SYSTEMATIC BIAS OF RADIOCARBON METHOD

Adam Walanus

Department of Geoinformatics, AGH University of Science and Technology, al. Mickiewicza 30, 30-059 Kraków, Poland.
Email: walanus@geol.agh.edu.pl.

ABSTRACT. Systematic bias of dates can became statistically significant regarding the growing global number of dates connected with the calibration curve plateau. For example, samples of true age in the span 800–700 BC are dated to be roughly 100 younger, on average. The curve of expected bias for a given age is presented. To avoid such a bias, the Bayesian paradigm probably must be modified in some way.

INTRODUCTION

A well-recognized feature of the radiocarbon calibration curve is present in the shape of its so-called plateaus, i.e. generally flat areas of the curve, which make the ^{14}C date to be the same, independent of the real age of the sample within the plateau span. The resulting probability distribution of the calendar age is roughly the same for samples of any true age, e.g. T_1, T_2, T_3 in Figure 1. For the oldest sample from the plateau (T_1), the user is confronted with the "mass" of probability moved to the right. The expected value of "date" ($E[T]$) for given probability distribution is far from the true age of sample (T_1), actually it is younger ($E[T] > T_1$). The age T_2 is in the middle of the calibrated probability distribution (Figure 1); therefore, $E[T] = T_2$, and T_2 is the (only) age from within the plateau to be dated "correctly."
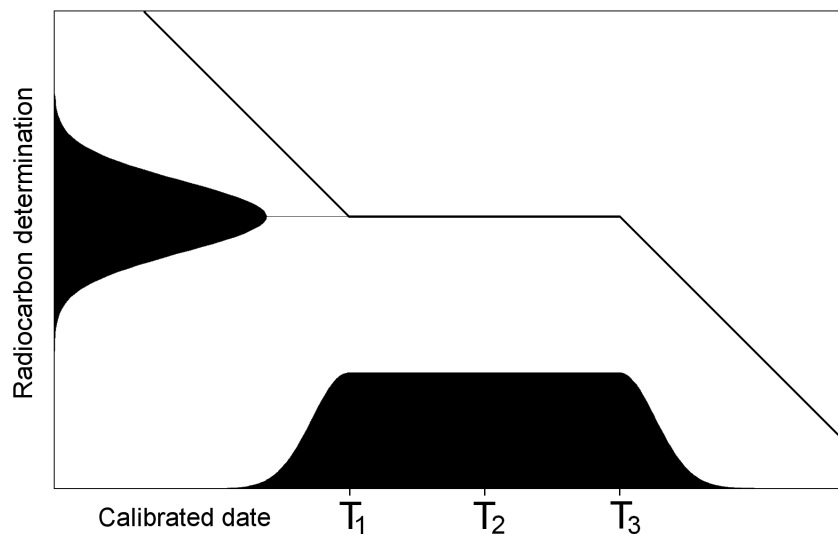

Figure 1 An idealized picture of the ^{14}C calibration curve plateau. One ^{14}C age is expected for a series of different ages. As a result, the calibrated age is biased for most true ages. Among the example ages shown here (T_1, T_2, T_3), only T_2 has a probability that is balanced, i.e. distributed equally, half and half on both sides (for probabilistic calibration; see e.g. Michczyńska et al. 1989.)

Thus, the question is not the poor precision of dating obtained in the plateau (Guilderson et al. 2005), but the systematic bias of dating results, which is much more dangerous.

THE PROBLEM

The difference between the expected value of the age, after dating, and the true age, i.e. the value of $(E[T] - T_1)$, can be called the expected bias. The probabilistic notion of the "expected value" of the probability distribution is unique. However, it would be easier to imagine a situation when the user takes a single number as a sample date (keeping in mind the difficulty with "point estimation," see Michczyński 2007). It is not so far from reality. A good example would be the confirmation of a precise archaeological date of an artifact. Assuming that the artifact has, in some sense, a random age, its age can be treated as "drawn from" the distribution, so far as the scientist's conviction about the date confirmation is proportional to the probability density value. The frequentist interpretation of probability, however, is best used here. If there are many samples of true age around T_1, most of them will be younger. The opposite is true for samples of age T_3. Samples of given age (e.g. T_1) are dated worldwide and an average systematically biased age is drawn from the ^{14}C dating procedure. As a result, the event or short period would be dated incorrectly by archaeology.

