SOME ASPECTS OF THE ACCURACY OF CARBON-14 DATING

Hugh McKerrell, National Museum of Antiquities of Scotland

This short communication arises directly from discussion during the Forum and is a brief attempt at outlining the essential features pertaining to the accuracy of carbon-14 dating. During the meeting a number of speakers introduced carbon-14 dates at various points in their papers and the use and accuracy of both the dates and derived chronological differences was the subject of much debate. This is not as it should be - a carbon-14 date along with its associated standard deviation is a fairly well defined scientific parameter which on its own should not provoke dissent. Related matters which are much more debatable would concern sampling and archaeological contexts and these we shall return to later. But for the moment it would be useful to consider first those aspects of statistics and measurement which together provide the data under consideration.

Statistical Error and Probability

Experimentally carbon-14 dating involves measurement of the number of C-14 disintegration occurring over a measured interval of time in a prepared sample of known weight. For accuracy, at least 10,000 such disintegrations (counts) are usually measured and with the usual small samples available this takes about twenty hours on material several thousand years old. The accuracy of a date is usually represented by the standard deviation (σ) associated with the overall measurement and the main contribution derives directly from the number of counts obtained - the standard deviation of this figure is simply calculated by raising it to the power one half. Thus for the case mentioned (10,000 counts) the standard deviation is equal to $(10,000)^{\frac{1}{2}}$ or 100. This latter value is exactly one per cent of the number of counts and it is the desire to achieve this level of precision in the resulting calculated date that dictates such lengthy measurements.

It is important to stress that the carbon-14 date as usually provided is not necessarily the <u>true</u> carbon-14 date (we shall discuss the difference between this latter and a dendrochronological or calendar date later). Repetitive measurement and calculation would provide a range of values and only by averaging say one hundred of these could we sensibly offer a true C-14 date. At this point it is necessary to introduce the normal (Gaussian) curve since this describes exactly the distribution about a true mean of randomly generated data such as the one hundred carbon-14 dates mentioned (Fig. 1).

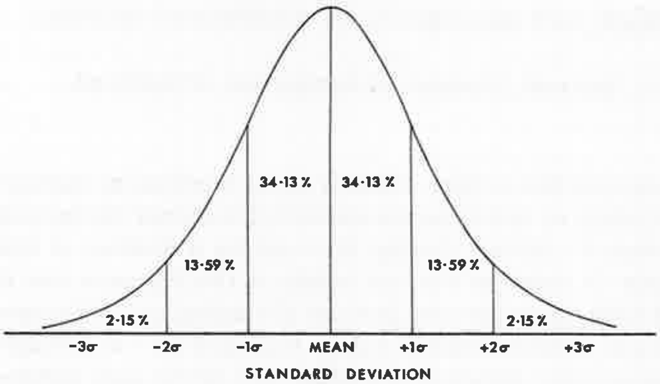

Figure 1. Percentages under the normal curve at various standard deviation units from the mean.

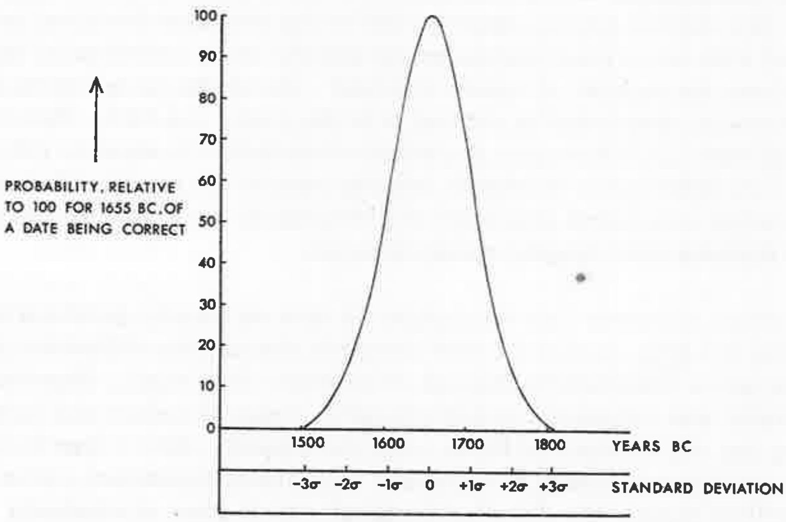

Figure 2. Leki male date 1655 ± 50 BC - Probability of a date being correct.

