University of Groningen

Developments in radiocarbon calibration for archaeology

Ramsey, Christopher Bronk; Buck, Caitlin E.; Manning, Sturt W.; Reimer, Paula; van der Plicht, Hans

Published in:
Quaternary Science Reviews

IMPORTANT NOTE: You are advised to consult the publisher's version (publisher's PDF) if you wish to cite from it. Please check the document version below.

Document Version
Publisher's PDF, also known as Version of record

Publication date:
2006

Link to publication in University of Groningen/UMCG research database

Citation for published version (APA):
Ramsey, C. B., Buck, C. E., Manning, S. W., Reimer, P., & van der Plicht, H. (2006). Developments in radiocarbon calibration for archaeology. *Quaternary Science Reviews*, *25*(5), 783-798.

Copyright
Other than for strictly personal use, it is not permitted to download or to forward/distribute the text or part of it without the consent of the author(s) and/or copyright holder(s), unless the work is under an open content license (like Creative Commons).

The publication may also be distributed here under the terms of Article 25fa of the Dutch Copyright Act, indicated by the “Taverne” license. More information can be found on the University of Groningen website: https://www.rug.nl/library/open-access/self-archiving-pure/taverne-amendment.

Take-down policy
If you believe that this document breaches copyright please contact us providing details, and we will remove access to the work immediately and investigate your claim.

Downloaded from the University of Groningen/UMCG research database (Pure): http://www.rug.nl/research/portal. For technical reasons the number of authors shown on this cover page is limited to 10 maximum.

Download date: 13-02-2023

Developments in radiocarbon calibration for archaeology

Christopher Bronk Ramsey[1], Caitlin E. Buck[2], Sturt W. Manning[3], Paula Reimer[4] & Hans van der Plicht[5]

This update on radiocarbon calibration results from the 19th International Radiocarbon Conference at Oxford in April 2006, and is essential reading for all archaeologists. The way radiocarbon dates and absolute dates relate to each other differs in three periods: back to 12 400 cal BP, radiocarbon dates can be calibrated with tree rings, and the calibration curve in this form should soon extend back to 18 000 cal BP. Between 12 400 and 26 000 cal BP, the calibration curves are based on marine records, and thus are only a best estimate of atmospheric concentrations. Beyond 26 000 cal BP, dates have to be based on comparison (rather than calibration) with a variety of records. Radical variations are thus possible in this period, a highly significant caveat for the dating of middle and lower Paleolithic art, artefacts and animal and human remains.

Keywords: Dating, radiocarbon, calibration, varves, ice-cores, speleothems

Introduction

Radiocarbon dating underpins most of the chronologies used in archaeology for the last 50 000 years. However, it is universally acknowledged that the radiocarbon 'ages' themselves (usually expressed in terms of ^{14}C years BP – because they are measured relative to the standard which corresponds to AD 1950) are not an accurate reflection of the true age (in calendar years) of samples, because the proportion of radiocarbon in the atmosphere has fluctuated in the past and because the half-life used for the calculation of radiocarbon ages is not correct. For this reason, where possible, radiocarbon dates are calibrated against material of known age (giving ages expressed in terms of cal AD, cal BC or cal BP – which is absolute relative to AD 1950). For recent periods (in practice, the Holocene) this is now standard practice amongst archaeologists. However, as we seek to extend the timescale over which calibration is possible, it is important to be aware of the diverse nature of calibration datasets and the limits to their reliability. It is also worth considering some of the reasons behind the controversy over the term 'calibration' (van Andel 2005).

[1] *Research Laboratory for Archaeology and the History of Art, University of Oxford, UK*

[2] *Department of Probability and Statistics, University of Sheffield, UK*

[3] *Department of Classics and The Malcolm and Carolyn Wiener Laboratory for Aegean and Near Eastern Dendrochronology, Cornell University, USA; School of Human and Environmental Sciences, University of Reading, UK*

[4] *14CHRONO Centre for Climate, the Environment and Chronology, Queen's University Belfast, Belfast, Northern Ireland*

[5] *Centre for Isotope Research, Rijksuniversiteit Groningen, Netherlands; Faculty of Archaeology, Leiden University, Netherlands*

Received: 28 June 2006; Accepted: 6 September 2006; Revised: 12 September 2006

ANTIQUITY 80 (2006): 783–798

Data for radiocarbon calibration

Until recently the main data that have been employed to generate the estimates of the radiocarbon calibration curve have been measurements of the radiocarbon concentration of wood which has been dendro-chronologically dated to the nearest year. This is ideal from the point of view of archaeologists since the wood in trees is laid down with carbon taken from the atmosphere. The same can be said for most plant fragments, and, through the food chain, for terrestrial animals. So, for the vast majority of archaeological material, the carbon in the samples should have a radiocarbon concentration very close to that of the tree rings used to generate the calibration curve. Only when there are samples from marine or fluvial environments, or other unusual situations (for example depleted $^{14}CO_2$ from volcanic sources, or significant oceanic upwelling in some coastal situations) do we have to worry about reservoirs of carbon with radiocarbon concentrations that are substantially different to those in the atmosphere.

Over the last couple of decades the extent of tree ring records available has been greatly expanded. In 1986 the firmly dated sections of the calibration curve extended back to about 7300 cal BP (Stuiver 1986), although floating sections could be used to infer its form back over the full extent of the Holocene. When the IntCal04 calibration curve (Reimer *et al.* 2004) was constructed the tree ring data extended back to about 12 400 cal BP. This record is in most places duplicated many times over, both in terms of the dendro-chronology and with dates measured at a number of different high-precision laboratories. This lends great strength to our conviction that, within the uncertainty quoted on IntCal04, the tree ring section of the IntCal04 curve closely represents a true record for the atmosphere of the mid-latitude Northern Hemisphere (see Figures 1 & 2). The 2004 estimate of the calibration curve for the past 1000 years from the Southern Hemisphere, which has a slightly different radiocarbon concentration (this difference equates to no more than *c*. 100 ^{14}C years in this time period), is also available in the form of the SHCal04 curve (McCormac *et al.* 2004) (see also Figure 2). Furthermore, more data are always being added to this corpus and floating sections of wood from Germany now extend well back into the late glacial. This will almost certainly allow us to extend the terrestrial calibration curve back further in time. Equally interesting is the fact that *kauri* trees from New Zealand are found with ages that extend right out beyond the range of radiocarbon and are currently being dated in the age range 25-55 000 BP (oral presentation by Chris Turney at the nineteenth International Radiocarbon Conference, Oxford). These do not, and perhaps never will, provide a continuous chronology that can be linked together to provide a chronology like the one we have for the Holocene. However, it is likely to give us insight into the way in which the radiocarbon concentration in the atmosphere fluctuated in the past.

