

NOVEL STATISTICAL MODEL FOR A PIECE-WISE LINEAR RADIOCARBON CALIBRATION CURVE

Delil Gómez Portugal Aguilar^{1,2} • Cliff D Litton³ • Anthony O'Hagan¹

ABSTRACT. The process of calibrating radiocarbon determinations onto the calendar scale requires the setting of a specific statistical model for the calibration curve. This model specification will bear fundamental importance for the resulting inference regarding the parameter of interest—namely, in general, the calendar age associated to the sample that has been ¹⁴C-dated.

Traditionally, the ¹⁴C calibration curve has been modelled simply as the piece-wise linear curve joining the (internationally agreed) high-precision calibration data points; or, less frequently, by proposing spline functions in order to obtain a smoother curve.

We present a model for the ¹⁴C calibration curve which, based on specific characteristics of the dating method, yields a piece-wise linear curve, but one which rather than interpolating the data points, smooths them. We show that with this specific model if a piece-wise linear curve is desired, an underlying random walk model is implied as covariance structure (and vice versa). Furthermore, by making use of all the information provided by the calibration data in a comprehensive way, we achieve an improvement over current models by getting more realistic variance values for the calibration curve.

INTRODUCTION

It is well known that due to violations in the original assumptions of the ¹⁴C dating method, ¹⁴C ages need to be calibrated onto the calendar scale in order to be of any use for archaeologists, climatologists and other users of the method. Thus, calibration has become a standard procedure, performed generally by means of one of the many available calibration programs.

A key element in the calibration process is the actual calibration curve, describing the relationship between ¹⁴C ages and calendar ages. Different approaches have been proposed for this matter. A simple, yet commonly accepted proposal consists of building the ¹⁴C calibration curve by joining the calibration data points with straight-line segments. Commonly accessible computer calibration programs such as CALIB (Stuiver and Reimer 1993) and that of Pazdur and Michczynska (1989) follow this proposal. However, although this has been traditionally considered as *the* calibration curve—in fact this is the way in which the high-precision calibration data is commonly presented (see Stuiver and Pearson 1986 and Pearson and Stuiver 1986, and the most recent internationally agreed high-precision calibration data set INTCAL98, in the time interval involving dendrochronologically established calendar ages, Stuiver et al. 1998), it has the shortcoming of failing to reflect the fact that the calibration data are themselves subject to uncertainty. Randomness in the data means that simply connecting the calibration data points produces a curve that is almost certainly “more wiggly” than the underlying true calibration curve.

A less frequent approach consists of adjusting a cubic spline to the calibration data. This kind of curve has the appeal of having continuity in their slope throughout the time scale, and is considered by some as the best representation of a natural process (McCormac and Baillie 1993). However, the use of cubic splines requires the additional input of a “smoothing parameter” S . This parameter takes values greater or equal to zero, and determines the degree of smoothing of the resulting spline. The majority of authors who suggest the use of a cubic spline for calibration curve set $S = 0$, leading to the “natural cubic spline”, which interpolates the data points—i.e. no smoothing (McCormac and

¹Department of Probability and Statistics, University of Sheffield, Sheffield S3 7RH, United Kingdom

²Corresponding author. Email: D.GomezPortugal@sheffield.ac.uk.

³School of Mathematical Sciences, University of Nottingham, Nottingham NG7 2RD, United Kingdom

Baillie 1993; van der Plicht and Mook 1989; van der Plicht 1993). Thus, once more the uncertainty in the calibration data is not reflected in the curve. In contrast, any value for S greater than zero implies that the resulting spline smooths rather than interpolates the calibration data; the greater the value of S , the smoother the resulting spline. Talma and Vogel (1993) utilize the value $S = 1$, suggesting that this is the ideal value, on considerations of counting statistics (given that deviations of the data points from the spline with $S = 1$ closely obey Gaussian statistics). Although not incorporated into a calibration program as such, Knox and McFadgen (1997) also aim for a curve such that there is continuity in the derivatives—i.e. a smooth calibration curve. Their approach is based on the use of Fourier transformations and other filtering techniques, and results in a curve that does not pass through the calibration data.

Our work constitutes a *model-based* approach for estimating the ^{14}C calibration curve. We provide a statistical model for the relationship between calendar ages and ages obtained through ^{14}C dating which will acknowledge the uncertainty in the data, as well as taking into account the uncertainty in the calibration curve itself. Throughout this paper, by “model” we mean *statistical* model, and by “data” we mean the observations which make part of the high-precision calibration data. We will work within the *Bayesian* framework of Statistics, so the results obtained will be a combination of all the information about the parameters in the statistical model available *a priori* (i.e., separately from having access to the calibration data points), and the information provided by the data. The importance of incorporating archaeological expert knowledge has been stressed in Buck et al. (1994). Indeed, this is an area where there will often be expert opinion available about the unknown quantities (or *parameters*) of the problem. The prior information may assume an infinitely wide range of forms, according to the specific problem at hand. Some examples of the incorporation of prior information in archaeological dating problems can be found in Naylor and Smith (1988), Buck et al. (1991), Christen and Litton (1995), and Zeidler et al. (1998).

We are aware that the processes underlying the theory behind ^{14}C dating are indeed complex. For instance, the fact that the calibration of a ^{14}C determination in order to determine the age of the corresponding sample depends on elements such as its geographical location, whether it is a marine or terrestrial sample, etc.

The model we propose relates calendar and ^{14}C ages in a very simple way, and allows for a logical incorporation of all the information provided by the data—including the uncertainty associated to each data point. It is based on representing the true calibration curve as a continuously evolving Gaussian Process, a modeling technique that has been used to represent unknown functions in a wide variety of statistical problems (e.g., O’Hagan 1992; O’Hagan et al. 1999; Neal 1999). The form of Gaussian Process that we use is informed by scientific understanding of how the atmospheric ^{14}C concentration varies over time.

Here we will compare the proposed approach with, on one hand, the approach of Christen (1994), which constitutes the statistical basis for the recently launched on-line calibration program BCal, and, on the other, with the much simpler approach taken by at least one of the most commonly used calibration programs—namely, CALIB—Stuiver and Reimer 1993. (OxCal [Bronk Ramsey 1995] provides options so that it can also be set to follow the same piece-wise linear approach as CALIB.)

In Section 2 we present the notation and conventions employed throughout the paper, as well as the format of the data.

Section 3 deals with the modeling of the ^{14}C calibration curve. Here we describe the characteristics of a Gaussian Process model, and we establish the specific model we propose for the ^{14}C calibration

curve. Of great importance in the model is the way we handle the uncertainty involved in the process. That is, the covariance structure chosen to reflect the influence of any point on the curve on any other. Hence, we devote attention to deriving a suitable structure and we present it at the end of this section. (The specific mathematical details are given in the appendix.)

