UCRL- JRNL-207975

NotCal04 - Comparison / Calibration ^{14}C records 26-50 cal kBP

J.van der Plicht[1*], J.W.Beck[2], E.Bard[3], M.G.L.Baillie[4], P.G.Blackwell[5], C.E.Buck[5], M.Friedrich[6], T.P.Guilderson[7,11], K.A.Hughen[8], B.Kromer[9], F.G.McCormac[4], C. Bronk Ramsey[10], P.J.Reimer[7,4], R.W.Reimer[11], S. Remmele[6], D. A. Richards[12], J.R.Southon[13], M.Stuiver[14], C.E. Weyhenmeyer[7]

To be submitted to *Radiocarbon*.

November 11, 2004

This document was prepared as an account of work sponsored by an agency of the United States Government. Neither the United States Government nor the University of California nor any of their employees, makes any warranty, express or implied, or assumes any legal liability or responsibility for the accuracy, completeness, or usefulness of any information, apparatus, product, or process disclosed, or represents that its use would not infringe privately owned rights. Reference herein to any specific commercial product, process, or service by trade name, trademark, manufacturer, or otherwise, does not necessarily constitute or imply its endorsement, recommendation, or favoring by the United States Government or the University of California. The views and opinions of authors expressed herein do not necessarily state or reflect those of the United States Government or the University of California, and shall not be used for advertising or product endorsement purposes.

NotCal04 - Comparison / Calibration ^{14}C records 26-50 cal kBP

J.van der Plicht[1*], J.W.Beck[2], E.Bard[3], M.G.L.Baillie[4], P.G.Blackwell[5], C.E.Buck[5], M.Friedrich[6], T.P.Guilderson[7,11], K.A.Hughen[8], B.Kromer[9], F.G.McCormac[4], C. Bronk Ramsey[10], P.J.Reimer[7,4], R.W.Reimer[11], S. Remmele[6], D. A. Richards[12], J.R.Southon[13], M.Stuiver[14], C.E. Weyhenmeyer[7]

[1]Center for Isotope Research, Groningen University, 9747 AG Groningen, Netherlands
[2]Department of Physics, University of Arizona, Tucson, AZ 85721 USA
[3]CEREGE, UMR-6635, Europole de l'Arbois BP80, 13545 Aix-en-Provence cdx 4, France
[4]School of Archaeology and Palaeoecology, Queen's University Belfast, Belfast BT7 1NN, UK
[5]Department of Probability and Statistics,University of Sheffield, Sheffield S3 7RH, UK
[6]Universität Hohenheim, Institut für Botanik-210, D-70593 Stuttgart, Germany
[7]Center for Accelerator Mass Spectrometry L-397, Lawrence Livermore National Laboratory, Livermore, CA 94550, USA
[8]Woods Hole Oceanographic Institution, Department of Marine Chemistry & Geochemistry, Woods Hole, MA 02543 USA
[9]Heidelberger Akademie der Wissenschaften,Im Neuenheimer Feld 229, D-69120 Heidelberg, Germany
[10]Oxford Radiocarbon Accelerator Unit, 6 Keble Rd, Oxford OX2 6JB, UK
[11]Department of Ocean Sciences, University of California - Santa Cruz, Santa Cruz, CA 95064 USA
[12]School of Geographical Sciences, University of Bristol, Bristol BS8 1SS, UK
[13] Department of Earth System Science, University of California - Irvine, Irvine, CA 92697 USA
[14]Quaternary Isotope Lab, University of Washington, Seattle, WA 98195 USA

* corresponding author, plicht@phys.rug.nl

Abstract

The radiocarbon calibration curve, IntCal04, extends back to 26 cal kBP. While several high resolution records exist beyond this limit, these data sets exhibit discrepancies one to another of up to several millennia. As a result, no calibration curve for the time range 26-50 cal kBP can be recommended as yet, but in this paper the IntCal04 working group compares the available data sets and offers a discussion of the information that they hold.

Keywords
calibration, comparison, radiocarbon dating, statistical models, random effects

Introduction

The need for calibration of the ^{14}C time scale was recognized soon after the discovery of the ^{14}C dating method (Libby 1955). Suess (1970) demonstrated that it is variations in the ^{14}C content of the atmosphere which cause the ^{14}C time scale to be different from the calendar time scale. Traditionally, the calibration of the radiocarbon to the calendar time scale has been based on ^{14}C measurements for tree-rings dated by dendrochronology. In general, dendrochronology can provide absolute dates, so that reliable ^{14}C calibration curves can be obtained. From 1998, until the publication of the present volume, the recommended calibration curve has been INTCAL98, published in a special volume of *Radiocarbon* (Stuiver and van der Plicht, eds., 1998). This curve comprises not only dendrochronologically obtained calibration data, but also paired ^{14}C / U-series dates from Pacific corals (Bard et al. 1998; Burr et al. 1998) and high resolution ^{14}C measurements from the laminated sediment from the Cariaco basin (Hughen et al. 1998a, 1998b). Both datasets are from marine environments so a reservoir effect has to be taken into account in order to allow such data to be incorporated into the atmospheric calibration curve. With the addition of such data, however, INTCAL98 extended back to 24,000 cal BP with only a few coral points beyond 15,585 cal BP (Stuiver et al. 1998).

Beyond the limit of 15,585 cal BP, many other datasets were available but were not selected for inclusion in INTCAL98 because of inconsistencies between the different records. At the time of INTCAL98, a variety of "calibration curves" could have been constructed based on ^{14}C dating paired with TL, U-series and varved layer counts. However, the INTCAL98 team did not feel that the data were sufficiently well understood to recommend a calibration curve for material older than 15585 cal BP. Nevertheless it was recognized that the data contained important information on past natural ^{14}C variations and hence should shed some light on calibration issues. As a result, these data were collected together in a separate volume of *Radiocarbon*, the so-called varve/comparison issue (van der Plicht 2000a).

Although attempts were made to construct a "calibration curve" based on such data (e.g. Vogel and Kronfeld 1997), both scatter and measurement error were much too large for useful calibration purposes. In addition, the temporal resolution for such data sets was typically very low.