QUANTIFICATION

The curve of bias (Figure 2) is obtained in a way that models a real situation of simple ^{14}C dating of a single sample. The Monte Carlo method is applied; however, it is numerically irrelevant to the obtained result as the random error is small. Modeling reality with some application of a random element seems to be a reasonable idea. The procedure of curve creation is repeated independently for each year from the span 20,000 BC to modern. The yearly resolution is too high in fact; however, it can be used for checking the statistical precision of the Monte Carlo calculations.

The procedure is as follows:

1. The measurement error σ is assumed (25, 50, or 100 yr).
2. For given (calendar) age T, the ^{14}C age T_r is read out from the calibration curve (IntCal04, Reimer et al. 2004). The error of the calibration curve is neglected. (For $T > 13,000$ BC, the error is far from being negligible; however, here not the absolute value of the calibration curve is critical but only its shape.)
3. The measurement result T_m is drawn from the Gaussian distribution $N(T_r, \sigma)$. The standard normal random variable (z) generator is used in the formula: $T_m = T_r + z \times \sigma$. (In Figure 1, the special case of $z = 0$ is adopted.)
4. The standard calibration procedure is applied to the ^{14}C date $T_m \pm \sigma$ to obtain the probability distribution $f_{cal}(T)$.
5. The (single) value T_d of a "date" is drawn randomly from the $f_{cal}(T)$. This should resemble the acceptance by the scientist of the value T_d as the real sample age. The acceptance is assumed to be in accordance with the ^{14}C dating result $f_{cal}(T)$. (It is worth mentioning that $f_{cal}(T)$ is obtained assuming a uniform *a priori* distribution of T.) Technically, the rejection sampling method (Robert and Casella 2004) is used to generate T_d. That method corresponds well with the ^{14}C user's intuition, as the chance for the age to be equal to T_d is simply proportional to $f_{cal}(T_d)$.
6. Points 3–5 are repeated 1000 times, producing the same number of T_d values, which are used to compute the average, T_d^{av}. The bias is assumed to be equal to $T_d^{av} - T$.
7. Next, we go to point 2 with the next T (21,940 times).
8. Then, go to point 1 with the next σ (3 times).
9. The resulting curves of bias as a function of age are smoothed by a 5-yr Gaussian filter, mainly for aesthetic reasons.
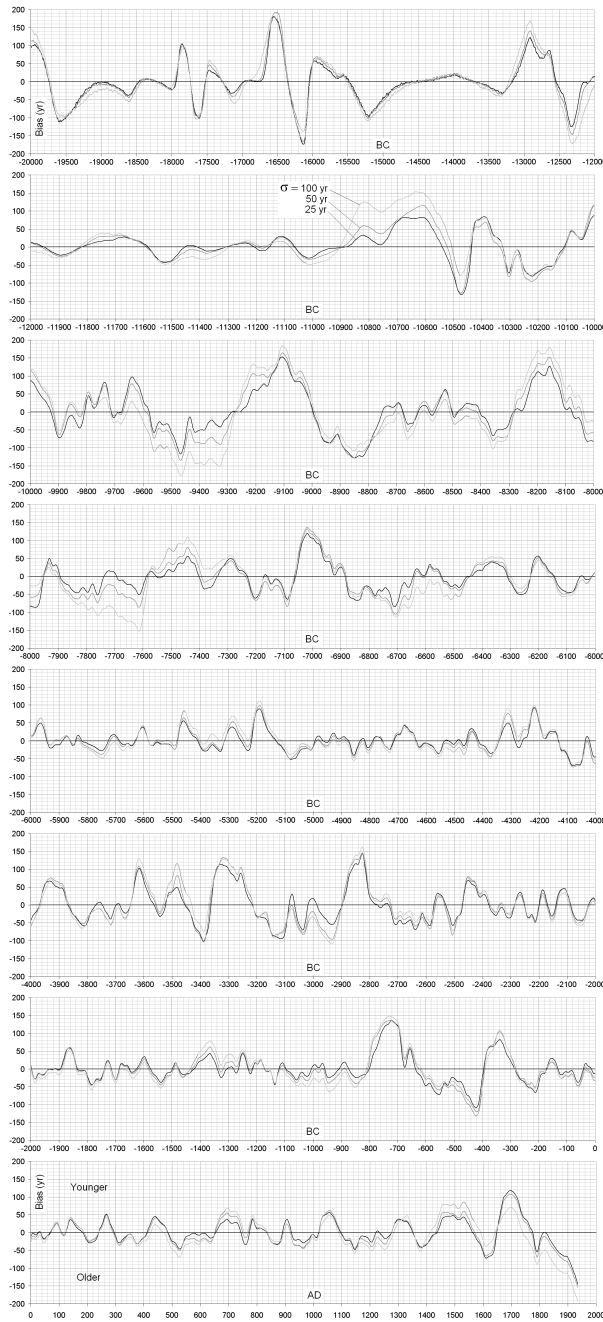
Figure 2 The curves of bias of the ^{14}C dating method. For a given true sample age (X axis), the difference is plotted (Y axis) between the average date and the true age. Positive values indicate that the accepted date will be younger on average; for negative values, the age will be older. Three curves correspond to the values of the measurement error $\sigma = 25$, 50, and 100 yr; however, the bias depends weakly on σ. IntCal04 is used (Reimer et al. 2004).

