

Mathematics and Archaeology ISBN 9781482226812 Copyright © 2015 by Taylor and Francis Group, LLC. Reproduced by permission of Taylor and Francis Group, LLC, a division of Informa plc.

This material is strictly for personal use only. For any other use, the user must contact Taylor & Francis directly at this address: [permissions.mailbox@taylorandfrancis.com](mailto:permissions.mailbox@taylorandfrancis.com). Printing, photocopying, sharing via any means is a violation of copyright.

# Bayesian approaches to the building of Archaeological Chronologies

Christopher Bronk Ramsey

January 20, 2015

## Abstract

Bayesian statistics provide a methodology for combining information from diverse sources, assuming that such information can be objectively quantified. Critically, the mathematical methods are also well suited to parameters that have non-normal probability distributions. In archaeological and related environmental contexts, there is usually a wide range of information that has bearing on chronology; this includes stratigraphic information, information on deposition processes and other constraints, which can be applied relative to the passage of calendar time. In addition, one of the most widely used dating techniques in archaeology, radiocarbon dating, requires a calibration process that inevitably leads to non-normal probability distributions. Bayesian statistics provide the ideal way to combine the information from these different sources and has been highly successful in improving the precision and accuracy of archaeological chronologies. The same is increasingly true in environmental science, where some of the methods first applied in archaeology are now finding application in different disciplines.

## 1 Introduction

Bayesian inference has become an important tool for the construction of archaeological chronologies. This has been made possible by the development of Bayesian statistical methods (see, for example, Buck et al., 1991, 1996; Buck and Millard, 2004; Nicholls and Jones, 2001) and the development of both computing power and algorithms (Gelfand and Smith, 1990; Gilks et al., 1996) able to put such methods to practical use. These have enabled the development of computer software able to perform chronological analysis along with development of new Bayesian models that cover a wide range of situations (Blaauw et al., 2003; Bronk Ramsey, 1995, 2008, 2009a,b; Buck et al., 1999; Christen, 2003; Haslett and Parnell, 2008; Jones and Nicholls, 2002).

The subject of Bayesian chronology building is thus a large one and this chapter is necessarily limited in its scope. The emphasis here is on the main mathematical ideas underlying much of the chronological work, and in particular how these are relevant to the interface between mathematics and archaeology. The approach taken here is focussed on applied mathematical methods suitable for the underlying processes rather than Bayesian statistics *per se*.

One major issue relevant to all dating methods, and often overlooked, is the basis for the measure of time we are using and the numerical value we attach to our time parameters. Most chronological systems are, for historical reasons, essentially counting methods based in integer arithmetic. The numerical approaches that we use assume parameters with real-number values and if we are to have a sound basis for our analyses this needs to be properly defined. The definitions that we adopt do not quite accord with the underlying archaeological processes (for example the notion of instantaneous events) nor strictly speaking physical realities (the notion

that there is a Newtonian universal time scale). We need to be aware of these limitations but in practice the definitions are usually good enough for what we are trying to do.

Probably the most important interface between archaeology and the mathematical approaches taken in chronology building is the way in which we group and relate different events to one another. The groups that we choose to define are a product of our understanding of the human activity reflected in the archaeological record. The way we relate our events also usually depends critically on our interpretation of site stratigraphy and sample taphonomy. For these reasons any Bayesian chronology building exercise is always part of the interpretation of the data rather than impartial analysis. That said, the aim of our mathematical models is to provide a degree of impartiality within that framework. The two most important classes of information that we wish to extract from any chronology are absolute date estimates and relative date estimates; both of these allow us to understand processes of change. Our models therefore aim to be unbiased in the estimates of age, and duration. What we mean by unbiased is critical in formulating our models and will be discussed in the following sections.

One of the main reasons that Bayesian methods have become so widely used is because the dating methods that we have all have uncertainty associated with them: in particular radiocarbon generated probability distribution functions which are non-normal. Despite this, because models are able to provide date and rate information for events which are not directly dated, the approaches discussed here would still be useful in cases where uncertainties are much less significant (such as is the case for dendrochronology). The outputs from our models are also in the form of probability distributions, and so mathematical methods which enable us to summarise these distributions, and compare them are also important elements of this endeavour.

## 2 Events and processes

In archaeology what we are usually interested in is the changing nature of human behaviour, and the interaction between humans in their environment. The way we gain this understanding is to look for material evidence for past human activity as recorded in the archaeological record. Archaeology is a complex subject because the nature of the record is such that we must apply sophisticated methods, often classed as scientific, for the analysis and quantification of the physical remains, and then use a whole range of approaches to interpret that evidence in terms of human behaviour.

When it comes to chronology, our interpretation of the evidence is mostly guided by information on the speed and order of changes, rather than information on exactly 'when' things happened. However, in practice, one of the easiest ways to work out rates of change, and relative orders of events in different regions is to put events and processes onto some common timescale. In order to do this quantitatively we need mathematical methods, and we need to define how we measure our processes in a formal way. We can draw an analogy with the capture of human locomotion needed for CGI film production. The movement of a person is a very complex and subtle process, but it can often be captured in a useful way by a relatively small number of marker points on the human body. We attempt to do the same thing in chronology, by associating events, whose timing we attempt to measure with those processes of change. The simplest example of this might be a phase of activity, which we normally formalise into two defining events: the start event and the end event of that phase.

So our task in building archaeological chronologies comes down to putting events onto some common timescale. It turns out that the useful information we need to achieve this, comes from a variety of different sources: most obviously we have methods for measuring age, but equally important, methods for giving relative age to events, principally through stratigraphy, and methods for determining age differences (for example through biological processes, like tree ring growth, or through physical processes, such as sedimentation). It will be apparent already that these different kinds of information are drawn, using a complex process of inference, from the archaeological evidence, and might rightly be considered to be part of the interpretation of data. This means that chronology cannot be seen as a stand-alone plug-in module within the

archaeological toolkit, but rather a key element within site interpretation. On a broader scale, the same can be said about the dating methods themselves: although they might be classes as scientific measurements, they also depend on interpretation of a complex range of environmental and archaeological information.

### 3 Bayesian Overview

Bayes theorem in its most formal terms allows us to make use of both prior and likelihood information. We will assume that we are interested in a whole set of parameters  $\mathbf{t}$ , for which we have a set of observed information  $\mathbf{y}$ . Bayes theorem can then be expressed as:

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

where  $\mathbf{t}$  are 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. Finally  $p(\mathbf{y})$  is the overall probability of the observations made. In most of the cases that we will be interested in  $p(\mathbf{y})$  can be treated as a normalisation constant and thus we have:

$$p(\mathbf{t}|\mathbf{y}) \propto p(\mathbf{y}|\mathbf{t})p(\mathbf{t}) \quad (2)$$

In a Bayesian chronological analysis we have to express temporal information in these terms. Most of our parameters  $\mathbf{t}$  are used to express information about time. Our observations  $\mathbf{y}$  may include measurements on isotope ratios, for example, and thus the likelihood terms are the way we relate these observations to the age parameters in the model; radiocarbon dates would thus be expressed as likelihoods in our overall scheme. Prior information is information which we know ahead of the analysis: the distinction here is slightly arbitrary, as much of this information, such as the ordering of events, is of course based on observations, and so we could consider the setting up of the prior as in itself a Bayesian exercise.

## 4 Measuring age

One clear aim of chronology is to put events onto a single timescale so we can compare what is happening in different places in a way that allows us to understand the human activity for which we have evidence in the archaeological record. In practice all methods we have for determining ages (or dates on a time-line) have uncertainties associated with them, uncertainties which might be independent between age estimates, or might be systematic between measurements of the same kind. This section will look at different ways of measuring age and some of the issues involved in incorporating such information into chronological models.

### 4.1 The timescale(s)

We have stated the aim in building chronologies as putting events onto a common timescale. We can fortunately don't have to worry about relativistic effects and can assume that time is a cartesian coordinate independent of position. There is no meaningful smallest increment of time so we consider it as real number  $t \in \mathbb{R}$ . In practice, however the most important two temporal cycles for humans are the day, and (at least for non-tropical regions) the year. Because of the resolution of most archaeological research the year is the unit usually used for the measurement of time. This raises several problems: firstly the year is not an SI unit and has a number of different definitions (Julian, Siderial, Tropical etc.), secondly the year cycle is different in the Northern and Southern hemispheres, and finally the counting of years in different calendars is

notoriously complicated. For most purposes this may not have much practical consequence but it does make any formal definition of the timeline very difficult.