In order to calibrate samples older than the extant tree-ring-based calibration curve, we need to make use of different kinds of records, and this is where things become more complicated (see Table 1). The reasons for these complications are obvious. Ideal calibration relates measurements of atmospheric radiocarbon (^{14}C years BP) to the absolute calendar timescale, and according to the strict definition, only the dendro-chronological record qualifies for this. Beyond the tree ring data, most radiocarbon samples in 'known-age' records are derived from non-terrestrial reservoirs, such as marine deposits and speleothems

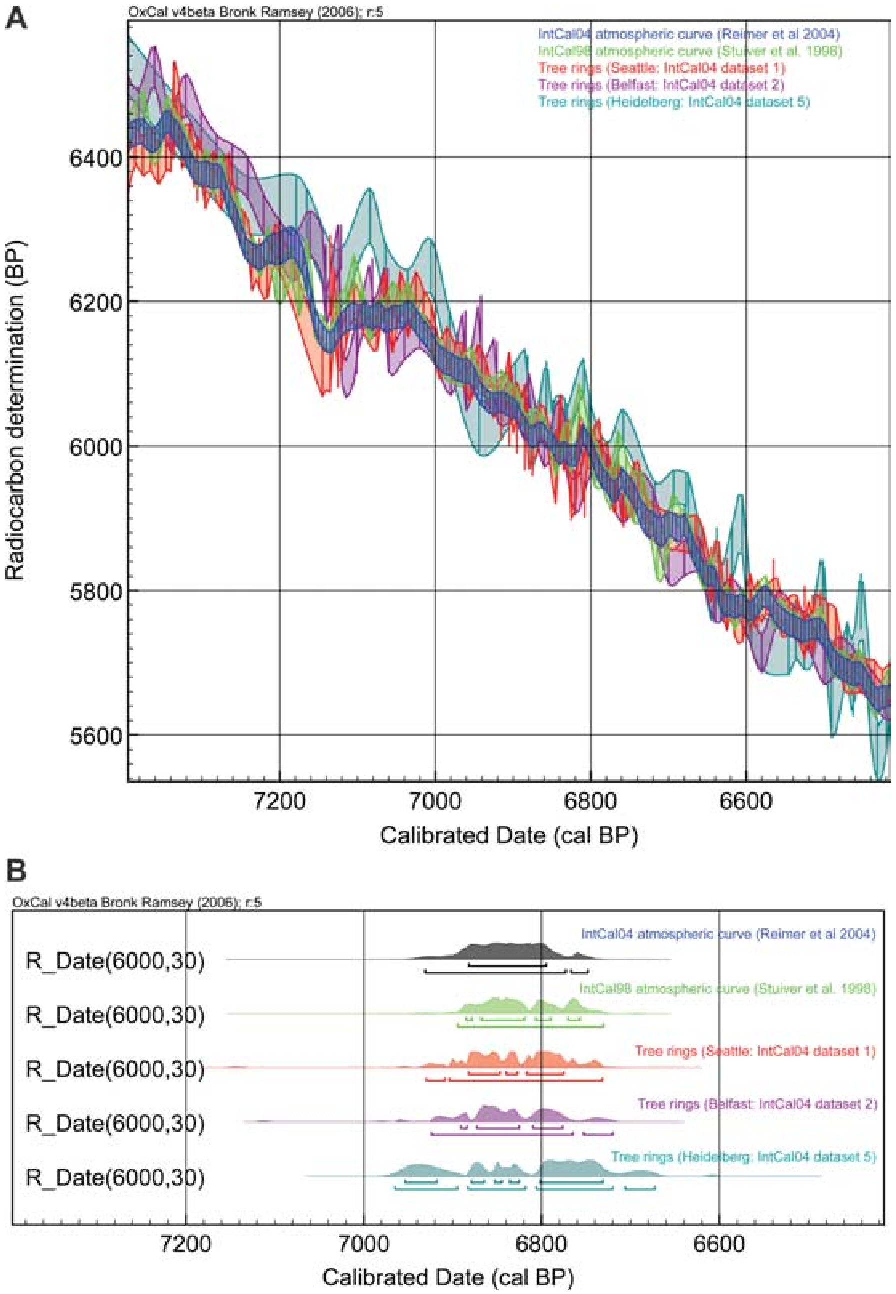

Figure 1. Calibration in the Holocene: the main calibration datasets for one section of this period are shown in panel A, overlain by the IntCal04 (blue) and IntCal98 (green) calibration curves. All curves are shown as a 1σ envelope with cubic interpolation at five year intervals between individual points as normally applied in OxCal (Bronk Ramsey 2001 and associated manual). In panel B you can see the distribution and ranges resulting from calibration of the date 6000 ± 30 BP against IntCal04 (grey distribution) or comparison to IntCal98 (green) and the other individual datasets; the concordance between the different datasets is clear. Both panels show outputs from the new version (v.4) of the OxCal software announced at the Oxford Conference by Bronk Ramsey and currently completing beta testing (June 2006) – this version incorporates a number of refinements and additions of relevance to both archaeologists and environmental scientists.

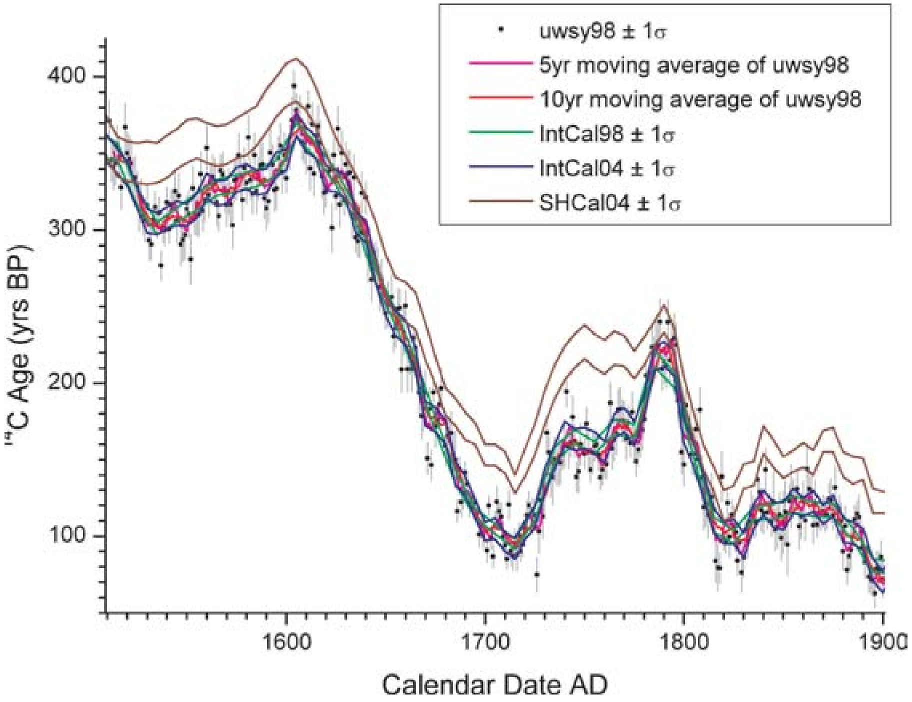

Figure 2. Comparison AD 1510-1900 of (i) single year northern hemisphere ^{14}C data (black circles with 1σ errors shown by the grey bars; uwsy98 dataset from Stuiver, Reimer & Braziunas 1998), versus (ii) a moving 5-year average of the single-year data of (i) in magenta, versus (iii) a moving 10-year average of the single-year data of (i) in red, versus (iv) the IntCal98 northern hemisphere calibration curve (Stuiver, Reimer, Bard et al. *1998) shown as a 1σ envelope in green, versus (v) the IntCal04 northern hemisphere calibration curve (Reimer* et al. *2004) shown as a 1σ envelope in blue, versus (vi) the SHCal04 southern hemisphere calibration curve (McCormac* et al. *2004) shown as a 1σ envelope in brown. We may note: (a) that annual variability/noise in the (non-replicated) single-year data lies closely around the longer-term 5-year or 10-year trend of the standard much-replicated calibration curves, with the 10-year moving average of the single-year data (red line) conforming very closely to the IntCal04 1σ envelope (blue lines); (b) the similarity of the IntCal98 and IntCal04 curves; (c) that IntCal04 picks up the short-term variability slightly better compared to IntCal98 where the underlying data have a resolution finer than the 5-year interpolated points of the IntCal04 curve – as the case AD 1511 and later where we have underlying 1-year resolution.*

