Random Snapshot on Radiocarbon Laboratory Dating Accuracy from a small Double Blind Test on Known Age Conifer Samples (and an issue regarding pretreatment).

Introduction

A recurrent question that archaeologists (especially) ask about radiocarbon (^{14}C) dates runs along the lines of: do we really know (referring to some set of a few dates from site or context *x*) that these dates are accurate? Some laboratories, notably Oxford, have published details of known ages tests run over given years of operation (e.g. Bronk Ramsey et al. 2002:1-4; 2004), and these reasonably inspire user confidence over the reliability of their measurements given their good performance. But, in many cases, such transparent data are either available only within the ^{14}C community, or on enquiry, or are not available at all. Nor of course (and unfortunately) is there a standard international radiocarbon databank where users might find and compare data (the failure of the original plan for a International Radiocarbon Data Bank – see e.g. Kra 1989 – to become a permanent repository and core element of the international radiocarbon community is a great failure which should be remedied as soon as possible). There is also the question of samples that need more pretreatment than the relatively radiocarbon-friendly oak (*Quercus sp.*) samples used, e.g., in the Oxford known-age tests. Although in the British Isles-European context much more oak is typically encountered – more attention is turning to conifers in recent work and for the medieval-early modern era (e.g. Tyers et al. 2009), and conifer-based data is of course relevant in many other regions of the world. As noted for many years, conifer samples require more extensive pretreatment to remove extraneous material from resins through lignins (since variations of the standard de Vries, AAA, pretreatment only remove about 90% of these materials: e.g. Stuiver and Quay 1981; generally on the topic, see Hoper et al. 1998), and thus it is important to know if radiocarbon dates on such samples are accurate (also Tyers et al. 2009: 386). Do the various solvent extraction and pretreatment methods employed on such conifer samples by different laboratories yield broadly consistent results, or are potential accuracy issues sometimes inadvertently introduced by these pretreatment methods?

With the support and participation of Henry Zemel and the CAENO Foundation (http://www.caeno.org), it was therefore decided that an interesting experiment would be to try a small double-blind ‘snapshot’ test on the dating of some known age conifer (*Pinus leukodermis*) samples with three Accelerator Mass Spectrometry (AMS) laboratories. Five samples were submitted to each laboratory – the sort of modest number of samples an archaeologist often submits, and is to rely on, whether to date a building or site or a specific context at a site (for example, compare the typical number of samples submitted or dates produced from the sites in the most recent Archaeometry datelist from the Oxford AMS system at the time of writing: Higham et al. 2011). Our aim was not to test especially difficult or variable or esoteric sample material – rather to test a sample typical of one large category of potential material submitted by archaeologists – conifer wood – which does however come with a little more complications than oak as noted above. The aim of the exercise was to gain some idea of, and, hopefully, reassurance about, the reliability of radiocarbon dating of random samples of conifer. This was a small test and with only three laboratories, but it provides a ‘snapshot’ test of dating accuracy from the viewpoint of a user. The dates were run on a standard basis: i.e. the samples were submitted and dated, and the user then billed for the regular charges per date which apply for each of the laboratories.

Methods

Known age tree-rings from three calendar intervals (of 10-years each) were carefully dissected from a section of wood from a *Pinus leukodermis* sample (from northern Greece from approximately 1600m elevation: PPG 3B from around 20km NNW of Metsovo in northern Greece). The calendar intervals employed were AD1269-1278, AD1420-1429 and AD1641-1650. The intervals were selected by Manning as they lie on distinct slopes in the radiocarbon calibration curve such that if the ^{14}C dates produced are accurate then they should clearly yield the correct calendar age range (without multiple intercepts/ambiguity issues: see Figure 1). The samples were placed into three bags (bag 1 = AD1420-1429, bag 2 = AD1641-1650 and bag 3 = AD1269-1278). Each sample was then given an alphabetic code and its own plastic sample bag by Manning and the original bag (1-3) it came from recorded by Manning. The 15 alphabetically only identified samples were then given to Zemel who randomly reassigned a different alphabetic code to each without telling Manning, with Zemel keeping a list of the original alphabetic codes and the newly assigned ones – witnessed by Barbara London. The newly labeled samples in separate plastic bags were then returned to Manning who sent five randomly selected bags with no information other than the alphabetic code and the fact the samples were pine wood to each of three laboratories for AMS ^{14}C dating: Beta Analytic (http://www.radiocarbon.com/), Oxford (http://c14.arch.ox.ac.uk/), and VERA (http://isotopenforschung.univie.ac.at/). The laboratories then ran the samples according their inspections of the samples received and procedures for such wood samples (the Beta report stated the pretreatment was acid/alkali/acid – so standard AAA; the Oxford pretreatment is discussed below). Two of the laboratories simply did as asked, and ran and returned 5 radiocarbon measurements on the five samples sent (Beta and Oxford). The VERA laboratory decided to run the samples twice using different pretreatment procedures. The VERA 'A' measurements come from their standard ABA pretreatment; the VERA 'B' measurements are after Soxhlet extraction and then processing to cellulose. Hence we end up with 20 data in the primary 'blind' part of this study. Results were then returned to Manning by each of the laboratories, and at a subsequent meeting Zemel provided his list of alphabetic codes and these were matched against the original lists of Manning to determine which samples were which. Figure 1 shows the calendar place of the 3 sample groups (bags 1-3) versus the IntCal09 radiocarbon calibration curve (Reimer et al. 2009) and the Manning, and then Zemel, alphabetic codes for the samples in each of these groups.