Since 1998, the INTCAL working group has been expanded considerably, as a result of the spectacular growth in new data. More than 1000 AMS dates containing information on calibration of the ^{14}C time scale are now available, with datasets covering the complete ^{14}C dating range of ca. 50,000 years. As a result, great progress has been

made in understanding the variations in natural ^{14}C level in the past which lead to the need for calibration of the ^{14}C time scale. Early in its work, the IntCal04 working group specified criteria on the basis of which data sets would be accepted or rejected for calibration purposes (Reimer et al. 2002). Data sets that meet these criteria were used to construct the calibration curves IntCal04 and Marine04. These are based on an increased number of measurements of ^{14}C vs. dendrochronology, U-series dated corals and laminated sediment from the Cariaco basin, and on improved statistical methods (Buck and Blackwell, 2004) for estimating the underlying calibration curves from the available data. The new curves now extend to 26 cal kBP (Reimer et al. 2004; Hughen et al. 2004a).

Beyond 26 cal kBP, the records from individual research projects deviate too much from one another to justify recommendation of a calibration curve for the timespan 26-50 cal kBP. Nevertheless, the IntCal04 working group undertook a range of investigations to see what options there are for understanding more about the underlying calibration curve in this period. The outcomes of our research in the range 26-50cal kBP are not to be used as calibration curves (hence the name NotCal04). Instead of calibration curve, the phrase "comparison curve" (van der Plicht 2000b; Richards and Beck 2001) is more appropriate. In what follows, despite the fact that we are not able to recommend a definitive estimate of the radiocarbon calibration curve for 26-50 cal kBP, we wish to highlight some of the approaches we considered. We do this mainly because researchers who make use of records beyond 26 cal kBP to derive calibrated ages, sometimes do so without proper regard for uncertainties and the work that we have done highlights just how enormous such uncertainties can be. The status of the present comparison effort, NotCal04, should be characterised as "work in progress" since zooming in towards a true calibration curve is a continuous process.

In the next section, we review the most important datasets currently available – taking into account some aspects of the criteria established by our working group (Reimer et al. 2002). Following on from that, we outline our investigations into options for constructing calibration curves in the presence of uncertainty and illustrate the difficulties we face in producing a definitive calibration curve for this period. Finally, some aspects of ^{14}C calibration beyond 26 cal kBP – crucial for the chronology of e.g. the Upper Palaeolithic - are discussed.

Datasets available in the range 26-50 cal kBP

Since the onset of ^{14}C dating, data for calibration have been obtained on the basis, for example, of TL vs. ^{14}C comparisons, α - spectometric (U/Th) vs. ^{14}C dated carbonates etc. Many of these records

are summarised in the 2000 *Radiocarbon* varve/comparison issue (van der Plicht 2000a) and are not further discussed here.

INTCAL98 (Stuiver et al. 1998) was the first Radiocarbon intercalibration to use ^{14}C archives other than tree rings, which include data from pacific corals dated using TIMS U/Th ages (Bard et al. 1998; Burr et al. 1998). Data from the Cariaco Basin varved sediments (Hughen et al. 1998a), whose age constraints were derived from varve counting as well as grey scale correlation with GISP2 oxygen isotopes were also added to this calibration.

Then (starting with the 16th International Radiocarbon Conference in Groningen in 1997) spectacular datasets with potential information on ^{14}C calibration have become available, some of them based on hundreds of AMS measurements yielding high temporal resolution. These datasets are illustrated in Figure 1 and discussed below. Readers should note that the error bars on the ^{14}C scale in Figure 1 represent one standard deviation and that we have not attempted to represent the calendar age uncertainties on this plot at all. For many of the data points, the calendar age uncertainties are substantial and the sources of these errors will be discussed for each data set below. All of the errors that we have been able to quantify, both on the calendar and ^{14}C scale, were taken into account in the statistical investigations summarized below and reported in greater detail in Buck and Blackwell (2004).

a). High temporal resolution datasets

a1). Lake Suigetsu

The first high temporal resolution radiocarbon calibration dataset measured by AMS arose from a laminated sediment from Lake Suigetsu, Japan (Kitagawa and van der Plicht 1998). For this lake sediment, a 29,100 year long varve chronology was constructed. More than 330 ^{14}C measurements were performed for terrestrial samples (mostly macrofossils, but also insects, branches and leaves) from the sediment. The varve chronology is not absolute but floating; the youngest part of the sediment overlaps with the oldest part of the tree-ring dataset and can therefore be matched using the ^{14}C measurements. Thus, the varve chronology was found to derive from the range 8,830 to 37,930 cal BP with an error in the wiggle match of ca. 5 years. The full date list of ^{14}C measurements from Lake Suigetsu (updated since 1998) can be found in Kitagawa and van der Plicht (2000). For this data set, in addition to the wiggle match error, other calendar age uncertainties (not shown in Figure 1) arise from the possible miscounting of varves and/or hiatuses in the varve sequences. Any such errors would be cumulative and greatest for the oldest part of the record. This error is estimated as not larger than 2000 varve years at the oldest part of the dataset.

Beyond 37,930 cal BP, there are a few more measurements (labelled "extension" in Figure 1); this part of the core however is not varve-counted; the calendar time scale for this part is based on sedimentation rate. Because of the uncertainty in the calendar time scale for this part of the record the extension is not used in NotCal04. Currently, the Lake Suigetsu record is the only atmospheric radiocarbon calibration data set available in this time range derived from terrestrial organic material.

a2). Bahamas Stalagmite

The next high temporal resolution calibration dataset measured by AMS arises from a speleothem from the Bahamas (Beck et al. 2001). For this archive, close to 300 ^{14}C and 81 TIMS/U-series dates have been measured. Like the coral and Cariaco data, this dataset is not a direct measure of atmospheric ^{14}C abundance, since a carbonate reservoir effect (called DCF, Dead Carbon Fraction) has to be corrected for. This DCF was determined to be 1450 ± 235 (1σ) ^{14}C years based on the offset with INTCAL98 between 11 and 16 ka. The DCF is assumed constant for the entire period of growth of this speleothem and is discussed in more detail in Richards et al. (2003). The U-series dates were used to generate a distance-age relationship along the growth axis of the speleothem using a weighted smoothing spline. The smoothing parameter (spar) used was automatically generated using the generalised cross-validation method. The resulting age model was used to define an absolute age scale (i.e. assumed equivalent to calendar age) for the ^{14}C measurements. The 95% confidence limits, which range from 25 to 320 years, were calculated for each calendar age by predicting a fit (spar =0) of the upper and lower limits of uncertainty for the period between 28 and 45 ka BP. These errors are not shown in Figure 1, but appear in the database compiled by the IntCal04 working group and are used in the statistical investigations discussed below. Note that there is a ca. 2 ka gap in the record at around 27 cal kBP because the drip that formed this speleothem shifted at that time. Because of the difficult geometry of this part of the sample, this interval of growth was not sampled.

a3) Cariaco Basin

The Cariaco Basin is an anoxic marine basin off the coast of Venezuela, separated from the open Caribbean Sea by shallow sills, possessing sediments used for continuous, high resolution AMS ^{14}C dating of high concentrations of planktonic foraminifera (Hughen et al. 1998a). The late Glacial section of the core is laminated, and the ^{14}C measurements from this section (Hughen et al. 1998b) were used in the construction of the calibration curves INTCAL98 (Stuiver et al. 1998) and IntCal04 (Reimer et al. 2004; Hughen et al. 2004a).

