Bayesian Analysis (2008) **3**, Number 2, pp. 255–262

Comment on article by Blackwell and Buck

Andrew R. Millard*

1 Introduction

Radiocarbon dating has come a long way since Willard Libby first conceived of the method (Libby et al. 1949; Arnold and Libby 1949) that was to earn him a Nobel Prize. Today we might consider Libby to be lucky that he could not detect the deviation from the expected radiocarbon decay curve. This was due to a combination of the imprecision of his measurements and his known age samples only spanning the last 5000 years. If Libby had been able to make more precise measurements the technique might never have been developed as the fundamental assumptions would have appeared to have been breached from the very start. Of course the deviation of radiocarbon content of known age samples from the expected values was discovered within a few years of Libby's initial publications and radiocarbon calibration curves have been in development ever since. The latest version of the calibration curve is INTCAL04 (Reimer et al. 2004), and owes its mathematical formulation to Buck and Blackwell (2004, Buck et al. 2006), and future curves will depend on their elaboration and extension of that description in the paper under discussion here.

2 Advances

The most important advances made by the method of Blackwell and Buck are: formal error estimation, a physically-based model, an explicit error structure on the calibration curve and the inclusion of calendar age uncertainty in the calculations.

2.1 Formal error structure

Blackwell and Buck provide for formal error estimation using rigorous mathematics, which is in contrast to the ad hoc methods previously used in the construction of curves. With IntCal04 we also have the advantage that the underlying data used to construct the curve are published as well as the method: consequently anyone can replicate the calculations if they find that the published curve expressed as means and standard deviations at five year intervals is insufficient for their needs.

2.2 Physical basis

The new model is partially based on the work of Gómez Portugal Aguilar et al. (2002) which sets out a physical basis for the model as representing the fluctuations in the

*Department of Archaeology, University of Durham, U.K., mailto:a.r.millard@durham.ac.uk

 DOI:10.1214/08-BA309B

production rate of ^{14}C in the upper atmosphere as random fluctuations around a mean value, which translate into a random walk description of the calibration curve. However, as some of the participants in the workshop observed, the calibration curve does not look like a simple random walk. This is true and we do know that there are non-random fluctuations in the production rate, caused by, for example, long term trends in the strength of the geomagnetic field and the 11-year sunspot cycle. Blackwell and Buck's model is therefore limited in its physical basis, but as these variations are complex and not fully understood, the random walk is a useful simple model, and, importantly, it is adaptable to include more structure in the future.

2.3 Calendar age uncertainty included

The incorporation of the uncertainty in calendar ages beyond the section of the curve dated exactly by dendrochronology is a significant advance, as these uncertainties can be as large as the radiocarbon measurement uncertainties. Previous estimates of the calibration curve did not include this.

3 Wiggle-matching

The approach to the calibration curve construction presented by Buck and Blackwell (2004, and the current paper) produces curves which are much smoother than previous curves, and which also involve the explicit calculation of a covariance structure on the curve. In this discussion I want to explore the implications of these two results for one of the applications of radiocarbon dating that yields the most precise ages – wiggle matching. Wiggle-matching (the title is really a misnomer, but now has wide acceptance in the radiocarbon community) involves taking a series of radiocarbon dates where the prior knowledge about the true calendar dates of the samples can be expressed as known differences in age between those samples, or occasionally as differences in age with some small uncertainty. The series of radiocarbon dates can then be matched to the calibration curve to provide a relatively precise estimate of age. When the results are plotted on a graph the 'wiggles' in the sample sequence of radiocarbon dates match the wiggles in the calibration curve - hence the name - though the method would also work if the calibration curve were a straight line.

The production of a smoother calibration curve raised fears among some users that as some wiggles disappeared, there would be a reduction in the precision of wiggle-matched results, and this is conceivably true, as the position of a sharp wiggle appearing in both calibration curve and data will tend to improve the precision of the match.

The new approach to the calibration curve demands some revisions to the wiggle-matching process. In constructing the curve Blackwell and Buck have explicitly accounted for the fact that most radiocarbon dates on dendrochronologically dated pieces of wood are 'blocked', and it is appropriate to do the same with the set of dates to be wiggle-matched. The computations discussed in this comment therefore allow for this. As IntCal04 is published without a covariance structure, it is allowed for here by using the published underlying data and the statistical model of the calibration curve implemented in WinBUGS 1.4 (Spiegelhalter et al. 2004) to generate a calibration curve at single-year resolution with full posterior covariance included.

Case study	Known date	Mode and 95% hpd region by different methods					
		Christen & Litton with IntCal98		IntCal04 with covariance		IntCal04 without covariance	
Stolford log	5781	5774	5766-5778	5773	5760-5780	5774	5765-5776
Hallstatt plateau	2400	2405	2360-2424	2397	2360-2427	2398	2361-2428
Thera	–	3567	3549-3577	3567	3550-3575	3567	3551-3576
Early	11943	11944	11748-11782, 11884-12148	11943	11898-11973	11943	11917-11967

Table 1: Modes and hpd regions from various wiggle-matches (all dates in cal BP).

