The Holocene 14,2 (2004) pp. 296–298

The intercept is a poor estimate of a calibrated radiocarbon age

R.J. Telford,[1]* E. Heegaard[1,2] and H.J.B. Birks[1,2,3]

([1]Bjerknes Centre for Climate Research, Allégaten 55, N-5007 Bergen, Norway; [2]Botanical Institute, University of Bergen, Allégaten 41, N-5007 Bergen, Norway; [3]Environmental Change Research Centre, University College London, 26 Bedford Way, London WC1H 0AP, UK)

Received 14 February 2003; revised manuscript accepted 12 June 2003

A HOLOCENE FORUM ARTICLE

Abstract: Intercept-based methods of generating a point estimate of a calibrated radiocarbon date are very popular, but exhibit undesirable behaviour. They are highly sensitive to the mean of the radiocarbon date and to adjustments of the calibration curve. Other methods give more stable results. The weighted average of the probability distribution function is recommended as the best central-point estimate, but more consideration should be given to using the full probability distribution rather than a point estimate in developing age-depth models.

Key words: Radiocarbon dating, calibration, intercept, weighted average, methodology, Holocene.

Introduction

For many aspects of palaeoecology, for example constructing age-depth models, it is convenient to have a single, central-point estimate of a date. For radiocarbon dates, the mean, around which the errors are normally distributed, is a suitable point estimate. However, after calibration, errors are often highly non-normal and multimodal, with non-continuous high probability fields. Palaeoecologists have used a variety of methods for generating point estimates, most frequently the intercept (Figure 1), where the mean of the radiocarbon date intercepts the calibration curve (also known as CALIB method A; Stuiver and Reimer, 1993), but little consideration has been given to the statistical properties of such estimates.

A survey of all articles published during 2002 in *The Holocene* reporting radiocarbon dates (n = 32) found that 12 papers used radiocarbon years as the basis of their chronology, contrary to the advice of Bartlein *et al.* (1995). Twelve papers used the intercept method, and one used the mid-point of the 2σ range. Of the remaining papers, three avoided making a central estimate by presenting either the 1σ or 2σ range, and it is not clear how four papers obtained a central estimate of their calibrated dates. Of the 12 papers using the intercept method, only two explicitly stated how multiple intercepts were treated.

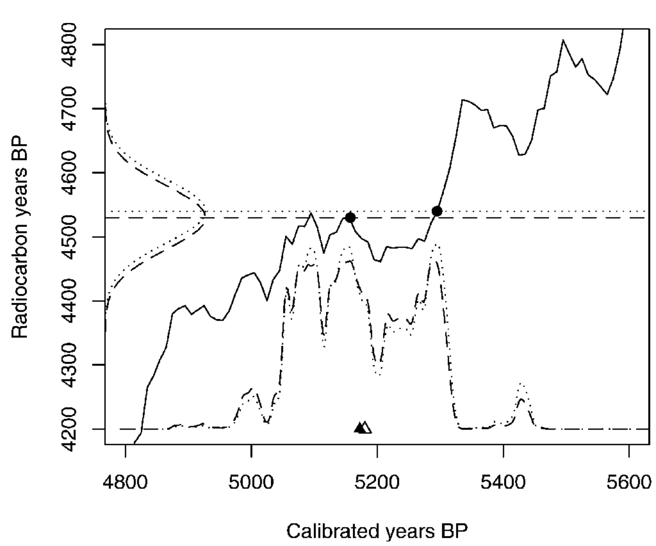

Figure 1 Comparison of intercept and weighted average methods for 4530 ± 50 (dashed) and 4540 ± 50 (dotted) ^{14}C yr BP. The horizontal lines show the radiocarbon date, with normally distributed errors shown against the y-axis. The average intercept of each date is marked by a circle and they differ by 138 years. The probability density function for each date is shown along the x-axis. The weighted averages are marked by a filled and an open triangle for 4530 and 4540 ^{14}C yr BP, respectively, and differ by eight years.

