IS RADIOCARBON DATING OBSOLESCENT FOR ARCHAEOLOGISTS?

BARBARA S OTTAWAY

Department of Archaeology, University of Edinburgh, Scotland

ABSTRACT. This paper will deal with two major points: 1) the lack of feedback between archaeologists supplying samples and using radiocarbon dating, and physicists carrying out ^{14}C dating measurements; and 2) the problem of calibrating groups of ^{14}C dates in a statistically meaningful way.

LACK OF FEEDBACK

These comments apply primarily to western Europe. However, in some of these countries it is still not generally accepted that ^{14}C dating has *any* usefulness for archaeologists, so that the points raised here are not by any means insular. Moreover, attendance at Archaeometry meetings and many similar scientific gatherings has strengthened my belief that many of the problems discussed here have a fairly general application to most of the non-Communist world.

There are two types of problems that archaeologists would like to solve with the help of ^{14}C dating: 1) they would like to date specific events such as the destruction horizon of a site; 2) they would like to date the duration of archaeologic phenomena such as the occupation of a settlement or a cemetery, or the duration of successive cultural groups in a geographically defined space. For the former, a precise date would be desirable, whereas for the latter, a range is what would be expected from the very nature of the question.

There are several problems that archaeologists have with the ^{14}C dating procedure. For example, an erroneous assumption is that there is always enough excavated material to obtain independent replicate dates, or to choose short-lived material for high-precision dates. Anyone who has excavated will know how difficult it is to find adequate material in sealed contexts.

There seems to be a widespread belief, at least in western Europe that money for radiocarbon dates is easy to come by; the opposite is, in fact, true. There is also the problem of adequate information. For example, what is really included in standard deviations quoted by laboratories, and why is there no internationally accepted agreement on this?

Why is there still no continuous quality control in which all laboratories publishing in say, *Radiocarbon*, are participating? Preliminary inter-laboratory comparisons (International Study Group, 1982; Ottaway, 1983; Waterbolk, 1983a) showed how badly such quality controls were needed on a continuous basis, yet there is still no sign of action.

Why are we archaeologists currently being advised by many radiocarbon physicists that we cannot compare dates from different laboratories, and that we should ignore dates measured more than 5 years ago? If these propositions were to be accepted, it would be a wasteful, not to say untenable, state of affairs, because some sites and their dating materials have gone forever.

CALIBRATION OF GROUPS OF ^{14}C DATES

Although the necessity to calibrate is accepted by all present at this meeting, this attitude is by no means uniform. It is not, for example, generally thought necessary by climatologists and Quaternary geologists, nor by many continental archaeologists. Particularly because of the interaction between climatologists and archaeologists, these divergent attitudes should be changed.

The fact that there has been no agreement up till now on any one calibration curve, or on the treatment of re-entrant wiggles, does not help. *Radiocarbon's* policy of deferring to the calibration preferred by individual scientists further confuses the issue; eg, in Linick (1979, p 196), La Jolla dates are calibrated on the La Jolla calibration curve and are given as, eg, 32nd, 33rd, and 34th millennium BC. In Meulengracht, McGovern and Lawn (1981, p 228) Pennsylvania dates are corrected on the MASCA calibration curve and are given as one date with standard deviation (eg, 400 ± 40 BC). In Felber (1982, p 227) Vienna dates are given using the de Vries correction again as individual dates with standard deviations (eg, $1500 \, {+\,150 \atop -\,\,\,50}$ BC).

Klein *et al* (1982, p 104) provide ". . . a calibration table suitable for the calibration of individual or 'single' radiocarbon dates. . . . Included in the category of 'single radiocarbon dates' are series of dates from samples thought to be coeval. . . ." There is no mention in the paper of how to calibrate a series of dates in a statistically meaningful way. Hence, archaeologists, encouraged to treat each date within a series as an individual sample, tend to calibrate each sample individually using the appropriate uncertainty and thus obtain a range which lies between the highest and the lowest calibrated age. This procedure is undesirable because 1) it neglects statistically meaningful information contained within a series of measurements, and 2) it mostly leads to the unacceptable result that the more samples we have for one archaeologic event, the larger the final range. In no other branch of science is this an acceptable procedure. Normally groups of measurements are made to confine the result, and to narrow down the range, thus increasing the precision of the overall estimate.