Foraminifera from the oldest part of the record were also dated by AMS. This part, however, is not varved so that only a floating chronology can be constructed. However, by comparing climate signals as they appear in the $\delta^{18}O$ values of the foraminifera, to the $\delta^{18}O$ values of the Greenland ice cores, this non-varved sequence can be tied to a calendar scale. The non-varved sequence was matched in this way to the ice core GISP2 (Meese et al. 1997), and an estimate of the calendar age of each sample was obtained. A marine reservoir age of 420 years was then subtracted (Hughen et al. 2004b). Thus, the non-varved Cariaco Basin data set consists of 225 AMS measurements with uncertainties on the ^{14}C ages (indicated as one standard deviation error bars on Figure 1), but also uncertainty in the calendar ages. The calendar age uncertainties arise from the uncertainty in the match between the GISP2 ice core and the Cariaco Basin $\delta^{18}O$, which results in a 180 year error, and the error estimate in the GISP2 ice core time scale (Meese et al. 1997).

a4) Arabian Speleothem

Socotra is an island in the Indian Ocean, south of the Arabian coast. From Moomi Cave located on the western side of the island, a stalagmite was collected, which grew between ca. 42,000 and 55,000 cal BP. The stalagmite has been thoroughly analysed for the stable isotopes $\delta^{18}O$ and $\delta^{13}C$ (Burns et al. 2003). Radiocarbon measurements were carried out on small pieces cut from the center of the stalagmite (Weyhenmeyer et al. 2003). The ^{14}C values were corrected for reservoir age (DCF, Dead Carbon Fraction) using the $\delta^{13}C$ values as a constraint for water-rock interaction. The calendar ages for this speleothem are measured by U-series isotopes (Burns et al 2003; 2004). The average DCF is ca. 15% and fairly constant during the 13,000 year growth period of the stalagmite. A correction for DCF of 1237 ± 300 ^{14}C years was used. A linear growth model based on nineteen ^{230}Th measurements provides the chronology for this record with an uncertainty of 500 years. The uncertainty of 250 years includes the uncertainty in the 230Th measurements and the least squares fit.

a5) North Atlantic marine data

Both benthic and planktonic foraminifera from North Atlantic marine cores were ^{14}C dated by AMS (Voelker et al. 2000; van Kreveld et al. 2000). No calendar age estimates are associated with these measurements. Nevertheless, data with potential for calibration were obtained by comparing climate signals as they appear in the $\delta^{18}O$ values of the foraminifera, to the $\delta^{18}O$ values of the nearby Greenland ice cores to provide calendar age estimates.

Since these are marine data, an appropriate correction for reservoir age also needs to be undertaken. For the North Atlantic marine data of

Voelker et al. (2000) and van Kreveld et al. (2000), the reservoir age is difficult to determine. The area is hydrologically very complex. For instance, very large age differences between the benthic and planktonic foraminifera were observed (Voelker et al. 2000). The data are plotted in Figure 1 assuming a constant 400 year reservoir correction, and using the Voelker et al. (2000) $\delta^{18}O$ correlation with the GISP 2 ice core chronology (Meese et al. 1997) in order to generate a calendar time scale. This dataset was not included in the NotCal04 comparison curve so we have not included error estimates for the calendar ages or for the reservoir ages.

b) Other datasets (lower temporal resolution)

b1) Corals

The new IntCal04 ^{14}C calibration curve based on dendrochronology has been extended into the Late Glacial using Pacific and Atlantic Corals, dated by both U-series isotopes and ^{14}C-AMS (Bard et al. 1998; Burr et al. 1998; Cutler et al. 2004; Fairbanks et al. 2004). These coral datasets qualify for ^{14}C calibration purposes because they meet the criteria as established by the IntCal04 working group (Reimer et al., 2002). Beyond 26 cal kBP, only a few data points from corals are currently available (Bard et al. 1998; Cutler et al. 2004). Nevertheless they are considered as "reference data points". Site specific reservoir ages (as detailed in Hughen et al. 2004a) have been subtracted.
More Atlantic coral data will be forthcoming (Fairbanks et al., in preparation).

b2) Iberian Margin

Foraminifera from Atlantic deep sea cores off the Iberian Coast have been dated by ^{14}C (Bard et al. 2004). Several independent proxies show marked variations (alkenones, ice rafted debris, magnetic susceptibility, $\delta^{18}O$ etc). A calendar age scale can be derived from the correlation of the temperature proxies (alkenones, $\delta^{18}O$) to the Greenland ice cores. For consistency with the Cariaco Basin record, the Iberian Margin data have been matched to the GISP2 timescale. Only a few data are available to date, but in general they follow the corals curve. The data are in good agreement with the new Cariaco Basin data. Note that both Iberian Margin and Cariaco Basin data are plotted along the GISP2 cal BP time scale.

A marine reservoir correction of 500 ± 100 ^{14}C years has been applied, based on pre-1950's molluscs and an assessment of the hydrological regime of the site (Bard et al. 2004).

b3) Lake Lisan

For Lake Lisan (the Last Glacial Dead Sea), paired ^{14}C/U-series dates for aragonite material are available (Schramm et al. 2000). The record shows a hiatus between ca. 43,000 an 50,000 years ago. To bring them onto the terrestrial radiocarbon scale, the aragonite samples can be corrected for reservoir age (DCF), which is determined from a single analysis of modern Dead Sea material has a value of ca. 1000 ^{14}C years. This reservoir correction was assumed constant and applied to all ^{14}C ages. The U-series measurements also need to be corrected, but in this case for detrital Th and U and authigenic (unsupported) Th, generated from dissolved U in the lake water. Figure 1 includes these data after all these corrections have been made; some data points (errors $^{14}C > 1$ ka BP) are not plotted.

