Bayesian Analysis (2006) **1**, Number 2, pp. 265–288

Bayesian nonparametric estimation of the radiocarbon calibration curve

Caitlin E. Buck*, Delil Gómez Portugal Aguilar†,
Cliff D. Litton‡, and Anthony O'Hagan§

Abstract. The process of calibrating radiocarbon determinations onto the calendar scale involves, as a first stage, the estimation of the relationship between calendar and radiocarbon ages (the radiocarbon calibration curve) from a set of available high-precision calibration data. Traditionally the radiocarbon calibration curve has been constructed by forming a weighted average of the data, and then taking the curve as the piece-wise linear function joining the resulting calibration data points. Alternative proposals for creating a calibration curve from the averaged data involve a spline or cubic interpolation, or the use of Fourier transformation and other filtering techniques, in order to obtain a smooth calibration curve. Between the various approaches, there is no consensus as to how to make use of the data in order to solve the problems related to the calibration of radiocarbon determinations.

We propose a nonparametric Bayesian solution to the problem of the estimation of the radiocarbon calibration curve, based on a Gaussian process prior structure on the space of possible functions. Our approach is model-based, taking into account specific characteristics of the dating method, and provides a generic solution to the problem of estimating calibration curves for chronology building.

We apply our method to the 1998 international high-precision calibration dataset, and demonstrate that our model predictions are well calibrated and have smaller variances than other methods. These data have deficiencies and complications that will only be unravelled with the publication of new data, expected in early 2005, but this analysis suggests that the nonparametric Bayesian model will allow more precise calibration of radiocarbon ages for archaeological specimens.

Keywords: Gaussian process, archaeology, chronology building, cross-validation

1 Introduction

Radiocarbon (^{14}C) has a number of properties that make it suitable for archaeological dating, including a half-life of the right order for much of prehistory, chemical features that ensure distribution throughout the biosphere, and of course, feasibility of measurement (Aitken, 1990). Besides its widespread use in archaeology, radiocarbon dating is also useful in other disciplines such as geology and climatology. Hence, radiocarbon dating is one of the most commonly used chronometric (or absolute) dating techniques.

*Department of Probability and Statistics. University of Sheffield, Sheffield, UK,
†Department of Probability and Statistics. University of Sheffield, Sheffield, UK,
‡School of Mathematical Sciences, University of Nottingham, Nottingham, UK,
§Department of Probability and Statistics. University of Sheffield, Sheffield, UK,

The production of radiocarbon occurs in the upper atmosphere, due to the interaction of cosmic rays with nitrogen. ^{14}C combines with oxygen to form carbon dioxide which is chemically indistinguishable from the carbon dioxide containing stable (non-radioactive) carbon isotopes (^{12}C and ^{13}C). By means of the photosynthesis process, this 'heavy' carbon dioxide enters plants and hence, via the food chain, all animals. Consequently, as long as plants and animals continue to metabolise, they exchange carbon with the atmosphere (or ocean) in which they live and so the ratio of radioactive to non-radioactive carbon in their cells is in equilibrium with their environment. There is only about one part of ^{14}C in a million million of modern carbon, whereas 99% of total carbon is ^{12}C and 1% is ^{13}C (Bowman, 1990). This balance is maintained by the carbon cycle. After the death of an organism, however, the assimilation of carbon ceases, and the amount of ^{14}C present in the organism — originally in equilibrium with the atmosphere — decreases in an exponential fashion, according to the law of radioactive decay. This states that given an initial (atmospheric) ^{14}C/^{12}C ratio $m(0)$ at time $t = 0$, an organism that ceases exchange with the atmosphere at that time will at a later time t have a remaining ^{14}C/^{12}C ratio given by

$$m(t) = m(0) \exp(-\lambda t), \tag{1}$$

where λ is the decay rate of radiocarbon.

The essence of radiocarbon dating is to use equation (1) to deduce t given $m(t)$ and $m(0)$. Thus, t is the unknown age of some organic archaeological object, and we measure its ^{14}C/^{12}C ratio $m(t)$ at the present time. The early use of the method assumed that the ^{14}C/^{12}C ratio in the carbon available to living organisms had remained constant over time, and so we could equate $m(0)$ to the known present ratio M_0. However, researchers soon had to accept that the rate of production of ^{14}C (and hence, the ^{14}C/^{12}C ratio in the atmosphere and oceans) has not remained constant over time, and the complex nature of this production process induces a relationship between calendar and radiocarbon ages that is non-monotonic, with many kinks (the so-called *wiggles*). This relationship is captured in the 1998 internationally-agreed high-precision calibration data sets which consist of observations (actually weighted averages of other observations) of the type

$$\theta_i^{(d)}, (y_i \pm \sigma_i); \qquad i = 1, \ldots n, \tag{2}$$

where for $i = 1, \ldots, n$, $y_i \pm \sigma_i$ represents the (average) radiocarbon-determined age assuming $m(0) = M_0$ and associated (average) standard error for a sample of known exact calendar age $\theta_i^{(d)}$. Since the raw data from which these averages were derived were not made publicly available, we follow other authors in treating these averages as observed data. A convention in radiocarbon dating, designed to avoid confusion over time and to allow comparison of dates, is to quote all dates in years *before present* (BP), where 'present' is defined to be the year 1950. Radiocarbon ages are always quoted in years BP, whereas calendar dates are expressed in years 'cal BP'.

2 The calibration data

There are currently two internationally-agreed calibration data sets (known as the INTCAL98 datasets): one suitable for calibrating radiocarbon ages for samples of organic material derived from organisms (such as terrestrial herbivores) that metabolised carbon from the atmospheric (or terrestrial) reservoir and the other for organisms (such as marine fish) that metabolised carbon from the marine (or oceanic) reservoir (Stuiver *et al.* 1998). Two calibration data sets are needed because there is a ^{14}C age offset between "terrestrial" and "oceanic" samples with identical calendar ages. Put simply, this offset arises because the (poorly understood) processes by which newly produced ^{14}C enters the ocean, mixes and becomes available to organisms that live in it is particularly slow (currently believed to be in the order of 400 years). Current knowledge about the mixing of ^{14}C in the atmosphere and oceans is summarised in the calibration issue of *Radiocarbon*, INTCAL98 (Stuiver *et al.* 1998).

In this paper we will focus on the atmospheric radiocarbon calibration data set and, more specifically, on the part of the data set that derives from measuring the ^{14}C/^{12}C ratio in tree-rings of known calendar age. Both the atmospheric and marine calibration data sets are publicly available at the University of Washington Quaternary Isotope Laboratory's World Wide Web site http://depts.washington.edu/qil/datasets. The calendar ages $\theta_1^{(d)}, \ldots, \theta_n^{(d)}$ of tree-ring samples are obtained via denrochronology (tree-ring dating), which is an extremely precise dating method that leads to reliable calendar age estimates (to the nearest year or even the nearest season in a given year). The INTCAL98 tree-ring data span 11,850 to zero cal BP (Stuiver *et al.* 1998) and are taken to have zero uncertainty on their calendar age estimates. Like all previous authors, we will take these data at face value. We are aware that they are not in reality raw data but have been subjected to a variety of preprocessing stages; however, the raw data are not readily available. Some discussion of deficiencies of these data, and of how they are being addressed in the ongoing INTCAL04 project, is given in Section 6.

3 Other information relevant to modelling

In addition to the fluctuation in the production rate of ^{14}C, there are other, more minor, problems which mean that calibration is essential. For instance, at the time of the development of the radiocarbon dating method, it was determined, to the levels of accuracy and precision then achievable, that the half-life $t_{\frac{1}{2}}$ for ^{14}C was 5568 years. This was used to set $\lambda = \ln 2/t_{\frac{1}{2}}$. It was subsequently discovered, however, that this value is incorrect, a better estimate of the true value being $t_{\frac{1}{2}} = 5730$. Once more, in order to facilitate references to results obtained with the original (incorrect) value, and to avoid the confusion regarding correcting those results, the convention is to continue the use of the initial value of 5568 years. This is known as Libby's value for the half-life of radiocarbon, and its corresponding decay rate will be denoted as $\lambda_L = \ln 2/5568$ throughout this paper.