In Section 4 we show that given the covariance structure derived from the properties of the ^{14}C dating method, and if we are willing to assume a simple random walk model (explained briefly in subsection 3.1) for the atmospheric ^{14}C concentration in time (the adequacy of this assumption is briefly discussed in Section 6), then the resulting calibration curve has the familiar (accepted and commonly used) feature of being piece-wise linear. However, and in recognition of the uncertainty present in the data, the proposed curve it *smooths* out the data points rather than interpolating them. Furthermore, by employing the proposed model, we obtain what we consider to be more realistic variance values than either those obtained by Christen’s methodology (Christen 1994) or those employed in CALIB (Stuiver and Reimer 1993).

In Section 5 we turn our attention to the calibration of a new ^{14}C determination. We stress the advantage of obtaining not only the posterior moments for its associated calendar year, but the whole posterior distribution. Through an example we illustrate the methodology, showing that by employing the model based on a Gaussian Process Prior we obtain a more satisfactory result, with more interpretability by eliminating the “wiggles” obtained by Christen’s method and attaining a smoother posterior distribution for the calendar age associated to the ^{14}C determination calibrated.

Section 6 contains a summary of our results and a discussion of the implications of the model.

Finally, we present a series of appendices with the mathematical details underlying various results in the paper.

2 NOTATION

Throughout this paper all ages referred to are expressed, as is conventional in the ^{14}C community, in *years before the present (BP)*, meaning—in practical terms—the number of years before 1950 AD. This convention prevents dependence of the ages upon the time of the actual dating. Other details of the notation follow.

Any quantity denoted by a θ will refer to a *calendar* age, whereas quantities denoted by y ’s will refer to *radiocarbon* ages. In order to stress whether any particular calendar age actually corresponds to a datum, a superscript “d” shall indicate, as will be explained in the next subsection, an element of the high-precision data set. When no superscript is given, it will therefore mean that that particular point in the calendar scale does not belong to the data set.

The ^{14}C calibration curve $r(\theta)$, represents the *true* ^{14}C age of the sample, as a function of its true calendar age θ .

2.1 The Data

We have access to a set of data providing information about the relationship between calendar and ^{14}C years. These data are obtained by means of ^{14}C -dating a series of tree-ring samples whose true calendar age is known exactly. This set is known as the “high-precision calibration data”. Considerable effort has been made by the ^{14}C laboratories in order to make their ^{14}C -determined ages as precise as possible, reporting standard laboratory errors even as small as 10 years, much smaller than ordinary ^{14}C determinations. (Note that we shall restrict our attention to the calibration data whose

absolute [i.e. calendar] age is determined by dendrochronology, although the most recent [extended] ^{14}C calibration data set INTCAL98 includes coral results and marine varve chronology as well.)

Denote data sets as $D = \{\underline{\theta}^{(d)}, (\underline{y}, \underline{\sigma})\}$. It has the form

$$D = \{\underline{\theta}^{(d)}, (\underline{y} \pm \underline{\sigma})\} = \left\{ \begin{array}{l} \theta_1^d, (y_1 \pm \sigma_1), \\ \theta_2^d, (y_2 \pm \sigma_2), \\ \mathbf{M} \\ \theta_n^d, (y_n \pm \sigma_n), \end{array} \right\}, \quad (1)$$

where $\theta_k^{(d)}$, $k = 1, \dots, n$ —obtained through dendrochronology, and assumed to be exact for the whole data set—represents the calendar age of the k -th sample in the high-precision calibration data set.

The values for $\theta_k^{(d)}$, $k = 1, \dots, n$, as mentioned above, are obtained through dendrochronology (hence the superscript (d) adopted throughout this paper), and assumed to be exact for the whole dataset.

The ^{14}C age of each of the samples used to build the dataset is reported as $y_k \pm \sigma_k$, $k = 1, \dots, n$ where y_k , $k = 1$ represents the age of the k -th tree ring as determined by the laboratory through the ^{14}C dating method, expressed again in years before the present⁴. Note that—strictly speaking—the calibration data correspond to dating of samples of a number (usually 10) of consecutive tree rings; the calendar age is then established as the time-midpoint of the sample. Therefore, each y_k represents the average ^{14}C age of the k -th group of tree rings. However, the ^{14}C content of such wood samples is not necessarily a perfect representation of the atmospheric ^{14}C level of the period they span. Factors such as variations in tree-ring thickness and the non-uniform formation of the wood throughout each year cause variable annual ^{14}C contributions to the sample average (Stuiver et al. 1998). The proposal of Knox and McFadgen (1997) is an attempt of dealing with this issue. Finding an acceptable way of incorporating this further complexity into the model for the ^{14}C calibration curve is a topic for future research. For the time being, we adhere to the common simplification that each ^{14}C determination corresponds to a single calendar age, for the sake of comparability with other approaches for the modeling of the ^{14}C calibration curve. Nevertheless, we are aware that this issue should be addressed in order to work towards a more comprehensive and realistic model for the calibration curve.

The values σ_k , $k = 1, \dots, n$, denote the experimental error associated to the determination reported by the laboratory. This quoted error typically represents \pm one standard deviation as determined by the total number of accumulated counts for that sample, and is not necessarily related to the spread of ages in mixed samples, to contamination or to other sample conditions (see Gillespie 1984). The use of so called “error multipliers” is a usual way of trying to overcome the fact that this reported error fails to reflect all the variability to which the ^{14}C measurements (i.e., the ^{14}C determinations) are really subject, with respect to the true (but unknown) ^{14}C ages of the samples. (Details of the specific use of “error multipliers” in the INTCAL98 calibration data can be found in Stuiver et al. 1998. See also Bowman 1990:41 for further details.) We will not deal with this specific problem; the proposed approach assumes reported errors which accurately reflect the variability of each ^{14}C determination. Ultimately, “only the laboratory itself can give the necessary information on the reliability of its error estimates and should be requested to do so,” (Bowman and Leese 1995:103).

⁴It is conventional to write “BP” for denoting radiocarbon years before the present, and “cal BP” to denote calendar years after calibration (before the present).

3 THE MODEL

As mentioned in the introduction, by *model* we mean a statistical model. In this sense, the problem under study can be explained in the following terms. We have access to a (finite) set of observations corresponding to a continuous function that relates calendar and ¹⁴C ages. This function is the calibration curve, and the observations are the high-precision calibration data. An additional aspect to consider is that the observations are subject to error—they are not the exact value of the function (i.e., the ¹⁴C age) for the given knots on the calendar scale.

With this information, the aim is to make inference about *the whole calibration curve*. That is, we want to make inference about the function given a set of (noisy) observations.

The basic model for the analysis of the relationship between calendar and ¹⁴C ages, both in the Classical and Bayesian frameworks, is based on the assumption that each ¹⁴C determination is normally distributed, having the “true” (but unknown) ¹⁴C age associated to the corresponding calendar year as mean and the reported laboratory error as standard deviation, and independent from the other determinations. This is, given the ¹⁴C determination $y \pm \sigma$, then

$$y|\theta, \sigma : N(r(\theta), \sigma^2), \tag{2}$$

where θ is the true calendar age of the object dated (or associated calendar year), and $r(\theta)$ is its “true” ¹⁴C age (as a function of θ , the true calendar age). This basic assumption implies—in particular—that each of the calibration observations in the high-precision calibration data can be translated into a similar model, so that

$$y_k|\theta_k^{(d)}, \sigma_k : N(r(\theta_k^{(d)}), \sigma_k^2), k=1, 2K \text{ } n; \text{ independent.} \tag{3}$$

3.1 Prior Modeling

We now require a link between the ¹⁴C determinations and the calendar ages. This link will be given precisely by the calibration curve, relating each calendar age θ to its corresponding (unknown) true ¹⁴C age $r(\theta)$. It is this function that we need to establish a model for.