As mentioned above, archaeologists want to date either specific events or the duration of a period. Each of these requires a different set of samples. For the first case, replicate measurements of samples from the same site and the same event should, wherever possible, be obtained. This event is usually a destruction horizon containing long-lived charred timbers. If short-lived material is available and a high-precision laboratory willing to date it can be found, one high-precision date could fill the place of a series of routine measurements, but only if the resulting date falls on a straight part of the calibration curve (see below). In both instances it is important to know that the laboratory is either producing unbiased results or that their bias is known.

To date an archaeologic period, the measured sample *must* come from different sites to carry a geographically meaningful message (*cf* Waterbolk, 1983b, p 641). Since excavating archaeologists usually have a working rela-

tionship with one dating laboratory, mostly in their own country, dates for a period of a culturally defined area, not necessarily honoring modern boundaries, usually come from several dating laboratories. The question of bias of the laboratory is less important here; one might be correct in assuming that the biases are randomly distributed and thus cancel each other out. However, until quality controls are published frequently this must remain an assumption.

<div align="center">EXAMPLES AND SUGGESTIONS</div>

I would now like to illustrate the calibration of groups of measurements, using data from my own excavation of a Late Neolithic enclosure in Bavaria (Ottaway, 1984, in press a).

Dating a Specific Event

Of the 9 dates in Table 1A, 7 could be used in dating 2 events: a) a fire causing a palisade fence to collapse into the ditch (* in Table 1A); b) a fire in a different part of the site (** in Table 1A). These groupings were made on archaeologic grounds. Neglecting uncertainties in the age of the wood after felling, the ^{14}C ages of a) and b) do not overlap, and using non-parametric procedures (Ottaway, 1972), a ca 90% probability that the two events were separated in time can be computed.

The Mann-Whitney statistical procedure (Campbell, 1974) cannot be used for calibrated dates because we are dealing with ranges. The best we can do is calculate weighted means for the two groups, to calibrate these means, and to see whether the ranges overlap. The results are summarized in Table 1B. Using the data of Klein *et al* (1982), the two ranges overlap significantly. Using the calibration curve of Pearson, Pilcher and Baillie (1983), the ranges also overlap, but only by ten years. The temptation to use the latter calibration is, for an archaeologist, very strong, but is it justified?

Taking into account the uncertainties of sample contamination and treatment which have been necessarily neglected here, I am very much inclined to agree with the conclusions of the International Study Group (1982), who ". . . urge increased caution in attempting to resolve ^{14}C differences of <200 years. "As we can see, however, this would imply that archaeologists should abandon any attempt to resolve, by ^{14}C dating, differences which are professionally very important in many excavations, and that physicists should cease to suggest that improvements in measuring techniques, *ie*, high-precision dating, could bring such resolution within reach, except in very exceptional circumstances.

Dating the Duration of a Period

To date the duration of an entire cultural group (using the 9 dates in Table 1A as an example) we might calculate the median and its confidence limits (Nair, 1940), quoted in Campbell (1974, p 37). This would be

$$P(\%) (4210 < 4280 < 4385) = 96.1$$

where P = confidence probability and 4210 and 4385 = M_L and M_U, the

TABLE 1

A. Radiocarbon dates relating to the Cham group at the Galgenberg,
Lower Bavaria (Ottaway, in press b)

Lab no.	^{14}C dates BP	Lab no.	^{14}C dates BP
GrN-12699	4510 ± 30*	GrN-12561	4255 ± 40
GrN-12702	4385 ± 35*	GrN-12700	4225 ± 30*
GrN-12562	4290 ± 45	GrN-12564	4210 ± 60**
UB-2551	4285 ± 85*	GrN-12563	4150 ± 60**
GrN-12701	4280 ± 35*		

B. Calibration of the two groups of 'single event' dates (see text) after two
different publications (using appropriate instructions)

Weighted mean of uncalibrated dates BP	Calibrated dates BC	
	(Klein *et al*, 1982)	(Pearson, Pilcher & Baillie, 1983)
Event a) 4350 ± 20 (5)*	3350–2890	3030–2900
Event b) 4180 ± 40 (2)**	3000–2540	2910–2590

C. Calibration of the Interquartile Range (IQR) of the Galgenberg dates
using three different methods

IQR of ^{14}C dates BP	Calibrated dates BC	
	(Klein *et al*, 1982)	(Pearson, Pilcher & Baillie, 1983)
4315–4220 (*cf* Fig 1A)	3150–2670	2920–2700
Each of 9 dates calibrated separately and the IQR derived as explained in the text (*cf* Fig 1B)		2906–2795

* Fire causing a palisade fence to collapse into the ditch
** Fire in a different part of the site

lower and upper bounds. When calibrated, M_L and M_U give very wide ranges (eg, 3360–2655 BC).