Table 1 summarises the posterior probability densities for the end date of the sequences from four case studies, with and without the covariance on the curve and, for comparison, using the method of Christen and Litton (1995) to match the sequence to IntCal98 using OxCal 3.10 (Bronk Ramsey 2001). The Stolford Log was used in the first paper on Bayesian wiggle-matching (Christen and Litton 1995). The last ring has a known date of 5781 cal. BP, which falls just outside the 95% highest posterior density (hpd) regions, but it is clear that the mode changes only slightly when the covariance is included, and though the hpd region shifts, it changes its length only slightly. Some regions of the calibration curve are quite flat, and the second example uses one of the samples used to construct the calibration curve from the part of the curve known as the Hallstatt Plateau (after the Hallstatt period in central Europe with which it coincides). As expected, the results are much less precise in this region, though shifts between the hpd regions from the various methods are minimal, at one year. The third example is a sample of olive wood of unknown date relating to one of the great archaeological dating controversies of the last 20 years: the date of the eruption of the volcano on the island of Thera (or Santorini) in the Aegean (Friedrich et al. 2006). Again we see just one or two year shifts in modes and hpd regions. The fourth example is from a much earlier period where the density of the data making up the calibration curve is much less, and where there has been more change between IntCal98 and IntCal04. Here including the covariance does not change the mode, but the hpd region is increased from 50 years to 75 years in length, and the result using IntCal98 differs significantly in its hpd, but not its mode.

In the first three examples the difference between the calibration curves and methods is minimal. In each of these cases a plot of the posterior probability distribution shows a slightly more peaked distribution when the covariance is accounted for than when it is not, but the changes in hpd (which is what archaeologists tend to rely on for interpretation) are small and insignificant. My statistical intuition suggests that including covariance should influence the uncertainty, so why is this the case? The first three examples are in a region of the curve which is dense in data, so perhaps the covariance is reduced relative to the example in Blackwell and Buck's paper, and relatively unimportant in determining the results. This suggestion is borne out by the fourth example where including the covariance does make a difference, with increased uncertainty. This small set of examples was not chosen to systematically explore areas of the calibration curve with differing features, so there may be variation in other periods, and especially in the early Holocene and late Pleistocene where there is a much lower density of data than for the first three examples examined here.

4 Beyond Dendrochronology

I will comment on two aspects of the calibration curve production where it goes beyond the use of the dendrochronological data.

First, in the use of radiocarbon dates and uranium-series dates, Blackwell and Buck assume a normal likelihood and an improper uniform prior for the uranium-series date. As I have shown (Millard 2006), a likelihood based on the dating equation is asymmetric and improper. If this is used then suitable priors for the true age of the sample which have a logical basis and yield a proper posterior distribution are bounded uniform or exponential. In practice however, there are prior constraints on the true age from stratigraphic ordering, and for high precision dates the normal distribution is a good approximation. These problems are therefore probably not really problems, but they could do with a more detailed consideration to confirm this. What is likely to be more important is the fact that with high precision uranium-series dates, the uncertainty on the decay constants of the radionuclides start to become a significant contribution to the uncertainty of the dates, and these will be correlated between all uranium-series dates.

Second, though we now have a good statistical method for estimating the calibration curve, much of the portion beyond 12 ka (i.e. 12 thousand years ago) is based on marine samples, such as corals. Due to features of the global carbon cycle, the calibration curve for marine samples differs to that for terrestrial samples, and so for this portion of IntCal04 and the whole of the Marine04 curve for marine environments, a model of the ocean carbon cycle is used to convert between the two curves. In fact, if one follows a chain of citations back to the original publication, the model being used is that of Oeschger et al. (1975). This rather old model has been superseded for almost all other purposes and we certainly have much more sophisticated models today, and also the computing power required to incorporate them into the construction of the calibration curve.

5 Future modelling of the calibration curve

The new statistical approach to the calibration curve is a major step forward which opens up possibilities for a variety of improved approaches to modelling the curve. Blackwell and Buck suggest that the random-walk prior could be modified to include solar cycles. This is certainly possible, but care should be taken that the knowledge of solar cycles used is not derived from analyses of the periodicities to be observed in the previous radiocarbon calibration curves. Additional prior information that could be incorporated includes predictions of radiocarbon production from information on ^{10}Be production (^{10}Be has a similar production mechanism to ^{14}C) and data on the history of the earth's magnetic field. Some work has already attempted forward modelling of atmospheric ^{14}C levels from such data using a carbon-cycle model (e.g. Laj et al. 2002). This resulted in a very imprecise envelope for predicted ^{14}C levels, but could provide a useful prior for the calibration curve in the period prior to 12 ka.