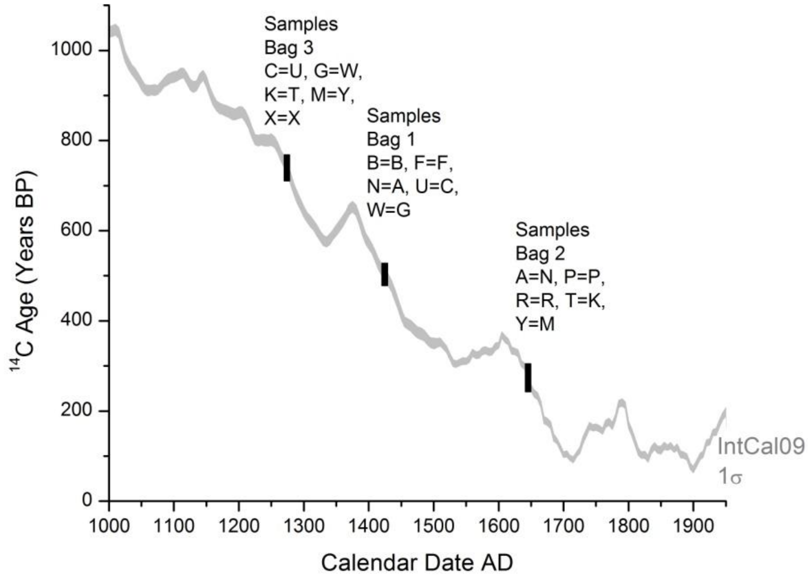

Figure 1. The 3 sample decades AD1269-1278 (bag 3), AD1420-1429 (bag 1) and AD1641-1650 (bag 2) shown in terms of placement on the IntCal09 (Reimer et al. 2009) international northern hemisphere radiocarbon calibration curve. The alphabet codes indicate (first letter) the 15 random letter codes applied to the samples by Manning, and then (second letter) the random re-coding of each sample by Zemel. The identity of each sample was only known by Manning and Zemel (or anyone else) after the experiment was completed when lists of codes were matched together.

Results

The radiocarbon measurements obtained from the three laboratories for the samples are shown in Figure 2 and Listed in Table 1. The ^{14}C ages obtained for each bag set are then shown separately against the modeled IntCal09 calibration curve and also against both the raw ^{14}C data on known age tree-ring samples from which IntCal09 was constructed (Reimer et al. 2009), and – where available – known age ^{14}C data on German Oak and Turkish Pine (Kromer et al. 2010; Manning et al. 2010) in Figures 3-5. Five measurements from a previous known-age test on a decade sample of Turkish Pine from AD 1640-1649 run by the VERA laboratory as part of the study reported in Manning et al. (2006: see Figure 1 lower right) are also included in the 17th century plot in Figure 3 for comparison. The results of this blind-test are then summarized in Figure 6 which shows the scatter of offsets from the central IntCal09 value for each sample (see also Figure 2 bottom) and so the biases evident in the blind-test dataset. By way of comparison, Figure 6 also shows the scatter of offsets between another recent set of AMS measurements on known age conifer samples and IntCal09 (data from Tyers et al. 2009).

