

---

# **Bradshaw and Bayes: Towards a Timetable for the Neolithic**

**Alex Bayliss, Christopher Bronk Ramsey,  
Johannes van der Plicht & Alasdair Whittle**

*The importance of chronology is reasserted as a means to achieving history and a sense of temporality. A range of current methods for estimating the dates and durations of archaeological processes and events are considered, including visual inspection of graphs and tables of calibrated dates and the summing of the probability distributions of calibrated dates. These approaches are found wanting. The Bayesian statistical framework is introduced, and a worked example presents simulated radiocarbon dates as a demonstration of the explicit, quantified, probabilistic estimates now possible on a routine basis. Using this example, the reliability of the chronologies presented for the five long barrows considered in this series of papers is explored. It is essential that the ‘informative’ prior beliefs in a chronological model are correct. If they are not, the dating suggested by the model will be incorrect. In contrast, the ‘uninformative’ prior beliefs have to be grossly incorrect before the outputs of the model are importantly wrong. It is also vital that the radiocarbon ages included in a model are accurate, and that their errors are correctly estimated. If they are not, the dating suggested by a model may also be importantly wrong. Strenuous effort and rigorous attention to archaeological and scientific detail are inescapable if reliable chronologies are to be built. The dates presented in the following papers are based on models. ‘All models are wrong, some models are useful’ (Box 1979, 202). We hope readers will find them useful, and will employ ‘worry selectivity’ to determine whether and how each model may be importantly wrong. The questions demand the timetable, and our prehistories deserve both.*

It may seem reactionary and perverse to reaffirm, as I do, at the beginning of a book on archaeology in the field that mere dates are still of primary and ultimate and unrelenting importance. And by dates I mean not simply those nebulous phases and sequences, those date-substitutes, with which archaeologists often try to bluff us. I mean time in hard figures. I mean Bradshaw (Wheeler 1956, 38).

A ‘Bradshaw’ was a colloquial designation of Bradshaw’s *Railway Guide*, a timetable of all railway trains running in Great Britain, printed annually from 1839–1961.<sup>1</sup> It is salutary, 50 years and several radiocarbon revolutions after Wheeler wrote, that we feel it

is necessary to introduce a series of papers on a group of long barrows in southern England with a justification of why we have devoted so much time and energy to unravelling the details of their chronology. And we will not bluff you with ‘date-substitutes’; by chronology we do mean Bradshaw — explicit, quantitative, probabilistic estimates of real dates when things happened by the agency of particular people in specific places in Neolithic southern England.

Chronology is fundamental to archaeological study. First, it allows variations in the archaeological record which depend upon time to be distinguished from those determined by other factors. In this case chronologies can be relative or absolute, since what

is required is sequence. This allows the comparison of contemporary sites or bodies of evidence, and allows changes in the material record to be tracked through time. Calendrical time is used simply to reveal sequence, with the advantage that it enables comparison of evidence which cannot be connected by stratigraphic, typological, or cultural series. Whereas Wheeler (1956, 39) was concerned with comparing 'cultures of different regions', a contemporary archaeologist might be more concerned with correlating environmental evidence from an offsite pollen core with the human activities revealed in an excavation trench. The questions change, but the fundamental concern with chronological sequence does not.

A date is just a number — a radiocarbon date just an expensive number! The elucidation of chronology as an end in itself is sterile; it is the use of chronology to enable fundamental questions of archaeological research to be addressed which is important. One influential account of the differences with which we are concerned has been given by Tim Ingold (1993). He made important distinctions between chronology and history on the one hand, and temporality on the other. Chronology is 'any regular system of dated time intervals, in which events are said to have taken place', whereas history is 'any series of events which may be dated in time according to their occurrence in one or other chronological interval' (Ingold 1993, 157). He goes on, 'In the mere succession of dates there are no events, because everything repeats; in the mere succession of events there is no time, as nothing does'. To both unite and supplant chronology and history, Ingold introduces the term of temporality, following the distinctions made by Gell (1992), after McTaggart (1908), between A-series time and B-series time (for a parallel distinction, between substantial and abstract time, see Shanks & Tilley 1987, 128; reviewed also in Lucas 2005). In the B-series, 'events are strung out in time like beads on a thread' and 'are treated as isolated happenings, succeeding one another frame by frame' (Ingold 1993, 157). In the A-series, by contrast, 'time is immanent in the passage of events', and each event 'is seen to encompass a pattern of retensions from the past and protentions for the future. Thus from the A-series point of view, temporality and historicity are not opposed but rather merge in the experience of those who, in their activities, carry forward the process of social life'. In other words, we need to move from the measurement of elapsed time, to a sense of successive events, and then to how people experienced the flow of time and saw themselves in time, both looking back to the past and forward to the future.

This raises tensions. What then if people's embodied experience of life makes little reference to measured or calculated time, but is rather 'an immersion in a substantial time ... a submission to the passage of time' (Shanks & Tilley 1987, 1278)? According to this view, 'chronology does not explain, nor does it provide a context. It is part of that which is to be understood' (Shanks & Tilley 1987, 127). The resolution of this dilemma may rest in Alfred Gell's claim (1992, 293) that every system of A-series time (whether linear or cyclical) is, in the end, underpinned by a B-series. Neither Ingold nor Shanks & Tilley explained how an A-series system of time was to be investigated. So here is the second way in which chronology is essential for archaeological understanding, but it has received much less attention. This is the need to determine the tempo of cultural change, the duration of activities seen in the archaeological record and (in agreement with both Ingold and Shanks & Tilley), the nature of temporality or the lived experience and marking of time. But it is hard to see how we can begin to attempt these goals without a framework of calendar dating — the order and spacing of the beads on the string, as it were. The calendrical time scale allows an assessment of elapsed time — how long is significant as well as when. It is significant because an assessment of the duration of a process or an activity can inform our interpretations. For example, the implications for social memory of a rite which can be shown to have endured for several centuries are different from those of a rite which endured for just a few decades. The social anthropologist Johannes Fabian some time ago critiqued the anthropological practice of distancing the Other as temporally remote, rather than simultaneously engaged in the present (Fabian 1983). He called this 'allochronism'. Archaeology has committed similar mistakes, but in different ways. Despite the avowed interest in change and long-term process, it has often in effect also distanced its object (or subjects) as temporally remote but timeless because they lack precise settings, thus tending to universalize the human experience and deny the particularity of individual contexts.

These are the reasons why we feel chronology is central to archaeology. Why do we feel they need to be articulated? Although much lip-service is paid to the importance of reliable chronology by archaeologists, particularly for prehistoric periods, this does not seem to translate into serious, scholarly, attention being paid to the subject. The provision of a handful of radiocarbon dates, hopefully from increasingly carefully selected samples, is the best that the vast majority of sites can hope for (see <http://ads.ahds.ac.uk/catalogue/>; <http://www.historicscotland.gov.uk>; Meadows

*et al.* forthcoming). Some attention is being given to the reassessment of the quality and interpretative value of radiocarbon dates which have already been obtained (e.g. Waterbolk 1971; Whittle 1988; Kinnes *et al.* 1991; Garwood & Brindley 1999; Schulting 2000; Pettitt *et al.* 2003; Scarre *et al.* 2003), but quantitative assessments of the chronological implications of the existing data are few and far between (Housley *et al.* 1997; Blackwell & Buck 2003; Ashmore 2004). As calibration has been extended, however, consideration of uncalibrated radiocarbon determinations is now rare in all but Palaeolithic archaeology (although see Russell 2004). Synthetic studies of prehistory have as yet done little more than calibrate the existing radiocarbon measurements and graphically interpret their ranges (e.g. Gibson & Kinnes 1997; Darvill 2004). This means that we are still working in a timescale which can divide a millennium into three or four distinct blocks. Effectively, this is the chronology established by the second radiocarbon revolution, when the need for calibration became apparent, in the early 1970s.

As a discipline we can do better than this (e.g. Cleal *et al.* 1995, 511–35). We have been able to do better than this for over a decade (Buck *et al.* 1991; Bronk Ramsey 1995). So why are we not doing so on a more routine basis?

On one level, the application of sophisticated mathematical approaches to the interpretation of radiocarbon dates is one of the success stories of quantitative methods in archaeology over recent decades. Nowadays almost everyone calibrates; there is a range of freely distributed software available which allows a variety of numeric approaches to be applied to radiocarbon (and other chronometric) data on a routine basis (<http://radiocarbon.pa.qub.ac.uk/calib>; <http://www.rlaha.ox.ac.uk>; <http://www.rug.nl/ees/informatieVoor/cioKlanten>; <http://bcal.sheffield.ac.uk/>; <http://www.datelab.org/>; and archaeology is privileged to collaborate actively with researchers in a variety of disciplines for the further development of methods (Nicholls & Jones 2001; Buck *et al.* 2003; Blaauw *et al.* 2003). In 2006, for one of the most widely used and most versatile programs, OxCal, there were 800 registered users in more than 50 countries.

And yet in much of English prehistory we are still floundering in a chronology which is precise (at best) to within a few hundred years. Why has it taken us so long to grasp the potential for better chronologies? In our view there is a complex of reasons for this failure: some practical and some theoretical. Practical obstacles include archaeological unawareness of the potential for precise dating, the level of detailed study required to build such chronologies, the multi-

disciplinary knowledge required to construct them, and potentially the cost of a statistically-viable suite of radiocarbon measurements.<sup>2</sup> These difficulties are to some extent country-specific and so will not be discussed in detail. Theoretical issues are, however, of more general interest.

### Method: the handmaiden of theory

Since the early 1980s, varieties of postprocessual theory have been dominant in much of the British archaeological literature (Johnson 1999; Hodder & Hutson 2003). Perhaps as a reaction to the emphasis on methodology and science which predominated in the preceding decades, these studies are generally weak on, or unconcerned with, method (but see also Hodder 1999). Detailed chronologies, and indeed modelling more widely, have simply been unfashionable. This is not the entire story, however, since archaeology in the German-speaking world, for example, has also failed to take full advantage of the methodological developments which enable more precise chronologies to be built on a routine basis. The Germanic tradition of ‘pre- and proto-history’ is undoubtedly centrally interested in chronology (Härke 1991) and has actively maintained research using quantitative methods during their decades of fallow in Britain (e.g. Jørgensen 1992; Müller & Zimmermann 1997; Czebreszuk & Müller 2001; Lüning 2005; Strien & Gronenborn 2005; cf. Dubouloz 2003, in the most rigorous French tradition). There seems to be some discomfort, however, with the concept of chronologies; from this perspective there is a single, correct answer — methods that can give you more than one are suspect! It should be noted, however, that in archaeology there is little concern for the conflicting principles underlying the debate between classical and Bayesian statistics that are so prevalent in some other areas of study (Steel 2001). Archaeological objections to the Bayesian approach are pragmatic, not philosophical (Reece 1994; Ashmore 2005).

Although in practice the theoretical perspectives of contemporary archaeologists may have affected the adoption of new methodology, quantitative methods are founded within the explicit theoretical framework of mathematics and so, in archaeological terms, the actual algorithms involved are almost by definition theory-free. This does not mean that the *application* of quantitative approaches in archaeology is not fundamentally tied to the theoretical framework in which the archaeologist is working. Theory-dependent choices are made at every step: the selection of the archaeological questions to be addressed by the

analysis, the selection of appropriate algorithms or variables, and in the archaeological interpretation of the output of an analysis. These theory-based decisions are thrown into sharper relief when an analysis attempts to model formally the specific parameters of a particular problem, rather than implement a more generalized approach.

### The questions

The research described in the following papers was designed to explore a series of site-specific archaeological questions about the development and use of five southern British long barrows and cairns. The explicit definition of these objectives at the start of each project was essential for the construction of the sampling strategy for radiocarbon dating, and the models described have been built specifically to address these questions. Consequently our models are fundamentally dependent on the interpretive framework which drove our selection of research questions. Our approach has been conditioned by debate on the sites in question over at least the last 30 years, and indeed going as far back as the 1930s, including sequences and circumstances of monument construction, the nature of rites of deposition of human remains, and the timescales over which monuments were in use.

The objectives of each dating programme are given in detail in each of the site-specific papers in this series, although they can be summarized as follows:

- to provide calendrical dating for pre-barrow occupation and to determine whether there was continuity between this activity and the subsequent barrow;
- to determine the structural sequence of monuments, especially to determine the construction dates of barrows and chambers;
- to explore spatial differences in the chronology of mortuary deposits (e.g. in different chambers, from back to front in linear chambered areas);
- to determine chronological differences in different types of mortuary rites (e.g. articulated/disarticulated, green bone: see Wysocki *et al.* this issue);
- to estimate the duration of the period when human remains were interred in the monuments.

This concentration on site-specific questions enables the construction of refined chronologies that allow discussion of context, meaning, agency, and remembrance within a timescale relevant to human experience (see again discussion above, and Ingold 1993). We are able to consider these sites within the context of human lifespans and even generations, examining differences in mortuary rites and the incidence of older,

ancestral, remains in these terms. This precise dating also allows us to take account of the order in which things happened at the five sites, comparing the dates of construction, use, and disuse and the duration of burial at each site. We are also able to compare each site with other, broadly contemporary, sites in their localities and regions.

In the final paper in the series, a first attempt is made to discuss the results of these dating projects within the context of broader issues in British Neolithic studies. Inevitably this is limited by the fact that we have produced relatively precise dating for only five sites, and there are few other sites within the southern British Neolithic that have been dated to a similar resolution (e.g. Hillam *et al.* 1990; Mercer & Healy forthcoming; Evans & Hodder 2006). Despite these limitations, the relative order of the five sites is considered in relation to the typology of long cairns and long barrows, and we estimate how many such monuments may have been in use in an area at any one time. The date of the earliest of these monuments is discussed in relation to the dates of causewayed enclosures and other, potentially earlier, monument types, in relation to the origins of the tradition of monuments as a whole, and in comparison with alternative models of the introduction of the Neolithic way of life into England.

### Thomas Bayes and the hermeneutic spiral

The basic idea behind the Bayesian approach to the interpretation of data is encapsulated in Bayes's Theorem (Bayes 1763; Fig. 1). In archaeological terms this simply means that we analyse the new data we have collected about a problem ('the standardized likelihood') in the context of our existing experience and knowledge about that problem (our 'prior beliefs'). This enables us to arrive at a new understanding of the problem which incorporates both our existing and our new data (our 'posterior belief'). This is not the end of the matter, however, since today's posterior belief becomes tomorrow's prior belief, informing the collection of new data and its interpretation as the cycle repeats (Fig. 2). In effect Bayes's Theorem allows us to implement the hermeneutic spiral (see Hodder 1992, fig. 22) in an explicit, quantitative, manner.

This approach is by no means confined to the interpretation of chronology, although in archaeology it has most frequently been applied in this area of study. Lindley (1985) provides a user-friendly introduction to the principles of Bayesian statistics, and Buck *et al.* (1996) provide an introduction to the approach from an archaeological viewpoint. For those interested in

the application of the Bayesian approach to chronological problems, the most accessible introduction is provided by the series of worked examples by Buck *et al.* (1991; 1992; 1994a–c), which introduced the technique to archaeologists. Archaeological problems tackled by integrating radiocarbon evidence with other forms of chronological information are becoming common in the literature (see for example Bayliss *et al.* 1997; Needham *et al.* 1998; Rom *et al.* 1999; Lu *et al.* 2001; Bruins *et al.* 2003; Higham *et al.* 2004; Kim *et al.* 2004; Friedrich *et al.* 2006; Manning *et al.* 2006).

### An introduction to constructing Bayesian chronologies

The use of the Bayesian approach for the interpretation of archaeological chronologies is based on the premise that, although the calibrated dates of radiocarbon measurements accurately estimate the calendar ages of the samples themselves, it is the dates of the archaeological events associated with those samples that are of importance. So, for example, we may obtain dates on nine antler picks from the base of the ditch at Stonehenge (Cleal *et al.* 1995, table 64). We are not, however, particularly interested in the time when these antlers happened to be shed, but rather in the time when the ditch was dug and they were placed on its base. Dating the antlers is a means to an end, not the end in itself.

Bayesian techniques can provide quantitative estimates of the dates of such events (the posterior beliefs) by combining scientific dating evidence, such as radiocarbon dates (the standardized likelihood), with relative dating evidence such as the stratigraphic relationships between contexts containing dated samples (the prior beliefs). The posterior beliefs are expressed as probability distributions known as ‘posterior density estimates’. These, by convention, are always given in *italics*, and are not absolute. They are interpretative estimates, which will change as additional data become available or as the existing data are modelled from different perspectives. This production of alternative models, none of them definitive, is simply a means of creating multiple pasts and is entirely congruent with the post-modern perspective adopted in this series of papers (Shanks & Tilley 1987).

