RADIOCARBON, Vol 55, Nr 4, 2013, p 1905–1922

THE BAYESIAN APPROACH TO RADIOCARBON CALIBRATION CURVE ESTIMATION: THE INTCAL13, MARINE13, AND SHCAL13 METHODOLOGIES

M Niu • T J Heaton[1] • P G Blackwell • C E Buck

School of Mathematics & Statistics, University of Sheffield, United Kingdom.

ABSTRACT. This article outlines the Bayesian models and methods used to facilitate construction of the 2013 internationally agreed radiocarbon calibration curves known as IntCal13, Marine13, and SHCal13. The models build on those used for the 2004 and 2009 estimates of the curves and, as in 2009, are implemented using Markov chain Monte Carlo sampling, specifically a Metropolis-within-Gibbs sampler. In addition to the data structures accounted for within the 2004 and 2009 models, the approach outlined here also allows for: the presence of additional uncertainty that the data providers have been unable to quantify; tree-ring data that derive their calendar age from wiggle-matching (in addition to ring counting); varve-counted data that exhibit zero increase in calendar age error between 2 or more consecutive layers; and any data source for which we have dependent calendar age uncertainties.

1. INTRODUCTION

Since 2004, the internationally agreed radiocarbon calibration curves have been estimated using tailor-made Bayesian statistical techniques specifically designed to model the diverse data sets used in its production. Such bespoke modeling is crucial for reliable curve production and is a significant part of what sets the IntCal curves apart from many of their competitors.

The 2004 versions of these curves published as Reimer et al. (2004), Hughen et al. (2004a), and McCormac et al. (2004) used a random walk model to represent the curves, for the first time. The random walk model expresses the prior belief that points on the curve that are close together in calendar age are likely to also be close together in ^{14}C age, and thus leads to one way of incorporating the uncertainty in the calendar ages of the calibration data into a coherent statistical analysis. This methodology is documented in Buck and Blackwell (2004).

When the Northern Hemisphere curves were updated in 2009 (Reimer et al. 2009), the statistical methodology was changed to a full Markov chain Monte Carlo (MCMC) approach, which allowed much greater flexibility in our modeling and a data-adaptive selection of important parameters, including the random walk variance, which was known to have a significant effect on the curve created. Descriptions of the changes made can be found in Blackwell and Buck (2008) and Heaton et al. (2009).

In this article, we describe the extensions to the Bayesian models and MCMC methods in Heaton et al. (2009) that have been added to facilitate the estimation of the 2013 internationally agreed curves described in Reimer et al. (2013) and Hogg et al. (2013a). These additions take 2 main forms: those due to a general change in our modeling approach to the errors in the ^{14}C determinations and those due specifically to new data structures for which the existing method cannot cater. The specific modifications made are implemented in C linked into the standard R software environment (R Core Team 2013) and can be summarized as follows:

- The recognition of potential irreducible error in the data corresponding to sources of error in addition to those measured in the ^{14}C laboratory—for example, the random process by which the ^{14}C initially entered the sample, possible contamination over the time between the record being created and extracted, and additional error sources within the various laboratories through

[1]Corresponding author: t.heaton@sheffield.ac.uk.

https://doi.org/10.2458/azu_js_rc.55.17222

which the sample passes that are not directly measurable. This is the only general change made to our method based on an improved understanding of the nature of the data collection.
- The introduction of a new approach for the representation of calendar age uncertainties within 4 sedimentary data sets that more accurately reflects their structure. These data sets contain age estimates that exhibit covariance, i.e. knowledge of the age at one depth within the record will affect age estimates at other depths. Such covariances are encoded within matrices and require new techniques to be incorporated properly into the curve-building procedure.
- Handling of data sets that are varve counted but where the accumulating counting errors do not increase continuously from 1 varve to the next but can instead have sections where no counting errors occur.
- Development of a method to allow inclusion of wiggle-matched tree-ring data sets that have an exactly known internal chronology but whose absolute age is not known. Often these data sets are described as floating and need to be matched to the calibration curve at a suitable point.

As detailed in Blackwell and Buck (2008) and Heaton et al. (2009), we use a Metropolis-within-Gibbs algorithm to implement a Bayesian statistical approach by sampling values from the joint posterior distribution of all unknown quantities. This enables inference for our models despite their being too complex for analytic calculations. The Metropolis-within-Gibbs structure means that we split the unknowns into blocks (which may be as small as a single variable) and, at each iteration, sample each block conditional on the values of the others. In particular, it makes it relatively easy to extend the algorithm, as we do here, simply by adding a new sampling step for each type of variable added to the model. As we describe the details of each of the extensions to the modeling, therefore, we also describe the additional steps in this algorithm needed to accommodate them.

The layout of our paper is as follows. In Section 2, we define our main notation. Section 3 then presents our main change to the general modeling approach to represent the potential irreducible errors within each data set. Next, in Section 4 we outline the specific challenges provided by the new data structures; Section 4.2 covers data sets with generally covarying age estimates, the modifications to include sections of varve-counted data in which the error does not increase are discussed in Section 4.3, while the floating tree rings are dealt with in Section 4.4. We describe the prior distributions actually used for the 2013 calibration curves in Section 5. We then discuss some additional computational considerations in our model implementation including the choice of grid on which to report the curve; the need to split curve estimation into 2 time sections—the first, and most recent, based only on tree rings and the second separate older section from other archive types; and finally assessing convergence of our Markov chain and hence the reliability of our posterior curve estimate. Details on all of these can be found in Section 6. Section 7 describes the method used to construct the Southern Hemisphere curve SHCal13. Finally, we discuss potential further work in Section 8 along with what we believe will be some of the future modeling challenges in calibration curve estimation.

2. NOTATION

2.1 Setting Out of Variables

We generally follow the notation of Blackwell and Buck (2008) and Heaton et al. (2009), restated here for convenience. Additional notation relating to the structural changes to the modeling is defined as it is introduced in Sections 3–4.

As in Heaton et al. (2009), we have a set $\mathcal{I}$ of calibration data (the IntCal database) that provides us with a set of observed variables:

https://doi.org/10.2458/azu_js_rc.55.17222

- $\mathcal{T} = \{T_i\}_{i \in \mathcal{I}}$ — the estimated calendar dates of the data. If the calibration datum is blocked (i.e. derives from material that dates to an average over several successive calendar years), this value refers to the most recent (start) date of the block.
- $\mathcal{X} = \{X_i\}_{i \in \mathcal{I}}$ — the estimated ^{14}C determinations of the (possibly blocked) data.

We also have the following parameters and unknown quantities in the model:

- $\mu(\theta)$ — the calibration curve, as a function of calibrated age θ. The curve is modeled as a random walk, as outlined in Section 1. This random walk model encodes the calibration curve as almost everywhere continuous with independent increments so that between any 2 ages s and t cal yr BP (with $t > s$), the change in the calibration curve $\mu(t) - \mu(s) \sim N(\beta(t-s), r^2(t-s))$.
- $\Theta^{Data} = \{\theta_i\}_{i \in \mathcal{I}}$ — the true calendar dates of the data. Again, if the calibration datum is blocked, this will refer to the most recent date of the block.
- $\mathcal{M}^{Data} = \{\mu_i\}_{i \in \mathcal{I}}$ — the values of the calibration curve at the dates Θ^{Data}. Based upon our random walk prior on the calibration curve μ, given any calendar dates Θ^{Data}, the calibration curve values $\mathcal{M}^{Data}$ will be multivariate normal. This property is crucial to the application of our model.
- Ω — the wiggle-match error in any data that possesses such an error structure. Such error occurs when the ages of a group of data are reliant upon estimation of a shared time shift. Typically, the age estimate of the youngest datum is matched to a particular calendar age and then ages of the other data are estimated relative to this initial datum. See Heaton et al. (2009) for more details.
- r^2 — the variance of the Gaussian random walk that models the calibration curve.
- β — the drift of the Gaussian random walk, as above.

Throughout this paper, we also use notation $\mathcal{K} \subset \mathcal{I}$ to represent a specific data set within the calibration database. For example, $\mathcal{T}_{\mathcal{K}} = \{T_i\}_{i \in \mathcal{K}}$ corresponds to all the estimated calendar ages of the data in set $\mathcal{K}$. Occasionally, we will need to exclude those data in a set $\mathcal{B}$, say, from the set $\mathcal{A}$ being considered; we write this as $\mathcal{A} \backslash \mathcal{B}$ (the set-theoretic difference or relative complement). If it is clear from the context what the set $\mathcal{A}$ is, this can be written more concisely; for example, the notation $\mathcal{T}_{(\mathcal{K})}$ represents $\mathcal{T}_{\mathcal{I} \backslash \mathcal{K}} = \{T_i\}_{i \notin \mathcal{K}}$, that is, all the estimated calendar ages *excluding* those in set $\mathcal{K}$.

