Proceedings of the British Academy, **92**, 39–59

Dating Stonehenge

A. BAYLISS, C. BRONK RAMSEY, & F.G. McCORMAC

Introduction

AS PART OF THE RECENT PROJECT to complete the analysis of the twentieth century excavations at Stonehenge (Cleal *et al.* 1995), a series of 46 new radiocarbon determinations was commissioned. The 16 results which had been obtained on material from the monument before 1994 were critically reassessed on the same basis as the new results. Full details of this programme are published elsewhere (Allen and Bayliss 1995; http://www.eng-h.gov.uk/stoneh). The results from two samples measured subsequently, with the consequent slight modifications to the interpretative model, will appear shortly (Bronk Ramsey and Bayliss forthcoming).

This paper attempts to take a wider view and addresses some of the scientific and archaeological problems which have been raised by the Stonehenge dating programme.

The concept

In 1763 a letter was sent to the Royal Society by the Revd. Thomas Bayes (Bayes 1763) introducing a concept which is fundamental to how we have approached the problem of dating Stonehenge over two centuries later. His ideas are encapsulated in Bayes' theorem (Fig. 1) which provides a coherent and logical framework for revising current beliefs in the light of new information (Buck *et al.* 1991).

$$\text{Prior beliefs} \times \text{Standardized likelihood} = \text{Posterior beliefs}$$

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

Figure 1. Bayes' Theorem.

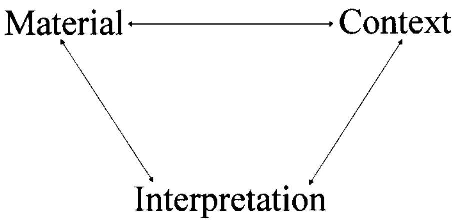

Figure 2. The relationships between data and interpretation.

The application of these mathematical techniques in association with Gibbs sampling to the dating of Stonehenge allows the radiocarbon determinations to be interpreted together with the stratigraphic and contextual evidence (Gelfand and Smith 1990; Bronk Ramsey 1995; http://www.rlaha.ox.ac.uk/). This enables us to formalise the links between our archaeological interpretations of the data and the data themselves (Fig. 2), to explore the effects of different interpretations, and so (hopefully) to produce more realistic estimates of the chronological parameters which are of interest to us.

The mathematical methods employed

The Bayesian analysis described here was performed using the program OxCal (Bronk Ramsey 1995) with a chronological model devised for this site (described below and more fully in Cleal *et al.* 1995 and Bronk Ramsey and Bayliss forthcoming). The exact mathematical methods employed are inevitably fairly complicated but the essential aspects of the method can be explained in terms of a number of simple stages.

The underlying assumption made is that without any of the archaeological information the events under investigation are equally likely to have occurred at any point in time (mathematically this can be expressed as a flat probability distribution extending over all time). Each piece of information at our disposal is then used in turn to modify this 'prior belief' according to Bayes' theorem. With material which has been radiocarbon dated the first stage in this process is to compare the measurement made to those on known age tree rings. This comparison is usually referred to as radiocarbon calibration and generates a new probability distribution.

Radiocarbon and other scientific dating methods rarely give the only information available on the chronology of a site. The Bayesian method allows other information to be treated in a similar way to modify further the probability distribution. We could include very broad-ranging assumptions, such as the site being pre-Roman or post-glacial, but these would not alter the distributions and so there is little point. At Stonehenge, as at many other sites, the most useful information concerns the relative ages of the various phases of the site—information which comes largely from stratigraphic relationships and the archaeological interpretation of these. Most relationships of this sort can be expressed in terms of one object being older than another (or one event occurring before another), but are more elegantly described as phases and sequences. Sometimes we can make more specific assumptions such as events within a phase being uniformly distributed throughout the phase.

It is mathematically possible to include all of this information analytically to produce modified probability distributions, but for anything but the most trivial examples this is impractical from a computational point of view. The program OxCal uses instead a method called Gibbs' Sampling (Buck *et al.* 1991). In this method a very large number of possible scenarios are randomly generated taking into account both the probability distributions from the radiocarbon evidence and the constraints imposed by the chronological relationships. These scenarios are then used to build up new probability distributions which take all of the information into account.

Phases 1 and 2

The first part of the monument to be constructed was the ditch and bank, with counterscarp bank (Fig. 3). The ditch gradually silted up, although there was a period of stabilisation during which an organic 'dark' layer formed on top of the primary silt. During the period of secondary silting, some features were cut into the ditch and there were some episodes of backfilling.

Only the ditch sequence provided material which could be dated in the recent campaign. The other structural elements from phases 1 and 2—the Aubrey Holes, the timber settings at the centre of the monument, the north-eastern entrance, and running towards the southern entrance, and the cremation cemetery—remain entirely undated except for a single measurement from Aubrey Hole 32 (C-602; 3798±275BP; 3020–1520 cal BC), which was measured by W.F. Libby in the early 1950s.

The model

Altogether there are 24 radiocarbon determinations from phases 1 and 2 which we consider reliable. Twenty-three of these are from the ditch sequence (see Table 1). From the stratigraphic sequence identified during excavation, some of these results can be placed in a relative order (Fig. 4).

When considering the stratigraphic constraints to be included in the model it is essential to consider the relationship between the samples dated and the archaeological parameters which are of interest. For example, in Phase 1 at Stonehenge we are very interested in the date of excavation of the main ditch. However we have actually dated a number of pieces of bone and antler from that ditch. The relationship between the date of the samples and the date of the archaeological event (the ditch digging) is both fundamental and interpretative.