In practice many archaeologists simply use the integer years as a timescale. Here years AD (or CE) form one set  $y_a$  and years BC (or BCE) form another  $y_b$ . In each case the integer is definite positive ( $y_a, y_b \in \mathbb{Z}^+$ ). Depending on the period in question, the years may be in the Gregorian calendar, or the Julian calendar. Normally for pre-history we really just assume that the years are a count of seasonal cycles from the present. This is very unsatisfactory formalism, and astronomers have long ago (18th Century) adopted an astronomical year numbering system with a zero ( $y_c \in \mathbb{Z}$ ). Furthermore to add complication, an integer year notation was introduced in 1950 which is termed 'before present' where the year AD 1950 is given the value 0 ( $y_d \in \mathbb{Z}^*$ ). And so:

$$y_c = y_a \quad (3)$$

$$y_c = 1 - y_b \quad (4)$$

$$y_d = 1950 - y_c \quad (5)$$

None of these really represents a reasonable basis for mathematical analysis. Even  $y_c$  in its normal definition has different year lengths for different periods, and none of the others cover the full time scale to the present. They are also all integer timescales with no ability to specify times at a finer interval without recourse to complexity of the calendar. To overcome these problems, it is much more convenient to use a real number time line with a unit of one year. Because we wish to retain synchrony with the seasons we need a year length to be a good approximation to the tropical year length and for this purpose it has been proposed (Bronk Ramsey, 2009a) to use the mean Gregorian year (31,556,952s), as this can be directly related to the SI unit of the second, with the reference point being 0001-01-01 00:00:00 Universal Time under ISO-8601. Under this scheme the floating point 1.0 is the start of the year 1AD (and consequently 2001.0 is the start of the year 2001AD). Within a day or two, the calendar year (Gregorian) in its different forms can be derived as:

$$y_c = \text{floor}(t) \quad (\text{for all } t) \quad (6)$$

$$y_a = \text{floor}(t) \quad (\text{for } t \geq 1) \quad (7)$$

$$y_b = 1 - \text{floor}(t) \quad (\text{for } t < 1) \quad (8)$$

$$y_d = 1950 - \text{floor}(t) \quad (\text{for } t < 1951) \quad (9)$$

Having defined our theoretical time-line, we now have to consider how we can use information to estimate the timing of events on that scale.

## 4.2 Events defined on a calendar

Having defined the timeline in terms of current time and date standards, we can use our knowledge of different calendar systems to put any event with a defined date (and time) onto our timeline. There is a convenient tool for doing this from many of the frequently used calendar systems at [http://c14.arch.ox.ac.uk/oxcalhelp/hlp\\_analysis\\_calend.html](http://c14.arch.ox.ac.uk/oxcalhelp/hlp_analysis_calend.html). Of course in practice we never know the time and date of an event perfectly and so usually we would define some uncertainty in the definition of that time range. This might commonly be in the form of a standard uncertainty (Normally distributed probability with a mean  $\sigma$ , a standard uncertainty or  $\sigma$ ) or as a range (uniform probability in the range  $t_a$  to  $t_b$ ).

$$p(t) \sim \mathcal{N}(\mu_t, \sigma_t^2) \quad (10)$$

$$p(t) \sim \mathcal{U}(t_a, t_b) \quad (11)$$

Other distributions, such as the Student-t distribution, might also be justified in some cases.

### 4.3 Scientific dating methods

Most scientific dating methods (Argon-Argon, Potassium-Argon, Uranium series, Luminescence methods, Amino-Acid Racemization etc) operate on the basis of measuring age. The uncertainties associated with the methods are often approximately Normal, and so it is possible to express the date from such measurements as a Normal probability distribution function. If the determined age is  $\mu_a \pm \sigma_a$  and the sample was measured at time  $t_0$ , then the probability distribution function for the date of the event in question is:

$$p(t) \sim \mathcal{N}(t_0 - \mu_a, \sigma_a^2) \quad (12)$$

In principle other less generic likelihood distributions could be derived. In particular for Uranium Series dating the likelihood distribution should be non-linear as you approach the limit to the technique (Ivanovich and Harmon, 1992). In optically stimulated dating, there is also much more information available that in principle might be used within a Bayesian model (Rhodes et al., 2003).

### 4.4 Layer counting methods

Three layer counting techniques are often used in the generation of timescales. These are dendrochronology, varve counting and ice layer counting. In all cases the method relies on the annual cycle of the seasons to leave an annual signal in the relevant record. These methods then normally give age intervals within the the archive with a precision of a year. If the archive can be extended to the present, these can be used to provide a date.

For varves and ice core layers, there are uncertainties involved in the counting process and so some uncertainty should be applied. For ice cores the normal practice is to to derive a maximum counting error. This can be interpreted as a uniform distribution as in equation 11, but probably more realistically can be interpreted as a  $2\sigma$  range of a normal distribution as in equation 10. Ideally, however, we should keep in mind that these uncertainties are highly correlated and we often know the relative age in such a record much better than we know the absolute age. This information can in principle be included in Bayesian models (using for example age difference priors) but a comprehensive methodology for doing this has not been developed.

Dendrochronology ideally brings in enough replication to provide a result for an extant tree ring which is precise to the year. However where there are missing rings, allowance needs to be made for this in estimating the use date (Miles, 2005).

### 4.5 Radiocarbon calibration

Radiocarbon dating is, mathematically, much more complex than many other dating techniques. What is measured is a radiocarbon isotope ratio, usually expressed as a fraction of the ‘modern’ value of reference material from 1950. This  $F^{14}C$  value,  $f$ , can be converted to a radiocarbon date  $a$  using the Libby half life of 5568 a:

$$a = \frac{5568}{\ln(2)} \ln(f) \approx 8033 \ln(f) \quad (13)$$

The uncertainty in the measurement of  $f$  should be a summary of all associated analytical uncertainties and is usually assumed to be Normal in distribution:

$$f \sim \mathcal{N}(\mu_f, \sigma_f^2) \quad (14)$$

To first order this gives a normal distribution in radiocarbon date:

$$a \sim \mathcal{N}(\mu_a, \sigma_a^2) \quad (15)$$

$$\mu_a \approx 8033 \ln(\mu_f) \quad (16)$$

$$\sigma_a \approx 8033 \frac{\sigma_f}{\mu_f} \quad (17)$$

However, this approximation is only valid when  $\sigma_f \ll \mu_f$ , which is true for younger dates.

The radiocarbon calibrations curves are generated using statistical methods that are designed to provide an estimate of the radiocarbon measurement that you would get for a sample of particular age. The curves are provided at discrete intervals but can be interpolated to give a continuous function covering the time range from 50000 calBP to 2010 (Reimer et al., 2013; Hogg et al., 2013; Hua et al., 2013). For the period after 1950, the data is circulated in terms of the parameter  $f$  but for the period before in terms of  $a$ . However, now that calibrations extends to the limit of the technique it is important that the calibration calculations are done in the parameter space of  $f$  rather than that of  $a$  given the approximation inherent in equation 17.

The curve is normally expressed as a function of time  $r(t)$  with an uncertainty which is also a function of time  $s(t)$ , so that the radiocarbon value  $\rho$  is assumed to take the value:

$$p(\rho|t) \sim \mathcal{N}(r(t), s^2(t)) \quad (18)$$

You will see that from the argument above, this is more valid if  $\rho$ ,  $r$  and  $s$  are expressed in terms of  $f$ , though in much of the literature they are expressed in terms of  $a$ . From this we can see that if we know the date of a particular event  $t_i$  and that the uncertainty in the measurement will be  $s_i$  it follows (Buck et al., 1992) that:

$$p(r_i|t_i, s_i) \sim \mathcal{N}(r_i - r(t_i), s^2(t_i) + s_i^2) \quad (19)$$

where  $r_i$  is the radiocarbon measurement made for that sample.