At some points in the analysis, rather than thinking of the model for the calibration curve as a random walk, it is more convenient to regard it as a continuous time stochastic process known as a Wiener process (or Brownian motion); if we consider the part of the process between 2 dates at which its values are known (hypothetically), this defines a new process known as a Brownian bridge (see e.g. Taylor and Karlin 1998 for more precise definitions and further explanation).

We use this terminology as necessary in defining the extensions to the Markov chain Monte Carlo algorithm in Section 4. Selecting this random walk prior with its resultant multivariate normal distributions and Brownian bridges is essential for the success of our method and so we wish to elucidate it here. Priors on the other variables do not impact on the setting out of the general methodology, thus we leave their discussion until Section 5.

2.2 Notational Inconsistencies within Calibration

It is important to note the inconsistency in the naming convention of the calibration curve that exists within ^{14}C calibration. In this paper, we denote by $\mu(\theta)$ the single true calibration curve. With this notation, the ^{14}C determination of any object with age θ_i is modeled as

$$X_i \sim N(\mu(\theta_i), \sigma_i^2)$$

https://doi.org/10.2458/azu_js_rc.55.17222

where σ_i^2 is the measurement variance. The true curve $\mu(\theta)$ is still unknown, but, through our Bayesian modeling, we are able to provide a posterior distribution for its value at any chosen time θ.

However, in much other calibration work, this uncertain value is not what is termed the calibration curve. Instead, these authors consider the calibration curve to be the posterior mean of this distribution but without any change of notation. This posterior mean is no longer uncertain and hence has an associated variance that is incorporated into the modeling of a ^{14}C determination. This inconsistency in notation can cause difficulties in understanding. We would therefore recommend that authors refer to the true uncertain calibration curve as $\mu(\theta)$, while the point estimate provided by the posterior mean should be referred to as $\hat{\mu}(\theta)$.

3. CHANGES TO THE GENERAL MODEL TO ALLOW FOR UNQUANTIFIED UNCERTAINTIES

3.1 Evidence for Further Uncertainty in Non-Tree-Ring Radiocarbon Determinations

3.1.1 Model Checking via Residuals

Due to a lack of tree rings extending throughout the entire range for which we wish to perform ^{14}C calibration, estimation of the older parts of all the calibration curves is undertaken using other forms of data. In 2004 and 2009, for all $i \in \mathcal{I}$, ^{14}C determinations were modeled as

$$X_i \sim N(\mu(\theta_i), \sigma_i^2)$$

where $\mu(\theta_i)$ is the calibration curve at calibrated age θ_i and σ_i^2 is the laboratory-reported uncertainty on the ^{14}C age estimate. It is possible to check this model assumption by calculating the residuals, i.e. the difference between each observed ^{14}C determination X_i and its underlying value $\mu(\theta_i)$. According to the above model, these should be distributed $N(0, \sigma_i^2)$.

For data from the non-tree-ring-based part of the curve, these residuals showed significantly heavier tails than suggested by a normal distribution with a variance of σ_i^2. In addition, as reported in Heaton et al. (2009), the posterior distribution for the random walk variance resultant from the non-tree data (if the only errors in their ^{14}C determinations were the σ_i^2 reported by the laboratory-supported values of r^2) was not consistent with that given by the tree-ring data. Both of these features suggest the possibility that the ^{14}C determinations of the older data contain more uncertainty than solely the σ_i^2 reported by the data providers when considered as observations of their true ^{14}C ages. This potential further uncertainty could be due to sources of error that are outside of the laboratory's control but still needs to be recognized in the modeling.

3.1.2 Justification for the Presence of Irreducible Error

In the older part of the curve, the data can exhibit considerable uncertainty in their ^{14}C determinations, which the data providers do their best to quantify and manage. There are, however, further sources of uncertainty that should be accounted for but which are not readily quantifiable and have not formally been recognized in previous estimates of the curve. These can be thought of as ^{14}C age uncertainties that occur for a range of reasons relating to sample preservation, selection, and handling prior to arrival at the ^{14}C laboratory. Given the many ways in which such uncertainties might be introduced, it is unlikely that data providers will ever be able to quantify (or even describe) all of them reliably, but it is still possible to introduce an extra term into the model to account for them. We refer to this term as the irreducible error term.

https://doi.org/10.2458/azu_js_rc.55.17222

Examples of such error sources might include the fact that the samples come from different data sets, are of different types, have experienced different preservation conditions, have been selected and packaged by different people, have been subsampled in different ways (both by the data providers and on arrival at the ^{14}C laboratory), and have reached the ^{14}C laboratories in different conditions. For a discussion of sources of such errors, see Christen (1994) and Bowman (1990:227–8).

As a result, for the 2013 curve estimates, for samples other than tree rings, we select a new model for the ^{14}C determinations X_i, allowing for a potential increase in the error so that

$$X_i \sim N(\mu(\theta_i),\ \xi_i^2)$$

where each $\xi_i \geq \sigma_i$ to represent all the combined errors (measured and irreducible) and is chosen in a data-adaptive manner within our MCMC sampler.

3.2 Choice of Irreducible Error Model

There are several possibilities for modeling the further uncertainty both in terms of what form the ξ's should take, and also whether the data should be grouped into sets that share the same further potential sources of error—associated for example with the particular ^{14}C laboratory, the geographic source of the sample, the type of material dated, or data set number within the IntCal database.

There has been previous work in the field of calibration on incorporating sources of irreducible uncertainty and the related problem of modeling outliers in ^{14}C determinations. Christen (1994) introduces a "shift outlier" model whereby a ^{14}C determination is believed, with a certain probability, to have been contaminated by an unknown shift from its true ^{14}C age; otherwise, the laboratory-reported uncertainty is the only source of uncertainty. After correcting Christen's prior on the shift from an impermissible vague prior to a normal prior, this corresponds to an additive model whereby the identified "outliers" have uncertainty $\xi_i^2 = \sigma_i^2 + \tau^2$. Here, τ^2 represents the uncertainty on the shift size caused by the contamination. This outlier work has been applied with alternative shift priors by Bronk Ramsey (2009), who also proposes a version where the shift occurs in the timescale as opposed to ^{14}C scale. An alternative approach, which as opposed to specifically modeling a proportion of the data as outliers instead considers all the data to possess some further error above that reported by the laboratory, can be found in Christen and Pérez (2009). These authors consider a multiplicative irreducible error model whereby the reported laboratory uncertainty is inflated by an unknown factor so that $\xi_i^2 = \alpha\sigma_i^2$ for some unknown error multiplier α.

We considered both the "shift outlier" and multiplicative error models but did not feel either accurately represented the true nature of the irreducible error-creating process. Firstly, we did not feel that only a proportion of the data would have experienced the additional error processes but instead that all data would possess irreducible error, albeit potentially small. Secondly, we did not desire a model whereby the size of the irreducible error should be proportional to the error as measured in the laboratory. Such a multiplicative model does not seem to be appropriate if it is believed that a significant factor to the irreducible error is not controllable within the laboratory but relates to the conditions in which the sample was preserved.

Instead, it was decided by the IntCal Working Group (at the Belfast Group Meeting in 2010) to use an additive model similar to the outlier work of Christen (1994) but modeling all data as having irreducible error with potential size varying according to data set. As such, those data that had been preserved together and may have experienced the same processes would be considered to have the same potential for irreducible error to have been introduced. However, if the data suggested that set had

https://doi.org/10.2458/azu_js_rc.55.17222

not experienced such additional error it could be estimated as small. This decision was taken after consideration of several other possible error groupings using the same model checking as discussed in Section 3.1.1 alongside discussion with data providers as to the form of the believed error-creating processes. With the introduction of an additive model for irreducible error and a grouping by data set, the model residuals showed a significantly improved goodness-of-fit over the version assuming the presence of solely laboratory-reported error in the ^{14}C determinations.