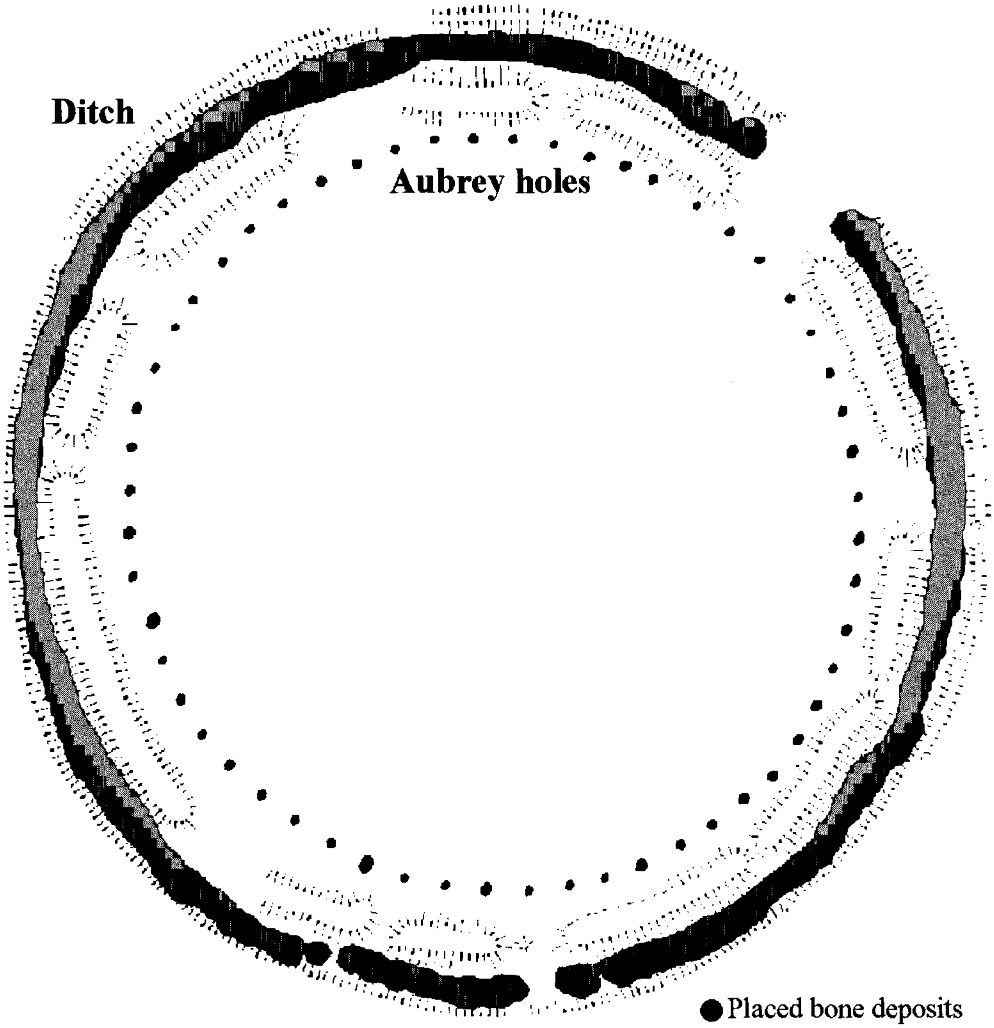

Figure 3. Plan of phase 1 at Stonehenge.

Taphonomy (functionality)

At the base of the ditch, beneath the primary silting, lay well over 100 antlers, many of which had been converted for use as picks or rakes and showed obvious signs of wear (Cleal *et al.* 1995, 414–26). The discovery of a lump of chalk with the tip of an antler tine embedded within it in Stone-hole 9, supports this functional interpretation (Cleal *et al.* 1994, 426; fig. 98). We interpret these antler tools as those which were used to dig the ditch and were placed on its base almost as soon as it was dug, since primary silt would have started to accumulate almost immediately (Bell *et al.* 1996). We postulate a functional relationship between the material to be dated (the antlers) and its context (the ditch cut).

Combining the radiocarbon results

The question then arises as to whether we can treat the date of these antlers as the date of the digging of the ditch, and so be justified in taking a weighted mean of the determinations (as replicate measurements on the same statistical population).

If we assume that all the antler tools are of exactly the same date, we can calculate a weighted mean of the measurements and use a χ^2 test to determine whether all the results are statistically consistent (Ward and Wilson 1978). If all nine determinations are taken together, they are not statistically significantly different at 95% confidence (4375±6BP; T'=15.0; T' for 5%=15.5; ν=8); although the seven high-precision measurements are significantly different at 95% confidence (4374±7BP; T'=14.6; T' for 5%=12.6; ν=6).

However we are not absolutely certain that all the antlers are of the same date; although they must have all been deposited at the same time because of the lack of primary silt beneath them, it is possible that antlers from several different growing seasons are represented. Consequently it is more appropriate to consider the spread of the radiocarbon measurements (Lyons 1991, 31 ff.). In this case the error on the weighted mean of all nine determinations becomes ±10BP, and that on the weighted mean of the seven high-precision measurements ±13BP. These errors, which may be regarded as an experimental measurement of how spread out the results are, are significantly larger than those calculated above, which are the theoretical errors that we expect on the basis of the accuracies of each measurement. This fact suggests that the results are not as consistent as would be expected if the samples were true replicates.

These figures suggest that the results from the antler tools cannot be considered to be from the same statistical population. Alternative explanations may be advanced for this.

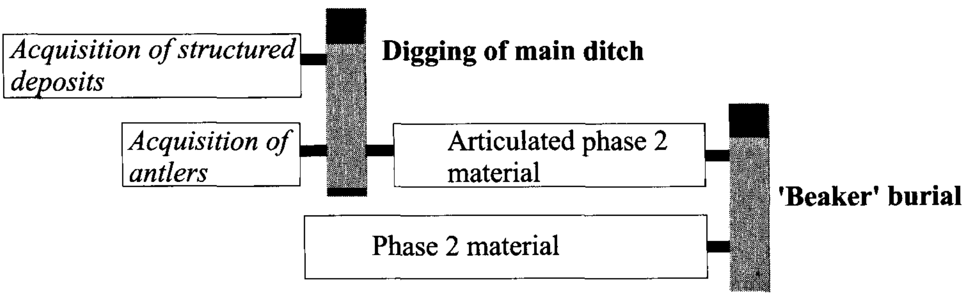

Figure 4. Summary of the chronological sequence of the principal phases and events in phases 1 and 2. The solid blocks represent events, the open blocks phases, and the horizontal lines stratigraphic relationships.

Table 1. Summary of reliable radiocarbon dates from Stonehenge

Context	*Material*	*Laboratory Reference*	*Radiocarbon Age (BP)*	*Calibrated date range (95% confidence)*
Mesolithic				
Post-pit WA9580	*Pinus* charcoal	OxA-4919	8520±80	7700–7420 cal BC
Post-pit WA9580	*Pinus* charcoal	OxA-4920	8400±100	7580–7090 cal BC
Post-pit WA9580	*Pinus* charcoal	GU-5109	8880±120	8090–7580 cal BC
Post-pit A	*Pinus* charcoal	HAR-455	9130±180	8820–7730 cal BC
Post-pit B	*Pinus* charcoal	HAR-456	8090±140	7480–6590 cal BC
Pre-phase 1				
Sarsen Circle	Animal bone	OxA-4902	5350±80	4360–3990 cal BC
Phase 1				
Ditch	Antler	UB-3787	4375±19	3085–2920 cal BC
Ditch	Antler	UB-3788	4381±18	3095–2920 cal BC
Ditch	Antler	UB-3789	4330±18	3030–2910 cal BC
Ditch	Antler	UB-3790	4367±18	3040–2915 cal BC
Ditch	Antler	UB-3792	4365±18	3040–2915 cal BC
Ditch	Antler	UB-3793	4393±18	3095–2920 cal BC
Ditch	Antler	UB-3794	4432±22	3305–2925 cal BC
Ditch	Antler	BM-1583	4410±60	3340–2910 cal BC
Ditch	Antler	BM-1617	4390±60	3330–2910 cal BC
Ditch	Animal bone	OxA-4833	4550±60	3500–3040 cal BC
Ditch	Animal bone	OxA-4834	4460±45	3350–2920 cal BC
Ditch	Animal bone	OxA-4835	4455±40	3340–2920 cal BC
Ditch	Animal bone	OxA-4842	4520±100	3510–2920 cal BC
Phase 1/2				
Aubrey Hole 32	Charcoal	C-602	3798±275	3020–1520 cal BC
Phase 2				
Ditch	Animal bone	OxA-4841	4295±60	3040–2700 cal BC
Ditch	Animal bone	OxA-4843	4315±60	3100–2700 cal BC
Ditch	Animal bone	OxA-4880	3875±55	2560–2140 cal BC
Ditch	Animal bone	OxA-4881	4300±60	3080–2700 cal BC
Ditch	Animal bone	OxA-4882	4270±65	3040–2660 cal BC
Ditch	Bone chisel	OxA-4883	4300±70	3100–2700 cal BC
Ditch	Antler	OxA-4904	4365±55	3300–2900 cal BC

