`Top_Hat(`μ`,`δ`)`	`Top_Hat(100,10)`	Uniform distribution in range $(\mu - \delta) \rightarrow (\mu + \delta)$
`T(`ν`)`	`T(5)`	Student-t distribution
`Exp(`τ`,`a`,`b`)`	`Exp(-1,0,10)`	Exponential distribution in range $a \rightarrow b$
`Prior(`*file*`)`	`Prior("Stockpiling")`	Pre-defined distribution
Special chronological functions		
`Date(`*expression*`)`	`Date(U(AD(1066),AD(1087)))`	Generic date likelihood
`Age(`*expression*`)`	`Age(N(29500,560)){Year=1995;}`	Age distribution
`R_Date(`r`,`s`)`	`R_Date(691,31)`	Calibrated radiocarbon date
`C_Date(`μ`,`σ`)`	`C_Date(1066,10)`	Normal likelihood
`R_Simulate(`μ`,`σ`)`	`R_Simulate(691,31)`	Simulated radiocarbon date
`C_Simulate(`μ`,`σ`)`	`C_Simulate(1066,10)`	Other simulated date
Parameter definition and naming with likelihood		
function`(...)`	`R_Date(3000,30)`	parameter with no name
function`("`*name*`",...)`	`R_Date("a",3000,30)`	parameter with name
name = *expression*	`a = R_Date(3000,30)`	parameter with name
Parameter definition and naming without likelihood		
function`(...)`	`Date()`	parameter with no name
function`("`*name*`",...)`	`Date("a")`	parameter with name
name = *expression*	`a = Date()`	parameter with name
Parameter referencing		
function`("=`*name*`",...)`	`Date("=a")`	cross-reference
name	`a`	cross-reference
function`("=`*name*`",...)`	`R_Date("=a",3000,30)`	likelihood also applied
name `&=` *expression*	`a &= R_Date(3000,30)`	likelihood also applied

Figure 2 Probability density functions in OxCal: OxCal models are all expressed in terms of a Chronological Query Language (CQL), which is a development of a system proposed in Bronk Ramsey (1998). Probability density functions are usually used as likelihoods for model parameters. OxCal will normally create a new parameter for each probability distribution defined unless the parameter is specifically cross-referenced.

Uniform Phase Model

The usual way to deal with this problem is to introduce the notion of the uniform phase, which has proved to be a very useful conceptual framework for this kind of analysis. There are different ways of going about deriving essentially the same equations (see Buck et al. 1991, 1992; Zeidler et al. 1998 for the original work on this). Perhaps the easiest starting point is to return to the 2-event situation, which is easy to deal with. We define 2 events that delineate the phase: a start event t_a and a finish event t_b. From now on, these will be referred to as boundary events. In this case, we have an

CQL code fragment		Schematic representation	Prior function
Usual	Alternative		

`Sequence()`
`{`
`Date("a");`
`Date("b");`
`Date("c");`
`};`
	`a=Date(); b=Date();`
	`c=Date();`
	`a < b < c;`
		$a \rightarrow b \rightarrow c$	$p_H(t_a, t_b, t_c)$

`Sequence()`
`{`
`Date("a");`
`Phase()`
`{`
`Date("b");`
`Date("c");`
`}`
`Date("d");`
`};`
	`a=Date(); b=Date();`
	`c=Date(); d=Date();`
	`a < ( b | c ) < d;`
		$a \rightarrow \begin{bmatrix} b \\ c \end{bmatrix} \rightarrow d$	$p_H(t_a, t_b, t_d) p_H(t_a, t_c, t_d)$

`Sequence()`
`{`
`Date("a");`
`Before()`
`{`
`Date("b");`
`};`
`Date("c");`
`Date("d");`
`};`
	`a=Date(); b=Date();`
	`c=Date(); d=Date();`
	`a < c < d;`
	`a < b;`
		$a \rightarrow \begin{bmatrix} c \rightarrow d \\ b \end{bmatrix}$	$p_H(t_a, t_c, t_d) p_H(t_a, t_b)$

`Sequence()`
`{`
`Date("a");`
`Date("b");`
`After()`
`{`
`Date("c");`
`};`
`Date("d");`
`};`
	`a=Date(); b=Date();`
	`c=Date(); d=Date();`
	`a < c < d;`
	`c < d;`
		$\begin{matrix} a \rightarrow b \\ c \end{matrix} \rightarrow d$	$p_H(t_a, t_b, t_d) p_H(t_c, t_d)$

Figure 3 Constraints in OxCal are usually expressed in terms of the commands `Sequence`, `Phase`, `Before`, and `After`; the examples shown above (without any likelihood information, for clarity) show how the CQL syntax is used to introduce constraint elements into the prior. They can also be added explicitly using the < and > operators. It should be stressed that such constraints should only ever be used in conjunction with some kind of a grouping model (see Groupings section). The individual fragments above do not form plausible models on their own.

additional constraint (see next section) that $t_a < t_b$, but we define a uniform prior for each of these otherwise:

$$p(t_a, t_b) \propto p_H(t_a, t_b) \tag{14}$$

If we then look at the effective prior for a parameter $s = t_b - t_a$, this is still uniform over the range 0 to ∞, which is suitably vague. Now let us look at a suitable prior for all of the other events t_i in the model. These must lie somewhere between the 2 outer boundary events and so:

$$p(t_i | t_a, t_b) = \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \tag{15}$$

The overall prior for the model then becomes:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} = \frac{\prod_i p_H(t_a, t_i, t_b)}{(t_b - t_a)^n} \tag{16}$$

where there are n events within the phase. This is the prior usually used for a single-phase model in any of the Bayesian analysis software packages, OxCal (Bronk Ramsey 1995), BCal (Buck et al. 1999), or DateLab (Jones and Nicholls 2002). It is sometimes referred to as a "uniform phase" model because of the uniformity of the prior in Equation 15. In most cases, the 2 outer boundary events are not directly dated, but this need not be the case. The important thing about it is that the effective prior for the span of all of the events within the model is constant—that is, the model does not favor short or long phases. However, it does make some assumptions, most notably that all of the events within the group are equally likely to occur anywhere between the start and the end.

Other Distributions

There will be circumstances where the uniform phase model is not appropriate. It is very easy to extend the same formalism to other distributions. If the distribution can be defined in terms of 2 boundary events in the same way as the uniform phase, then we only need to alter the form of Equation 15. As a first example, suppose we know that the events within the phase are likely to grow in number throughout the phase; instead of a uniform prior within the phase we can use a ramped one:

$$p(t_i | t_a, t_b) = 2p_H(t_i, t_b)(t_i - t_a)/(t_b - t_a)^2 \tag{17}$$

and the overall prior for the model becomes:

$$p(\mathbf{t}) \propto p_H(t_a, t_b) \prod_i 2p_H(t_i, t_b)(t_i - t_a)/(t_b - t_a)^2 \tag{18}$$

In this model, all of the dates within the phase are still constrained to lie within the outer boundaries and again the prior for this model is independent of the span of all of the events.