For these reasons, the *radiocarbon age* of a specimen (obtained assuming $m(0) = M_0$

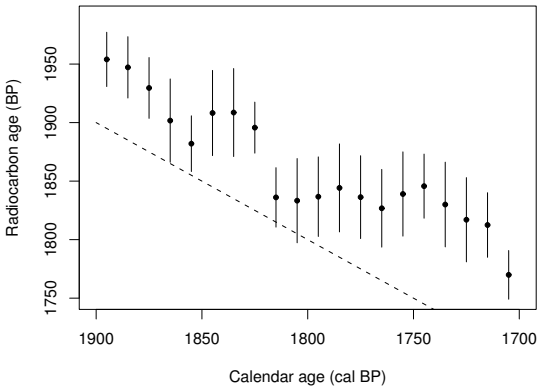


Figure 1: Part of the high-precision atmospheric calibration data (INTCAL98, Stuiver *et al.* 1998). The midpoints along each bar are the data points $(\theta_k^{(d)}, y_k)$, and the vertical lines represent the intervals $y_k \pm 2\sigma_k$. The dotted line represents the identity relationship between calendar and radiocarbon years BP.

and $\lambda = \lambda_L$) will differ from its true *calendar age*. This is evident in Figure 1, which shows a section of the tree-ring derived part of the INTCAL98 data (Stuiver *et al.* 1998) and the identity relationship between the two time scales (dotted line). Notice that the calendar-ages axis runs from right to left. This is yet another convention amongst the radiocarbon community. It simply implies that the 'flow of time' follows the usual direction from left to right; i.e. older ages are located further left in the plot. Thus, Figure 1 suggests that a specimen that truly dates from 1800 cal BP will be given a radiocarbon age of around 1830 (radiocarbon) years BP.

In principle, there is a deterministic relationship between the two timescales, in the sense that organisms that ceased metabolising in year θ will all have the same radiocarbon age $r(\theta)$. Thus, one of the objectives of radiocarbon calibration is to determine the calibration function $r(\cdot)$. A further objective of radiocarbon calibration is to transform (or *calibrate*) radiocarbon ages obtained for samples of material of unknown age onto the calendar timescale. Notice, however, that in Figure 1 radiocarbon ages in the range 1900–1700 years BP cannot be uniquely matched to calendar ages, and this is generally true throughout the radiocarbon dating range. Uncertainty about the true calendar age of a new radiocarbon sample is therefore a product of the interaction between uncertainty about $r(\cdot)$ and its nonlinear form.

3.1 Outline of paper

The calibration process consists of two phases. The first is the 'estimation phase', in which $r(\cdot)$, the functional relationship between calendar and radiocarbon years, is estimated from a set of calibration data. The second is the 'prediction phase', in which, from a new observation $y_0 \pm \sigma_0$, interest lies in making inference about the corresponding calendar age θ_0.

In this paper we are interested in the first stage of the calibration process for radiocarbon dating, namely inference about $r(\cdot)$, the radiocarbon calibration curve, describing the relationship between calendar and radiocarbon years. The problem effectively corresponds to making inference about an unknown function, some of whose values are observed with error.

Several approaches have been proposed throughout the history of radiocarbon dating for 'constructing' the radiocarbon calibration curve. The radiocarbon community seem to share the view that the calibration curve should be a smooth curve that nevertheless follows some of the 'wiggles' in the calibration data. There is currently no agreement as to how to achieve this, but commonly used models for $r(\cdot)$ are given by either the piecewise linear function that joins the calibration data (we refer to this as the 'traditional' radiocarbon calibration curve, since this was the earliest method adopted), or by spline functions or cubic interpolation which produce a smoother curve (for example, cubic interpolation is implemented in popular calibration software known as OxCal: Bronk Ramsey, 1995). These methods produce calibration curves that pass exactly through the data points $(\theta_i^{(d)}, y_i)$. Knox and McFadgen (1997) also argue for a smooth function and illustrate a method based on least-squares fitting of Fourier transformed and filtered calibration data that additionally smooths the data points. The latter feature is justified by the acknowledged measurement error in the calibration data. Gómez Portugal Aguilar *et al.* (2002) proposed modelling the calibration curve as a random walk, resulting in an estimate that is piece-wise linear but smooths the data points.

In what follows, we develop a model that attempts to incorporate scientific knowledge about the process of radiocarbon production. Naturally, the radiocarbon community is reluctant to embrace any proposal for modelling the calibration curve without clear evidence of the benefits and rationale behind them, and without making sure that the real problem is being properly addressed. Hence we proceed in stages, stressing the implications and significance of each assumption. We adopt a Bayesian approach to tackle this problem, based on the use of a Gaussian process prior structure for $r(\cdot)$, and incorporating our knowledge relating the calibration curve to the flux of ^{14}C in the atmosphere. We present the proposed model in Section 4. We emphasize the importance of the specification of the covariance structure, and illustrate its effects through the Gómez Portugal Aguilar *et al.* (2002) structure that yields a piece-wise linear calibration curve, but one that does not pass exactly through the data points $(\theta_i^{(d)}, y_i)$. This first step moves away from approaches that exactly interpolate the calibration data, whilst retaining the familiar feature of piece-wise linearity of the 'traditional' calibration curve. In Section 5 a more general process is introduced. The resulting calibration curve is

smooth, as well as allowing for the observation error in not simply interpolating the data points. Particular attention is paid to validating the model's predictions by cross-validation. The new model is shown to estimate $r(\cdot)$ more accurately than previous methods, and hence to lead to more accurate calibration of radiocarbon ages for new specimens. The results of this analysis are discussed in Section 6, where we also outline possible further refinements of the model.

4 Gaussian Process Model for the Radiocarbon Calibration Curve

We adopt a Bayesian approach to inference about the radiocarbon calibration curve $r(\cdot)$, relating calendar and radiocarbon ages. We interpret the calibration data (2) statistically as asserting that conditional on the true curve the values y_i are independently and normally distributed with known variances σ_i^2, thus

$$y_i \,|\, r(\cdot) \sim N(r(\theta_i^{(d)}), \sigma_i^2); \quad i = 1, 2, \ldots, n; \qquad \text{independent.}$$

This yields the likelihood function for the statistical analysis, and we now formulate a prior distribution for $r(\cdot)$.

The model we propose for $r(\cdot)$ is a Gaussian process (GP). There is a substantial literature concerning the use of GPs to represent prior distributions for unknown functions in a variety of situations; see for example O'Hagan (1978, 1992), Neal (1999), Kennedy and O'Hagan (2001), Schmidt and O'Hagan (2003) and Oakley and O'Hagan (2004). Formally, we suppose that the function $r(\cdot)$ has a GP prior distribution with mean function $m(\cdot)$ and covariance function $c(\cdot, \cdot)$, denoted by

$$r(\cdot) \,|\, \beta, \tau^2, \eta \sim GP(m(\cdot), c(\cdot, \cdot)) \ , \tag{3}$$

meaning that the joint distribution of $r(\theta_1), r(\theta_2), \ldots, r(\theta_N)$, for any $\theta_1, \theta_2, \ldots, \theta_N$ and any $N = 1, 2, \ldots$, is multivariate normal with parameters given by $E[r(\theta_j)] = m(\theta_j)$ and $Cov[r(\theta_j), r(\theta_{j'})] = c(\theta_j, \theta_{j'})$. The rationale for such modelling in general can be found in the previously cited references; the choice of the mean and covariance functions will be developed here to represent appropriate prior beliefs about the radiocarbon calibration function $r(\cdot)$. Note that the model (3) is expressed conditionally on three hyperparameters β, τ^2 and η, which will appear in our formulations of $m(\cdot)$ and $c(\cdot, \cdot)$.