(mineral cave deposits), and are therefore subject to reservoir effects. The 'known-ages' in these records also depend on deposition models and measurement errors. All of these issues lead to varying degrees of uncertainty, depending on the nature of the dataset, as discussed below (see also the list of 'pros and cons' given by van der Plicht *et al.* 2004).

First of all we have the different kinds of sample that can be used for measurements. The main samples that have been used for this kind of study are: wood, plant remains, foraminifera, corals and speleothems. The first two of these reflect atmospheric radiocarbon concentration and so are potentially ideal for calibration purposes. However foraminifera and corals are marine organisms, and so reflect the radiocarbon concentration in particular regions of the ocean. We know the radiocarbon concentration of the surface oceans today, but there is increasing evidence that the difference between the oceans and the atmosphere has varied (and perhaps very considerably if we look at the late glacial and earlier periods).

Table 1. Summary of different calibration records showing the sample types and the methods used to assess independently the (calendar) ages; the examples given are not intended to be an exhaustive list

	Sample material				
Independent dating method	**Wood (terrestrial)**	**Plant fragments (terrestrial; assumed young on deposition)**	**Corals (surface ocean)**	**Foraminifera (oceanic; depth depends on species)**	**Speleothems tufas, etc. (mixed terrestrial and geological carbon)**
Tree rings (accurate to the year)	Tree ring records (see main records in Reimer *et al.* 2004)				
Uranium series (quality depends on samples)			Coral records (e.g. Bard *et al.* 1998; Chiu *et al.* 2006; Cutler *et al.* 2004; Fairbanks *et al.* 2005)		Speleothems and Tufa records (e.g. Beck *et al.* 2001; Stein *et al.* 2004; Vogel & Kronfeld 1997)
Ice cores (subject to modelling or counting errors)				Ocean sediment records (e.g. Bard *et al.* 2004; Hughen *et al.* 2004b)	
Varved sediments (susceptible to missing varves and counting errors)		Varved lake records (e.g. Kitagawa & van der Plicht 1998)		Varved ocean sediments (Hughen *et al.* 2004b)	

This should not surprise us since one of the main phenomena of the glacial fluctuations in climate is major change in ocean circulation (Dansgaard *et al.* 1993). Speleothem records are even more complex: they contain a mixture of carbon from the atmosphere and from ground water, which is likely to have a component of carbon from geological deposits that are essentially free of radiocarbon.

Secondly, we have different methods of estimating the true age of the samples that are to be used for calibration. In the Holocene we have the luxury of tree ring dates that are accurate usually to the exact year. We do not have this in earlier periods and so we must use other methods, the main ones being varve counting, ice-core timescales and uranium series dating. Varve counting of lakes (such as Lake Suigetsu, Japan) is susceptible to error for a number of reasons – although such sequences do usually provide a fairly good relative chronology. Ice-core timescales are either based on direct counting of ice layers (as in the case of the GISP2 chronology and the new NGRIP chronology back to *c.* 40 000 BP) or based on age/depth models (as in the case of GRIP and GISP2 beyond 40 000 BP). In principle, these records suffer some of the same problems as varved lakes (for one discussion of problems in the chronology of the well-known GISP2 ice core, see Southon 2004) but due to the concentration of effort in these records and the degree of duplication they are, at their best, considerably better than varves (presentation of J.P. Steffensen at the Oxford Radiocarbon Conference). They also have the benefit of being the timescale against which much palaeoclimate data are generated, and so, even if the absolute ages are not correct, the relationships to these data will be. However, one further complication is that in order to use these timescales it is necessary to make assumptions about the synchronicity of global climate signals that may not be fully justified. Finally, we have uranium series dating, in this case either of corals or speleothems. This is a very precise and accurate technique if correctly applied. However, it does require very careful analysis to ensure that the samples dated have not suffered from detrital contamination or post-depositional re-crystallisation. These caveats aside, the timescale derived is independent and so provides a very useful method for radiocarbon calibration, when proven absolute (Chiu *et al.* 2006).

So we can see that all of the records we might use for calibration of earlier timescales do have their problems – often complicated and often interwoven. There is some strength in the diversity of the methods employed and this is why for the IntCal04 calibration curve some of these records were used to extend the calibration curve back to 26 000 cal BP on the basis that there was sufficiently good agreement between the different datasets (see Figure 3). However, it should be stressed that beyond the tree ring data this curve is essentially based on marine data and therefore relies on assumptions about the relationship between the radiocarbon concentration of the oceans and the atmosphere. Thus, this part of the atmospheric calibration curve is 'marine derived'. Further back in time the records, in part because of the various problems outlined above, showed poor agreement when IntCal04 was compiled (see Figure 4). Research in this area is, however, very active and the situation is changing rapidly. Research programmes and investigations in different areas are bringing the marine calibration datasets into much closer agreement. For example, the Cariaco basin data (Hughen *et al.* 2004a; Hughen *et al.* 1998; Hughen *et al.* 2004c), for which the initial calendar ages were based on the GISP2 timescale, agrees much better with the coral data if either the new NGRIP chronology is used or the chronology from Hulu Cave (Wang *et al.*