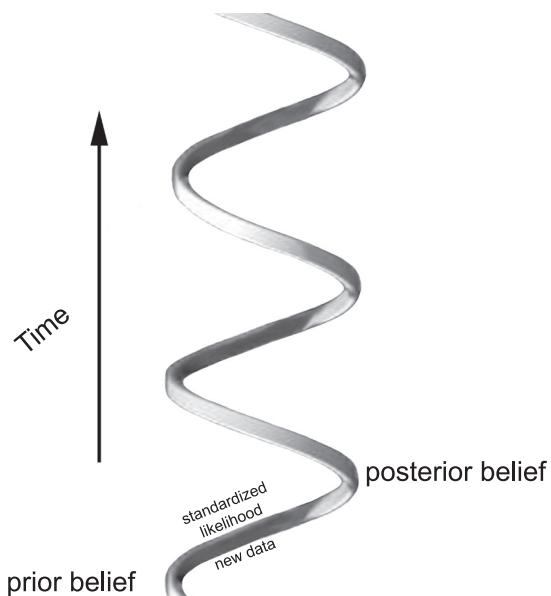
The technique used here is a form of Markov Chain Monte Carlo (MCMC) sampling, which has been applied using the program OxCal v3.10 (Bronk Ramsey 1995; 1998; 2001). The model is defined in OxCal, detailing the radiocarbon results and explicitly specifying the known relative ages of the radiocarbon samples. If the dated material died at the same time as

$$P(\text{parameters}) \times P(\text{data}|\text{parameters}) = P(\text{parameters}|\text{data})$$

$$P(\text{data})$$

Prior beliefs x Standardized likelihoods = Posterior beliefs

**Figure 1.** Bayes's theorem.



**Figure 2.** Bayes and the hermeneutic spiral.

the context from which it was recovered was deposited, these relationships may be summarized in the site's Harris matrix. These relative age relationships between radiocarbon samples are based on strong archaeological evidence and should affect the output of the model strongly. They are known as ‘informative prior beliefs’.

A second form of prior belief, however, is also incorporated into the model. This is an assumption about the mathematical distribution of the archaeological events in the phase of activity which has been sampled for radiocarbon dating. It is necessary to implement such a distribution to counteract the statistical scatter on the radiocarbon measurements. Scatter occurs because radiocarbon dates come with errors, and so within any group of dates relating to an archaeological phase, a proportion of the probability distributions of the calibrated radiocarbon dates will lie outside – earlier or later than – the actual calendar span of that phase. If this scatter is not taken into account, then it will appear that the archaeological activity started earlier, finished later, and continued for longer than was actually the case (Steier & Rom 2000; Bronk Ramsey 2000; and see below). The approach adopted by OxCal v3.10 is to assume that the

archaeological events which have been sampled for radiocarbon dating are distributed uniformly (Buck *et al.* 1992). It is thought that this is an ‘uninformative prior belief’, where the output from the model is relatively robust against a violation of this assumption (see below for investigation of this point).

Once the model has been explicitly defined in OxCal, the program calculates the probability distributions of the individual calibrated radiocarbon results (Stuiver & Reimer 1993), and then attempts to reconcile these distributions with the prior information by repeatedly sampling each distribution to build up the set of solutions consistent with the model structure. This is done using a random sampling technique (MCMC) which generates a representative set of possible combinations of dates, in the case of OxCal specifically utilizing a mixture of the Metropolis-Hastings algorithm and the more specific Gibbs sampler (Gelfand & Smith 1990; Gilks *et al.* 1996). This process produces a posterior density estimate of each sample’s calendar age, which occupies only part of the calibrated probability distribution. In this series of papers these posterior density estimates are shown in solid black in the figures with the distribution name in italics; the calibrated radiocarbon date from which it has been sampled is shown in outline. It should be noted, however, that the model is also able to produce estimates for the dates of events that have not been dated directly by single radiocarbon determinations, such as the beginning and end of a phase of activity or an archaeological event which must have happened after the date of one sample and before the date of another. By comparing these probability distributions it is also possible to provide quantitative estimates of the durations of phases of activity or hiatuses between such phases. For example, in our paper on the West Kennet long barrow, the start of primary burial in the chambers (*start\_primary*; Bayliss, Whittle & Wysocki this issue, p. 94, fig. 6) is calculated from the radiocarbon measurements on all the dated individuals in the primary mortuary deposits, as is the estimate for the end of this phase of primary deposition (*end\_primary*; Bayliss, Whittle & Wysocki this issue, p. 94, fig. 6). These posterior beliefs do not depend on any particular radiocarbon date, but rather on the entire assemblage of dates from the phase. To go further, comparison of these two posterior density estimates allows a formal estimate to be made of the elapsed time between the start and end of the primary deposits. This estimate, measured in numbers of years, is shown in Bayliss, Whittle & Wysocki this issue, p. 94, fig. 7 (*primary use*).

Two statistics are calculated by OxCal which aid the archaeologist in an assessment of the reliability

and stability of a model. The first of these is known as the index of agreement ( $A$ : Bronk Ramsey 1995, 429), and relates to individual standardized likelihoods, i.e. usually calibrated radiocarbon dates. This index provides a measure of how well any posterior distribution agrees with the prior distribution; if the posterior distribution is situated in a high-probability region of the prior distribution, the index of agreement is high, if it falls in a low-probability region, it is low. If the index of agreement falls below 60% (a threshold value analogous to the 0.05 significance level in a  $\chi^2$  test), then the radiocarbon result is regarded as in some way problematic. Sometimes this merely indicates that the radiocarbon result is a statistical outlier, although a very low index of agreement may suggest that a sample is residual or intrusive (i.e. that the calendar age of the sample is different to that implied by the stratigraphic position of the context from which it was recovered). This statistic provides a useful indication of samples whose archaeological taphonomy may be unexpected, or where contamination may be suspected. An overall index of agreement is then calculated from the individual agreement indices ( $A_{\text{overall}}$ : Bronk Ramsey 1995, 429), providing a measure of the consistency between the prior information and the radiocarbon results.<sup>3</sup> Again, the overall index of agreement has a threshold value of 60%, and models which produce values lower than this should be subject to critical re-examination.

The second statistic is known as convergence. This is a measure of how quickly the MCMC sampler is able to produce a representative and stable solution to the model. Details of the measure used in OxCal ( $C$ ) may be found in Bronk Ramsey (1995, 429). In practice a model whose convergence is poor (less than 95%) is unstable and the results should not be used. In such a case the sampling procedure can be examined in more detail to determine whether particular parts of the model are unstable (see, for example, Bayliss, Whittle & Wysocki this issue, p. 98, fig. 11).

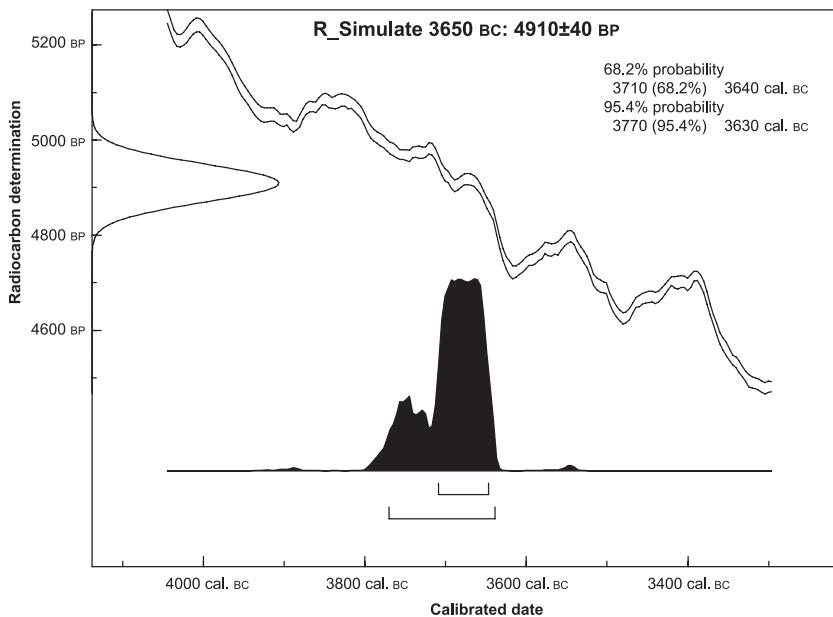
#### **‘All models are wrong, some models are useful’ (Box 1979, 202)**

Models attempt to provide a simplified view of data which is archaeologically meaningful on the basis of a limited sample of information (Voorrips 1987, 68). They are therefore, almost by definition, wrong, as it would be very remarkable if any complex society from the past – or indeed system existing now in the real world – could be exactly represented by a simple model. Having admitted that ‘all models are wrong’, we need to consider which ‘are useful’. This means we

need to assess what George Box (1976, 792) has called ‘worry selectivity’. All models are wrong, but we need to decide what about them is *importantly* wrong. If we vary each element of a chronological model, one by one, which are critical to the date estimates produced? Some elements in a model can be erroneous, but if they do not affect the date estimates, the model is not importantly wrong. Other elements, however, may significantly affect the model outcomes, and so if they are erroneous, the model is importantly wrong.

There are a number of ways in which this can be gauged. We have met two statistical criteria that may aid in our selection of the most appropriate model in the previous section (convergence and indices of agreement). Some models may be preferred to others, judged more plausible, simply on the basis of archaeological evidence. For example, in considering the chronology of Fussell’s Lodge long barrow (Wysocki *et al.* this issue), we have selected the authors’ preferred model on the basis of the proportion of hand and foot bones recovered from the mortuary deposits. This purely archaeological criterion is used to distinguish between a range of statistically plausible models. We may also compare models which incorporate different archaeological interpretations of the same sequence, to determine where they provide similar estimates and where not. For example, again in considering the chronology of Fussell’s Lodge long barrow, alternative readings provide very similar estimates for the date when the barrow was raised (*build\_barrow*; Wysocki *et al.* this issue, p. 79, fig. 12), but very different estimates for when the mortuary structure which it contained and covered was first constructed (*build\_box*; Wysocki *et al.* this issue, p. 79, fig. 12). On this basis we may choose to place more faith in the reliability of the former estimate. As these comparisons measure the sensitivity of our estimates to different models, they are known as ‘sensitivity analyses’.

A final approach for assessing the validity of different models is to use simulation to determine whether different approaches produce the correct answer when presented with known-age data. This method is used in this section to examine three aspects of chronological modelling which occur frequently in the literature.

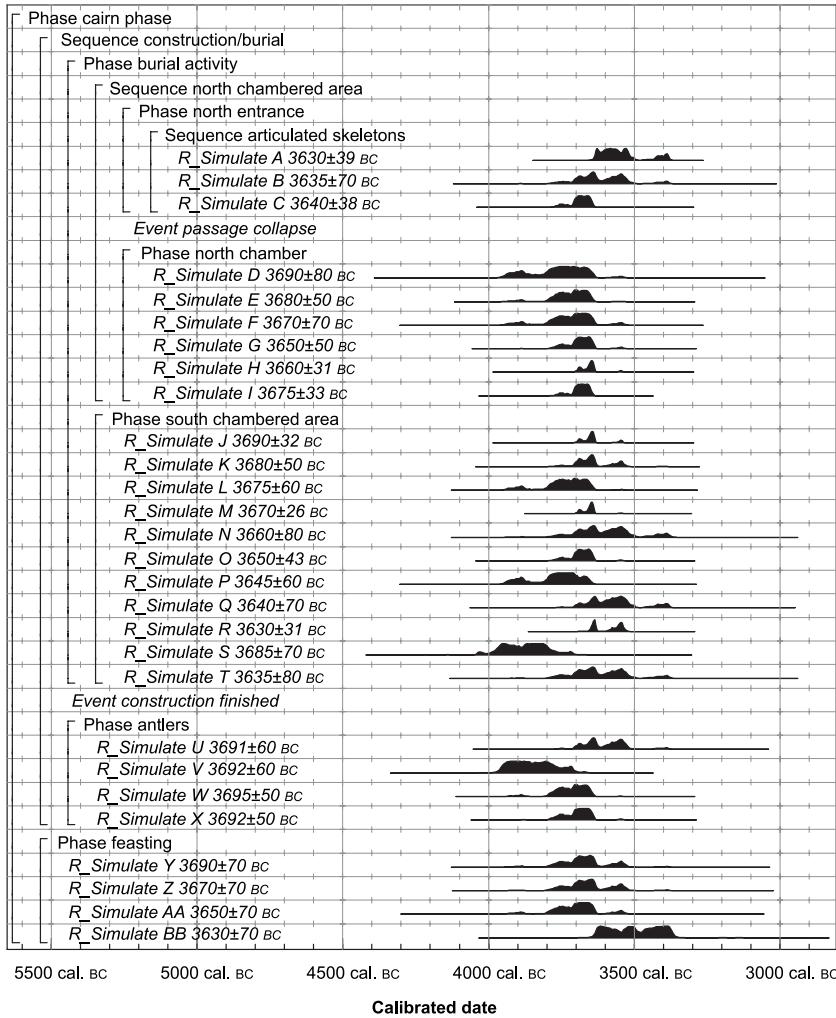


**Figure 3.** A simulated radiocarbon measurement for a sample whose real calendar age is 3650 BC.

This is done using a model for a fictitious long barrow (in fact, this is based on a simplified version of the chronological model for Hazleton North long cairn presented in Meadows *et al.* (this issue, p. 54, fig. 8)). Here the radiocarbon age determinations obtained from Hazleton have been replaced by determinations, with the same error terms, simulated by OxCal from samples of known calendrical date.

Figure 3 shows an example of the simulation process. The archaeologist estimates a date on the calendrical timescale for the sample in question, in this case 3650 BC (dates of this kind are cited as ‘BC’). A radiocarbon age which would be expected from a sample whose real calendar date was 3650 BC is then produced, in this case  $4910 \pm 40$  BP (following international convention, these ages are cited as ‘BP’). This age is then calibrated, producing the probability distribution for the calibrated date shown in Figure 3 and a range of 3770–3630 cal. BC (at 95% confidence) (following international convention, these dates are cited as ‘cal. BC’).<sup>4</sup>

In this simulation, the 28 age determinations in the Hazleton model have been replaced by simulated ages as shown in Table 1. The actual dates of the antlers used in the construction of the cairn range from 3695 BC to 3691 BC, the construction of the cairn was completed in 3690 BC, the passage to the north chamber collapsed between 3645 BC and 3640 BC, and the last interment occurred in 3630 BC. Burial occurred continuously for 60 years, between 3690 BC and 3630 BC and all the dated events cover only 65 years.



**Figure 4.** Calibrated probability distributions of the simulated radiocarbon ages from a long barrow (the actual dates of these samples range from 3695 BC to 3630 BC).

#### Why model at all? The dangers of visual inspection

The probability distributions of the calibrated simulated radiocarbon ages listed in Table 1 are shown in Figure 4. Visual inspection of graphs of this kind is probably the most popular method for interpreting radiocarbon dates. The simplest approach is to inspect the distributions, taking their widest limits, and simply to say that activity on this site happened sometime between c. 4000 cal. BC and c. 3400 cal. BC. A slightly more sophisticated variant of this approach recognizes that estimating radiocarbon ages is a probabilistic process and attempts to account for outliers, either by excluding low parts of probability distributions from the edges of the graph or by determining a time when most of the probability of most of the dates is concentrated. With these data, such an approach might

**Table 1.** Simulated radiocarbon ages and actual calendar dates for samples from the fictitious long barrow discussed in the simulations shown in Figures 4, 8, 13, 14, 19 and 20.

Sample identifier	Calendar age	Simulated radiocarbon age	Error
<i>North entrance</i>			
A	3630 BC	4759	39
B	3635 BC	4840	70
C	3640 BC	4904	38
<i>North chamber</i>			
D	3690 BC	4952	80
E	3680 BC	4933	50
F	3670 BC	4919	70
G	3650 BC	4897	50
H	3660 BC	4862	31
I	3675 BC	4910	33
<i>South chambered area</i>			
J	3690 BC	4856	32
K	3680 BC	4859	50
L	3675 BC	4941	60
M	3670 BC	4863	26
N	3660 BC	4834	80
O	3650 BC	4898	43
P	3645 BC	4969	60
Q	3640 BC	4811	70
R	3630 BC	4823	31
S	3685 BC	5093	70
T	3635 BC	4846	80
<i>Antlers</i>			
U	3691 BC	4832	60
V	3692 BC	5049	60
W	3695 BC	4931	50
X	3692 BC	4907	50
<i>Feasting</i>			
Y	3690 BC	4881	70
Z	3670 BC	4871	70
AA	3650 BC	4904	70
BB	3630 BC	4714	70

suggest that activity on this site happened sometime between c. 3800 cal. BC and c. 3500 cal. BC (Fig. 5).

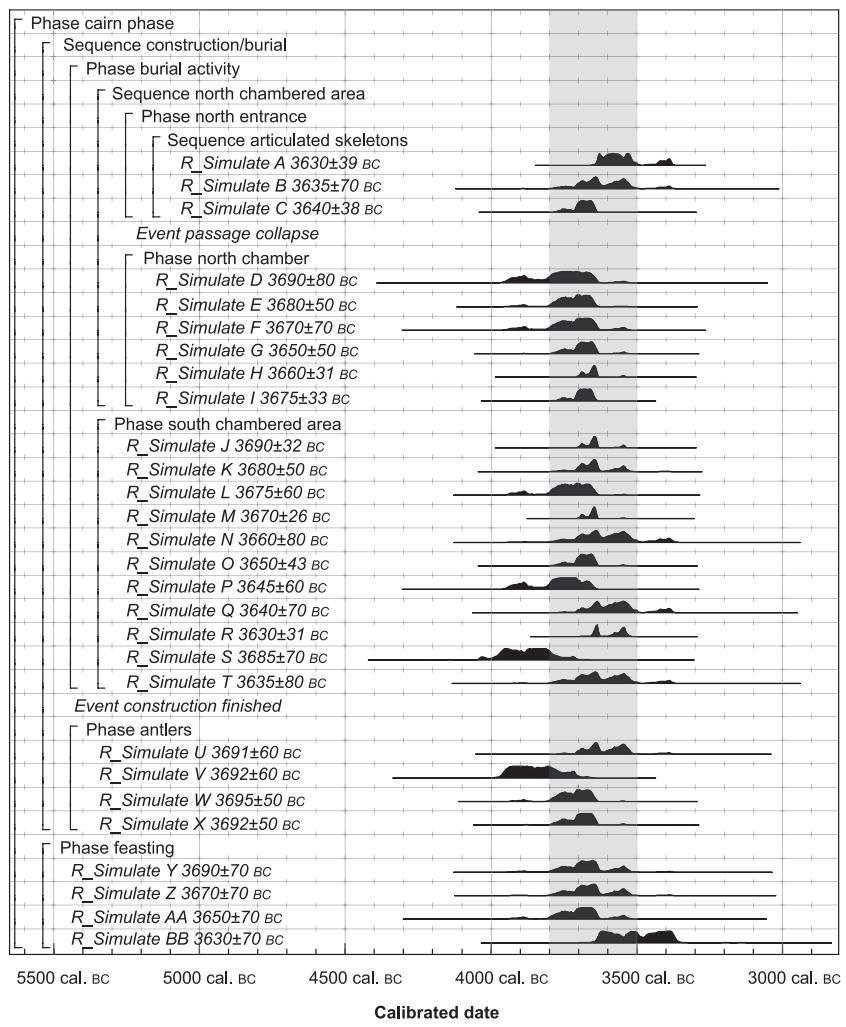
As long as such graphs are interpreted as ‘something happened for some period of time (of unknown duration) at some point between this date and that date’, then we would not be importantly misled. Fundamentally there is nothing importantly wrong about these interpretations. In these data, our site was in use between 3695 BC and 3630 BC — well within the limits of both estimates. The approach is informal and probably only broadly reproducible between workers, but we are only attempting to produce a broad chronology.

