

© 2020 by the Arizona Board of Regents on behalf of the University of Arizona. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

## THE INTCAL20 APPROACH TO RADIOCARBON CALIBRATION CURVE CONSTRUCTION: A NEW METHODOLOGY USING BAYESIAN SPLINES AND ERRORS-IN-VARIABLES

Timothy J Heaton<sup>1\*</sup>  • Maarten Blaauw<sup>2</sup>  • Paul G Blackwell<sup>1</sup>  • Christopher Bronk Ramsey<sup>3</sup>  • Paula J Reimer<sup>2</sup>  • E Marian Scott<sup>4</sup> 

<sup>1</sup>School of Mathematics and Statistics, University of Sheffield, Sheffield S3 7RH, UK

<sup>2</sup>The <sup>14</sup>CHRONO Centre for Climate, the Environment and Chronology, School of Natural and Built Environment, Queen's University Belfast, Belfast BT7 1NN, UK

<sup>3</sup>Research Laboratory for Archaeology and the History of Art, University of Oxford, 1 South Parks Road, Oxford OX1 3TG, UK

<sup>4</sup>School of Mathematics and Statistics, University of Glasgow, Glasgow G12 8QS, UK

**ABSTRACT.** To create a reliable radiocarbon calibration curve, one needs not only high-quality data but also a robust statistical methodology. The unique aspects of much of the calibration data provide considerable modeling challenges and require a made-to-measure approach to curve construction that accurately represents and adapts to these individualities, bringing the data together into a single curve. For IntCal20, the statistical methodology has undergone a complete redesign, from the random walk used in IntCal04, IntCal09 and IntCal13, to an approach based upon Bayesian splines with errors-in-variables. The new spline approach is still fitted using Markov Chain Monte Carlo (MCMC) but offers considerable advantages over the previous random walk, including faster and more reliable curve construction together with greatly increased flexibility and detail in modeling choices. This paper describes the new methodology together with the tailored modifications required to integrate the various datasets. For an end-user, the key changes include the recognition and estimation of potential over-dispersion in <sup>14</sup>C determinations, and its consequences on calibration which we address through the provision of predictive intervals on the curve; improvements to the modeling of rapid <sup>14</sup>C excursions and reservoir ages/dead carbon fractions; and modifications made to, hopefully, ensure better mixing of the MCMC which consequently increase confidence in the estimated curve.

### 1 INTRODUCTION

One of the core challenges in creating a reliable radiocarbon (<sup>14</sup>C) calibration curve is ensuring that the statistical approach used for the curve's production is able to accurately synthesize the various datasets which make up the curve, specifically recognizing and incorporating their diverse and unique aspects in doing so. This is particularly critical in light of the recent increase in the availability of high-precision radiocarbon determinations, and the consequent demand by users for ever more precise and accurate calibration. For IntCal20, not only is the underlying data available for use in curve construction considerably more numerous and detailed than in previous curves but, due to the advances in our wider understanding of the Earth's systems, our insight into the specific and individual characteristics of much of that data has also improved. If we are able to harness and more accurately represent these specific features within our curve construction then this will hopefully improve the resultant calibration. To achieve this, for the new calibration curve we have completely revised the statistical methodology from a random walk to a Bayesian spline based approach. This new approach allows us to take better advantage of the underlying data and also to provide output which is more useful and relevant for calibration users.

Over the years, the IntCal statistical methodology has developed and improved alongside our understanding of the constituent data and as we gain more knowledge of the underlying Earth

\*Corresponding author. Email: [t.heaton@shef.ac.uk](mailto:t.heaton@shef.ac.uk).