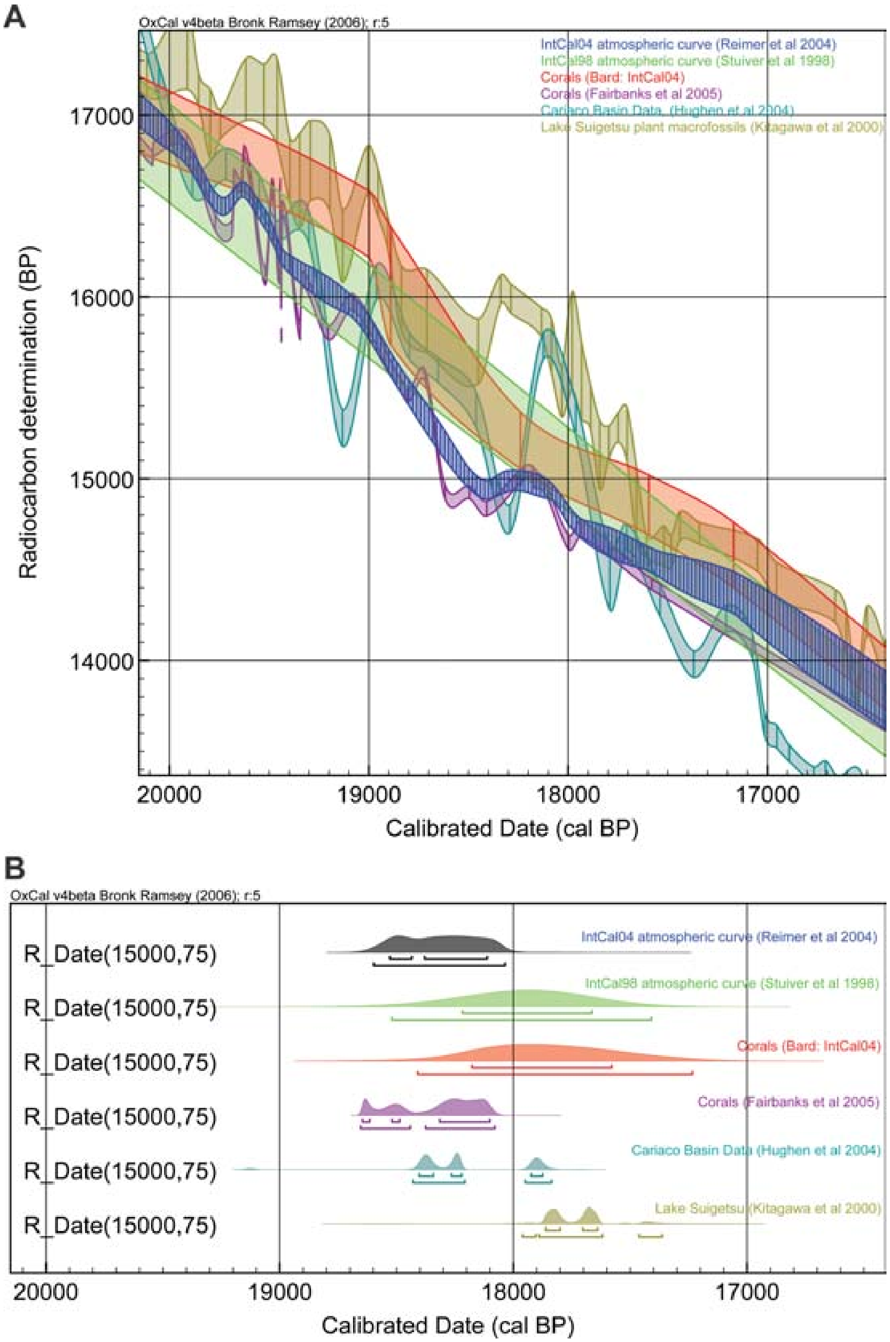

Figure 3. Calibration at around 15 000 ^{14}C years BP. Panel A shows the main calibration datasets for this period and interpolated as for Figure 1A. The plots are overlain by the IntCal04 curve (blue) and the IntCal98 curve (green). In panel B you can see the distribution and ranges resulting from calibration of the date 15 000 ± 75 BP against IntCal04 (grey distribution) or comparison to IntCal98 (green) and the other individual datasets. There is reasonable concordance with the marine datasets although the Lake Suigetsu data (which is terrestrial and not used in IntCal) suggests slightly younger ages. This may be due to the uncertainties in the Suigetsu timescale or because of slightly larger than expected oceanic reservoir effects.

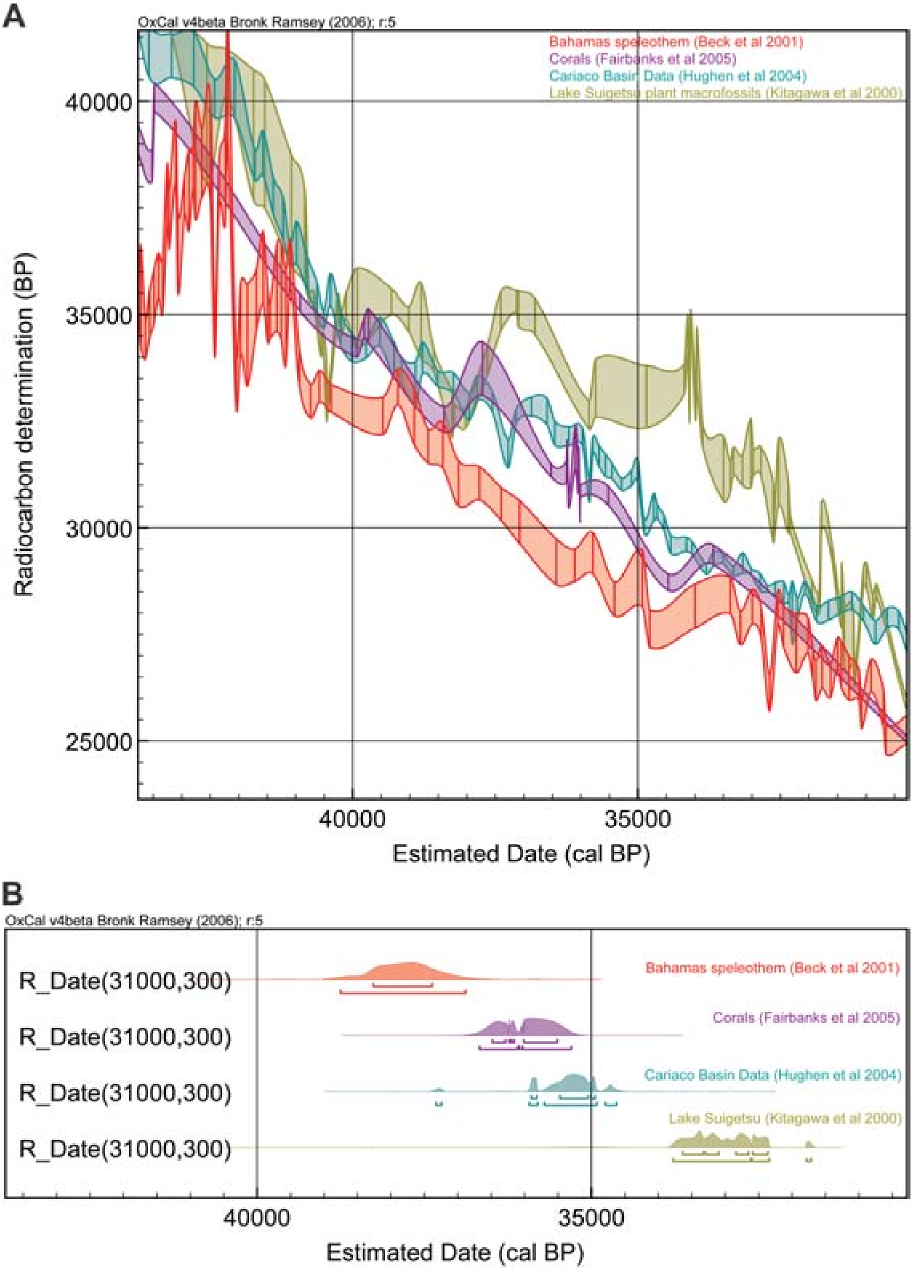

Figure 4. Comparison against datasets at around 31 000 ^{14}C years BP (approximate radiocarbon age of samples from Chauvet: Valladas et al. *2001). Panel A shows the main radiocarbon datasets for this period and interpolated as for Figure 1A. In panel B you can see the distribution and ranges resulting from comparison of the date 31 000 ± 300 BP to the individual datasets. There is virtually no overlap between the results of any of the analyses. The concordance between the two marine datasets plotted here is better although even this becomes significantly worse earlier than 38 000 cal BP, probably largely because of the problems with the GISP2 timescale. The terrestrial records, which are more directly applicable to archaeological samples, show very substantial offsets. Thus, at present, considerable caution is appropriate when estimating any calendar age range for examples like this.*