Ditch	Antler	UB-3791	4397±18	3095–2920 cal BC
Ditch	Animal bone (articulated)	OxA-5981	4220±35	2920–2660 cal BC
Ditch	Animal bone (articulated)	OxA-5982	4405±30	3300–2920 cal BC
Phase 3				
Sarsen Circle	Antler	UB-3821	4023±21	2655–2485 cal BC
Sarsen Trilithon	Antler	OxA-4839	3860±40	2470–2200 cal BC
Sarsen Trilithon	Antler	OxA-4840	3985±45	2850–2400 cal BC
Sarsen Trilithon	Antler	BM-46	3670±150	2480–1680 cal BC
Bluestone Circle	Animal bone	OxA-4878	3740±40	2290–2030 cal BC
Bluestone Circle	Antler	OxA-4900	3865±50	2480–2140 cal BC
Bluestone Horseshoe	Antler	OxA-4877	3695±55	2280–1940 cal BC
Stone-hole E	Antler	OxA-4837	3995±60	2860–2350 cal BC
Stone-hole E	Antler	OxA-4838	3885±40	2490–2200 cal BC
Z Hole 29	Antler	OxA-4836	3540±45	2030–1740 cal BC
Y Hole 30	Antler	UB-3822	3341±22	1735–1530 cal BC
Y Hole 30	Antler	UB-3823	3300±19	1675–1520 cal BC
Y Hole 30	Antler	UB-3824	3449±24	1880–1690 cal BC
'Beaker' burial	Human bone	BM-1582		
'Beaker' burial	Human bone	OxA-4886		
'Beaker' burial	Human bone	OxA-5044	3817±27*	2460–2140 cal BC
'Beaker' burial	Human bone	OxA-5045		
'Beaker' burial	Human bone	OxA-5046		
Avenue				
Stonehenge terminal	Antler	OxA-4884	3935±50	2580–2300 cal BC
Stonehenge terminal	Antler	BM-1164	3678±68	2290–1890 cal BC
Nr Avon terminal	Animal bone	OxA-4905	3865±40	2470–2200 cal BC
N side of A344	Antler	HAR-2013	3720±70	2350–1930 cal BC
Post-monument				
Palisade Ditch	Human bone	UB-3820	2468±27	775–410 cal BC
Sarsen Circle	Bone point	OxA-4885	2840±60	1260–840 cal BC

* weighted mean of 3960±60BP, 3785±70BP, 3825±60 BP, 3775±55 BP, and 3715±70 BP.

1 the antlers are of different dates;

2 the concentration of radiocarbon in the samples is not the same as the concentration of radiocarbon in the atmosphere when the deer died (e.g. following exchange in the burial environment);

3 the errors on the radiocarbon measurements have been estimated incorrectly.

The first of these options appears to be the most likely, although, if it is true, it means that a very subtle difference in calendar date has been detected using radiocarbon. The period of collection of the antlers must be quite short because antler tools can only be used for a limited time to dig a ditch in chalk before they wear out, and because antler becomes brittle if curated for many years before use (Cleal *et al.* 1995, 414–25).

The second explanation is also possible. Bone chemistry and diagenesis are not fully understood (e.g. Gillespie 1989; Hedges and Millard 1995; Sobel and Berger 1995). However the samples involved are extremely large—approximately 1000 g of antler producing 15 g of benzene—so the effect of such problems should not be significant, especially as collagen preservation was relatively good. In addition the burial environment of all the material was very similar, and so, even if a difference were to be detectable, we would expect this option to affect the accuracy, rather than the consistency, of the results.

The final option is not considered very likely because of the rigorous and extensive quality assurance programmes routinely undertaken by the laboratories concerned (Otlet *et al.* 1980; International Study Group 1982; Scott *et al.* 1990; Rozanski *et al.* 1992; Scott *et al.* forthcoming), and the specific quality assurance measurements undertaken concurrent with the Stonehenge programme itself (Allen and Bayliss 1995, 516–18; Bronk Ramsey and Bayliss forthcoming, table 2). In particular it should be noted that the high-precision results are inconsistent as replicate measurements on the same population. This is important because the errors on these measurements are estimates of *total* error (including indeterminate error), unlike the previous measurements from the British Museum laboratory where the errors were calculated on the basis of the counting statistics alone (Allen and Bayliss 1995, 519). An error multiplier (Stuiver 1982), which has been determined empirically from the reproducibility of dates on replicate samples of cellulose within the laboratory, is used to account for this indeterminate error. This estimate is normally distributed, although we have no evidence whether the indeterminate error of a particular sample is so distributed. However this concern is, by its very nature, not measurable.

For the reasons given above we have not taken a weighted mean of the measurements from the antlers from the base of the ditch at Stonehenge. In addition we have no positive evidence that they must all have been of exactly the same date. Instead we have chosen to regard the end of the phase of acquisition of all the material which was deposited in the ditch below the primary silt as the most realistic estimate of the date of the ditch digging. This approach is conservative, and provides an estimate for the date of digging the ditch of *3015–2935 cal BC (95% confidence)* (Fig. 5). Because this range is based