Alternatively (Bronk Ramsey, 2009a) we can use the calibration curve to provide a two dimensional probability density for  $\rho$  and  $t$ . Assuming that the marginal prior for  $t$  should be uniform, this is:

$$p(\rho_i, t_i) \propto \frac{1}{s(t_i)} \exp(-(\rho_i - r(t_i))^2 / (2s^2(t_i))) \quad (20)$$

And from the likelihood expression in equation 14 we have:

$$p(r_i|\rho_i, s_i) = \frac{1}{s_i \sqrt{2\pi}} \exp(-(r_i - \rho_i)^2 / (2s_i^2)) \quad (21)$$

Combining these two and integrating over the parameter  $\rho_i$  we have:

$$p(r_i|\rho_i, t_i, s_i) \propto \frac{1}{s_i s(t_i)} \exp(-(r_i - \rho_i)^2 / (2s_i^2)) \exp(-(\rho_i - r(t_i))^2 / (2s^2(t_i))) \quad (22)$$

$$p(r_i|t_i, s_i) \propto \int_{\rho_i=-\infty}^{\infty} p(y_i|\rho_i, t_i) d\rho_i \quad (23)$$

$$\propto \frac{\exp\left(-\frac{(r_i - r(t_i))^2}{2(s_i^2 + s^2(t_i))}\right)}{\sqrt{s_i^2 + s^2(t_i)}} \quad (24)$$

$$\sim \mathcal{N}(r_i - r(t_i), s^2(t_i) + s_i^2) \quad (25)$$

Which gives the same result as in equation 19. The value of this approach lies in understanding the probability distribution in equation 20. It this two dimensional probability distribution is marginalised over  $t_i$  it is uniform, but not if it is marginalised over  $\rho_i$ . The non-linearity of the calibration ‘curve’ implies that the probability density in only one of these two parameters can be uniform and clearly for dating material which might equally be any age this formulation makes most sense of the available information.

## 4.6 Application of dating methods to single events

Using Bayes theorem:

$$p(t_i|y_i) = \frac{p(y_i|t_i)p(t_i)}{\int p(y_i|t_i)p(t_i)dt_i} \quad (26)$$

with a uniform  $p(t_i)$ , and making use of the fact that  $\int p(y_i|t_i)p(t_i)dt_i$  does not depend on  $t_i$ , we derive the result that:

$$p(t_i|y_i) \propto p(y_i|t_i) \quad (27)$$

What this means is that, for any dating method, the likelihood function when normalised over  $t_i$  can be used as a marginal posterior for the date of the event in question. This is why, for example, for simple calibration of single radiocarbon results (from equation 24) we can see that:

$$p(t_i|r_i, s_i) \propto \frac{\exp\left(-\frac{(r_i-r(t_i))^2}{2(s_i^2+s^2(t_i))}\right)}{\sqrt{s_i^2+s^2(t_i)}} \quad (28)$$

This is the probability distribution used for calibration of radiocarbon dates in all widely used software packages (such as Bronk Ramsey, 2001; Buck et al., 1999; van der Plicht, 1993).

## 4.7 Dealing with outliers

In practice not every age measurement is correct. The reasons for this are manifold: there may be problems with the associated measurements, or it may be that the age measured cannot really be associated with the event envisaged. For example, a cereal grain deposited in a particular context might be residual and so the chronologist expects the measured radiocarbon date to be close to the time of deposition but the radiocarbon date reflects the growth of the grain some years earlier. There have been three main approaches taken to such outliers.

The first approach is to identify outliers, either ‘by eye’ or using the measures discussed in section 6.2 below.

The second method, within the context of a Bayesian model, is to introduce a new boolean parameter for each measurement  $\phi_i$  which is 1 if the sample is an outlier and 0 if it is not. We assign a prior probability  $q_i$  that the sample’s measurement is an outlier for some reason. This approach has the advantage that we can look at the marginal posterior for  $\phi_i$  and thereby identify probable outliers. It was first formulated by Christen (1994, 2003) and then further elaborated by Bronk Ramsey (2009b) to include a number of different types of outlier.

The final method is to modify the distribution used for the calibration from the Normal distribution in equation 28 to a student-t distribution: this is the approach taken by Christen and Pérez (2009).

## 5 Groupings, Processes and Models

The previous section dealt with how we treat individual events and the simplest approach to dealing with multiple events is just to treat them as a series of completely independent and unconnected events. Such statistical models are relevant when, for example, we are dealing with indistinguishable particles. However with events we always know something about them and, if we are studying them in relation to an archaeological problem, they do by definition have some relationship to one another.

### 5.1 The null model

At this stage we will introduce the notion of the null model. This is one under which we do assume full independence of all events. This is unrealistic but it does provide a baseline from which to consider other models. In this null model the prior can be factorized as a uniform distribution in each event parameter. This prior is thus improper in that it cannot be normalised. However under this prior the marginal posterior for each event is just directly proportional to the likelihood function (as with a single event).

The problem with such a model as a realistic description of an archaeological process comes principally from the fact that although the model has a uniform prior for each individual event



(out of a total of  $n$ , where  $n \geq 2$ ), the effective prior for the derived quantity of the span  $s$  of those events is proportional to  $s^{n-2}$ . Thus, for example 12 events would be  $10^{10}$  times more likely to span 100 years than 10 years. This is not a reasonable assumption to make for most archaeological processes.

## 5.2 Constraints

Before we come on to consider our solution to the problem with the ‘null model’ we will introduce an important aspect of the prior and that is constraints. One of the pieces of information that we often have from archaeological sites or process is information on the order of events. These constraints might come from stratigraphy (Harris, 1989) or from the definition of parameters describing a process. To introduce a single constraint  $t_a < t_b$  between two event parameters we introduce an element in the prior:

$$p_H(t_a, t_b) = \begin{cases} 1, & \text{if } t_a < t_b \\ 0, & \text{otherwise.} \end{cases} \quad (29)$$

The function  $p_H$  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:

$$p_H(t_a, t_b, t_c, \dots) = \begin{cases} 1, & \text{if } t_a < t_b < t_c < \dots \\ 0, & \text{otherwise.} \end{cases} \quad (30)$$

Any constraints that we wish to include in a prior model can be expressed as factors of this form.

## 5.3 Two parameter processes

Next we will consider processes which can be described by two independent parameters. The most commonly considered of these is the single phase of activity (Buck et al., 1991, 1992; Zeidler et al., 1998; Bronk Ramsey, 2009a). Here we have two parameters which describe the phase, a start event  $t_a$  and a finish event  $t_b$ . Because these describe the limits of the phase they are sometimes referred to as phase boundaries (Bronk Ramsey, 2009a). The prior for these boundaries is uniform but we constrain  $t_a$  to be before  $t_b$  on our time-line. That is the prior is given by:

$$p(t_a, t_b) \propto p_H(t_a, t_b) \quad (31)$$

As there are only two boundary events the prior for the difference between them  $s = t_b - t_a$  is uniform over the range  $0 - \infty$  which is suitably uninformative. This prior is improper but this will not normally matter in practice. All of the events within the phase then have a normalised prior of:

$$p(t_i | t_a, t_b) = \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \quad (32)$$

Thus the overall prior for the model is:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} = \frac{\prod_i p_H(t_a, t_i, t_b)}{(t_b - t_a)^n} \quad (33)$$

where there are  $n$  events within the phase.

This is the most widely used type of model for describing archaeological processes. The model has  $n + 2$  parameters and an overall scale which we can define as  $s = t_b - t_a$ . Because the prior contains a factor of  $s^{-n}$  this exactly cancels the bias towards longer scales discussed in section 5.1 above. Thus the prior for the model is scale invariant which is suitably uninformative.