Our model is statistical, representing expectations and uncertainties about the true calibration curve $r(\theta)$ that might reasonably apply *before* making use of the high-precision calibration data. Expectations about the curve are described by

$$E(r(\theta)|\beta) = \beta\theta. \tag{4}$$

Thus, it would be reasonable to estimate $r(\theta)$ in the absence of any of the high-precision data, by some multiple of θ . A natural assumption would be to set $\beta = 1$, but radiocarbon determinations are based on a conventional half-life value for ¹⁴C that is known to be in error by approximately 3%, and potentially greater systematic deviations may be caused by a drift over time in atmospheric ¹⁴C concentration. We therefore allow β to be uncertain *a priori*. β is technically known as a hyperparameter⁵ of the model, and we will in effect use the data to learn about its value. This is why the expectation is written as conditional upon β . We suppose, *a priori*, that the prior expectation of β is 1, but with a variance of 1/36, expressing reasonable prior uncertainty about the combined effect of the

⁵*Hyperparameters* are those parameters (i.e. unknown quantities) that index the family of prior distributions of the parameters about which we directly wish to make inference (O’Hagan [1994]).

wrong value for the half-life for ^{14}C used for the determination of ^{14}C dates, and the possible violation of the assumption that the atmospheric ^{14}C concentration has remained constant through time. The vast amount of calibration data means that the prior settings will have little weight, compared to the weight given *a posteriori* to the information provided by the data.

Of course, we know that the calibration curve is not linear, and that on the contrary it wiggles, sometimes departing from the overall trend by many years. The possible departure of $r(\theta)$ from its prior expectation of $\beta\theta$ is what lends uncertainty to its actual value, and this uncertainty is described by a variance. Naturally, uncertainty increases with years cal BP, and we represent this by

$$\text{Var}(r(\theta)|\tau^2) = \tau^2\theta. \quad (5)$$

Formally, this linear increase of variance with time corresponds to an assumption of an underlying random walk model, which is developed in the Appendix by drawing on established science concerning the variation of atmospheric ^{14}C concentrations. The constant τ^2 in this formula is another hyperparameter that fixes the overall degree of variation of the curve about its mean. We shall use the high-precision data to learn about its value, and give it an uninformative prior distribution, reflecting our prior ignorance about it but give it a prior estimate⁶ of 25, with considerable initial uncertainty expressed through $\text{Var}(\tau^2) > 1,000,000$. The same underlying random walk model that leads to the variance formula also gives an expression for the prior correlation between the value of $r(\theta)$ at two points on the curve $\theta_1 < \theta_2$, via

$$\text{Corr}(r(\theta_1), r(\theta_2)) = \sqrt{\theta_1/\theta_2}. \quad (6)$$

We can see from this that at points close together in time the correlation between the corresponding values of the calibration curve is high, whereas the correlation is low between distant points. This is again a natural expression of initial knowledge about the calibration curve prior to seeing any of the high-precision data. Simply speaking, it says just that we know that $r(\theta)$ should be a smoothly varying (though possibly widely varying if τ^2 is large) function over time.

The statistical model is completed by assuming that uncertainty at all points on the curve can be described by the normal, or Gaussian, distribution. The assumption is made in the same spirit as normal distributions are commonly assumed in practical statistics. That is, the assumption is convenient mathematically and will not be implausible in practice. Our analysis will be driven primarily by the high-precision calibration data, secondarily by the forms of expectation, variance and correlation that we have presented, and only slightly by the assumption of normality. Technically, assuming normality at all points of the curve in this way leads to a model known as a Gaussian Process, which has been employed in a wide variety of other contexts to model unknown functions (see, e.g., O'Hagan 1992; O'Hagan et al. 1999; Neal 1999; Schmidt and O'Hagan 2000). In all such applications, the key task is to formulate expectations, variances and correlation. Our model in this case is intended to do so in a way that accurately reflects prior knowledge about the ^{14}C calibration curve.

4 RESULTING RADIOCARBON CALIBRATION CURVE

In this section we present the posterior mean and variance for any point $r(\theta)$ on the calibration curve. We show that there are definite advantages in the methodology we propose, in comparison with both

⁶These values are chosen arbitrarily; as mentioned above, the vast amount of calibration data means that the prior setting will have little influence on the posterior results.

Christen’s approach (Christen 1994) and with the methodology behind the calibration program CALIB (Stuiver and Reimer 1993).

In the preceding section we have presented the key features of the prior information, that is, information available about the calibration curve $r(\theta)$ prior to making use of the high-precision calibration data. The Bayesian statistical method now combines this prior model with the data to yield a *posterior* distribution that synthesizes those two sources of information. Since the whole calibration curve is an unknown function, this posterior distribution also expresses statistical information jointly about all points on the curve.

Mathematical details of the derivation of the posterior distribution will be presented in a forthcoming technical paper. However, the most important results are again expectations and variances. In the next subsection, we describe the posterior mean function which comprises our new estimate of the ^{14}C calibration curve. This new calibration curve is piecewise linear but smooths the high-precision data points in a new way.

A crucial benefit of a statistical approach to the problem is that we have not just an estimated calibration curve but also a posterior variance to quantify uncertainty about every point $r(\theta)$ on the curve. The nature of our posterior variance function is discussed in subsection 4.2 and compared with the alternative methodologies of Christen (1994), that is incorporated in the BCal on-line calibration software, and of Stuiver and Reimer (1993), that underlies the CALIB software.

4.1 New Piece-Wise Linear Curve

To describe the new curve, it is convenient to think first about the estimation of the calibration $r_i = r(\theta_i^{(d)})$ at the high-precision calibration data points $\theta_i^{(d)}$, $i = 1, 2, \dots, n$, and then to consider the estimate of $r(\theta)$ between two adjacent data points.

Although the data points are described as “high-precision”, there is still appreciable error in calibration ages y_i , as described by σ_i . The standard approach of joining the points $(\theta_i^{(d)}, y_i)$ will, because of random error, tend to produce a curve that “wiggles” more, with sharper turns and more deviation from a straight line than the underlying true curve. Our Bayesian approach naturally smooths out the curve somewhat by posterior estimated values

$$\hat{r}_i = E(r(\theta_i^{(d)}) | \underline{y}) \tag{7}$$

that differ from the data values y_i . The differences recognize the stated errors σ_i on those data values, and of course this limits the degree of smoothing in a way which naturally respects the quality of the data.

Between data points, it is a property of the random walk model for the prior variation and correlations that the estimated curve interpolates the \hat{r}_i values linearly. That is, the smoothed data points \hat{r}_i are connected by straight line segments.