*Author for correspondence (e-mail: richard.telford@bjerknes.uib.no)

10.1191/0959683604hl707fa

In this paper we consider how different estimates of the central point of a calibrated radiocarbon year behave as the mean is moved, the standard deviation of an age increases, and the calibration curve is adjusted.

Eight estimates of the central point of the calibrated radiocarbon date were made.

(1) Intercept (Stuiver and Reimer, 1993), using the mean intercept if there is more than one intercept.
(2) As above but using the median intercept (e.g., Seierstad *et al.*, 2002).
(3) Mode.
(4) Median.
(5) Weighted average or moment.
(6) Weighted average of 2σ ranges using the range mid-points (extension of Bennett, 1994).
(7) Weighted average of 2σ ranges using the range mode.
(8) Weighted average of 2σ ranges using the range intercept (mean intercept if more than one) or mid-point if no intercept in that range (Brown *et al.*, 2002).

Methods

OxCal3.5 (Bronk Ramsey, 1995) was used to generate probability density fields of calibrated dates, using a cubic spline interpolation of the INTCAL98 (Stuiver *et al.*, 1998) calibration curve and a resolution of two years. Intercepts were calculated from the INTCAL98 calibration curve. All calculations were done using R (Ihaka and Gentleman, 1996).

Results and discussion

The true calibration curve is a smooth continuous function (Gómez Portugal Aguilar *et al.*, 2002), although our estimate of it (Stuiver *et al.*, 1998) has a decadal resolution and is somewhat jagged. Despite this, the probability density function of calibrated dates changes smoothly as the radiocarbon date mean is varied (compare the probability density functions of 4530 ± 50 and 4540 ± 50 ^{14}C yr BP in Figure 1). The complex shape of the probability density function prohibits a single value summarizing it (cf. a normal distribution). Despite this caveat, it is often useful to summarize the probability distribution by a single value. All of the methods examined here (with the exception of the mode) can fall in a low-probability region in multimodal distributions, for example, after calibration the weighted average of 670 ± 15 ^{14}C yr BP has a minute probability of being the true age. There can be no test of how good a single value is. What can be examined, however, is the behaviour of the method as the radiocarbon mean, standard deviation and calibration curve are altered.

Given the reported errors on radiocarbon dates (typically 30–100 years), the dates 4530 ± 50 and 4540 ± 50 ^{14}C yr BP are indistinguishable. This is reflected by the similarity of the probability density functions of each date (Figure 1). A well-behaved central-point estimate should mimic this similarity. The weighted averages of the two probability density distributions differ by only eight years: the mean intercepts differ by 138 years. This behaviour is not limited to this pair of dates: in the period 3000–5000 ^{14}C yr BP the difference between successive calibrated dates at 10-year intervals is greater than 50 cal. years 14 times (Figure 2). These large differences reflect the gain and loss of intercepts as wiggles in the calibration curve are passed. The small changes in the weighted average reflect the small changes in the probability density distribution between successive dates. The median and the 95% quantile of the absolute difference between successive dates, minus the mean, for the period 3000–5000 ^{14}C yr BP is given in