2001). Other records are also being revised as new data and methods become available (such as that of Beck *et al.* 2001) and it looks as if it will not be long before a marine calibration curve can be constructed for the last 40 000 (or even 50 000-55 000) years – as evident in presentations by both Konrad Hughen and Richard Fairbanks at the Oxford Radiocarbon Conference.

However, other discrepancies remain. These probably arise from three major factors:

- Increasing uncertainty in the calendar age estimates for the samples undergoing radiocarbon dating. Ice-core timescales become increasingly uncertain with increasing age because of thinning of the annual layers and concatenation of errors through the record. Correlation with the oxygen isotope records also becomes more complicated in some periods. Uranium series dates are in principle still very precise over this time range but there is increasing chance of post-depositional change and complications of changing (or unknown) environmental conditions.
- Increasing difficulty in measuring the radiocarbon concentration of the samples accurately, especially as the records get back before 30 000 ^{14}C years BP, where the corrections for modern contamination in processing and more recent environmental contamination in the samples are issues which can be difficult to resolve fully (at this age only about 2 per cent of the radiocarbon remains in the sample and even low levels of contamination become significant).
- Increasing difficulty in assessing the state of the global carbon cycle, including particularly the ocean circulation, deep ocean ventilation and the radiocarbon production rate in these periods.

Of greatest significance are indications in some of the terrestrial (but not atmospheric) records (such as the Bahamas speleothem; Beck *et al.* 2001) that there may be some considerable offsets between the atmosphere and the oceans at particular periods and possibly major age inversions at or just before 40 000 cal BP which may be related to major geomagnetic excursions such as the Laschamp event. If this is the case, then caution will still be needed in using marine records for the calibration of terrestrial samples.

What is calibration

Much debate centres on the use of the word calibration. There are of course many uses of the word 'calibrate' in the English language, but the sense in which it is most often used in science is 'to set an instrument so that readings taken from it are absolute rather than relative' (Simpson & Weiner 1989). The mathematical methods employed by radiocarbon calibration programs such as BCal (Buck *et al.* 1999), CALIB (Stuiver & Reimer 1993), CalPal (Joris & Weninger 1998), the Groningen radiocarbon calibration program (WinCal25/Cal25; van der Plicht 1993), or OxCal (Bronk Ramsey 2001) are essentially methods for mapping radiocarbon ages and their associated laboratory uncertainties through a mathematical function with its own uncertainty (often known as a calibration curve) onto the calendar scale. It is the view of many in the radiocarbon community that this mapping process should really only be called 'calibration' if the mathematical function or calibration curve we use is derived in such a way that we can be fairly sure that by using it we are putting our samples (with a known degree of accuracy) onto an absolute timescale.

The reason for this caution is essentially in order to prevent too much confusion in the disciplines served by radiocarbon dating. Archaeology has suffered too much over the last five decades from 'radiocarbon revolutions' without having to experience further ones every time a new 'calibration' record emerges. For this reason it seems sensible to base our estimates of calibration curves solely on data that are well corroborated and to avoid data which (although potentially useful for other purposes) are currently seen as provisional for calibration purposes. In this respect it would also seem sensible to draw a semantic distinction between 'calibration' as such and 'comparison' of radiocarbon dates to particular records. The same kinds of mathematical method can be used to undertake both 'calibration' and 'comparison' and the data are almost always made freely available by the scientific community, so there is no question of curtailing freedom as suggested by van Andel (2005). We simply urge everyone to make it clear whether they are undertaking true calibration or a comparison and draw their readers' attention to the difference between the two.

There is an argument that 'calibration' need not be very precise and that even a rough calibration may be useful. This is certainly true. However, if the calibration is to be useful it must have a statement of uncertainty attached to it and this must accurately reflect the true uncertainty in the absolute age estimate generated. Herein lies a problem. Each group of researchers who provide data with potential utility for radiocarbon calibration curve estimation do their best to quantify their own internal sources of error and uncertainty and to report these in a standard form. What they do not and cannot do is to allow for sources of error or uncertainty that they are completely unaware of. If we look at the currently available data for the pre-tree-ring timescale we find that there are substantial uncertainties that have simply not been quantified. Buck and Blackwell (2004) provide a statistical method to estimate the scale of unquantified uncertainties that must be present if all of the records they considered relate to the same underlying radiocarbon calibration record and found offsets as large as 2500 years (van der Plicht *et al.* 2004). Given this (and other observations about the data), the IntCal group felt that they could not provide a reliable estimate of the radiocarbon calibration curve beyond 26 000 cal BP in 2004.

In the absence of an internationally agreed calibration curve beyond 26 000 cal BP, it is natural for researchers to compare one record to another (exactly as the IntCal team did). In doing this, however, it is wise to avoid use of the term 'calibration' since this does suggest an absolute scale, and instead use alternatives, for example 'comparison' as proposed previously (Richards & Beck 2001; van der Plicht 2000; van der Plicht *et al.* 2004).

Implications for archaeologists

So how should archaeologists treat the data that are currently available? The data are there to be used and studied and no-one wishes to stifle speculation about what those data mean for very important archaeological issues. Indeed, the calendar timescale created by radiocarbon largely shapes a number of questions and debates in the later Palaeolithic period. It is thus not realistic to assume that those working in the area will wait until the research is complete before starting to look at such issues (as for example in Mellars 2006 and the discussion with

Turney *et al.* 2006). However, it is important that the archaeological community is aware of the different nature of the radiocarbon records.