Systems processes. For IntCal98 (Stuiver et al. 1998), the curve was produced via a combination of linear interpolation of windowed tree-ring averages and frequentist splines. In 2004, the approach was updated to a random walk model (Buck and Blackwell 2004) which introduced an approximate Bayesian method, and began to incorporate some of the important additional aspects of the constituent data such as potential uncertainty in the calendar ages for the older determinations, and that many tree-ring observations related to multiple years of metabolization. This random walk was developed further for IntCal09 (Blackwell and Buck 2008; Heaton et al. 2009) and IntCal13 (Niu et al. 2013) into a fully Bayesian MCMC approach that enabled the inclusion of more of the unique structures in the calibration datasets such as floating sequences of tree rings and more complex covariances in the calendar age estimates provided by wiggle-matching or palaeoclimate tie-pointing.

The random walk approach however had several disadvantages. While in principle, it allowed for modeling flexibility, the details of the implementation required a very large number of parameters that, despite being highly dependent, were predominantly updated individually. This made it extremely slow to run; difficult to ensure it explored the full space of possible curves; and also hard to assess whether the MCMC had reached its equilibrium distribution. Furthermore, this restricted the ability to explore the impact of specific modeling assumptions or individual data on the final curve. As the volume of the data used to generate IntCal has increased, alongside the need for further bespoke modeling of key data structures, this random walk approach to curve creation has consequently become computationally impractical for further use.

For the updated IntCal20 curve, the IntCal working group therefore requested a new approach to curve creation be developed. This new approach needed to be equally rigorous, from a statistical perspective, as the previous methodology; to incorporate the advances in both geoscientific understanding and radiocarbon precision that have occurred since 2013; and yet overcome the implementational difficulties of the previous random walk approach. Specifically, to increase confidence in the curve's robustness and reliability the new approach needed to run at a speed which allowed investigation of the effect of key modeling choices not possible with the previous methodology. The new approach is based upon Bayesian splines with errors-in-variables as introduced by Berry et al. (2002) but requires multiple bespoke innovations to accurately adapt to the specific data complexities. We believe it offers a significant improvement over previous approaches not just in the modeling but also the additional output it can provide for both calibration users and geoscientific users more widely.

The paper is set out as follows. In Section 2 we list the main advances made in the statistical methodology, data understanding and modeling since the last set of IntCal curves (Reimer et al. 2013; Hogg et al. 2013). We then provide, in Section 3, a short non-technical introduction to three key practical ideas for calibration users: an explanation of Bayesian splines; the importance of recognizing potential calendar age uncertainty in curve construction; and the modeling of potential over-dispersion in the data. Section 4 then describes the main technical details of curve construction. The IntCal20 curve is created in two sections with somewhat different statistical concerns. The more recent part of the curve, back to approximately 14 cal kBP, is based entirely upon tree-ring determinations, that are predominantly dendrodated with exact, known calendar ages. Here the main challenges are to incorporate the large amount of data, provide high resolution in the curve

with appropriate uncertainties, and accurately incorporate blocked determinations that represent the measurement of multiple years. This is described in Section 4.3. Further back in time, the curve is based upon a wider range of  $^{14}\text{C}$  material where, as well as tree rings, we use direct and indirect records of atmospheric  $^{14}\text{C}$  in the form of corals, macrofossils, forams and speleothems. We describe the modifications required to incorporate these kinds of data in Section 4.4. This being a Bayesian approach we are able to incorporate prior information into our model and also provide posterior information on particular outputs of independent interest. In Section 5 we provide an indication of the kind of additional output the new methodology is able to provide. In particular the new approach generates not only pointwise means and variances for the calibration curve but, for the first time, sets of complete posterior realizations from 0–55 cal kBP which also allow access to covariance information. This covariance has the potential to improve future calibration of multiple radiocarbon determinations, for example in wiggle matches or more complex modeling. We also gain information on the level of additional variation (beyond laboratory reported uncertainty) present in tree-ring  $^{14}\text{C}$  determinations arising from the same calendar year. Finally, we discuss potential future work in Section 6 along with areas we have identified for further improvement for the next IntCal iteration. Note that the statistical approaches to the creation of SHCal20 (Hogg et al. 2020 in this issue) and Marine20 (Heaton et al. 2020 in this issue) are presented within those papers and not discussed in detail here.