A more satisfactory non-parametric treatment is the interquartile range (Ottaway, 1972, 1973), since it is the *range,* or *floruit,* of the period that is required. This treatment is particularly effective in that the interquartile range and the median are very stable, being little affected by additional values once a (small) critical number has been attained. For example, until recently, only 6 dates of the late Neolithic Cham group of Bavaria were known. Recent excavations and dating have increased this number to 16, yet the interquartile range and the median dates have changed very little (Ottaway, in press b).

The interquartile range, moreover, contains within it statistical information by means of which the contemporaneity or otherwise of several cultural groups can be tested (Hypergeometric Distribution (Siegel, 1956)).

Is it more sensible to construct an interquartile range with uncalibrated dates (Fig 1A) and then to calibrate the range, or to calibrate each date and then to construct an interquartile range (Fig 1B)? Figure 1 and
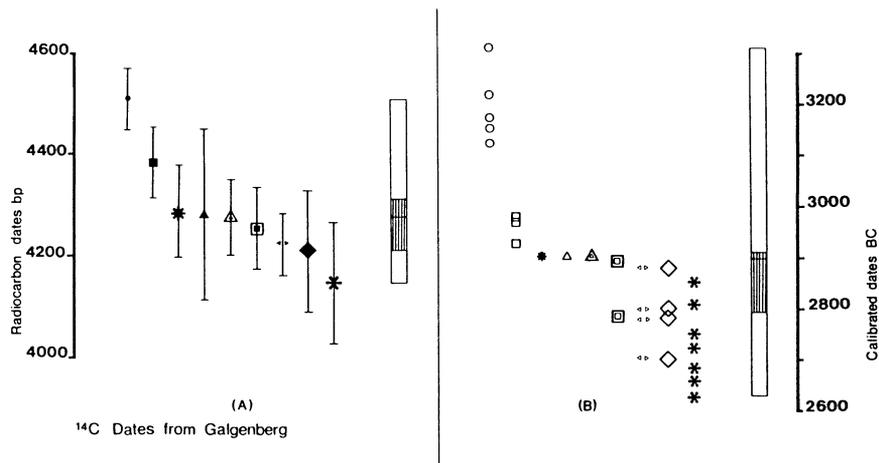
Fig 1A. Nine ^{14}C dates from the Galgenberg (*cf* Table 1A) with two standard deviation ranges and their Interquartile Range

B. The same nine dates individually calibrated on 'the high-precision curve' (Pearson, Pileher & Baillie, 1983); the multiple symbols denote fractionally weighted values used for the construction of the Interquartile Range.

Table 1C show that the two methods give notably different values. Moreover, construction of the interquartile range with individual calibrated dates is slow and involves fractional weighting to allow for the probability of multiple calibrated dates in regions of the calibration curve where there are re-entrant wiggles. Although this technique uses the fact that there is a low probability of the calibrated dates occurring outside the error bounds, *ie*, between the 'wiggles', it is novel and still subject to discussion. It has been included because Table 1C shows that the question of the 'true' or reliable range of the duration of the cultural group is still far from settled, although it is of vital interest to practicing archaeologists.

As a final comment, it is interesting to calculate from the table of Distribution-free Tolerance Limits (Diem, 1962, p 28) how many samples we would need to have a confidence probability, P, of 0.99 that 99% of all possible values of the population would lie between the lowest and the highest values of the samples actually measured. Table 2 shows that 661 samples would be needed! Even to have P = 0.95, that 90% of the ages of the total population lie between the lowest and the highest values, we would need 45 ^{14}C dates. Only the very exceptional excavation would have funds or facilities to acquire sample dates on such a scale. Thus, most archaeologists will be forced, by the nature of the excavated material or the limitations of funding, to be satisfied with lower confidence limits, *ie*, a higher degree of uncertainty. Also, all available dates *must* be used, even if they are from different laboratories and archaeologists *must* treat such collections of data statistically. This can only be done if all laboratories carry out, and publish, routine quality controls. Finally, for statistical tests, either simple or sophisticated (eg, Efron, 1979), point values after calibration, not just ranges, are essential.