Based as they are upon radioactive decay C-14 dates are normal data and if we were to plot out these one hundred values we should expect to find that this would be exhibited in their distribution about the mean or true C-14 date. The area under a normal curve between any chosen range of σ values, when expressed as a percentage of the total area, is the per cent probability of data being between those limits of σ. Thus we should expect 68 of our 100 date measurements to fall between -1σ and $+1\sigma$ and 95 to fall between -2σ and $+2\sigma$. These, and percentages for any other σ ranges, are more exactly tabulated in standard works[1]. Clearly then there is a 68 per cent probability that any one date will fall between $\pm 1\sigma$ and (by difference) a 27 per cent probability for ± 1 to 2σ. It is fairly clear where this line of reasoning leads - that it is entirely feasible and correct to predict the probability of a date lying within any (narrow or wide) range of standard deviation and thus to calculate the relative probability for any two ranges (for example that it is 68/27 = 2.4 times more probable that any one date of the group considered will fall between 0-1σ compared with 1-2σ). However, before proceeding to these considerations, it is necessary to examine the data as we normally receive it, i.e. not one hundred related dates but merely one - no mean or true C-14 date - and a standard deviation that is relevant <u>not</u> to the mean or true value but simply to the one date provided. If first we examine the standard deviation of the mean of one hundred dates and each individual σ we find that both are in fact very similar (both in theory and by experimental measurement)[2, 3] and it introduces no significant error to take the one experimental σ to be equal to that which would have been obtained from one hundred or so measurements. The next point to consider is the fact that we have no true mean around which we can consider standard deviation ranges and this might at first glance seem a serious obstacle to any probability considerations. However we know that the distribution of many dates would be statistically normal and since we can use the one experimental standard deviation to relate to that of a true C-14 date, there is in fact no actual difficulty. We can thus correctly predict that there is a 68 per cent <u>probability</u> that the date obtained is within $\pm 1\sigma$ of the true C-14 date and a 27 per cent probability that the date is within the range ± 1 to 2σ of the true date. And the relative probability of 2.4 applies. There is of course no necessity to consider only whole units of σ nor either these expressed in such units rather than years and Fig. 2 illustrates for a practical case, the well known Unetice date from Leki Male of 1655 ± 50 BC[4], the relative probability of the true date being of any particular year. In general it is true to say that the date provided (here 1655 BC) is the most likely figure to be correct and is about 1.7 times more <u>probably</u> correct than a value calculated by adding or subtracting one standard deviation - and about 7.3 times more so for two standard deviations. It is thus incorrect to regard the standard deviation term (or worse, two standard deviations) simply as a blanket error - it can and should be interpreted more exactly as we have shown. (Table 1 illustrates these points). Probability is unfortunately by its very nature open to subjective interpretation and whilst it would be unwise to draw any significant conclusions from the above relative probability at the 1σ level, it would be entirely reasonable to do so at the 2σ level.

Table 1

Leki Male Date 1655±50 BC

Standard Deviation from Date	Relative Probability of a date being correct
0	100
0.5	88
1	60
1.5	31
2	14
2.5	4
3	0.8

The foregoing is to a considerable degree related to consideration of significance in date differences. However the straightforward calculation of the standard deviation of such a difference (calculated as $[(\sigma_1)^2 + (\sigma_2)^2]^{\frac{1}{2}}$) is the most direct procedure[2]. A difference greater than 2σ is very probably significant. For example two dates of 1655±50 BC and 1905±100 BC, yielding a difference of 250 years with a standard deviation of $[(50)^2 + (100)^2]^{\frac{1}{2}}$ = 112 years, would be taken to be significantly different.