Far too frequently, however, archaeologists attempt to estimate when activity on a site started and ended (and thus by implication determine for how long it was in use) from such graphs. If this approach is interpreted as ‘my site was occupied between this date and that date’, then it is importantly wrong. In our simulated data our site starts in 3695 BC — more than 300 years later than ‘c. 4000 cal. BC’ and more than a century later than ‘c. 3800 cal. BC’. It ends in 3630 cal. BC — more than 200 years earlier than ‘c. 3400 cal. BC’ and more than a century earlier than ‘c. 3500 cal. BC’. Not only does it appear, incorrectly, that our site started earlier and carried on later than it did in reality, but this interpretation gives a thoroughly misleading impression of the duration of activity on the site. It seems to have continued for 300 years or more, when in fact the dates only span a period of 65 years.<sup>5</sup>

The problem is that this approach has no mechanism to allow for the statistical scatter on the radiocarbon dates, and so it looks as if archaeological activity started earlier, continued for longer, and ended later than was actually the case. In effect, the scatter and uncertainties on the radiocarbon dates are being confused with longevity of ancient activity. This is a very real problem, as much of our current understanding of the chronology of British prehistory derives from the informal interpretation of graphs or tables of simple calibrated radiocarbon dates.

#### *The dangers of summing calibrated dates*

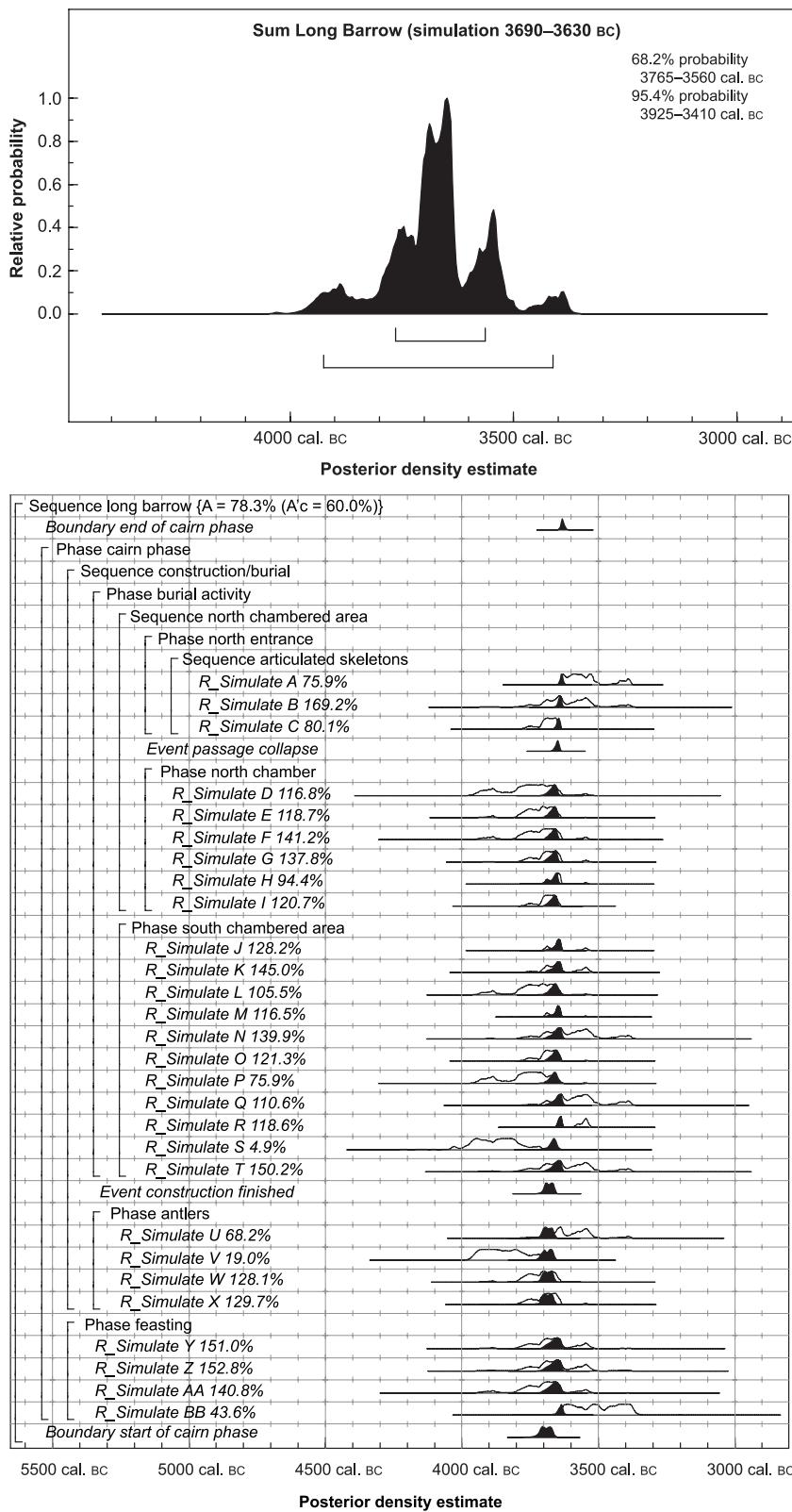
The limitations of simple inspection of radiocarbon dates to determine the duration of an archaeological phenomenon have long been appreciated, and led to the concept of a *floruit* (initially for uncalibrated radiocarbon ages: Ottaway 1973; Geyh & de Maret 1982; and then for calibrated radiocarbon dates: Aitchison *et al.* 1991). The method involves adding together all the probability distributions of the calibrated dates; the probability distribution then is normalized to an integral of 1.0; and then the distribution is integrated from each end until a certain proportion of the curve



**Figure 5.** *Calibrated probability distributions of the simulated radiocarbon ages from a long barrow, with a visual assessment of the most likely period when the site was in use (the actual dates of these samples range from 3695 BC to 3630 BC).*

has been excluded (15.9% or 2.3% from each end as appropriate). The range defined is then the part of the distribution between these two points. Details of this approach are provided in Aitchison *et al.* (1991).

The sum of the probability distributions of the calibrated simulated radiocarbon ages listed in Table 1 is shown in Figure 6, along with the *floruits* at 68% and 95% confidence. As a method for determining the duration of an archaeological phase, it is apparent that this technique is importantly wrong. As these radiocarbon age estimates derive from simulations of known date, we know that all the dated samples span 65 years (from 3695 BC to 3630 BC). This approach estimates that the period in which 95% of these events took place was 205 years (at 68% confidence) or 515 years (at 95% confidence), and so misleads us to be-



**Figure 6.** Sum of the calibrated probability distributions of the simulated radiocarbon ages from a long barrow, and floruit ranges for the site (Aitchison *et al.* 1991) (the actual dates of these samples range from 3695 BC to 3630 BC).

lieve that the archaeological activity continued for much longer than it did in reality.

Although this approach does make some attempt to allow for the presence of a few outliers in the data (Aitchison *et al.* 1991, 116), again the problem is that the method does not allow sufficiently for the statistical scatter on the radiocarbon dates. The more dates are included in the analysis, the more scattered will be the results and the longer the estimates for the duration of the activity. This is exacerbated if the radiocarbon measurements have large error terms, as recognized by some users who exclude such measurements from the analysis (e.g. Czebreszuk & Szmyt 2001). This method is widely used. For example, over half the papers in Czebreszuk & Müller (2001) utilize the approach, and it is particularly popular for estimating the date of duration of different cultures in European prehistory. The trouble is that it does not give the correct answer.

Mathematically what this approach does is to provide a frequency distribution modulated by the un-

**Figure 7.** Probability distributions of simulated dates from a long barrow. Each distribution represents the relative probability that an event occurs at a particular time. For each of the dates two distributions have been plotted: one in outline, which is the result of simple radiocarbon calibration, and a solid one, based on the chronological model used; the 'event' associated with, for example, 'V', is the growth of the dated antler. Distributions other than those

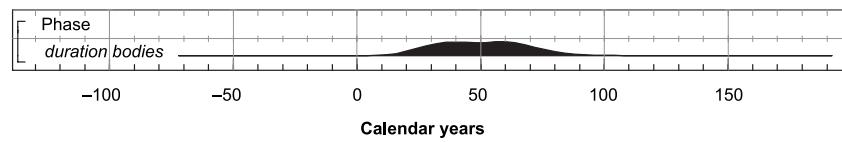
relating to particular samples correspond to aspects of the model. For example, the distribution 'construction finished' is the estimated date when the barrow was built. The large square brackets down the left-hand side along with the OxCal keywords define the overall model exactly. The actual dates of these samples range from 3695 BC to 3630 BC.

certainty on the calibrated date of the sample. This means that the technique attempts to provide a view of the spread of the actual calendar dates of the dated material in a phase, although, as this view is folded together with uncertainty caused by the statistical spread of the radiocarbon dates, we are looking at it through blurred spectacles. If this method is used in isolation, all of the uncertainties are implicitly assumed to be independent — an assumption which is manifestly untrue for a coherent archaeological phase or culture. In fact this is the same problem alluded to earlier (see above p. 5; Steier & Rom 2000; Bronk Ramsey 2000). Summing probability distributions can be used to provide an estimate for the frequency distribution of dated events in a phase, but only in combination with a full Bayesian analysis, which takes into account the interdependencies between events, and in situations where the dating resolution is sufficient to discriminate between the distribution of actual dated events and the distribution imposed by the model. In archaeological interpretation, however, the distribution of the dated events is rarely of interest. What we wish to know is when the activity started, when it finished, and for how long it continued. Summing probability distributions, even within a Bayesian framework, does not answer these questions (*contra* Michczyński & Pazdur 2003). It is of more utility as a diagnostic tool for testing models.

#### *A Bayesian model*

Finally, a Bayesian model combining the calibrated simulated radiocarbon ages listed in Table 1 with the appropriate prior information summarized in Meadows *et al.* (this issue, p. 56, fig. 11) is shown in Figures 7 and 8. The posterior density estimates for key parameters are given in Table 2. It is apparent that the formal modelling of the problem has produced quantified estimates of the actual archaeological dates of interest, and that these estimates are accurate (the known dates/duration of each of the five key parameters lie within both the 68% probability and 95% probability estimated date ranges).

The effect of the statistical scatter of the simulated radiocarbon measurements which are included in this model is shown in Figures 9 and 10 (for the same five key parameters). The model was run ten times, in each case with a different suite of radio-



**Figure 8.** Probability distributions showing the number of calendar years during which bodies were interred in a long barrow according to the simulation shown in Figure 7 (these burials actually span 60 years).

carbon ages simulated by OxCal from the calendar dates of the samples defined in Table 1. These graphs illustrate the variation in our estimates for the dates of these key parameters, simply based on the statistical scatter of radiocarbon age estimates. In all cases, the resulting posterior density estimate includes the actual date of the event in question.

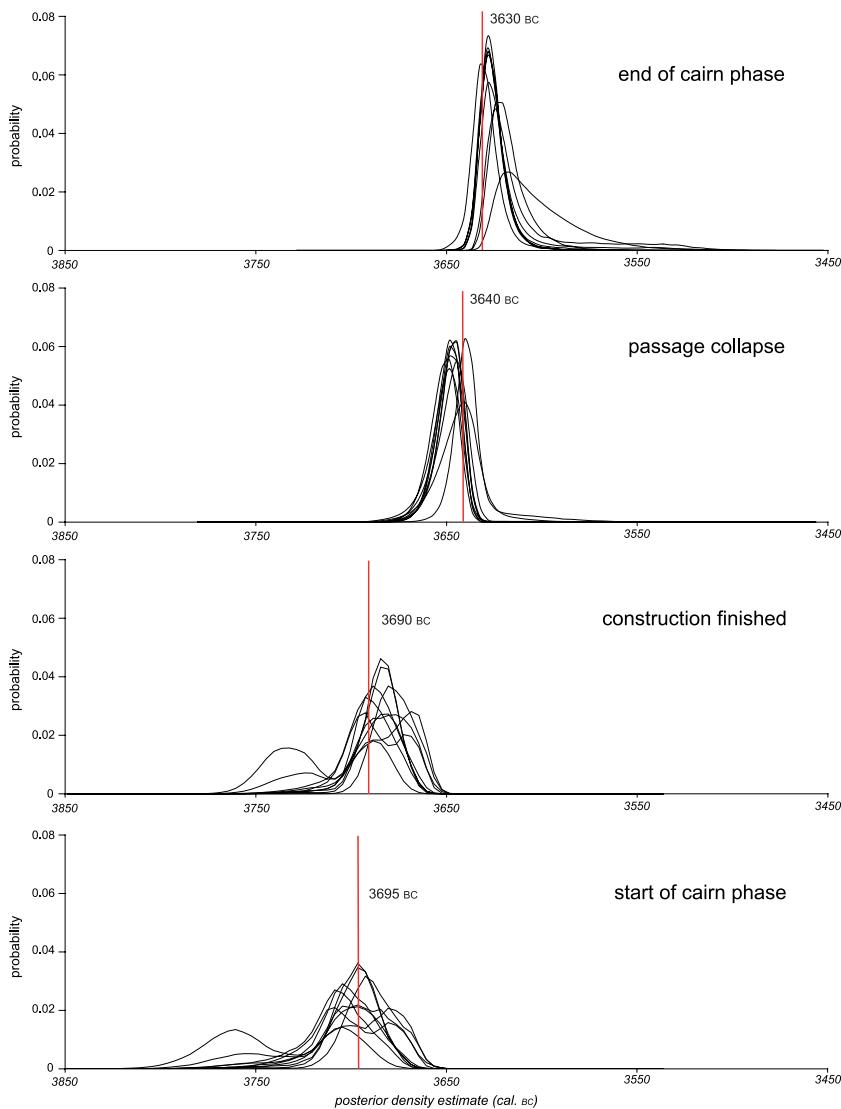
This analysis demonstrates that if the informative and uninformative prior information included in a model and the radiocarbon ages are both accurate, then the Bayesian approach can produce the correct answer (see also Hamilton *et al.* in prep). But have we simply been lucky with the actual dates selected in this case? Is it possible to produce precise and accurate calendar date ranges when the shape of the calibration curve is unhelpful (*contra* Ashmore 2004, 130)?<sup>6</sup>

#### Some sensitivity analyses

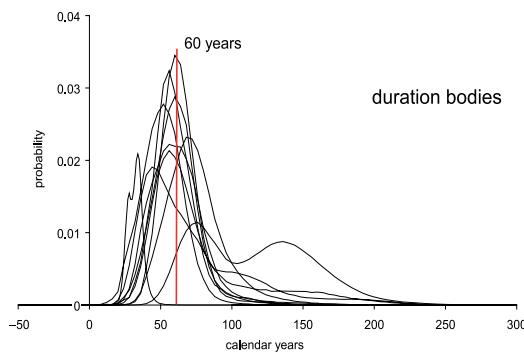
To investigate this issue, the model was re-run with sets of simulated dates running from 3900–3840 cal. BC every 10 years until 3400–3340 cal. BC. The five key parameters from the model shown in Table 2 were compared against the actual dates simulated for each parameter in the respective model. Overall, 51 models were calculated, in which only two of the true values lay outside the 95% probability range of the relevant posterior density estimate (2/255 = <1%). Thirty-nine of the true values lay outside the relevant 68% probability range (39/255 = 15%). These figures suggest that, if the prior information is correct, then a Bayesian model should give an accurate age estimate no matter where on the calibration curve a particular site falls. On the

**Table 2.** Posterior density estimates for key parameters from the chronological model shown in Figures 7 and 8.

Parameter	Calendar date/ duration	Posterior density estimate (95% probability)	Posterior density estimate (68% probability)
<i>start of cairn phase</i>	3695 BC	3735–3655 cal. BC	3715–3665 cal. BC
<i>construction finished</i>	3690 BC	3710–3650 cal. BC	3700–3660 cal. BC
<i>passage collapse</i>	3645–3640 BC	3670–3635 cal. BC	3660–3640 cal. BC
<i>end of cairn phase</i>	3630 BC	3645–3610 cal. BC	3640–3620 cal. BC
<i>duration bodies</i>	60 years	15–90 years	30–70 years



**Figure 9.** Graphs showing the variation in the posterior density estimates for the key chronological parameters in ten runs of the model described in Figure 7 (different measurements are simulated in each run of the model).