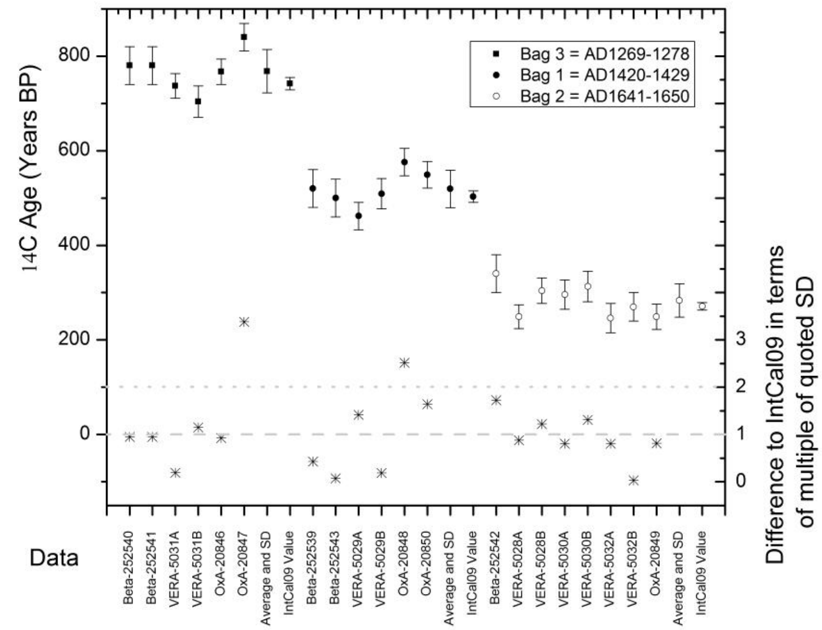

Figure 2. Top: the ^{14}C ages measured on the samples from (left to right) the bag 1 samples, then the bag 2 samples and finally the bag 3 samples. Data from Table 1; 1σ errors shown. The (non-weighted) average ^{14}C age and standard deviation (SD) for each bag set is also shown along with the value interpolated from IntCal09 for the mid-point of the decade represented by the samples in each bag (linear interpolation). Bottom: the difference in terms of multiples of the quoted SD for each measurement between the measured ^{14}C age reported and the IntCal09 central ^{14}C value (see also Figure 6).

	Lab ID	**δ^{13}C ‰**	**SD (if reported)**	**^{14}C Age BP**	**SD**	
Bag 3	Beta-252540	-21.2		780	40	
	Beta-252541	-21.6		780	40	
	VERA-5031A	-23.6	0.5	737	26	
	VERA-5031B	-23.7	0.9	704	33	
	OxA-20846	-23.08	0.3	767	27	
	OxA-20847	-23.62	0.3	840	29	
	Bag 3 Weighted Average			767	13	BUT: T = 11.55 > χ^2 df5 =11.1 at 5% level
	IntCal09 Value			742	13	
Bag 1	Beta-252539	-21.7		520	40	
	Beta-252543	-21.8		500	40	
	VERA-5029A	-24.1	0.6	462	29	
	VERA-5029B	-21.5	0.7	509	32	
	OxA-20848	-21.72	0.3	576	29	

	OxA-20850	-21.12	0.3	549	28	
	Bag 1 Weighted Average			522	14	T = 9.1 < χ^2 df5 =11.1 at 5% level
	IntCal09 Value			503	12	
Bag 2	Beta-252542	-21.9		340	40	
	VERA-5028A	-24.1	0.5	249	25	
	VERA-5028B	-23.4	0.5	304	27	
	VERA-5030A	-26	0.7	296	31	
	VERA-5030B	-23.8	0.6	313	32	
	VERA-5032A	-25.9	0.9	246	31	
	VERA-5032B	-23.5	0.6	270	30	
	OxA-20849	-21.88	0.3	249	27	
	Bag 2 Weighted Average			278	11	T = 8.4 < χ^2 df7 =14.1 at 5% level
	IntCal09 Value			271	8	

Table 1. Results returned for the submitted samples from bags 1-3. A Chi-square test on the hypothesis that all the data for each bag (same 10 years of tree-rings) are consistent with being the same age within 95% probability (so agreement at better than the 5% level) from Ward and Wilson (1978) is also shown.