### Notation

All ages in this paper and the database are reported relative to mid-1950 AD (= 0 BP, before present). Conventional, pre-calibration,  $^{14}\text{C}$  ages are given in units “ $^{14}\text{C}$  yrs BP.” Calendar, or calibrated, ages are denoted as “cal BP” or “cal kBP” (thousands of calibrated years before present).

### Data and Code

As in previous updates to the curve, the constituent data is available from the IntCal database <http://www.intcal.org/>. Other inputs are available on request, for example covariance matrices for the calendar ages of the tie-pointed records of the Cariaco Basin (Hughen and Heaton 2020 in this issue), Pakistan and Iberian margin (Bard et al. 2013); and Lake Suigetsu (Bronk Ramsey et al. 2020 in this issue). Coding was performed in R (R Core Team 2019), using the **fd**a (Ramsay et al. 2018), **mvtnorm** (Genz et al. 2019), **mvnfast** (Fasiolo 2016) and **Matrix** (Bates and Maechler 2019) packages to efficiently create and fit the spline bases; and **doParallel** (Corporation and Weston 2019) to implement parallelization for the MCMC tempering. This code is available on request from the first author.

## 2 DEVELOPMENTS IN THE INTCAL20 STATISTICAL METHODOLOGY

The main differences/improvements in the updated IntCal20 methodology over the previous random walk can be split into three broad categories: those relating to improvements in the statistical implementation itself; progress in the detailed modeling of unique data aspects that are enabled by the updated methodology; and finally, advances in the curve output that are both relevant to users of the calibration curve and of potential interest in their own right.

## Improvements in Statistical Implementation

1. Bayesian Splines—the change from the random walk of Niu et al. (2013) to splines allows a much more computationally feasible fitting algorithm. We retain the Bayesian aspect since it allows us to incorporate the various unique aspects of the calibration data; provides useful posterior information, e.g. the posterior calendar ages of the constituent data; and maintains consistency with calibration itself which is now universally implemented under that paradigm. This Bayesian spline approach is equally conceptually rigorous as the random walk but considerably more flexible and can be run much more quickly.
2. Change in modeling and fitting domain—function estimation via splines is based on a trade-off between creating a curve that passes close to the observed data yet is not too *rough*. Previously all aspects of the curve's construction occurred in the radiocarbon age domain. However, this is not the natural space in which to either model the curve roughness or decide on fidelity of the data to the curve. In the older section of the  $^{14}\text{C}$  record, the measurement uncertainties on the radiocarbon age scale become non-symmetric. This causes difficulties in fairly assessing the fit of a model to observed data if we judge it in this radiocarbon age domain. Instead, it is more natural to assess model fit in  $F^{14}\text{C}$  where measurement uncertainty remains symmetric. Similarly, a more natural domain in which to model the curve roughness is in  $\Delta^{14}\text{C}$  space. We therefore change our modeling domain to  $\Delta^{14}\text{C}$  and our fitting domain to  $F^{14}\text{C}$  when creating the curve. See Section 3.1.2 for definitions of the radiocarbon age,  $F^{14}\text{C}$  and  $\Delta^{14}\text{C}$  domains.
3. Over-dispersion in dendrodated trees—as the volume of data entering IntCal increases, it is key to make sure we do not produce a curve which is over-precise as this would give inaccurate calibration for a user. While laboratories attempt to quantify all  $^{14}\text{C}$  uncertainty in their measurements, including through intercomparison exercises such as Scott et al. (2017), there remain some sources of additional  $^{14}\text{C}$  variation which are difficult for any laboratory to capture. For tree rings, potential examples include variation between local region, species, or growing season. Consequently, when we bring together  $^{14}\text{C}$  measurements from the same calendar year, they may potentially be over-dispersed, i.e. more widely spread than would be expected given their laboratory quoted uncertainty. The new approach incorporates a term allowing for potential over-dispersion within  $^{14}\text{C}$  determinations of the tree rings so that, if additional variability is seen in the underlying IntCal data, the method will account for it and prevent excessive confidence in the resultant curve.
4. Heavy tailed errors—in the older portion of the curve, where data come from a range of different  $^{14}\text{C}$  reservoirs, we aim to reduce the influence of potential outliers by permitting each dataset to have heavier tailed errors. These tails are adaptively estimated during curve construction.
5. Improved model mixing and parallel tempering—when using any MCMC method, it is crucial to ensure that the chain has reached its equilibrium distribution and that it is not stuck in one part of the model space. This was a significant concern with the previous random walk approach since the curve was updated one calendar year at a time. Conversely, Bayesian splines enable updates of the complete curve simultaneously via Gibbs sampling. To address any additional concerns about mixing of the new MCMC and to ensure we explore the space of possible curves more freely, we also implement parallel tempering whereby we run multiple modified/tempered chains simultaneously in