3.3 The Additive Model

For any non-tree-ring data set $\mathcal{K} \in \mathcal{I}$, we consider

$$X_i \sim N(\mu(\theta_i),\ \tau_{\mathcal{K}}^2 + \sigma_i^2)$$

where, in a slight abuse of notation, $\tau_{\mathcal{K}}^2$ represents the single unquantified potential uncertainty for all points in set $\mathcal{K}$. For ease of understanding, this model can alternatively be written as

$$X_i = \mu(\theta_i) + \varepsilon_i + \zeta_i$$

where

$$\varepsilon_i \sim N(0,\ \sigma_i^2) \text{ and } \zeta_i \sim N(0,\ \tau_{\mathcal{K}}^2)$$

This perhaps shows more clearly that for each sample in the IntCal13 database there exists an error ε_i corresponding to the laboratory quantified uncertainty and also a further error ζ_i representing the additional effect of the unquantifiable error-creating processes. Note that while these errors ζ_i are independent of one another, within any individual data set $\mathcal{K}$ all observations share the same potential for additional error represented by the same value of $\tau_{\mathcal{K}}$. This represents our belief that the data within each set $\mathcal{K}$ will have experienced similar shared processes. Note also that the size of this additional error is not dependent upon the individual σ_i's.

3.4 Incorporating Irreducible Error into the Sampler

We work with the variable $\tau_{\mathcal{K}}$ to simplify updating in light of the constraint that the variance of the additive irreducible error must be positive; by updating $\tau_{\mathcal{K}}$, we can consider any $\tau_{\mathcal{K}} \in \mathbb{R}$. We then wish to update $\tau_{\mathcal{K}}$ via a separate step within our full curve sampler. We have

$$p(\tau'_{\mathcal{K}} \mid \{X_i, \mu(\theta_i)\}_{i \in \mathcal{K}}) \propto \prod_{i \in \mathcal{K}} \left[\frac{1}{\sqrt{\sigma_i^2 + \tau_{\mathcal{K}}^2}} \exp\left\{ -\frac{(x_i - \mu_i)^2}{2(\sigma_i^2 + \tau_{\mathcal{K}}^2)} \right\} \right] \pi(\tau_{\mathcal{K}})$$

Here $\pi(\tau_{\mathcal{K}})$ represents our prior on the value of $\tau_{\mathcal{K}}$. A Gibbs algorithm is not feasible since the normalizing constant above is not easily found. As a result, we take a Metropolis-Hastings approach as follows:

1. Propose new $\tau'_{\mathcal{K}} \sim N(\tau_{\mathcal{K}}, \psi^2)$.
2. Find acceptance probability

$$\nu(\tau_{\mathcal{K}}, \tau'_{\mathcal{K}}) = \min\left[\frac{p(\tau'_{\mathcal{K}} \mid \{X_i, \mu(\theta_i)\}_{i \in \mathcal{K}})}{p(\tau_{\mathcal{K}} \mid \{X_i, \mu(\theta_i)\}_{i \in \mathcal{K}})}, 1 \right]$$

where $p(\tau_{\mathcal{K}} \mid \{X_i, \mu(\theta_i)\}_{i \in \mathcal{K}})$ is as described above and then accept or reject appropriately.

https://doi.org/10.2458/azu_js_rc.55.17222

There is little constraint on the form of the prior distribution for $\tau_{\mathcal{K}}$. However, we are seldom likely to be able to elicit a very informative prior since this irreducible error is, by its nature, difficult for data providers to quantify. Hence, for estimation of IntCal13 and Marine13, we felt it appropriate to be cautious and adopt an improper, uniform prior for $\pi(\tau_{\mathcal{K}})$. In practice, the amount of data within each group $\mathcal{K}$ was sufficient to be informative despite this prior and, as the posteriors show, the estimated size of these errors are highly variable dependent upon data set. Use of such an improper prior allows us to drop the $\pi(\tau)$ terms from the calculations. For completeness, in our proposal distribution we set $\psi^2 = 20^2$.

3.5 Implications for Calibration Users

By adding this extra parameter to the model and the MCMC sampler, we can allow for and quantify the irreducible errors associated with the data sets that make up the IntCal database. For IntCal13 and Marine13, we found that the posterior means for $\tau_{\mathcal{K}}$ varied from close to zero (with associated posterior standard deviations of the order of 10 yr) to several hundred years (with associated posterior standard deviations of the order of 100 yr).

Users of the 2013 curves should note the range of the irreducible errors associated with the curve estimates, since the samples used to provide the data in the IntCal database are a high-quality subset of the types of samples that users are likely to be working with themselves. Users should thus consider carefully whether there are possible, previously unquantified errors associated with their own samples, seek to quantify any that they can, and consider allowing for these in the calibration process.

4. CHANGES TO ACCOMMODATE NEW DATA STRUCTURES

4.1 Introduction to New Data Structures

In addition to the changes made to modeling the ^{14}C errors discussed in the previous section, several other additions to the 2009 methodology were all necessitated by features of the new data sets that make up the IntCal13 database (documented in Reimer et al. 2013). Specifically, these differences relate to tree-ring data that derive their calendar age not just from ring counting but also from wiggle-matching, varve-counted data that exhibit zero increase in calendar age error between 2 or more consecutive layers, and data sources for which we have dependent calendar age uncertainties.

The changes required to accommodate the wiggle-matched tree-ring data of Hua et al. (2009) are primarily implementational since appropriate models and close-to appropriate algorithms were discussed in Heaton et al. (2009). The difference for IntCal13 and Marine13 is that we have added code to allow for data where the only source of calendar error is a wiggle-match. In 2009, all of the data sets that exhibited wiggle-match errors also exhibited some additional calendar age uncertainty (other than that derived from the wiggle-match error). This is not present in any of the tree-ring data, and so we have used a simplified version of the 2009 methodology as documented in Section 4.4.

As in 2009, data sets that derive their calendar ages from varve counting form a key part of the IntCal13 database. What we did not do in 2009, however, was to discuss the special case where varve-counted data sets exhibit zero increase in calendar age error between 2 or more consecutive layers. This case is addressed in Section 4.3.

The most significant change to the statistical modeling approach used for IntCal13 and Marine13 (over that used for IntCal09 and Marine09) was precipitated by the adoption of new age-depth modeling approaches (Bronk Ramsey and Lee 2013; Heaton et al. 2013) by some of the data providers. Data sets that derive their calendar age estimates from such methods—i.e. those relating to non-

varved Cariaco (Hughen et al. 2006), to the Pakistan and Iberian margins (Bard et al. 2013, this issue), and to Suigetsu (Bronk Ramsey et al. 2012)—are not provided, as they would have been in 2004 and 2009, simply in the form of a vector of calendar age estimates ($\mathcal{T}_\mathcal{K}$) and independent uncertainties ($\varepsilon_\mathcal{K}$). Instead, within each data set, the calendar ages covary and so the uncertainties are conveyed via a covariance matrix. This new data structure requires an appropriate model to be formed and implemented that enables us to update several points on the calibration curve simultaneously. We discuss this in the next section.

4.2 Dependent Errors from Age-Depth Models

For any data set $\mathcal{K}$ that exhibits dependent calendar age uncertainties, this information is conveyed in the form of a covariance matrix $Z_\mathcal{K}$. The diagonal elements of the matrix are variances on the estimates of the calendar age and the off-diagonal elements are the covariances between pairs of calendar ages. Four of the data sets in the 2013 IntCal database have such dependence within their calendar age estimates. Representing the vectors of true and estimated calendar ages from the relevant data set as $\Theta_\mathcal{K}$ and $\mathcal{T}_\mathcal{K}$, respectively, we model these age estimates via a multivariate normal distribution as $\mathcal{T}_\mathcal{K} \sim MVN(\Theta_\mathcal{K}, Z_\mathcal{K})$.

4.2.1 Updating Calibration Curve Parameters ($\Theta_\mathcal{K}$, $\mathcal{M}_\mathcal{K}|\mathcal{T}, \mathcal{X}, \Omega, r^2$)

A Metropolis-within-Gibbs Approach

In principle, such a data structure need not affect our Metropolis-within-Gibbs sampler; we could follow Blackwell and Buck (2008) and Heaton et al. (2009) and update each individual pair $(\mu_j, \theta_j)_{j\in\mathcal{K}}$ in the following way.

a) Propose new θ'_j from a symmetric distribution centered at current value θ_j.

b) Accept with probability $\min\{HR, 1\}$ where

$$HR = \frac{p(x_j|\theta'_j, \mathcal{M}_{(j)}, \Theta_{(j)}, \Omega, r, \beta)p(\theta'_j|\mathcal{T}, \mathcal{M}_{(j)}, \Theta_{(j)}, \Omega, r)}{p(x_j|\theta_j, \mathcal{M}_{(j)}, \Theta_{(j)}, \Omega, r, \beta)p(\theta_j|\mathcal{T}, \mathcal{M}_{(j)}, \Theta_{(j)}, \Omega, r)}$$

is the Hastings ratio and where $\mathcal{Y}_{(i)}$ denotes a set $\mathcal{Y}$ with element y_i removed.

c) If accepted, sample new μ_j from distribution $\mu_j|\theta'_j, \mathcal{M}_{(j)}, \Theta_{(j)}, \mathcal{X}, \Omega, r, \beta$.