We begin by setting the mean function as

$$m(\theta) = \beta\theta \tag{4}$$

describing the belief that radiocarbon ages are proportional to the true calendar ages. There are several reasons for this choice. First, it is certainly the simplest possible relationship between calendar ages and radiocarbon ages, since we really do not believe *a priori* that radiocarbon ages should be, for example, a quadratic or higher-order-polynomial function of their corresponding calendar ages. Another reason relates to

a general belief in the approximate stationarity of the atmospheric ^{14}C concentration. Stationarity (and using the correct half-life value for ^{14}C) would suggest $\beta = 1$, whereas allowing $\beta \neq 1$ encompasses the effects of both using an 'incorrect' value of the half-life of ^{14}C and of a non-stationarity drift over archaeological time of the production rate of ^{14}C. Ultimately, the Gaussian Process model allows the curve to deviate from linearity according to the information provided by the data, but the linear prior mean helps to damp spurious wiggles in the curve due to observation error.

The covariance function will take the form

$$c(\theta, \theta^*) = \tau^2 v_\eta(\theta, \theta^*) \ , \tag{5}$$

depending on a scale parameter τ^2 and a smoothness hyperparameter η in a form that will be developed in Section 4.1.

This stage of the model formulation is completed by establishing prior distributions for the hyperparameters β, τ^2 and η. As discussed above, there is some genuine prior information about β, in that the known inaccuracy in the Libby half-life λ_L means that $\beta = \lambda/\lambda_L = 5568/5730 = 0.97$ would be a better initial expectation than $\beta = 1$. However, this correction is small compared to uncertainty about the possibility of a drift over time in ^{14}C production that would lead to β deviating from this value. We chose the prior distribution $\beta \sim N(1, \frac{1}{36})$. This adopts the mean of 1 for convenience, and accommodates a prior judgement that β might easily deviate by as much as $\frac{1}{6}$ from unity due to non-stationarity in ^{14}C production.

It is, however, difficult to propose any meaningful prior information about the magnitude of the scale parameter τ^2 or the smoothness parameter η. We adopt the Jeffreys log-uniform prior for τ^2 and a uniform prior for η, independently of β, so that the prior distribution for the model parameters is

$$p(\beta, \tau^2, \eta) = p(\beta)p(\tau^2)p(\eta) \propto \frac{1}{\tau^2} \exp\left\{-18(\beta - 1)^2\right\} \ .$$

As in any Bayesian analysis, the posterior results will be a synthesis of the information contained in the prior structure and that provided by the data. The relative weight given to each of these components depends on the strength of belief stated in the prior, and on the 'amount' of data (the number of data points and their precision). The fact that we have access to a large set of data means that the choice of the prior settings for β, τ^2 and η will in fact have negligible influence on the posterior analysis.

4.1 Specification of the Covariance Structure

We now consider the formulation of the covariance function $c(.,.)$. Its definition is of fundamental importance in the resulting calibration curve. The choice of covariance structure will determine the degree of smoothness and differentiability of the resulting curve. Furthermore, a suitable choice for the covariance function can represent specific characteristics of the radiocarbon calibration curve, such as the existence of a starting

point (zero point) in the chronology, and the fact that the uncertainty in the relationship between calendar and radiocarbon ages grows the further we move back in time from this starting point.

This second issue deserves some elaboration since two separate factors contribute to the uncertainty on the radiocarbon timescale. Firstly, there is the uncertainty related to a particular estimate of the radiocarbon content of a sample. Due to practical restrictions (not least the extremely tiny proportion of ^{14}C remaining in old specimens), this uncertainty is greater for older samples. This is represented in the standard errors reported for each of the radiocarbon determinations that make up the calibration data sets, and we assume here that $\sigma_1^2, \ldots, \sigma_n^2$ adequately reflect this aspect of the uncertainty.

The other source of uncertainty is represented by the covariance structure in the Gaussian process model and relates to our *current knowledge* about the true relationship between calendar and radiocarbon ages. In recent years, substantial effort has been made to extend the calibrated ^{14}C record and to offer us greater insight into the range and nature of the variations in ^{14}C levels over time (van der Plicht, 2002; Beck *et al.* 2001). Beyond the limits of INTCAL98, data relating calendar and ^{14}C ages are mainly based on foraminifera trapped in laminated sediments and on various carbonates dated by U-series isotopes (van der Plicht, 2002). The results provide valuable qualitative information about the relationship between the two timescales. This work makes it clear that our present knowledge about the relationship between calendar and radiocarbon ages is subject to a higher level of uncertainty for older periods of time than for more recent ones.

In order to derive a suitable covariance structure for the radiocarbon calibration curve, we propose making use of the link that exists between the process of generation of ^{14}C and the relationship between calendar and radiocarbon ages.

In practice, the determination of the radiocarbon age $r(\theta)$ of a sample of (unknown) calendar age θ that has a current ^{14}C/^{12}C ratio m is performed, as mentioned earlier, using Libby's half-life value for ^{14}C, so that

$$r(\theta) = -\frac{1}{\lambda_L} \ln\left(\frac{m}{M_0}\right).$$

Apart from the measurement error represented by σ^2, m is given by (1):

$$m = M(\theta) \exp\left(-\lambda\theta\right) \ ,$$

and depends upon the original (but unknown) atmospheric ^{14}C/^{12}C ratio at time θ cal BP, $M(\theta)$, and the correct half-life value for ^{14}C, λ. Hence,

$$r(\theta) = \frac{\lambda}{\lambda_L}\theta - \frac{1}{\lambda_L} \ln\left(\frac{M(\theta)}{M_0}\right).$$

Consider the rate of change

$$r'(\theta) = \frac{d}{d\theta} r(\theta) = \frac{\lambda}{\lambda_L} - \frac{1}{\lambda_L}\frac{M'(\theta)}{M(\theta)} \ . \tag{6}$$

It is well known that $M(\cdot)$ is not strictly stationary, and many factors, such as geomagnetism and sunspot activity, may affect its behaviour (Bowman, 1990). In particular, $M(\theta)$ represents the accumulated effects to time θ of both ^{14}C generation in the upper atmosphere and its dissipation through the biosphere and decay. We propose to model $r'(\cdot)$ as a stationary Gaussian process, and refer to it as the flux process.

Thus, we now redefine our model for the radiocarbon calibration curve as the integrated version of the process $r'(\cdot)$, using the fact that $r(0) = 0$, i.e.

$$r(\theta) = \int_0^\theta r'(t)\, dt,$$

where we now suppose that

$$r'(.) \sim GP\left(\beta, \tau^2 w_\eta(\cdot,\cdot)\right) .$$

We specify that $w_\eta(\cdot,\cdot)$ is a correlation function through $w_\eta(\theta,\theta) = 1, \forall\theta$, so that τ^2 is the variance of the process. Then it is straightforward to show that $r(\cdot)$ is also a GP with mean function

$$E(r(\theta)) = \int_0^\theta \beta dt = \beta\theta,$$

as in (4). As for the covariance structure for $r(\cdot)$, in terms of (5)

$$v_\eta(\theta,\theta^*) = \int_0^\theta \int_0^{\theta^*} w_\eta(t,u)\ dt\ du, \tag{7}$$

and we need to define an appropriate correlation function $w_\eta(\cdot,\cdot)$ for $r'(\theta)$. From equation (6) we see that the uncertainty for $r'(\theta)$ is directly related to the uncertainty of $M'(\theta)$, which corresponds to the flux that generates changes in ^{14}C levels. Notice also from (7) that uncertainty about $r(\theta)$ increases with θ, reflecting the fact that, prior to incorporating the high-precision calibration data, the uncertainty about the radiocarbon age should be an increasing function of calendar age.

4.2 A piecewise linear curve

An initial possibility for establishing the full model for the radiocarbon calibration curve is by proposing a pure random noise model for this flux. This was the approach in Gómez Portugal Aguilar *et al.* (2002) (henceforth, GPA, L & O'H). That is,

$$w_\eta(t,u) = w_0(t,u) = \begin{cases} 1 & ; \text{ if } t = u, \\ 0 & ; \text{ otherwise.} \end{cases}$$

This gives rise to the following covariance structure for the calibration curve $r(\theta)$

$$Cov(r(\theta), r(\theta^*)) = \tau^2 v(\theta,\theta^*) = \tau^2 \min\{\theta,\theta^*\}. \tag{8}$$

Thus, $r(\cdot)$ is a random walk or Wiener process, with no hyperparameter η in the correlation function. It can be proved that the resulting posterior mean for the radiocarbon calibration curve is piece-wise linear.