The Tree-Ring Correction of Carbon-14 dates

By convention all carbon-14 dates issued by dating laboratories are based upon a half-life for C-14 of 5568 years[5]. It is generally agreed that this figure is too low and the most accurate value in use is 5730 years[6]. Prior to the emergence of the bristlecone pine correction curve it was usual to multiply the 5568 value by 1.029 (using the BP date) to correct to the more accurate half-life. With dendrochronological calibration curves which use the old half-life this is now not necessary and we may derive a tree-ring based calendar date directly by appropriate interpolation. The use of such calibration curves is now more or less standard and the most recent version due to Suess[7], published separately by both Libby[8] and Berger[9], has been carefully examined to provide the data in Table 2. Interpolation error is inevitable but should not amount to more than ±20 years for the data listed. The table yields the bristlecone pine calendar date or dates for any conventional carbon-14 date at intervals of fifty years. Two points derive from examination of this list - the magnitude of the deviation from conventional dates increases gradually from about the middle of the first millennium BC and there is for about one quarter of the conventional dates listed no unique dendrochronological equivalent. This latter problem derives from the short term fluctuations observed in the curve and very much reduces the accuracy obtainable in these regions - the range of dates available can span as much as four hundred years in some cases. The availability of these calibration curves has been widely heralded and they have already been used, perhaps somewhat prematurely, to derive important and far-reaching archaeological conclusions. It is, however, essential to point out that the final form of the curve has not yet been established and that slight (but important) differences do exist between different workers[10]. The curve generally used is derived from the work of Suess - here no actual dates have been published only Δ C-14 values and curves[7,11,12].

In general there has not been much critical discussion of the use of such curves for archaeological purposes. The fact that good agreement exists between early Dynastic Egyptian historical dates (first half of the third millennium BC) and dendrochronologically corrected carbon-14 dates, suggests clearly that the general trend of the correction is right[14]. But a more rigorous comparison with historical Egyptian dates is made with material from the early second millennium BC in and around the Twelfth Dynasty. Here, with the recorded heliacal rising of Sirius in 1872 BC, during the seventh year of the reign of Sesostris III, historical dating can be of extreme accuracy[13]. A careful examination of the dates in this region determined recently by the British Museum and the University of California at Los Angeles would suggest that the bristlecone pine curve could be in error by about three centuries for some dates. The detailed implications of this will be discussed elsewhere[15] but it may be noted that the 'historical' calibration curve constructed, using the historical dates to correct the carbon-14 dates, closely parallels a curve drawn using just the 5730 half-life. Archaeological dates from the Wessex-related Úňetice culture[4] and the Breton First Series graves[16] instead of extending back to about 2100 BC[17] would correct to about 1800 BC and Mycenean dates from the MH period[18] would also correct to quite acceptable values. So clearly the final correction

Table 2: Bristlecone Pine Corrections

Conventional Carbon-14 Date $t\frac{1}{2}$ 5568 Years	Bristlecone Pine Corrected Date Calendar Years	Conventional Carbon-14 Date $t\frac{1}{2}$ 5568 Years	Bristlecone Pine Corrected Date Calendar Years
1900 AD	AD 1830, 1800,	550 AD	AD 610
	1720, 1680	500	570
1850	1770, 1720,	450	530
	1670	400	440
1800	1760, 1730,	350	410
	1660	300	330
1750	1650	250	290
1700	1640, 1540,	200	240
	1510	150	180
1650	1630, 1580,	100	110
	1500	50	70
1600	1480	0 AD	AD 60
1550	1450	50 BC	AD 50, 0-30 BC,
1500	1430		70, 100
1450	1400	100	AD 60, 30-50 BC,
1400	1370		130
1350	1330	150	BC 150
1300	1290	200	170, 330
1250	1260	250	200, 270, 380
1200	1230	300	400
1150	1210	350	420
1100	1210, 1170,	400	460
	1130	450	520, 690, 770
1050	1210, 1190,	500	530, 650, 780
	1070	550	560, 590,780
1000	1030	600	790
950	1010	650	860
900	970	700	870
850	860	750	880
800	840	800	890, 940, 980
750	830, 770	850	1070
700	710	900	1120
650	670	950	1140, 1220
600	640	1000	1230, 1290, 1330