For the youngest part of the record (up to ca. 30,000 years ago), a high temporal resolution dataset for Lake Lisan became available recently (van der Borg et al. 2004). Analysis of terrestrial material in aragonitic layers indicates that the DCF may have varied considerably.
The Lake Lisan dataset was not included in NotCal04 so we have not included error estimates for te calendar ages or for the DCF correction.

We emphasize, once again, that all the data shown in Figure 1 are plotted with errors only along the vertical axis (i.e. one standard deviation about the mean on ^{14}C ages in BP). The errors along the horizontal axis (cal BP) are not plotted because (apart from the fact that the plot would be unreadable) the errors in cal BP between the records arise in several different ways. In some cases they derive from matters like varve counting and wiggle matching and are thus correlated (Lake Suigetsu). In others they arise as a result of matching to ice core records which themselves have errors on the calendar scale, are correlated one to another and propagate through to the calibration data set (Iberian Margin, Cariaco Basin, North Atlantic). In the remainder, we have estimates of laboratory errors on the calendar ages because they derive from U-series dating, but we often do not have estimates of the errors induced by other parts of the data collection process such as the estimates of reservoir correction and the like.

In other words the horizontal axis does not represent "absolute calendar age" and we thus need to be sure to allow for uncertainty on this axis in any interpretations we make on the basis of these data. Uncertainty in the ^{14}C scale should also be considered. Even in hydrologically simple areas, marine reservoir corrections are likely to have varied over time, particularly during major climatic changes such as the last glaciation. Estimates of the uncertainty on reservoir corrections has been calculated from the overlap of the records with the tree-ring dataset as discussed in Hughen et al. (2004a). For the Bahamas and Arabian speleothem records the error in DCF has been discussed above.

Discussion of the available data

Up to 26 cal kBP, the agreement between the high temporal resolution datasets obtained from Lake Suigetsu and the Bahamian Speleothem is reasonable. The general trends of their comparison curves, including wiggles and plateaux, agree. A detailed inspection, however shows discrepancies of typically a few hundred ^{14}C years. For this reason neither of these records are included in IntCal04. Beyond 26 cal kBP, the records start deviating strongly from each other – up to several millennia. This deviation could be due either to errors in calendar age scale, or to variability in the DCF assumed constant for the speleothem record. The larger deviations observed prior to 32 cal kBP cannot be explained by DCF errors, however, as they would require negative DCF corrections. The most dramatic feature of the Bahamian Speleothem data is the large excursion (>1000‰) in $\Delta^{14}C$ at 44 cal kBP. Switches in the mode of ocean circulation are required to enable the observed abrupt and high amplitude shifts (Beck et al. 2001). Such excursions, if confirmed, would have severe consequences for any calibration effort between 40 and 50 cal kBP. This large excursion is not observed in either the Cariaco data or the Arabian speleothem record, which raises the question whether this is an artifact. In the case of the Cariaco record, the magnitude of such an excursion might be subdued because of the marine nature of the record, though this does not explain why the excursion is not observed in the Arabian speleothem.

The new, extended data from the Cariaco Basin neither confirm nor refute the reliability of either the Lake Suigetsu or Bahamas Stalagmite data. The Cariaco Basin data actually fall between the Suigetsu and Bahamas data. Note here that the Cariaco data are tied to the GISP2 ice core time scale. The Cariaco Basin record does not show large excursions in $\Delta^{14}C$, but one has to consider that this dataset represents a marine reservoir. The other data discussed here – North Atlantic foraminifera, and the sparser data collection from Lake Lisan, Pacific Corals and Iberian Margin foraminifera also fall generally within the

"envelope" that can be constructed from the Lake Suigetsu and Bahamas Speleothem data.

The new speleothem data from Arabia (Weyhenmeyer et al. 2003) range between 42 and 55 cal kBP and are not at all consistent with the data from the Cariaco Basin and North Atlantic for the same time period, but might be seen as an extension of the speleothems from the Bahamas (Beck et al. 2001) with an overlap of a few millennia.

The sources of these discrepancies between the available data sets are not currently understood, but may be due to some combination of artefacts relating to varve counting, ^{230}Th dating, uncertainties in the GISP2 chronology or correlation with GISP2 δ^{18}O, unaccounted for variability in reservoir correction or dead carbon fraction, or undetected overprinting from secondary alteration or authigenic mineral growth. For the moment, all we can do is note and describe the differences between the records and hope that in the near future we will be able to resolve them.

In the mean time, note that we should not be very surprised about these discrepancies since many of the datasets available (as collected in van der Plicht 2000a, as well as almost all datasets discussed in this paper) do not meet the criteria for use as radiocarbon calibration data as established by the IntCal04 team. Reimer et al. (2002) outline criteria for the use of the following archives for radiocarbon calibration: tree-rings, corals, carbonates (non-corals), laminated sediments and marine sediments. Since the available records all fall into one of these categories, it is instructive to summarize the "pros and cons" of the records discussed in this paper – see Table 1.

Table 1. Overview ("pros and cons") of the comparison records 26-50 cal kBP with main differences concerning calibration (i.e. BP vs. cal BP).

record	*^{14}C axis*	*calendar axis*
	BP	cal BP
Lake Suigetsu	atmospheric	varve errors unknown not multicore; no tiepoints
Bahamas Speleothem	DCF = 1470 assumed constant	U/Th dates taken as absolute but with large errors
Cariaco Basin	marine (forams) reservoir correction	GISP linked (δ^{18}O)
Corals*	marine reservoir correction	U/Th dates taken as absolute but with large errors
Iberian Margin	marine (forams) reservoir correction	GISP linked (δ^{18}O)
Arabian Speleothem	DCF DCF modelled (^{13}C)	U/Th dates taken as absolute but with large errors
North Atlantic	marine (forams) complex reservoir	GISP linked (δ^{18}O)

Lake Lisan	DCF = 1000 assumed constant	U/Th dates considered absolute but with large errors

* meet criteria as established by the IntCal04 working group

A true calibration curve is a plot of the atmospheric ^{14}C content as a function of calendar time (with an associated indication of our uncertainties) and we cannot reliably offer that yet for the period 26-50 cal kBP. Nonetheless, the IntCal04 team did consider the options available to us for estimating the underlying curve from such data and we summarize them here in the hope that they are of value to those building curves in the future.