**Figure 10.** Graph showing the variation in the posterior density estimate for the duration of burial in ten runs of the model described in Figure 7 (different measurements are simulated in each run of the model).

other hand, it appears that the true ages fall outside the posterior density estimates less frequently than would be expected. This is in part because the overall index of agreement is poor when a simulated radiocarbon age is an extreme outlier. These runs of the simulation were discarded because, in a real-world application, such a model would be rejected and rerun with any measurements identified as misfits removed.<sup>7</sup> It may also be because the distribution of dated events (one every ten years) is more even than would be expected in a real random sample from a uniformly distributed group of events. In fact these results should not surprise us too much as the Bayesian approach is to find the most probable solution rather than solutions that will be correct a certain proportion of the time. It is reassuring, however, that the answers in these simulations are right more often than one might expect them to be.

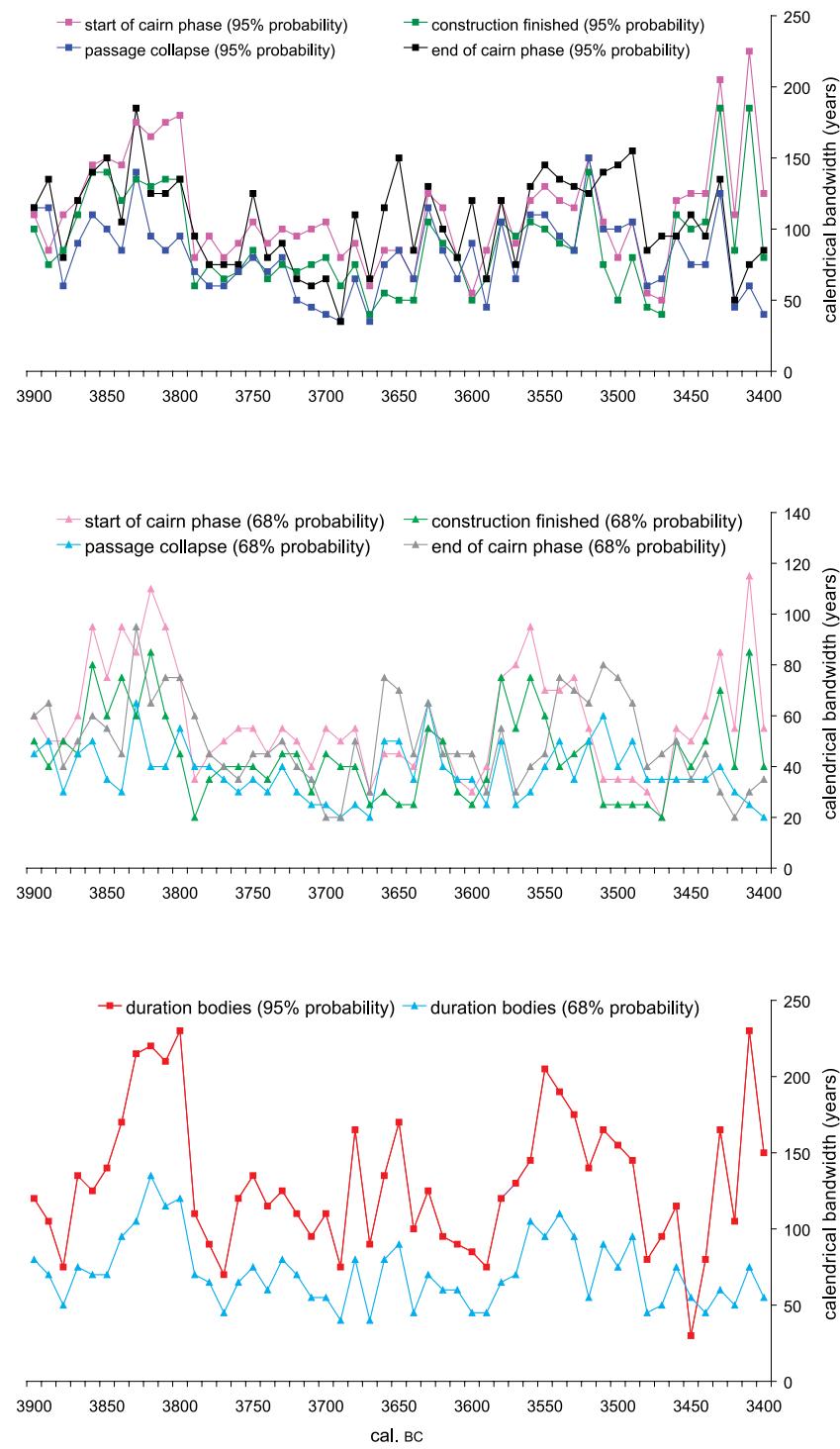
Figure 11 shows the calendrical bandwidths<sup>8</sup> of the posterior density estimates of the key parameters provided by the simulations between 3900–3840 cal. BC and 3400–3340 cal. BC (bandwidths are plotted again the earlier date in this range). It can be seen that, for the 28 measurements incorporated in this model, all estimates are more precise in the thirty-eighth and thirty-seventh centuries cal. BC (generally averaging 100 years or less at 95% probability and 50 years or so at 68% probability), and less precise in part of the thirty-ninth, thirty-fifth, and thirty-fourth centuries (generally averaging more than 100 years at 95% probability and more than 50 years at 68% probability). This does not mean, however, that more precision is not obtainable in these centuries — just that more measurements or prior information are required to obtain more precision. Of course, there may well be a limited number of suitable samples available with a limited number of stratigraphic relationships between them. In practice this may limit the precision achievable for a particular problem. For example, in the burial chamber of Wayland's Smithy I, the remains of only 14 human individuals were recovered (see Whittle,

Bayliss & Wysocki this issue). Once these 14 samples have been measured to the limits of the accuracy and precision currently available, and the stratigraphic relationships between the burials incorporated in the model, in practice no further precision is obtainable for this application from this approach.

### Further sensitivity analyses: the prior beliefs

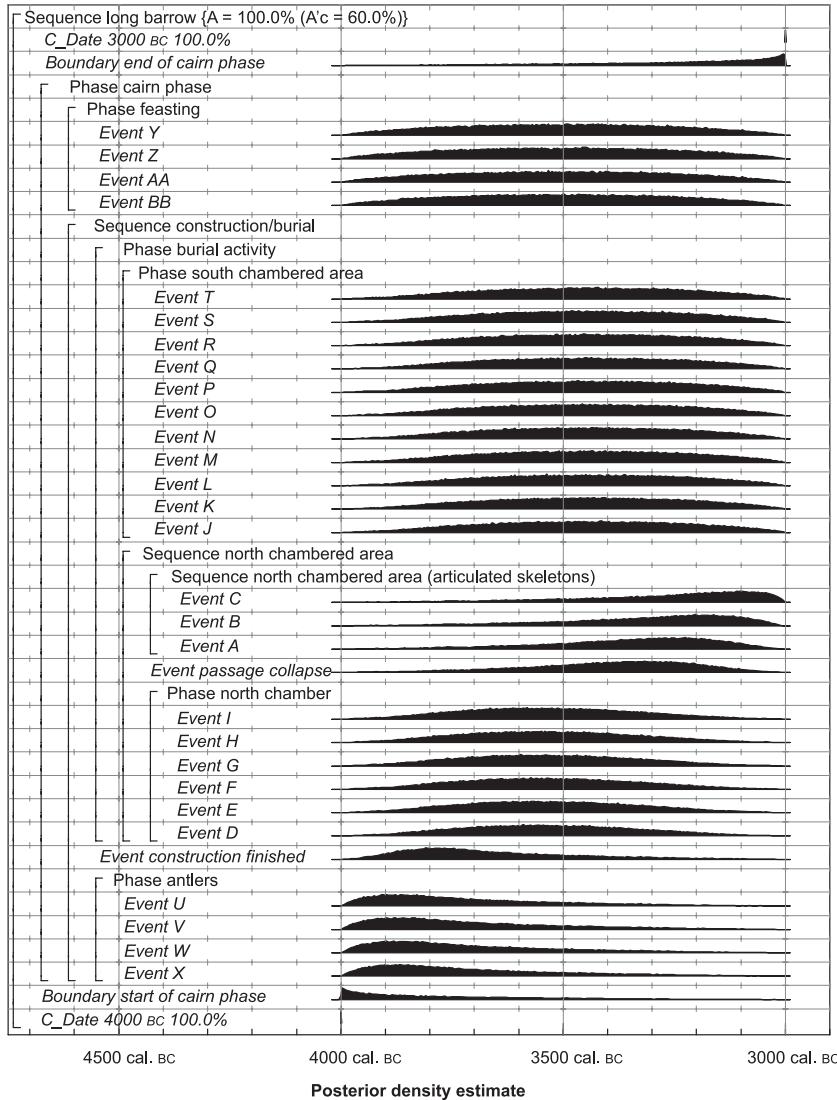
The series of simulations discussed in the previous section has demonstrated that, if the inputs into a Bayesian model are correct, then the analysis should produce accurate estimates for the chronology of the site. But what if those inputs are wrong? How wrong do they have to be for the outputs of a model to be invalid? How can we worry selectively and judge how importantly wrong a particular model may be? This can be investigated by varying the inputs into the simulation model defined in Figure 7 (and Table 1) one by one, and by comparing the results with those of the simulation where all the inputs are known to be correct (i.e. Fig. 7 and Table 2).

The first component of our model to be considered is the prior information (Fig. 1). As explained above, this can be divided into two categories: information which is intended to strongly influence the outputs of the model, informative prior information; and the assumptions of the model against which outputs are intended to be robust, uninformative prior information. For the chronological modelling of archaeological data, informative prior information usually consists of the stratigraphic relationships between dated samples or archaeological events inferred from excavation. This is the case for the fictitious long barrow considered in Figure 7. In other circumstances, other data, such as calendar dates from dendrochronology, coins, or documentary sources, or



**Figure 11.** Graphs showing the variation in the precision of the posterior density estimates for the key parameters in the simulation model (see Table 2) by calendar date.

relative dating derived from typology or seriation may be included in a model.



**Figure 12.** Probability distributions of undated events from a long barrow. Each distribution represents the relative probability that an event occurs at a particular time, given the prior information included in the model (see Meadows *et al.*, *this issue*, p. 56, fig. 11). Events are constrained to have occurred between 4000 cal. BC and 3000 cal. BC for clarity, in reality the distribution of dated events is less strongly constrained by these inputs. The large square brackets down the left-hand side along with the OxCal keywords define the overall model exactly.

#### Informative prior information

The informative prior information included in the model shown in Figure 7 is summarized in Meadows *et al.* (*this issue*, p. 56, fig. 11). The effect of this information on the results of the analysis is illustrated in Figure 12. This shows that events unrelated to the stratigraphic sequence in this barrow (e.g. AA, BB, Y, and Z), before any scientific dating information is included in the model, have very broad potential

distributions. These dates have been constrained to lie between 4000 cal. BC and 3000 cal. BC simply to make this graph readable, but fundamentally any of the potential dates for this barrow may be the actual dates of these samples. In contrast, events which have stratigraphic relationships with many other events in the model are strongly constrained to date to one part of the potential date range of the barrow. So, for example, the antlers used in the construction of the monument (U, V, W, and X) are much more likely to fall in the earlier part of this period. Equally the sequence of articulated human remains in the north chambered area (A, B, and C) are much more likely to fall in the later part of this period.

Figure 12 demonstrates that stratigraphic information included in a Bayesian chronological model strongly affects the outputs of the model. This is as expected, since this information is intended to be informative. This does place a burden of rigour, however, on the archaeologist. Not only must the stratigraphic sequence included in a model be accurate, but this relative chronological sequence must also apply to the dated samples. None of them can be reworked, curated, residual, or intrusive, or the inputs to the model will not be correct. As these inputs are informative, if they are wrong, then the chronology suggested by the model will also be wrong! It is for this reason that the taphonomy and sequence of the dated samples are considered in such detail in the site-specific papers in this series. These basic archaeological interpretations

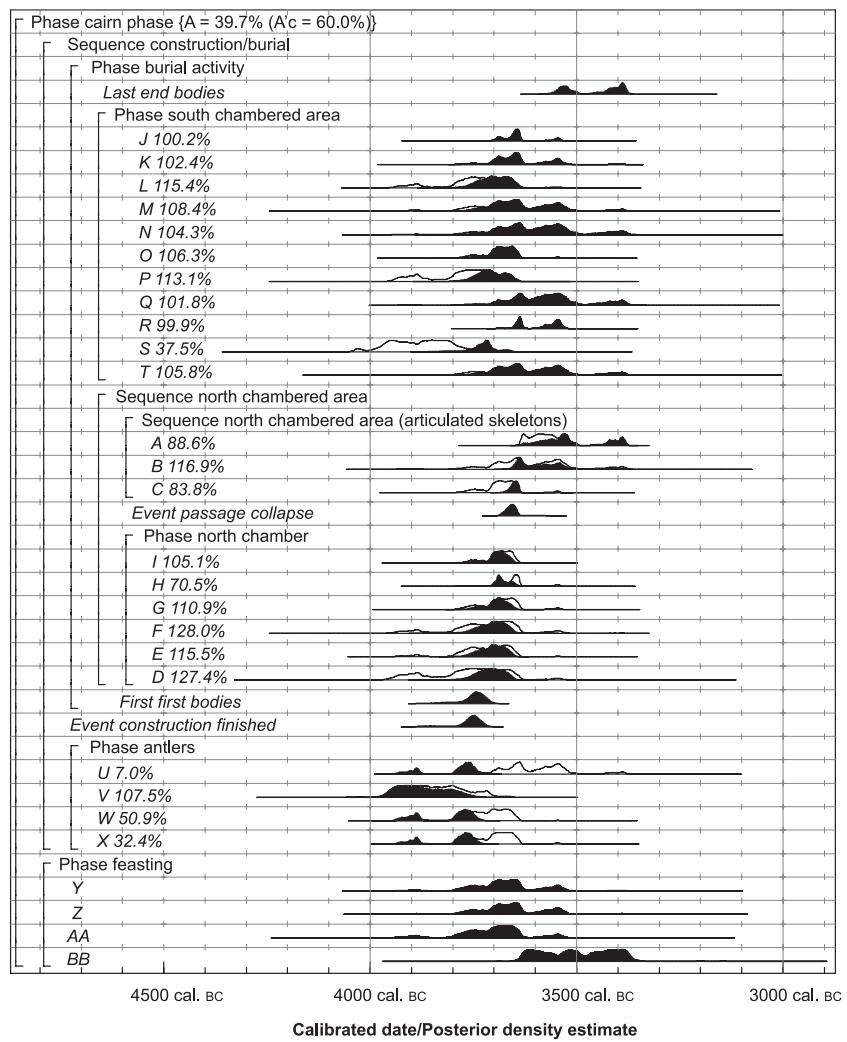
of the excavated evidence are critical to the validity of our models.

#### Uninformative prior information

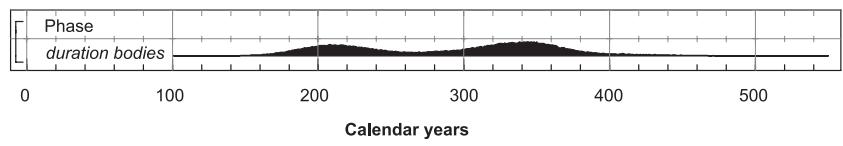
The effect of the uninformative prior information on the results of the simulation model shown in Figure 7 has been examined in two ways. First, the model was re-run including the informative prior information summarized in Meadows *et al.* (*this issue*, p. 56, fig. 11)

but excluding any assumption about the distribution of dated events. In this case, the cairn is estimated to have been built between the date of the antler tools used in its construction and the interment of the first dated individual within the monument, the passage in the north chamber is estimated to have collapsed between the date of the last dated individual in the north chamber and the date of sample C, and the end of the use of the cairn is estimated by the date of the last dated individual interred within it. The duration of burial is provided by the difference between the dates of the first and last dated individuals within the monument. In this analysis the barrow again actually dates to 3695–3640 BC.