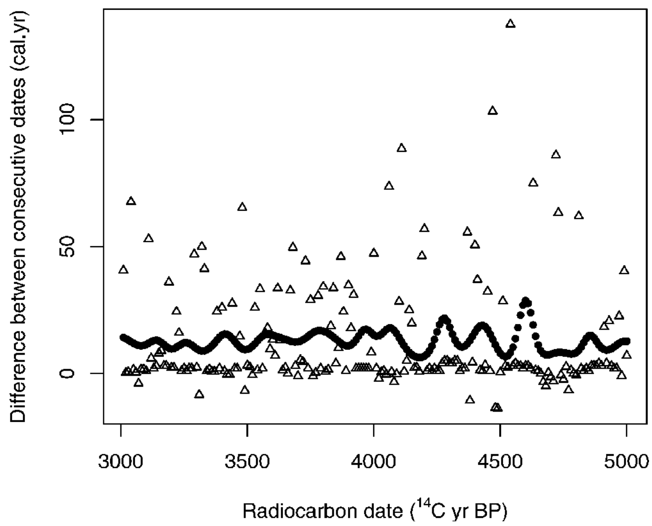

Figure 2 Difference in calibrated years between consecutive calibrated radiocarbon dates (at 10-year spacing) for the period 3000–5000 ^{14}C yr BP σ = 50 yr. Open triangle = mean intercept; filled circle = weighted average.

Table 1. The methods can be divided into two groups: the first, containing the weighted average, median, and weighted average of the mid-points of the ranges, has low median differences, indicating that they are stable. A second group based on intercepts or modes has substantially higher median differences between adjacent samples. This group is sensitive to small changes of the ^{14}C date. The median is normally regarded as a robust statistic, so it was unexpected to find that the mean intercept is slightly more stable than the median intercept. This behaviour can be explained by what happens when a ^{14}C date with a single intercept, so the mean and median are co-incident, is moved until it crosses a wiggle in the calibration curve. It will now have three intercepts; one of the two new intercepts will be the new median, but the new mean intercept will have changed less.

The magnitude of the standard deviation of a radiocarbon date on a particular sample will depend on the sample mass and counting duration (Aitken, 1990). A low standard deviation will give a tight calibrated probability distribution (although possibly multimodal); a large standard deviation, for the same mean, will give a broader probability distribution, with a different shape. A well-behaved central-point estimate should respond to this changed shape. Methods that just use the intercept or mode give the same value whatever the standard deviation. Weighted average- and median-based methods respond to this changed distribution, but it is not possible to rank their response. Weighted average of the ranges based on their mode is very sensitive to ranges merging as the standard deviation increases.

Table 1 Absolute difference between consecutive decadal radiocarbon dates minus the mean difference

Method	median	95% quantile
Weighted average	2.4	6.4
Median	3.3	11.3
Weighted average of ranges using range mid-points	4.7	14.1
Weighted average of ranges using range intercept or mid-point	8.8	44.2
Weighted average of ranges using range mode	9.6	77.7
Mean intercept	11.7	44.5
Median intercept	14.5	68.4
Mode	10.8	82.1

The errors reported for radiocarbon errors do not always reflect the full uncertainty attached to a date, as shown by the radiocarbon laboratory intercomparisons (Boaretto *et al.*, 2002). This has led to the occasional use of 'error multiplier terms'. The use of these multipliers will have no impact on intercept-based methods of estimating the central point, but the weighted average will change to reflect the changing probability distribution.

The true calibration curve is not precisely known and further updates or refinements (e.g., Gómez Portugal Aguilar *et al.*, 2002) are likely. A well-behaved central-point estimate should change only slightly if small adjustments are made to the calibration curve. The intercept method will be highly sensitive to changes in the calibration curve as new intercepts will occur and old ones will be lost. Other methods of estimating the central point should be much less sensitive to this problem.

When radiocarbon dates are taken in sequence, for example down a core, this information can be used in the Bayesian statistical framework to reduce the uncertainty around each date (Biasi and Weldon, 1993). After such constraints have been taken into consideration, the intercept of the radiocarbon date with the calibration curve has no relevance; it may even be outside the 95% confidence limits of the adjusted date. The mode, the year with the highest probability, may appear to be an obvious alternative, but it exhibits unstable behaviour in multimodal distributions, switching between nodes if there are small changes in the prior information. A weighted average of the probability distribution can still be made and is stable.