So far, we have considered single-phase models where the underlying processes are defined by 2 events. However, consider now a phase where there is a definite end point t_b, but the events preceding it are distributed exponentially with an unknown time constant. Such might be the case with residual material in a destruction layer, though this is yet to be tested archaeologically. In such a case, we can still define 2 controlling boundary events but use the earlier one to define the time constant $\tau = t_b - t_a$ rather than to limit the distribution. We now replace Equation 15 with:

$$p(t_i | t_a, t_b) = p_H(t_i, t_b) \frac{\exp(-(t_b - t_i)/(t_b - t_a))}{(t_b - t_a)} \tag{19}$$

and the overall prior for the model becomes:

$$p(\mathbf{t}) \propto p_H(t_a, t_b) \prod_i p_H(t_i, t_b) \frac{\exp(-(t_b - t_i)/(t_b - t_a))}{(t_b - t_a)} \tag{20}$$

As a final example, we will consider the case where events are normally distributed, with a gradual onset and tailing off. Again, we can use the same overall formulation, but we will define the 2 char-

acteristic boundaries t_a and t_b to provide the 1σ limits of the distribution. We now replace Equation 15 with:

$$p(t_i|t_a, t_b) = \frac{1}{(t_b - t_a)\sqrt{\pi/2}} \exp\left(\frac{-(2t_i - t_b - t_a)^2}{2(t_b - t_a)^2}\right) \quad (21)$$

and the overall prior for the model becomes:

$$p(\mathbf{t}) \propto p_H(t_a, t_b) \prod_i \frac{1}{(t_b - t_a)} \exp\left(\frac{-(2t_i - t_b - t_a)^2}{2(t_b - t_a)^2}\right) \quad (22)$$

To some extent, these models with different functional forms can be used to deal with underlying processes that do not have definite start and end events. Figure 4 shows the different ways in which such single-phase models can be specified in OxCal and Figure 5 shows an example application.

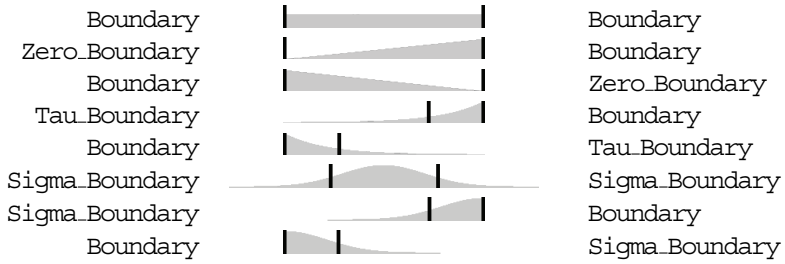


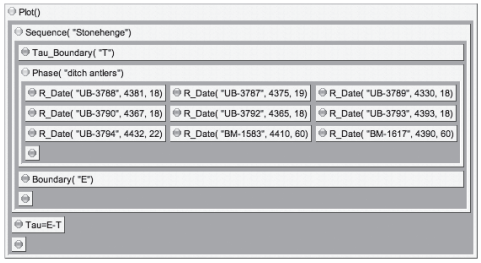
Figure 4 Range of grouping distributions supported in OxCal v4.0. Groupings are defined by using `Boundary` commands within a `Sequence`. All events sandwiched between 2 boundaries (including those in nested phases) are treated as a single group. The type of group is defined by the type of boundary used. A simple `Boundary` at the start and end of a group defines a uniform phase. A `Zero_Boundary` is used to define the start or end of a group where the event rate has a ramped distribution. A `Tau_Boundary` can be used to define an exponentially distributed group and a pair of `Sigma_Boundary` statements, a normal distribution. The latter two types of group allow the events to spill beyond the dates of the boundaries themselves and allow the creation of models of processes that do not have definite start and end events.

The grouping distributions described here are all expressed in terms of 2 parameters. In principle, more complicated distributions can also be defined, which require more parameters for their specification. The most significant of these is perhaps the trapezium model (Karlsberg 2006), which requires 4 parameters and might be appropriate in a number of archaeological situations where you have a slow start and end to a phase. These more complicated models are not yet implemented in OxCal.

Any coherent group of dates can be treated as a single phase if no more information is available. Such a treatment is much more realistic, in almost all cases, than the constant prior $p_0(\mathbf{t})$, which assumes that all of the events are independent. This is because under any of these models, the effective prior for the span of the events is neutral (that is, there is no bias to a longer or shorter span). The constant prior, on the other hand, is very strongly biased to longer spans. In cases where otherwise modeling might not have been considered, the main modeling decision to make is what the most realistic distribution of events is. In many cases, uniform between some start and end event might be most appropriate. In other cases, we might expect our group of events to be clustered but to have a slow onset

and tail, in which case a normal distribution would be a better approximation. In cases where there is ambiguity, it is useful to look at the sensitivity of the model outputs to these different priors.

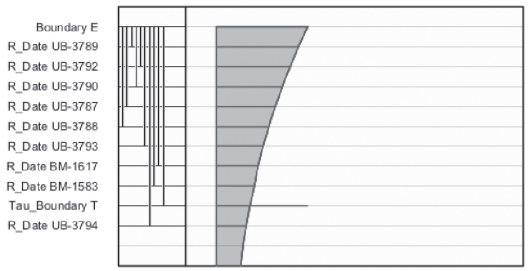
These single-phase models are also important because they form the basis for most of the more complex models.



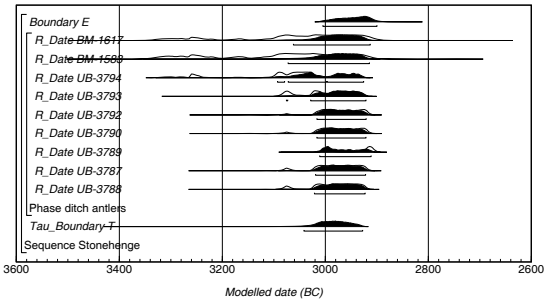
a)

```
Plot()
{
 Sequence( "Stonehenge")
 {
  Tau_Boundary("T");
  Phase("ditch antlers")
  {
   R_Date("UB-3788", 4381, 18);
   R_Date("UB-3787", 4375, 19);
   R_Date("UB-3789", 4330, 18);
   R_Date("UB-3790", 4367, 18);
   R_Date("UB-3792", 4365, 18);
   R_Date("UB-3793", 4393, 18);
   R_Date("UB-3794", 4432, 22);
   R_Date("BM-1583", 4410, 60);
   R_Date("BM-1617", 4390, 60);
  };
  Boundary("E");
 };
 Tau=E-T;
};
```

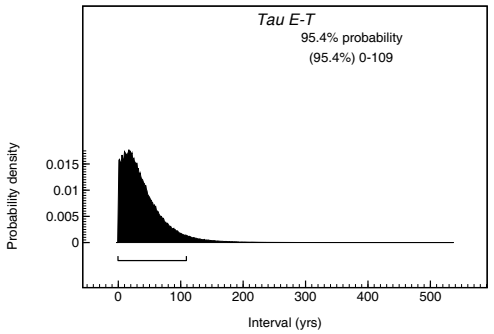
b)



c)



d)



e)

Figure 5 Example of the application of a non-uniform phase model in OxCal. The dates are for antlers used to dig the main enclosure ditch at Stonehenge (Cleal et al. 1995). The assumption here is that the construction is dated by *E* and the antlers are most likely to just predate this event, some possibly being somewhat older. An exponential distribution has been assumed with an unknown time constant τ, which is estimated in the analysis. This also illustrates the main aspects of the modeling process: a) shows the model input as shown on the screen; b) the CQL code for the model; c) the schematic for the model (stratigraphic order); d) the age output plots from the model (stratigraphic order); and e) the estimate for the timescale τ for the acquisition of the antlers. This model is shown for illustration purposes only as it does not include all of the other information from the site.

