

**HIGH PRECISION RADIOCARBON DATING BY LIQUID  
SCINTILLATION COUNTING APPLIED TO  
RADIOCARBON TIME-SCALE CALIBRATION**

GORDON W PEARSON

Palaeoecology Laboratory, The Queen's University of Belfast,  
Northern Ireland

**ABSTRACT.** The precision quoted on a radiocarbon date is often misleading. Improvement in the accuracy of measurement is discussed together with the use of appropriate precisions. The effect of high accuracy in  $^{14}\text{C}$  dating and meaningful precisions are illustrated and applied to radiocarbon time-scale calibration. Intercomparison of such calibrations reveals a significant continuous saw-tooth pattern of short-term periodicity.

INTRODUCTION

The errors on a radiocarbon date can be divided into three main groups, 1) those due to the selection and isolation of a sample, 2) those due to the  $^{14}\text{C}$  measurement, including all counting errors and those introduced by experimental technique, and 3) those over which we have no control, eg, variation of  $^{14}\text{C}$  concentration in the atmosphere.

For optimal use of the dating technique, an attempt should be made to equate the overall error for each group. Although the errors in group 1) are not normally included in the precision quoted on a date, it is obvious that the reliability of the sample will need to be considered in the interpretation of the evaluated date. This error can be reduced by careful choice of materials, opting primarily for contemporary material or contemporary components of material. This paper is directed to the overall error for group 2) *ie*, that generally quoted on a radiocarbon date, and its effect on the derivation of errors due to radiocarbon time-scale calibration, group 3).

The error introduced by conversion of radiocarbon years to calendar years, must, as far as genuine 'wiggles' allow, be no worse than the greatest accuracy of group 1) and 2), above; it should not be a limiting factor, or excuse, to any improvement being made in either sampling or measurement techniques.

*Errors in dating*

Interlaboratory measurement of identical samples, under normal laboratory conditions and experimental procedures has not yet been achieved on a sufficiently large scale to give a reliable estimate of the actual spread of results from the dating laboratories. However, a number of laboratories have been involved, in part, in radiocarbon time-scale calibration using dendrochronologically dated wood samples. A comparison of these measurements shows that the reproducibility is outside the statistical limits quoted, indicating some inaccuracy in experimental technique. Although there may be a small error, <0.1 percent (Stuiver, 1978) introduced by the non-removal of mobile fractions, *ie*, resins which give a 'memory' effect to individual year growth, perhaps the greatest cause of the differences in the fine detail of calibration between laboratories is due to the inaccuracy of measurement. In most cases, the quoted

error, usually about  $\pm 0.6$  percent (ca 50 yr) is too small and needs to be increased to ca  $\pm 0.8$  percent to cover the whole sequence of processing and measurement (Clark, 1975), in order to agree with the observed spread of duplicated measurements. In general, then, additional error sources almost equal in magnitude to the precision quoted are inferred.

The error quoted on a radiocarbon date should be a realistic assessment of the overall error, including all the individual errors resulting from any factor affecting the accuracy of measurement or causing any deviation from standard conditions. Furthermore, this estimated or evaluated error should be confirmed by internal duplication of samples that take account of extreme conditions of operation. Such a set of duplicate measurements is given in table 1. An overall estimated error, *ie*, not propagated, should be avoided since it may lead to a wrong estimate of accuracy and, in some instances, allow misinterpretation (see below, *Assessment of precision*). A detailed breakdown and evaluation of these error sources for liquid scintillation counting of benzene is available (Pearson, 1979). All factors that could significantly affect the accuracy of measurement were investigated and any deviation from standard conditions was resolved by correction, resulting in an overall standard deviation error now calculated to be ca  $\pm 0.25$  percent (*ie*, 20 yr) on the measurement of a one half-life sample under standard conditions. If corrections are ignored, or not investigated without allowing for omission in evaluating the overall error, then the distribution of results will fall outside of statistical expectancy. The effect of these additional errors, unaccounted for in the quoted precision, could be interpreted incorrectly as a 'wobble' when applied to the measurement of radiocarbon time-scale calibration samples.