Options for combining data to form a calibration curve beyond 26 ka

Perhaps the most obvious option is to make a "calibration envelope" encompassing all data. Unfortunately, this only yields limited information (just as the "stippled curve" that was drawn through the sparse set of coral data points, Stuiver et al. 1998) and, given all the sources of uncertainty articulated above, is unlikely ever to lead us to a useful estimate of the underlying calibration curve that we want to learn about. Alternatively, we could use a weighted averaging procedure of the sort adopted for the construction of the INTCAL98 estimate of the calibration curve but (because of the need to allow for a range of sources of calendar age uncertainty) we choose to reject such an approach for IntCal04 and so it seems desirable to do the same here. As a result, we sought an extension to the approach used for IntCal04 to help us begin to devise methods for curve construction beyond 26cal kBP too. The approach we have devised is detailed in Buck and Blackwell (2004) and is a random effects extension to the random walk model used for IntCal04.

In Figure 2 (plotted using black dotted lines which represent one standard deviation about the mean) is an estimate of the underlying curve obtained using the random effects model detailed in Buck and Blackwell (2004). The NotCal04 data clearly exhibit more complex error structure than the IntCal04 data and so the random effects model adds another level of sophistication to the statistical model to accommodate this.

We assume that each of the data sets in the range 26-50ka represents an estimate of the underlying calibration curve, but with the possibility of an offset between the "comparison curve" that we could build from any single data set and the underlying (or true) "calibration curve" that we want to learn about. In doing this, we are effectively allowing for the possibility that the corrections and error assessments that have been made for the NotCal04 data sets are not as reliable as the

equivalent assessments for the IntCal04 data sets. As noted above, such extra uncertainty might arise from artefacts relating to varve counting, ^{230}Th dating, uncertainties in the GISP2 chronology or correlation with GISP2 δ^{18}O, unaccounted for variability in reservoir correction or dead carbon fraction, or undetected overprinting from secondary alteration or authigenic mineral growth. Consequently, there is potential for offsets between individual data sets and the underlying radiocarbon calibration curve. By formally allowing for the possibility of such offsets in a tailored statistical model and by allowing them to vary for each data set over time (as in the Buck and Blackwell random effects model), we can estimate offsets as part of the curve building process and take account of them when we estimate the underlying curve. In constructing the NotCal04 curve, we included parts of the IntCal04 data from the range 24-26ka. The IntCal04 data sets are assumed to be reliable (simply because they meet the criteria for IntCal04) and thus do not have offsets from the underlying curve. Since we are also using the same methods we adopted for IntCal04 to allow for the other sources of uncertainty within each data set (outlined above), we are taking account of a wide range of sources of uncertainty most of which have been ignored in previous attempts to build curves from disparate data sets.

Some readers will be surprised to find that the resulting curve (shown with dotted black lines in Figure 2) has such a narrow range and so it is worth saying a bit more about the random effects model and the estimated offsets for each of the raw data sets from the underlying curve. In particular, we have made some assumptions to produce these results (for details see Buck and Blackwell 2004). Mostly these are the same as the ones that we made when constructing the IntCal04 curve. In addition, however, we have a parameter in the model (detailed in Buck and Blackwell, 2004) which reflects our *a priori* belief about the size of the offsets between the individual data sets and the underlying true curve. We can think of this as the *a priori* standard deviation of the offset between any single comparison data set and the underlying curve. In obtaining the black dotted curve in Figure 2, we assumed that each of the data sets under consideration is equally likely to exhibit the same size of offset and so the same *a priori* standard deviation (i.e. 1000 ^{14}C years) was taken for all. In future, we could easily use a different prior offset for each data set, but from the discussion of data above and the plot in Figure 1 we can see that (in the absence of information that one data set or another is likely to suffer from a larger offset than the others) 1000 ^{14}C years is a broadly reasonable figure. The actual (*a posteriori*) offset needed for each data set is, of course, estimated from the data as part of the process of estimating the calibration curve. Experiments showed that, in practice, the precise value we choose for the *a priori* standard deviation did not have an important impact on the estimate of the

underlying curve we obtain provided that it is of the appropriate order of magnitude.

By allowing for a) the possibility of such offsets and b) for the possibility that they might vary over time, we are acknowledging that the individual data sets may not reliably inform us about the underlying radiocarbon calibration curve and that there is thus an additional source of uncertainty which is not recorded in the database. This possibility has conventionally been ignored when trying to build radiocarbon calibration curves and this has resulted in enormous uncertainty about the underlying "true" curve. Thus, the random effects model (used to produce the estimate of the curve shown with black dotted lines in Figure 2) allows for and explicitly models possible offsets which are a major potential source of error that previous methods have had to roll into the uncertainty on the estimate of the calibration curve. We plot the means of our estimates of these offsets for the NotCal04 data sets in Figure 3 in which we can see that there is a great deal of variability (both within and between data sets) in the size of offset needed and that for some of the data sets at some points in time these offsets look to be very large (i.e. greater than 2000 years). By separating out our uncertainty about the offsets in this way, the resulting estimate of the radiocarbon calibration curve has a much smaller error envelope than we would get if we assumed that all the NotCal04 data sets reliably informed us about the "true" underlying curve.

Some further observations about the random effects curve are probably worth making to help readers understand why the error envelope is so small. Firstly, we are assuming that the observational error on potential future observations are zero (we do this, not because we believe we will obtain data with no error, but because those errors will vary and it does not seem helpful to simply guess what they would be). Secondly, the dotted black curve in Figure 2 is our current best estimate of the underlying calibration curve (with a one standard deviation error range). It arises only from the 'comparison' data sets included in the NotCal04 database and it does not tell us anything (on its own) about data sets we are likely to obtain in the future. Thirdly, as detailed in Buck and Blackwell (2004), we can also use our random effects model, the current data and the offsets in Figure 3 to help predict future data sets. Predicting individual future 'comparison' datasets in the presence of so many sources of uncertainty (and relatively few current datasets) is not a very useful or reliable exercise in its own right. What is potentially useful, however, is the fact that if we can predict individual datasets we can also estimate a predictive distribution from which future datasets are likely to be drawn. Such a distribution allows us to encompass not just the uncertainty in the data, but the uncertainty arising from the offsets too. As a consequence, it has a much larger one standard deviation range than the

estimate of the underlying curve and allows us to convey uncertainty from a range of sources all at the same time. In Figure 2 (plotted using a green shaded area) we represent our estimate of the predictive distribution by plotting a one standard deviation envelope around its mean.