The only change required here would be that $p(\theta_j|\mathcal{T}, \mathcal{M}_{(j)}, \theta_{(j)}, \Omega, r)$ would now depend upon all the other calendar age information within data set $\mathcal{K}$, i.e. $\Theta_{\mathcal{K}\setminus j}$ and $\mathcal{T}_{\mathcal{K}\setminus j}$. This is different from the case where there is no covariance and it would simply depend upon the individual value T_j, i.e. the other data do not provide any information about the errors.

However, on implementation, we found that the level of covariance within our data sets was at such a level that this approach did not allow our Markov chain to mix well. This was due to the restrictive nature of $p(\theta_j|\mathcal{T}, \mathcal{M}_{(j)}, \Theta_{(j)}, \Omega, r)$ for those points j that had a lot of correlation with other points in the data set, so that, in reality, they could only move together. As a consequence, it was not possible for the calendar ages to move sufficiently freely to fully explore the possible range of values they might take. To overcome this issue, it was necessary to update multiple pairs (μ_j, θ_j) simultaneously, requiring finer consideration of the Brownian bridges formed in construction of $\mu_j|\theta_j, \mathcal{M}_{(j)}, \Theta_{(j)}, \mathcal{X}, \Omega, r, \beta$.

4.2.2 Updating Multiple Points on the Curve Simultaneously

To allow our sampler to mix, we updated all the pairs (μ_j, θ_j) within each covarying data set currently under consideration (i.e. non-varved Cariaco, Pakistan Margin, Iberian Margin, or Suigetsu) simultaneously. To do this requires alterations to our algorithm. Suppose that we wish to update a complete set of pairs $\{(\mu_j, \theta_j)\}_{j\in\mathcal{K}}$ simultaneously where $\mathcal{K}$ denotes a particular data set. We would then perform the following:

a) Propose new multidimensional $\Theta'_{\mathcal{K}}$ from $MVN(\Theta_{\mathcal{K}}, Z_{\mathcal{K}})$ distribution. Here, $Z_{\mathcal{K}}$ is the covariance matrix for the observed $\mathcal{T}_{\mathcal{K}}$.

b) Accept with probability $\min\{HR, 1\}$ where

$$\begin{aligned} HR &= \frac{p(\mathcal{X}_{\mathcal{K}}|\Theta'_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta)p(\Theta'_{\mathcal{K}}|\mathcal{T}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r)}{p(\mathcal{X}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta)p(\Theta_{\mathcal{K}}|\mathcal{T}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r)} \\ &= \frac{p(\mathcal{X}_{\mathcal{K}}|\Theta'_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta)p(\Theta'_{\mathcal{K}}|\mathcal{T}_{\mathcal{K}})}{p(\mathcal{X}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta)p(\Theta_{\mathcal{K}}|\mathcal{T}_{\mathcal{K}})} \end{aligned}$$

c) If accepted, sample new $\mathcal{M}_{\mathcal{K}}$ from distribution $\mathcal{M}_{\mathcal{K}}|\Theta'_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \mathcal{X}, \Omega, r, \beta$.

This requires 2 modifications to our method. Firstly, we sample from a multidimensional proposal density for $\Theta'_{\mathcal{K}}$ and calculate the multidimensional density $p(\Theta'_{\mathcal{K}}|\mathcal{T}_{\mathcal{K}})$ for the acceptance ratio. Secondly, we must consider the distribution of $\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \mathcal{X}, \Omega, r, \beta$, which is now also multidimensional. Exactly as in the single-pair update of Heaton et al. (2009), these latter calculations are used in finding the Hastings ratio as well as the final sampling step.

In order to determine this latter density, as in Heaton et al. (2009), we first write

$$f(\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \mathcal{X}, \Omega, r, \beta) \propto \pi(\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta)f(\mathcal{X}_{\mathcal{K}}|\mathcal{M}_{\mathcal{K}}) \qquad (1)$$

and consider each term on the RHS separately. This can be seen as first creating a prior for $\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta$ and then updating in light of the observed data $\mathcal{X}_{\mathcal{K}}$.

To calculate the prior, we also still use the properties of a Wiener process but are now required to form a series of independent Brownian bridges between all of those data points considered known for this update step, i.e. $\mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}$. This is illustrated in Figure 1. We partition the sampled data set $\Theta_{\mathcal{K}}$ values according to which interval (and hence Brownian bridge) they lie in. Two values that lie in the same interval will have μ's with covariance, while those that lie in different intervals will be independent conditional on the observed data. Suppose, as shown in Figure 1, we have 2 values $\theta_{\mathcal{K}_2}$ and $\theta_{\mathcal{K}_3}$, which lie in the same interval bounded above and below by the known pairs (θ^B, μ^B) and (θ^A, μ^A), respectively. Then, by the properties of the Brownian bridge, the joint distribution of $\mu_{\mathcal{K}_2}$ and $\mu_{\mathcal{K}_3}$ will be multivariate normal so that

$$E[\mu_{\mathcal{K}_i}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta] = \mu^B + (\mu^A - \mu^B)\frac{\theta_{\mathcal{K}_i} - \theta_B}{\theta_A - \theta_B},$$

$$Var[\mu_{\mathcal{K}_i}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta] = r^2\frac{(\theta_A - \theta_{\mathcal{K}_i})(\theta_{\mathcal{K}_i} - \theta_B)}{\theta_A - \theta_B}$$

$$Cov[\mu_{\mathcal{K}_i}, \mu_{\mathcal{K}_j}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta] = r^2\frac{(\theta_A - \theta_{\mathcal{K}_j})(\theta_{\mathcal{K}_i} - \theta_B)}{\theta_A - \theta_B} \text{ if } \theta_{\mathcal{K}_i} < \theta_{\mathcal{K}_j}.$$

https://doi.org/10.2458/azu_js_rc.55.17222

However, the μ's at these points will be conditionally independent from the value at any point $\theta_{\mathcal{K}_k}$, which does not lie in the same Brownian bridge. Hence, in Figure 1, the distribution of the μ's at $\theta_{\mathcal{K}_2}$ and $\theta_{\mathcal{K}_3}$ will exhibit covariance but will be conditionally independent of the values at $\theta_{\mathcal{K}_1}$ and $\theta_{\mathcal{K}_4}$. We can extend this argument to the entire data set $\mathcal{K}$ in order to generate the complete prior distribution of $\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta$ as a multivariate normal:

$$\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta \sim MVN(\mu_{RW}, \Sigma_{RW})$$

where the covariance matrix Σ_{RW} will be of a block diagonal form with the blocks corresponding to the partitioning of $\Theta_{\mathcal{K}}$ into its independent Brownian bridges given the known $\mathcal{M}_{(\mathcal{K})}$ and $\Theta_{(\mathcal{K})}$.

Now to calculate the second term on the RHS of Equation 1, we note that $\mathcal{X}_{\mathcal{K}}|\mathcal{M}_{\mathcal{K}} \sim MVN(\mathcal{M}_{\mathcal{K}}, \Sigma_{\mathcal{X}_{\mathcal{K}}})$ where $\Sigma_{\mathcal{X}_{\mathcal{K}}}$ is the covariance matrix for the ^{14}C data set $\mathcal{K}$. This matrix is diagonal since we assume that the ^{14}C determinations are independent. Putting these 2 terms together and using conjugacy, we find the posterior for $\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \mathcal{X}, \Omega, r, \beta$ to be

$$MVN((\Sigma_{RW}^{-1} + \Sigma_{\mathcal{X}_{\mathcal{K}}}^{-1}) \times (\Sigma_{RW}^{-1}\mu_{RW} + \Sigma_{\mathcal{X}}^{-1}\mathcal{X}_{\mathcal{K}}), (\Sigma_{RW}^{-1} + \Sigma_{\mathcal{X}_{\mathcal{K}}}^{-1})^{-1}).$$