The results of this model are shown in Figures 13 and 14. The posterior density estimates for key parameters are given in Table 3. It is clear that this approach is importantly wrong, since this is a simulation and we know the actual answers. The estimate for the date of construction is too early — the correct date does not lie within the range of this estimate at either 95% or 68% probability. The estimate for the time when burial finished is too late; again the correct date does not lie within the range of this estimate at either 95% or 68% probability. In contrast, the estimate for the date when the passage in the northern chambered area collapsed is accurate, with the true value lying within the range of this estimate at both 95% and 68% probability. This may be because this parameter is so strongly constrained by stratigraphic information that any other prior information is proportionately unimportant. The estimate for the duration of burial in the cairn is, however, also inaccurate, as the model suggests that burial continued for two or three times as long as it did in reality. This analysis clearly demonstrates that an assumption about the distribution of dated events in a phase of activity is essential for accurate chronological modelling. Archaeologically we may be

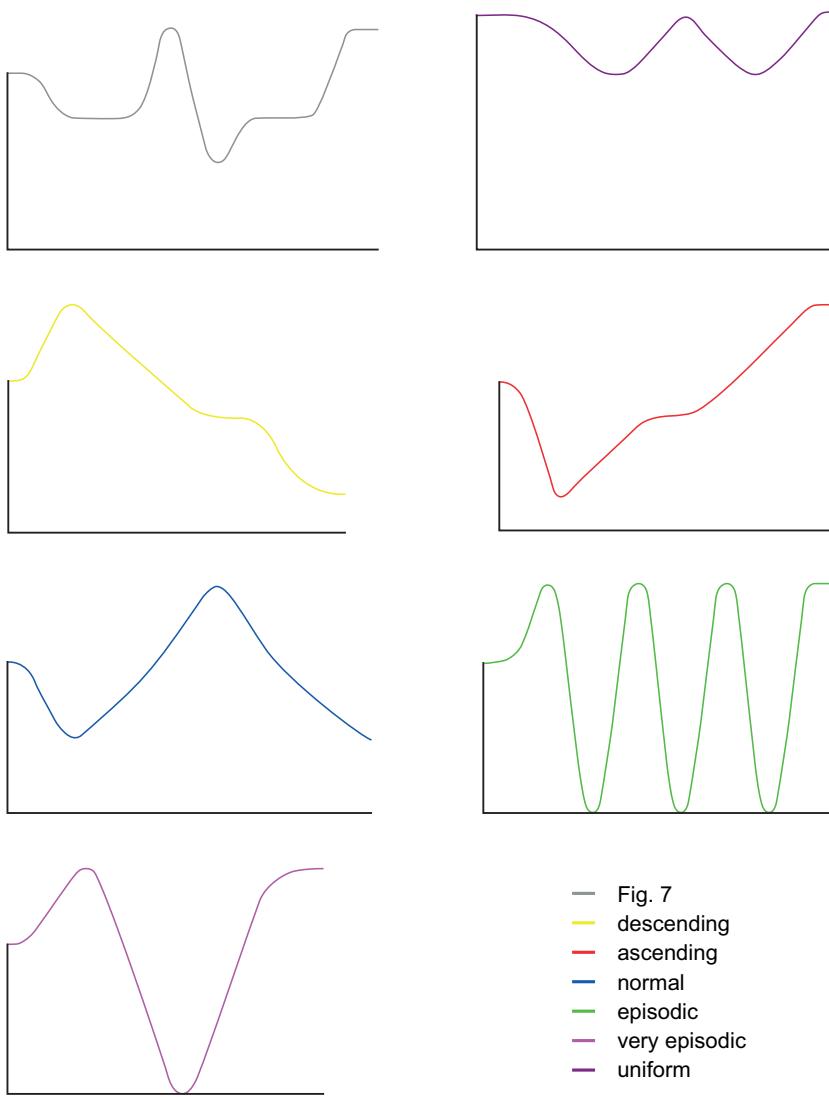


**Figure 13.** Probability distributions of simulated dates from a long barrow. The format is identical to that of Figure 7. In this model no assumption has been made about the distribution of the dated events within the phase of activity during which the barrow was constructed and in use. The large square brackets down the left-hand side along with the OxCal keywords define the overall model exactly. The actual dates of these samples range from 3695 BC to 3630 BC.



**Figure 14.** Probability distributions showing the number of calendar years during which bodies were interred in a long barrow according to the simulation shown in Figure 13 (these burials actually span 60 years).

uncertain of the most realistic form for this distribution to take, but not making an assumption at all is, demonstrably, importantly wrong (see Steier & Rom 2000 for further discussion of this point).



**Figure 15.** A variety of distributions of dated events used within simulations of the chronology of a long barrow (3690–3630 cal. BC).

**Table 3.** Posterior density estimates for key parameters from the chronological model shown in Figures 13 and 14.

Parameter	Calendar date/duration	Posterior density estimate (95% probability)	Posterior density estimate (68% probability)
construction finished	3690 BC	3885–3805 cal. BC (13%) or 3800–3710 cal. BC (82%)	3775–3725 cal. BC
passage collapse	3645–3640 BC	3695–3600 cal. BC	3675–3640 cal. BC
end bodies	3630 BC	3565–3490 cal. BC (40%) or 3465–3370 cal. BC (55%)	3550–3515 cal. BC (15%) or 3425–3375 cal. BC (43%)
duration bodies	60 years	160–420 years	185–240 years (26%) or 300–375 years (42%)

The effect of the uninformative prior information on the results of the simulation model shown in Figure 7 has been examined in a second way. In this case, the aim of the model is to determine the importance of the form of the distribution assumed for the dated events in a phase. The simulated dates in this model have been varied so that the distribution of dated events is no longer uniform (although in all cases the barrow still dates to 3695–3640 cal. BC). Figure 15 illustrates the distributions of dated events included in these models. In each case four samples are available relating to the initial construction of the barrow, but thereafter the use of the chambered areas for burial varies. Either burial is broadly constant and continuous (Fig. 7; Fig. 15, uniform), or it is more frequent when the monument was newly constructed and tails off gradually (Fig. 15, descending), or it starts with infrequent burial which gradually becomes more frequent until burial ceases (Fig. 15, ascending). Burial in the chambers could also rise to a peak and then decline again (Fig. 15, normal), or be episodic. In the simulations considered here burial occurs in one decade, but not the next (Fig. 15, episodic), or occurs in one twenty-year period, but not the next (Fig. 15, very episodic). In the case of chambers which were used for only 60 years, this means that burial did not occur at all in the middle third of the period of use. The point of these analyses is not to suggest that any of these approaches are more archaeological plausible than the default assumption of a relatively constant and continuous phase of burial, but to explore the implications of this assumption being incorrect — to determine whether these models are importantly wrong.

The results of these analyses for the five key parameters of the model shown in Figure 7 are shown in Figures 16 and 17. It can be seen that in almost all cases these models

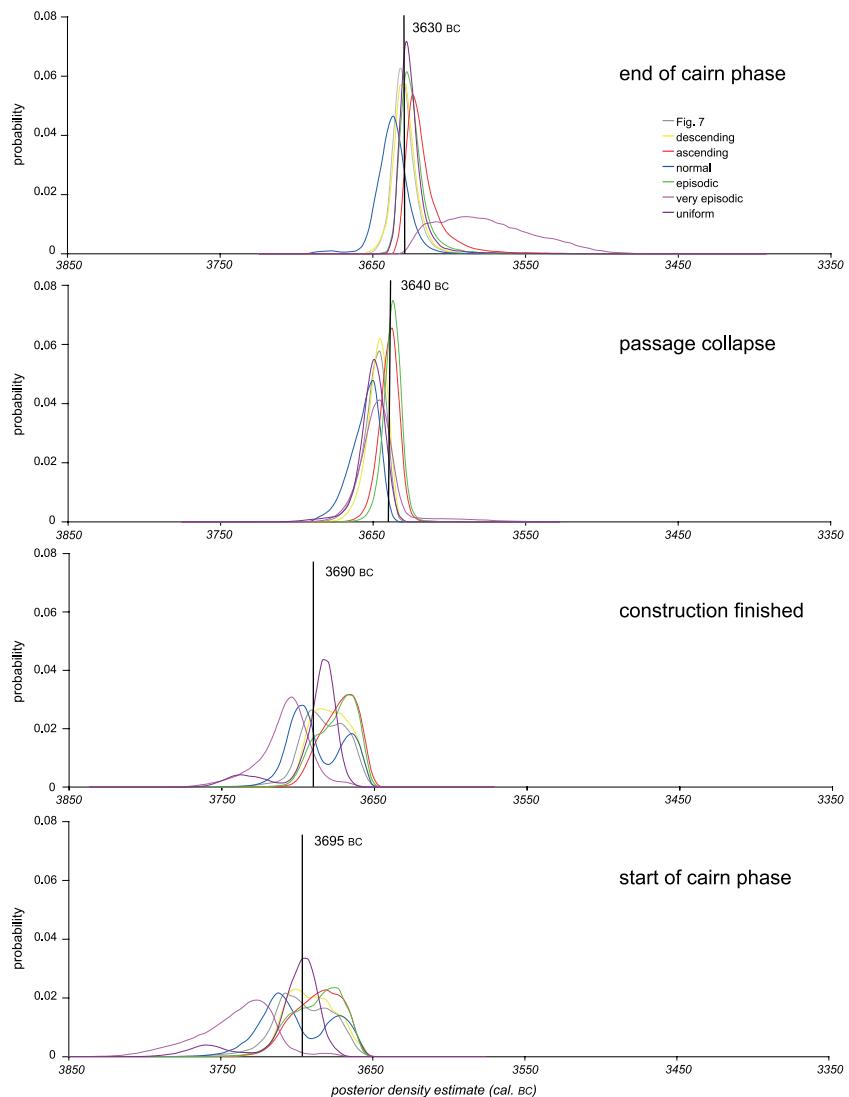
produce posterior density estimates which include the actual dates for the archaeological events concerned. The only exception is the analysis where the distribution of dated events is very episodic, where the analysis suggests that the barrow started earlier, ended later, and was in use for longer than is actually the case. It should be noted that the estimates for the well-constrained events, *construction finished* and *passage collapse*, are accurate even in this model. This seems to be because the estimates for these events are so strongly affected by the informative prior information available from the stratigraphic sequence that, even when the uninformative prior information is grossly incorrect, its effect is out-weighted. Basically the results of the analyses are robust against violations of the assumption that the dated events are uniformly distributed. Only when the actual distribution of dated events is drastically different are the outputs of the model affected substantively and importantly wrong.

#### Further sensitivity analyses: the standardized likelihoods

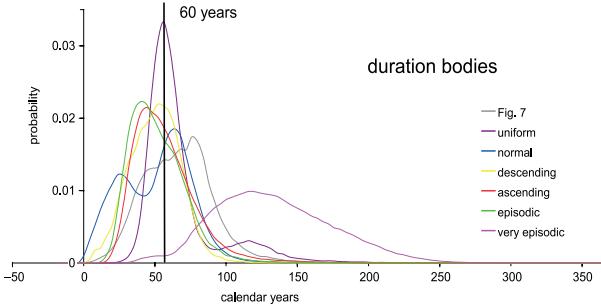
The second component of our model to be considered is the standardized likelihoods (Fig. 1). For the chronological modelling of archaeological data, this usually takes two forms: the age estimates from the archaeological material submitted for scientific dating, and any calibration data required for the conversion of these measurements to the calendar timescale. For the long barrows and long cairns considered in this series of papers, these data are all radiocarbon ages.

#### *The calibration curve*

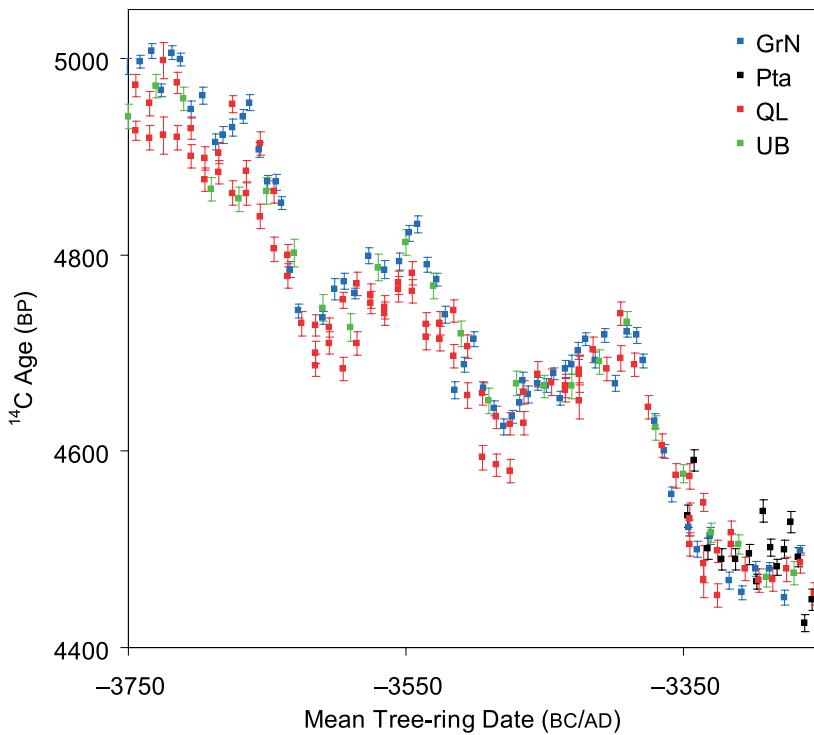
The data included in the relevant section of the current internationally-agreed calibration curve (INTCAL04; Reimer *et al.* 2004) is shown in Figure 18. This curve is based on blocks of wood whose calendar date is known by dendrochronology. It should be noted that, in addition to the decadal data of the University of Washington, Seattle (Stuiver & Becker 1993; corrected



**Figure 16.** Graphs showing the variation in the posterior density estimates for the key chronological parameters in versions of the model described in Figure 7 where the dated events are distributed as shown in Figure 15 (see text).



**Figure 17.** Graph showing the variation in the posterior density estimate for the duration of burial in versions of the model described in Figure 7 where the dated events are distributed as shown in Figure 15 (see text).



**Figure 18.** Radiocarbon measurements included in INTCAL04 (Reimer *et al.* 2004) on wood dated by dendrochronology to between 3750 bc and 3250 bc (GrN = Rijksuniversiteit Groningen; PtA = National Physical Research Laboratory, Pretoria; QL = University of Washington, Seattle; UB = The Queen's University, Belfast).

as described by Stuiver *et al.* 1998a) and the bi-decadal data of the Queen's University, Belfast (Pearson *et al.* 1986), this section of the curve also contains data with sub-decadal (1–3 calendar years) resolution on sections of European oak from the Rijksuniversiteit Groningen (de Jong *et al.* 1986; corrected as described by de Jong *et al.* 1989). This means that the part of the calibration curve of relevance to this series of papers is unusually well replicated, and contains data of a resolution uncommon in the prehistoric period (but see Vogel & van der Plicht 1993). This curve was constructed by a group of laboratories who exchanged standard materials and known-age wood, minimizing inter-laboratory offsets to within the range 0–20 bp (Stuiver *et al.* 1998b). For example, the difference between the Groningen and Pretoria conventional laboratories (GrN and PtA) on the same material is on average  $7.1 \pm 6.4$  bp.

To examine the effect of the calibration data on the model of the chronology of the fictitious long barrow defined in Figure 7, it was re-calculated five times using different calibration data. This is a sensitivity analysis to examine the effect of varying the calibration data; this does not mean that the authors of this

paper advocate the use of any curve than that currently agreed by the international scientific community for routine use in archaeological applications. All the models presented in this series of papers have been calculated using INTCAL04 (Reimer *et al.* 2004), and in Table 1 the radiocarbon ages have been simulated from the estimated calendar ages using INTCAL04 which thus provides the base-line against which other data are compared. Here the model has been run using INTCAL04 (Reimer *et al.* 2004), as published and with the quoted errors arbitrarily halved and doubled. It has also been run using the previous internationally-agreed curve INTCAL98 (Stuiver *et al.* 1998b). These curves include the same radiocarbon measurements, but differ in the mathematical methods used for their construction (Stuiver *et al.* 1998b; Buck & Blackwell 2004). Finally, the model has been run using the Groningen data alone (extended where this runs out using INTCAL98). This is the highest resolution data available for this period at present, although as a

single record it is not as robust as curves constructed from replicate data sets.

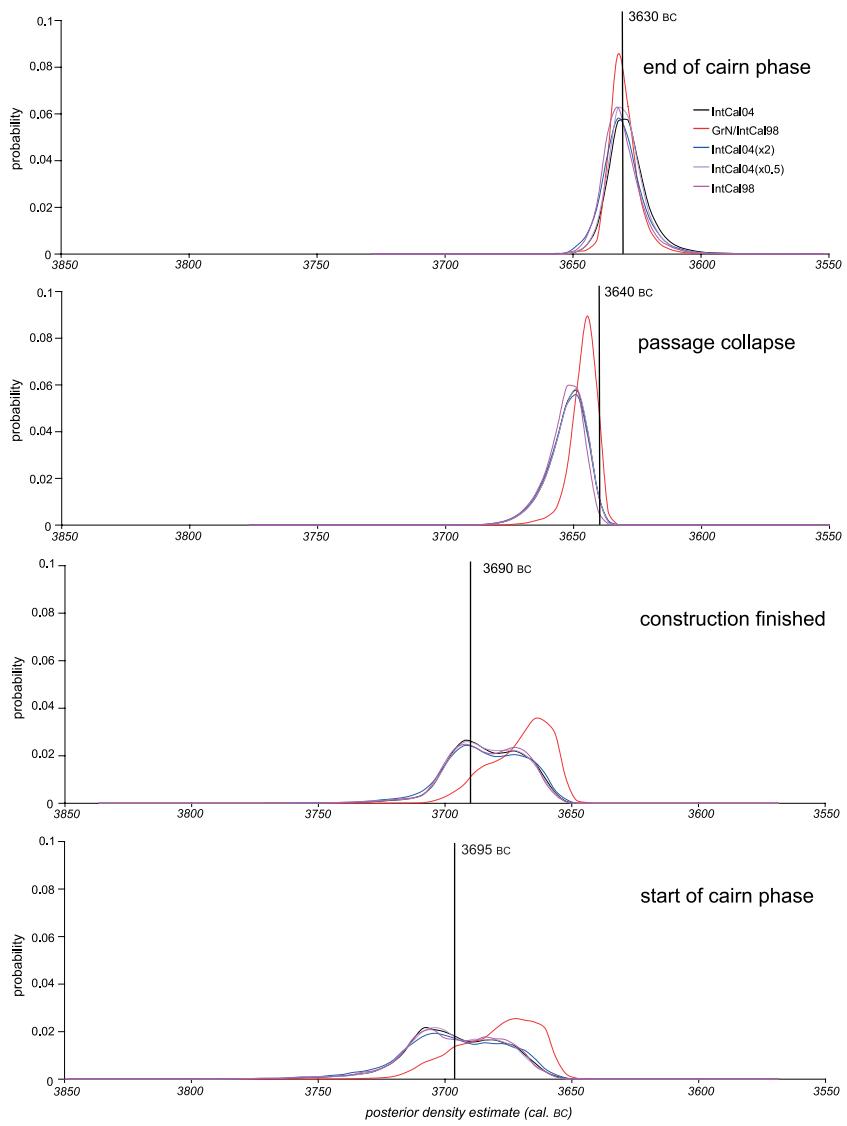
The results of these analyses for the five key parameters of the model shown in Figure 7 are shown in Figures 19 and 20. It can be seen that in all cases these models produce posterior density estimates which include the actual dates for the archaeological events concerned. The outputs for the model using the different calibration data sets are remarkably robust, suggesting that the ongoing refinements to the calibration data are unlikely to affect our results substantively. The largest, although still subtle, differences are seen when using the high-resolution Groningen data. This may be because there is further structure in the shape of the calibration curve which is masked when using measurements made on decadal or bi-decadal wood samples. Differences of this scale have been observed when refining the calibration data from bi-decadal to decadal measurements in the mid first millennium AD (McCormac *et al.* 2004). This may suggest that it is a detailed knowledge of the shape of the calibration curve, rather than the precision of the measurements on which it is based, which is of practical importance for archaeological applications.