Figure 1 shows a section of the calibration curve. Our curve follows the major features of the data over this region but has smoothed out the sharp localized peaks and troughs that it is reasonable to suppose may be partly the result of random error in the high-precision data.

We suggest that this smoothed piece-wise linear calibration curve is more realistic than the one that simply interpolates the calibration data, employed by CALIB, and resulting from Christen’s approach as well. By making use of the information drawn from all the calibration data (including

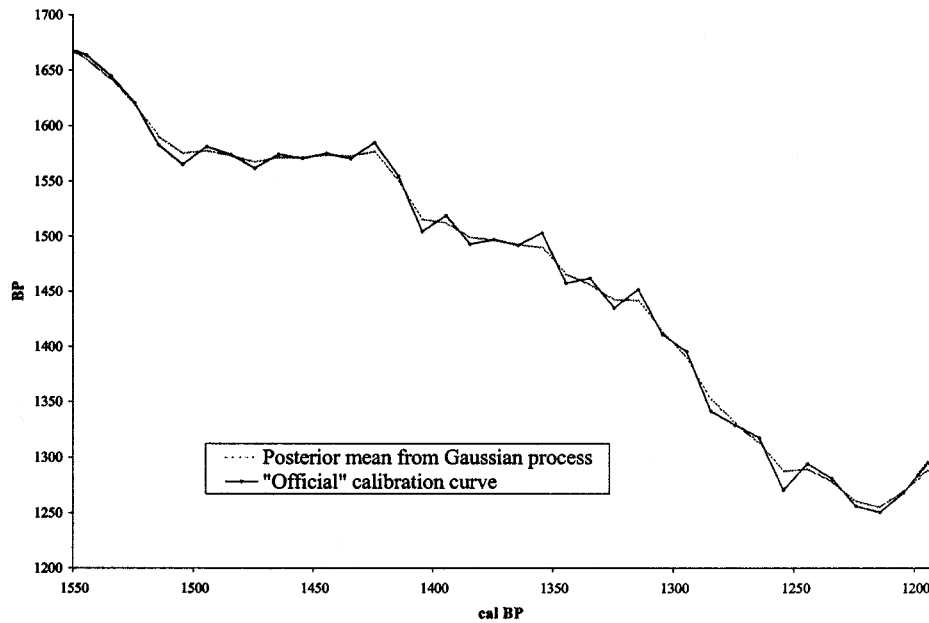


Figure 1 Comparison of posterior ^{14}C calibration curves (fragment). The proposed approach yields a posterior mean (dotted line) which, in contrast to the one used by CALIB and OxCal and also obtained by Christen (1994) (solid line with marks representing the data points), smooths out the calibration data points. This results from the incorporation of the uncertainty in the calibration data into the model. (High-precision calibration data from Stuiver et al. [1998])

the uncertainty associated to each point) through the Gaussian Process model, it is possible to infer a more reasonable location of the true ^{14}C ages as a function of the calendar ages.

It is arguable that a completely smooth curve might be even more realistic than a piece-wise linear one, and indeed we are exploring such a curve via an elaboration of our approach. However, we believe that it is important for the radiocarbon community to see and evaluate separately the consequence of smoothing out the wiggles in the data at the calibration points while retaining the familiar piece-wise linear form between those points.

4.2 Posterior Variance for the Radiocarbon Calibration Curve

To describe the posterior uncertainty around the new calibration curve, it is again convenient to think first about uncertainty at the data points $\theta_i^{(d)}$ and then at how the uncertainty varies between those points.

Remember that the Bayesian approach is characterized by synthesizing two sources of information—the data and the prior information. This typically results in less uncertainty than would result from using the data alone. We find in this analysis that the posterior variances at the high-precision data points is less than the reported error variances of the data alone. That is,

$$\text{Var}(r_i | \underline{y}) < \sigma_i^2. \quad (8)$$

We believe that this is a distinct advantage of our method. In contrast, both BCal and CALIB assume variances at the data points equal to the reported variances σ_i^2 of the radiocarbon determinations. This is because the methods of Christen (1994) and Stuiver and Reimer (1993) estimate r_i simply by y_i . Our method produces lower variances because our model allows the estimation of r_i to be based not just on y_i but also on neighboring points, and so uses more information.

Now consider the uncertainty about the curve between data points, and suppose that $\theta_i^{(d)} \leq \theta_{i+1}^{(d)}$. The formula for the posterior variance of $r(\theta)$ is complex but can be readily understood in terms of three components of uncertainty in a piece-wise linear estimating function. We will express these in general terms that allow us to compare the variance produced by our method with those produced within BCal and CALIB.

1. First, suppose that we believe that the true calibration curve must be a straight line segment between $r(\theta_i) = r_i$ and $r(\theta_{i+1}) = r_{i+1}$. We will be uncertain about r_i and r_{i+1} , and represent this in general terms by variances V_i and V_{i+1} .
2. Even if we believe that the true curve is definitely piece-wise linear, we should also account for correlation between r_i and r_{i+1} , that we represent by a covariance $C_{i,i+1}$. This affects the uncertainty at an intermediate point θ because it measures the extent to which we would expect r_i and r_{i+1} to be either both above their estimated values or both below, rather than one above and one below.
3. Of course, nobody believes that the true radiocarbon calibration curve happens to be piece-wise linear with the linear pieces joined at exactly the dates we have high-precision data for! So there should be additional uncertainty about how far the curve will deviate from linearity between adjacent data points. We represent this third component of uncertainty by T .

Using these three uncertainty components, it is possible to express the variance of $r(\theta)$ under all three methods in terms of the general variance formula

$$Var[r(\theta)|y] = \alpha^2 V_{i+1} + (1-\alpha)^2 V_i + \alpha(1-\alpha)[S_i T + 2C_{i,i+1}] \tag{9}$$

where we additionally define $S_i = \theta_{i+1}^{(d)} - \theta_i^{(d)}$ to be the spacing between the two data points (generally 10 in the high-precision data) and $\alpha = (\theta - \theta_i^{(d)}) / S_i$ is the relative position of θ between those points.

In our method, $V_i = Var(r_i|y)$ is the posterior variance of r_i , $C_{i,i+1} = Cov(r_i, r_{i+1}|y)$ is the posterior covariance between r_i and r_{i+1} , and $T = E(\tau^2|y)$ is the posterior expectation of the hyperparameter τ^2 .

Christen’s method, incorporated in BCal, has $V_i = \sigma_i^2$, and as already mentioned this means that our method produces lower variances at the data points. Because Christen estimates each r_i on the basis only of its associated y_i , and because the y_i s are independent, his variance formula correctly corresponds to (9) with $C_{i,i+1} = 0$. Finally, Christen sets $T = \lambda^2$, where λ is a parameter that accounts for short-term variability in the atmospheric ^{14}C concentration. Christen directly estimates $\lambda \approx 19$. Since in our analysis $T = E(\tau^2|y) \approx 52$ and $C_{i,i+1}$ is never larger than 250, it is readily seen that Christen’s method produces very much larger variances than ours.