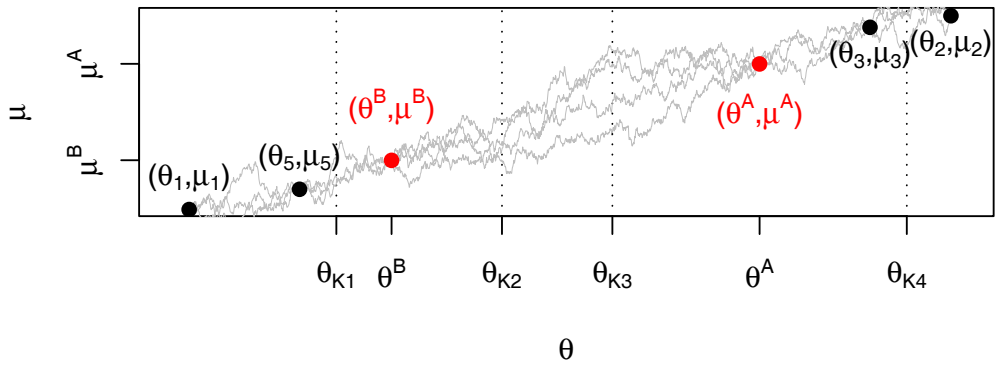

Figure 1 Updating multiple pairs with a Brownian bridge. This plot illustrates a single step of our sample to update the values $\mu_{\mathcal{K}}$ corresponding to the calendar ages $\theta_{\mathcal{K}}$ shown with the dotted vertical lines. Within this step, we condition on knowing all the other pairs (θ_i, μ_i) shown by the solid dots. To perform the update, we form independent Brownian bridges between each of these known pairs. Any calendar ages $\theta_{\mathcal{K}}$ lying in the same interval will have μ's with covariance while those lying in different intervals will be independent. As such, the distribution for $\mu_{\mathcal{K}}|\theta_{\mathcal{K}}, \mathcal{M}_{(\mathcal{K})}, \Theta_{(\mathcal{K})}, \Omega, r, \beta$ will be multivariate normal with a covariance matrix that is blocked in nature. In this illustration, the calendar ages $\theta_{\mathcal{K}_2}$ and $\theta_{\mathcal{K}_3}$ both lie within the same bridge bounded by the pairs (θ^A, μ^A) and (θ^B, μ^B) shown in red and so will have distributions for their μ's that covary. Conversely, $\theta_{\mathcal{K}_1}$ and $\theta_{\mathcal{K}_4}$ lie in different bridges and so will be conditionally independent.

4.3 Dependent Errors from Varve Counting

As outlined above, in the 2013 IntCal database, some consecutive layers in the varve-counted portion of the Cariaco record (Hughen et al. 2000, 2004b,c) share the same varve-count errors (i.e. no extra uncertainty accumulates within the subset of samples in question). Continuing to represent any single data set within the IntCal database as $\mathcal{K}$, if the data set in question exhibits varve-count errors, we assume it can be subdivided into $N_{\mathcal{K}}$ disjoint subsets, $\mathcal{K}_k$, $k = 1, 2,..., N_{\mathcal{K}}$, each of which contains consecutive observations and within any one of which all samples share the same true accumulated varve-count error, $\phi_{\mathcal{K}_k}$.

https://doi.org/10.2458/azu_js_rc.55.17222

In what follows, we use $\#\mathcal{K}$ and $\#\mathcal{K}_k$ to represent the number of samples in data set $\mathcal{K}$ and subset $\mathcal{K}_k$, respectively. Note that we only require $\#\mathcal{K}_k \geq 1$ and so the situation where each varve accumulates extra uncertainty over those that have gone before can simply be seen as a special case in which all $\#\mathcal{K}_k = 1$ and $N_{\mathcal{K}} = \#\mathcal{K}$.

To ease computation and notation, we take advantage of the fact that in a varved sequence the chronological order of the samples is known *a priori* and thus assume that the elements of $\mathcal{K}$ are ordered such that $\mathcal{K}_1$ is the youngest and $\mathcal{K}_{\#\mathcal{K}}$ is the oldest and then, within $\mathcal{K}_k$, the youngest sample is $\mathcal{K}_{k_1}$.

The estimated date of the lth element of $\mathcal{K}_k$ then has calendar age

$$\mathcal{T}_{\mathcal{K}_{k_l}} = \Theta_{\mathcal{K}_{k_l}} + \phi_{\mathcal{K}_k}, \text{ for } l = 1, 2, ..., \#\mathcal{K}_k$$

where

$$\phi_{\mathcal{K}_k} = \sum_{c=1}^{k} \varepsilon_{\mathcal{K}_c}, \text{ and } \varepsilon_{\mathcal{K}_c} \sim N(0, \sigma^2_{\mathcal{K}_c}).$$

Here, $\sigma^2_{\mathcal{K}_c}$ represents the varve-count error for subset $\mathcal{K}_c$ as recorded in the IntCal database. In other words, the error within subset $\mathcal{K}_k$ is the cumulative sum of all the varve-counting errors made in moving between all the previous subsets up to the set in question. i.e. between $\mathcal{K}_1$, $\mathcal{K}_2$,..., $\mathcal{K}_k$.

4.3.1 Updating $(\Theta_{\mathcal{K}}, \mathcal{M}_{\mathcal{K}})$ in the Presence of Accumulated Varve-Count Error $\phi_{\mathcal{K}}$

Since the varve-count error increases only as we move from one subset $\mathcal{K}_k$ to the next $\mathcal{K}_{k+1}$, if we know the true calendar age for the first datum in any subset $\mathcal{K}_k$, then we also know the calendar age of any of the other samples in subset $\mathcal{K}_k$ and the value of $\phi_{\mathcal{K}_k}$. As a consequence, when we update any individual varve-counting error $\Theta_{\mathcal{K}_k}$, this corresponds to updating the whole set of values $\phi_{\mathcal{K}_k}$ together. Therefore, within our MCMC sampler, the easiest way to update a varve-counted data set is to consider $\mathcal{M}_{\mathcal{K}_k}$ and $\Theta_{\mathcal{K}_k}$ for each subset k and update all the values in each simultaneously using a very similar approach to that outlined in Section 4.2.1. The algorithm for this update step becomes the following:

a) Propose new $\Theta'_{\mathcal{K}_k}$ by shifting the complete set together by a value $v_k \sim N(0, \varsigma^2)$, i.e. so that $\Theta'_{\mathcal{K}_{k_l}} = \Theta_{\mathcal{K}_{k_l}} + v_k$ for $l = 1,..., \#\mathcal{K}_k$. For the 2013 updates of the calibration curves, we took $\varsigma^2 = \sigma^2_{\mathcal{K}_k}$.

b) Accept with probability $\min\{HR, 1\}$ where

$$HR = \frac{p(\mathcal{X}_{\mathcal{K}_k} | \Theta'_{\mathcal{K}_k}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r, \beta) p(\Theta'_{\mathcal{K}_k} | \mathcal{T}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r)}{p(\mathcal{X}_{\mathcal{K}_k} | \Theta_{\mathcal{K}_k}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r, \beta) p(\Theta_{\mathcal{K}_k} | \mathcal{T}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r)}$$

c) If accepted, sample new $\mathcal{M}_{\mathcal{K}_k}$ from distribution $\mathcal{M}_{\mathcal{K}_k} | \Theta'_{\mathcal{K}_k}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \mathcal{X}, \Omega, r, \beta$.

All of these steps have already been explained except in this case the calculation of the term $p(\Theta_{\mathcal{K}_k} | \mathcal{T}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r)$ in the Hastings ratio is somewhat different and worthy of further explanation. With a uniform prior for $\Theta_{\mathcal{K}_k} | \mathcal{T}_{(\mathcal{K}_k)}, \mathcal{M}_{\mathcal{K}_k}, \Theta_{(\mathcal{K}_k)}, \Omega, r$ subject to the required known ordering constraints on the calendar ages (i.e. that $\Theta_{\mathcal{K}_k}$ does not overlap with $\Theta_{\mathcal{K}_{k-1}}$ or $\Theta_{\mathcal{K}_{k+1}}$), we write

https://doi.org/10.2458/azu_js_rc.55.17222

$$\begin{aligned} p(\Theta_{\mathcal{K}_k}|\mathcal{T}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r) &\propto p(\mathcal{T}_{\mathcal{K}_k}|\Theta_{\mathcal{K}_k}, \mathcal{T}_{(\mathcal{K}_k)}, \mathcal{M}_{(\mathcal{K}_k)}, \Theta_{(\mathcal{K}_k)}, \Omega, r) \\ &= p\phi_{\mathcal{K}_k}|\phi_{\mathcal{K}_1}, \ldots, \phi_{\mathcal{K}_{k-1}}, \phi_{\mathcal{K}_{k+1}}, \ldots, \phi_{\mathcal{K}_{N_\mathcal{K}}} \\ &= p\phi_{\mathcal{K}_k}|\phi_{\mathcal{K}_{k-1}}, \phi_{\mathcal{K}_{k+1}} \\ &= p(\varepsilon_{\mathcal{K}_k}|(\varepsilon_{\mathcal{K}_k} + \varepsilon_{\mathcal{K}_{k+1}})) \end{aligned}$$