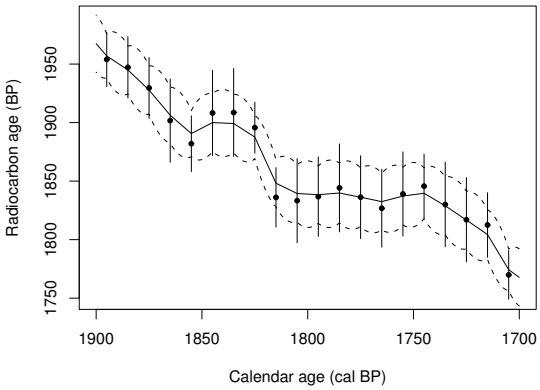


Figure 2: A section of the GPA, L & O'H piece-wise linear curve resulting from the random-walk covariance structure in equation (8). The solid line is the posterior mean, and the dashed lines show the mean plus and minus two standard deviations. The dots and vertical bars show the calibration data points plus and minus two standard errors as in Figure 1.

Our piece-wise linear (posterior mean) radiocarbon calibration curve, however, differs from the traditionally employed curve in that it smooths the data rather than interpolating them (see Figure 2). This is a fully desirable feature of our approach, reflecting the fact that the calibration data are themselves subject to uncertainty as expressed through the reported errors $\sigma_1, \sigma_2, \ldots \sigma_n$. Furthermore, we can assess the posterior uncertainty associated with the calibration curve through the computation of the posterior variance for any point $r(\theta)$ on the calibration curve. GPA, L & O'H carried out a detailed comparison of this assessment, with those corresponding to the calibration programs CALIB (Stuiver and Reimer, 1993) and BCal (Buck, Christen and James, 1999). It emerged from this comparison that the posterior variance values for any point on the calibration curve used in those programs are much larger than the results arising from our formulation (for further details refer to GPA, L & O'H.)

This constitutes the first, motivating proposal for modelling the radiocarbon calibration curve. It retains the familiarity of straight-line segments whilst breaking away from the data points themselves, which is justified in terms of the acknowledged error in the data points, plus the resulting estimate of uncertainty in the flux process.

5 Smooth Calibration Curve

A clear advantage of our modelling strategy for the radiocarbon calibration curve is that the proposed (integrated) Gaussian process structure is a generic one. In particular, the

random-walk choice for the covariance structure that yields a piece-wise linear posterior calibration curve arises by modelling the flux of ^{14}C as white noise. However, it is arguable that the flux should be less erratic than white noise, and this is matched by a feeling in at least part of the radiocarbon dating community that the calibration curve should be smooth rather than piece-wise linear (as indicated by the fact that some researchers favour cubic interpolation, e.g. Bronk Ramsey, 1995). The same basic structure we have proposed can accommodate this possibility if we establish a correlation structure for the flux of ^{14}C, $w_\eta(\theta, \theta^*)$, that is smoother than the integrated white noise (the random-walk model) of GPA, L & O'H. This is reasonable, not only because we have no reason to believe *a priori* that this process does not have a smooth behaviour over time, but also because even if production of ^{14}C did change in a stepwise manner, the atmospheric mixing time and continuous exchange between atmosphere and ocean would smooth out any jumps.

We propose an integrated correlation structure for the radiocarbon calibration curve $r(\cdot)$, given by equation (7) with

$$w_\eta(\theta, \theta^*) = \exp\left\{-\frac{1}{2\eta^2}(\theta - \theta^*)^2\right\} .$$

This form implies that $r'(\cdot)$ is infinitely differentiable everywhere, provided the prior mean function is too (O'Hagan, 1992). That is, this correlation function expresses the belief that the curve of interest has derivatives of any order; it is therefore appropriate when the curve is believed to vary smoothly. The hyperparameter η describes how rough or smooth the function is. The larger the value of η, the higher the correlation between two values of the function, hence the smoother the curve is. As $\eta \to 0$, the correlation between any two values goes to zero and the process approaches white noise. Hence this formulation will include the previous analysis as a limiting case.

Thus, the model we propose for the atmospheric radiocarbon calibration curve is given by

$$\begin{aligned} y_i \,|\, r(\cdot) &\sim N\left(r(\theta_i^{(d)}), \sigma_i^2\right), \quad i = 1, 2, \ldots, n \ \text{ indep.} \\ r(\theta) \,|\, \theta, \beta, \tau^2, \eta &\sim GP\left(\beta\theta, \tau^2 v_\eta(\theta, \theta)\right) \qquad (9) \\ p(\beta, \tau^2, \eta) &\propto \left(\tau^2\right)^{-1} \exp(-18(\beta - 1)^2) . \qquad (10) \end{aligned}$$

The integrated covariance function in (9) can be evaluated as

$$\begin{aligned} v_\eta(\theta, \theta^*) &= \int_0^\theta \int_0^{\theta^*} \exp\left(-\frac{1}{2\eta^2}(t - u)^2\right) dt\, du \\ &= \sqrt{2\pi}\eta\,[\eta\{\phi(\theta/\eta) + \phi(\theta^*/\eta) - \phi((\theta - \theta^*)/\eta) - \phi(0)\} \\ &\qquad + \theta\{\Phi(\theta/\eta) - \Phi((\theta - \theta^*)/\eta)\} \\ &\qquad + \theta^*\{\Phi(\theta^*/\eta) + \Phi((\theta - \theta^*)/\eta) - 1\}] \quad \forall \theta \geq \theta^* , \qquad (11) \end{aligned}$$

where $\phi(\cdot)$ is the standard normal density function and $\Phi(\cdot)$ is the standard normal distribution function. Although the explicit form (11) may appear complex, it is computationally straightforward to evaluate.

The prior distribution (10) for the hyperparameters combines the $N(1, 1/36)$ prior distribution for β with a noninformative log-uniform prior distribution for τ^2 and a uniform prior distribution for the smoothness parameter η. In fact we shall see that the data are sufficiently substantial to identify these hyperparameters quite accurately enough for prior information to be unimportant.

5.1 Posterior distributions of hyperparameters

For computational reasons, it is convenient to let $y_0 = 0$, $\theta_0^{(d)} = 0$, and $\sigma_0^2 = 0$, and to work in terms of the differences

$$z_i = y_i - y_{i-1} \ , \qquad \delta_i = r(\theta_i^{(d)}) - r(\theta_{i-1}^{(d)}) \ , \qquad d_i = \theta_i^{(d)} - \theta_{i-1}^{(d)}$$

for $i = 1, 2, \ldots, n$. Thus we regard the data as being $z_1, z_2, \ldots, z_n$, instead of $y_1, y_2, \ldots, y_n$, with distributions

$$z_i \,|\, r(\cdot) \sim N(\delta_i, \sigma_i^2 + \sigma_{i-1}^2) \ , \quad i = 1, 2, \ldots, n \ . \tag{12}$$

Note that the z_is are not independent, as

$$\mathrm{cov}(z_{i+1}, z_i \,|\, r(\cdot)) = -\sigma_i^2 \ , \tag{13}$$

but the covariances between z_is whose subscripts differ by more than 1 are zero.