The two parameter approach can be used for other types of process other than the uniform phase: Figure 1 shows some examples implemented within the OxCal package. All of these two

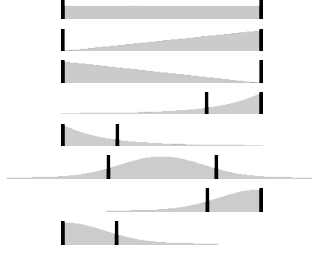


Figure 1: Range of two-parameter processes or groupings supported in OxCal v4.0 from Bronk Ramsey (2009a).

parameter process models are effectively a way to mathematically describe a group of events. As an example if we wish to model an exponential rise we can characterise this with two parameters. We have choice over what these parameters are, but to keep the model similar to that for a uniform phase we choose to take a midpoint,  $t_a$ , of the process (that is one time constant before the end) and the end-point,  $t_b$ , as being the two event parameters. The overall prior becomes:

$$p(\mathbf{t}) \propto p_H(t_a, t_b) \prod_i p_H(t_i, t_b) \frac{\exp(-(t_b - t_i)/(t_b - t_a))}{(t_b - t_a)} \quad (34)$$

Conceptually the two parameter process is central to all of the more complex models. The reason for this is that in a model which has two independent parameters  $t_a$  and  $t_b$  each with a uniform unconstrained prior  $\sim \mathcal{U}(-\infty, \infty)$  the effective prior for both the centroid  $(t_a + t_b)/2$  is  $\sim \mathcal{U}(-\infty, \infty)$  and for the span  $s = |t_b - t_a|$  is  $\sim \mathcal{U}(-0, \infty)$ . This property will be referred to here as ‘linear scale invariance’.

#### 5.4 Constrained processes

We can put together our two formalisms for constraint and two parameter process to provide a grouped sequence. Most commonly we have a uniform phase of events for which the order is pre-defined (usually by stratigraphic constraint). Adapting the prior from equation 33, where we take  $i = 1 \dots n$ , we have:

$$p(\mathbf{t}) \propto \frac{p_H(t_a, t_1, \dots, t_n, t_b)}{(t_b - t_a)^n} \quad (35)$$

In principle there is no mathematical reason why this cannot be extended to other types of process, though it is hard to see the archaeological situation where this would be relevant.

#### 5.5 Multiple phases

The model for the two-parameter process which describes a single uniform phase can easily be expanded to include the case of multiple phases. Where the phases are entirely independent, the prior is proportional to a product of the priors for the individual phases. Thus if, for example, we have two phases, one of which is from  $t_a$  to  $t_b$  and another from  $t_c$  to  $t_d$ , then the prior would be given by:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \quad (36)$$

Here there is no relationship between the phases at all (Figure 2A). We might also know that the second phase follows the first with an unknown gap (the phases are sequential, as in Figure 2B). This would give us an additional element to the prior of  $p_H(t_b, t_c)$ :

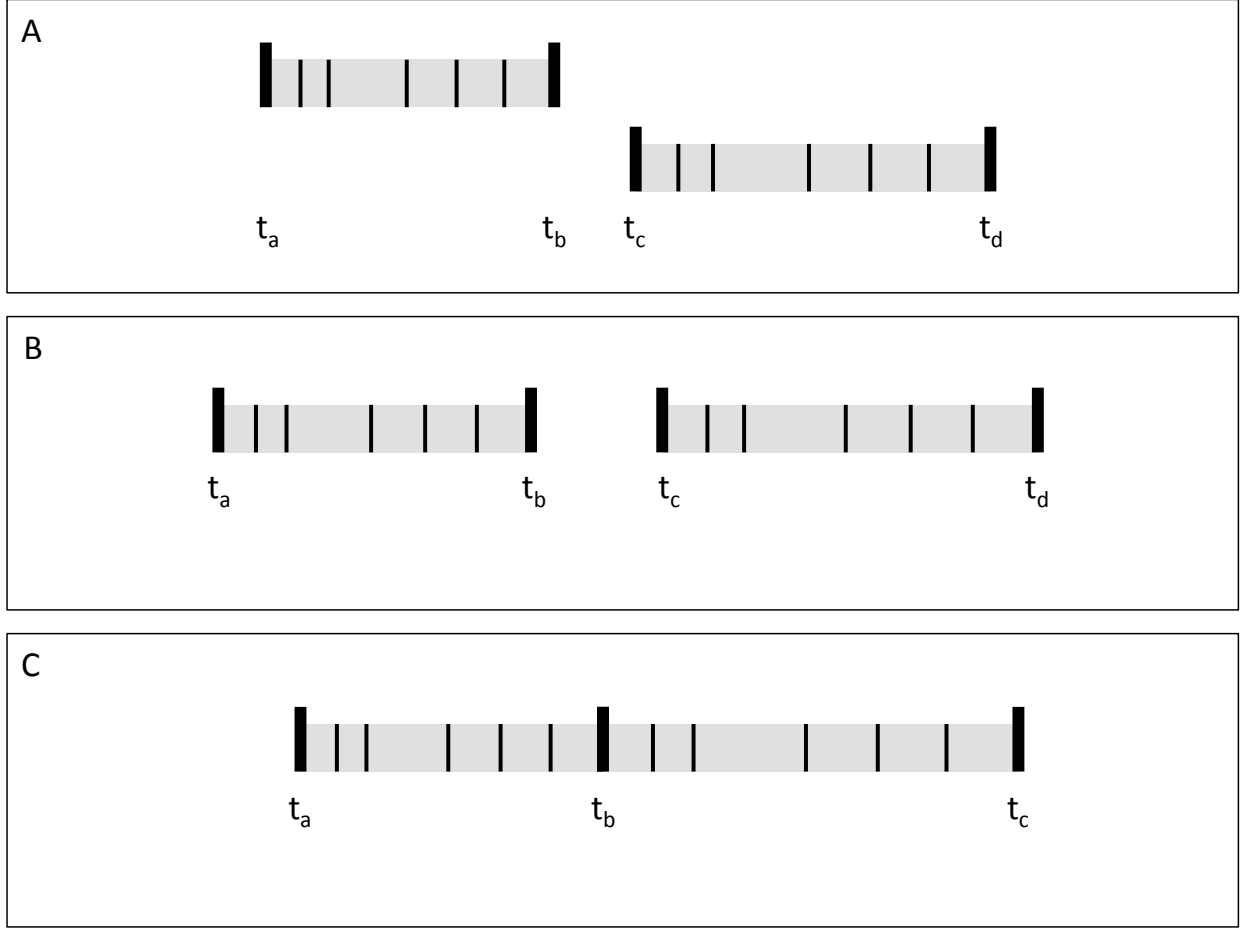


Figure 2: Examples of multiple phase models with just two phases. Two such phases can be modelled to be independent (A), sequential (B) or abutting (C).

$$p(\mathbf{t}) \propto p_H(t_b, t_c) \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \quad (37)$$

Another common interpretation of the archaeological evidence might be that the phases follow directly on from one another (they are abutting as in Figure 2C). In this case we have one less boundary parameter because the end of the first phase  $t_b$  would be the start of the second phase, which ends at  $t_c$ . In this case our prior is:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_b, t_j, t_c)}{(t_c - t_b)} \quad (38)$$

The first two of these (equations 36 and 37) are effectively four parameter process models, whereas the last (equation 38) is a three-parameter process model.

Now in the case of the dependent phases (examples of which are given in equations 37 and 38), we no longer have linear scale invariance between the earliest and latest boundaries in the model. This is exactly the same problem as we saw for independent events discussed in section 5.1 above. We can solve this problem 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 normalised prior of:

$$p(t_k|t_a, t_m) = \frac{p_H(t_a, t_k, t_m)}{(t_m - t_a)} \quad (39)$$

Thus equation 37 for the sequential phase example (Figure 3B) becomes:

$$p(\mathbf{t}) \propto p_H(t_b, t_c) \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \frac{1}{(t_d - t_a)^2} \quad (40)$$

and equation 38 for the abutting phase example (Figure 3C), becomes:

$$p(\mathbf{t}) \propto \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_b, t_j, t_c)}{(t_c - t_b)} \frac{1}{(t_c - t_a)} \quad (41)$$

Thus what we have done is to add another two process model at a higher level which shares parameters with the start of the first phase and the end of the last. This process model is used for all of the intervening phase boundaries.