Conventional Carbon-14 Date t $\frac{1}{2}$ 5568 Years	Bristlecone Pine Corrected Date Calendar Years
1050 BC	BC 1340
1100	1350
1150	1460
1200	1490
1250	1510
1300	1530, 1610
1350	1650
1400	1670
1450	1690
1500	1720
1550	1760, 1780, 2040
1600	2070
1650	2090
1700	2120
1750	2140
1800	2160
1850	2180
1900	2210, 2290, 2370, 2470
1950	2410, 2480
2000	2490
2050	2510
2100	2530, 2710, 2740
2150	2550, 2680, 2760, 2810, 2840, 2880, 2930
2200	2940
2250	2940
2300	2950
2350	2960
2400	2980, 3040, 3120, 3340
2450	3210, 3310, 3370
2500	3250, 3380
2550	3390

Conventional Carbon-14 Date t $\frac{1}{2}$ 5568 Years	Bristlecone Pine Corrected Date Calendar Years
2600	3390
2650	3400, 3460
2700	3400, 3430, 3490
2750	3510
2800	3530, 3590, 3640
2850	3670
2900	3680
2950	3690
3000	3710
3050	3740
3100	3820-3900
3150	3950
3200	3970
3250	3990
3300	4030, 4110, 4220
3350	4230, 4310, 4330
3400	4340
3450	4350
3500	4360
3550	4370
3600	4400
3650	4440
3700	4480
3750	4520
3800	4580
3850	4680
3900	4810
3950	4830
4000	4860
4050	4890
4100	4940
4150	5000
4200	5070
4250	5220
4300	5290

curve to be used is not yet available. However the data in Table 2 represents the present best information available.

The Error from Non-contemporaneity of Samples and Archaeological Context

It is important here to make brief mention of an aspect of carbon-14 accuracy generally appreciated but little considered. This concerns the contemporaneity of the death of the material dated and the archaeological context. Almost certainly this is the most underestimated source of error from the archaeological point of view. Peat and bog wood are two serious potential sources of error in North Britain and re-use of building timber could be a far greater problem than usually assumed. Even artifacts made from freshly cut oak, if constructed from heart wood, could be one or two centuries older than their associated archaeological context.

The correction of Wessex culture related dates to about 1800 BC has already been mentioned and if one considers carefully the possibility of the grave wood or charcoal being older than the date of burial, as could well be the case here[4], then adjustment of those events to a date as recent as 1600 BC is entirely feasible. We cannot of course be certain about this adjustment, but equally we cannot dismiss it. Clearly the accuracy of the dating of the archaeological event is here necessarily very poor and there is obviously little justification for dismissal of the generally accepted links between Wessex and Mycenae, as has been suggested[15].

A similar problem is demonstrated in the series of carbon-14 dates from Pylos[16] (Palace of Nestor), the palace destruction dating to Late LH IIIB with a conventional carbon-14 date average of 1105 BC. This would adjust to about 1280 BC, using the 'historical' curve correction mentioned, which is just about acceptable for Late LH IIIB. However the palace construction timbers from Mid LH IIIB, with conventional carbon-14 dates as old as 1500 BC, are impossible to reconcile with Aegean chronology. The situation is, however, understandable when one appreciates that these massive timbers are in fact squared beams which could well derive from the heart wood of trees centuries older than the period in which they were actually used.