Clearly, estimating the underlying calibration curve and predicting the interval within which new 'comparison' datasets will lie are two different but related tasks. We include plots of both here a) to emphasize that they are different and b) to help readers to see the benefits that can be gained when different sources of uncertainty are separated out, carefully articulated and handled appropriately. What we are most interested in is estimating the underlying calibration curve, but it is also useful to see how much variability we can expect in future 'comparison' curves and still obtain the same estimate of the underlying curve.

Conclusion

In summary, calibration of the ^{14}C time scale for the range 26-50 ka is still problematic and this should be clearly stated to the users of radiocarbon, in particular in archaeology. For example, the famous Chauvet cave in France, dates to approximately 31,000 BP, which would "calibrate" to around 31,000 BC using a comparison curve formed from the data from Lake Suigetsu, around 38,000 BC using the Bahamian Stalagmite data, and around 36,000 BC using the Cariaco Basin data (Bard 2001). Furthermore, a key issue in modern human evolution is the Late Neanderthal/Early modern human transition. This important scientific issue takes place in the time range discussed here and, when ^{14}C dates are used, erroneous conclusions can easily be made. For example, the co-existence of Late Neanderthal and Cro-Magnon humans is discussed using ^{14}C dates, calibrated with a combination of Lake Suigetsu and the North Atlantic datasets (Stringer and Davies 2001). In another example, series of Neanderthal ^{14}C dates are compared with TL/ESR dates from the site by simply subtracting 3000 years from the latter in order to "calibrate them into ^{14}C terms" (Mellars 1998). Also, a smooth, structureless "radiocarbon calibration curve" back to 45 cal kBP has been proposed recently, deduced from geomagnetic records - thus based on indirect data and ignoring wiggles (van Andel 1998; commented by van der Plicht 1999).

Calibration of ^{14}C requires calibration curves that are absolute, or very close to absolute. In this sense, calibration of ^{14}C is not possible until the detailed differences between data sets are resolved or explained in such a way that they can be (statistically) modelled and included in the curve building process. Thus, the currently available data are not to be used as calibration curves (hence the name NotCal04, as opposed to

IntCal04). Instead of calibration curve, the phrase "comparison curve" (van der Plicht 2000b; Richards and Beck 2001) is advocated.

It is obvious that there must be errors in at least some of the datasets (and/or our understanding of the errors associated with them) since by definition, there can only be one ^{14}C calibration curve. For the "classical" ^{14}C calibration archive (dendrochronologically dated tree-rings) we have seen continuous progress since the publication of the first calibration tables and graphs. With IntCal04, the tree-ring calibration curve is now reaching well into the Younger Dryas. Beyond the tree-ring limit, corals and marine varves currently provide data for a calibration curve (marine derived) back to 26,000 cal BP.

Beyond 26 cal kBP, more work is clearly needed. We advocate a) the making of independent measurements to resolve discrepancies between the current data sets and b) more thought about our options for the statistical modelling of the underlying curve.

Ideally, an archive which is truly continuous, absolute and atmospheric/terrestrial cross dated is needed, but even without this improved statistical methods along with further independent measurements may one day lead to the calibration curve we seek.

We believe that the term "radiocarbon calibration curve" should be reserved for internationally agreed curves built from undisputed data; otherwise, the term "radiocarbon comparison curve" is advocated.

Acknowledgements: A portion of this work was produced by the University of California's Lawrence Livermore National Laboratory under the auspices of the DOE contract W-7405-Eng-48.

Figure captions

Figure 1
Comparison of ^{14}C calibration datasets obtained from Lake Suigetsu (red & orange), Bahamas Speleothem (blue), Corals (black), Cariaco Basin (purple), Iberian Margin (dark green), Arabian Speleothems (brown), North Atlantic (grey) and Lake Lisan (light green). Errors plotted in the ^{14}C scale are one standard deviation from the mean ; errors on the cal BP scale are *not* plotted (see the main text for a description of the wide range of sources of uncertainty on this scale).

Figure 2
Summary plots of curves calculated in an attempt to derive the underlying "true" calibration curve. The black dotted lines represent one standard deviation about the mean of the estimate of the calibration curve ("random effects model"). The green shaded area represents one standard deviation about the mean of the predictive distribution within which new "comparison" datasets are likely to lie (so-called "Python" curve). For details of methodology see Buck and Blackwell, 2004. For illustration, selected current "comparison" datasets (same legend as Fig.1) are shown as well.

Figure 3
Plot of the estimated offsets between each of the NotCal04 comparison data sets and the random effects curve (shown with dotted lines in Figure 2). These estimates arise from the random effects calculations and help to explain why the error envelope on the associated estimate of the curve is so small when compared with those obtained by other methods which roll this possible offset into the uncertainty on the estimate of the curve.

References

van Andel TH. 1998. Middle and Upper Palaeolithic environments and the calibration of ^{14}C dates beyond 10,000 BP. *Antiquity* 72:26-33.

Bard E, Arnold M, Hamelin B, Tisnerat-Laborde N, Cabioch G. 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* 40:1085-92.

Bard E. 2001. Extending the Calibrated Radiocarbon record. *Science* 292:2443-4.

Bard, E., Ménot-Combes, G. Rostek F. 2004. Present status of Radiocarbon Calibration and Comparison Records Based on Polynesian Corals and Iberian Margin Sediments. *Radiocarbon* (this issue).

Beck, JW, Richards, DA, Edwards, RL, B. W. Silverman, Smar, PL, Donahue, DJ, Herrera-Osterheld, S, Burr, GS, Calsoyas, L, Jull, AJT, and Biddulph., D. 2001. Extremely large variations of atmospheric ^{14}C concentration during the last glacial period. *Science* **292**, 2453-58.

van der Borg K, Stein M, de Jong AFM, Waldmann N, Goldstein SL. 2004. Near-zero $D^{14}C$ values at 32 kyr calBP observed in the high-resolution ^{14}C record from U-Th dated sediment of Lake Lisan. *Radiocarbon* 46:785-95.