Stuiver and Reimer (1993) follow the approach presented earlier in Stuiver and Pearson (1986) of interpolating standard deviations linearly. This leads to a version of (9) in which $V_i = \sigma_i^2$, $C_{i,i+1} = \sigma_i \sigma_{i+1}$ and $T = 0$. The value for $C_{i,i+1}$ corresponds to assuming perfect correlation between r_i and r_{i+1} , which conflicts markedly with Christen’s apparently indisputable analysis that these should

be independent if r_i is estimated solely from y_i . Furthermore, assuming $T = 0$ implies the equally indefensible belief that the true curve is piece-wise linear. We therefore believe that the variance formula adopted by Stuiver and Reimer, and used in CALIB, has no statistical justification.

Summary of the Variance Calibration

Figure 2 clearly shows the much higher variances given by Christen's method. We believe that these are seriously over-pessimistic. Stuiver and Reimer, on the other hand, employ a formula that is unsupported statistically and fails to recognize interpolation uncertainty properly. We argue that our approach gives more realistic variances than either of these other methods.

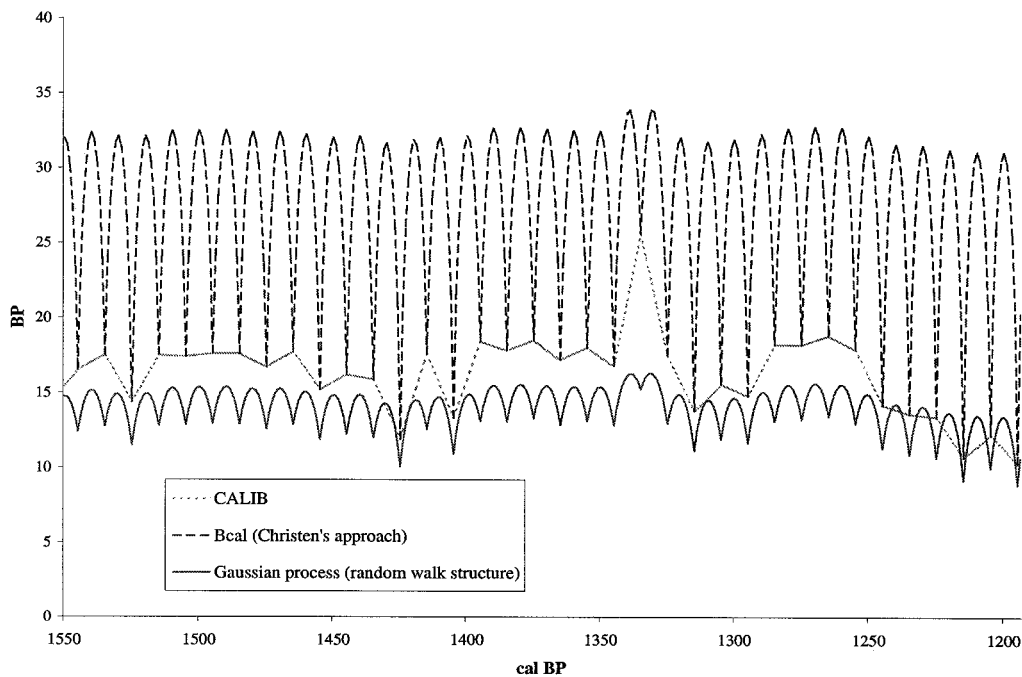


Figure 2 Comparison of posterior standard deviations for the ^{14}C calibration curve (fragment). Notice that the standard deviation resulting from our model (continuous line) is always more moderate than that from Christen's model (dashed line). On the other hand, the comparison with the values employed by CALIB and OxCal (dotted line) depends on the proximity of the point on the calendar scale to any calendar age included in the high-precision calibration data. (High-precision calibration data from Stuiver et al. 1998.)

5 CALIBRATION OF A NEW RADIOCARBON DETERMINATION

The previously described modeling of the calibration curve itself enables us to perform the calibration of a new radiocarbon determination. (See Naylor and Smith 1988; Buck et al. 1996.)

Consider the case when we learn the value of a new ^{14}C determination $y_0 \pm \sigma_0$. The interest lies in making inference about the calendar age θ_0 associated to the object that is being ^{14}C -dated. It is well recognized (Bowman 1990) that due to the nature of the calibration curve, the posterior distribution for the calendar year associated to any particular ^{14}C age $y_0 \pm \sigma_0$ is not normal or any other standard probability distribution, and indeed its form depends on the corresponding section of the calibration curve (as well as on the reported precision of the new ^{14}C determination). Therefore, simple summa-

ries such as the mean and variance of the posterior distribution for θ_0 , the associated calendar age, would fail to convey all the relevant information about its true value, in light of the high-precision calibration data.

Since we are working within the Bayesian framework, we need to involve all the prior information that the archaeologist has about the true calendar year in which the organism from which the sample under analysis is taken, may have died. We believe that in this way the Bayesian approach provides archaeologists the opportunity to take advantage of all the knowledge they may have about specific events of interest—in this case, the calendar age of a specific sample. The process of elicitation of the archaeologists' information and knowledge concerning the parameter of interest into an appropriate statistical prior distribution constitutes itself a whole subject for study and discussion. (See Buck et al. [1996:359–60] for general comments on prior elicitation.) For the purposes of these examples, however, we use flat priors for θ_0 , representing very weak prior information, to facilitate comparison between our method and others.

The distribution for $(y_0 | \theta_0, \sigma_0^2, D)$ is required to perform the calibration of the new ^{14}C determination, and obtaining it involves the use of numerical techniques. A more technical paper (in preparation) will discuss this issue in greater depth.

5.1 Examples

As a final illustration of the proposed methodology, we present the calibration of two ^{14}C determinations obtained from the same archaeological site. As in Section 4, we will compare our results to those arising from Christen's approach, and to the results obtained when using the model employed by the calibration program CALIB.

The ^{14}C determinations correspond to samples of bone retrieved from the archaeological site known as the Buttermarket Cemetery in Eastern England (Scull and Bayliss 1999).

Example 1. The first ^{14}C determination corresponds to the sample identified as *UB 4074*. For this sample, the ^{14}C -determined age is given as 1419 ± 23 .

Figure 3 shows the posterior distribution for the calendar age associated to ^{14}C determination 1419 ± 23 , together with that obtained when using Christen's approach and the model from CALIB.

Although not yielding a radically different result, we notice that with the new approach, the rather undesirable “wiggles” in the posterior density for the associated calendar year which are particularly pronounced for Christen's approach are smoothed out. This can be explained by looking back at Figure 1, the corresponding section of the (posterior) calibration curve. By making use of the whole set of data points through the Gaussian Process model to make inference about the ^{14}C calibration curve, the curve's posterior variance is smaller than when using the data pairwise. This yields a smoother posterior distribution for the calendar year associated to a particular ^{14}C determination to be calibrated.

The “wiggles” in the posterior distribution obtained through Christen's method (less markedly resulting from performing the calibration with the methodology from CALIB) may contain an important part of the posterior probability. To illustrate this, we present in Table 1 the highest posterior density intervals resulting from the three approaches.