From (9) and (11), the δ_is have a multivariate normal prior distribution with the following structure.

$$\begin{aligned} E(\delta_i \,|\, r(\cdot), \beta, \tau^2, \eta) &= \beta d_i \ , \\ \mathrm{var}(\delta_i \,|\, r(\cdot), \beta, \tau^2, \eta) &= \tau^2\eta^2\sqrt{2\pi}[2g(\gamma_i - \gamma_{i-1}) \\ &\qquad -2g(0) - (\gamma_i - \gamma_{i-1})] \ , \qquad (14) \\ \mathrm{cov}(\delta_i, \delta_{i-k} \,|\, r(\cdot), \beta, \tau^2, \eta) &= \tau^2\eta^2\sqrt{2\pi}[g(\gamma_i - \gamma_{i-k-1}) \\ &\qquad -g(\gamma_i - \gamma_{i-k}) \\ &\qquad - g(\gamma_{i-1} - \gamma_{i-k-1}) \\ &\qquad +g(\gamma_{i-1} - \gamma_{i-k})] \ , \qquad (15) \end{aligned}$$

for $i = 1, 2, \ldots, n$ and $k = 1, 2, \ldots, i-1$, where

$$\gamma_i = \theta_i^{(d)}/\eta \ ,$$

$$g(x) = \phi(x) + x\Phi(x) \ .$$

Therefore, letting $\mathbf{z} = (z_1, z_2, \ldots, z_n)^T$ and $\mathbf{d} = (d_1, d_2, \ldots, d_n)^T$, we have that

$$\mathbf{z} \,|\, \beta, \tau^2, \eta \sim N(\beta\mathbf{d}, \tau^2\mathbf{A}_\eta + \mathbf{S}) \ , \tag{16}$$

where $\tau^2\mathbf{A}_\eta$ is the variance matrix of the δ_is, a function of η with diagonal elements given by (14) and off-diagonal elements by (15), and where

$$\mathbf{S} = \begin{pmatrix} \sigma_1^2 & -\sigma_1^2 & 0 & 0 & \dots & 0 \\ -\sigma_1^2 & \sigma_1^2+\sigma_2^2 & -\sigma_2^2 & 0 & \dots & 0 \\ 0 & -\sigma_2^2 & \sigma_2^2+\sigma_3^2 & -\sigma_3^2 & & 0 \\ 0 & 0 & -\sigma_3^2 & \sigma_3^2+\sigma_4^2 & & \\ \vdots & \vdots & & & \ddots & \\ 0 & 0 & 0 & & & \sigma_{n-1}^2+\sigma_n^2 \end{pmatrix}$$

is the variance matrix of $\mathbf{z}$ given the δ_is, from (12) and (13).

The purpose of working with differences z_i is to make the variance matrix $\tau^2\mathbf{A}_\eta + \mathbf{S}$ better behaved than it would have been had we worked with the y_is. In particular, the elements of $\mathbf{A}_\eta$ are smaller and this matrix is much nearer to being diagonal than the corresponding variance matrix of the $r(\theta_i^{(d)})$s. Although the variance matrix of the y_is would be diagonal, the change to tri-diagonal $\mathbf{S}$ still leaves $\tau^2\mathbf{A}_\eta + \mathbf{S}$ much less ill-conditioned.

Combining the likelihood from (16) with the prior distribution (10), the posterior distribution of the hyperparameters is

$$\begin{aligned} p(\beta, \tau^2, \eta \,|\, \mathbf{z}) \quad &\propto \quad \tau^{-2} \left|\tau^2\mathbf{A}_\eta + \mathbf{S}\right|^{-1/2} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{z}-\beta\mathbf{d})^T(\tau^2\mathbf{A}_\eta+\mathbf{S})^{-1}(\mathbf{z}-\beta\mathbf{d}) - 18(\beta-1)^2\right\}. \end{aligned} \tag{17}$$

It is now clear that the conditional posterior distribution of β given τ^2 and η is normal, given by

$$\beta \,|\, \tau^2, \eta, \mathbf{z} \sim N(m_\beta, v_\beta) \ , \tag{18}$$

where

$$\begin{aligned} m_\beta &= v_\beta \left(\mathbf{d}^T(\tau^2\mathbf{A}_\eta+\mathbf{S})^{-1}\mathbf{z} + 36\right) \ , \\ v_\beta &= 1/\left(\mathbf{d}^T(\tau^2\mathbf{A}_\eta+\mathbf{S})^{-1}\mathbf{d} + 36\right) \end{aligned}$$

are functions of both τ^2 and η. We can therefore integrate out β from (17), with the result

$$\begin{aligned} p(\tau^2, \eta \,|\, \mathbf{z}) \quad &\propto \quad \tau^{-2} \left|\tau^2\mathbf{A}_\eta + \mathbf{S}\right|^{-1/2} \sqrt{v_\beta} \\ &\quad \times \exp\left\{-\frac{1}{2}(\mathbf{z}-m_\beta\mathbf{d})^T(\tau^2\mathbf{A}_\eta+\mathbf{S})^{-1}(\mathbf{z}-m_\beta\mathbf{d}) - 18(m_\beta-1)^2\right\}. \end{aligned} \tag{19}$$

5.2 Posterior distribution of the calibration curve given the hyperparameters

The posterior distribution of $r(\cdot)$ conditional on the hyperparameters can be constructed in two steps. First, letting $\boldsymbol{\delta} = (\delta_1, \delta_2, \ldots, \delta_n)^T$, the joint prior distribution of any collection of $r(\theta)$s and $\boldsymbol{\delta}$ is multivariate normal, from the definition of the GP (9). It is then straightforward to derive from standard conditioning results that

$$r(\cdot) \,|\, \boldsymbol{\delta}, \beta, \tau^2, \eta \sim GP\left(m^*(\cdot), \tau^2 v^*(\cdot,\cdot)\right) \ ,$$

where

$$m^*(\theta) = \beta\theta + \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1}(\boldsymbol{\delta} - \beta\mathbf{d}) \ ,$$
$$v^*(\theta, \theta^*) = v_\eta(\theta, \theta^*) - \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1}\mathbf{t}(\theta^*) \ ,$$

$\mathbf{t}(\theta)$ is the vector of covariances between $r(\theta)$ and $\boldsymbol{\delta}$, with elements

$$\begin{aligned} t_i(\theta) &= v_\eta(\theta, \theta_i) - v_\eta(\theta, \theta_{i-1}) \\ &= \begin{cases} \eta^2\sqrt{2\pi}\left\{g(\gamma_i) - g(\gamma_{i-1}) - g(\gamma - \gamma_i) + g(\gamma - \gamma_{i-1}) - (\gamma_i - \gamma_{i-1})\right\} \\ \qquad\qquad \text{if } \theta \geq \theta_i \\ \eta^2\sqrt{2\pi}\left\{g(\gamma_i) - g(\gamma_{i-1}) - g(\gamma_i - \gamma) + g(\gamma - \gamma_{i-1}) - (\gamma - \gamma_{i-1})\right\} \\ \qquad\qquad \text{if } \theta_i \geq \theta \geq \theta_{i-1} \\ \eta^2\sqrt{2\pi}\left\{g(\gamma_i) - g(\gamma_{i-1}) - g(\gamma_i - \gamma) + g(\gamma_{i-1} - \gamma)\right\} \\ \qquad\qquad \text{if } \theta_{i-1} \geq \theta \end{cases} \end{aligned}$$

and $\gamma = \theta/\eta$.

Second, we have from (12) and (13),

$$\mathbf{z} \,|\, r(\cdot) \sim N(\boldsymbol{\delta}, \mathbf{S}) \ .$$

We then find that the posterior distribution of $\boldsymbol{\delta}$ given the hyperparameters is

$$\boldsymbol{\delta} \,|\, \beta, \tau^2, \eta, \mathbf{y} \sim N(\mathbf{m}, \mathbf{V}) \ ,$$

where

$$\begin{aligned} \mathbf{V} &= (\tau^{-2}\mathbf{A}_\eta^{-1} + \mathbf{S}^{-1})^{-1} \ , \\ \mathbf{m} &= \mathbf{V}(\tau^{-2}\mathbf{A}_\eta^{-1}\mathbf{d}\beta + \mathbf{S}^{-1}\mathbf{z}) \ . \end{aligned}$$

Therefore, putting the two steps together,

$$r(\cdot) \,|\, \beta, \tau^2, \eta, \mathbf{y} \sim GP(m^{**}(\cdot), c^{**}(\cdot,\cdot)).$$

where

$$m^{**}(\theta) = \beta\theta + \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1}(\mathbf{m} - \beta\mathbf{d}) \ ,$$
$$c^{**}(\theta, \theta^*) = \tau^2\left\{v_\eta(\theta, \theta^*) - \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1}\mathbf{t}(\theta^*)\right\} + \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1}\mathbf{V}\mathbf{A}_\eta^{-1}\mathbf{t}(\theta^*) \ .$$