Multiple Phases

There is very little difficulty in extending the formulation of the single-phase models outlined above to multiple phases or groups. In the simple case of overlapping phases, the different groups can be treated entirely independently. There is no additional mathematical formulation needed. If, for example, we have 2 such phases, the first of which starts at time t_a and ends at time t_b with internal events t_i, and the second of which starts at t_c and ends at t_d with internal events t_j, then the prior is just:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \tag{23}$$

More often, however, we have some relationship between the different groups of events. Sometimes the groups follow one after another with some kind of transition event (contiguous phases). In such a case, the mathematical formulation is essentially the same for each phase, but the same boundary parameter is used for the end of one phase and the start of the next. For 2 such phases, the first of which starts at time t_a and ends at time t_b with internal events t_i, and the second of which starts at t_b and ends at t_c with internal events t_j, the prior is:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_b, t_j, t_c)}{(t_c - t_b)} \tag{24}$$

In other cases, we might expect there to be a hiatus between the phases—that is, they are sequential but not contiguous. In this case, the prior is similar to that for overlapping phases but with a constraint that the start of the second phase should come after the end of the first as in:

$$p(\mathbf{t}) \propto p_H(t_b, t_c) \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \tag{25}$$

In this way, one can see that, with some additional constraints, any multiphase model can be built up by simply multiplying the priors for all of the individual groupings within the model (Figure 6).

At this stage, it is worth discussing the types of applications where these models can be used. The most obvious is single sites where the stratigraphy and taphonomy of the samples is such that well-defined constraints can be imposed on events known to have taken place at a site. This is probably the most robust and powerful use of the method and can be seen in examples such as those discussed in Bayliss and Whittle (2007). In such instances, the prior information is derived from concrete evidence giving information about the chronology. The technique can also be applied in other cases where the chronological framework is much more general. In such cases, the overall framework of the chronology is derived from a much wider range of evidence. This is the case, for example, in the material for the Iron Age of Israel (Boaretto et al. 2005; Mazar and Bronk Ramsey 2008), the chronology of the Aegean Late Bronze Age (Manning et al. 2006), or that of the British Bronze Age (Needham et al. 1998). In such cases, the overall chronological framework is used as a working hypothesis on which the analysis is based. To some extent, the analysis can be used to test whether such a hypothesis is tenable, but more often the results must be seen as depending on the assumptions built into the chronological framework. Often we need to return to the very nature of the models, event-based as they are, and ask whether they really reflect the underlying reality. For example, we might ask, is the transition from Iron I to Iron II in Israel really a synchronous event? On one level, it clearly is not, but since the resolution of our chronologies are typically only decadal at best, if it is a process complete within a few years it might be a reasonable working model to adopt.

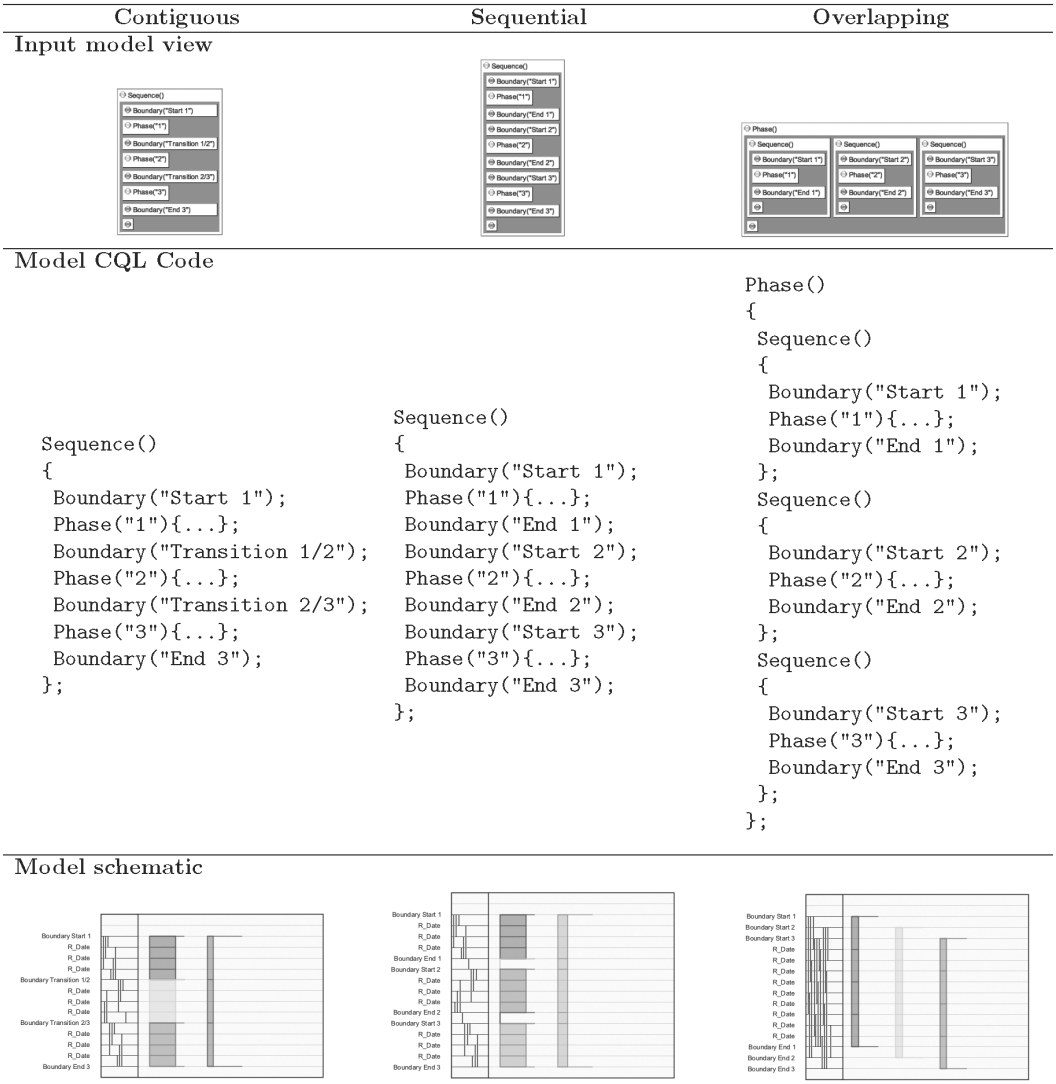


Figure 6 The main types of uniform multiphase model that can be defined in OxCal. The top of this figure shows how the models are usually specified in the program. Below this are the main SQL codes associated with such models (the actual dates within the phases have been left out here for clarity). At the bottom, the schematics for the models are shown. In each schematic, the left-hand side shows the parameters of the model. The lines in the center show the constraints imposed on the parameters and on the right is shown the phasing and the associated distributions. In addition to the 3 phases in the contiguous and sequential examples, one can see how the inner boundaries are treated as a uniform phase within the 2 outer boundaries. These schematic diagrams provide a quick check of whether the model is arranged as intended.