TABLE 1  
Comparison of duplicate measurements

Sample id.	$^{14}\text{C}$ age BP $\pm 1$ SD	Diff in $^{14}\text{C}$ Age (yr)	Sample id.	$^{14}\text{C}$ age BP $\pm 1$ SD	Diff in $^{14}\text{C}$ Age (yr)
740 A	4067 $\pm$ 20	4	HAR A-2	-5476 $\pm$ 15	28
740 B	4071 $\pm$ 17		HAR A-4	-5448 $\pm$ 18	
640 A	4111 $\pm$ 20	49	HAR B-1	2104 $\pm$ 20	23
640 B	4062 $\pm$ 20		HAR B-7	2081 $\pm$ 24	
640 C	4091 $\pm$ 20		HAR C-1	5075 $\pm$ 24	
500 A	4055 $\pm$ 20	4	HAR C-3	5089 $\pm$ 29	
500 B	4051 $\pm$ 17		HAR D-5	10,080 $\pm$ 30	34
420 A	4024 $\pm$ 20	21	HAR D-9	10,046 $\pm$ 36	
420 B	4003 $\pm$ 20		HAR E-1	20,020 $\pm$ 68	11
340 A	3880 $\pm$ 20	23	HAR E-9	20,009 $\pm$ 82	
340 B	3857 $\pm$ 20		OX, 1-30 A	-2459 $\pm$ 13	5
60 A	3754 $\pm$ 20	9	OX, 1-30 B	-2454 $\pm$ 13	
60 B	3763 $\pm$ 20		OX, M-3 A	-2458 $\pm$ 13	9
1460 A	404 $\pm$ 17	17	OX, M-3 B	-2449 $\pm$ 13	
1460 B	387 $\pm$ 17				
1840 A	99 $\pm$ 17	20	Interval between duplicate measurements varies from 1 week to 4 years.		
1840 B	119 $\pm$ 17				

An additional disadvantage is that such dates will give a wrong weighting if used in combination with other calibration measurements, thereby giving incorrect calendrical conversion of  $^{14}\text{C}$  dates. Thus, it is particularly important in calibration not only to improve the accuracy of measurement but to ensure that the precision quoted is a true overall estimate. A comparison of our results with those of Stuiver (1978, fig 1) illustrates the advantages of having both high and realistic evaluated precisions.

Although the above discussion has been confined to radiocarbon time-scale calibration samples, interlaboratory variability in measuring archaeological samples might be worse due to greater variability in the pre-treatment process. The above comments must, therefore, apply.

#### *Assessment of precision in series measurement*

The overall precision quoted on a radiocarbon date calculated as in Pearson (1979), can be misleading in certain circumstances as it may prove to be over-estimated.

The high precision radiocarbon dating of some 50 bi-decade samples of Irish subfossil oak that had been dendrochronologically matched but remained as a 'floating' chronology, covering the time period ca 4000 to 5200 BP, was reported (Pearson and others, 1977). These radiocarbon dates (95 percent of which, although not perfectly 'normally' distributed, fell within  $\pm 2$  standard deviations of a linear regression line drawn from the data) seemed to indicate that, over the period investigated, no significant 'wiggles' were possible in excess of 0.3 percent. This indicated a contradiction of the findings of Suess (1970), inferring that the spread of Suess' dates was due solely to imprecise measurement and that his illustrated wiggles were not statistically sound. Commenting on our data, Suess said, "The authors, however, seem to have underestimated the accuracy of their own results . . ." It would, however, have necessitated a substantial reduction in our quoted precision of  $\pm 25$  years to justify any attempt at interpreting our findings with an irregular curve. Nevertheless, a re-evaluation of our error sources was made and a slightly reduced overall error was calculated, primarily because of updated information. This error reduction, although significant (ca 2 yr) was not sufficient to radically change the interpretation of our data. However, the possibility of greater error reduction was consistent with results obtained for duplicate sample measurement. A re-appraisal was made to determine the appropriate overall error that should be used when considering results determined as a series, *ie*, all samples measured within the same time period as that used to evaluate a grand mean value for the reference standard, providing the use of such a grand mean value is statistically valid.

$$\text{If } Z = \frac{N_s - N_b}{N_m - N_b} \text{ and } 8033 \log_e \frac{Z}{0.95} = \text{radiocarbon date BP}$$

where  $N_s$  = gross sample count rate (cpm),  $N_m$  = gross reference standard count rate (cpm) and  $N_b$  = background count rate (cpm).