using the facts that $\mathcal{T}_{\mathcal{K}_i}$ and $\Theta_{\mathcal{K}_i}$ inform us of the value of $\phi_{\mathcal{K}_i}$, and that this $\phi_{\mathcal{K}_i}$ is conditionally independent of all other $\phi_{\mathcal{K}_j}$ given $\phi_{\mathcal{K}_{i-1}}$ and $\phi_{\mathcal{K}_{i+1}}$. Now since each $\varepsilon_{\mathcal{K}_i} \sim N(0, \sigma^2_{\mathcal{K}_i})$ we can find this density since

$$\varepsilon_{\mathcal{K}_k}|\varepsilon_{\mathcal{K}_k} + \varepsilon_{\mathcal{K}_{k+1}} = a \sim N\left(\frac{\sigma^2_{\mathcal{K}_k} a}{\sigma^2_{\mathcal{K}_k} + \sigma^2_{\mathcal{K}_{k+1}}}, \frac{\sigma^2_{\mathcal{K}_k}\sigma^2_{\mathcal{K}_{k+1}}}{\sigma^2_{\mathcal{K}_k} + \sigma^2_{\mathcal{K}_{k+1}}}\right).$$

4.4 Dependent Errors from Wiggle-Matching

Our final consideration is the updating of a new wiggle-matched data set within the tree-ring-based part of the IntCal database (Hua et al. 2009). The updating of the calendar ages for this data set is done slightly differently to those modifications already described in that we choose to work directly with the wiggle-match error Ω as a parameter in our MCMC model. Given this value, all the calendar ages $\Theta_\mathcal{K}$ within this data set are known and so the corresponding $\mathcal{M}_\mathcal{K}$ can then be updated as for the tree rings with a simple Gibbs only step.

The wiggle-match for this data set required specific updated coding since, unlike the instance described in Heaton et al. (2009), it is not combined with any other type of error. Also, due to the splitting of the curve into 2 distinct sections described in Section 6, this data set is older in age than all other data in its section. which simplifies greatly the updating process.

4.4.1 Updating ($\Omega|\mathcal{M}, \mathcal{X}, \Theta_{(\mathcal{K})}, \mathcal{T}, r$)

To update $\Omega|\mathcal{M}, \mathcal{X}, \Theta_{(\mathcal{K})}, \mathcal{T}, r$ within our sampler, we perform the following:

a) Propose new $\Omega' \sim N(\Omega, \eta^2)$ where, for the 2013 curve estimates, $\eta = 10$.

b) Accept with probability $\min\{HR, 1\}$ where

$$HR = \frac{p(\Omega'|\mathcal{M}, \mathcal{X}, \Theta_{(\mathcal{K})}, \mathcal{T}, r)}{p(\Omega|\mathcal{M}, \mathcal{X}, \Theta_{(\mathcal{K})}, \mathcal{T}, r)}.$$

We can initially simplify this if we note that

$$\begin{aligned} p(\Omega|\mathcal{M}, \mathcal{X}, \Theta_{(\mathcal{K})}, \mathcal{T}, r) &\propto \pi(\Omega)p(\mathcal{M}_\mathcal{K}|\Omega, \mathcal{M}_{(\mathcal{K})}, \mathcal{X}, \Theta_{(\mathcal{K})}, \mathcal{T}, r) \\ &\propto \pi(\Omega)p(\mathcal{X}_\mathcal{K}|\mathcal{M}_\mathcal{K})p(\mathcal{M}_\mathcal{K}|\Theta_\mathcal{K}, \Theta_{(\mathcal{K})}, \mathcal{M}_{(\mathcal{K})}, r) \end{aligned}$$

using the fact that proposing a value Ω is equivalent to proposing a value for $\Theta_\mathcal{K}$. The $p(\mathcal{X}_\mathcal{K}|\mathcal{M}_\mathcal{K})$ terms do not depend upon Ω and so will cancel in the HR calculation.

This already reduces the calculation of the Hastings ratio somewhat, but we can go much further in our implementation since, as explained, we create the curve in 2 sections and this data set is known

https://doi.org/10.2458/azu_js_rc.55.17222

to be older than all of the other data in its section, i.e. all $\Theta_{\mathcal{K}}$ are greater than $\Theta_{(\mathcal{K})}$. As such, when we are required to find the likelihood of $p(\mathcal{M}_{\mathcal{K}}|\Theta_{\mathcal{K}}, \Theta_{(\mathcal{K})}, \mathcal{M}_{(\mathcal{K})}, r)$ it is the single continuation of a Wiener process as opposed to a series of Brownian bridges. Consider this full likelihood in terms of the conditional likelihood of each $\mu_{\mathcal{K}_i}$ sequentially:

$$p(\mathcal{M}_{\mathcal{K}}|\Theta'_{\mathcal{K}}, \Theta_{(\mathcal{K})}, \mathcal{M}_{(\mathcal{K})}, r) = p(\mu_{\mathcal{K}_1}|\theta_{\mathcal{K}}, \theta_{(\mathcal{K})}, \mu_{\mathcal{K}}, r) \times p(\mu_{\mathcal{K}_2}|\theta_{\mathcal{K}}, \theta_{(\mathcal{K})}, \mu_{\mathcal{K}_1}, r) \times \ldots$$
$$\ldots \times p(\mu_{\mathcal{K}_n}|\theta_{\mathcal{K}}, \theta_{(\mathcal{K})}, \mu_{\mathcal{K}_1}, \ldots, \mu_{\mathcal{K}_{n-1}}, r).$$

Now, as a continuation of a Wiener process, any $p(\mu_{\mathcal{K}_i}|\theta_{\mathcal{K}}, \theta_{(\mathcal{K})}, \mu_{\mathcal{K}_1}, \ldots, \mu_{\mathcal{K}_{i-1}}, r)$ will depend only upon 3 values: the random walk variance r^2; the value of the curve $\mu^B_{\mathcal{K}_i}$ at the earlier neighboring time $\mu^B_{\mathcal{K}_i}$; and $\mu^B_{\mathcal{K}_i} - \theta_{\mathcal{K}_i}$, the time elapsed between then and the new point.

For all but the most recent sample in the wiggle-matched data set, these likelihoods will not depend upon Ω since the internal relative differences $\mu^B_{\mathcal{K}_i} - \theta_{\mathcal{K}_i}$ within this data set are known exactly. Hence, the only term in the likelihood that does depend upon the value of Ω is the single value $\mu_{\mathcal{K}_i}$, corresponding to the most recent calendar age in the set. We can therefore write the Hastings ratio as

$$HR = \frac{\pi(\Omega')p(\mu_{\mathcal{K}_1}|\theta'_{\mathcal{K}_1}, \theta^B_{\mathcal{K}_1}, \mu^B_{\mathcal{K}_1}, r)}{\pi(\Omega)p(\mu_{\mathcal{K}_1}|\theta_{\mathcal{K}_1}, \theta^B_{\mathcal{K}_1}, \mu^B_{\mathcal{K}_1}, r)}$$

where $\theta^B_{\mathcal{K}_i}$ is the oldest calendar age in $\Theta_{(\mathcal{K})}$ and $\mu^B_{\mathcal{K}_1}$ is the corresponding value of the calibration curve at that point. Here, we have a $N(0, \rho^2)$ prior on $\Omega_{\mathcal{K}}$ as specified by the data providers and the likelihood $p(\mu_{\mathcal{K}_1}|\theta'_{\mathcal{K}_1}, \theta^B_{\mathcal{K}_1}, \mu^B_{\mathcal{K}_1}, r)$ is also a normal density formed from the Wiener process.

5. PRIORS DISTRIBUTION AND VALUES

We summarize here the prior distributions used for the parameters of the model, or where applicable the fixed values assumed.