We can now marginalise this with respect to the distribution (18) of β given τ^2, η and $\mathbf{z}$, to give

$$r(\cdot) \,|\, \tau^2, \eta, \mathbf{y} \sim GP(m^{***}(\cdot), c^{***}(\cdot,\cdot)) \;, \tag{20}$$

where (after some simplification)

$$m^{***}(\theta) = m_\beta \theta + \tau^2 \mathbf{t}(\theta)^T (\tau^2 \mathbf{A}_\eta + \mathbf{S})^{-1} (\mathbf{z} - m_\beta \mathbf{d}) \tag{21}$$

and

$$\begin{aligned} c^{***}(\theta, \theta^*) \;\; = \;\; & \tau^2 \left\{ v_\eta(\theta, \theta^*) - \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1} \mathbf{t}(\theta^*) \right\} + \mathbf{t}(\theta)^T \mathbf{A}_\eta^{-1} \mathbf{V} \mathbf{A}_\eta^{-1} \mathbf{t}(\theta^*) \\ & + v_\beta \left\{ \theta - \tau^2 \mathbf{t}(\theta)^T (\tau^2 \mathbf{A}_\eta + \mathbf{S})^{-1} \mathbf{d} \right\} \left\{ \theta^* - \tau^2 \mathbf{t}(\theta^*)^T (\tau^2 \mathbf{A}_\eta + \mathbf{S})^{-1} \mathbf{d} \right\} . \end{aligned} \tag{22}$$

5.3 Estimating τ^2 and η

The marginal posterior distribution of $r(\cdot)$ is in principle now obtained by marginalising its conditional posterior (20) with respect to the posterior distribution (19) of τ^2 and η (representing, respectively, the variability and smoothness of the flux process $r'(\cdot)$). However, the latter is complex, and even though it is only two-dimensional involves a substantial computational burden. The reason is that we have $n = 1149$ data points, covering the most recent (tree-ring derived) part of the INTCAL98 data (11,850 to zero cal BP). To compute (19) at any point requires the inversion of the 1149×1149 matrix $\tau^2 \mathbf{A}_\eta + \mathbf{S}$. The use of differences was designed to make it possible to do these inversions without numerical instability, but the simple computing time of constructing and inverting the matrix, followed by several other matrix operations with the inverse, is non-trivial. Rather than formally marginalising with respect to this distribution, we have chosen to adopt the approach of finding point estimates of τ^2 and η and then to substitute these into (20). The justification of this approach lies in the following observations.

1. Exploration of (19) shows that the posterior density falls to less than one-hundredth of its modal value for τ^2 and η outside a fairly narrow region. Trying a variety of values in this region suggests that the posterior conditional distribution (20) is quite insensitive to the choice. Therefore, although conditioning rather than marginalising will understate the posterior uncertainty, it seems that the understatement will be very small compared with the posterior uncertainty in $r(\theta)$ expressed in (20) conditional on any plausible values of τ^2 and η.

2. The predictive variances produced by this conditional approach validate well with the data, as will be shown below, suggesting again that any understatement of uncertainty is immaterial.

3. There is a need to provide a calibration curve for the radiocarbon community that is as simple as possible. We have described in Section 3.1 that estimating the calibration curve is just the first of two tasks, the second of which is to use

the curve to calibrate new radiocarbon ages obtained for archaeological samples of unknown calendar age. The second task is itself computationally intensive, meaning that there is a benefit in having a relatively simple computation of the curve and its uncertainty that is gained for every subsequent use in practice.

The joint posterior mode is at $\tau^2 = 3.33$ and $\eta = 6.7$. We therefore propose to use the GP (20) conditioned on these values as the posterior distribution for the calibration curve.

It is worth noting that η is defined such that the correlation between $r'(\theta)$ and $r'(\theta + \eta)$ equals $\exp(-0.5) = 0.6$. This therefore determines the time lapse over which the flux process $r'(\cdot)$ is reasonably highly correlated. A value of 6.7 can therefore be interpreted as suggesting stability of the flux process over a period of a few years but not tens of years. It is clear that this will lead to quite different estimation of the process $r(\cdot)$ than the implicit value of $\eta = 0$ in the analysis of GPA, L & O'H.

5.4 Model validation

An alternative method for estimating the smoothness parameter η is cross-validation (Oakley, 1999). In view of the importance of the smoothness parameter in GP modelling, it is sensible to use cross-validation as a check on both the model estimates and the model fit.

For this purpose, we leave out each observation in turn and use the remaining data to fit the model and predict the missing observation. If $\hat{y}_i$ is the predictive expectation of y_i using all observations except y_i the cross-validation fit criterion is $\sum_{i=1}^{n}(y_i - \hat{y}_i)^2$. We can then choose τ^2 and η to minimise this criterion. (We omitted the twentieth century calibration data points in both the posterior mode calculation and cross-validation, since it is clear that these are affected by human activity.)

This is again a computationally intensive process. We need to invert a 1148×1148 matrix for each of the $n = 1149$ missing observations. However, it is easy to see that this is not strictly necessary. Conditional on τ^2 and η, the predictive distribution for y_i will only depend on the calibration data for $\theta_j^{(d)}$ points reasonably close to $\theta_i^{(d)}$. The range depends on η, but we found that for values of η of interest it is certainly adequate to use a window of ten points on each side of the missed out data point, a range covering at least 100 years on each side. This makes it computationally feasible to explore the (τ^2, η) space to minimise the cross-validation criterion. The optimal values were found to be $\tau^2 = 3.37$, $\eta = 6.6$, very close to the modal values.

The cross-validation can also be used as a direct model check. Conditional on the chosen values of τ^2 and η, the predictive distribution of y_i is normal with mean the posterior mean (21) of $r(\theta_i^{(d)})$ and variance σ_i^2 plus the posterior variance (22) of $r(\theta_i^{(d)})$. We can therefore standardise the predictive residuals $y_i - \hat{y}_i$ by dividing by the predictive standard deviation. The resulting standardised residuals should look like a sample from a standard normal distribution. Figure 3 shows a normal probability plot of the

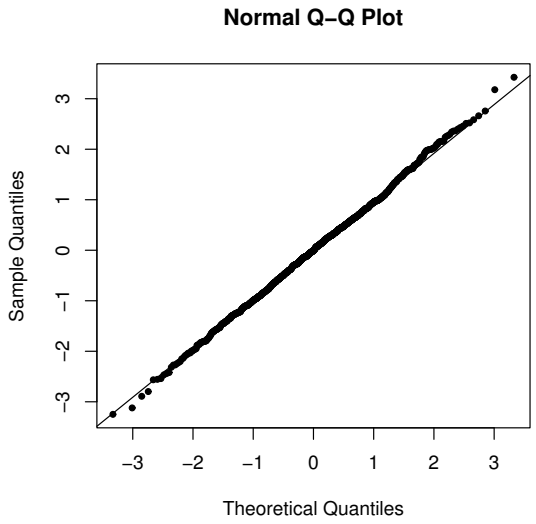


Figure 3: Normal probability plot of standardised cross-validation residuals with $\tau^2 = 3.33$ and $\eta = 6.7$. The straight line corresponds to the standard normal distribution.

residuals. The straight line is the theoretical $N(0, 1)$ line, and it is clear that the model fit is almost perfect in this respect. The very slight curvature at the ends indicates a slightly heavier tail than the normal, and is consistent with having ignored uncertainty in τ^2, but corresponds roughly to a Student t distribution with many hundreds of degrees of freedom.

Figure 4 shows the standardised residuals plotted over time. There is an indication here of heteroscedasticity, with larger variances for more recent values, but this impression is entirely due to a very small number of residuals out of 1149. Averages of 20 consecutive residuals show no such heteroscedasticity. We also fitted the model to the data for the most recent 5000 years and found almost identical estimates of τ^2 and η, indicating that the underlying model is stable over the whole time period.

Overall, we consider that the assumptions implied by the model are adequately verified.

5.5 Posterior inference for the calibration curve

Figure 5 shows the posterior mean and one standard deviation bounds for $r(\theta)$ over the range of θ from 1900 to 1700 cal BP. The way that the estimated curve smooths the data points is clear, and the resulting curve is smoother and more plausible than the GPA, L & O'H curve shown in Figure 2.