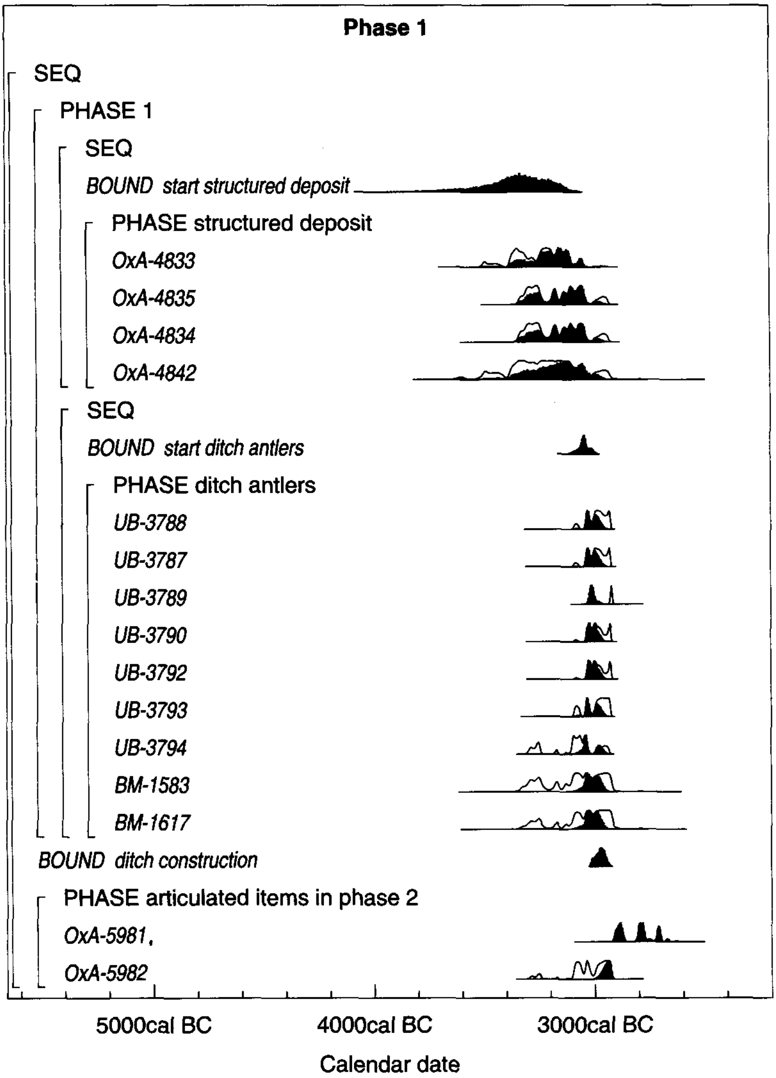

Figure 5. Probability distributions of dates from phase 1: each distribution represents the relative probability that an event occurs at some particular time. For each of the radiocarbon dates two distributions have been plotted, one in outline which is the result of simple radiocarbon calibration and a solid one which is based on the chronological model used; the 'event' associated with, for example, the radiocarbon date UB-3788 is the growth of the antler concerned. The other distributions correspond to aspects of the model: we have assumed that the material in the structured deposit started to be accumulated at some point defined as 'start structured deposit' and must have finished by the point at which the ditch was actually dug ('ditch construction'); the acquisition of the antlers also forms a similar sub-phase starting at 'start ditch antlers'; we also use the additional information that the articulated material from phase 2 must post-date the 'ditch construction' event. The large square brackets down the left hand side along with the OxCal keywords define the overall model exactly.

not only on the radiocarbon measurements but also on our chronological model which has changed with new radiocarbon determinations on articulated material found within the ditch fill (see Fig. 5 and Bronk Ramsey and Bayliss forthcoming), it is slightly different from that published in Cleal *et al.* 1995.

Precision, calibration, and accuracy

The precision of this estimate, with a range which covers only 80 years at 95% confidence, justifies the major effort which has been made towards quality assurance in the project to demonstrate accuracy of radiocarbon measurements against the relevant calibration data (Pearson *et al.* 1986; McCormac *et al.* 1995). The statistical consistency of the radiocarbon results with each other, and with the stratigraphic sequence, also supports their accuracy. Because the mathematical model of the site's dating includes all the radiocarbon results and archaeological evidence together, the statistical scatter of particular radiocarbon measurements (both from Stonehenge and from the calibration dataset) is counterbalanced by the overall picture, producing a stable and believable estimate of the site's chronology.

Taphonomy (curation)

In addition to the antler tools discussed above, there are also a number of large bones which were placed on the base of the ditch beneath the primary fill (Cleal *et al.* 1995, 422–5). These appear to be concentrated in the terminals of the ditch segments (Fig. 3).

A point of note emerged from the analysis of the measurements from these bones and from the antler tools, both of which were found on the base of the ditch. The placed bone deposits are significantly earlier than the digging of the ditch, although obviously they must have been deposited after it was dug. The mathematical model of the chronology of this phase is statistically significantly inconsistent if the structured deposits are constrained to be later than the ditch digging (Bronk Ramsey and Bayliss forthcoming, fig. 5). However the samples date to when the animal of which they were part died, not to their deposition, so this can be explained by the curation of material for some time before it was deposited. Analysis of the information currently available suggests that this was for between *70 and 420 years (95% confidence).* As well as raising a significant issue in prehistoric archaeology, this emphasises the importance of functionality in the interpretation of the taphonomic relationship between sample and context.

An alternative, more prosaic, interpretation of these results would be that there was some methodological reason for the measurements on these bones being so much earlier. There are essentially two major possibilities which could explain this. Firstly is that, since the measurements on the antler tools and on the placed bone deposits were measured at different laboratories, there is a laboratory offset between the radiocarbon measurements. However the inter-laboratory comparisons (Allen and Bayliss 1995, 516–18) show that this is not the case, certainly not to the extent necessary to explain the scale of the observed difference between the measurements on the placed bones and those on the antler tools. Secondly, it is possible that in either case the pretreatment of the samples may have been in some way insufficient. The close clustering of the large number results on the antler tools suggests that their dating is secure, which leaves us to question the validity of the bone dates.

Given that the measurements give an age which is older than the context, any error would have to be the result of radiocarbon-depleted contamination of some sort. Such an effect could, for example, be due to the incomplete removal of preservatives. There are, however, three internal checks which can be used to test for such an occurrence. First is collagen preservation, since poorly preserved collagen is much more difficult to purify reliably—of the dates in question (OxA-4833, OxA-4835, OxA-4834 and OxA-4842) three had very good collagen preservation (at about 20% of the level found in modern bone) while one (OxA-4842) was a low collagen bone with levels some twenty times lower. Secondly we can use the CN ratio to give us some measure of the purity of the collagen produced—the same three measurements give very similar values of 3.00–3.06 (atomic ratio) which is exactly what one would expect, while the poorly preserved bone (OxA-4842) gave a value of 3.47 indicating that the collagen was not so pure. Finally, the $\delta^{13}C$ measurements on the stable carbon isotopes are another indication both of environment and of sample purity—again the three well preserved bones give a consistent set of results in the range –22.4 to –23.1, suggesting a semi-woodland environment, while the poorly preserved bone gives an anomalous value of –23.8. In conclusion we can see that any pretreatment problems would be most likely to show up in sample OxA-4842, although this sample has been given a much larger uncertainty than the others because of this. Nevertheless the radiocarbon measurement still lies within the other three, providing us with an unintentional internal quality control check. It should also be noted that OxA-4842 is not the oldest of the placed bone deposits, OxA-4833, about which there can be little doubt, may well be even older. Although it is never possible to rule out the effect of some methodological offset entirely, this information taken as a whole does imply that the dating is secure and that the archaeological interpretation of the age difference is most likely to be correct.

Residuality and multiple sampling

To return to the main ditch, after the primary silt had accumulated, there seems to have been a period when a soil partially developed (Evans 1984, 10). Hawley noted the recurrent appearance of 'wood ash' in this layer, suggesting that burning was a feature of the activity in the ditch at this time. Unfortunately none of this material survives in the archive, so we were unable to date it.

Material from the secondary silts which accumulated above it does survive however. The sampling strategy for these silts was to submit a relatively large number of samples from throughout the profile. Bone samples were chosen in preference to antler because there is relatively little bone known from the primary silts (37 fragments). Therefore the bone fragments found higher up the profile are less likely to be residual from phase 1 than antler samples from the same position, since there are hundreds of antler fragments from the base of the ditch. Items which were uneroded and reasonably large were also selected to minimise the possibility of residuality. Unfortunately the provenance of some of the dated material was shown to be unreliable when additional archival material became available to the project team on the death of Professor Atkinson in October 1994. Four results have been excluded from the analysis for this reason (Allen and Bayliss 1995, 520–1).