It is now recognized that, when radiocarbon dates are calibrated, both the wiggles in the calibration curve and the uncertainty in the radiocarbon result need to be fully accounted for. The intercept method fails to do this (Bowman and Leese, 1995). A Bayesian methodology, as employed by most calibration programs (e.g., OxCal; Bronk Ramsey, 1995), is required to incorporate both these issues. It would be ironic if these advanced statistical tools were only used to find the 1σ range of the date, but the intercept used for all subsequent calculations.

Conclusions

No single value can adequately describe the complex shape of a calibrated radiocarbon probability density function, and wherever possible this full distribution should be used. When a single estimate must be used, a robust estimate, such as the weighted average or median, should be used and the method specified. Intercept-based methods should be avoided as they are sensitive to small changes in the mean of the radiocarbon date. The large difference between the central points given by the different methods (median = 47, maximum = 200 years, testing every 10th year between 3000 and 5000 ^{14}C yr BP, $\sigma = 50$ yr) means this is not a trivial issue and has a substantial impact on the derived chronology in calibrated years.

Acknowledgements

This research was supported by NORPAST-2, funded by the Norwegian Research Council, and PACLIVA, an EU Framework 5 Programme project (EVK2-CT2002–000143). This is publication no. A0016 of the Bjerknes Centre for Climate Research.

References

Aitken, M.J. 1990: *Science-based dating in archaeology*. London: Longman.

Bartlein, P.J., Edwards, M.D., Shafer, S.L. and **Barker, E.D.** 1995: Calibration of radiocarbon ages and the interpretation of paleoenvironmental records. *Quaternary Research* 44, 417–24.

Bennett, K.D. 1994: Confidence intervals for age estimates and deposition times in late-Quaternary sediment sequences. *The Holocene* 4, 337–48.

Biasi, G.P. and **Weldon, R.** 1993: Quantitative refinement of calibrated ^{14}C distributions. *Quaternary Research* 41, 1–18.

Boaretto, E., Bryant, C., Carmi, I., Cook, G., Gulliksen, S., Harkness, D., Heinemeier, J., McClure, J., McGee, E., Naysmith, P., Possnert, G., Scott, M., van der Plicht, H. and **van Strydonck, M.** 2002: Summary findings of the Fourth International Radiocarbon Intercomparison (FIRI) (1998–2001). *Journal of Quaternary Science* 17, 633–37.

Bowman, S.G.E. and **Leese, M.N.** 1995: Radiocarbon calibration: current issues. *American Journal of Archaeology* 99, 102–105.

Bronk Ramsey, C. 1995: Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37, 425–30.

Brown, S., Bierman, P., Lini, A., Davis, P.T. and **Southon, J.** 2002: Reconstructing lake and drainage basin history using terrestrial sediment layers: analysis of cores from a post-glacial lake in New England, USA. *Journal of Paleolimnology* 28, 219–36.

Gómez Portugal Aguilar, D., Litton, C.D. and **O'Hagan, A.** 2002: Novel statistical model for a piece-wise linear radiocarbon calibration curve. *Radiocarbon* 44, 195–212.

Ihaka, R. and **Gentleman, R.** 1996: R: a language for data analysis and graphics. *Journal of Computational and Graphical Statistics* 5, 299–314.

Seierstad, J., Nesje, A., Dahl, S.O. and **Simonsen, J.R.** 2002: Holocene glacier fluctuations of Grovabreen and Holocene snow-avalanche activity reconstructed from lake sediments in Grøningstølsvatnet, western Norway. *The Holocene* 12, 211–22.

Stuiver, M. and **Reimer, P.J.** 1993: Extended ^{14}C database and revised CALIB 3.0 ^{14}C age calibration program. *Radiocarbon* 35, 215–30.

Stuiver, M., Reimer, P.J., Bard, E., Beck, J.W., Burr, G.S., Hughen, K.A., Kromer, B., McCormac, G., van der Plicht, J. and **Spurk, M.** 1998: INTCAL98 radiocarbon age calibration, 24000–0 cal. BP. *Radiocarbon* 40, 1041–83.