Buck CE and Blackwell PG. 2004. Formal statistical models for estimating radiocarbon calibration curves, *Radiocarbon* (this issue).

Burns SJ, Fleitmann D, Matter A, Kramers J, Al-Subbary AA. 2003. Indian Ocean Climate and Dansgaard/Oeschger Events 9 to 13. *Science* 301:1365-7.

Burns SJ, Fleitmann, D, Matter, A, Kramers, J, Al-Subbary AA. 2004. Corrections and clarifications. *Science* 305:1567.

Burr GS, Beck JW, Taylor FW, Recy J, Edwards RL, Cabioch G, Correge T, Donahue DJ, O'Malley JM. 1998. A high-resolution radiocarbon calibration between 11,700 and 12,400 calendar years BP derived from

Th-230 ages of corals from Espiritu Santo Island, Vanuatu. *Radiocarbon* 40:1093-105.

Cutler KB, Gray SC, Burr GS, Edwards RL, Taylor FW Cabioch G, Beck JW, Récy J, Cheng H, Moore J. 2004. Radiocarbon calibration to 50 ky BP with paired ^{14}C and ^{230}Th dating of corals from Vanuatu and Papua New Guinea. *Radiocarbon* (this issue).

Fairbanks, RG, Mortlock, RA, Chiu, T-C, Guilderson, TP, Cao, L, Kaplan, A, Bloom, A. 2004. Marine Radiocarbon Calibration Curve Spanning 7,000 to 50,000 Years B.P. Based on Paired $^{230}Th/^{234}U/^{238}U$ and ^{14}C Dates on Pristine Corals. *Quaternary Science Reviews*, submitted.

Hughen KA, Overpeck JT, Lehman SJ, Kashgarian M, Southon JR, Peterson LC, Alley R, Sigman DM. 1998a. Deglacial changes in ocean circulation from an extended radiocarbon calibration. *Nature* 391:65-8.

Hughen KA, Overpeck JT, Lehman SJ, Kashgarian M, Southon JR, Peterson LC. 1998b. A new ^{14}C calibration data set for the last deglaciation based on marine varves. *Radiocarbon* 40:483-94.

Hughen, KA, Baillie, MGL, Bard, E, Bayliss, A, Beck, JW, Bertrand, CJH, Blackwell, PG, Buck, CE, Burr, GS, Cutler, KB, Damon, PE, Edwards, RL, Fairbanks, RG, Friedrich, M, Guilderson, TP, Kromer, B, McCormac, FG, Manning, SW, Ramsey, CB, Reimer, PJ, Reimer, RW, Remmele, S, Southon, JR, Stuiver, M, Talamo, S, Taylor, FW, van der Plicht, J, and Weyhenmeyer, CE. Marine04 Marine radiocarbon age calibration, 26 - 0 ka BP. *Radiocarbon* **46** (this issue a).

Hughen KA, Lehman S, Southon J, Overpeck J, Marchal O, Herring C, Turnbull J et al. 2004b. C-14 activity and global carbon cycle changes over the past 50,000 years. *Science* 303 (5655): 202-207.

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

Kitagawa H, van der Plicht J. 2000. Atmospheric radiocarbon calibration beyond 11,900 cal BP from Lake Suigetsu laminated sediments. *Radiocarbon* 42:369-80.

van Kreveld S, Sarntheim M, Erlenkeuser H, Grootes PM, Jung S, Nadeau MJ, Pflaumann U, Voelker A. 2000. *Paleoceanography* 15:425-42.

Libby WF, 1955. Radiocarbon Dating, Chicago, University of Chicago Press, 2nd ed.

Meese DA, Gow AJ, Alley RB, Zielinski GA, P.M. Grootes, Ram M, Taylor KC, Mayewski PA, Bolzan JF. 1997. *Journal of Geophysical Research* 102,C12:26411-24.

Mellars P. 1998. The fate of the Neanderthals. *Nature* 395:539-40.

van der Plicht J. 1999. Radiocarbon calibration for the Middle/Upper Palaeolithic: a comment. *Antiquity* 279:119-23.

van der Plicht J. 2000a (ed.). Varve/Comparison Issue. *Radiocarbon* 42: 313-452.

van der Plicht J. 2000b. Introduction: the 2000 radiocarbon varve/comparison issue. *Radiocarbon* 42: 313-22.

Reimer PJ, Hughen KA, Guilderson TP, McCormac FG, Baillie MGL, Bard E, Barratt P, Beck WJ, Buck CE, Damon PE, Friedrich M, Kromer B, Ramsey CB, Reimer RW, Remmele S, Southon JR, Stuiver M, van der Plicht J. 2002. Preliminary Report on the First Workshop of the IntCal04 Radiocarbon Calibration/Comparison Working Group. *Radiocarbon* 44:653-61.

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Herring C, Hughen KA, Kromer B, McCormac G, Manning S, Bronk Ramsey C, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer C. 2004. IntCal04 Terrestrial radiocarbon age calibration, 26 – 0 ka BP. *Radiocarbon* (this issue).

Richards DA, Beck JW. 2001. Dramatic shifts in atmospheric radiocarbon during the last glacial period. *Antiquity* 289:482-5.

Richards DA, Beck JW, Smart P, Hoffman D, Mattey D, Hawkesworth C. 2003. Speleothems from the Bahamas and atmospheric radiocarbon during the last glacial period: recent developments. 18th International Radiocarbon Conference, Wellington, Abstracts p.81.

Schramm A, Stein M, Goldstein SL. 2000. Calibration of the C-14 time scale to > 40 ka by U-234-Th-230 dating of Lake Lisan sediments (last glacial Dead Sea). *Earth and Planetary Science Letters* **175**, 27-40.

Stringer C, Davies W. Those elusive Neanderthals. *Nature* 413:791-2.

Suess HE. 1970. Radiocarbon variations and absolute chronology. Proceedings of the XII Nobel Symposium. New York, Wiley, 303-11.