### The errors on the radiocarbon measurements

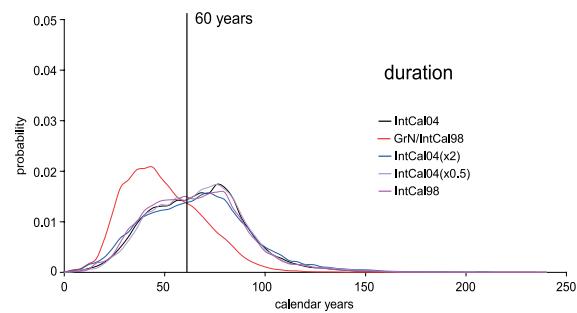
Turning now to the radiocarbon measurements made on archaeological samples by radiocarbon-dating laboratories, the model of the chronology of the fictitious long barrow defined in Figure 7 has been re-calculated three times. The quoted errors are in this case the same as the true errors on the measurements, as the scatter of radiocarbon ages for these samples has been calculated based on them (see Table 1). The outputs from the model using these errors are therefore the standard against which the other models are judged. It should be noted, however, that in reality the true error of a radiocarbon age is never known, rather it is estimated by the radiocarbon laboratory. These uncertainty estimates are based on the irreducible statistical uncertainty of the measurement itself, known measurable blank corrections, and information on the repeatability of the whole dating process, including sample pre-treatment. To test how important the accurate estimation of laboratory errors is for this application, the model was re-calculated with the same radiocarbon ages (see Table 1), but with the errors terms either halved or doubled. In the case of the doubled error, this means that the error quoted by the laboratory would have been conservative and the measurement is actually more precise than admitted. In the case of the halved error, the error quoted by the laboratory would have been too optimistic and the measurement is in reality less precise than claimed.

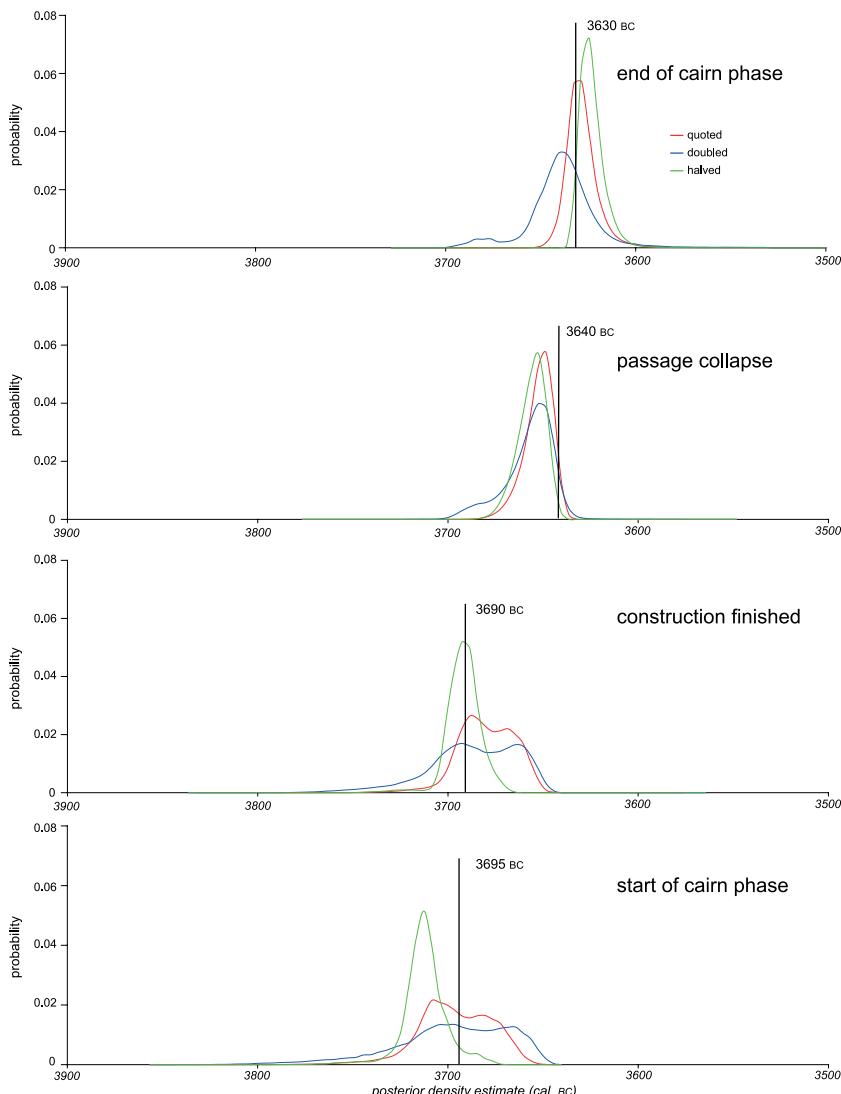
Figures 21 and 22 show the results of these analyses for the five key parameters of the model described in Figure 7. For the actual dates of the key events in the history of the barrow (Fig. 21), only the estimate for the start of the cairn phase is sensitive to the quoted error.

**Figure 20.** Graph showing the variation in the posterior density estimate for the duration of burial for the model described in Figure 7, calculated using the calibration data described in Figure 19.

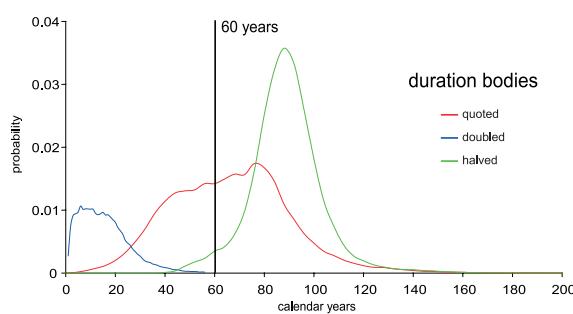


**Figure 19.** Graphs showing the variation of the posterior density estimates for the key chronological parameters of the model described in Figure 7, calculated using INTCAL04 (Reimer et al. 2004), INTCAL98 (Stuiver et al. 1998b), INTCAL04 with error arbitrarily halved and doubled, and the high-resolution from the Rijksuniversiteit Groningen (de Jong et al. 1989) extended using INTCAL98 where necessary.





**Figure 21.** Graphs showing the variation in the posterior density estimates for the key chronological parameters in versions of the model described in Figure 7, where the quoted errors are accurate, half, or double the actual error.



**Figure 22.** Graph showing the variation in the posterior density estimate for the duration of burial for the model described in Figure 7, where the quoted errors are accurate, half, or double the actual error.

For this parameter, when the error is under-quoted, the estimate is too early. The estimate for the duration of burial is, however, extremely sensitive to the accuracy of the quoted error (Fig. 22). If it is under-quoted the model estimates that burial continued for longer than it did in reality, and if it is over-quoted the model estimates that burial continued for a shorter period than it did in reality. This is because the amount of scatter on the radiocarbon ages is determined by the true error on the measurements, and so if the quoted error is too large the model will estimate that there is less scatter than there is in reality (so estimating a shorter duration), and if the quoted error is too small the model will estimate that there is more scatter than there is in reality (so estimating a longer duration). If the quoted errors on the radiocarbon ages of the archaeological samples are not accurate estimates of the true errors of the radiocarbon measurements, then the results of a Bayesian model can be importantly wrong. The estimates of duration in such an application will be importantly wrong.

#### *Biased radiocarbon measurements*

In the previous analysis, the radiocarbon ages estimated by the laboratory were accurate but the quoted errors were not. What happens if the situation is reversed and the radiocarbon ages are biased, although the quoted errors are realistic? To examine this scenario the model defined in Figure 7 was re-calculated five times, in all cases using the error terms listed in Table 1, but using the radiocarbon ages in that table, and the same ages shifted systematically by 40 radiocarbon years (BP) older and younger and by 80 radiocarbon years (BP) older and younger.

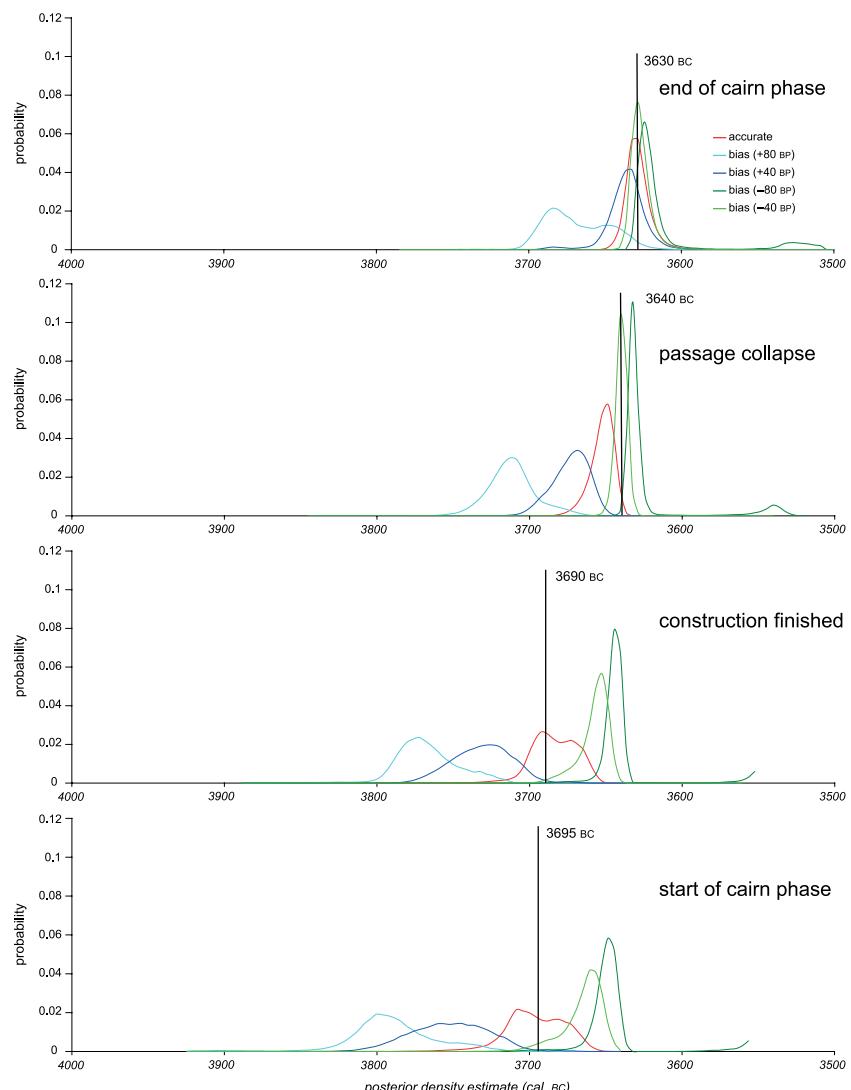
The results of these analyses are shown in Figures 23 and 24. For all cases radiocarbon results which are biased towards older ages provide estimated dates for the key parameters of this model which are too old. The radiocarbon results which are biased towards younger ages provide estimates which are too recent (Fig. 23). The estimates of duration provided by the models including systematically biased ages are also

incorrect, with longer estimates provided by models including results biased towards older ages and shorter estimates by models including results biased towards younger ages (Fig. 24). The direction of the inaccuracies in the estimates of durations provided in this instance may relate to the shape of the calibration curve for the particular application under consideration. It is clear, however, that systematic bias in the radiocarbon ages included in a model will cause the outputs of that model to be importantly wrong.

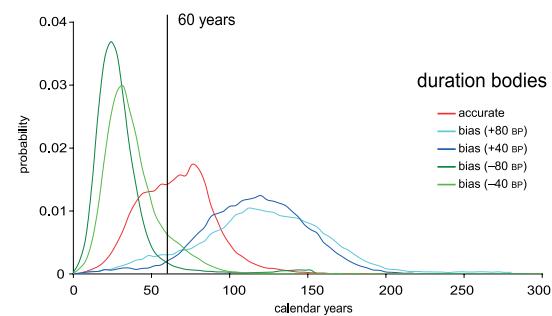
#### A digression on laboratory bias and accuracy

A bias in radiocarbon ages of the type described in the previous section is uncommon, but is of particular relevance to several of the papers included in this special supplement. Late in 2002, a technical problem was discovered with the ultrafiltration protocol used for processing bone and antler samples at the Oxford Radiocarbon Accelerator Unit which resulted in some bone samples giving ages which were about 100–300 radiocarbon years (BP) too old (Bronk Ramsey *et al.* 2004). This protocol was in use at Oxford between 2000 and 2002 (bone and antler samples in the range OxA-9361 to OxA-11851 and OxA-12214 to OxA-12236).

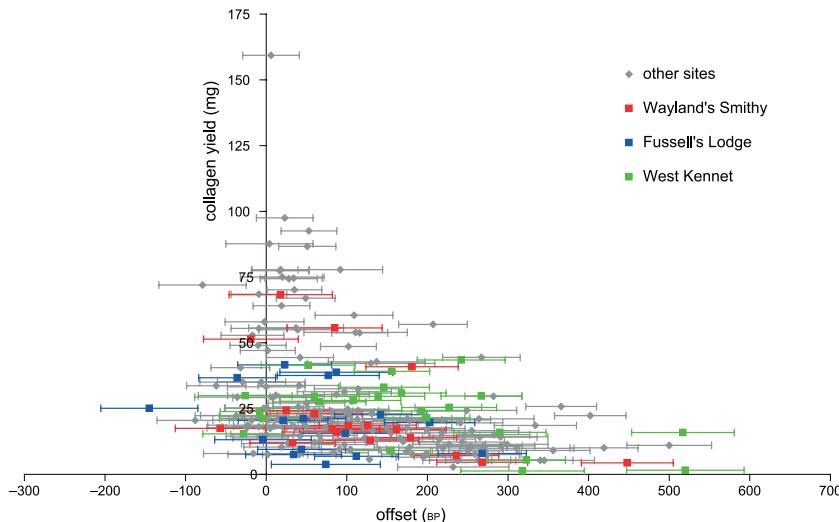
In total 19 measurements from Wayland's Smithy, 22 measurements from Fussell's Lodge, and 23 measurements from West Kennet were made using this method. As part of a programme to investigate further the scale of the problem, as many of the samples funded by English Heritage and originally dated using this protocol as possible have been re-dated. This has been done either by reprocessing stored gelatin from the original analysis using the revised protocol described by Bronk Ramsey *et al.* (2004), or by redating from samples of new bone (at Oxford or another laboratory). Figure 25 shows the offsets between the original ages quoted for these samples (almost all of which have now been withdrawn) and the new measurements, against the collagen yield of the original sample. The significant bias towards older



**Figure 23.** Graphs showing the variation in the posterior density estimates for the key chronological parameters in versions of the model described in Figure 7, where the measurements are accurate or systematically biased by the amounts indicated.



**Figure 24.** Graph showing the variation in the posterior density estimate for the duration of burial for the model described in Figure 7, where the measurements are accurate or systematically biased by the amounts indicated.



**Figure 25.** Offsets between radiocarbon results on bones produced using the ultrafiltration protocol documented in Bronk Ramsey *et al.* (2000a) and repeat measurements on the same samples ( $n = 204$ ). At Oxford these were made either by reprocessing stored gelatin from the original analysis using the revised protocol outlined in Bronk Ramsey *et al.* (2004) ( $OxA$  (NRC);  $n = 88$ ), or by redating from new samples of bone using this method ( $OxA$  (AF);  $n = 63$ ), sometimes with additional cleaning by a sequence of acetone, methanol, chloroform, and water ( $OxA$  (AF\*);  $n = 12$ ). Results were also available from the same samples measured previously in Oxford, using the methods of Bronk Ramsey *et al.* (2000b) ( $n = 9$ ) and Ward *et al.* (1984) ( $n = 1$ ). Replicate measurements were available from five other laboratories (AA,  $n = 9$ ; GrA,  $n = 35$ ; GU,  $n = 2$ , NZA,  $n = 10$ , and UB,  $n = 9$ ). Weighted means have been taken of repeat measurements when there is more than one on the same sample (Ward & Wilson 1978).

ages suggested by Bronk Ramsey *et al.* (2004, 156) is confirmed, and it can be seen that this bias includes most of the original results quoted for samples from these three long barrows. It appears that samples from West Kennet were most severely affected by this problem (average bias,  $177 \pm 151$  BP), with Wayland's Smithy and particularly Fussell's Lodge less severely affected (average bias respectively  $119 \pm 118$  BP and  $65 \pm 95$  BP).