Back to around 12 400 cal BP, the period for which we have multiple records that are in good agreement, including tree rings, it seems very likely that the calibration curve will not change significantly as new data come to light and calibration can in most cases be used as a tool in studying archaeological chronology even in a fairly fine-grained manner (see Figure 1). This period of relative certainty is likely to reach back to about 18 000 cal BP once the new work extending the tree ring record reported by Mike Friedrich at the Oxford Radiocarbon Conference is (eventually) completed. In this time period there are some minor issues that are still to be sorted out for very high precision work. These centre on how the different calibration sets are compiled into a single curve. Such a compilation is undoubtedly the best policy since it ensures that no one dataset, with its inevitable possible faults, is given too much weight. All of the indications are that within any one hemisphere there are no significant regional effects although some very minor differences between records have been attributed to differences in growth seasons (Kromer *et al.* 2001) or proximity to ocean upwelling regions (Stuiver & Braziunas 1998). Probably more significant is the fact that most of the calibration data are measured on ten- or twenty-year sections of wood and therefore average out shorter-term to annual variations (see Figure 2 – this visible noise will usually in effect cancel itself out over even a few years and especially within the range of many typical radiocarbon measurements and their associated errors – minor exceptions may occur at times of major peaks or troughs in the radiocarbon record – e.g. AD 1788-92 – but it should also be remembered that this single-year record is not replicated and clearly contains substantial noise as well as signal). There are also questions over what the best statistical methods are for combining the datasets; the IntCal04 curve (Buck & Blackwell 2004; Reimer *et al.* 2004) uses a statistical model which introduces a small amount of smoothing to the data (though no more than is apparently justified by the expected random noise – and indeed this model better reflects underlying data when we have annual scale input when compared to IntCal98 – see Figure 2). Since such methods cannot distinguish between random outliers and real extreme values there are some real short-term fluctuations that may be attenuated in this compilation (especially when the underlying data are only decadal or bidecadal). There is scope for further work to refine these statistical methods. However, from the point of view of a user of calibration, IntCal04 provides the most comprehensive and up-to-date estimate of the Northern Hemisphere calibration curve and should always be the first choice for calibration. Comparison of the results with those obtained against the IntCal98 (Stuiver, Reimer, Bard *et al.* 1998) calibration curve, which used a simple binning and weighted average of the data then available, can be valuable as can comparison against individual datasets. Such a degree of complexity is however only really warranted in large-scale Bayesian models (when the results are usually insensitive to such changes) or wiggle-matching of tree ring sequences (where differences are occasionally significant if the match relies predominantly on one or two fluctuations in radiocarbon levels). For normal calibration the IntCal04 curve is all that is required.

Between 12 400 and 26 000 cal BP, the current situation is slightly different. Here the calibration curve is based on multiple records in good agreement, but these are all marine records and therefore represent our best estimate of the atmospheric concentration. There may however be changing marine reservoir offsets that could mean the curve in some sections of this time period is out by as much as 250 ^{14}C years BP (Bondevik *et al.* 2006; Kromer *et al.* 2004). It is very unlikely to be worse than this given the agreement of IntCal04 with other records not used in the calibration curve, such as the terrestrial macrofossil record from Lake Suigetsu (Kitagawa & van der Plicht 2000). In this time range calibration for archaeological purposes is possible. However, such calibration is more provisional and there could be some minor changes as new data accumulate, particularly from terrestrial records, which might be significant in certain contexts (see Figure 3).

Further back than 26 000 cal BP, the situation is radically different. Here the records are neither based on purely terrestrial material, nor do they agree with one another (see Figure 4). As stated above, some of these discrepancies are being addressed actively and within a couple of years the situation is likely to be much better. However, the possible major discrepancies between the marine and atmospheric data need to be viewed with particular caution as they imply that even with consistent marine records we may still not understand how to interpret the records in the context of terrestrial archaeological samples. Given this, it is clear why the radiocarbon community does not think that calibration as such is possible in this time range, since it is not clear which, if any, of the present records provide a good indication of the atmospheric radiocarbon concentration. Thus far there is only one record that represents true atmospheric ^{14}C measurements (Lake Suigetsu); however, this record stands alone in the sense that it is not confirmed by others. We know for the period in which we do have an atmospheric record that there are many short-term fluctuations, which are missing from the marine record. This is likely to be even more significant in periods where the climate is much less stable, there may be major magnetic excursions, and the resolution of the marine measurements we do have is poorer.

The highest resolution record in this time range, that from the Cariaco basin (Hughen, Lehman *et al.* 2004), illustrates many of these points clearly and also shows what can and cannot be done with the current data. The samples for this record are marine, and they are absolutely dated by matching changes in the characteristics of the sediments to changes in the climate as recorded by the Greenland ice cores, in this case GISP2. This means that the timescale used is the GISP2 timescale, which is based on a model of ice accumulation beyond 41 000 cal BP. Recent work on the NGRIP core (presentation of J.P. Steffensen at the Oxford Radiocarbon Conference) suggests that the GISP2 timescale has non-linear errors, which means that not only are the absolute ages wrong, but that rates of change estimated from this timescale may be significantly incorrect too. This in turn then significantly impacts archaeological assessments made using the 2004 Cariaco record (as in Mellars 2006). As reported at the Oxford Radiocarbon Conference, this particular problem is likely to be addressed by linking to other absolutely dated records, most likely that at Hulu Cave (Wang *et al.* 2001). However, the uncertainty in the difference between the atmospheric and marine radiocarbon concentration will not be so easily addressed. Even though these differences seem to be fairly well behaved in the late glacial we cannot assume that this is always the case. That said, comparison of radiocarbon dates to this record can undoubtedly be valuable, particularly if what is of interest is how the dates lie in relation to the changes in climate as recorded in the GISP2 δ^{18}O record – but where this is done it should always be made clear that the comparison is made against this record and is on the GISP2 or NGRIP timescale (as discussed in Gravina *et al.* 2005).

Other records are also valuable for archaeologists. Coral data, although also marine, link into a more absolute uranium-series-based chronology. This is better from the point-of-view of absolute ages – though not as useful if you wish to compare them to the oxygen isotope records of the Greenland ice cores. Furthermore, the coral-based records such as that of Fairbanks *et al.* (2005), are not continuous records, since they are based on chance finds of corals; nor is it likely that they are an entirely random sample since formation factors linking to climate and environmental changes are likely to bias the recovered sample set. Thus any curve generated from such datasets looks smooth. But we must remember that absence of evidence is not evidence of absence and such a curve almost certainly fails to show even the scale of fluctuations in the radiocarbon concentration of the oceans, let alone the levels of variability in the atmosphere. Available climate indicators suggest similar (e.g. Roig *et al.* 2001) or greater (e.g. Bond & Lotti 1995; Dansgaard *et al.* 1993) periods and cycles of change for the later Pleistocene compared to the Holocene (e.g. Bond *et al.* 2001). These would be reflected in an atmospheric ^{14}C record giving at least as many, and very likely more, variations and cyclical features than for the record available for the Holocene. At present we are largely lacking such information for the periods before terrestrial tree-based records, and measurement errors on very old radiocarbon ages will anyway tend to mask some of the expected century-scale variation. The record for Lake Suigetsu is potentially very useful as it is purely terrestrial, but it lacks a good absolute timescale. The speleothem records are partly terrestrial and so also provide useful information on the possible scale of differences between the radiocarbon concentration of the surface oceans and the terrestrial groundwater. No one record is right in all respects but all give information that is potentially useful. Because their problems are all different it is also potentially misleading to compile them into a composite curve for calibration since this merely serves to mask the underlying complications. This is the reason for the ironically named NOTCal curve (van der Plicht *et al.* 2004), and the criticisms levelled at aspects of the CalPal program referred to by van Andel (2005).