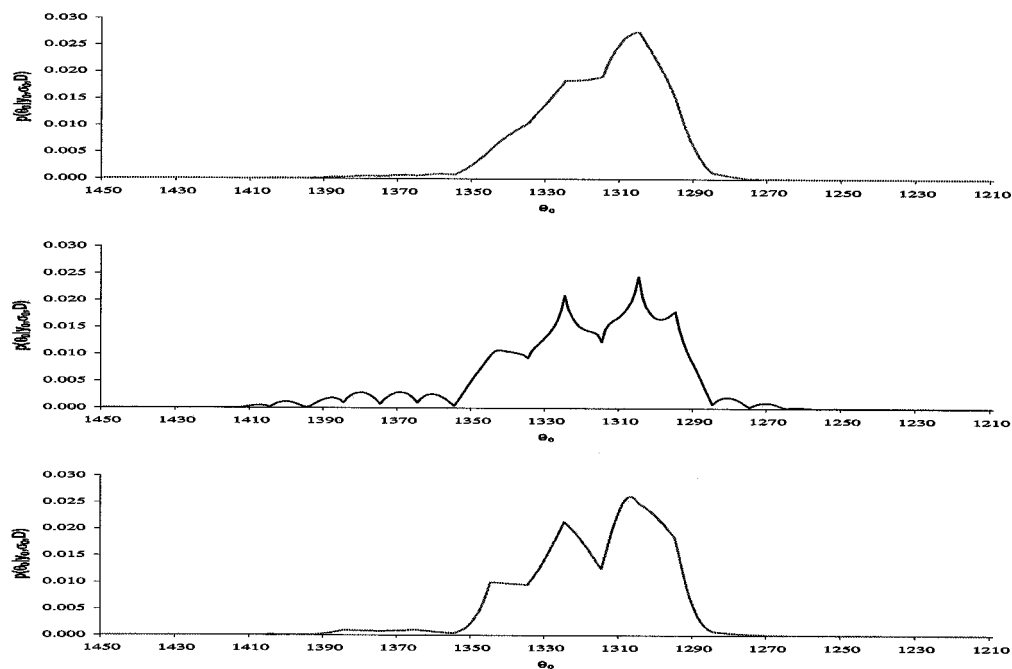


Figure 3 Posterior distribution for the calendar year θ_0 associated to the ^{14}C determination UB 4074, 1419 ± 23 (example 1). The top plot corresponds to the proposed approach; the middle one results from Christen's methodology, and the bottom one corresponds to the results from CALIB.

Table 1 Highest-posterior-density intervals for the calendar year associated to the determination *UB 4074*

Posterior probability	New approach		
	99%	95%	90%
HPD interval (cal BP)	(1280–1372)	(1289–1347)	(1292–1342)
Length	93 yr	59 yr	51 yr
Posterior probability	Christen's approach		
	99%	95%	90%
HPD interval (cal BP)	$(1268-1273) \cup (1277-1354) \cup (1356-1393) \cup (1398-1403)$	$(1279-1282) \cup (1286-1353) \cup (1358-1363) \cup (1367-1373) \cup (1377-1383) \cup (1388-1389)$	$(1287-1352) \cup (1368-1371) \cup (1380-1381)$
Length	128 yr	94 yr	72 yr
Posterior probability	CALIB		
	99%	95%	90%
HPD interval (cal BP)	$(1283-1354) \cup (1360-1387)$	(1289–1348)	(1293–1345)
Length	100 yr	60 yr	53 yr

The “wiggles” in the posterior distribution, thus, make inference about the calendar age associated to the new ^{14}C determination difficult, due to the resulting disjoint intervals containing the desired posterior probability.

Example 2. The second example, taken from the same archaeological site, allows for similar conclusions. In fact, in this case the resulting HPD’s for both Christen’s approach and that from CALIB are made up of even a greater number of disjoint subintervals.

The sample in this case is identified as *UB 4077*, and the ^{14}C determination to be calibrated is given by 1476 ± 24 .

Notice that although both ^{14}C determinations had a very similar reported standard error, in this second case the associated calendar age has a larger posterior variance. Indeed, the posterior distribution for the calendar age associated to ^{14}C determination 1476 ± 24 is more spread out, as is shown in Figure 4. This is due to the section of the calibration curve corresponding to each of these ^{14}C determinations; the section which includes the first determination is more steep than the section corresponding to the second determination shown here. Thus, the calibration onto the calendar scale is more spread out for this second example.

Again, we present the resulting highest posterior density intervals resulting from the three approaches (Table 2).

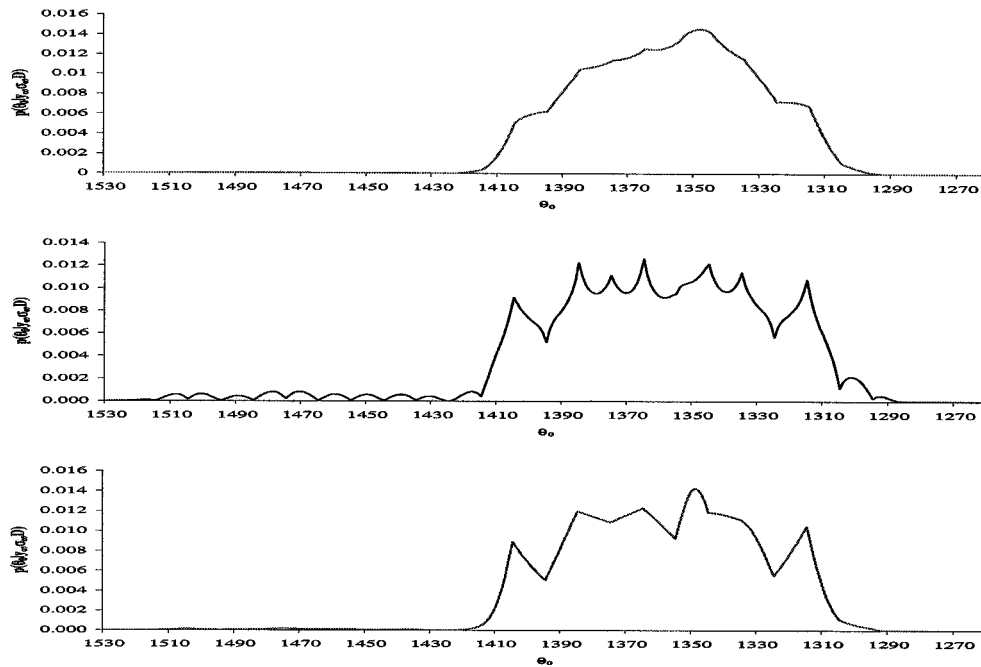


Figure 4 Posterior distribution for the calendar year θ_0 associated to the ^{14}C determination *UB 4077*, 1476 ± 24 (example 2). The top plot corresponds to the proposed approach; the middle one results from Christen’s methodology, and the bottom one corresponds to the results from CALIB.