## 5.6 Model heirarchies

In deriving the multiple parameter process models in the previous section, we have already introduced the notion of a hierarchy in that the process parameters (or boundaries) are taken to be higher level elements of the chronology under consideration. This can in principle be generalised to cover the nesting of processes. As an example we will consider the simple case of two independent phases described by the prior in equation 36. We would introduce a two-parameter ( $t_e$  and  $t_f$ ) process model to cover the four boundary events.

$$p(\mathbf{t}) \propto \frac{p_H(t_e, t_a, t_b, t_f)p_H(t_e, t_c, t_d, t_f)}{(t_f - t_e)^4} \prod_i \frac{p_H(t_a, t_i, t_b)}{(t_b - t_a)} \prod_j \frac{p_H(t_c, t_j, t_d)}{(t_d - t_c)} \quad (42)$$

This model (Figure 3A) will now be scale invariant again once more which the independent phase model is not on its own. In general parameterising the model for overall linear scale invariance makes most difference when there are a large number of phases in the model, but it is a good principle on which to base model definitions. Within OxCal such model hierarchies can be realised by nesting one set of phase models within another (Bronk Ramsey, 2009a).

Now if we look back at the modifications made in equations 40 and 41, they are actually just a special case of the nested model where the start and end boundaries of the outer model are constrained to be the same as the first and last boundaries within the sequence (compare Figures 2 and 3). Our justification for building models in this way needs ultimately to come from our understanding of human activity as a complex system which goes through periods of equilibrium and change on a range of different temporal scales (Bronk Ramsey, 2003).

## 5.7 The trapezium process model

So far we have considered only processes which in their individual elements can be described by two parameters as illustrated in Figure 1. However, there are some archaeological processes that might better be described with a more complex process model. The case of this kind which has been studied in most detail is the trapezium process model (Karlsberg, 2006; Lee and Bronk Ramsey, 2012). In this model we assume that the underlying archaeological activity has a period where the activity grows (at a linear rate), a period of stasis, followed by a period of decline (again linear) as shown in Figure 4. The process can be described by four parameters (Karlsberg, 2006), either as the four change points ( $t_a, t_b, t_c, t_d$ ) under the constraint  $t_a < t_b < t_c < t_d$  where the activity follows a prior:

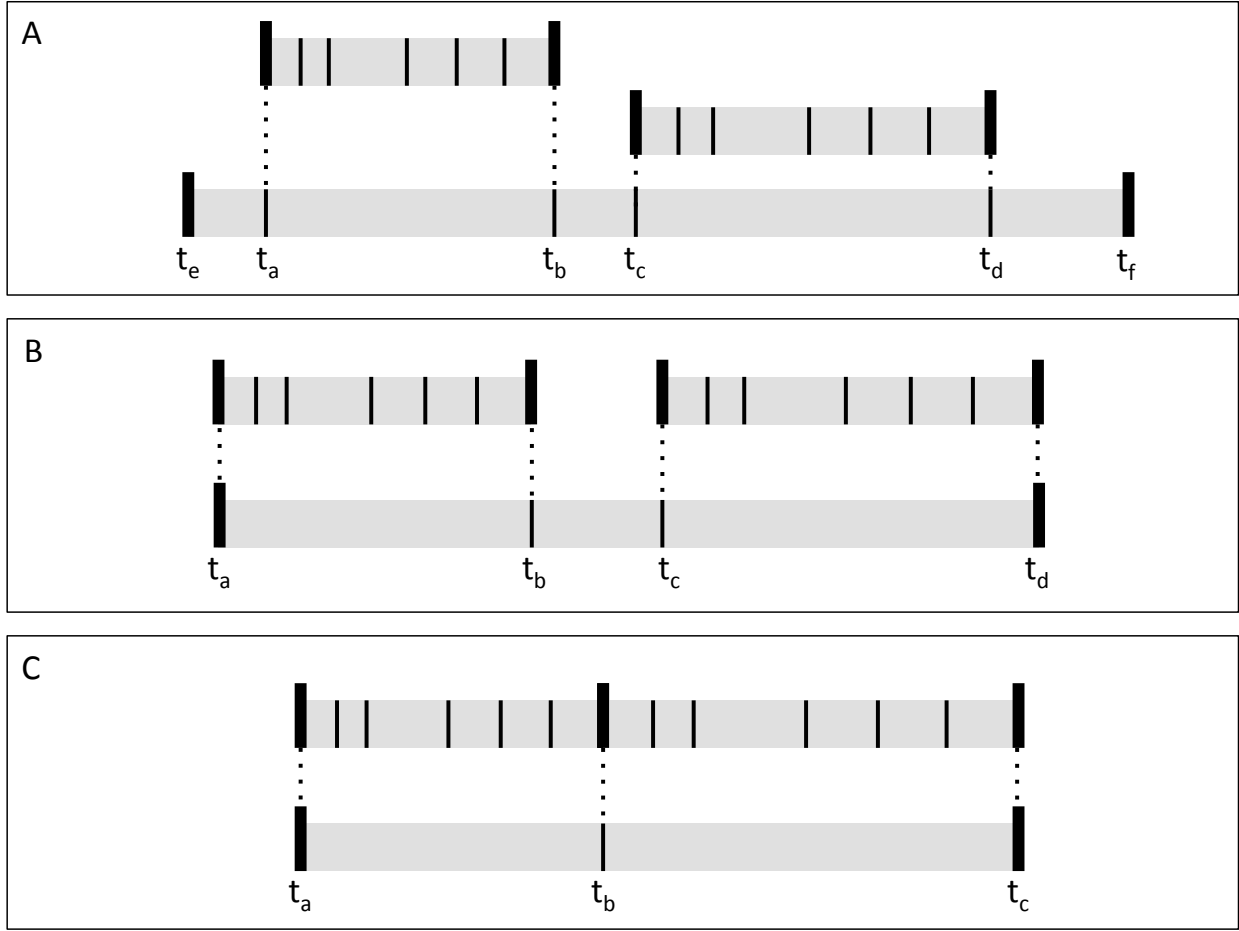


Figure 3: Examples of multiple phase models with just two phases but modelled using a two parameter process to describe the phase boundaries. The independent (A), sequential (B) or abutting (C) phases should be compared to Figure 2.

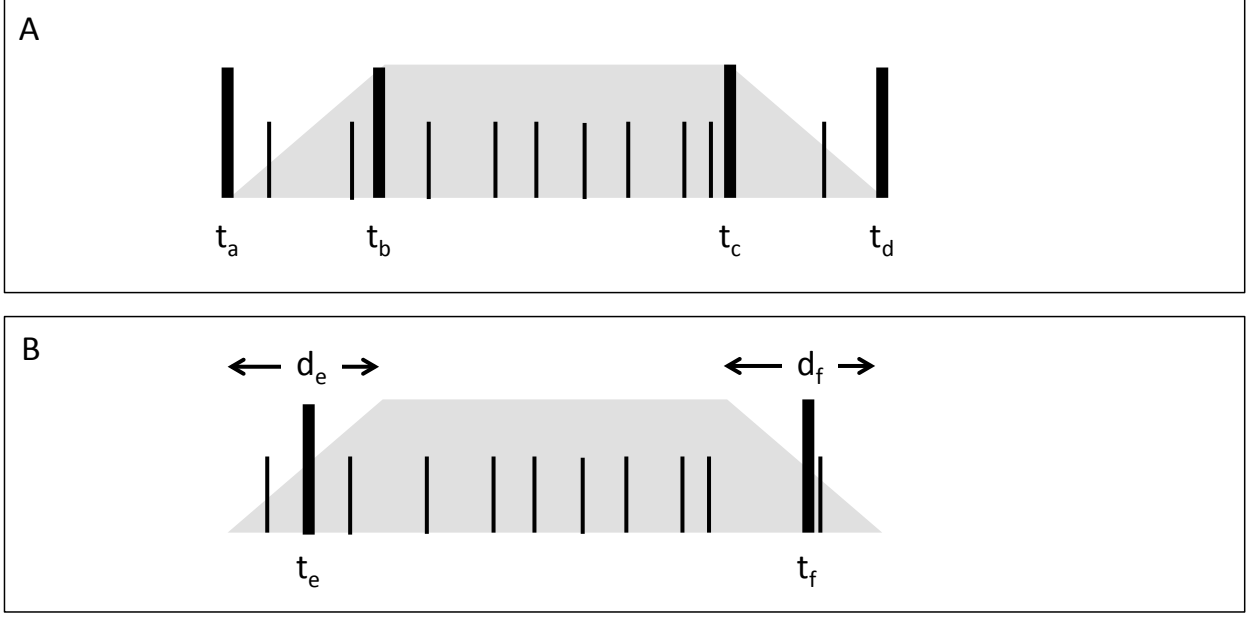


Figure 4: Formulations of the trapezium model (see text).