It can be shown (Pearson, 1979) that

$$\frac{\text{Var } Z}{Z^2} = \frac{\text{Var}(N_s)}{(N_m - N_b)^2} + \frac{\text{Var}(N_m)}{(N_m - N_b)^2} + \text{Var}(N_b) \left[ \frac{N_m - N_s}{(N_m - N_b)(N_s - N_b)} \right]^2$$

Consider  $\frac{\text{Var}(N_m)}{(N_m - N_b)^2}$ : this fractional variance is made up from two types of error; first, the counting error, *ie*, the poisson statistics of radioactive decay (Type 1) and second, all other errors that are the result of deviations from standard conditions (Type 2). The magnitude of the second type can be reduced by correcting to standard conditions (Pearson, 1979) but the error remaining is still significant ( $\cong 0.07$  percent). The overall standard deviation error (Type 1 + Type 2) on the reference standard can be evaluated from the distribution obtained from repeated reference standard measurement, *ie*, (quality control) over the appropriate period. By root mean square subtraction of the counting error, Type 2 error can be evaluated. It is the distinction between these two types of error that allows different precisions to be used for the reference standard fractional variance when estimating the overall precision for the comparison of dates within or outside a series.

The reference standard measurement can, therefore, be considered as having two distinct functions; first, it provides a value (within the counting statistical limits) to convert a  $^{14}\text{C}$  measurement into a radiocarbon date, and, second, it enables the additional errors (Type 2) to be evaluated from its quality control distribution. Consequently, all samples measured in a series are only subject to the second type of error until the  $^{14}\text{C}$  measurement is converted into a radiocarbon date. Thus, it is appropriate to consider these  $^{14}\text{C}$  measurements as having an overall precision appropriately reduced to take account of the reference standard counting errors. If these  $^{14}\text{C}$  measurements are multiplied by the same constant, to convert to a quasi-radiocarbon date, using as this constant, the grand mean value of the reference standard distribution, no additional error has to be included. The degree of confidence around this constant, *ie*, counting error, need only be considered for dates removed in isolation or compared to dates outside the series. Further reduction in the overall error is appropriate, and should be made, if a grand mean value of background is used from data accumulated over the same period as for the series measurement. If  $^{14}\text{C}$  measurements of samples within the series are close, then the inaccuracy introduced by assuming a constant background equal to its grand mean value is much less than the counting error of the background. But as with the reference standard error, all errors other than the counting error on the background measurement, need to be propagated into the overall error on the series date.

It is deduced from above that two correct precisions apply to radiocarbon dates depending on the relationship within which they are being considered. For results presented by Pearson and others (1977) the standard deviation precision that applied to the dates within the series should have been  $\cong \pm 20$  years, although the series itself would have been able

to move to the limits defined by  $\pm 25$  years, *ie*, one date with full precision and all other dates compared in relation. Because it was no longer possible with this reduced precision to conclude a straight line as being a good fit to our data, it was important for correct interpretation to complete the measurement of the missing bi-decade samples over this period. These measurements together with those from Pearson and others (1977), are shown in figure 2.

#### METHOD

The methods for measuring samples presented in this paper are essentially as described previously (Pearson and others, 1977; Pearson, 1979).

A change in the derivation of mean values for count-rate, weight-loss, and barometric pressure has given greater flexibility to the counting schedule. Instead of using an arithmetic mean value for the determination of these factors (necessitating a rigorous adherence to a fixed schedule to avoid any weighting), a mean value based on linear regression is now used. This allows for considerable variation in 'cycle time'. Tests showed that, over a period of several weeks, weight losses for vials were linear. Thus, it was appropriate to use a linear regression. Routine monitoring of weight loss is carried out by weighing all vials at regular intervals to confirm linearity.