- $\mu(\theta)$ — the calibration curve at calibrated age θ. The random walk model can be thought of as defining the prior distribution for $\mu(\cdot)$, given values of r^2 and β.
- $\Theta^{Data} = \{\theta_i\}_{i \in \mathcal{I}}$ — the true calendar dates of the data. We assume that little is known *a priori* about the calendar date of each sample and thus use independent improper uniform prior distributions, except in cases where the ordering of observations is known, in which case we also condition on that ordering.
- $\mathcal{M}^{Data} = \{\mu_i\}_{i \in \mathcal{I}}$ — the values of the calibration curve at the dates Θ^{Data}. The prior on $\mu(\theta)$ is a random walk, as explained in Sections 1 and 2, inducing a multivariate normal prior on $\mathcal{M}^{Data}$ conditional on Θ^{Data}.
- r^2 — variance of the random walk. As discussed in Blackwell and Buck (2008), our prior knowledge about this parameter is informative; here we use the same Inverse Gamma prior for the tree-ring-based estimate of the 2013 curves that we used in 2009 (Heaton et al. 2009). As detailed in Reimer et al. (2013), for the non-tree-based estimate of the 2013 curves, this parameter is fixed at the posterior mean derived from the tree-based estimate, i.e. $r^2 = 55$.
- β — the drift of the random walk. While in principle this can be treated as an unknown random variable, here (as in 2004 and 2009) we treat it as fixed with a value of 1. As explained in Buck and Blackwell (2004), this is based upon the expectation that the value of the calibration curve changes by approximately 1 ^{14}C yr per calendar year. The authors of that paper also found that

https://doi.org/10.2458/azu_js_rc.55.17222

the resultant curve is very insensitive to the specific value of β used, a more precise value based upon current knowledge of ^{14}C decay rates making very little difference to the curve.
- $\tau_{\mathcal{K}}$ — the standard deviation of the irreducible error on a particular data set $\mathcal{K}$. We use an improper, uniform prior here.
- Ω — wiggle-match error. The prior for this is normally distributed with a mean of zero and a variance supplied by the data providers.

6. ADDITIONAL CONSIDERATIONS

In this section, we document some additional considerations relating to the implementation of the curve estimation for IntCal13, Marine13, and SHCal13, all three of which have been completely recalculated using the updated methodology outlined here. Unless otherwise stated, these implementation details are additions to those documented in Heaton et al. (2009).

6.1 Choice of Grid

As in 2009, for the purposes of reporting, a grid was selected on the calibrated age scale at which posterior estimates of the mean and standard deviation of the calibration curve are provided. For the 2013 curves, however, the grid intervals are somewhat different, as follows. For IntCal13, the grid on the calendar scale is at intervals of 5 yr in the range 0 to 13.90 cal kyr BP, 10 yr in the range 13.91 to 25.00 cal kyr BP, and 20 yr in the range 25.02 to 50.00 cal kyr BP. For Marine13, the grid is at intervals of 5 yr in the range 0 to 10.495 cal kyr BP, 10 yr in the range 10.5 to 25.0 cal kyr BP, and 20 yr in the range 25.02 to 50.00 cal kyr BP. For SHCal13, the grid is at the same ages as IntCal13, as explained in Section 7.

6.2 Splitting the Curve

As with IntCal09, IntCal13 was created in 2 distinct sections. For 2013, however, we have more tree-ring data and so the tree-ring-based section now extends from 0–13.9 cal kyr BP and the older section, from 13.91–50.00 cal kyr BP, was constructed using data from other archive types (as documented in Reimer et al. 2013). Exploratory analysis indicated significant differences between the features of these 2 sources of data and suggested it was inappropriate to model them identically. In 2009, it was argued that because the data from non-tree sources exhibit considerably increased variability from that suggested by the tree-ring data, the curves should be created for each section independently, assuming different and uncertain variances in the 2 parts, subject to the condition of continuity at the juncture. However, since 2009, we have had access to tree-ring data from beyond the end of the trees that currently make up the IntCal database, in particular, the Fuji and ND113 data sets from van der Plicht et al. (2012) and a Late Glacial Huon pine data set from Hogg et al. (2013a). We do not have reliable calendar age estimates for these trees, but we do know that they are older than the trees in the IntCal database and so they provide potentially valuable information about the variation in ^{14}C levels in the Earth's atmosphere before the Holocene. Investigation of the underlying variation suggested by these data did not provide any conclusive evidence of variances as large as that obtained as part of the 2009 curve construction.

As a result, for IntCal13 and Marine13, we used only tree-ring data from the IntCal database to learn about the random walk variance in the following way. When estimating the earlier (tree-ring) part of IntCal13, the random walk variance, r^2, was allowed to vary, thus allowing us to obtain a posterior estimate for it (via the same approach used in 2009). The resulting posterior mean, of 55, was then used as a fixed input when estimating the older parts of both IntCal13 and Marine13.

https://doi.org/10.2458/azu_js_rc.55.17222

6.3 Convergence and Pointwise Summaries

As in 2009, quantitative assessment of convergence was difficult due to the large number of variables that are updated in the sampler. Instead, we used a heuristic approach whereby, having let the sampler run for a burn-in period, independent subsamples of the later output were selected and the corresponding calibration curve estimate for each was plotted. No visible differences were observed, thus suggesting that the Markov chain had reached and adequately explored its equilibrium distribution.

The 0–13.9 cal kyr BP (tree-based) segment of IntCal13 was estimated using an MCMC run of 500,000 iterations with the first 5000 iterations treated as burn-in and then saved every 10th iteration so that the final estimates were based on 49,500 joint samples from the posterior. For the 13.91–50.00 cal kyr BP (non-tree) segment of IntCal13, for which a substantial proportion of the data points exhibit dependent errors, an MCMC run of 10 million iterations was used, with the first 100,000 treated as burn-in and then (because the chains exhibit very marked autocorrelation and slow convergence) thinned, retaining every 100th iteration, so that the final parameter estimates are based on 99,000 iterations. The 10.51–50.00 cal kyr BP segment of Marine13 (for which autocorrelation is less of a problem because the Suigetsu data set is not used) was estimated using an MCMC run of 5,000,000 iterations of which the first 50,000 were treated as burn-in and then the remainder were thinned down to every 50th iteration, so that the final estimates were based on 99,000 iterations.

Histograms of the pointwise posterior distributions for $\mu(\theta)$ were plotted at a range of ages. The resultant density estimates were unimodal and relatively symmetric indicating they could be reasonably approximated by the normal distribution with appropriate mean and variance presented in IntCal13 and ShCal13. These pointwise summaries do not, however, utilize the covariance information available, which is an avenue for further work as discussed in Section 8.

7. CALCULATING SHCal13

The construction of SHCal is heavily dependent on IntCal, but needs to take full advantage of the limited amount of true Southern Hemisphere (SH) data available (Hogg et al. 2013a). To combine the two, we use a simple statistical model of the dynamics of the North–South (NS) offset between IntCal and SHCal, as follows:

a) Firstly, we analyze the available SH tree-ring data (covering approximately 0–1000 and 12,100–12,700 cal BP) detailed in Hogg et al. (2013a) using the same method as described above for the recent part of IntCal13. This gives an initial estimate of the SH curve, albeit one that is very imprecise over most of the required range.

b) Secondly, we treat the differences between this SH estimate and IntCal13, at the "grid points" (see Section 6.1) on which IntCal13 is defined within the range of the SH data, as observations of the NS offset. This observed offset behaves as might be expected given previous work (McCormac et al. 2004; Hogg et al. 2011; see also discussion in Hogg et al. 2013b), with a typical value of about 43 yr and substantial autocorrelated variability around that. We capture these features by modeling the offset as a stationary first-order autoregressive or AR(1) process (e.g. Brockwell and Davis 2002) with a 5-yr time-step and with parameters, including the mean level, to be estimated from the observations. Estimation is carried out using standard software (R Core Team 2013) using maximum likelihood, which we interpret as Bayesian maximum *a posteriori* (MAP) estimation with an uninformative prior distribution.

c) Thirdly, we use the predictions of the fitted AR(1) model to estimate the offset for those dates (on the same grid as IntCal13) for which we do not have a direct SH estimate. This means that at dates far from

https://doi.org/10.2458/azu_js_rc.55.17222

the SH data, this estimated offset is constant, and is simply given by the estimated long-run mean level of 43 yr; but at dates close—on the timescale defined by the estimated autocorrelation—to the SH data, the predicted offset is influenced by the "local" observations on the offset as well as the "global" parameters, with smaller uncertainty as a consequence.

d) Finally, we use these elements to construct SHCal13. Within the range of the actual SH data, we use the estimates based purely on those data. Elsewhere, we take SHCal13 to be the sum of IntCal13 and the estimated offset at that date; except as noted below, its variance is calculated as the sum of the variance on IntCal13 and the variance on the offset (based on the prediction error of the AR(1) model, using the MAP point estimate of its parameters).