Finally, Blackwell and Buck suggest that a fully Bayesian approach would allow most radiocarbon dating samples to be used so that they were calibrated jointly with updating a continuously revised calibration curve. This would mean that each user of the calibration curve would end up with their own personal posterior estimate of the curve (dependent on the particular dates they had calibrated), and, importantly, unless the posterior probability distributions of all previously calibrated dates were updated when a new date came along, the results of calibration would depend on the order in which dates were calibrated. The prior information provided by users would influence the updated calibration curve, but sometimes such information appears strong initially and is then found to be wrong. Further, the prior information for what the user conceives of as separate calibrations can be strongly correlated, which would have to be taken into account in the updating of the calibration curve. It is also far from clear how model comparison or choice could be conducted as the chronological models usually differ in their expression of the prior information available. I think that the type of joint inference proposed, whilst philosophically satisfying, could not be implemented in practice, as it presumes a Bayesian statistical sophistication on the part of users which is unlikely and may also be unwanted by some who see the Bayesian formulation of the calibration curve and of the calibration process as merely a convenient method and not a philosophical approach. Conversely, it is also quite possible that the calibration data are so numerous and precise for the last 12 ka that the changes introduced this way would be small.

6 Conclusions

The IntCal04 model and the development of it laid out by Blackwell and Buck is a major advance in the construction of the radiocarbon calibration curve, which lays the basis for future developments. Although it is good to know about the covariance structure of the posterior estimate of the curve, it appears to have only insignificant effects on the results of calibration for Holocene samples. Future effort in the development of this model should be directed towards the portion beyond 12 ka, with more detailed consideration of calendar age uncertainties and especially incorporation of more sophisticated prior information to improve the estimates.

References

Arnold, J. R. and Libby, W. F. (1949). "Age Determinations by Radiocarbon Content: Checks with Samples of Known Age." *Science*, 110: 678–680. 255

Bronk Ramsey, C. (2001). "Development of the Radiocarbon Program OxCal." *Radiocarbon*, 43(2A): 381–389. 257

Buck, C. and Blackwell, P. G. (2004). "Formal Statistical Models for Estimating Radiocarbon Calibration Curves." *Radiocarbon*, 46: 1093–1102. 255, 256

Buck, C., Gómez Portugal Aguilar, D., Litton, C., and O'Hagan, A. (2006). "Bayesian non-parametric estimation of the calibration curve for radiocarbon dating." *Bayesian Analysis*, 1(2): 265–288. 255

Christen, J. A. and Litton, C. D. (1995). "A Bayesian approach to wiggle-matching." *Journal of Archaeological Science*, 22: 719–725. 257

Friedrich, W. L., Kromer, B., Friedrich, M., Heinemeier, J., Pfeiffer, T., and Talamo, S. (2006). "Santorini eruption radiocarbon dated to 1627-1600 BC." *Science*, 312(5773): 548–548. 257

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(1): 195–212. 255

Laj, C., Kissel, C., Mazaud, A., Michel, E., Muscheler, R., and Beer, J. (2002). "Geomagnetic field intensity, North Atlantic Deep Water circulation and atmospheric Δ14C during the last 50 kyr." *Earth and Planetary Science Letters*, 200: 177–190. 259

Libby, W. F., Anderson, E. C., and Arnold, J. R. (1949). "Age determination by radiocarbon content: world wide assay of natural radiocarbon." *Science*, 109: 227–228. 255

Millard, A. R. (2006). "Bayesian analysis of Pleistocene Chronometric Methods." *Archaeometry*, 48(2): 359–375. 258

Oeschger, H., Siegenthaler, U., Schotterer, U., and Gugelmann, A. (1975). "A box diffusion model to study the carbon dioxide exchange in nature." *Tellus*, 27: 168–192. 258

Reimer, P. J., Baillie, M. G., Bard, E., Bayliss, A., Beck, J. W., Bertrand, C. J., Blackwell, P. G., Buck, C. E., Burr, G. S., Cutler, K. B., Damon, P. E., Edwards, R. L., Fairbanks, R. G., Friedrich, M., Guilderson, T. P., Hogg, A. G., Hughen, K. A., Kromer, B., McCormac, G., Manning, S., Ramsey, C. B., Reimer, R. W., Remmele, S., Southon, J. R., Stuiver, M., Talamo, S., Taylor, F., van der Plicht, J., and Weyhenmeyer, C. E. (2004). "IntCal04 Terrestrial Radiocarbon Age Calibration, 0-26 Cal Kyr BP." *Radiocarbon*, 46: 1029–1058. 255

Spiegelhalter, D., Thomas, A., Best, N., and Lunn, D. (2004). *WinBUGS Version 1.4 User Manual*. Cambridge: MRC Biostatistics Unit.
URL http://www.mrc-bsu.cam.ac.uk/bugs 257

Acknowledgments

I would like to thank the organisers of *Case Studies in Bayesian Statistics Workshop 9* for inviting me to Pittsburgh to be a discussant and providing funds towards my travel, as well as for arranging an excellent conference.