It must be emphasized that all original measurements from these sites which may have been biased by the original ultrafiltration protocol have been withdrawn. In 40 cases, however, measurements have been made on re-purified gelatin (denoted by \* in Bayliss, Whittle & Wysocki this issue, p. 90, table 1; Whittle, Bayliss & Wysocki this issue, p. 110, table 1; and Wysocki, Bayliss & Whittle this issue, p. 72, table 1). Evidence for the accuracy of the results from this purified material is provided in Figure 26. This shows the offset between the results on repurified gelatin and results on fresh samples of bone measured by either at Oxford or by the Rijksuniversiteit Groningen as part of

this initiative (average bias,  $4 \pm 52$  BP), and demonstrates the accuracy of the results on the refiltered material. Full details of this work can be found in Bronk Ramsey *et al.* (forthcoming).

The accuracy of the measurements from the sites discussed in this series of papers is also supported by consideration of the replicate radiocarbon results. Sixteen samples have more than one determination, and in thirteen cases the results are by two different laboratories. In only one case are the replicate results statistically significantly different at 95% confidence (WK16 from West Kennet; Whittle, Bayliss & Wysocki this issue, p. 110, table 1), which is consistent with expectation as, if the ages and quoted errors are accurate, you would expect one in twenty to fail this test just due to random statistical error. This overall reproducibility is particularly convincing as evidence of the accuracy of the quoted measurements because such a high proportion of replicates are from more than one laboratory (Scott 2003).

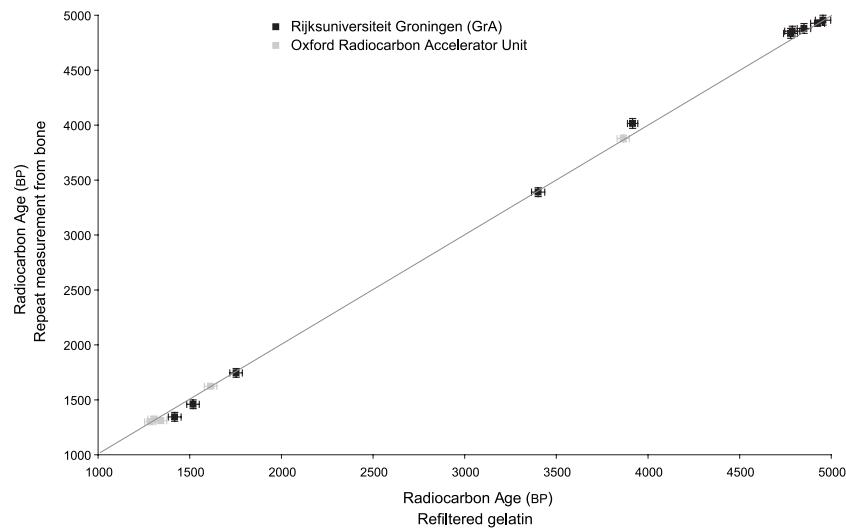
### Accurate beliefs?

These sensitivity analyses are a means to worry selectivity, providing an indication of the extent to which the chronologies presented in the following series of papers may be *importantly* wrong. They also allow us to assess which components of the models are most vulnerable to incorrect information.

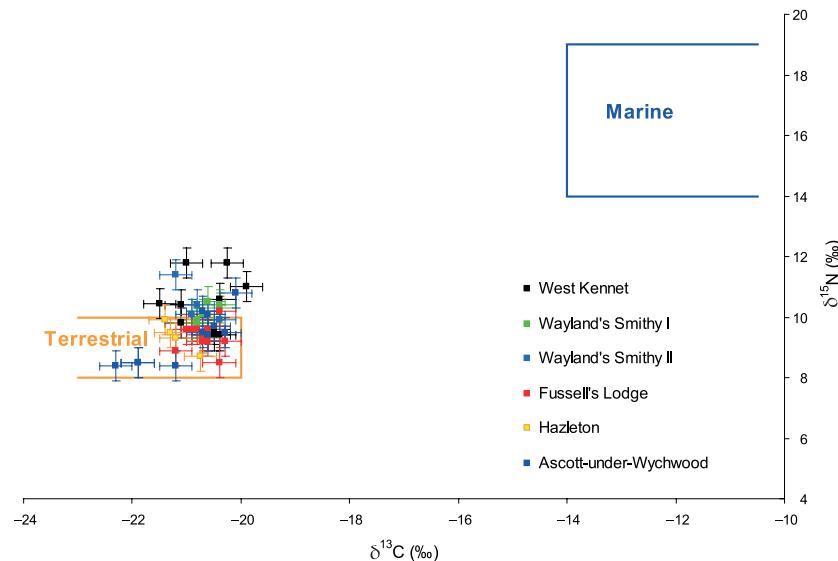
It is essential that the informative prior beliefs — the archaeological information and particularly the stratigraphic sequence of samples included in a model — are accurate or the resulting date estimates will be importantly wrong. The explicit definition of this information, critical assessments of its reliability, and alternative readings of excavation records are extensively discussed in each site-specific paper in this series. In contrast, the uninformative prior beliefs, i.e. assumptions about the distributions of dated events, have to be grossly wrong before they are problematic for the estimation of accurate chronologies. All of the models presented do, however, make some assumption about the distribution of dated events and so one pitfall which may make a model importantly wrong has been avoided.

The accuracy of the radiocarbon ages included in a model is also essential. Models based on dates whose errors are over- or under-estimated can be importantly wrong (particularly in regard to estimates of duration), and models based on systematically biased ages will produce chronologies which are importantly wrong. Considerable efforts are made by radiocarbon dating laboratories to ensure the necessary accuracy (e.g. Scott 2003), including in this case strenuous efforts to resolve a problem identified with a chemical protocol (see above and Bronk Ramsey *et al.* 2004). The consistency of the radiocarbon results produced by more than one laboratory, both when comparing replicate ages obtained on the same sample and in the good agreement of ages on stratigraphically related samples, strongly suggests that the quoted determinations are accurate. Although it is likely that calibration data will be refined in the future, this is unlikely to produce substantial differences in our chronologies in this period. In general, it seems that it is the shape, rather than the precision, of the calibration curve which is important in applications such as those presented in this series of papers.

Assessing the accuracy of our posterior beliefs in the light of methodological and scientific development is much more difficult to quantify. Only the uniform distribution of dated events is currently available for routine use as uninformative prior information (Buck *et al.* 1992). Although it is now possible to implement other distributions for a series of dated events, it is not yet clear how much difference their implementation will make in practice (OxCal v4.0β; [www.rlaha.ox.ac.uk/orau](http://www.rlaha.ox.ac.uk/orau)). In due course, formal mathematical criteria may be able to aid the archaeologist in the selection of the most plausible model. The comparison of models in a Bayesian setting, however, is yet to be made an exact science (Gilks *et al.* 1996, ch. 6). Bayes factors represent the relative weight in favour of one model over another and, although some attempt is being



**Figure 26.** Radiocarbon results on bones produced by reprocessing stored gelatin from the original analysis using the revised protocol outlined in Bronk Ramsey *et al.* (2004) and results from the same samples redated from replicate bone samples ( $n = 15$ ).



**Figure 27.** Graph of  $\delta^{13}\text{C}$  and  $\delta^{15}\text{N}$  values of bone collagen from adult skeletons from West Kennet, Wayland's Smithy, Fussell's Lodge, Hazleton, and Ascott-under-Wychwood related to the values expected for archaeological populations consuming pure C3 and marine diets (after Mays 1998).

made to implement them for chronological modelling (Millard & Brooks forthcoming), their computation is currently prohibitively expensive for routine use. There are also technical issues to be resolved for this approach, particularly for the comparison of models incorporating different levels of constraint (Lindley 1957). Alternative approaches, such as the Deviance Information Criterion, have also been suggested, al-

though again all are still the subject of further research (Spiegelhalter *et al.* 2002).

The dating programmes for the five long barrows and long cairns reported in this series of papers were initiated before the dating of cremated bone became widely available (Lanting *et al.* 2001). The implementation of this technique would enable further archaeological questions to be addressed at these sites, and also increase the archaeological reliability of our estimates for the time when these sites were in use for burial. In common with other human remains of Neolithic date from Britain, the carbon and nitrogen stable isotope values for the human remains sampled for dating from these monuments show little evidence for the consumption of marine or riverine resources (Richards *et al.* 2003; Fig. 27). The potential for dietary offsets between the radiocarbon in the bone gelatin from the dated individuals and the contemporary atmospheric concentration of radiocarbon has therefore not been considered (Bayliss *et al.* 2004). Neither has a potential time-lag in the turnover of the carbon in human bone (Barta & Stolc forthcoming), although many of the human remains dated were from adults. It is unclear whether quantitative consideration of these scientific factors would affect the results of the chronological models importantly.

## Conclusions

Future-proofing apart, we feel that the chronological models presented in the following series of site-specific papers represent a real advance in our understanding of the Neolithic of southern England. Such models can only be constructed by detailed and multi-disciplinary analysis, and it is the production of the detail, not the radiocarbon measurements, that has been the major cost in these studies. We argue that such investment is amply justified by the archaeological results. Timetables of the sort presented in these papers are now not just achievable on a routine basis, but are a necessity if we are to address fundamental questions about our pasts, including the experience of the flow of life, the social marking of time, and the 'pattern of retensions from the past and protentions for the future' (Ingold 1993, 157) discussed at the start of this paper.

The dates presented in the following papers are based on chronological models. 'All models are wrong, some models are useful' (Box 1979, 202). We hope readers will find them useful, and will employ worry selectivity to determine whether and how each model may be importantly wrong. Imperfect and provisional as they undoubtedly are, nonetheless these chronologies are based on formal statistical approaches which

handle the probabilistic nature of radiocarbon age estimates. These models are models, and are wrong. They may, however, be useful and they may not be importantly wrong.

In contrast, the less formal approaches to interpreting radiocarbon dates which are very widely used by prehistorians are very frequently importantly wrong and misleading. Not only does it appear that human activities which may in fact have been separated by several centuries were contemporary, but it also appears, erroneously, that activities lasted for much longer than they did in reality. The prehistoric past deserves better, especially as it 'exists not as the past studied in itself but represents a project in the present' (Shanks & Tilley 1987, 211).

## Acknowledgements

We would like to thank Ian Dennis and Derek Hamilton for preparation of the figures, and Geoff Nicholls for discussions inspiring Figure 12. Paula Reimer kindly answered queries relating to the precise data sets included in IntCal04. The introduction to the Bayesian method has benefited from discussion with John Meadows. Helpful comments on an earlier version of this paper were provided by Andrew David, Derek Hamilton, Frances Healy, Jonathan Last, Peter Marshall, Jane Sidell, the editor and one anonymous referee.

## Notes

1. 'Bradshaw' has also been used more recently to denote a group of Australian Aboriginal rock art, of indeterminate but possibly early date: *Times Literary Supplement*, July 28 and September 1, 2006.
2. The cost of the actual radiocarbon measurements is probably the least significant of these factors. Substantial resources are available for radiocarbon dating in English archaeology (since 1992 in London alone 885 measurements have been obtained, almost 80% by the commercial sector: Meadows *et al.* forthcoming). The cost-effectiveness of these approaches has also been demonstrated by English Heritage, increasing the precision of dating produced by around 25% (Bayliss & Bronk Ramsey 2004, 26). However, the real cost of the additional time required for sample selection, modelling and interpretation is probably at least as substantial as the costs of the radiocarbon measurements themselves.
3. This is a transform of a pseudo-Bayes factor and, although the mathematical formulation is not entirely rigorous, in practice it does provide a good working indication of when a statistical model is inconsistent with the age measurements used (Bayliss & Bronk Ramsey 2004, 39).
4. This format has been followed throughout this paper: so, dates 'bc' are estimated or actual dates on the cal-

- endar scale and dates 'cal. BC' derive from calibrated radiocarbon ages. Dates '*cal. BC*' in italics are posterior density estimates derived from Bayesian modelling. Radiocarbon ages are given as 'BP', a unit used for radiocarbon measurement. This is defined as the  $^{14}\text{C}$  radioactivity measured relative to a standard activity (oxalic acid), with ages calculated using a conventional half-life and corrected for isotopic effects using the stable carbon isotope  $^{13}\text{C}$ .
5. An audience of experienced archaeologists from the commercial and academic sectors, which met in Cardiff in August 2006 to discuss results from the dating programme on causewayed enclosures, were asked to estimate the dates of construction and abandonment, and duration of use, from the same simulation (Fig. 4). The vast majority of the audience got the answers *importantly* wrong – 81% estimated the start date inaccurately, all estimated the end date inaccurately, and 94% significantly over-estimated the duration!
  6. For example, at the start of the fourth millennium BC the concentration of radiocarbon in the atmosphere changes quite rapidly, producing a steep gradient to the calibration curve. Between c. 3900 BC and c. 3700 BC this flattens off, before two pronounced wiggles appear in the curve in the middle of the millennium (Fig. 18). From c. 3350 BC the amount of  $^{14}\text{C}$  in the atmosphere remained reasonably constant until the end of the millennium, resulting in the 'middle Neolithic' plateau. In practice, this means that a calibrated radiocarbon age for a sample dating to 3950 BC will have a more precise calibrated date range, than an age with the same error term for a sample dated to 3250 BC.
  7. In practice outliers may be identified by formal statistical outlier detection (Christen 1994), low indices of agreement, on archaeological grounds, or on the basis of laboratory data.
  8. So, for example, a posterior density estimate with a range of 3710–3650 cal. BC (95% probability) has a calendrical bandwidth of 60 years (95% probability).

## References

- Aitchison, T., B. Ottaway & A.S. AlRuzaiza, 1991. Summarizing a group of  $^{14}\text{C}$  dates on the historical time scale: with a worked example from the late Neolithic of Bavaria. *Antiquity* 65, 108–16.
- Ashmore, P., 2004. Absolute chronology, in *Scotland in Ancient Europe: the Neolithic and Early Bronze Age of Scotland in their European Context*, eds. I.A.G. Shepherd & G.J. Barclay. Edinburgh: Society of Antiquaries of Scotland, 125–36.
- Ashmore, P., 2005. Dating Barnhouse, in *Dwelling Among the Monuments: the Neolithic Village of Barnhouse, Maeshowe Passage Grave and Surrounding Monuments at Stenness, Orkney*, ed. C. Richards. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research, 385–8.
- Barta, P. & S. Stolc, forthcoming. HBCO correction: its impact on archaeological absolute dating. *Radiocarbon*.
- Bayes, T.R., 1763. An essay towards solving a problem in the doctrine of chances. *Philosophical Transactions of the Royal Society* 53, 370–418.
- Bayliss, A. & C. Bronk Ramsey, 2004. Pragmatic Bayesians: a decade integrating radiocarbon dates into chronological models, in *Tools for Constructing Chronologies: Tools for Crossing Disciplinary Boundaries*, eds. C.E. Buck & A.R. Millard. London: Springer, 25–41.
- Bayliss, A., C. Bronk Ramsey & F.G. McCormac, 1997. Dating Stonehenge, in *Science and Stonehenge*, eds. B. Cunliffe & C. Renfrew. Oxford: British Academy, 35–59.
- Bayliss, A., E. Shepherd Popescu, N. Beavan-Athfield, C. Bronk Ramsey, G.T. Cook & A. Locker, 2004. The potential significance of dietary offsets for the interpretation of radiocarbon dates: an archaeological significant example from medieval Norwich. *Journal of Archaeological Science* 31, 563–75.
- Bayliss, A., A. Whittle & M. Wysocki, this issue. Talking about my generation: the date of the West Kennet long barrow.
- Blaauw, M., G.B. Heuvelink, D. Mauquay, J. van der Plicht & B. van Geel, 2003. A numerical approach to  $^{14}\text{C}$  wigglematch dating of organic deposits: best fits and confidence intervals. *Quaternary Science Reviews* 22, 1485–500.
- Blackwell, P.G. & C.E. Buck, 2003. The Late Glacial human re-occupation of northwestern Europe: new approaches to spacetime modelling. *Antiquity* 77, 232–40.
- Box, G.E.P., 1976. Science and statistics. *Journal of the American Statistical Association* 71, 791–9.
- Box, G.E.P., 1979. Robustness in scientific model building, in *Robustness in Statistics*, eds. R.L. Launer & G.N. Wilkinson. New York (NY): Academic Press, 201–36.
- Bronk Ramsey, C., 1995. Radiocarbon calibration and analysis of stratigraphy. *Radiocarbon* 36, 425–30.
- Bronk Ramsey, C., 1998. Probability and dating. *Radiocarbon* 40, 461–74.
- Bronk Ramsey, C., 2000. Comment on 'The use of Bayesian statistics for  $^{14}\text{C}$  dates of chronologically ordered samples: a critical analysis'. *Radiocarbon* 42, 199–202.
- Bronk Ramsey, C., 2001. Development of the radiocarbon calibration program. *Radiocarbon* 43, 355–63.
- Bronk Ramsey, C., P.B. Pettitt, R.E.M. Hedges, G.W.L. Hodgins & D.C. Owen, 2000a. Radiocarbon dates from the Oxford AMS system: *Archaeometry* datelist 30. *Archaeometry* 42, 459–79.
- Bronk Ramsey, C., P.B. Pettitt, R.E.M. Hedges, G.W.L. Hodgins & D.C. Owen, 2000b. Radiocarbon dates from the Oxford AMS system: *Archaeometry* datelist 29. *Archaeometry* 42, 243–54.
- Bronk Ramsey, C., T. Higham, A. Bowles & R.E.M. Hedges, 2004. Improvements to the pretreatment of bone at Oxford. *Radiocarbon* 46, 155–63.
- Bronk Ramsey, C., T. Higham & J. Pearson, forthcoming. *Bone Pretreatment by Ultrafiltration: a Report on Unintended Age Offsets Introduced by the Method*. English Heritage Research Department Report.
- Bruins, H.J., J. van der Plicht & A. Mazar, 2003.  $^{14}\text{C}$  dates from Tel Rehov: Iron Age chronology, pharaohs, and