Refinements to Multiphase Models

There are a couple of refinements to the prior that should be considered, especially if there are a large numbers of phases. It should be stressed that these refinements will only make very slight differences in practice to most models. They are included for completeness to maintain the original intention that the prior should be independent of the overall span of the dated events in the model. Our starting point above (“Uniform Phase Model” section) was that the boundaries events them-

selves was taken to be uniform apart from an ordering constraint. The rationale behind this was that there were only 2 of them. If we now have more than 2 boundary events, the prior for the overall span of our modeled events is no longer uniform. We can deal with this issue by treating all of the inner boundaries in a model as part of a single uniform phase for the chronology as a whole, starting with the first boundary t_a and ending with the last one t_m. Each of these inner boundary events t_k has a normalized prior of:

$$p(t_k|t_a, t_m) = \frac{p_H(t_a, t_k, t_m)}{(t_m - t_a)} \tag{26}$$

Thus, in the simple 2-phase example above with contiguous phases, the prior becomes:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_b, t_j, t_c)}{(t_c - t_b)} \frac{1}{(t_c - t_a)} \tag{27}$$

And the 2-phase example with hiatuses becomes:

$$p(\mathbf{t}) \propto p_H(t_b, t_c) \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \frac{1}{(t_d - t_a)^2} \tag{28}$$

This refinement of the multiphase model was first suggested by Nicholls and Jones (2001) and implemented in DateLab (Jones and Nicholls 2002) and OxCal (Bronk Ramsey 2001).

The other factor in the prior of single- or multiphase model is only relevant if there are chronological constraints on the upper and lower phase boundary events (something we assumed was not the case in the "Uniform Phase Model" section). If there are upper and lower limits $ulim_a$, $llim_a$ for t_a and an upper and lower limits $ulim_m$, $llim_m$ for t_m the effective prior becomes proportional to:

$$g(t_a, t_m, llim_a, llim_m, ulim_a, ulim_m) = \min((ulim_m - llim_a) - (t_m - t_a)$$

$$(t_m - t_a) - (llim_m - ulim_a), ulim_a - llim_a, ulim_m - llim_m) \tag{29}$$

and so by adding a factor of the reciprocal of this quantity:

$$p(t_a, t_m|llim_a, llim_m, ulim_a, ulim_m) = \frac{1}{g(t_a, t_m, llim_a, llim_m, ulim_a, ulim_m)} \tag{30}$$

to the prior we maintain a uniform effective prior for the span of all the events $(t_m - t_a)$. A simplified form of this factor was first suggested by Nicholls and Jones (2001), then generalized by Bronk Ramsey (2008) for other types of constraint. A simplified form of this factor is included in models calculated in DateLab and OxCal v3, and the full function used in OxCal v4. In OxCal v4, setting the `UniformSpanPrior` option off (the default is on) will remove these refining elements to the prior.

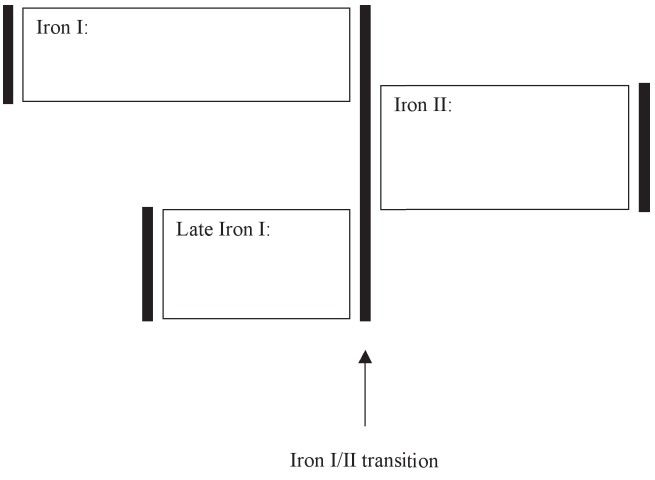
The latter 2 refinements aside, most of the Bayesian analysis computer programs implement the same models for single or multiple uniform phases and, with or without these, can be expected to give very similar results.

Beyond Simple Phases Sequences

In many cases, a multiphase model can be used to describe the activity on a site, the deposition of samples, or more generally the chronology of a region. However, reality is often more complicated than this, and it is useful to be able to construct models that better reflect the underlying processes. The modularity of the formulation of the phase prior helps with this, and in practice most situations can be reasonably well described by linking variables and cross-referencing events between different elements in a model. As an example of this, we will consider the case where we have 2 phases that end at the same point but are different lengths. A schematic for the model is shown in Figure 7 and is taken from Figure 2 of Mazar and Bronk Ramsey (2008). It is useful in this context to consider the mathematical formulation of the prior for the boundaries in such a model. We define the start of the main sequence as t_a, the transition from Iron Age I/II as t_b, and the end of the main sequence as t_c. The other independent boundary is the start of the Late Iron Age I material, which we will call t_d. The main Iron I material has multiple events t_i, and likewise multiple events t_j within Iron II. The late Iron I events are designated as t_k. The basic main model is a 2-phase model as in Equation 27 and we add to that a single-phase model from t_d to t_b and so the overall prior is:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_b, t_j, t_c)}{(t_c - t_b)} \frac{1}{(t_c - t_a)} \prod_k \frac{p_H(t_d, t_k, t_b)}{(t_d - t_b)} \tag{31}$$

The right-hand side of Figure 7 shows how such a model could be specified in OxCal.




```
Plot()
{
 Sequence("Main")
 {
  Boundary("Start");
  Phase("I"){...};
  Boundary("Transition I/II");
  Phase("II"){...};
  Boundary("End");
 };
 Sequence("Late I")
 {
  Boundary("Start Late I");
  Phase("Late I"){...};
  Boundary("=Transition I/II");
 };
};
```

Figure 7 The model described in Figure 2 of Mazar and Bronk Ramsey (2008) for Iron Age material is shown on the left, and on the right how such a model can be expressed in OxCal's CQL notation. The model is essentially a main 2-phase model and a second 1-phase model. The two are tied together by the cross-referenced boundary for the Transition I/II. This ensures that only 1 independent parameter is used for this event.