Table 2 HPD intervals for the calendar year associated to the determination *UB 4077*

Posterior probability	New approach		
	99%	95%	90%
HPD interval (cal BP)	(1304–1412)	(1313–1405)	(1315–1398)
Length	109 yr	93 yr	84 yr
Posterior probability	Christen's approach		
	99%	95%	90%
HPD interval (cal BP)	(1291–1294) \cup (1296–1421) \cup (1430–1432) \cup (1437–1442) \cup (1448–1452) \cup (1457–1462) \cup (1467–1474) \cup (1476–1483) \cup (1488–1491) \cup (1498–1503) \cup (1507–1512)	(1299–1304) \cup (1306–1413)	(1312–1394) \cup (1396–1408)
Length	182 yr	114 yr	98 yr
Posterior probability	CALIB		
	99%	95%	90%
HPD interval (cal BP)	(1301–1414)	(1311–1394) \cup (1396–1407)	(1312–1323) \cup (1327–1393) \cup (1399–1406)
Length	114 yr	96 yr	87 yr

6 DISCUSSION

We propose the use of a Gaussian Process model for the ^{14}C calibration curve that takes into account all the information provided by the data while at the same time acknowledging the uncertainty present in the calibration data. This produces posterior results that seem more realistic than those resulting from the approaches considered here (corresponding to Christen's methodology and to that used by the computer calibration program CALIB). Firstly, because the resulting posterior mean calibration curve smooths the data points rather than interpolating them; and secondly because the posterior variances for each point on the ^{14}C calibration curve seem more realistic by incorporating information from all the data points. The considerable reduction in variance values for points on the calibration curve (compared to those from Christen's method) is the result of the comprehensive use of the data implied by the proposed model. The way in which the variance values grow as we move away from the calendar ages included in the calibration data set is due to the specific correlation structure presented in Section 3, which determines a piece-wise linear mean for the calibration curve; this is a characteristic missing from the variance estimates from CALIB.

As a consequence of this, the related problem of calibrating a new ^{14}C determination has also a more satisfactory solution, as the resulting posterior distributions for the calendar (i.e., calibrated) age associated to the new ^{14}C determination appear easier to interpret without the "wiggles" that commonly arise from Christen's approach and, to a lesser extent, from the model used by CALIB, as in the examples presented here.

It is clear that the covariance function plays a fundamental role in the model, and it determines features of the resulting posterior calibration curve such as smoothness and differentiability. The particular choice of covariance structure presented here is derived from a Random Walk model for the auxiliary process given by the atmospheric ^{14}C concentration in time. We are fully aware that in reality such process has a much more complex structure. Nevertheless, our choice, on the one hand, constitutes the simplest possible model accounting for the non-stationarity of the ^{14}C concentration in time. On the other hand, it determines the piece-wise linear nature of the resulting posterior calibra-

tion curve, which makes it straightforward to perform comparisons between the proposed approach and those corresponding to the currently most commonly used calibration programs. Thus, we have illustrated—through a simple structure—the advantages of adopting the proposed modeling approach for the ^{14}C calibration curve. Ongoing research will examine the sensitivity and consequences of alternative choices for the covariance structure.

ACKNOWLEDGMENTS

We are particularly grateful for comments and useful suggestions by J A Christen and G Nicholls, which were instrumental in helping us to locate and correct a significant error in an earlier draft of this paper. We are also grateful to Andrew Millard for many useful comments and to Caitlin Buck for patient discussions towards the putting together of this paper. We also acknowledge the useful and detailed comments from the referees.

The first author is sponsored by Consejo Nacional de Ciencia y Tecnología (CONACYT), México (Ref. 115,703).

REFERENCES

- Aitken MJ. 1990. *Science-based dating in archaeology*. London: Longman Archaeology Series.
- Bowman S. 1990. *Radiocarbon dating (Interpreting the past)*. London: British Museum Publications Ltd.
- Bowman SGE, Leese MN. 1995. Radiocarbon calibration: current issues. *American Journal of Archaeology* 99(1):102–5.
- Bronk Ramsey C. 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2):425–30.
- Buck CE, Cavanagh WG, Litton CD. 1996. *Bayesian approach to interpreting archaeological data*. Chichester: John Wiley & Sons.
- Buck CE, Litton CD, Scott EM. 1994. Making the most of ^{14}C dating: some statistical considerations. *Antiquity* 68:252–63.
- Christen JA. 1994. Bayesian interpretation of radiocarbon results. PhD thesis. University of Nottingham.
- Christen JA, Litton CD. 1995. A Bayesian approach to wiggle-matching. *Journal of Archaeological Science* 22:719–25.
- Gillespie R. 1984. *Radiocarbon user's handbook*. Oxford University. Committee for Archaeology. Monograph Number 3.
- Grimmett GR, Stirzaker DR. 1992. *Probability and random processes*. Second edition. Oxford Science Publications. Clarendon Press, Oxford.
- Knox FB, McFadgen BG. 1997. Least-squares fitting a smooth curve to radiocarbon calibration data. *Radiocarbon* 39(2):193–204.
- McCormac FG, Baillie MGL. 1993. Radiocarbon to calendar date conversion: calendrical band widths as a function of ^{14}C precision. *Radiocarbon* 35(2):311–6.
- Naylor JC, Smith AMF. 1988. An archaeological inference problem. *Journal of the American Statistical Association* 83(403): 588–95.
- Neal RM. 1999. Regression and classification using Gaussian process priors. In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian statistics 6*. Oxford University Press. p 69–95.
- Nicholls G, Christen JA. 2000. Random-walk ^{14}C calibration. Technical report number 457. Mathematics Department, University of Auckland.
- O'Hagan A. 1992. Some Bayesian numerical analysis (with discussion). In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian statistics 4*. Oxford University Press. p 345–363.
- O'Hagan A. 1994. *Kendall's advanced theory of statistics*. Volume 2B, Bayesian inference. London: Edward Arnold.
- O'Hagan A, Kennedy MC, Oakley JE. 1999. Uncertainty analysis and other inference tools for complex computer codes (with discussion). In: Bernardo JM, Berger JO, Dawid AP, Smith AFM, editors. *Bayesian statistics 6*. Oxford University Press. p 503–24.
- Pazdur MF, Michczynska DJ. 1989. Improvement of the procedure for probabilistic calibration of ^{14}C results. *Radiocarbon* 31(3):824–32.
- Pearson GW, Stuiver M. 1986. High-precision calibration of the time scale, 500–2500 BC. *Radiocarbon* 28(2B):839–62.
- Schmidt AM, O'Hagan A. 2000. Bayesian inference for nonstationary spatial covariance structure via spatial deformations. Research report no. 498/00. Department of Probability and Statistics, University of Sheffield.
- Scull C, Bayliss A. 1999. Radiocarbon dating and Anglo-Saxon graves. In: von Freeden, Koch and Wiczorek, editors. *Volker an Nord- und Ostsee und die Franken*.
- Stuiver M, Pearson GW. 1986. High-precision calibration of the ^{14}C time scale, AD 1950–500 BC. *Radiocarbon* 28(2B):805–38.
- Stuiver M, Reimer P. 1989. Histograms obtained from computerized ^{14}C age calibration. *Radiocarbon* 31(3):

- 817–23.
- Stuiver M, Reimer PJ. 1993. Extended ^{14}C data base and revised CALIB 3.0 ^{14}C age calibration program. *Radiocarbon* 35(1):215–30.
- Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac G, van der Plicht J, Spurk M. 1998. INTCAL98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon* 40(3):1041–83.
- Talma AS, Vogel JC. 1993. A simplified approach to calibrating ^{14}C dates. *Radiocarbon* 35(2):317–22.
- Taylor RE. 1987. *Radiocarbon dating, an archaeological perspective*. Orlando: Academic Press.
- van der Plicht J, Mook WG. 1989. Calibration of ^{14}C ages by computer. *Radiocarbon* 31(3):805–16.
- van der Plicht J. 1993. The Groningen radiocarbon calibration program. *Radiocarbon* 35(1):231–7.
- Zeidler JA, Buck CE, Litton CD. 1998. Integration of archaeological phase information and radiocarbon results from the Jama river valley, Ecuador: a Bayesian approach. *Latin American Antiquity* 9(2):160–79.