The question of residuality is crucial because we know that the secondary silts must have accumulated after the ditch was dug and after the primary silting. If the material can be shown to be contemporary with the silting, then the analysis of the radiocarbon determinations can be constrained by the stratigraphic sequence. In fact if this is done, the model is statistically consistent (Bronk Ramsey and Bayliss forthcoming, fig. 8). Again we have chosen the conservative course however, and have been unwilling to make this assumption because of the lingering doubts over the contextual integrity and taphonomy of the dated material from phase 2 raised by the Atkinson archive.

Taphonomy (articulation)

Two further samples were submitted in October 1995, both from partially articulated skeletons within the secondary fill. One sample was from a piglet, and the other from cattle vertebrae. The crucial point is that both must have had tendons at least attached when they were buried, or they would not have been recovered together. This provides a strong argument against residuality or post-depositional disturbance. These two results have been used as an additional constraint for the estimate of the date of the ditch digging, to provide the model illustrated in Figure 5, and the estimated date for the construction of the ditch of *3015–2935 cal BC (95% confidence).*

Uniform distributions

This model treats the material from the base of the ditch in two groups, the animal bone deposits and the antlers. It is assumed that each group of dated material was gathered at a fairly constant rate over the period of collection (uniform phase; cf. Buck *et al.* 1992). In practice this assumption makes little difference to the estimated dates. The date of construction of the ditch is particularly robust, because it is so well constrained by the measurements on the articulated bone from phase 2 and by multiple high-precision measurements. The estimate for the start of the accumulation of the structured bone deposits is less robust, because there are only four measurements in this phase. However using the assumption of a uniform phase allows us to overcome the problem that the natural scatter of radiocarbon results would otherwise tend to give an unrealistically early estimate for this boundary.

Other distributions

The secondary silting of the ditch at Stonehenge is well constrained. It is preceded by the construction of the ditch which has been discussed above, and succeeded by the burial of an adult human male (Evans 1984) which cuts through the top of the secondary silting. The first dated event in these silts has been calculated using only the dates from the articulated items referred to above, which cannot be residual. The estimate of the last dated event has been calculated using all the reliable results from the secondary fills (Fig. 6). The difference between these distributions can be calculated, suggesting that the infilling of the ditch took between *460 and 740 years (95% confidence).*

Although this seems to be a reasonable estimate for these archaeological events, there are questions which are still to be answered. It would be particularly useful to be able to estimate the dates when the silting had reached different heights in the ditch. We have not modelled the rate of infill of the ditch, since it can be demonstrated that rates calculated for equivalent chalkland ditches are not uniform (Crabtree 1990; Evans 1990). Recently

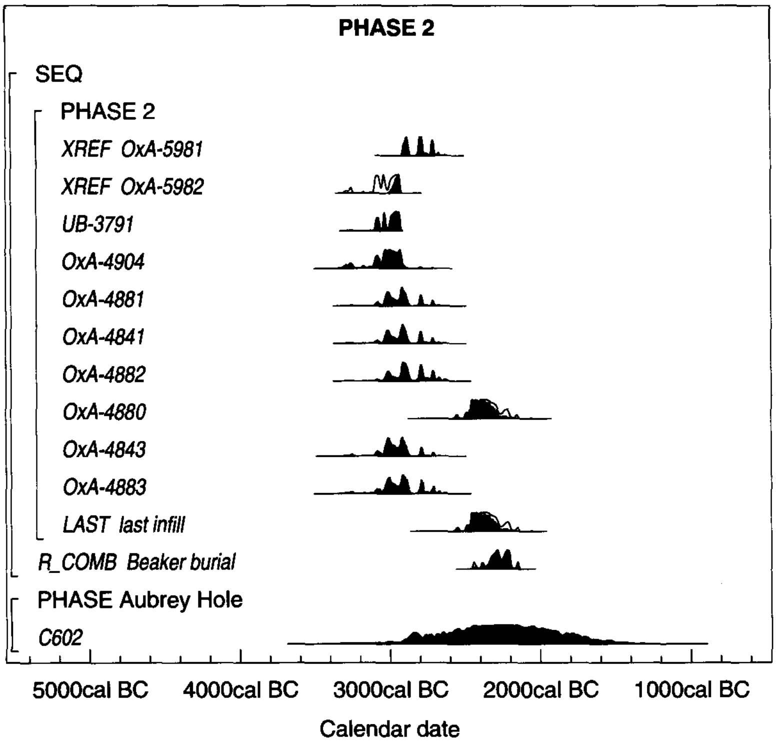

Figure 6 Probability distributions of dates from phase 2: the format is identical to that for Figure 5. In this phase the model imposed is simply that all of the material in the infill must pre-date the 'Beaker' burial. The two items labelled 'XREF' are also constrained within the model for phase 1 (see Fig. 5). The 'Beaker' burial has the keyword 'R_COMB' because it is based on the combination of several radiocarbon determinations.

work on pollen sequences has used several other possible assumptions about the distribution of dated events (Christian *et al.* 1995), although these have yet to be applied to the silting of ditches. An approach which modelled the rate of infilling using an exponential distribution could provide much better estimates for dating of this process. This would be particularly important on the chalk, where the major source of environmental evidence is from currently undated land snail sequences through the ditches of monuments.

Our interpretation of the chronology of phases 1 and 2

To summarise our results so far; the main ditch at Stonehenge was cut in *3015–2935 cal BC (95% confidence).* A number of bone deposits were placed on the bottom of the ditch immediately after construction, along with the antler tools used in the excavation. At least some of these bone deposits were already old when placed in the ditch, being between *70 and 420 years old (95% confidence)* on deposition. After a short period of between *0 and 75 years (95% confidence)* during which the primary silt and organic 'dark' layer accumulated, the secondary silt started to accumulate. This infilling took between *460 and 740 years (95% confidence)*, and was complete by the time a burial was inserted into the top of the secondary silts in *2400–2140 cal BC (95% confidence).*

Phase 3

In this phase the stone settings were constructed (Fig. 7). They seem to echo the timber settings of phase 2 and their pattern developed over many years, with one plan superseding the next. A complex and poorly understood sequence of erection of paired and single stones occurred within the ditched enclosure at this time and the Avenue was constructed.

The model

Unfortunately there was very little datable material available from the stone-holes of this phase. In total only 13 new samples could be measured, all from primary silts of stone-holes. A maximum of three items were measured from any stone setting and some settings had no suitable samples at all.

Although the centre of the monument was designed as a concentric setting of stones, allowing little stratigraphic overlap, excavation has nevertheless recovered a partial sequence. This has allowed us to place some of the results from the phase 3 settings into a relative order (Fig. 8).