Cross-referencing between otherwise independent models is a powerful tool as it allows synchronous events across sites or cultures to be modeled using chronological information from a range of sources. As with multiphase models more generally, however, it is important to consider exactly what is meant by synchronous: this is fine for well-characterized tephra horizons, for example, but probably not for vegetation changes related to climate or cultural changes that may spread more gradually. Caution is also needed from a mathematical point of view since it is important that the same elements of the prior are not duplicated by defining an event to be in what are essentially 2 ver-

sions of the same phase. Also, extensive use of cross-referencing can make the MCMC analysis slower. For these reasons, the numbers of cross-references should be kept to the minimum necessary to formulate a suitable model.

SPECIFIC MODELS

So far we have looked at generic, general-purpose models that are applicable to a wide variety of sites and environmental deposits. The Bayesian formalism also allows much more specific models to be generated, which can be used in particular situations such as models that allow for spacial modeling (such as Blackwell and Buck 2003). This is likely to be an area of active development for some time to come and really goes beyond the scope of this paper (a useful general reference is Buck and Millard 2004). However, in the context of general-purpose models it is worth mentioning the case of deposition sequence models, which are already in widespread use.

Defined Deposition Rate

The first really widely used deposition model is the one where the rate of deposition is well known. This is especially useful in the case of wood with well-defined tree rings (see Christen and Litton 1995) and the `D_Sequence` command in OxCal (Bronk Ramsey et al. 2001; Galimberti et al. 2004), and is regularly used to fit elements of floating tree-ring calibration data. The mathematics associated with the various approaches to this, including classical least-squares fits, are discussed in detail in Bronk Ramsey et al. (2001). For the Bayesian approach, the prior is essentially very simple. If we have a series of dated events t_i, the earliest of which is t_a and the latest of which is t_b, then we have a known gap between them such that:

$$t_{i+1} = t_i + g_i \quad \text{for all } i \text{ such that } a \le i < b \tag{32}$$

There is therefore only 1 independent variable in the model (which is typically taken to be t_b the felling date for the tree). Because there is effectively only 1 variable, this is one of the few Bayesian models where the posterior probability densities can be directly calculated without having to use Monte Carlo methods.

Other Deposition Models

There is also increasing use of other kinds of Bayesian deposition models. These models vary in type. It is possible to simply use the order of deposition on its own using the models discussed above (the OxCal `Sequence` command, Bronk Ramsey 1995). This is surprisingly effective in some cases but makes no use of depth information at all. A much more prescriptive assumption of constant deposition rate is also sometimes made (Christen et al. 1995; Kilian et al. 1995; or the `U_Sequence` command of OxCal, Bronk Ramsey 2008).

Real sedimentary deposition is not completely uniform, and there have been a number of methods developed for dealing with this situation. One approach is to assume that the deposition is constant for short periods with periodic changes in deposition rate (Aguilar et al. 2002; Blaauw et al. 2003; Blaauw and Christen 2005). Other methods (sometimes not Bayesian) allow for non-linear deposition rates (Christen et al. 1995; Walker et al. 2003; Telford et al. 2004; Heegaard et al. 2005). In OxCal, a randomly variable deposition model has been developed (the `P_Sequence` command, Bronk Ramsey 2008).

The priors for most of these deposition models are mathematically more complicated than the generalized constraint and group models described above. Ideally, they should incorporate prior information about the deposition process itself.

ANALYSIS METHODS

In theory, once the model has been defined (the *prior*) and the measurements made (the *likelihood*), the posterior probabilities can be directly calculated using Bayes' theorem. So it is very easy to calculate the relative probability of each specific outcome. However, in practice most chronological models have sufficient numbers of independent parameters that the number of possible outcomes to consider at a reasonable resolution is very high. For example, a model with 20 events with a possible span of 500 yr or so that we wish to explore at 5-yr resolution has $(500/5)^{20} = 10^{40}$ possible outcomes. For this reason, Markov Chain Monte Carlo (MCMC) methods are usually used. Each iteration of the MCMC analysis provides a possible solution set of all of the parameters of the model. The probability of a particular solution appearing in the MCMC analysis should be directly proportional to its probability as defined by the posterior probability density, and in most cases after about 10^5 to 10^6 iterations a reasonably representative solution can be generated.

Assuming the MCMC algorithm is correctly implemented, it should eventually give representative distributions. If there are too few iterations in the analysis, the resulting probability distributions will usually be noisy and will vary from run to run. The degree to which a truly representative solution set has been generated is called "convergence" (see section "Problems with the MCMC Algorithm" below). One form of MCMC analysis is particularly useful and that is the Metropolis-Hastings algorithm (see e.g. Gilks et al. 1996), since it only requires relative probability information and in many cases full normalization of the posterior probability is difficult. This is the algorithm used exclusively in OxCal v4. Some other methods such as Gibbs sampling (Gelfand and Smith 1990) can give more rapid convergence but can only efficiently be applied in specific types of model and are no longer used in OxCal.

MODEL OUTPUTS

Of course in order to be useful, any analysis has to have outputs that can be easily interpreted. The most commonly used type of output from Bayesian analyses is the marginal posterior probability distribution function, which essentially pulls out the probability distribution for a single parameter, based on the entire model. Based on this, ranges of likely values can also be derived. However, there are other possible outputs from any analysis that can be useful in some circumstances.

Probability Distribution Functions

Mathematically, the marginal probability distribution function can be considered as the integral over all other parameters in the posterior probability density. However, since in practice we are usually finding the posterior distribution using MCMC analysis, all that we have to do is to build up a distribution of sample values for the parameter as the analysis proceeds. These distributions are what are usually plotted as the output of the model (as in Figure 5d,e). It is often convenient to see how the marginal posterior distributions relate to the original likelihood, so these are often plotted together. For example in OxCal, the likelihood is shown in outline, or light grey with the posterior marginal distribution overlain. This allows the effect of the modeling to be visually assessed.

Ranges

It is also frequently useful, both for likelihoods and marginal posterior distributions, to calculate ranges of probable values. These are usually quoted at 68% and 95% for comparability with 1 and 2 standard deviations in normal distributions. It should be noted though that the probability distributions are not usually normally distributed and these ranges should not be referred to as 1 σ and 2 σ, the latter notation being reserved for the intercept method of calibration. The ranges calculated from

probability distribution functions are calculated on the basis of highest probability density (HPD). So if we have a 95% range, it is the shortest range that includes 95% of the probability in the probability density function.

Interrogating the Analysis

In any MCMC analysis, there is much more information that can be extracted than just the marginal probability distributions. Bayesian analysis programs normally include ways to retrieve such information. For example, if we have a series of events we might wish to know the date of the earliest or latest of them. This can be performed during the MCMC by, at each iteration, taking a sample of whichever event parameter has the minimum or maximum values (`First` and `Last` CQL commands in OxCal). Likewise, it is easy at each stage to sample the span of all of the events in a group (`Span` command), or their order (the `Order` command builds up a matrix of the relative order of all pairs of events). You can also calculate the time interval between 2 events, which is sometimes even more interesting than the calendar age (using the `Interval` or `Difference` commands or directly as in the calculation of τ in Figure 5).

The marginal posterior densities sometimes hide useful information. Parameters are often highly correlated and so, for example, if the start of a phase is late it may be that the end will also be late within the possible ranges. In such instances, it can be useful to plot the probability density for 2 parameters relative to one another (`Correlation` function in CQL). As a last resort, the iterations of the MCMC can be inspected directly (`MCMC_Sample` command in OxCal v4.1 and later).