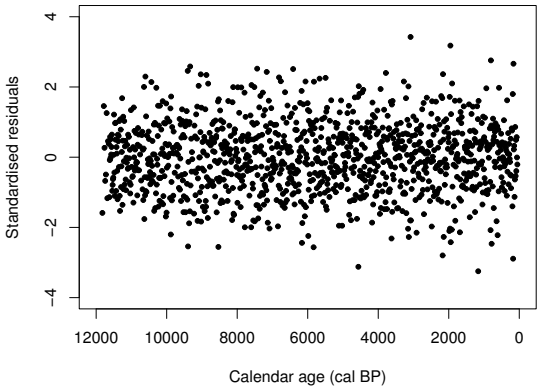


Figure 4: Standardised cross-validation residuals plotted against calendar age.

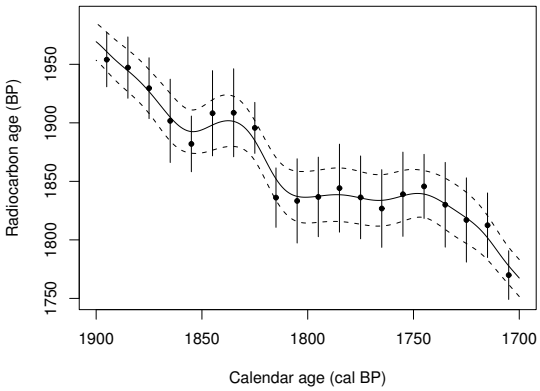


Figure 5: The new smooth calibration curve over the time period 1900–1700 cal BP. The posterior mean is shown as a solid line, with dashed lines at plus and minus two posterior standard deviations. The dots and vertical lines show the calibration data y_i plus and minus $2\sigma_i$.

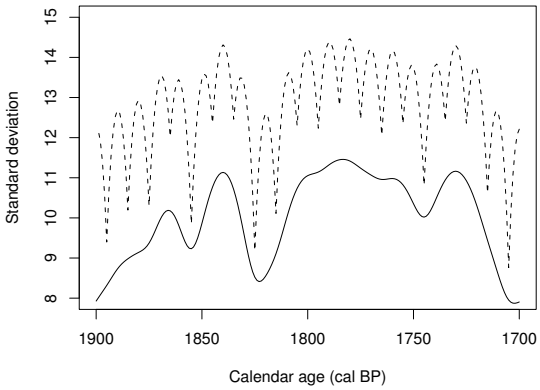


Figure 6: Posterior standard deviations for the new smooth calibration curve (solid line) and the piecewise linear curve of GPA, L & O'H (dashed), over the time period 1900–1700 cal BP.

The posterior variance is also important. Figure 5 shows that the calibration data have not identified the calibration curve without some appreciable posterior uncertainty. Nevertheless, the standard deviation of the curve from this new analysis is much smaller than has been obtained by other approaches. Figure 6 compares the standard deviation with that of GPA, L & O'H over the same time interval. We see that the new methodology predicts with distinctly greater precision. GPA, L & O'H showed that their estimates generally had lower standard deviations than any other plausible methods.

It is important to emphasise that this claim of lower predictive uncertainty is backed up by the cross-validation analysis and Figure 3, showing that, as far as it is possible to verify, the new method does indeed predict points on the curve to within the claimed margin of error.

5.6 Example of calibration

Figure 6 shows that the reduction in uncertainty is appreciable, but Figure 5 demonstrates that even a modest improvement can have a disproportionate effect on the accuracy with which we can calibrate new radiocarbon determinations. Suppose that we obtain a radiocarbon age of 1870 BP. Then although the most probable calibration of this is to about 1820 cal BP there is enough uncertainty around the curve to allow a possibility that the true calendar age is anywhere from 1830 cal BP to almost 1730 cal BP. The range of uncertainty for the calibration is already much less with our analysis than would be the case with previous methods. Figure 7 demonstrates this by showing the posterior probability distribution for the true calendar age θ_0 of a sample with

radiocarbon age $y_0 = 1870$ BP and standard error $\sigma_0 = 30$ years, from three different calibration curves. These are based on a weak prior distribution; see Buck et al (1996) for details of how these posterior distributions are derived.

We see that the uncertainty about θ_0 is substantially less in Figure 7(a) using the new smooth calibration curve than in Figure 7(b) which results from the GPA, L &O'H piece-wise linear curve. The posterior distribution in Figure 7(c) arises from calibration of the same radiocarbon age y_0 and standard error σ_0 using the traditional piecewise linear interpolation of the INTCAL98 data. This shows dramatically wider uncertainty about θ_0. All three curves have basically four modes at the same places but, primarily because of its lower variance, the new smooth curve discriminates far more strongly between them. The 'lumpiness' of the density in Figure 7(b) is linked to the behaviour of the standard deviation (the dashed curve in Figure 6), which in turn arises from the non-smoothness of the flux process in the model of GPA, L & O'H.

It is possible that some of the improvement that we have achieved with the new radiocarbon calibration curve is due to our having to take the INTCAL98 data at face value, whereas we know that they ignore uncertainty in the calendar ages and may have been over-smoothed. Nevertheless, we believe that the nonparametric Bayesian modelling has the potential to provide a major advance for the radiocarbon dating community.

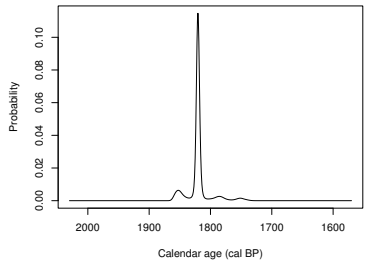
5.7 The slope parameter β

We argued in Section 4 that the prior settings for the model hyperparameters would, given the vast amount of calibration data, have little impact on the posterior results. In particular, we claimed that the posterior inference about β would not be materially affected by the $N(1, 1/36)$ prior distribution. Using (18) and the estimated values of τ^2 and η (but the distribution is quite insensitive to variations of these parameters), we find that the posterior distribution of β is $N(0.862, 0.00476)$. A 99% posterior credible interval is therefore from 0.75 to 0.97, showing that not only is β almost certainly less than 1 but also that it is very probably less than the ratio $\lambda/\lambda_L = 0.97$.

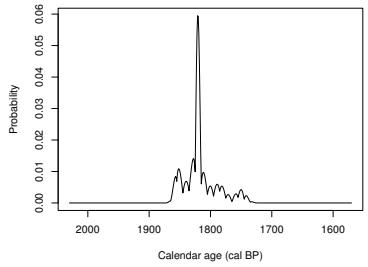
6 Discussion

The proposed methodology provides a sound statistical framework for making inference about the radiocarbon calibration curve. Within our Bayesian approach, the calibration curve is dealt with as an unknown function some of whose values are observed, subject to observation error.

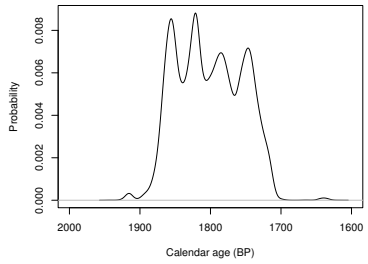
The Gaussian process prior distribution is a very flexible structure that allows the incorporation of the information provided by the calibration data, in accordance with their quality. The structure also allows the incorporation of prior information not only about a plausible mean function, but also about features such as continuity and differentiability of the calibration curve.



(a)



(b)



(c)

Figure 7: Posterior densities for true calendar age θ_0 of a sample with radiocarbon age 1870 ± 30. (a) Density from the new smooth radiocarbon curve. (b) Density from the piece-wise linear curve of GPA, L & O'H. (c) Density from the traditional piecewise linear interpolation of the INTCAL98 data.

Specific characteristics of the process underlying the generation of ^{14}C led to the idea of having an integrated process in order to reflect the nonstationarity of the ^{14}C level. Thus, prior beliefs about the ^{14}C flux in the atmosphere have been incorporated through a stationary covariance structure for $r'(\cdot)$. As a result of this, the fact that there is growing uncertainty in the dating method as we move further back in the timescale has also been acknowledged in the model.