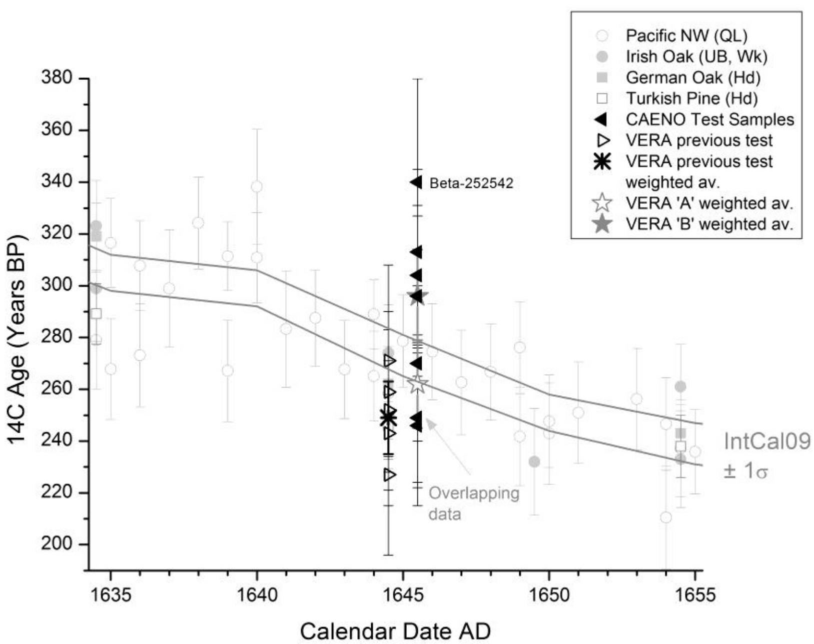

Figure 3. Bag 2 blind-test data ("CAENO Test Samples") shown against IntCal09 (1σ band), the raw ^{14}C data on known age tree-ring samples from which IntCal09 was constructed, ^{14}C data on known age samples of German Oak and Turkish Pine, and a previous known age test run on a decade of wood with mid-point 1 year older by the VERA laboratory (Manning et al. 2006: Figure 1 lower right). The overall bag 2 weighted average is 278±11 ^{14}C years BP versus the approximate value of 271±8 ^{14}C years BP from IntCal09. The VERA 'A' weighted average is 262±17 ^{14}C years BP, the VERA 'B' weighted average is 296±18 ^{14}C years BP, and the previous

VERA known age test weighted average was 249±14 ^{14}C years BP. All error bars shown are 1σ. Beta-252542, >1.7x multiples of its quoted error away from the IntCal09 central value, despite its already generous (±40 ^{14}C years) quoted error, is labeled.

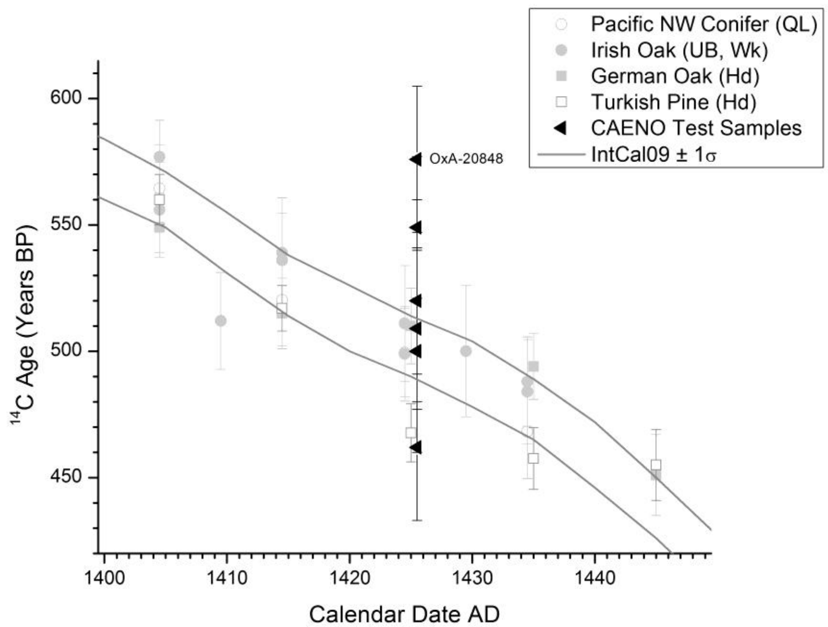

Figure 4. Bag 1 blind-test data ("CAENO Test Samples") shown against IntCal09 (1σ band), the raw ^{14}C data on known age tree-ring samples from which IntCal09 was constructed, and ^{14}C data on known age samples of German Oak and Turkish Pine. OxA-20848, >2.5 multiples of its quoted error away from the IntCal09 central value, is labeled.