The simple modeling of the dynamic offset, as described above, gives a smooth transition between the direct and indirect estimates of SHCal. However, it overestimates the variance immediately outside the range of the true SH data, even with the offset modeled appropriately, since over a short range—at 12,065 BP and for about 50 yr beyond 12,680 BP—the variance as calculated in (d) would exceed that of the SH-only version in (a). Combining these 2 variances formally would require additional assumptions about the correlation structure; however, experimentation under a range of different plausible assumptions shows that the variances obtained are rather robust, and are close to those obtained by linear interpolation over the ranges identified, so the actual variances reported are those from that linear interpolation.

More detailed modeling of the dynamics of the offset, taking into account more of our understanding of the relevant physical processes, was not feasible for SHCal13, since the analysis could not begin until IntCal13 was known, but is an important aim for future SH updates.

8. FUTURE WORK

There are several areas that we would like to investigate further. Firstly, our MCMC methodology allows a set of complete posterior realizations of the curves to be created at every desired time point. Currently, these complete realizations are reduced to pointwise summaries of individual means and variances on a chosen grid for publication. Such a pointwise approach loses a great deal of information about the curves and the covariance information between values at different times that is contained in the realizations themselves. This covariance information has the potential to affect calibration if one is interested in anything other than the dating of a single ^{14}C determination—for example, it may alter estimation of the time elapsed between multiple determinations or the wiggle-matching of a set of ^{14}C determinations.

Blackwell and Buck (2008) provide an example of the noticeable difference in posterior distributions that can be obtained if this covariance information is incorporated into the estimation of the time elapsed between 2 ^{14}C determinations; Millard (2008) shows that the magnitude of the effect can vary greatly. Investigation of the impact this covariance information can have on a wider range of practical calibration problems using the current internationally agreed calibration curves would be a interesting avenue for further investigation.

An initial obstacle to this goal is the current approach that splits the curve creation into 2 distinct sections—that based upon tree-ring data up to 13.9 cal kyr BP, and that based on data from other archive types that extends up to 50 cal kyr BP. This means that, while we can access covariance information in each section separately, we are not able to record covariance information across the join and hence the whole time range. To fully use curve realization rather than pointwise summaries, our estimation procedure would first need to be extended to create the entire curve from 0–50 cal kyr BP simultaneously.

https://doi.org/10.2458/azu_js_rc.55.17222

A second avenue worthy of consideration relates to the inclusion of further data types within the calibration database as more data sets are considered to be of calibration quality. In the view of the authors, one area of particular interest is the addition of data sets (particularly those of terrestrial origin) that have only very weak, perhaps only relative, prior information on their calendar ages (such as those from large-scale archaeological dating projects or from pre-Holocene floating tree-ring sequences, for example). While such data sets may not provide much information about the calibration curve on their own, when combined with other data sets possessing more age information, their relative stratigraphic order may be able to add finer scale detail to the calibration curves.

ACKNOWLEDGMENTS

The authors wish to thank the UK's Natural Environmental Research Council who funded the work described in this paper via grant number NE/E019129/1.

REFERENCES

Bard E, Ménot G, Rostek F, Licari L, Böning P, Edwards RL, Cheng H, Wang YJ, Heaton TJ. 2013. Radiocarbon calibration/comparison records based on marine sediments from the Pakistan and Iberian margins. *Radiocarbon* 55(4), this issue.

Blackwell PG, Buck CE. 2008. Estimating radiocarbon calibration curves. *Bayesian Analysis* 3(2):225–48.

Bowman S. 1990. *Radiocarbon Dating*. London: British Museum Publications.

Brockwell P, Davis R. 2002. *Introduction to Time Series and Forecasting*. 2nd edition. New York: Springer.

Bronk Ramsey C. 2009. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* 51(3):1023–45.

Bronk Ramsey C, Lee S. 2013. Recent and planned developments of the program OxCal. *Radiocarbon* 55(2–3):720–30.

Bronk Ramsey C, Staff RA, Bryant CL, Brock F, Kitagawa H, van der Plicht J, Schlolaut G, Marshall MH, Brauer A, Lamb HF, Payne RL, Tarasov PE, Haraguchi T, Gotanda K, Yonenobu H, Yokoyama Y, Tada R, Nakagawa T. 2012. A complete terrestrial radiocarbon record for 11.2 to 52.8 kyr B.P. *Science* 338(6105): 370–4.

Buck CE, Blackwell PG. 2004. Formal statistical models for estimating radiocarbon calibration curves. *Radiocarbon* 46(3):1093–102.

Christen JA. 1994. Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics* 43(3):489–503.

Christen JA, Pérez S. 2009. A new robust statistical model for radiocarbon data. *Radiocarbon* 51(3): 1047–59.

Heaton TJ, Blackwell PG, Buck CE. 2009. A Bayesian approach to the estimation of radiocarbon calibration curves: the IntCal09 methodology. *Radiocarbon* 51(4):1151–64.

Heaton TJ, Bard E, Hughen K. 2013. Elastic tie-pointing—transferring chronologies between records via a Gaussian process. *Radiocarbon* 55(4), this issue.

Hogg A, Palmer J, Boswijk G, Turney C. 2011. High-precision radiocarbon measurements of tree-ring dated wood from New Zealand: 195 BC–AD 995. *Radiocarbon* 53(3):529–42.

Hogg AG, Hua Q, Blackwell PG, Niu M, Buck CE, Guilderson TP, Heaton TJ, Palmer JG, Reimer PJ, Reimer RW, Turney CSM, Zimmerman SRH. 2013a. SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon* 55(4), this issue.

Hogg A, Turney C, Palmer J, Southon J, Kromer B, Bronk Ramsey C, Boswijk G, Fenwick P, Noronha A, Staff R, Friedrich M, Reynard L, Guetter D, Wacker L, Jones R. 2013b. The New Zealand kauri (*Agathis australis*) research project: a radiocarbon dating intercomparison of Younger Dryas wood and implications for IntCal13. *Radiocarbon* 55(4), this issue.

Hua Q, Barbetti M, Fink D, Kaiser K, Friedrich M, Kromer B, Levchenko V, Zoppi U, Smith A, Bertuch F. 2009. Atmospheric ^{14}C variations derived from tree rings during the early Younger Dryas. *Quaternary Science Reviews* 28(25–26):2982–990.

Hughen KA, Southon JR, Lehman SJ, Overpeck JT. 2000. Synchronous radiocarbon and climate shifts during the last deglaciation. *Science* 290(5498):1951–4.

Hughen KA, Baillie MGL, Bard E, Beck JW, Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Kromer B, McCormac G, Manning S, Bronk Ramsey C, Reimer PJ, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE. 2004a. Marine04 marine radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1059–86.

Hughen KA, Lehman SJ, Southon JR, Overpeck J, Marchal O, Herring C, Turnbull J. 2004b. ^{14}C activity and global carbon cycle changes over the past 50,000 years. *Science* 303(5655):202–7.

Hughen KA, Southon JR, Bertrand CJH, Frantz B, Zermeño P. 2004c. Cariaco Basin calibration update: revisions to calendar and ^{14}C chronologies for core

PL07-58PC. *Radiocarbon* 46(3):1161–87.

Hughen K, Southon J, Lehman S, Bertrand C, Turnbull J. 2006. Marine-derived ^{14}C calibration and activity record for the past 50,000 years updated from the Cariaco Basin. *Quaternary Science Reviews* 25(23–24): 3216–27.

McCormac FG, Hogg AG, Blackwell PG, Buck CE, Higham TFG, Reimer PJ. 2004. SHCal04 Southern Hemisphere calibration, 0–11.0 cal kyr BP. *Radiocarbon* 46(3):1087–92.

Millard AR. 2008. Comment on article by Blackwell and Buck, "Estimating radiocarbon calibration curves." *Bayesian Analysis* 3(2):255–61.

R Core Team. 2013. R: A Language and Environment for Statistical Computing. Vienna: R Foundation for Statistical Computing. URL: http://www.R-project.org.

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

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Burr GS, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kaiser KF, Kromer B, McCormac FG, Manning SW, Reimer RW, Richards DA, Southon JR, Talamo S, Turney CSM, van der Plicht J, Weyhenmeyer CE. 2009. IntCal09 and Marine09 radiocarbon age calibration curves, 0–50,000 years cal BP. *Radiocarbon* 51(4): 1111–50.

Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Haflidason H, Hajdas I, Hatté C, Heaton TJ, Hoffman DL, Hogg AG, Hughen KA, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney CSM, van der Plicht J. 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55(4), this issue.

Taylor H, Karlin S. 1998. *An Introduction to Stochastic Modeling*. 3rd edition. Oxford: Academic Press.

van der Plicht J, Imamura M, Sakamoto M. 2012. Dating of Late Pleistocene tree-ring series from Japan. *Radiocarbon* 54(3–4):625–33.