$$p(t) \propto \begin{cases} 0 & \text{for } t < t_a \\ \frac{t-t_a}{t_b-t_a} & \text{for } t_a < t < t_b \\ 1 & \text{for } t_b < t < t_c \\ \frac{t_d-t}{t_d-t_c} & \text{for } t_c < t < t_d \\ 0 & \text{for } t > t_d \end{cases} \quad (43)$$

Alternatively it can be re-parameterised to two time parameters ( $t_e = (t_a + t_b)/2$ ,  $t_f = (t_b + t_c)/2$ ) which give the centre point of the onset and end of the phase of activity, and two more ( $d_e = (t_b - t_a)$ ,  $d_f = (t_d - t_c)$ ) which give the period for the rise and fall in activity respectively (Lee and Bronk Ramsey, 2012). The main point of this parameterisation is for cases where you have a succession of such phases, but it also emphasises the underlying similarity with the two-parameter uniform phase. The issues involved in ensuring linear scale invariance in such a model are more complicated than the uniform phase case and discussed in detail in Lee and Bronk Ramsey (2012).

This type of model lends itself better to changes in cultural activity assumed to be gradual in nature.

## 5.8 Sedimentation

So far the models that we have considered are based either on our understanding of archaeological processes, combines with constraints which normally come from stratigraphy. However, stratigraphic order is only one type of information which can be derived from considering structured deposits. The two most relevant examples for archaeology are tree-rings in wood (or derived charcoal) and sedimentary sequences.

In the case of wood with countable tree-rings we have a very special situation where we know the relationship between different radiocarbon samples measured for radiocarbon. This means that effectively there is only one independent variable in the model (typically taken to be the age of the outer ring). So if  $t_i$  and  $t_{i+1}$  are the ages of two samples taken from the wood where

there are  $g_i$  whole rings separating then, then:

$$t_{i+1} = t_i + g_i \quad \text{for } i < n \quad (44)$$

where there are  $n$  samples in total. This simple relationship provides an easy way to ‘wiggles-match’ radiocarbon sequences measured on wood onto the calibration curve (Bronk Ramsey et al., 2001; Christen, 2003). Given that in this particular case, we do know the overall length of the group of dates, and because there is only one independent variable involved, there are no issues of linear scale invariance involved here. There is a somewhat analogous situation in varved lakes and ice-cores, except that in those instances, there is often some uncertainty associated with the gap between samples (Bronk Ramsey, 2008). This is rarely directly relevant to archaeology and so will not be discussed further here.

The much more common situation in archaeological contexts is sedimentary deposits including peat formation. In these cases we have stratigraphic ordering, but we also expect the deposition rate to be reasonably regular. This is to be contrasted with the situation in many archaeological sites (urban settlements, cave shelters etc) where there may be very sporadic changes in level, sometimes with large increments in level in short periods and at other times erosion or stasis.

For this discussion we will consider the timing of deposition  $t$  at a vertical position in the final deposit (after compaction etc.) of  $z$ . In order to simplify the discussion here we will consider the model implications of interpolating between two points in a sequence with timing  $t_i$  and  $t_{i+1}$  at positions  $z_i$  and  $z_{i+1}$  respectively. We will consider the timing of the deposition in the range  $z_i < z < z_{i+1}$ .

At one extreme, if we expect very variable deposition rates, then as above, all we can use is stratigraphic constraint which tells us that:

$$p(t|z) = \begin{cases} \frac{1}{t_{i+1}-t_i} & t_i < t < t_{i+1} \\ 0 & \text{otherwise} \end{cases} \quad (45)$$

or this could be included in the prior as a constraint:

$$p(t) \propto p_H(t_i, t, t_{i+1}) \quad (46)$$

However, the implication of this model is that we get no useful chronological information from the depth at all: a sample near the top of this range is just as likely to be similar in age to the bottom as the top. This is really only reasonable if we think that there is only one deposition event between  $z_i$  and  $z_{i+1}$ .

Another extreme assumption is that the deposition between the two points proceeds perfectly uniformly. In this case only  $t$  can be directly calculated from the  $t_i$  and  $t_{i+1}$  and there are only two independent variable in the model:

$$t = t_i + (t_{i+1} - t_i) \frac{(z - z_i)}{(z_{i+1} - z_i)} \quad (47)$$

In practice, however, no deposition is truly continuous or constant in rate, and so although this might be a convenient mathematical model it is not realistic. What is required is a process model for the deposition itself. One approach is to assume that the deposition is actually governed by an underlying Poisson process with discrete deposition events occurring at random (Bronk Ramsey, 2008). For this we assume that there are a certain number of events per unit length given by the parameter  $k$ . So if for example our  $z$  is measured in meters and  $k = 1000$ , this model is like having a simple sequence deposition model with an event every 1mm. Under this assumption our interpolated age probability for position  $z$  is:

$$p(t|z, k) \propto \frac{(t - t_i)^{k(z-z_i)} (t_{i+1} - t)^{k(z_{i+1}-z)}}{(t_{i+1} - t_i)^{k(z_{i+1}-z_i)}} \quad (48)$$

which reaches a maximum when:

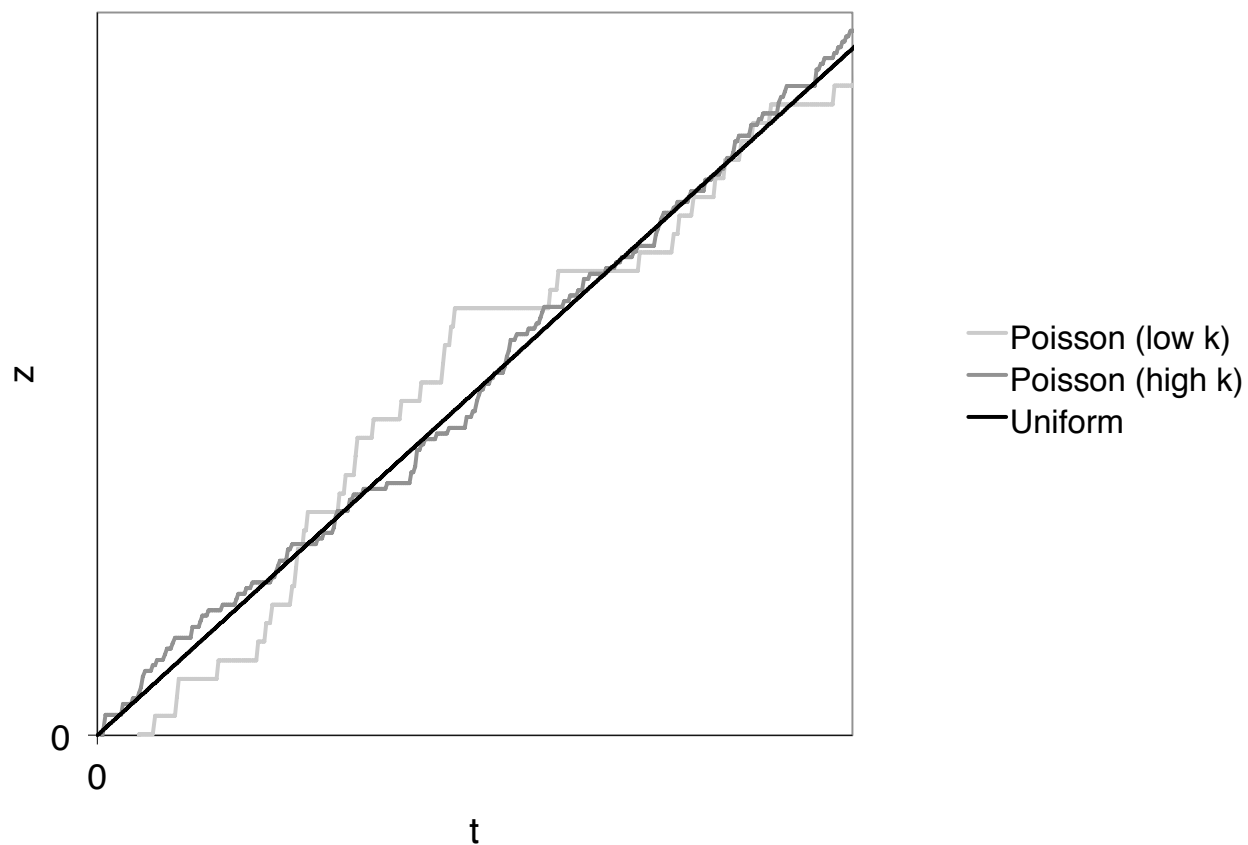


Figure 5: Comparison of deposition model assumptions from Bronk Ramsey (2008). The Poisson process model allows for variability in deposition rate by assuming that the underlying deposition process is incremental rather than continuous.