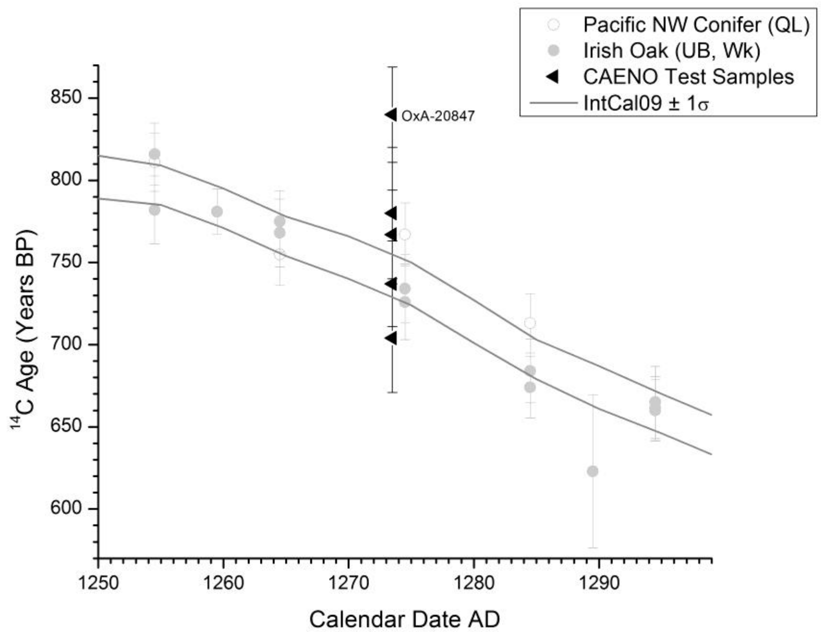

Figure 5. Bag 3 blind-test data ("CAENO Test Samples") shown against IntCal09 (1σ band) and the raw ^{14}C data on known age tree-ring samples from which IntCal09 was constructed. OxA-20847, >3.3 multiples of its quoted error away from the IntCal09 central value, is labeled. Primarily because of OxA-20847, the bag 3 data are not compatible at 95% probability with the hypothesis that they all belong to the same real radiocarbon age (see Table 1).

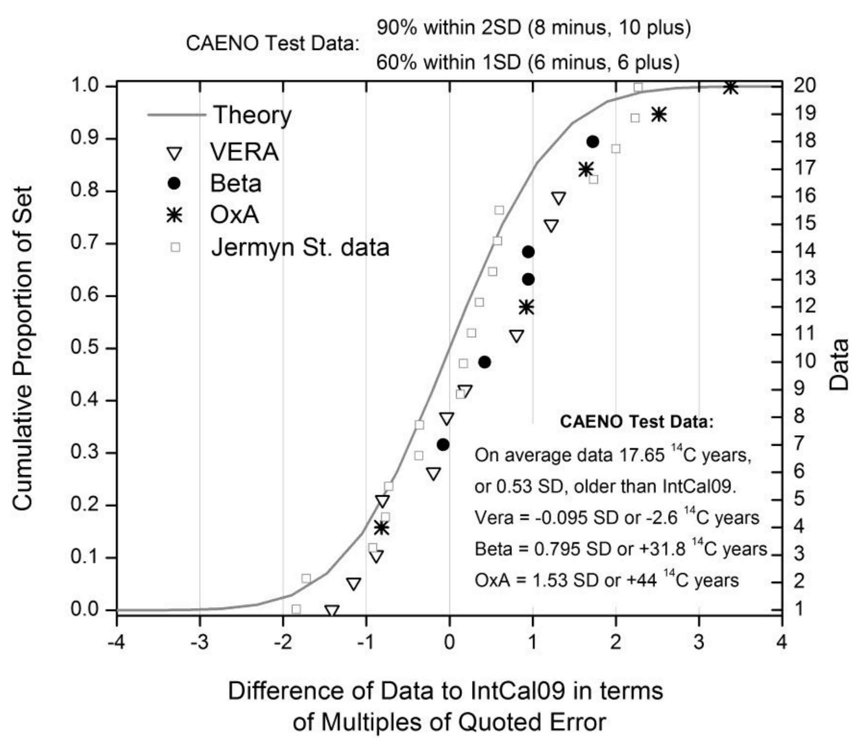

Figure 6. Blind date test set results by laboratory showing the offset from the central IntCal09 value for the mid-point of the decade dated for each sample in terms of the multiple of the quoted measurement error. The VERA data indicate very little bias. If we further break this down into the VERA ‘A’ and VERA ‘B’ measurements (see above), the VERA ‘A’ set has an average difference of -0.498 SD or -13.6 ^{14}C years and the VERA ‘B’ set has an average difference of 0.307 SD or +8.4 ^{14}C years. The Beta and Oxford data indicate some bias to older ^{14}C ages (two Oxford measurements in particular: OxA-20847 and OxA-20848) – see also Figure 2. Grey squares: these plot the equivalent offsets in terms of the multiple of the quoted measurement error from the central IntCal09 value for the single-year data on *Pinus sylvestris* samples from Jermyn Street (Tyers et al. 2009) – these data offer AMS measurements on known age conifer samples and thus an analogue for the blind test data in this paper.

Discussion

In general terms performance in this blind known-age test on a conifer sample was reasonable. If the ^{14}C measurements obtained are calibrated (with IntCal09) then only 3 of 20 dates do not include even part of the known age range within their most likely 68.2% calibrated ranges, and only two of these dates do not include any part of the known age range in their 95.4% calibrated range: Figure 7.