Much nearer the interests of most participants of the Forum meeting was the date series mentioned there by Mr Greig for his site at Cullykhan. The conventional date of 1186±60 BC (BM 444) from the charred beams inside the vitrified fort wall illustrated exactly the point being made here. Namely that old wood can indeed be a very serious source of error, here perhaps as much as eight centuries are involved. And it is of course only with gross discrepancies of this kind, or when unacceptable deviation from historical chronology is observed, that the problem is usually even suspected. It is necessary then to interpret with caution dates from materials of this kind and one may well wonder if there really is any point in dating doubtful samples expected to be in regions of severe short term dendrochronological fluctuation - an example would clearly be oak from LBA and Iron Age times.

Laboratory Accuracy and Comparisons

Present day laboratory technique for carbon-14 dating has evolved, in general, good and reliable procedures. Concern here resolves itself into two parts. Firstly the use of old and published dates and secondly concerning a current evaluation. For the latter possibility it is perfectly feasible to ascertain all the information needed for a personal estimation of the laboratory procedure. One can enquire concerning sample pretreatment, the number of dates determined, the statistical error associated with the counting, the laboratory technique used, the correction, if any, for isotopic fractionation, and the general comments by the laboratory staff as to their personal evaluation of the sample and the date obtained. However, in the case of published data, it is by no means always possible to ascertain all the above information. Here the topic can indeed become subjective and a generally pessimistic approach would seem desirable.

For present purposes, a useful comparison between two completely separate series of laboratory measurements, upon identical samples, is provided by the collaborative dating programme between the British Museum and the University of California, Los Angeles. Here the same samples were dated by two entirely different procedures and the bulk of the results at present available are listed in Table 3. From the earlier remarks concerning the reliability of date differences we can conclude that in terms of probability theory we would expect 68 per cent of these results to differ by no more than $\pm 1\sigma$ (calculated from the two individual standard deviations as described). A further 27 per cent should differ by between 1 and 2σ and only 5 per cent should differ by more than 2σ. There are of course only twenty-three samples listed so that even one sample falling into the wrong bracket changes these figures by about 4 per cent. Thus the derived data, 52 per cent for $0-1\sigma$, 39 per cent for 1 to 2σ and 9 per cent above 2σ, are in fact in really remarkably good agreement. There is clearly no serious systematic error between both series of experimental measurements. This would of course be the expected conclusion from departments with the reputations here enjoyed and the main result of interest to archaeologists is in fact the apparently very large differences that can occur but which are entirely expected from the very nature of the measurement. From the point of view of the reliability of one single date or a series of dates covering expected cultural phases these considerations must be applied before real significance can be attached to any agreement or divergence observed with an archaeological model.

Thermoluminescence

Finally, although not strictly related to carbon-14 work, it would here be appropriate to consider the technique of thermoluminescence dating[19]. At present this work is applied only to pottery and has an accuracy of about ten per cent (±400 years at 2000 BC). The great attraction of this work, however, is not only that it can be applied to the most usefully dated of archaeological materials but also that it is an absolute method quite independent of half-life accuracy and the earth's past magnetic field variations[20]. Also of course the dates obtained should relate clearly to the context of interest. These considerable advantages are offset somewhat by the difficulties associated with accurate determination of the effective archaeological radiation dose that the pottery suffers whilst buried. Surrounding soil or rock and their overall annual wetness, considered for a sphere of about fifteen inches diameter around each specimen, provides about one-fifth of the total radiation dose received by the pottery and has to be carefully considered for each sample. Preferably then, pottery that has been well buried from very early on and which is immersed in a reasonably homogeneous matrix is selected for dating. A sample of the surrounding soil is essential for radioactivity measurements and since both this and each sherd need to have their moisture contents determined they are necessarily placed in a sealable polythene bag immediately after excavation. With this essential degree of concern for burial conditions it is clearly advisable to seek the advice of the dating laboratory if there is any doubt about moisture or homogeneity. The pottery itself needs to be at least a quarter inch in thickness and a few square inches usually suffice. A minimum of five specimens from each context is essential since the mean date will provide a much closer approach to the true date of firing than any one sherd alone. None of these requirements is particularly restrictive and we may expect to see increasing use of the technique over the next few years.