$$t = t_i + (t_{i+1} - t_i) \frac{(z - z_i)}{(z_{i+1} - z_i)} \quad (49)$$

as for the uniform deposition case. Indeed the two extreme cases of unstructured deposition and uniform deposition are simply limits of this model in the limit where  $k$  tends to zero or infinity respectively (see Figure 5). The details of this model, including ways to ensure that the overall model is scale in variation, and dealing with the case where the underlying sedimentation is not constant is dealt with in Bronk Ramsey (2008). One difficulty of this model is the need to work out the most appropriate  $k$  value. However, this problem can be overcome by allowing  $k$  to be a parameter in the model itself, and allow the data to inform us on the uniformity of deposition (Bronk Ramsey and Lee, 2013).

The Poisson process is however, only one approach that can be taken. The other most widely used approach is to assume a gamma process (Blaauw and Christen, 2005; Haslett and Parnell, 2008). In practice, given that all of these models provide a simplified model of the underlying processes, there is a lot to be said for testing several different models to check for the robustness of the findings.

## 6 Model implementation

So far we have considered only the mathematical formulation of Bayesian models relevant to archaeological research. The fact that such models have become so widely used only over the last couple of decades, is in large part because they cannot be solved analytically and require Markov Chain Monte Carlo methods to calculate posterior probability densities.

Several statistical packages have been developed which allow the user to use Bayesian models together with chronological data, including BCal (Buck et al., 1999), OxCal (Bronk Ramsey, 1995, 2008, 2009a,b), BWigg (Christen, 2003), BChron (Haslett and Parnell, 2008) and BPeat (Blaauw et al., 2003). Each of these packages have slightly different approaches to the formulation of models and different algorithms for the implementation of the MCMC. However, in many cases the underlying mathematical models are similar, though many packages only implement certain types of model.

### 6.1 Reporting model outputs

The main point of any Bayesian modelling is to obtain posterior estimates for the model parameters. Most often this is in the form of marginal posterior density estimates for events within the chronological model. Such outputs are normally given as probability distributions (binned at a suitable resolution), and summarised in terms of highest probability density (HPD) ranges. Unlike many applications, it has been common practice to generate split ranges rather than just a single HPD range. This means that, for example, the 95.4% range is defined as the set of subranges with the shortest combined length which contains 95.4% of the probability in the distribution. The reason that this is common practice is partly as a legacy of the intercept method of calibrating radiocarbon dates, and also because the extra information in these split ranges is often quite useful in interpreting chronologies. However, there are some disadvantages in this approach in that range fragmentation can result from noise in the probability distribution functions derived from MCMC analysis (to an extent which depends on bin-size).

Other information can also be pulled out from Bayesian models of this kind, as part of the MCMC analysis. It is for example possible to generate probability distributions for the first of the group of events, the difference between two events or the covariance between model parameters (Bronk Ramsey, 2009a; Bronk Ramsey and Lee, 2013). Sometimes this information on the relationship between model parameters can be as important as the age estimates themselves.

## 6.2 Model output comparison

Given the output of any analysis is primarily given as probability distributions, it is worth considering, from a mathematical perspective, what the most appropriate methods are for comparing such distributions.

One application of such a comparison might be to compare the model outputs from two different models, or from two runs of the same model to see how similar they are. Supposing we have two different marginal posterior density estimates  $p_i(t)$  and  $p_j(t)$ . For comparison purposes a suitable mathematical approach might be to look for a scalar product (or inner product). As the functions are real, the inner product will require, symmetry, linearity (on either argument) and that the product is positive definite. Since the probability is always positive this is easily achieved as:

$$\langle p_i, p_j \rangle = \frac{\int p_i(t)p_j(t)dt \int p_i(t)dt \int p_j(t)dt}{\sqrt{\int p_i(t)p_i(t)dt \int p_j(t)p_j(t)dt}} \quad (50)$$

The properties of this scalar product are:

$$\langle p_i, p_j \rangle = \langle p_j, p_i \rangle \quad (51)$$

$$\langle ap_i, p_j \rangle = \langle p_i, ap_j \rangle = a \langle p_i, p_j \rangle \quad (52)$$

$$\langle p_i, p_j \rangle > 0 \quad (53)$$

$$\langle p_i, p_i \rangle = \left( \int p_i(t)dt \right)^2 \quad (54)$$

$$\langle p_i, p_j \rangle = 0 \quad \text{if } \int p_i(t)p_j(t)dt = 0 \quad (55)$$

Furthermore in the usual situation in which  $p_i(t)$  and  $p_j(t)$  are each normalised we have:

$$\langle p_i, p_j \rangle = \frac{\int p_i(t)p_j(t)dt}{\sqrt{\int p_i(t)p_i(t)dt \int p_j(t)p_j(t)dt}} \quad (56)$$

$$\langle p_i, p_i \rangle = 1 \quad (57)$$

This inner product would be useful in comparing model outputs. One place where it is used is for the convergence measure used in the OxCal program (Bronk Ramsey, 1995) which is just  $\langle p_i, p_j \rangle^2$  from equation 56.

The other application of distribution comparison is for evaluating the likelihood distributions in relation to the posteriors. It is important that we do this because if the overlap between the likelihood and posterior distributions is very small it implies that the model does not agree with the observations. In this case we do not expect the distributions to be the same, and the likelihood and marginal posterior are different elements of the model. For this reason the symmetrical inner product is not appropriate. One approach which has been found to be very useful in practice is the integral of the likelihood over the posterior. For an individual parameter we can compare this integral for marginal density from our full Bayesian model  $p_1(t_i|\mathbf{y})$  to that for the marginal density from the null model described above in section 5.1 which is  $p_0(t_i|y_i) = p(y_i|t_i)$ . The ratio that we get from this:

$$F_i = \frac{\int p(y_i|t_i)p_1(t_i|\mathbf{y})dt_i}{\int p(y_i|t_i)p_0(t_i|\mathbf{y})dt_i} = \frac{\int p(y_i|t_i)p_1(t_i|\mathbf{y})dt_i}{\int p(y_i|t_i)p(y_i|t_i)dt_i} \quad (58)$$

has the property that it is 1 if the marginal posterior is the same as the likelihood. It might be higher than 1 if the marginal posterior is concentrated in the highest portions of the likelihood distribution. This factor gives us the ratio of the likelihood of this particular parameter under the two models and so if the factor is very low it implies that the model is unlikely. This measure can be used to identify outliers and in Bronk Ramsey (1995), the agreement index is  $100F_i$ . The

factor  $F_i$  only makes use of the information for one parameter in the model, but the approach can be generalised to provide a factor for the model as a whole:

$$F_{model} = \frac{\int p(\mathbf{y}|\mathbf{t})p_1(\mathbf{t}|\mathbf{y})d\mathbf{t}}{\int p(\mathbf{y}|\mathbf{t})p_0(\mathbf{t}|\mathbf{y})d\mathbf{t}} \quad (59)$$

Although we might be tempted to use these factors to compare two models (since by taking a ratio between them we eliminate the null model from the equation), it should be noted that they are not formal Bayes factors and in general more constrained (but possibly more realistic models) will often have lower  $F_{model}$  values. In particular many models have an  $F_{model} < 1$  and yet are certainly more appropriate than the null model which assumes complete independence of all the events.