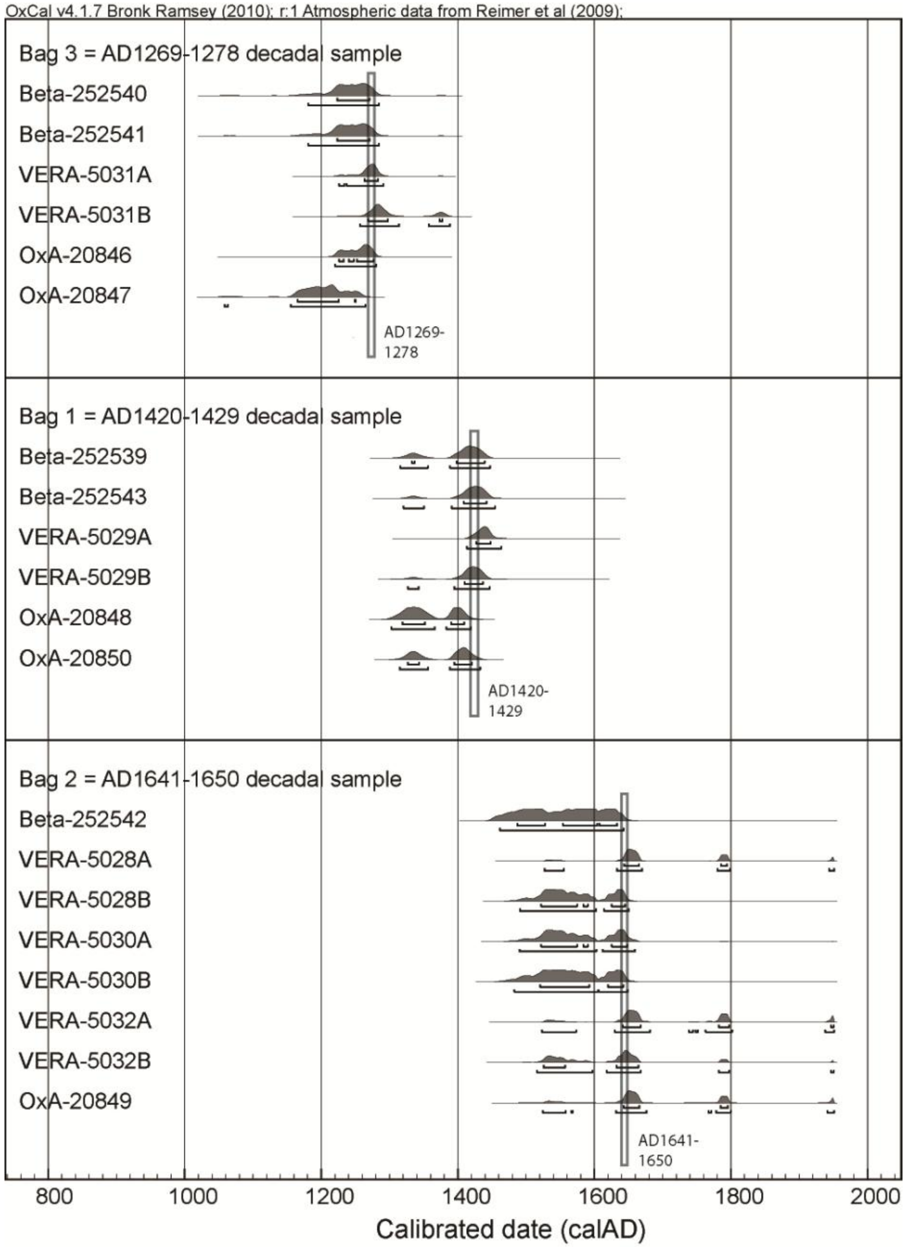

Figure 7. Calibrated calendar age ranges and probability histograms for the ^{14}C data in Table 1 from OxCal (Bronk Ramsey 1995; 2009 – curve resolution set at 1) and IntCal09 (Reimer et al. 2009). The upper and lower lines under the histograms indicate the 68.2% and 95.4% probability ranges respectively. Bag 3: OxA-20847 fails to include the known age in either its 95.4% or 68.2% calibrated ranges. All the other samples include at least part of the known age range in both the 68.2% and 95.4 calibrated ranges. Bag 1: OxA-20848 fails to include the known age in either its 95.4% or 68.2% calibrated ranges (missing the former by 1 year); all the other samples include at least part of the known age range in both the 68.2% and 95.4 calibrated ranges (with OxA-20850 overlapping for just 1 year at 68.2% probability). Bag 2: Beta-252542 fails to include the known age in its 68.2% calibrated ranges, but scrapes a 2-year overlap for its 95.4% calibrated range; all the other samples include at least part of the known age range in both the 68.2% and 95.4 calibrated ranges.