Residuality and multiple sampling

The very small number of samples from each setting raises concerns about the reliability

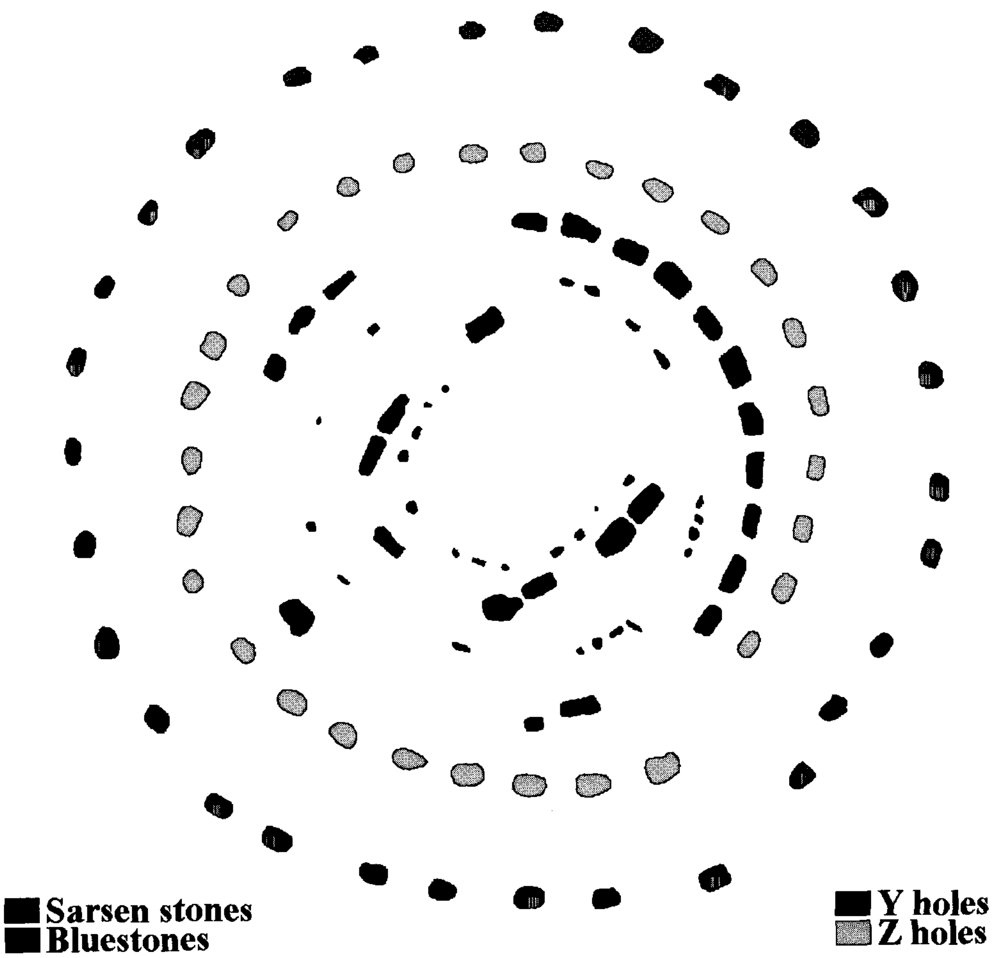

Figure 7. Plan of phase 3 at Stonehenge.

of our estimates. For example, we may be able to say that the results produced from the Bluestone Circle give an estimate of the last dated event of *2280–2030 cal BC (95% confidence)* (Fig. 9). However although we may be fairly confident in this estimate, it is not necessarily very reliable because it is based on only two dated items. If these happen to be residual (and we know that residual material is included in these primary silts, for example OxA-4902 (5350±80BP) from Stone-hole 27 of the Sarsen Circle), then our estimate will not be reliable.

Obviously the more dated items we have from a given setting the more reliable our estimate of the date of interest will be. Because the taphonomic relationship between the samples dated and the archaeological event which is of interest is not clear—there is no articulation or functional relationship demonstrable, then these measurements each provide

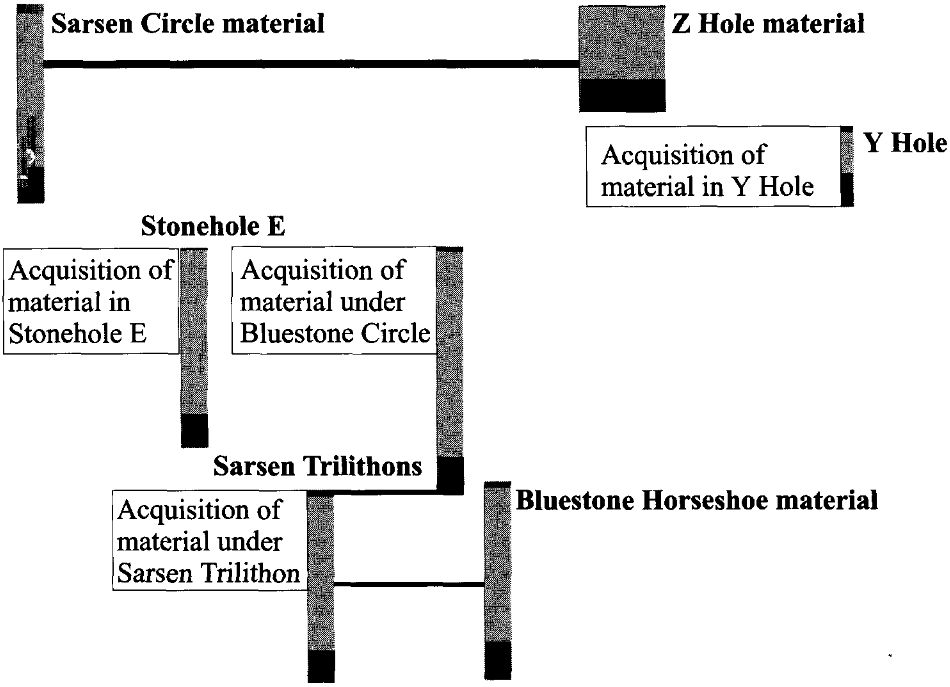

Figure 8. Summary of the chronological sequence of the principal phases and events in phase 3. The solid blocks represent events, the open blocks phases, and the horizontal lines stratigraphic relationships.

a *terminus post quem* for the construction of the setting. We hope that by submitting a number of samples from contexts relating to each event, the difference between the actual date of construction and the last *terminus post quem* becomes insignificant. Since this strategy is forced on us by the type of material available, the small number of items which can be dated is a significant problem.

Reliability vs confidence

Unfortunately the reliability of our estimates cannot be measured quantitatively; it is up to our archaeological judgement to decide how far we wish to trust the reliability of the estimates given the number of items it has been possible to date. In contrast, the level of confidence quoted is an expression of our confidence in the estimate given the data currently available and the model which we have described. This will change (and has already!).