So what should the archaeological researcher working in this period do? Ignoring the problem, either by assuming that radiocarbon ages in this period can be treated as some approximate proxy for age, or by using some *ad hoc* compilation of data into a ‘comparison’ curve as if it were a ‘calibration’ curve cannot be regarded as good scholarship. It is almost bound to result in conclusions and assertions which will have to be changed (and quite possibly significantly) within a very few years – indeed often before the research is physically published. The uncertainties need to be fully acknowledged. The correct approach will depend very much on the application. In many cases it may be appropriate to compare dates to a number of different specific records – unless there are very particular reasons for one record being most appropriate. The timescale to which the comparison has been made (for example Uranium Series or NGRIP ice core) should be made explicit and the ages deduced would be better referred to as ‘estimated’ rather than ‘calibrated’ dates. Most crucially all should be aware that these estimates may well change significantly as our understanding of the Earth’s system during the last glaciation improves. If absolute ages are the primary interest then there is not really much of a substitute for comparison against all of the main records since this demonstrates the range of possible true ages depending on which of the records most closely reflect the relevant reality. As the datasets improve, this exercise will hopefully provide a narrower and narrower range of possibilities.

Conclusions

There has been considerable progress in recent years in the level of information available for assessing the past radiocarbon concentration of the atmosphere and oceans. This information is very valuable for archaeologists since it helps them to interpret their radiocarbon dates in terms of absolute chronology. However, the cost of this progress is increasing complexity in the nature of the data, and this means that archaeologists need to have a critical understanding of what sort of analyses the data can and cannot support.

Back to about 12 400 cal BP, the data are fairly robust and the IntCal04 calibration curve should provide accurate calibration for most purposes. Where very high precision is required, with large Bayesian models or wiggle-matching of tree ring sequences, it may also be valuable to compare the results of such analyses against the IntCal98 curve because it is compiled differently (even though it does have known deficiencies), or against individual datasets (such as Irish oak in the case of British sites, for example).

In the period between 12 400 and 26 000 cal BP any calibration is more provisional since the data used for construction of the calibration curve are marine-derived. However, given the level of agreement between records in this region, such calibration is likely to be fairly accurate and for most purposes the IntCal04 calibration curve can be used as it is. In some critical applications, it may also be useful to compare such calibration to estimates from individual records.

Beyond 26 000 cal BP, there is no accepted calibration curve simply because of the disparity in the records we have for this time period as of June 2006, and so comparison should be made to a range of individual records to estimate ages on the timescale relevant to the specific records. The records used for such comparisons will depend on the details of the application. If climatic correlations are important, then records that link to climatic data will be most useful. On the other hand, if absolute ages are the main issue, then the full range of datasets should be considered to see the range of possibilities.

In order to prevent confusion, it makes a lot of sense to reserve the terms 'calibration' and 'calibrated dates' for analyses based on the recognised calibration curves (IntCal04, SHCal04 & Marine04). In the periods covered by these curves it may also be useful to make a 'comparison' against other records. The term 'estimated dates' for the results of such analyses seems most appropriate. Where calibration is not yet possible, 'comparison' against the different records now available may still be useful but the provisional nature of such analyses should be fully appreciated. As with the paper by Mellars (2006), speculation about the implications of the data as they emerge are entirely appropriate but the caveat at the end of that piece is important to keep fully in mind: 'A final, definitive calibration curve for this time range will depend on the results of new calibration studies, at present being pursued in several different laboratories. The full implications of these studies for the interpretation of the human archaeological and evolutionary record will need to be kept under active and vigilant review'. If we always remember this, we should avoid the inevitable disappointment when new facts emerge to overturn a beautiful and elegant hypothesis constructed on the basis of preliminary data.

References

Bard, E., M. Arnold, B. Hamelin, N. Tisnerat-Laborde & G. Cabioch. 1998. Radiocarbon calibration by means of mass spectrometric Th-230/U-234 and C-14 ages of corals: An updated database including samples from Barbados, Mururoa and Tahiti. *Radiocarbon* 40(3): 1085-92.

Bard, E., F. Rostek & G. Menot-Combes. 2004. Radiocarbon calibration beyond 20 000 C-14 yr BP by means of planktonic foraminifera of the Iberian Margin. *Quaternary Research* 61(2): 204-14.

Beck, J.W., D.A. Richards, R.L. Edwards, B.W. Silverman, P.L. Smart, D.J. Donahue, S. Hererra-Osterheld, G.S. Burr, L. Calsoyas, A.J.T. Jull & D. Biddulph. 2001. Extremely large variations of atmospheric C-14 concentration during the last glacial period. *Science* 292(5526): 2453-58.

Bond, G., B. Kromer, J. Beer, R. Muscheler, M.N. Evans, W. Showers, S. Hoffmann, R. Lotti-Bond, I. Hajdas & G. Bonani. 2001. Persistent solar influence on north Atlantic climate during the Holocene. *Science* 294(5549): 2130-36.

Bond, G.C. & R. Lotti. 1995. Iceberg Discharges into the North-Atlantic on Millennial Time Scales during the Last Glaciation. *Science* 267(5200): 1005-10.

Bondevik, S., J. Mangerud, H.H. Birks, S. Gulliksen & P. Reimer. 2006. Changes in North Atlantic radiocarbon reservoir ages during the Allerod and Younger Dryas. *Science* 312(5779): 1514-17.

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

Buck, C.E. & P.G. Blackwell. 2004. Formal statistical models for estimating radiocarbon calibration curves. *Radiocarbon* 46(3): 1093-1102.

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

Chiu, T.C., R.G. Fairbanks, R.A. Mortlock, L. Cao, T.W. Fairbanks & A.L. Bloom. 2006. Redundant ^{230}Th/^{234}U/^{238}U, ^{231}Pa/^{235}U and ^{14}C dating of fossil corals for accurate radiocarbon age calibration. *Quaternary Science Reviews* 25(17-18): 2431-40.

Cutler, K.B., S.C. Gray, G.S. Burr, R.L. Edwards, F.W. Taylor, G. Cabioch, J.W. Beck, H. Cheng & J. Moore. 2004. Radiocarbon calibration and comparison to 50 kyr BP with paired C-14 and Th-230 dating of corals from Vanuatu and Papua New Guinea. *Radiocarbon* 46(3): 1127-60.

Dansgaard, W., S.J. Johnsen, H.B. Clausen, D. Dahljensen, N.S. Gundestrup, C.U. Hammer, C.S. Hvidberg, J.P. Steffensen, A.E. Sveinbjornsdottir, J. Jouzel & G. Bond. 1993. Evidence for General Instability of Past Climate from a 250-Kyr Ice-Core Record. *Nature* 364(6434): 218-20.

Fairbanks, R.G., R.A. Mortlock, T.C. Chiu, L. Cao, A. Kaplan, T.P. Guilderson, T.W. Fairbanks, A.L. Bloom, P.M. Grootes & M.J. Nadeau. 2005. Radiocarbon calibration curve spanning 0 to 50 000 years BP based on paired Th-230/U-234/U-238 and C-14 dates on pristine corals. *Quaternary Science Reviews* 24(16-17): 1781-96.

Gravina, B., P. Mellars & C. Bronk Ramsey. 2005. Radiocarbon dating of interstratified Neanderthal and early modern human occupations at the Chatelperronian type-site. *Nature* 438(7064): 51-6.