Model comparison is an area where further research and development is certainly needed.

## 7 Conclusion

The application of mathematical methods to this aspect of archaeology has been a very fruitful one. Not only has it been useful in providing quantitative archaeological chronologies, but also in making us think about the very nature of those chronologies and the assumptions that underly them. The process of formalising our approach to chronologies makes us define what we mean by the time-line, which may be very different to the way people perceive age and date in much non-quantitative analysis. It also helps us to think what is an appropriately neutral set of assumptions to apply when we come to study the processes revealed in archaeology. The notion of the two parameter process which has ‘linear scale invariance’ is something which is central to almost all the approaches discussed here; in the end the reason that we choose this as an approach is that both absolute time and duration of process are central to much that we wish to study through chronology. A hierarchical approach to model building is also something which ultimately comes from treating human activity as an inherently complex process with scales of structure visible at many different levels. The approaches taken in this field also highlight how we think about comparison of results with uncertainty more generally, and specifically in cases where the usual classical statistical approaches (based for example on assumptions of normality) cannot be used.

There are still many areas where further research and ideas are needed. These are particularly in the areas of model comparison where we would like much more powerful tools to allow us to test different ideas through chronological models. A related issue is what to do when we don’t have a model of what is going on archaeologically and we just want the data to inform us about the chronology of underlying changes. In such cases the null model is still probably a poor representation of the processes and thus simply looking at summed probability distributions is not a good solution.

There is now a good framework for analysis of chronology using Bayesian models. This provides a very useful basis for archaeologists wishing to develop novel models relevant to their research, and for mathematicians and statisticians interested in extending these methods and improving our understanding of the fundamental nature of the models that we employ.

## References

- Blaauw, M., Christen, J. A., 2005. Radiocarbon peat chronologies and environmental change. *Journal of the Royal Statistical Society Series C-Applied Statistics* 54, 805–816.
- Blaauw, M., Heuvelink, G. B. M., Mauquoy, D., van der Plicht, J., van Geel, B., 2003. A numerical approach to C-14 wiggle-match dating of organic deposits: best fits and confidence intervals. *Quaternary Science Reviews* 22 (14), 1485–1500.

- Bronk Ramsey, C., 1995. Radiocarbon calibration and analysis of stratigraphy: The OxCal program. *Radiocarbon* 37 (2), 425–430.
- Bronk Ramsey, C., 2001. Development of the radiocarbon calibration program OxCal. *Radiocarbon* 43 (2A), 355–363.
- Bronk Ramsey, C., 2003. Punctuated dynamic equilibria: a model for chronological analysis. In: Bentley, R. A., Maschner, H. D. G. (Eds.), *Complex Systems and Archaeology*. University of Utah Press, pp. 85–92.
- Bronk Ramsey, C., 2008. Deposition models for chronological records. *Quaternary Science Reviews* 27 (1-2), 42–60.
- Bronk Ramsey, C., 2009a. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51 (1), 337–360.
- Bronk Ramsey, C., 2009b. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* 51 (3), 1023–1045.
- Bronk Ramsey, C., Lee, S., 2013. Recent and Planned Developments of the Program OxCal. *Radiocarbon* 55 (2-3), 720–730.
- Bronk Ramsey, C., van der Plicht, J., Weninger, B., 2001. 'Wiggle matching' radiocarbon dates. *Radiocarbon* 43 (2A), 381–389.
- Buck, C. E., Cavanagh, W. G., Litton, C. D., 1996. *Bayesian approach to interpreting archaeological data*. Wiley, Chichester.
- Buck, C. E., Christen, J. A., James, G. N., 1999. BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeology* 7.
- Buck, C. E., Kenworthy, J. B., Litton, C. D., Smith, A. F. M., 1991. Combining Archaeological and Radiocarbon Information - a Bayesian-Approach to Calibration. *Antiquity* 65 (249), 808–821.
- Buck, C. E., Litton, C. D., Smith, A. F. M., 1992. Calibration of Radiocarbon Results Pertaining to Related Archaeological Events. *Journal of Archaeological Science* 19 (5), 497–512.
- Buck, C. E., Millard, A., 2004. *Tools for constructing chronologies : crossing disciplinary boundaries*. Springer, London.
- Christen, J. A., 1994. Summarizing a Set of Radiocarbon Determinations - a Robust Approach. *Applied Statistics-Journal of the Royal Statistical Society Series C* 43 (3), 489–503.
- Christen, J. A., 2003. Bwigg: an internet facility for Bayesian radiocarbon wiggle matching. *Internet Archaeology* 7.
- Christen, J. A., Pérez, E. S., 2009. A new robust statistical model for radiocarbon data. *Radiocarbon* 51 (3), 1047–1059.
- Gelfand, A. E., Smith, A. F. M., 1990. Sampling-Based Approaches to Calculating Marginal Densities. *Journal of the American Statistical Association* 85 (410), 398–409.
- Gilks, W. R., Richardson, S., Spiegelhalter, D. J., 1996. *Markov chain Monte Carlo in practice*. Chapman & Hall, London.
- Harris, E. C., 1989. *Principles of archaeological Stratigraphy*. Academic Press, London.

- Haslett, J., Parnell, A., 2008. A simple monotone process with application to radiocarbon-dated depth chronologies. *Journal of the Royal Statistical Society: Series C (Applied Statistics)* 57 (4), 399–418.  
URL <http://dx.doi.org/10.1111/j.1467-9876.2008.00623.x>
- Hogg, A. G., Hua, Q., Blackwell, P. G., Niu, M., Buck, C. E., Guilderson, T. P., Heaton, T. J., Palmer, J. G., Reimer, P. J., Reimer, R. W., Turney, C. S. M., Zimmerman, S. R. H., 2013. SHCal13 Southern Hemisphere Calibration, 0-50,000 Years cal BP. *Radiocarbon* 55 (4).
- Hua, Q., Barbetti, M., Rakowski, A. J., 2013. Atmospheric Radiocarbon for the Period 1950-2010. *Radiocarbon* 55 (4).
- Ivanovich, M., Harmon, R. S., 1992. Uranium-series disequilibrium: Applications to earth, marine, and environmental sciences. Oxford University Press.
- Jones, M., Nicholls, G., 2002. New radiocarbon calibration software. *Radiocarbon* 44 (3), 663–674.
- Karlsberg, A. J., 2006. Statistical modelling for robust and flexible chronology building. Ph.D. thesis, University of Sheffield.
- Lee, S., Bronk Ramsey, C., 2012. Development and Application of the Trapezoidal Model for Archaeological Chronologies. *Radiocarbon* 54 (1), 107–122.
- Miles, D. H., 2005. New Developments in the Interpretation of Dendrochronology as Applied to Oak Building Timbers. Ph.D. thesis, University of Oxford.
- Nicholls, G., Jones, M., 2001. Radiocarbon dating with temporal order constraints. *Journal of the Royal Statistical Society Series C-Applied Statistics* 50, 503–521.
- Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Bronk Ramsey, C., Grootes, P. M., Guilderson, T. P., Haffidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M., van der Plicht, J., 2013. IntCal13 and Marine13 Radiocarbon Age Calibration Curves 0-50,000 Years cal BP. *Radiocarbon* 55 (4).
- Rhodes, E. J., Bronk Ramsey, C., Outram, Z., Batt, C., Willis, L., Dockrill, S., Bond, J., 2003. Bayesian methods applied to the interpretation of multiple OSL dates: high precision sediment ages from Old Scatness Broch excavations, Shetland Isles. *Quaternary Science Reviews* 22 (10-13), 1231–1244.
- van der Plicht, J., 1993. The Groningen Radiocarbon Calibration Program. *Radiocarbon* 35 (1), 231–237.
- Zeidler, J. A., Buck, C. E., Litton, C. D., 1998. Integration of Archaeological Phase Information and Radiocarbon Results from the Jama River Valley, Ecuador: A Bayesian Approach. *Latin American Antiquity* 9, 160–179.