#### RESULTS AND DISCUSSION

An additional advantage in adopting the techniques described above (Pearson, 1979) results from the standard conditions of operation. Consequently, all samples are directly comparable with each other as  $^{14}\text{C}$  measurements. Long-term stability tests on efficiency (quality control) have shown no deviation (after correction) from the expected poisson distribution of radioactive decay over a period of ca 4 years. The first 9 results (Pearson and others, 1977) however, were calculated using an individual reference standard value. The remaining 43 results used the grand mean value. Because of subsequent proof of stability, these results have been recalculated using a single updated grand mean value, for the oxalic acid reference standard, which changed the 43 previously published results by ca 5 years, and the first 9 published by ca 30 years. All results presented here are calculated using the oxalic acid reference standard grand mean value of  $116.6544 = (0.95 \times \text{nett cpm})$ .

Table 1 gives the results of duplicate measurements, each being a full duplication of experimental method from unpretreated samples, except for the interlaboratory check samples (prefixed HAR) provided as benzene from the  $^{14}\text{C}$  dating laboratory AERE Harwell. All dates are calculated as BP, -ve values indicating dates after present, *ie*, after AD 1950. The differences between duplicates are quite consistent with the evaluated standard deviation error and support the validity of the applied corrections. The results of the new oxalic acid reference standard calibration (prefixed OX) illustrate the agreement for quadruplicates; an excellent example of short term series measurement where the precision on

each measurement for inter-comparison is ca  $\pm 0.15$  percent or,  $\pm 0.11$  percent on duplicate mean, *ie*, ca  $\pm 9$  years standard deviation.

Figure 1 gives the results obtained for bi-decade and decade samples of Irish oak wood, for the years AD 1410 to AD 1840. All tree-ring ages refer to the mid-age point of each sample and makes no allowance for variation in tree-ring width. These results are compared with those of Stuiver (1978) and show remarkable agreement, especially when considering that two techniques—liquid scintillation versus gas counting, and two tree species—oak versus pine were used, and the samples were from Ireland and America, respectively. The close agreement between calibrations shows that little additional error would be introduced by using a single-time scale calibration (at least for the Northern Hemisphere), providing such agreement in detail can be sustained. This will only be possible with high accuracy in measurement.

Figure 2b gives the re-calculated results of samples (Pearson and others, 1977), together with additional measurements; the negative tree-ring years have been used so as not to change the original table relativities. All measurements are in years before AD 1950 and are based on the Libby half-life of 5568 years. These measurements are compared with those of Suess (1978), figure 2a, and de Jong, Mook, and Becker (1979), figure 2c. Our results agree in detail sufficiently with those of Suess (1978)

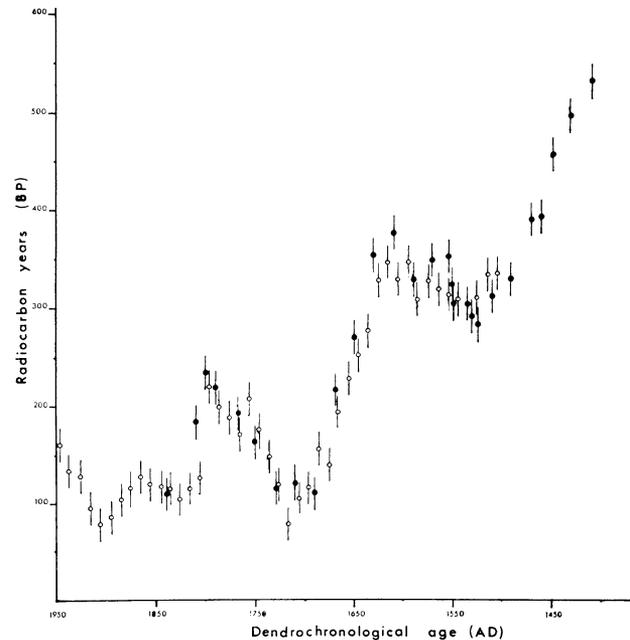


Fig 1. Comparison of radiocarbon time-scale calibrations. Radiocarbon dates  $\bullet$  for bi-decade and decade samples of tree-ring dated Irish and Scottish oaks. Radiocarbon dates  $\circ$  for decade samples of American pine (Stuiver, 1978).