Taphonomy (more curation)

The curation which has been demonstrated for the items placed on the base of the main ditch is also apparent later in the use of the monument. The pile of antlers which were

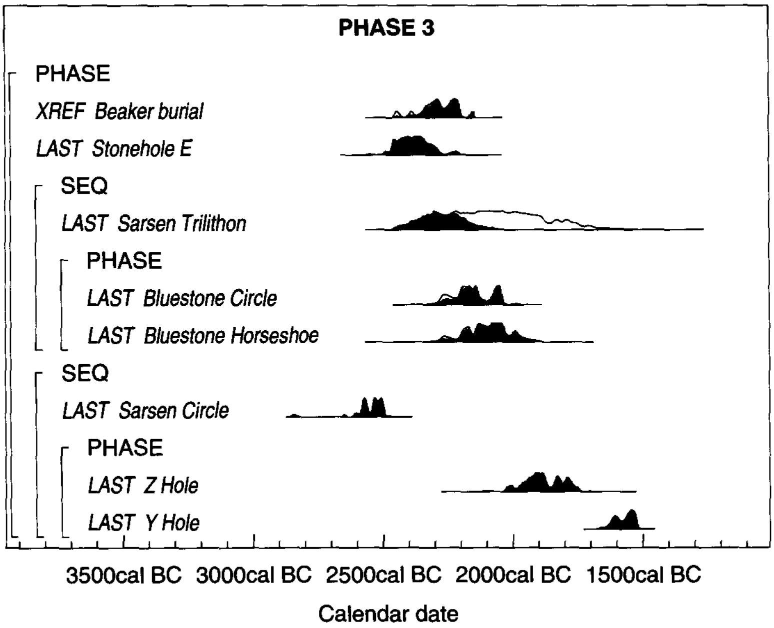

Figure 9. Probability distributions of dates from phase 3: the format is identical to that for Figure 5. Here we have made use of the stratigraphic information that the Sarsen Trilithon must pre-date the Bluestone settings and the Sarsen Circle must pre-date the Y and Z holes.

stacked on the base of Y Hole 30 produced radiocarbon determinations which are statistically significantly different (T'=24.1; T for 5%=6.0; ν=2; Ward and Wilson 1978). Although this is based on only three measurements—the other two antlers from the base of this Y Hole have been retained in the archive to allow further analyses in the future —the difference is so great in this case that there can be no question that the phenomenon observed is real, and that the antlers grew over a period of between *90 and 255 years (82% confidence) or 260 and 330 years (13% confidence)* before deposition.

It should be noted that these antlers are very different in character from those from the base of the main ditch (Cleal *et al.* 1995, 426) and have not been modified into tools. They also appear to have been gathered and placed in a pile on the bottom of the Y Hole, as 'they were entangled and difficult to remove' (Cleal *et al.* 1995, 260). Although it is not impossible that a residual antler could have been excavated and re-interred, the difference in character between these antlers and the others recovered from the monument, along with the apparently deliberate act of placing them on the base of the pit, makes this unlikely. It seems more probable that the deposition of these antlers was a significant activity.

LAST dated events

Where there is more than a single dated item from a stone setting, we have taken the estimate of the last dated event from a setting as the best estimate of its construction. This is on the principle that a context dates to the latest material within it (see above). The construction of each setting is assumed to be an event, with all the stones erected as a unitary whole. However this method of analysis may in fact suggest that the erection date is later than it actually was because of the inevitable statistical scatter of radiocarbon measurements. The small number of measurements available from any one setting probably counteracts the problem in this case, however, and so it is unlikely to be significant. We have not chosen to use the assumption of a uniform distribution for the dates of the items within the stone-holes because, as for the ditch filling, this is almost certainly not the case.

Distributions

In the future it may be better to attempt to model the distribution of the dates of residual material in a context, and so to provide an estimate for the end of the phase rather than of the last dated event. Again a uniform distribution may not be the most appropriate model, there being a much larger chance that material is residual by a few years than by many millennia. This is another area where more research is required, although again an exponential distribution may be a more appropriate way of modelling the process of the accumulation of material within a context.

Our interpretation of the chronology of phase 3

With all of this caution in mind, we estimate that it took between *850 and 1090 years (95% confidence)* for the stone monument at Stonehenge to reach its final form. The earliest stone settings, the Q and R holes, remain undated as do the earlier phases of the bluestone settings. However the Sarsen Circle was in place by *2620–2480 cal BC (92% confidence)*, and the Sarsen Trilithons by *2440–2100 cal BC (95% confidence)*. The remodelling of the bluestones into the circle and horseshoe occurred by *2280–2030 cal BC (95% confidence)* and *2270–1930 cal BC (95% confidence)* respectively. The last major modifications were completed by *1640–1520 cal BC (95% confidence)* when the Y holes, and possibly also the Z holes, were excavated.

Summary of issues raised by Stonehenge dating programme

The recent dating programme from Stonehenge has substantially enhanced our knowledge of the chronology of the monument. It has clearly demonstrated the potential of integrating archaeological and scientific information in a rigorous manner to produce precise and robust estimates of dates of archaeological interest.

The importance of the archaeological interpretation of the relationship between the items to be dated and the archaeological date which is to be estimated is crucial and cannot be overstated. At Stonehenge we have proposed very close links between samples and the events of interest, where articulated bone samples were recovered by excavation and where a functional relationship between sample and context can be determined. However residuality and the deposition of material which was already old when placed in the ground has also been demonstrated. Multiple sampling has been proposed as a method to address these problems, but this raises the question of how many items need to be dated before an estimate can be regarded as archaeologically reliable.

Although exhaustive efforts have been made to ensure that the radiocarbon measurements produced as part of the dating programme are accurate, their interpretation is still limited by problems which are not fully understood. These are areas of ongoing research, and we can do little more than recognise that they exist and attempt to err on the side of caution when interpreting the data.

Undoubtedly future research will also refine the mathematical model which has been proposed for the chronology of the site. In addition to methodological developments, such as the use of exponential distributions, future excavation may well recover more stratigraphic information and more material which can be dated. However, above all, the implementation of the ideas communicated to the Royal Society by the Revd. Thomas Bayes in the eighteenth century has allowed us to propose a model for the dating of Stonehenge which is analytical and interpretative. Other researchers may choose to take our data and reinterpret them against different hypotheses and within different conceptual frameworks, but Bayes' legacy has come to fruition.

Note on the calibration and citation of radiocarbon dates

Dates in Table 1 are cited in accordance with Mook (1986) and calibrated using the maximum intercept method of Stuiver and Reimer (1986). The date ranges cited *in italics* in the text, and the probability distributions shown in Figs 5, 6, and 9, have been calculated as part of the mathematical analysis presented in Allen and Bayliss (1995) and Bronk Ramsey and Bayliss (forthcoming). All calibrations use data from Stuiver and Pearson (1986), Pearson and Stuiver (1986), Pearson *et al.* (1986), and Kromer and Becker (1993). Further details of the methods of citation and calibration can be found in Cleal *et al.* (1995, 6).

References

ALLEN, M.J. and BAYLISS, A. 1995: Appendix 2: The radiocarbon dating programme. In Cleal, R.M.J., Walker, K.E. and Montague, R. 1995, 511–35.

BAYES, T.R. 1763: An essay towards solving a problem in the doctrine of chances. *Phil. Trans. Roy. Soc.* 53, 370–418.