Hughen, K.A., M.G.L. Baillie, E. Bard, 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, B. Kromer, G. McCormac, S. Manning, C. Bronk Ramsey, P.J. Reimer, R.W. Reimer, S. Remmele, J.R. Southon, M. Stuiver, S. Talamo, F.W. Taylor, J. van der Plicht & C.E. Weyhenmeyer. 2004a. Marine04 marine radiocarbon age calibration, 0-26 cal kyr BP. *Radiocarbon* 46(3): 1059-86.

Hughen, K.A., S. Lehman, J. Southon, J. Overpeck, O. Marchal, C. Herring & J. Turnbull. 2004b. C-14 activity and global carbon cycle changes over the past 50 000 years. *Science* 303(5655): 202-7.

Hughen, K.A., J.T. Overpeck, S.J. Lehman, M. Kashgarian & J.R. Southon. 1998. A new C-14 calibration data set for the last deglaciation based on marine varves. *Radiocarbon* 40(1): 483-94.

Hughen, K.A., J.R. Southon, C.J.H. Bertrand, B. Frantz & P. Zermeno. 2004c. Cariaco basin calibration update: Revisions to calendar and C-14 chronologies for core PL07-58PC. *Radiocarbon* 46(3): 1161-87.

Joris, O. & B. Weninger. 1998. Extension of the C-14 calibration curve to ca. 40 000 cal BC by synchronizing Greenland O-18/O-16 ice core records and North Atlantic foraminifera profiles: A comparison with U/Th coral data. *Radiocarbon* 40(1): 495-504.

Kitagawa, H. & J. van der Plicht. 1998. Atmospheric radiocarbon calibration to 45 000 yr BP: Late glacial fluctuations and cosmogenic isotope production. *Science* 279(5354): 1187-90.

–2000. Atmospheric radiocarbon calibration beyond 11 900 cal BP from Lake Suigetsu laminated sediments. *Radiocarbon* 42(3): 369-80.

Kromer, B., M. Friedrich, K.A. Hughen, F. Kaiser, S. Remmele, M. Schaub & S. Talamo. 2004. Late glacial C-14 ages from a floating, 1382-ring pine chronology. *Radiocarbon* 46(3): 1203-9.

Kromer, B., S.W. Manning, P.I. Kuniholm, M.W. Newton, M. Spurk & I. Levin. 2001. Regional (CO2)-C-14 offsets in the troposphere: Magnitude, mechanisms, and consequences. *Science* 294(5551): 2529-32.

McCormac, F.G., A.G. Hogg, P.G. Blackwell, C.E. Buck, T.F.G. Higham & P.J. Reimer. 2004. SHCal04 Southern Hemisphere calibration, 0-11.0 cal kyr BP. *Radiocarbon* 46(3): 1087-92.

Mellars, P. 2006. A new radiocarbon revolution and the dispersal of modern humans in Eurasia. *Nature* 439(7079): 931-5.

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, G. McCormac, S. Manning, C. Bronk Ramsey, R.W. Reimer, S. Remmele, J.R. Southon, M. Stuiver, S. Talamo, F.W. Taylor, J. van der Plicht & C.E. Weyhenmeyer. 2004. IntCal04 terrestrial radiocarbon age calibration, 0-26 cal kyr BP. *Radiocarbon* 46(3): 1029-58.

Richards, D.A. & J.W. Beck. 2001. Dramatic shifts in atmospheric radiocarbon during the last glacial period. *Antiquity* 75(289): 482-5.

Roig, F.A., C. Le-Quesne, J.A. Boninsegna, K.R. Briffa, A. Lara, H. Grudd, P.D. Jones & C. Villagran. 2001. Climate variability 50 000 years ago in mid-latitude Chile as reconstructed from tree rings. *Nature* 410(6828): 567-70.

Simpson, J.A. & E.S.C. Weiner. 1989. *The Oxford English dictionary*. 2nd ed. 20 vols. Oxford: Clarendon Press.

Southon, J. 2004. A radiocarbon perspective on Greenland ice-core chronologies: Can we use ice cores for C-14 calibration? *Radiocarbon* 46(3): 1239-59.

Stein, M., C. Migowski, R. Bookman & B. Lazar. 2004. Temporal changes in radiocarbon reservoir age in the dead sealake Lisan system. *Radiocarbon* 46(2): 649-55.

Stuiver, M. 1986. Proceedings of the 12th International Radiocarbon Conference – Held at Trondheim Norway 24-28 June 1985. *Radiocarbon* 28(2B): R2-R2.

Stuiver, M. & T.F. Braziunas. 1998. Anthropogenic and solar components of hemispheric C-14. *Geophysical Research Letters* 25(3): 329-32.

Stuiver, M. & P.J. Reimer. 1993. Extended C-14 Data-Base and Revised Calib 3.0 C-14 Age Calibration Program. *Radiocarbon* 35(1): 215-30.

Stuiver, M., P.J. Reimer, E. Bard, J.W. Beck, G.S. Burr, K.A. Hughen, B. Kromer, G. McCormac, J. Van der Plicht & M. Spurk. 1998. INTCAL98 radiocarbon age calibration, 24 000-0 cal BP. *Radiocarbon* 40(3): 1041-83.

Stuiver, M., P.J. Reimer & T.F. Braziunas. 1998. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* 40(3): 1127-51.

Turney, C.S.M., R.G. Roberts & Z. Jacobs. 2006. Progress and pitfalls in radiocarbon dating. *Nature* 443: E3-E4.

Valladas, H., J. Clottes, J.M. Geneste, M.A. Garcia, M. Arnold, H. Cachier & N. Tisnerat-Laborde. 2001. Palaeolithic paintings – Evolution of prehistoric cave art. *Nature* 413(6855): 479.

van Andel, T.H. 2005. The ownership of time: approved C-14 calibration or freedom of choice? *Antiquity* 79(306): 944-8.

van der Plicht, J. 1993. The Groningen Radiocarbon Calibration Program. *Radiocarbon* 35(1): 231-7.

–2000. Introduction: The 2000 Radiocarbon varve/comparison issue. *Radiocarbon* 42(3): 313-22.

van der Plicht, J., J.W. Beck, E. Bard, M.G.L. Baillie, P.G. Blackwell, C.E. Buck, M. Friedrich, T.P. Guilderson, K.A. Hughen, B. Kromer, F.G. McCormac, C. Bronk Ramsey, P.J. Reimer, R.W. Reimer, S. Remmele, D.A. Richards, J.R. Southon, M. Stuiver & C.E. Weyhenmeyer. 2004. NotCal04 – Comparison/calibration C-14 records 26-50 cal kyr BP. *Radiocarbon* 46(3): 1225-38.

Vogel, J.C. & J. Kronfeld. 1997. Calibration of radiocarbon dates for the late Pleistocene using U/Th dates on stalagmites. *Radiocarbon* 39(1): 27-32.

Wang, Y.J., H. Cheng, R.L. Edwards, Z.S. An, J.Y. Wu, C.C. Shen & J.A. Dorale. 2001. A high-resolution absolute-dated Late Pleistocene monsoon record from Hulu Cave, China. *Science* 294(5550): 2345-8.