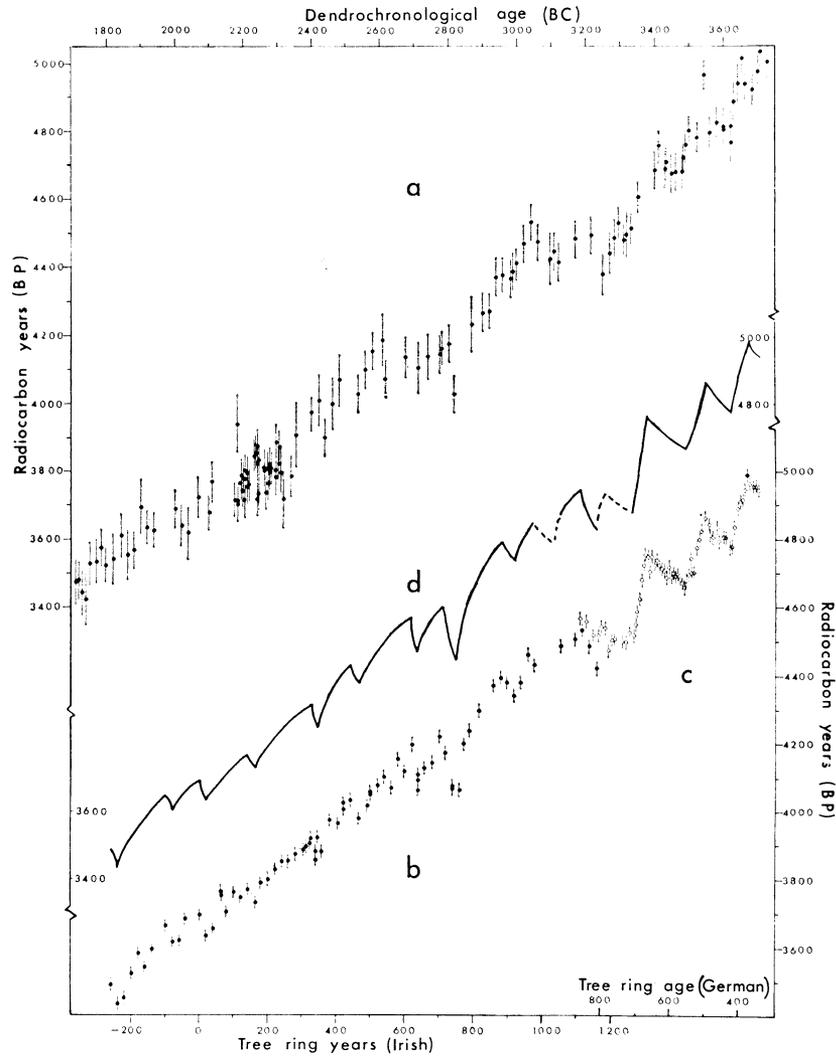


Fig 2. Relationship radiocarbon time-scale calibrations. a) Radiocarbon dates for bristlecone pine samples (Suess, 1978) corrected to the proper standard activity ( $0.95 \times$  NBS oxalic acid). b) Radiocarbon dates  $\bullet$  for bi-decade samples from an Irish oak floating chronology. c) Radiocarbon dates  $\circ$  for single year samples of a South German floating chronology (de Jong, Mook, and Becker, 1979). d) A non-mathematical curve drawn by eye to represent a probable fit to dates from 2b and 2c and provide an estimate of error for  $^{14}\text{C}$  date conversion over this 2000-year period.