APPENDIX

The treatment we propose for the variability of the calibration curve (expressed as the covariance function $v(\theta, \theta^*)$) involves the modeling of an auxiliary process. In order to produce the covariance structure required in the model, we notice (Christen 1994) that there exists a strong relationship between the calibration curve and an auxiliary process given by the atmospheric ^{14}C concentration through time. Therefore a model for the atmospheric ^{14}C concentration will provide us with a useful structure to incorporate as covariance structure in the original model for the calibration curve. Our proposal for the covariance structure arises from assuming that the process $\Delta(\theta)$ ⁷, which describes the variations in the atmospheric ^{14}C concentration, behaves as a Random Walk. That is, we assume that the variation from one period to the next (e.g. from year to year) in the level of $\Delta(\theta)$ is purely random (or “white noise”). We are aware that many elements affect the behavior of this $\Delta^{14}\text{C}$ process, such as geomagnetism and the 11-year sunspot cycle. However, our position here is to keep the model as simple as possible in order to facilitate the analysis. On the other hand, the idea of considering the variations in the atmospheric ^{14}C concentration as a purely random element might not be unreasonable to such extent as to dismiss this initial structure. Furthermore, the consequences of this choice of covariance structure (namely, the piece-wise linearity of the resulting calibration curve) can contribute to its appeal.

From the law of exponential decay, we know that

$$r(\theta) = \theta - \xi \log \left(\frac{M(\theta)}{M_0} \right), \quad (1)$$

where ξ is the mean-life value for ^{14}C ,⁸ $M(\theta)$ represents the atmospheric $^{14}\text{C}/^{12}\text{C}$ ratio at time θ cal BP, and M_0 is the value of the modern standard for this ratio. A common way of expressing $M(\theta)$ is through the auxiliary process $\Delta^{14}\text{C}$, which we denote $\Delta(\theta)$ for (mathematical) formality, defined as

$$\Delta(\theta) = \left(\frac{M(\theta)}{M_0} - 1 \right) \times 10^3. \quad (2)$$

Therefore,

$$r(\theta) = \theta - \xi \log [10^{-3} \Delta(\theta) + 1]. \quad (3)$$

⁷This process is known in the ^{14}C community as $\Delta^{14}\text{C}$; here, we are only making explicit the time dependence of this process.

⁸For Libby’s value of $T_{1/2} = 5568$ for the half-life of ^{14}C , the corresponding mean-life is $\xi = 8033$.

Given that the values for $10^{-3}\Delta(\theta)$ “small” (Christen 1994), the Taylor series approximation $\log [10^{-3}\Delta(\theta) + 1] \approx [10^{-3}\Delta(\theta)]$ can be used. So,

$$r(\theta) \approx \theta - \xi [10^{-3}\Delta(\theta)]. \quad (4)$$

This auxiliary process $\Delta(\theta)$ represents the variation of the atmospheric ^{14}C concentration at moment θ BP with respect to the modern level (M_0). It is clear from the expression above that it constitutes a strong influence for the relationship between calendar ages (θ) and ^{14}C ages ($r(\theta)$).

The simplest model for incorporating a non-constant behavior assumes that the variations in the level (magnitude) of the variable occurring from year to year can be considered as pure random fluctuations. This assumption defines a Random Walk Process model for $\Delta(\theta)$.

$$\Delta(\theta) = \Delta(\theta-1) + \varepsilon_\theta; \quad (5)$$

where $\{\varepsilon_t\}$ represents the random error process such that each ε_t follows a Normal distribution with mean zero and variance σ_ε^2 . Hence,

$$\{\nabla(\Delta(\theta))\}: N(0, \sigma_\varepsilon^2) \quad (6)$$

where ∇ represents the first difference operator $\nabla(Z_t) = Z_t - Z_{t-1}$.

Now we incorporate this information into the expression for the first difference of $r(\theta)$:

$$\begin{aligned} \nabla(r(\theta)) &= r(\theta) - r(\theta-1) \\ &= [\theta - \xi 10^{-3}\Delta(\theta)] - [(\theta-1) - \xi 10^{-3}\Delta(\theta-1)] \\ &= 1 - \xi 10^{-3}[\Delta(\theta) - \Delta(\theta-1)] \\ &= 1 - \xi 10^{-3}[\nabla(\Delta(\theta))]. \end{aligned} \quad (7)$$

Therefore

$$\{\nabla(r(\theta))\}: N(1, \xi^2 10^{-6} \sigma_\varepsilon^2), \quad (8)$$

and so, assuming the existence of a starting point $r(0)$ in the series, we can write

$$\begin{aligned} r(\theta) &= r(0) + \sum_{i=1}^{\theta} \nabla(r(i)) \\ &= r(0) + \sum_{i=1}^{\theta} a_i, \end{aligned} \quad (9)$$

where now $a_i = \nabla r(i) : N(1, \xi^2 10^{-6} \sigma_\varepsilon^2)$. We know that $r(0) = 0$; therefore

$$\begin{aligned} E(r(\theta)) &= \theta \\ \tau^2 \text{Var}(r(\theta)) &= \theta(\xi^2 10^{-6} \sigma_\varepsilon^2) \\ \tau^2 \text{Cov}(r(\theta), r(\theta-k)) &= (\theta-k) \times (\xi^2 10^{-6} \sigma_\varepsilon^2); k > 0, \end{aligned} \quad (10)$$

where σ_ε^2 is in fact the variance of the white noise process defined as the first difference of the series $\Delta(\theta)$. So, the proposal for the covariance function $v(\theta, \theta^*)$ arises from this time series analysis, and is given by $\tau^2 v(\theta, \theta^*) = \min\{\theta, \theta^*\} \times \xi^2 10^{-6} \sigma_\varepsilon^2$. Notice that σ_ε^2 , the variance of the process $\nabla[\Delta(\theta)]$, is formally unknown. In section 3.1, we have denoted the whole factor $\xi^2 10^{-6} \sigma_\varepsilon^2$ by the unknown parameter τ^2 . Thus,

$$v(\theta, \theta^*) = \min\{\theta, \theta^*\}. \quad (11)$$