It should be stressed, however, that in a sense none of these interrogations, or queries, of the model actually add anything or change in any way the model itself. They do not affect the prior or the likelihoods on which the model is based.

DIAGNOSIS

Problems can arise with any statistical analysis, and the more complicated the analysis is the more likely there are to be problems. It is worth subdividing this into 3 distinct areas: the model, the measurements, and the MCMC analysis itself.

Testing Models

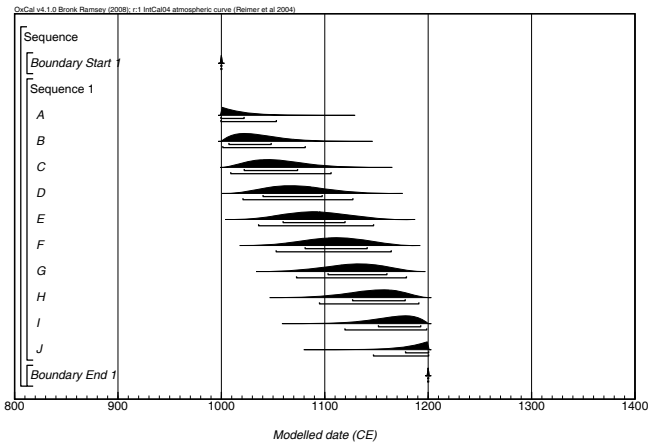
The prior can be divided into two parts, the informative prior (e.g. the sequence of events) and the uninformative prior (which is the rest of the mathematical construct required for the model). Ideally, the uninformative prior used in an analysis of this sort should be what is usually termed "vague." That is, the main information on the dating of events and the intervals between them should come from the data itself and not from the model. Sometimes, especially with a more complex or unusual model, it would be a good idea to check if the model performs as expected. There are two main approaches that can be useful for this.

The first is to use simulated ^{14}C dates (`R_Simulate`) for a scenario that mimics what is believed to be going on. For example, if we have a single phase that is expected to have a span of 100 yr, we might choose a start and end date for the phase, simulate ^{14}C dates for that phase, run a Bayesian analysis, and see if the output matches (within the uncertainty of the analysis) the range originally defined. This way it is possible to test the precision and accuracy of the analysis. If the precision is not high enough, then further simulations might indicate how many ^{14}C dates are needed to achieve the precision required. If the analysis is inaccurate, then there must be something wrong with the model itself.

The second approach is to run a model with very little dating information at all (use the `Date` command with no likelihood in OxCal). For this, we need some dating information at the start and finish of the model, to provide some information, but that is all. The analysis will then demonstrate how "vague" the model is, or if, for some reason, it is strongly biased towards one kind of solution. This type of analysis is useful in determining the unintended consequences of a particular model.

Figure 8 shows an example of both types of simulation. For a more comprehensive set of examples, see Bayliss et al. (2007).

```
Sequence()
{
 Boundary("Start 1",1000);
 Sequence("1")
 {
  Date("A");
  Date("B");
  Date("C");
  Date("D");
  Date("E");
  Date("F");
  Date("G");
  Date("H");
  Date("I");
  Date("J");
 };
 Boundary("End 1",1200);
};
```




```
Sequence()
{
 Boundary("Start 1");
 Sequence("1")
 {
  R_Simulate("A",1010,25);
  R_Simulate("B",1030,25);
  R_Simulate("C",1050,25);
  R_Simulate("D",1070,25);
  R_Simulate("E",1090,25);
  R_Simulate("F",1110,25);
  R_Simulate("G",1130,25);
  R_Simulate("H",1150,25);
  R_Simulate("I",1170,25);
  R_Simulate("J",1190,25);
 };
 Boundary("End 1");
};
```

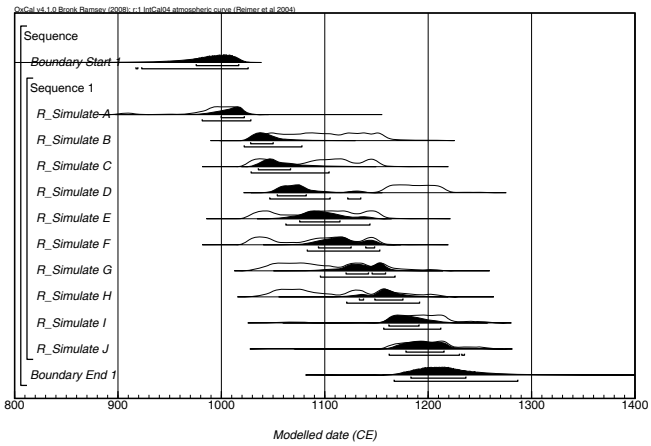


Figure 8 This figure shows two kinds of simulation that can be useful. At the top, we have a simulation of a sequence of 10 events assumed to be from the period 1000–1200 CE. In this simulation, we include no dating information but only the order of the events. On the left is the model definition in OxCal's CQL and on the right the effect of the model prior on the dates within the sequence. Even without any dating information, a sequential prior does have a considerable effect on the intervening dates. Below this, there is a simulation of a series of 10 ^{14}C dates, assumed to come from the same period. The plot on the right shows the kind of resolution one might expect to get from such a model, in this period, with uncertainty terms on the ^{14}C dates of ±25. Comparison of this plot with the one above shows the degree to which the ^{14}C dates and the model define the final posterior distributions.

Inconsistencies Between Measurements and the Model

Even if we have an accurate model, it may be that the measurement data are incompatible with it. For individual measurements, this is most easily spotted by seeing where the overlap between the likelihood distribution and the marginal prior is very small. If it is, then this means that it is very unlikely that we would have made this observation (e.g. ^{14}C measurement) if the model, and all of the other data, are correct. This might be because the model is wrong (e.g. a sample may be residual in a context) or because there is something wrong with the measurement itself: the effect is essentially the same.

There are two different approaches that can be taken to measurement data that do not agree with a model. One approach is to use formal outlier analysis (Christen 1994, 2003; Bronk Ramsey 2009). This has the advantage of bringing the treatment of the outliers into the model itself. It has the disadvantage that it brings complication to the model and perhaps makes it harder to intuitively understand what the assumptions behind the model are. In formal outlier analysis, each measurement is given a prior probability of being an outlier (typically a low probability like 0.05) and the sample is further down-weighted in the model if it is inconsistent with the rest of the available information. The output from the model is affected by this down-weighting, and in addition to the normal model outputs, a posterior probability for the sample being an outlier is generated. Further discussion of this approach and its implementation in OxCal v4.1 is given in Bronk Ramsey (2009). The other approach is to use the overlap between the likelihood and marginal posterior distributions as an indicator of problems: this is the approach that has been taken with the agreement index in Bronk Ramsey (1995). This has the disadvantage that it is not a formal statistical approach with a well-defined cut-off, but has the advantage that the model itself is unaffected. In practice, extreme outliers will be identified by either method and even in the formal methods the cut-off is arbitrary.