Stuiver M, van der Plicht J. 1998 (eds.). INTCAL98: Calibration Issue. *Radiocarbon* 40:1041-1159

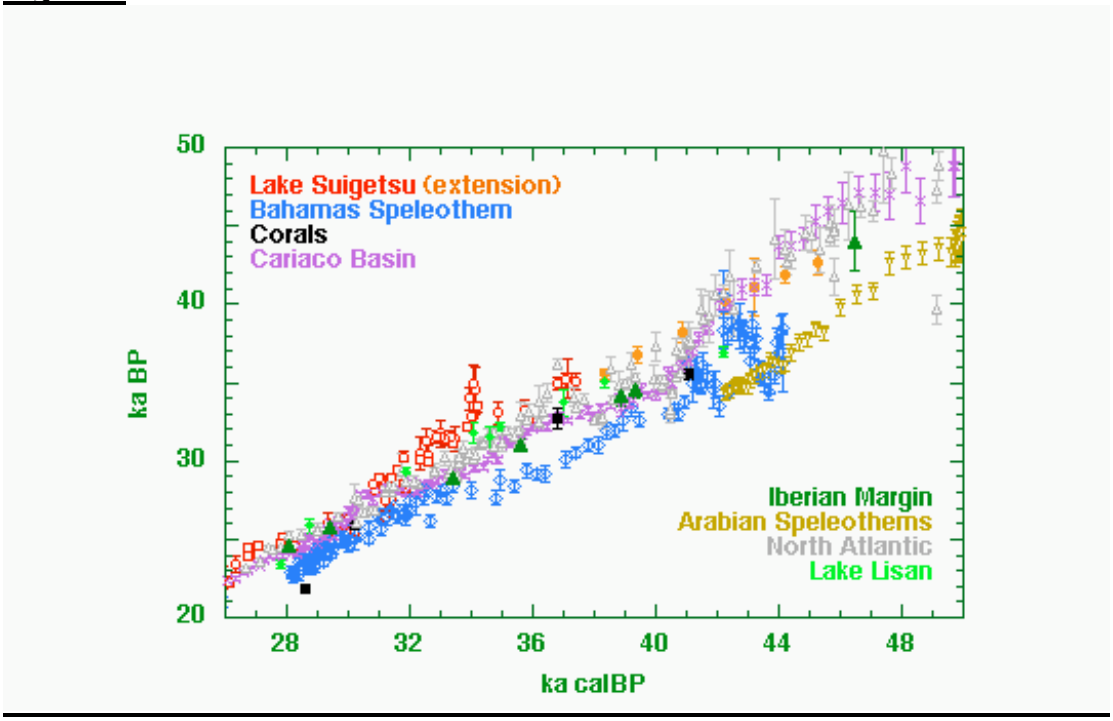
Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac G, van der Plicht J, and Spurk M. 1998. INTCAL98 radiocarbon age calibration, 24,000-0 cal BP. *Radiocarbon* 40:1041-83.

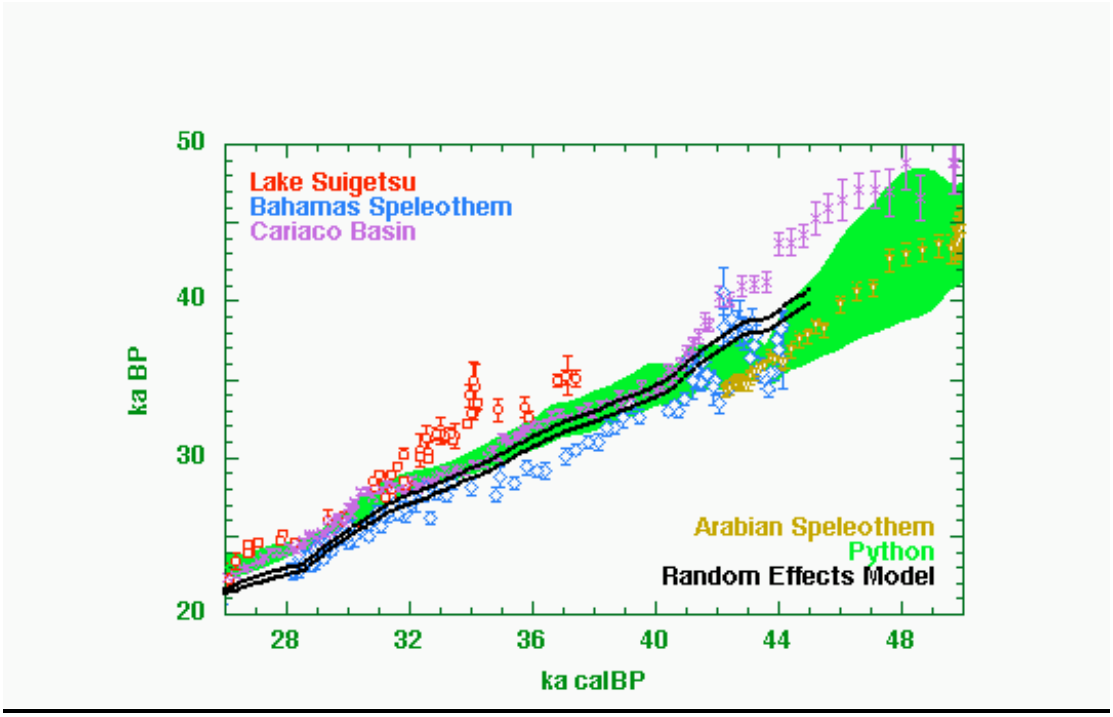
Voelker AHL, Grootes PM, Nadeau MJ, Sarntheim M. 2000. Radiocarbon levels in the Iceland sea from 25-53 kyr and their link to the Earth's magnetic field intensity. *Radiocarbon* 42:437-52.

Vogel JC, and Kronfeld J. 1997. Calibration of Radiocarbon dates for the Late Pleistocene using U/Th dates on stalagmites. *Radiocarbon* 39:27-32.

Weyhenmeyer CE, Burns S, Fleitmann D, Kramers, JD, Matter, A, Waber, HN, Reimer PJ. 2003. Changes in atmospheric ^{14}C between 55 and 42 ky BP recored in a stalagmitefrom Socotra Island, Indian Ocean. EOS Trans. AGU, 84 (46) Fall Meet. Suppl. Abstract PP32B-0298.

Figure1


Lake Suigetsu (extension)
Bahamas Speleothem
Corals
Cariaco Basin
Iberian Margin
Arabian Speleothems
North Atlantic
Lake Lisan
ka BP
ka calBP
50
40
30
20
28
32
36
40
44
48


Lake Suigetsu
Bahamas Speleothem
Cariaco Basin
Arabian Speleothem
Python
Random Effects Model
ka BP
ka calBP
50
40
30
20
28
32
36
40
44
48