TABLE 2

Distribution—free tolerance limits*

Confidence probability = 0.99				
	99%	95%	90%	50%
N	661	130	64	10

Confidence probability = 0.95				
N	473	93	45	7

* After Documenta Geigy (Diem, 1962)

SUMMARY

I believe that if archaeologists are to go on using radiocarbon dates in a meaningful way, we need

1) better information exchange between ^{14}C laboratories and archaeologists

2) one universally accepted calibration curve

3) that this should contain point estimates rather than ranges to allow for viable statistics on group data. This would facilitate employment of non-parametric methods, but would assume that the best possible accuracy within each ^{14}C laboratory is achieved

4) more research on the statistical treatment of group data and of their calibration, and on the treatment of re-entrant wiggles

5) agreement and uniformity of publication of uncalibrated radiocarbon dates with their standard deviations

6) regular publication of quality controls of all dating laboratories

7) the final universally accepted calibration data and procedure should be put on floppy disks, to be distributed freely and updated frequently.

REFERENCES

Campbell, R C, 1974, Statistics for biologists, 2nd ed: Cambridge, Cambridge Univ Press, p 35–39.

Diem, K, ed, 1962, Documenta Geigy scientific tables, 6th ed: Manchester, England, Geigy Pharmaceutical Co.

Efron, B, 1979, Bootstrap methods: another look at the jackknife: Annals Statistics, v 7, p 1–26.

Felber, H, 1982, Vienna Radium Institute radiocarbon dates XII: Radiocarbon, v 24, no. 2, p 222–228.

International Study Group, 1982, An inter-laboratory comparison of radiocarbon measurements in tree rings: Nature, v 298, p 619–623.

Klein, J, Lerman, J C, Damon, P E and Ralph, E K, 1982, Calibration of radiocarbon dates: Tables based on the consensus data of the Workshop on Calibrating the Radiocarbon Time Scale: Radiocarbon, v 24, no. 2, p 103–150.

Linick, T W, 1979, La Jolla natural radiocarbon measurements VIII: Radiocarbon, v 21, no. 2, p 186–202.

Meulengracht, A, McGovern, P and Lawn, B, 1981, University of Pennsylvania radiocarbon dates XXI: Radiocarbon, v 23, no. 2, p 227–240.

Ottaway, B S, 1972, Dispersion diagrams: a new approach to the display of ^{14}C dates: Archaeometry, v 15, p 5–12.

————— 1973, Estimating the duration of cultures: Antiquity, v 47, p 231–233.

Ottaway, B S, ed, 1983, Archaeology, dendrochronology and the radiocarbon calibration curve: Occasional papers no. 9, Dept Archaeol, Edinburgh Univ.

——————— 1984, Zwei neolithische Siedlungsgrabungen in Niederbayern: Archäol Korrespondenzblatt, v 14, no. 1, p 23–29.

——————— in press a, The Galgenberg, a late Neolithic enclosure in Bavaria, *in* Burgess, C, ed, Enclosures and defences in the Neolithic of western Europe: Oxford, Br Archaeol Repts.

——————— in press b, Neue Radiokarbondaten Altheimer und Chamer Siedlungsplätze in Niederbayern: Archäol Korrespondenzblatt.

Pearson, G W, Pilcher, J R and Baillie, M, G, L, 1983, High-precision ^{14}C measurement of Irish oaks to show the natural ^{14}C variations from 200 BC to 4000 BC *in* Stuiver, M, and Kra, R S, eds, Internatl ^{14}C conf, 11th, Proc: Radiocarbon, v 25, no. 2, p 179–186.

Siegel, S, 1956, Non-parametric statistics for the behavioral sciences: London,

Waterbolk, H T, 1983a, Ten guidelines for the archaeological interpretation of radiocarbon dates: PACT, p 57–70.

——————— 1983b, The integration of radiocarbon dating in archaeology *in* Stuiver, M, and Kra, R S, eds, Internatl ^{14}C conf, 11th, Proc: Radiocarbon, v 25, no. 2, p 639–644.