The data from the VERA laboratory are impressively accurate and precise in this study. If there is an observation to make it is that the ‘B’ samples with additional pretreatment – aimed at removing possibly extraneous material from conifer samples like those submitted – seem to lead typically to slightly older ages in 4 of 5 cases (and, on average across the 5 cases, to +22 ^{14}C years, so around 3‰, with a SD of c.35 ^{14}C years). Either the VERA ‘A’ or ‘B’ data (in isolation) offer an accurate and precise ^{14}C wiggle-match: Table 2.

However, beyond the VERA data, some biases are evident in the data obtained in our blind test. Although 60% of the ^{14}C ages obtained were within 1SD of the IntCal09 central value, and 90% within 2SD (simply using the quoted errors for each measurement), there is an old-age bias evident in the set as a whole, and this is caused largely by the Oxford and Beta data (see especially Figure 6). If we compare our blind test data against another set of the known age AMS data run on conifer samples from the 16th-17th centuries AD from the Tyers et al. (2009) study – see Figure 6, grey squares – then we see that although there is scatter in the data in the Tyers et al. (2009) study, and two (11%) of the data are more than 2x their measurement error distant from the central IntCal09 value, nonetheless, these data are less biased as a set compared to our blind test data in this study. The Tyers et al. (2009) data are offset on average by -7.1 years, versus the +17.65 years for the blind test dataset, and all outliers are less than 2.27x their quoted measurement error. The individual laboratories in the Tyers et al. (2009) study show a range of offsets, but they are rather smaller than those identified in the blind test for Beta and Oxford (+31.8 and +44 ^{14}C years respectively): Groningen -19.83 ^{14}C years, Oxford -9.17 ^{14}C years, and SUERC 10.17 ^{14}C years.

In the blind test four of the five Beta data are older than the IntCal09 value but – helped by the much larger quoted error of (routinely) ±40 ^{14}C years on each of these measurements – all are within 2SD of the IntCal09 central value and 4 are within 1SD. The results are thus reasonable within the stated errors. A wiggle-match of the Beta data in isolation yields a placement which is too old by 12.5 calendar years: see Table 2. The correct mid-point age is outside the SD on the mean of the calculated distribution, but only by 3.5 years. Two of the Oxford data are more than 2x their quoted error from the IntCal09 central value and 4 of 5 of the Oxford data are older than the IntCal09 central value with an average factor of over 1.5x the quoted errors or +44 ^{14}C years. The higher-precision errors quoted for these Oxford data highlight this issue. If the 5 Oxford data are ^{14}C-wiggle-matched in isolation, then a notably poor result emerges: see Table 2. The wiggle match fails a Chi-squared test at 95% level (T=7.26 > χ^2 value at 5% for df2 of 6.0) and yields a poor OxCal Acomb value of 27.7 <An 40.8. The best fit is -19.5 calendar years from the correct fit (although the correct age just sneaks in within the large SD around the mean of the calculated distribution). A problem is clearly indicated, even in terms of the Oxford set in isolation.

	Midpoint, known age date AD	**Wigglematch Placement, date AD μ±σ**	**Difference (calendar years)**	**OxCal Agreement**	**OxCal Acomb**
Oxford Data - fails χ^2: T=7.26 > 6.0 at df2 at 5% level					
AD1268-1279 Decade	1273.5	1254±20	-19.5	74.6	
AD1420-1429 Decade	1224.5	1405±20	-19.5	94.8	
AD1641-1650 Decade	1645.5	1626±20	-19.5	15.4	27.7 < An 40.8

Beta Data					
AD1268-1279 Decade	1273.5	1261±9	-12.5	114.4	
AD1420-1429 Decade	1224.5	1412±9	-12.5	97.9	
AD1641-1650 Decade	1645.5	1633±9	-12.5	87.9	99.1 < An 40.8
VERA 'A' Data					
AD1268-1279 Decade	1273.5	1277±4	3.5	128.5	
AD1420-1429 Decade	1224.5	1428±4	3.5	90.8	
AD1641-1650 Decade	1645.5	1649±4	3.5	132	128.3 < An 40.8
VERA 'B' Data					
AD1268-1279 Decade	1273.5	1271±4	-2.5	83.3	
AD1420-1429 Decade	1224.5	1422±4	-2.5	140	
AD1641-1650 Decade	1645.5	1643±4	-2.5	136.6	130.8 < An 40.8

Table 2. Wiggle-match results for the data from each laboratory (with the VERA data separated out into 'A' and 'B' data – see text above) employing the OxCal D_Sequence function (Bronk Ramsey et al. 2001) and IntCal09 with curve resolution set at 1. Where there were two or more measurements for the exact same sample (decade) the weighted average is employed (from OxCal's R_Combine function from Ward and Wilson 1978). The OxCal agreement index values for each decade placement should be ≥60 (the one exception is shaded grey). The OxCal Acomb value should be higher than the An value shown and the one exception is shaded grey.