- Hebrew kings. *Science* 300, 315–18.
- Buck, C.E. & P.G. Blackwell, 2004. Formal statistical models for radiocarbon calibration curves. *Radiocarbon* 46, 1093–102.
- Buck, C.E., J.B. Kenworthy, C.D. Litton & A.F.M. Smith, 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65, 808–21.
- Buck, C.E., C.D. Litton & A.F.M. Smith, 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19, 497–512.
- Buck, C.E., J.A. Christen, J.B. Kenworthy & C.D. Litton, 1994a. Estimating the duration of archaeological activity using  $^{14}\text{C}$  determinations. *Oxford Journal of Archaeology* 13, 229–40.
- Buck, C.E., C.D. Litton & E.M. Scott, 1994b. Making the most of radiocarbon dating: some statistical considerations. *Antiquity* 68, 252–63.
- Buck, C.E., C.D. Litton & S.J. Shennan, 1994c. A case study in combining radiocarbon and archaeological information: the early Bronze Age settlement of St. VeitKlinglberg, Land Salzburg, Austria. *Germania* 72, 427–47.
- Buck, C.E., W.G. Cavanagh & C.D. Litton, 1996. *Bayesian Approach to Interpreting Archaeological Data*. Chichester: Wiley.
- Buck, C.E., T.F.G. Higham & D.J. Lowe, 2003. Bayesian tools for tephrochronology. *Holocene* 13, 639–47.
- Christen, J.A., 1994. Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics* 43, 489–503.
- Cleal, R.M.J., K.E. Walker & R. Montague, 1995. *Stonehenge in its Landscape: Twentieth-century Excavations*. London: English Heritage.
- Czebreszuk, J. & J. Müller (eds.), 2001. *Die absolute Chronologie in Mitteleuropa 3000–2000 v. Chr.* Studien zur Archäologie in Ostmitteleuropa 1, Studian nad Pradziejami Europy Środkowej 1. Poznań, Bamberg, Rahden: Leidorf.
- Czebreszuk, J. & M. Szmyt, 2001. The 3rd millennium BC in Kujawy in the light of  $^{14}\text{C}$  dates, in *Die absolute Chronologie in Mitteleuropa 3000–2000 v. Chr.*, eds. J. Czebreszuk & J. Müller. Poznań, Bamberg, Rahden: Leidorf, 177–208.
- Darvill, T., 2004. *Long Barrows of the Cotswolds and Surrounding Areas*. Stroud: Tempus.
- de Jong, A.F.M., W.G. Mook & B. Becker, 1986. High-precision calibration of the radiocarbon time scale, 3930–3230 cal. BC. *Radiocarbon* 28, 954–60.
- de Jong, A.F.M., W.G. Mook & B. Becker, 1989. Corrected calibration of the radiocarbon time scale, 3904–3203 cal. BC. *Radiocarbon* 31, 201–5.
- Dubouloz, J., 2003. Datation absolue du premier Néolithique du Bassin parisien: complément et relecture des données RRBp et VSG. *Bulletin de la Société Préhistorique Française* 100, 671–89.
- Evans, C. & I. Hodder, 2006. *A Woodland Archaeology: Neolithic Site at Haddenham*. (McDonald Institute Monographs.) Cambridge: McDonald Institute for Archaeological Research.
- Fabian, J., 1983. *Time and the Other: How Anthropology Makes its Object*. New York (NY): Columbia University Press.
- Friedrich, W.L., B. Kromer, M. Friedrich, J. Heinemeier, T. Pfeiffer & S. Talamo, 2006. Santorini eruption radiocarbon dated to 1627–1600 BC. *Science* 312, 548.
- Garwood, P. & A. Brindley, 1999. Grooved Ware in southern Britain: chronology and interpretation, in *Grooved Ware in Britain and Ireland*, eds. R. Cleal & A. MacSween. Oxford: Oxbow, 145–76.
- Gelfand, A.E. & A.F.M. Smith, 1990. Sampling approaches to calculating marginal densities. *Journal of the American Statistical Association* 85, 398–409.
- Gell, A., 1992. *The Anthropology of Time: Cultural Constructions of Temporal Maps and Images*. Oxford: Berg.
- Geyh, M. & P. de Maret, 1982. Histogram evaluation of  $^{14}\text{C}$  dates applied to the first Iron Age sequence from West Central Africa. *Archaeometry* 24, 158–63.
- Gibson, A. & I. Kinnes, 1997. On the urns of a dilemma: radiocarbon and the Peterborough problem. *Oxford Journal of Archaeology* 16, 65–72.
- Gilks, W.R., S. Richardson & D.J. Spiegelhalter, 1996. *Markov Chain Monte Carlo in Practice*. London: Chapman and Hall.
- Hamilton, W.D., A. Bayliss, C. Bronk Ramsey, S. Freeman & A. Menuge, in preparation. Validating a Bayesian chronology: dating the north wing of Baguley Hall, Greater Manchester, UK.
- Härke, H., 1991. All quiet on the Western front? Paradigms, methods, and approaches in West German archaeology, in *Archaeological Theory in Europe: the Last Three Decades*, ed. I. Hodder. London: Routledge, 187–222.
- Higham, T.F.G., C. Bronk Ramsey, F.J. Petchey, C. Tompkins & M. Taylor, 2004. AMS radiocarbon dating of *Rattus exulans* bone from the Kokohuia site (New Zealand), in *Radiocarbon and Archaeology: Proceedings of the 4th Symposium, Oxford 2002*, eds. T. Higham, C. Bronk Ramsey & C. Owen. (Oxford University School of Archaeology Monograph 62.) Oxford: Oxford University, 135–51.
- Hillam, J., C.M. Groves, D.M. Brown, M.G.L. Baillie, J.M. Coles & B.J. Coles, 1990. Dendrochronology of the English Neolithic. *Antiquity* 64, 210–20.
- Hodder, I., 1992. *Theory and Practice in Archaeology*. London: Routledge.
- Hodder, I., 1999. *The Archaeological Process*. Oxford: Blackwell.
- Hodder, I. & S. Hutson, 2003. *Reading the Past: Current Approaches to Interpretation in Archaeology*. 3rd edition. Cambridge: Cambridge University Press.
- Housley, R.A., C.S. Gamble, M. Street & P. Pettitt, 1997. Radiocarbon evidence for the Late glacial human recolonisation of northern Europe. *Proceedings of the Prehistoric Society* 63, 25–54.
- Ingold, T., 1993. The temporality of the landscape. *World Archaeology* 25, 152–74.
- Johnson, M., 1999. *Archaeological Theory: an Introduction*. Oxford: Blackwell.
- Jørgensen, L. (ed.), 1992. *Chronological Studies of Anglo-Saxon*

- England, Lombard Italy and Vendel Period Sweden.* Copenhagen: University of Copenhagen.
- Kim, I.C., J.C. Kim, J.H. Park, M.Y. Young, S.B. Kim & E.S. Lee, 2004. Dating the King's tombs of the kingdom of old Silla, Korea, in *Radiocarbon and Archaeology: Proceedings of the 4th Symposium, Oxford 2002*, eds. T. Higham, C. Bronk Ramsey & C. Owen. (Oxford University School of Archaeology Monograph 62.) Oxford: Oxford University, 185–92.
- Kinnes, I., A. Gibson, J. Ambers, S. Bowman, M. Leese & R. Boast, 1991. Radiocarbon dating and British Beakers: the British Museum programme. *Scottish Archaeological Review* 8, 35–68.
- Lanting, J.N., A.T. Aerts-Bijma & J. van der Plicht, 2001. Dating of cremated bones. *Radiocarbon* 43, 249–54.
- Lindley, D.V., 1957. A statistical paradox. *Biometrika* 44, 187–92.
- Lindley, D.V., 1985. *Making Decisions*. 2nd edition. London: Wiley.
- Lu, X., Z. Guo, H. Ma, S. Yuan & X. Wu, 2001. Data analysis and calibration of radiocarbon dating results from the cemetery of the Marquises of Jin. *Radiocarbon* 43, 55–62.
- Lucas, G., 2005. *The Archaeology of Time*. London: Routledge.
- Lüning, J., 2005. Bandkeramische Hofplätze und die absolute Chronologie der Bandkeramik, in *Die Bandkeramik im 21. Jahrhundert*, eds. J. Lüning, C. Frirdich & A. Zimmermann. Rahden: Marie Leidorf, 49–74.
- Manning, S.W., C. Bronk Ramsey, W. Kutschera, T. Higham, B. Kromer, P. Steier & E.M. Wild, 2006. Chronology for the Aegean Late Bronze Age 1700–1400 BC. *Science* 312, 565–9.
- Mays, S., 1998. *The Archaeology of Human Bones*. London: Routledge.
- Meadows, J., A. Barclay & A. Bayliss, this issue. A short passage of time: the dating of the Hazleton long cairn revisited.
- Meadows, J., J. Sidell, H. Swain & B. Taylor, forthcoming. *Absolute Dating: a Regional Review for London*, vol. I: *Physical and Chemical Methods*. London: Museum of London.
- Mercer, R. & F. Healy, forthcoming. *Hambledon Hill, Dorset, England: Excavation and Survey of a Neolithic Monument Complex and its Surrounding Landscape*. Swindon: English Heritage.
- McCormac, F.G., A. Bayliss, M.G.L. Baillie & D.M. Brown, 2004. Radiocarbon calibration in the Anglo-Saxon period: AD 495–725. *Radiocarbon* 46, 1123–5.
- McTaggart, J.E.M., 1908. The unreality of time. *Mind* 17, 457–74.
- Michałski, A. & A. Pazdur, 2003. The method of combining radiocarbon dates and other information in application to study the chronologies of archaeological sites. *Geochronometria* 22, 41–6.
- Millard, A.R. & S.P. Brooks, forthcoming. Age-depth models for sediment cores with changes in accumulation rate. *Radiocarbon*.
- Müller, J. & A. Zimmermann (eds.), 1997. *Archäologie und Korrespondenzanalyse. Internationale Archäologie* 23.
- Needham, S., C. Bronk Ramsey, D. Coombs, C. Cartwright & P.B. Pettitt, 1998. An independent chronology for British Bronze Age metalwork: the results of the Oxford Radiocarbon Accelerator Programme. *Archaeological Journal* 154, 55–107.
- Nicholls, G. & M. Jones, 2001. Radiocarbon dating with temporal order constraints. *Applied Statistics* 50, 503–21.
- Ottaway, B., 1973. Dispersion diagrams: a new approach to the display of <sup>14</sup>C dates. *Archaeometry* 15, 5–12.
- Pearson, G.W., J.R. Pilcher, M.G.L. Baillie, D.M. Corbett & F. Qua, 1986. High-precision <sup>14</sup>C measurements of Irish oaks to show the natural <sup>14</sup>C variations from AD 1840 to 5210 BC. *Radiocarbon* 28, 911–34.
- Pettitt, P.B., S.W.G. Davies, C.S. Gamble & M.B. Richards, 2003. Palaeolithic radiocarbon chronology: quantifying our confidence beyond two half-lives. *Journal of Archaeological Science* 30, 1685–93.
- Reece, R., 1994. Are Bayesian statistics useful to archaeological reasoning? *Antiquity* 68, 848–50.
- Reimer, P.J., M.G.L. Baillie, E. Bard, A. Bayliss, J.W. Beck, C.J.H. Bertrand, P.G. Blackwell, C.E. Buck, G.S. Burr, K.B. Cutler, P.E. Damon, R.L. Edwards, R.G. Fairbanks, M. Friedrich, T.P. Guilderson, A.G. Hogg, K.A. Hughen, B. Kromer, F.G. McCormac, S. Manning, C. Bronk Ramsey, R.W. Reimer, S. Remmelle, J.R. Southon, M. Stuiver, S. Talamo, F.W. Taylor, J. van der Plicht & C.E. Weyhenmeyer, 2004. IntCal04 Terrestrial Radiocarbon Age Calibration, 026 cal kyr BP. *Radiocarbon* 46, 1029–58.
- Richards, M.P., R.J. Schulting & R.E.M. Hedges, 2003. Sharp shift in diet at onset of Neolithic. *Nature* 425, 366.
- Rom, W., R. Golser, W. Kutschera, A. Priller, P. Steier & E.M. Wild, 1999. AMS <sup>14</sup>C dating of equipment from the Iceman and of spruce logs from the prehistoric salt mines of Hallstatt. *Radiocarbon* 41, 183–97.
- Russell, T.M., 2004. *The Spatial Analysis of Radiocarbon Databases: the Spread of the First Farmers in Europe and of the Fat-tailed Sheep in Southern Africa.* (British Archaeological Reports International Series 1294.) Oxford: BAR.
- Scarre, C., P. Arias, G. Burenhult, M. Fano, L. Oosterbeek, R. Schulting, A. Sheridan & A. Whittle, 2003. Megalithic chronologies, in *Stones and Bones: Formal Disposal of the Dead in Atlantic Europe during the Mesolithic–Neolithic interface 6000–3000 BC*, ed. G. Burenhult. (British Archaeological Reports International Series 1201.) Oxford: BAR, 65–111.
- Schulting, R.J., 2000. New AMS dates from the Lambourn long barrow and the question of the earliest Neolithic in southern England: repacking the Neolithic package? *Oxford Journal of Archaeology* 19, 25–35.
- Scott, E.M. (ed.), 2003. The Third International Radiocarbon Intercomparison (TIRI) and the Fourth International Radiocarbon Intercomparison (FIRI) 1990–2002: results, analysis, and conclusions. *Radiocarbon* 45, 135–408.
- Shanks, M. & C. Tilley, 1987. *Social Theory in Archaeology*. Oxford: Polity Press.
- Spiegelhalter, D., N. Best, B. Carlin & A. van der Linde,

2002. Bayesian measures of model complexity and fit. *Journal of the Royal Statistical Society (Series B)* 64, 583–639.
- Steel, D., 2001. Bayesian statistics in radiocarbon calibration. *Philosophy of Science* 68, S153–64.
- Steier, P. & W. Rom, 2000. The use of Bayesian statistics for  $^{14}\text{C}$  dates of chronologically ordered samples: a critical analysis. *Radiocarbon* 42, 183–98.
- Strien, H.C. & D. Gronenborn, 2005. Klima und Kulturwandel Während des Mitteleuropäischen Altneolithikums (58./57.–51.50. Jahrhundert v. Chr.), in *Climate Variability and Culture Change in Neolithic Societies in Central Europe, 6700–2200 cal. bc*, ed. D. Gronenborn. Mainz: Verlag des Römisch-Germanischen Zentralmuseums, 131–49.
- Stuiver, M. & B. Becker, 1993. High-precision decadal calibration of the radiocarbon time scale, AD 1950–6000 BC. *Radiocarbon* 35, 35–66.
- Stuiver, M. & P.J. Reimer, 1993. Extended  $^{14}\text{C}$  data base and revised CALIB 3.0  $^{14}\text{C}$  age calibration program. *Radiocarbon* 35, 215–30.
- Stuiver, M., P.J. Reimer & T.T. Braziunas, 1998a. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* 40, 1127–51.
- Stuiver, M., P.J. Reimer, E. Bard, J.W. Beck, G.S. Burr, K.A. Hughen, B. Kromer, F.G. McCormac, J. van der Plicht & M. Spurk, 1998b. INTCAL98 radiocarbon age calibration, 24,000–0 cal bp. *Radiocarbon* 40, 1041–84.
- Vogel, J.C. & J. van der Plicht, 1993. Calibration curve for short-lived samples, 1900–3900 BC. *Radiocarbon* 35, 87–91.
- Voorrips, A., 1987. Formal and statistical models in archaeology, in *Quantitative Research in Archaeology*, ed. M.S. Aldenderfer. Beverly Hills (CA): Sage, 61–72.
- Wand, J.O., R. Gillespie & R.E.M. Hedges, 1984. Sample preparation for Accelerator-based radiocarbon dating. *Journal of Archaeological Science* 11, 159–63.
- Ward, G.K. & S.R. Wilson, 1978. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20, 19–31.
- Waterbolk, H.T., 1971. Working with radiocarbon dates. *Actes du VIII<sup>e</sup> congrès international des sciences préhistoriques et protohistoriques* 1, 11–25.
- Wheeler, R.E.M., 1956. *Archaeology from the Earth*. Harmondsworth: Penguin.
- Whittle, A., 1988. *Problems in Neolithic Archaeology*. Cambridge: Cambridge University Press.
- Wysocki, M., A. Bayliss & A. Whittle, this issue. Serious mortality: the date of the Fussell's Lodge long barrow.
- Whittle, A., A. Bayliss & M. Wysocki, this issue. Once in a lifetime: the date of the Wayland's Smithy long barrow.