Clearly the overall picture here emerging is one of steady progress in the field of absolute dating. No doubt several decades will pass before all main difficulties are resolved satisfactorily but it is clear that, providing due consideration be taken of the existing known problems, then adequate interpretation and use may be made of much of the information at present available.

Table 3: Interlaboratory Comparison

Published Carbon-14 Dates BC[9,14] 5730 half-life				Date Difference (UCLA minus BM) and Standard Deviation	Date Difference as a fraction of the Standard Deviation
British Museum		Univ. of California			
Ref. no.	Date	Ref. no.	Date		
228	2480±65	1200	2685±60	205±88	+2.3
229	2710±65	1201	2470±60	-240±88	-2.7
230	2560±65	1202	2401±60	-159±88	-1.8
231	2450±65	1203	2315±60	-135±88	-1.5
232	2410±65	1204	2365±60	-45±88	-0.5
233	2170±65	1205	2225±60	55±88	+0.6
234	1950±65	1206	2130±60	180±88	+2.0
235	2240±65	1207	2220±60	-20±88	-0.2
333	1070±100	1390	1220±60	150±117	+1.3
336	1020±100	1393	1210±60	190±117	+1.6
338	1170±85	1395	1015±60	-155±104	-1.5
344	730±70	1401	720±60	-10±92	-0.1
334	570±70	1391	655±60	85±92	+0.9
340	430±80	1397	455±60	25±100	+0.3
337	1220±75	1394	1170±60	-50±96	-0.5
332	2150±105	1389	2385±60	235±121	+1.9
346	2030±80	1403	2105±60	75±100	+0.8
331	1940±115	1388	2120±60	180±130	+1.4
330	1930±115	1387	2030±60	100±130	+0.8
342	1820±70	1399	1775±60	-45±92	-0.5
343	1880±85	1400	1860±60	-20±104	-0.2
347	1800±80	1413	1935±60	135±100	+1.4
238	1740±65	1212	1800±60	60±88	-0.7

Notes

1. Downie, N.M. & Heath, R.W. Basic Statistical Methods (1965).
2. Barker, H. The Scientist and Archaeology (1963).
3. Experimental verification if readily obtained from radioisotope generated x-ray counting measurements. One hundred identically timed runs, each producing about 10,000 counts, had an overall standard deviation agreeing (within 5 per cent) with the σ from any one run.
4. Bakker, J.A., Vogel, J.C. & Wislanski, T. Helinium IX (1969), 231.
5. Libby, W.F. Radiocarbon Dating (1949).
6. Godwin, H. Nature 195 (1962), 984.
7. Suess, H.E. Proc. XII Nobel Symposium, Uppsala (1969).
8. Libby, W.F. Phil. Trans. Roy. Soc. 269A (1970), 1.
9. Berger, R. Phil. Trans. Roy. Soc. 269A (1970), 23.
10. Damon, P.E. & Ralph, E.K. Proc. XII Nobel Symposium, Uppsala (1969).
11. Suess, H.E. Radioactive Dating and Methods of Low Level Counting, A.E.C. Vienna (1967).
12. Stuiver, M. & Suess, H.E. Radiocarbon 8 (1966), 534.
13. Parker, R.A. The Calendars of Ancient Egypt (1950).
14. Edwards, I.E.S. Phil. Trans. Roy. Soc. 269A (1970), 11.
15. McKerrell, H. PPS forthcoming.
16. Briard, J. Gallia Prehistoire XI (1968), 256.
17. Renfrew, C. PPS XXXVI (1970), 280
18. Kohler, E.L. & Ralph, E.K. Amer. J. Archaeol. 65 (1961), 357.
19. Zimmerman, D.W. Archaeometry 13 (1971), 29.
20. Bucka, V. Nature 224 (1969), 681.