DISCUSSION

The curve presented in Figure 2 is simply another point of view on the shape of the calibration curve. The calibration curve can easily be presented in such a way that its shape will relatively closely resemble the presented curves. However, the interpretation of its values obtained by subtracting the trend line would not be clear. Also, the interference of the measurement error is lost in such a presentation. The approach proposed here seems to be close to archaeological practice. Having many samples dated by ^{14}C and another, independent method could result in significant (statistical and practical) discrepancies. For example, the systematic difference is about 50 yr for samples of real age between 400 and 300 BC. It must be noted that proper interpretation of single or many calibrated probability distributions is necessary; however, when some kind of meta-analysis is performed, the bias will emerge.

The value of bias approaches 150 yr in a few cases only; however, 100 yr is a typical maximum value. Such a value seems to be far from being negligible when compared to the contemporary ^{14}C measurement precision (σ). Of course, at time positions where the bias-curve approaches 100 yr or more, the calibrated distribution is wider than $4\,\sigma$ (95%). It must be mentioned, however, that even a 20–50 yr bias can influence archaeological reasoning in cases when many samples are already dated at more or less pronounced plateaus. This is due to the simple statistical rule that the expected random error of the average value is inversely proportional to the square root of the number of dates. Therefore, theoretically, the bias will become statistically significant at some point with a growing number of dates.

The difficulties with the frequency distribution of ^{14}C dates, as applied to climate research (Stolk et al. 1989; Michczyńska et al. 2007), resulting from the calibration curve shape, can be resolved with the proposed approach. The presented curve is obtained for precisely equally spaced "true" dates (1 sample per year), and resulting ages are biased left or right, resulting in a non-uniform shape of the cumulative distribution. The Monte Carlo method, applied here, is numerically nearly equivalent to the summation of calibrated distributions.

To overcome the problem of bias, the typically uniform prior distribution in the standard Bayesian calibration procedure (Buck et al. 1996) should probably be modified in some way, according to the calibration curve. The idea of a robust Bayesian analysis (Steier et al. 2001) may be useful here.

REFERENCES

Buck CE, Cavanagh WG, Litton CD. 1996. *The Bayesian Approach to Interpreting Archaeological Data.* Chichester: Wiley. 402 p.

Guilderson TP, Reimer PJ, Brown TA. 2005. The boon and bane of radiocarbon dating. *Science* 307(5708): 362–4.

Michczyńska DJ, Pazdur MF, Walanus A. 1989. Bayesian approach to probabilistic calibration of radiocarbon ages. *PACT* 29:69–79.

Michczyńska DJ, Michczyński A, Pazdur A. 2007. Frequency distribution of radiocarbon dates as a tool for reconstructing environmental changes. *Radiocarbon* 49(2):799–806.

Michczyński A. 2007. Is it possible to find a good point estimate of a calibrated radiocarbon date? *Radiocarbon* 49(2):393–401.

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW,

Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Hogg AG, Hughen KA, Kromer B, McCormac G, Manning S, Bronk Ramsey C, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE. 2004. IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–58.

Robert CP, Casella G. 2004. *Monte Carlo Statistical Methods.* 2nd edition. New York: Springer-Verlag. 645 p.

Steier P, Rom W, Puchegger S. 2001. New methods and critical aspects in bayesian mathematics for ^{14}C calibration. *Radiocarbon* 43(2A):373–80.

Stolk AD, Hogervorst K, Berendsen H. 1989. Correcting ^{14}C histograms for the non-linearity of the radiocarbon time scale. *Radiocarbon* 31(2):169–78.