to suggest that on a purely visual match our oldest date occurs at the bristlecone pine dendrochronological date of 3235 bc. Also, an overlap with the results of de Jong, Mook, and Becker (1979), occurs over the period 3190 to 3235 bc, but, as yet, this has not been confirmed by the dendrochronologic matching of the South German Neolithic and the Irish oak floating chronologies (Baillie, pers commun). The use of Suess' results is, therefore, considerably important for wiggle matching, to line up our floating chronologies, making dendrochronologic gap-bridging easier. When this visual match is confirmed by dendrochronology, the high precision measurements of de Jong and those from Irish oaks presented here will together provide more than 2000 years (2500, if figure 1 is included) of high precision calibration. When the results of Campbell and Baxter (1979) Borth 4 samples are compared, it is difficult to see any such pronounced <sup>14</sup>C depletion type deviations in our data for matching, although 3 single-decade samples were selected and measured over the most likely period of a match, ca 2400 bc (fig 2b). These single-decade results were not significantly different from the bi-decade results and it is therefore unfortunate that the more extreme Borth 4 measurements were not duplicated or confirmed when such large variations were apparent.

The saw-tooth type pattern of de Jong, Mook, and Becker (1979) would also appear to best fit our data from 1815 to 3235 bc, occurring with a slightly irregular frequency (fig 2d), between 110 to 180 years. This pattern is in excellent agreement with that shown in figure 1 for the years, AD 1410 to 1950. It would appear therefore, that there is a continuous cycle of abrupt change in the <sup>14</sup>C production process—first, enrichment followed by partial collapse in this production, then, the <sup>14</sup>C equilibrates back towards a smooth base line defined by the magnitude of, and the time periods between enrichments. This on/off switching pattern may well prove to be the final shape of the radiocarbon time-scale calibration, where these short-term saw-toothed irregularities are superimposed on the enrichment side of a smooth sinusoidal curve that has a periodicity of about 10<sup>4</sup> years. Unfortunately, this regularity could lead to tentative wiggle matches every cycle and the matching would become dependent on the magnitude of deviation, unless the bias on dates between laboratories could be guaranteed small compared to a 1/2 cycle value, requiring once more high accuracy in measurement.

In conclusion, it now seems likely that in the near future a high precision radiocarbon time-scale calibration will be universally acceptable. Many inaccuracies of measurement will have been eliminated and the quoted precision on a <sup>14</sup>C date will become meaningful.

#### ACKNOWLEDGMENTS

I wish to thank Mrs F Qua and other laboratory staff for their help in sample analysis, M G L Baillie and J R Pilcher, Palaeoecology Laboratory, The Queen's University of Belfast, for providing dendrochronologically dated wood samples, and SRC for a grant made to the author in 1979.

REFERENCES

- Campbell, J A and Baxter, M S, 1979, Radiocarbon measurements on submerged forest floating chronologies: *Nature*, v 278, p 409-413.
- Clark, R M, 1975, A calibration curve for radiocarbon dates: *Antiquity*, v 49, p 251-266.
- Damon, P E, Lerman, J C, and Long, Austin, 1978, Temporal fluctuations of atmospheric  $^{14}\text{C}$ : causal factors and implications: *Ann Rev Earth and Planetary Sci.* v 6.
- de Jong, A J M, Mook, W G, and Becker, Bernd, 1979, Confirmation of the Suess wiggles: 3200-3700 bc: *Nature*, v 280, p 48-49.
- Pearson, G W, 1979, Precise  $^{14}\text{C}$  measurement by liquid scintillation counting: *Radiocarbon*, v 21, p 1-21.
- Pearson, G W, Pilcher, J R, Baillie, M G L, and Hillam, J, 1977, Absolute radiocarbon dating using a low altitude European tree-ring calibration: *Nature*, v 270, p 25-28.
- Stuiver, Minze, 1978, Radiocarbon timescale tested against magnetic and other dating methods: *Nature*, v 272, p 271-274.
- Suess, H E, 1970, Bristlecone-pine calibration of the radiocarbon time-scale 5200 bc to the present, in Olsson, I U, ed, Radiocarbon variation and absolute chronology, Nobel symposium, 12th, Proc: New York, John Wiley & Sons, p 303-309.
- 1978, La Jolla measurements of radiocarbon in tree-ring dated wood: *Radiocarbon*, v 20, p 1-18.