The Oxford result especially is surprising, as extensive known age testing by the laboratory over the same period on British Isles oak yielded almost no bias (C Bronk Ramsey, personal communication), and the Oxford data in the Tyers et al. (2009) study on conifer samples fairly similar to the blind test samples also indicated a good outcome. This leads to the hypothesis that the some part of the pretreatment applied to the blind-test samples at Oxford, because they were conifer, has in fact led to older ^{14}C ages – and to (incorrectly) too old ages divergent from IntCal09. We may already note that the 'B' pretreatment VERA data are typically a little older, but nonetheless in good agreement with the IntCal09 central value. In the Tyers et al. (2009) study we are told (p.387) that "The samples dated by AMS at the Oxford Radiocarbon Accelerator Unit were prepared following the AAA protocol with additional bleaching to holocellulose (T Higham, personal communication) and dated as described by Bronk Ramsey et al. (2004)." A key question to be discussed in the next section of the paper – after consultation with the laboratories and their seeing the results reported above – is whether a different pretreatment was followed in our blind test by Oxford, and whether a specific problem can be identified.

At this point in the present study, Manning contacted each of the laboratories and shared the findings above, and asked for comments. In particular, it seemed that a possible problem with the Oxford pretreatment for these particular samples had been identified. How the Beta samples had been pretreated is also relevant since these data also seem to lie on the older side. The next section presents the outcome of these discussions ……

Discussion with Laboratories and Input regarding Issues

VERA – comments….**

Beta – comments….**

Oxford – comments….**

Conclusions

References

Bronk Ramsey C. 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2): 425-430.

Bronk Ramsey C. 2009. Bayesian Analysis of Radiocarbon Dates. *Radiocarbon* 51(2): 337-360.

Bronk Ramsey C, Higham T, and Leach P. 2004. Towards high-precision AMS: progress and limitations. *Radiocarbon* 46(1): 17-24.

Bronk Ramsey C, Higham TFG, Owen DC, Pike AWG and Hedges REM. 2002. Radiocarbon dates from the Oxford AMS System: Datelist 31. *Archaeometry* 44(3) Suppl. 1: 1-149.

Bronk Ramsey C, van der Plicht, J, Weninger, B. 2001. 'Wiggle matching' radiocarbon dates, *Radiocarbon* 43(2A): 381-389.

Higham TFG, Bronk Ramsey C, Brock F, Baker D, Ditchfield P. 2011. Radiocarbon Dates from the Oxford AMS System: *Archaeometry* Datelist 34. *Archaeometry* 53(5): 1067-1084.

Kra R. 1989. The International Radiocarbon Data Base: A Progress Report. *Radiocarbon* 31: 1067-1076.

Kromer B, Manning SW, Friedrich M, Talamo S, and Trano N. 2010. ^{14}C calibration in the 2nd and 1st millennia BC - Eastern Mediterranean Radiocarbon Comparison Project (EMRCP). *Radiocarbon* 52: 875-886.

Manning SW, Kromer B, Bronk Ramsey C, Pearson CL, Talamo S, Trano N and Watkins JD. 2010. ^{14}C Record and Wiggle-Match Placement for the Anatolian (Gordion Area) Juniper Tree-Ring Chronology ~1729 to 751 Cal BC, and typical Aegean/Anatolian (growing season related) regional ^{14}C offset assessment. *Radiocarbon* 52(4): 1571-1597.

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Burr GS, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kaiser KF, Kromer B, McCormac FG, Manning SW, Reimer RW, Richards DA, Southon JR, Talamo S, Turney CSM, van der Plicht J, Weyhenmeyer CE. 2009. IntCal09 and Marine09 Radiocarbon Age Calibration Curves, 0–50,000 Years cal BP. *Radiocarbon* 59(4):1111-1150.

Stuiver M, Quay PD. 1981. A 1600-year-long record of solar change derived from atmospheric ^{14}C levels. *Solar Physics* 74:479-81.

Tyers C, Sidell J, van der Plicht J, Marshall P, Cook G, Bronk Ramsey C and Bayliss A. 2009. Wiggle-Matching Using Known-Age Pine from Jermyn Street, London. *Radiocarbon* 51(2): 385-396.

Ward GK, Wilson SR. 1978. Procedures for Comparing and Combining Radiocarbon Age Determinations – A Critique. *Archaeometry* 20*:* 19-31.