In this paper we have outlined the case of a piece-wise linear calibration curve, arising from assuming a white noise model for the flux of ^{14}C in the atmosphere and previously analysed fully by GPA, L & O'H. We also offered a proposal for a smooth calibration curve by adjusting beliefs about the flux process from a random-walk to a smoothly-varying process. Just how smoothly this process varies is dictated by a further hyperparameter η whose value is estimated from the data. The tree-ring subset of the INTCAL98 calibration data support a value of $\eta = 6.7$ that yields a posterior mean which indicates smoothness of the flux process over a time-scale of several years. The resulting radiocarbon calibration curve varies more smoothly than the piecewise linear form and represents a more plausible representation of $r(\cdot)$.

Our method can in principle take account of uncertainty in the hyperparameters η and τ^2 (the variance of the flux process). However, we have argued that the substantial extra computational effort of this is not necessary. Instead we simply estimate these parameters and then treat them as fixed. Our model validation procedure has suggested that the posterior variances are not being underestimated, confirming our claim that to account for the extra uncertainty in η and τ^2 would have negligible effect. Furthermore, the fact that we obtain substantially lower variances than other methods can be attributed to more accurate statistical modelling and analysis.

There is clearly scope for incorporating even more of modern scientific opinion on the production of radiocarbon, for instance by modelling the flux as linked to the sunspot cycle. The finding that the flux process is coherent on a scale comparable with the 11-year sunspot cycle suggests the possibility to reduce uncertainty in $r(\cdot)$ still further. In addition, there is further structure to the data that has not thus far been well handled during the curve building process. For example, due to physical restrictions related to the measurement of the proportion of ^{14}C remaining in a sample, the construction of the calibration data set often involves the need to handle tree-ring samples whose calendar age estimate arises from a number of adjacent tree-rings (as opposed to a single one as we have assumed here). Thus, the resulting value $y_i \pm \sigma_i$ paired with $\theta_i^{(d)}$ (usually taken as the calendar age of the middle ring) actually consists of an average of the radiocarbon ages corresponding to the calendar ages of all the tree rings included in the sample. Clearly, this relationship could be modelled and accounted for using an appropriate covariance structure. The international radiocarbon community is currently revising its high-precision dataset, and one of the authors (CEB) is a member of the INTCAL04 team that is charged with identifying the definitive data and estimated calibration curve, and in particular is addressing the aggregated nature of much of the data.

Finally, since there is now a considerable body of data which do not derive from

tree-ring dated timbers, there are a lot of potential calibration data that we have not considered at all in the present paper. Other sources of calibration data include plant and animal remains trapped in lake sediments, corals from marine deposits and stalactites from caves. Many such samples now meet the criteria for inclusion in the international radiocarbon calibration data set and will form part of the raw data released upon publication of the new internationally-agreed calibration curves (Reimer et al. 2002). Other than the tree-ring dated samples, all such samples have uncertainty associated with their calendar age estimates. This is because the calendar ages derive from chronometric methods such as uranium-series dating which are much less precise than tree-ring dating. As part of the INTCAL04 project, all such sources of error are being identified and quantified so that, in future, we will have substantial, publicly available, unprocessed calibration data with imprecise calendar age estimates. In order to take account of this, the INTCAL04 estimate of the calibration curve will use tailored random walk models (similar to those in GPA, L & O'H) with extensions that take account of a range of sources of uncertainty on both the calendar and radiocarbon scales. Clearly such extensions result in greater analytical complexity and computational load, but once the INTCAL04 work is published and the new data released, it will be worth returning to the smooth calibration curve model presented here and to consider options for more sophisticated curve building in the future. We hope that the methodology presented here may in due course develop into the standard approach for the radiocarbon community.

References

Aitkin, M. (1990). *Science-based Dating in Archaeology*. London: Longman Archaeological Series.

Beck, J., Richards, D., Lawrence, R., Edwards, L., Silverman, B., Smart, P., Donahue, D., Herrera-Osterheld, S., Burr, G., Calsoyas, L., Timothy, A., Jull, T., and Biddulph, D. (2001). "Extremely Large Variations of Atmospheric ^{14}C Concentration During the Last Glacial Period." *Science*, 292: 2453–2458.

Bowman, S. (1990). *Radiocarbon Dating (Interpreting the Past)*. London: British Museum Publications, Ltd.

Bronk Ramsey, C. (1995). "Radiocarbon Calibration and Analysis of Stratigraphy: The OxCal Program." *Radiocarbon*, 37: 425–430.

Buck, C., Cavanagh, W., and Litton, C. (1996). *The Bayesian Approach to Interpreting Archaeological Data*. Chichester: Wiley.

Buck, C., Christen, J., and James, G. (1999). "BCal: An On-line Bayesian Radiocarbon Calibration Tool." *Internet Archaeology*, 7. http://intarch.ac.uk/journals/issue7/buck/.

Gómez Portugal Aguilar, D., Litton, C., and O'Hagan, A. (2002). "Novel Statistical Model for a Piece-wise Linear Radiocarbon Calibration Curve." *Radiocarbon*, 44: 195–212.

Grimmett, G. and Stirzaker, D. (1992). *Probability and Random Processes*. Oxford: Clarendon Press, 2nd edition.

Kennedy, M. and O'Hagan, A. (2001). "Bayesian Calibration of Computer Models (with discussion)." *Journal of the Royal Statistical Society, Series B*, 63: 425–464.

Knox, F. and McFadgen, B. (1997). "Least-Squares Fitting a Smooth Curve to Radiocarbon Calibration Data." *Radiocarbon*, 39: 193–204.

Neal, R. (1999). "Regression and Classification Using Gaussian Process Priors." In Bernardo, J., Berger, J., Dawid, A., and Smith, A. (eds.), *Bayesian Statistics 6*, 475–501. Oxford: University Press.

Oakley, J. (1999). "Bayesian Uncertainty Analysis for Complex Computer Codes." Ph.D. thesis, Department of Probability and Statistics, University of Sheffield, UK.

Oakley, J. and O'Hagan, A. (2004). "Probabilistic Sensitivity Analysis of Complex Models: A Bayesian Approach." *Journal of the Royal Statistical Society, Series B*, 66: 751–769.

O'Hagan, A. (1978). "Curve Fitting and Optimal Design for Prediction (with discussion." *Journal of the Royal Statistical Society, Series B*, 40: 1–42.

— (1992). "Some Bayesian Numerical Analysis." In Bernardo, J., Berger, J., Dawid, A., and Smith, A. (eds.), *Bayesian Statistics 4*, 345–363. Oxford: University Press.

Reimer, P., Hughen, K., Guilderson, T., McCormac, G., Baillie, M., Bard, E., Barratt, P., Beck, J., Buck, C., Damon, P., M., F., Kromer, B., Ramsey, C., Reimer, R., Remmele, S., Southon, J., Stuiver, M., and van der Plicht, J. (2002). "Preliminary Report of the First Workshop of the IntCal04 Radiocarbon Calibration /Comparison Working Group." *Radiocarbon*, 44: 653–661.

Schmidt, A. and O'Hagan, A. (2003). "Bayesian Inference for Non-stationary Spatial Covariance Structure via Spatial Deformations." *Journal of the Royal Statistical Society, Series B*, 65: 745–758.

Stuiver, M. and Reimer, P. (1993). "Extended ^{14}C Data Base and Revised CALIB 3.0 ^{14}C Age Calibration Program." *Radiocarbon*, 35: 215–230.

Stuiver, M., Reimer, P., Bard, E., Beck, W., Burr, G., Hughen, K., Kromer, B., McCormac, G., van der Plicht, J., and Spurk, M. (1998). "INTCAL98 Radiocarbon Age Calibration 24,000–0 Cal BP." *Radiocarbon*, 40: 1041–1083.

van der Plicht, J. (2002). "Calibration of the ^{14}C Time Scale: Towards the Complete Dating Range." *Netherlands Journal of Geosciences*, 81: 85–96.

Acknowledgments

DGPA was sponsored by Consejo Nacional de Ciencia y Tecnología (CONACYT) of Mexico, ref. 115 703. We are grateful to Andrew Millard for his constructive comments on an earlier draft of this paper.