The rationale of the agreement index, as defined in Bronk Ramsey (1995), is that a comparison can be made between the marginal posterior distribution under the complete model with that under the null prior $p_0(\mathbf{t}) \propto constant$. Under this prior, all of the likelihoods and posterior distributions are the same since:

$$p_0(\mathbf{t}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{t})p_0(\mathbf{t}) \propto p(\mathbf{y}|\mathbf{t}) \tag{33}$$

Under this model, therefore, any marginal posterior density $p_0(t_i|\mathbf{y}) = p(t_i|y_i) \propto p(y_i|t_i)$. This is not usually true for the full model:

$$p_1(\mathbf{t}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{t})p_1(\mathbf{t}) \not\propto p(\mathbf{y}|\mathbf{t}) \tag{34}$$

In this case, we write the marginal posterior density depends on all the measurements and is written as $p_1(t_i|\mathbf{y})$. For each likelihood, we can then define a ratio:

$$F_i = \frac{\int p(y_i|t_i)p_1(t_i|\mathbf{y})dt_i}{\int p(y_i|t_i)p_0(t_i|\mathbf{y})dt_i} = \frac{\int p(y_i|t_i)p_1(t_i|\mathbf{y})dt_i}{\int p(y_i|t_i)p(t_i|y_i)dt_i} \tag{35}$$

This is the ratio of the mean likelihoods of making the measurement under the two models (the full model and the null model). The individual agreement index is just taken to be $A_i = 100F_i$ with a reasonable minimum acceptable value being in the region of 60% (Bronk Ramsey 1995). On average, one might expect about 1 in 20 A_i values to drop below 60%, but if they are substantially lower or a large proportion fall below this level there may well be something internally inconsistent between

the model and the data. Note that the ratio F_i can be greater than 1 (and therefore A_i can be greater than 100%) as this means the measurement is more likely under the full model than under the zero model.

To get an idea of how all the measurements together agree with the model, we can factor all the ratios F_i together:

$$F_{overall} = \prod_i F_i \tag{36}$$

But this does not take account of the correlation between the parameters, so a better overall measure is:

$$F_{model} = \frac{\int p(\mathbf{y}|\mathbf{t}) p_1(\mathbf{t}|\mathbf{y}) d\mathbf{t}}{\int p(\mathbf{y}|\mathbf{t}) p_0(\mathbf{t}|\mathbf{y}) d\mathbf{t}} \tag{37}$$

Such factors are not strictly speaking Bayes' factors, but they can be compared between different models in a similar way. For a single model, the overall agreement indices are then defined as:

$$A_{overall} = 100(F_{overall})^{1/\sqrt{n}} \tag{38}$$

$$A_{model} = 100(F_{model})^{1/\sqrt{n}} \tag{39}$$

These indices have the property that their values are usually of roughly the same magnitude as the individual agreement indices, and so again if they drop much below 60%, it is worth considering if there are any problems with the model as a whole, or with some of the individual measurements. For OxCal v4 and later, A_{model} is to be preferred because it accounts for the correlation between parameters. In earlier versions of OxCal, only $A_{overall}$ was calculated.

In Bayesian analysis more generally, model choice is usually performed using Bayes' factors. This approach has been taken in ^{14}C dating too (Nicholls and Jones 2001) though, as yet, it has not been implemented for a wide range of different models, and this is an area where further work is likely to be useful. An alternative to such model choice is model-averaging, an example of which is outlier analysis, where a number of different possible models are analyzed at the same time (Christen 1994, 2003; Bronk Ramsey 2009).

Problems with the MCMC Algorithm

In some cases, models are so inconsistent with their measurement data that there are no possible solutions. In such cases, the MCMC chain cannot run. In other cases, the solution space may be very fragmentary. In such situations, the MCMC solutions will be different with each different random starting position. These cases of poor convergence are not always easy to detect. The algorithms are usually set up to try various different starting positions and then test if the marginal posterior densities are the same for each run (see e.g. Bronk Ramsey 1995). However, although multiple runs always giving the same distribution is a necessary condition for good convergence, it is not, in theory, a sufficient one.

In some instances when the MCMC algorithm fails to start, it may be possible to provide extra information that will help to find a first reasonable solution. In particular, it can be difficult to automati-

cally estimate a reasonable starting point for boundary events that are not in themselves directly dated. In such instances, it may be useful to define a reasonable range as part of the model definition (in OxCal v4 this can be done by adding a likelihood distribution to the Boundary definition as in `Boundary("End",U(AD(1000),AD(1400)))`).

Where more complex models fail to converge, or indeed to find a suitable starting point, it is often worth breaking them down into smaller units and testing these separately.

Other Potential Problems

There are other potential problems that cannot be diagnosed so easily. Perhaps the most important one to bear in mind is that any analysis of this kind is very strongly dependent on the information that goes into it. In particular, the calibration curve underpins the linkage between the ^{14}C measurements and their place in the timescale. Any correlated or consistent mismatch between the measurement set and the calibration curve is difficult to diagnose and could have a significant effect on the output of the analysis. There is more discussion of ways to deal with such issues in Bronk Ramsey (2009).

CONCLUSIONS

Bayesian analysis is now widely used in the interpretation of ^{14}C dates. The models can be very simple, in many cases involving only a single group of events. Even in these cases, a surprising amount of information can be extracted from the analysis, and this allows a much better interpretation of the data than is possible by "eye-balling" a set of calibrated dates.

In other cases, more complex models are needed, but these can usually be built up from simple components and the mathematics underlying such models is essentially the same. Most such models are based around the notion of "uniform" groupings, but other distributions of events can be treated just as easily and this enables us to use the same approach for chronologies that do not have a definite start and end.

Bayesian analysis of ^{14}C dates is now a mature methodology with several programs providing implementation of many of the most widely used models and diagnostic methods for exploring their implications. That said, there are still issues that we will always need to keep at the forefront of our minds. In particular, however much statistical analysis we do, ^{14}C dates are still reliant on the underlying assumptions of the ^{14}C method—any problems with the samples, their context, their association with each other, or with the calibration curve, will have implications for the accuracy of our chronologies.

ACKNOWLEDGMENTS

Thanks are due to English Heritage for funding the development of OxCal v4 though a research grant (Historic Environment Enabling Programme Grant 3164) and the Natural Environment Research Council for ongoing institutional support through Services and Facilities. I am also very grateful to all of those who have contributed to the ideas behind this paper through discussions and previous publications; in particular I would like to thank Alex Bayliss, Maarten Blaauw, Simon Blockley, Caitlin Buck, Andres Christen, Mike Dee, Zhiyu Guo, Sharen Lee, Dan Miles, Andrew Millard, Geoff Nicholls, Richard Staff, and Peter Steier all of whom have been especially helpful.

REFERENCES

Aguilar DGP, Litton CD, O'Hagan A. 2002. Novel statistical model for a piece-wise linear radiocarbon calibration curve. *Radiocarbon* 44(1):195–212.

Bayliss A, Whittle A, editors. 2007. Histories of the dead: building chronologies for five southern British long barrows [special issue]. *Cambridge Archaeological Journal* 17(Supplement S1).