BELL, M.G., FOWLER, P.J. and HILLSON, S. 1996: *The experimental earthwork project 1960–1992* (CBA Res. Rep. 100).

BRONK RAMSEY, C. 1995: Radiocarbon calibration and analysis of stratigraphy. *Radiocarbon* 37, 425–30.

BRONK RAMSEY, C. and BAYLISS, A. forthcoming: Dating Stonehenge. In Lockyer, K. and Mihăilescu-Bîrliba, V. (eds), Computer applications and quantitative methods in archaeology (BAR Int. Ser.).

BUCK, C.E., KENWORTHY, J.B., LITTON, C.D. and SMITH, A.F.M. 1991: Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65, 808–21.

BUCK, C.E., LITTON, C.D. and SMITH, A.F.M. 1992: Calibration of radiocarbon results pertaining to related archaeological events. *J. Archaeol. Sci.* 19, 497–512.

CHRISTEN, J.A., CLYMO, R.S. and LITTON, C.D. 1995: A Bayesian approach to the use of ^{14}C dates in the estimation of the age of peat. *Radiocarbon* 37, 431–42.

CLEAL, R.M.J., WALKER, K.E. and MONTAGUE, R. 1995: *Stonehenge in its Landscape: Twentieth Century Excavations* (London, Engl. Heritage Archaeol. Rep. 10)..

CRABTREE, K. 1990: Experimental earthworks in the United Kingdom. In Robinson, D.E. (ed.), *Experimentation and reconstruction in environmental archaeology*, 225–35.

EVANS, J.G. 1984: Stonehenge – the environment in the late Neolithic and early Bronze Age and a Beaker-age burial. *Wiltshire Archaeol. Natur. Hist. Soc. Mag.* 78, 7–30.

EVANS, J.G. 1990: Notes on some late Neolithic and Bronze Age events in long barrow ditches in southern and eastern England. *Proc. Prehist. Soc.* 56, 111–16.

GELFAND, A.E. and SMITH, A.F.M. 1990: Sampling based approaches to calculating marginal densities. *J. Amer. Stat. Assoc.* 85, 398–409.

GILLESPIE, R. 1989: Fundamentals of bone degradation chemistry: collagen is not "the way". *Radiocarbon* 31, 239–46.

HEDGES, R.E.M. and MILLARD, A.R. 1995: Bones and groundwater: towards the modelling of diagenetic processes. *J. Arch. Sci.* 22, 155–64.

INTERNATIONAL STUDY GROUP 1982: An inter-laboratory comparison of radiocarbon measurements in tree rings. *Nature* 298, 619–23.

KROMER, B. and BECKER, B. 1993: German oak and pine ^{14}C calibration, 7200–9439 BC. *Radiocarbon* 33, 125–35.

LYONS, L. 1991: *A practical guide to data analysis for physical science students* (Cambridge).

McCORMAC, F.G., BAILLIE, M.G.L., PILCHER, J.R. and KALIN, R.M. 1995: Location-dependent differences in the ^{14}C content of wood. *Radiocarbon* 37, 395–407.

MOOK, W.G. 1986: Business meeting: Recommendations/Resolutions adopted by the Twelfth International Radiocarbon Conference. *Radiocarbon* 28, 799.

OTLET, R.L., WALKER, A.J., HEWSON, A.D. and BURLEIGH, R. 1980: ^{14}C interlaboratory comparison in the UK: experiment design, preparation, and preliminary results. *Radiocarbon* 22, 936–46.

PEARSON, G.W. and STUIVER, M. 1986: High-precision calibration of the radiocarbon time scale, 500–2500 BC. *Radiocarbon* 28, 839–62.

PEARSON, G.W., PILCHER, J.R., BAILLIE, M.G.L., CORBETT, D.M. and QUA, F. 1986: High-precision ^{14}C measurement of Irish oaks to show the natural ^{14}C variations from AD 1840–5210 BC. *Radiocarbon* 28, 911–34.

ROZANSKI, K., STICHLER, W., GONFIANTINI, R., SCOTT, E.M., BEUKENS, R.P., KROMER, B. and VAN DER PLICHT, J. 1992: The IAEA ^{14}C intercomparison exercise 1990. *Radiocarbon* 34, 506–19.

SCOTT, E.M., HARKNESS, D.D., COOK, G.T., MILLER, B.F., BEGG, F.H. and HOLTON, L. forthcoming: The TIRI project: a status report. *Radiocarbon.*

SCOTT, E.M., LONG, A. and KRA, R.S. (eds) 1990: Proceedings of the international workshop on intercomparison of radiocarbon laboratories. *Radiocarbon* 32, 253–397.

SOBEL, H. and BERGER, R. 1995: Studies on selected proteins of bone in archaeology. *Radiocarbon* 37, 331–5.

STUIVER, M. 1982: A high-precision calibration of the AD radiocarbon time scale. *Radiocarbon* 24, 1–26.

STUIVER, M. and PEARSON, G.W. 1986: High-precision calibration of the radiocarbon time scale, AD 1950–500 BC. *Radiocarbon* 28, 805–38.

STUIVER, M. and REIMER, P.J. 1986: A computer program for radiocarbon age calculation. *Radiocarbon* 28, 1022–30.

WARD, G.K. and WILSON, S.R. 1978: Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20, 19–31.

ALEX BAYLISS, CHRISTOPHER BRONK RAMSEY and F. GERRY McCORMAC

Dating Stonehenge

As part of the recent research programme on the twentieth-century excavations at Stonehenge (Cleal *et al.* 1995), a series of nearly fifty new radiocarbon determinations was commissioned. A chronological model of the site has been developed which combines the evidence of the radiocarbon measurements with the stratigraphic sequences recovered during excavation. This has enabled much more precise estimates of dates of archaeological interest to be calculated.

A number of points of archaeological and scientific interest have been raised by this programme of work; in particular the importance and complexities of archaeological taphonomy are seen as crucial. Some of the choices which were encountered when building the model are also discussed. Above all this work is seen as both analytical and interpretative, and will inevitably be modified as more data become available, different questions are asked, and different interpretative frameworks adopted.

DAVE BATCHELOR

Mapping the Stonehenge World Heritage Site

This paper describes the work of the Central Archaeology Service in creating an integrated and dynamic database that encompasses geographic and textual data from a number of disparate sources. It will concentrate on the physical and cultural landscape that surrounds Stonehenge rather than the monument itself.

A. DAVID and A. PAYNE

Geophysical surveys within the Stonehenge landscape: a review of past endeavour and future potential

The techniques of archaeological geophysics now have a very widespread currency in British archaeology. Those most commonly in use, magnetometry and resistivity surveying, can be particularly effective for the mapping of the buried outlines of domestic, industrial and funerary sites from later prehistory until the present day. Given the pre-eminent reputation of Stonehenge and its surroundings it is perhaps surprising that such techniques have not been used more exhaustively to explore the area for hidden detail. However, in recent years, fuelled both by research initiatives and the modern pressures now affecting this World Heritage Site, geophysical survey has indeed been applied with increasing determination. This paper provides an overview of this recent work, both at Stonehenge itself and at neighbouring sites, and will confront both its present limitations as well as its future potential.