Bayliss A, Bronk Ramsey C, van der Plicht J, Whittle A. 2007. Bradshaw and Bayes: towards a timetable for the Neolithic. *Cambridge Archaeological Journal* 17(Supplement S1):1–28.

Blaauw M, Christen JA. 2005. Radiocarbon peat chronologies and environmental change. *Applied Statistics* 54(4):805–16.

Blaauw M, Heuvelink GBM, Mauquoy D, van der Plicht J, van Geel B. 2003. A numerical approach to ^{14}C wiggle-match dating of organic deposits: best fits and confidence intervals. *Quaternary Science Reviews* 22(14):1485–500.

Blackwell PG, Buck CE. 2003. The Late Glacial human reoccupation of north-western Europe: new approaches to space-time modelling. *Antiquity* 77(296): 232–40.

Boaretto E, Jull AJT, Gilboa A, Sharon I. 2005. Dating the Iron Age I/II transition in Israel: first intercomparison results. *Radiocarbon* 47(1):39–55.

Bowman SGE, Leese MN. 1995. Radiocarbon calibration: current issues. *American Journal of Archaeology* 99(1):102–5.

Bronk Ramsey C. 1994. Analysis of chronological information and radiocarbon calibration: the program OxCal. *Archaeological Computing Newsletter* 41:11–6.

Bronk Ramsey C. 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2):425–30.

Bronk Ramsey C. 1998. Probability and dating. *Radiocarbon* 40(1):461–74.

Bronk Ramsey C. 2000. Comment on 'The use of Bayesian statistics for ^{14}C dates of chronologically ordered samples: a critical analysis.' *Radiocarbon* 42(2):199–202.

Bronk Ramsey C. 2001. Development of the radiocarbon calibration program. *Radiocarbon* 43(2A):355–63.

Bronk Ramsey C. 2008. Deposition models for chronological records. *Quaternary Science Reviews* 27(1–2): 42–60.

Bronk Ramsey C. 2009. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* 51(3).

Bronk Ramsey C, van der Plicht J, Weninger B. 2001. 'Wiggle matching' radiocarbon dates. *Radiocarbon* 43(2A):381–9.

Buck CE, Millard A. 2004. *Tools for Constructing Chronologies: Crossing Disciplinary Boundaries*. London: Springer. 257 p.

Buck CE, Kenworthy JB, Litton CD, Smith AFM. 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65(249):808–21.

Buck CE, Litton CD, Smith AFM. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19(5): 497–512.

Buck CE, Litton CD, Scott EM. 1994. Making the most of radiocarbon dating: some statistical considerations. *Antiquity* 68(259):252–63.

Buck CE, Cavanagh WG, Litton CD. 1996. *Bayesian Approach to Interpreting Archaeological Data*. Chichester: Wiley. 402 p.

Buck CE, Christen JA, James GN. 1999. BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeol*ogy 7: http://intarch.ac.uk/journal/issue7/buck_index.html.

Christen JA. 1994. Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics* 43(3):489–503.

Christen JA. 2003. Bwigg: an internet facility for Bayesian radiocarbon wiggle-matching. *Internet Archaeology* 13: http://intarch.ac.uk/journal/issue13/christen_index.html.

Christen JA, Litton CD. 1995. A Bayesian approach to wiggle-matching. *Journal of Archaeological Science* 22(6):719–25.

Christen JA, Clymo RS, Litton CD. 1995. A Bayesian approach to the use of ^{14}C dates in the estimation of the age of peat. *Radiocarbon* 37(2):431–41.

Cleal R, Walker KE, Montague R. 1995. *Stonehenge in Its Landscape: Twentieth-Century Excavations*. London: English Heritage. 640 p.

Galimberti M, Bronk Ramsey C, Manning SW. 2004. Wiggle-match dating of tree-ring sequences. *Radiocarbon* 46(2):917–24.

Gelfand AE, Smith AFM. 1990. Sampling-based approaches to calculating marginal densities. *Journal of the American Statistical Association* 85(410):398–409.

Gilks WR, Richardson S, Spiegelhalter DJ. 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman & Hall. 486 p.

Heegaard E, Birks HJB, Telford RJ. 2005. Relationships between calibrated ages and depth in stratigraphical sequences: an estimation procedure by mixed-effect regression. *The Holocene* 15(4):612–8.

Jones M, Nicholls G. 1999. New radiocarbon calibration software. *Radiocarbon* 44(3):663–74.

Karlsberg AJ. 2006. Flexible Bayesian methods for archaeological dating [unpublished PhD dissertation]. University of Sheffield.

Kilian MR, van der Plicht J, van Geel B. 1995. Dating raised bogs: new aspects of AMS ^{14}C wiggle matching, a reservoir effect and climatic change. *Quaternary Science Reviews* 14(10):959–66.

Manning SW, Bronk Ramsey C, Kutschera W, Higham T,

Kromer B, Steier P, Wild EM. 2006. Chronology for the Aegean Late Bronze Age 1700–1400 BC. *Science* 312(5773):565–9.

Mazar A, Bronk Ramsey C. 2008. ^{14}C dates and the Iron Age chronology of Israel: a response. *Radiocarbon* 50(2):159–80.

Needham S, Bronk Ramsey C, Coombs D, Cartwright C, Pettitt P. 1998. An independent chronology for British Bronze Age metalwork: the results of the Oxford Radiocarbon Accelerator Programme. *Archaeological Journal* 154:55–107.

Nicholls G, Jones M. 2001. Radiocarbon dating with temporal order constraints. *Applied Statistics* 50(4):503–21.

Reimer PJ, Brown TA, Reimer RW. 2004. Discussion: reporting and calibration of post-bomb ^{14}C data. *Radiocarbon* 46(3):1299–304.

Steier P, Rom W. 2000. The use of Bayesian statistics for ^{14}C dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42(2):183–98.

Stuiver M, Polach HA. 1977. Discussion: reporting of ^{14}C data. *Radiocarbon* 19(3):355–63.

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

Telford RJ, Heegaard E, Birks HJB. 2004. The intercept is a poor estimate of a calibrated radiocarbon age. *The Holocene* 14(2):296–8.

van den Bogaard P. 1995. ^{40}Ar/^{39}Ar ages of sanidine phenocrysts from Laacher See Tephra (12,900 yr BP): chronostratigraphic and petrological significance. *Earth and Planetary Science Letters* 133(1–2):163–74.

van der Plicht J. 1993. The Groningen radiocarbon calibration program. *Radiocarbon* 35(1):231–7.

Walker MJC, Coope GR, Sheldrick C, Turney CSM, Lowe JJ, Blockley SPE, Harkness DD. 2003. Devensian lateglacial environmental changes in Britain: a multi-proxy environmental record from Llanilid, South Wales, UK. *Quaternary Science Reviews* 22(5–7):475–520.

Zeidler JA, Buck CE, Litton D. 1998. Integration of archaeological phase information and radiocarbon results from the Jama River Valley, Ecuador: a Bayesian approach. *Latin American Antiquity* 9(2):160–79.