such a way that some can move around the space more easily. By appropriate switching between these tempered chains we can further improve model mixing.

### **Improvements in Data Modeling**

These steps forward in the statistical implementation, while still maintaining computational feasibility, also enable us both to incorporate new data structures and to advance our modeling of the processes from which the data arise:

1. Large increase in volume of data throughout the curve—the new IntCal20 is based upon a much greater number of <sup>14</sup>C determinations. These include many annual tree-ring determinations such as remeasurement within various laboratories of data from the period of Thera (e.g. Pearson et al. 2018); new wiggle-matched and floating sequences of late-glacial tree rings (e.g. Capano et al. 2020 in this issue); and new measurements of Hulu Cave extending further back in time (Cheng et al. 2018). In total, the new IntCal20 curve is based upon 12,904 raw <sup>14</sup>C measurements compared to the 7019 used for IntCal13 (Reimer et al. 2013). This rapid increase means a faster curve construction method is essential.
2. Blocking in dendrodated trees—with the new radiocarbon revolution providing an increased ability to measure annual tree rings, there has been a concern that the inclusion within IntCal of determinations relating to decadal or bi-decadal averages alongside these new annual measurements could mean we lose critical information on short-term variation in atmospheric radiocarbon levels. Our approach fully recognizes the number of years each determination represents, meaning there is no loss of information as a result of such blocked determinations.
3. Variable marine reservoir ages—IntCal09 (Reimer et al. 2009) and IntCal13 (Reimer et al. 2013) modeled marine reservoir ages beyond the Holocene as constant over time. For IntCal20, we incorporate time-varying marine reservoir ages using the LSG model of Butzin et al. (2020 in this issue) and a separate adaptive spline for the Cariaco Basin.
4. Inclusion of new floating tree-ring sequences—the new curve includes several new late-glacial trees which have calendar age estimates obtained via wiggle matching (Reinig et al. 2018; Capano et al. 2020 in this issue), as well as three entirely floating tree chronologies around the time of the Bølling-Allerød<sup>1</sup> (Adolphi et al. 2017) and two older Southern Hemisphere floating kauri tree-ring sequences (Turney et al. 2010, 2017) on which we have no absolute dating information and which need to be placed accurately amongst the other data.
5. Rapid <sup>14</sup>C excursions—there are several specific times when the level of atmospheric <sup>14</sup>C is known to vary extremely rapidly. For IntCal20 we have identified three such events around 774–5 AD, 993–4 AD and ~660 BC (Miyake et al. 2012, 2013; O'Hare et al. 2019). Such rapid <sup>14</sup>C changes will typically not be modeled well by standard regression which will tend to smooth them out. However, by increasing the density of knots forming our spline basis in the vicinity of these rapid excursions we can better represent these significant features.

---

<sup>1</sup>While Adolphi et al. (2017) do provide potential age estimates for their Bølling-Allerød trees obtained by comparison to ice-core <sup>10</sup>Be, to maintain independence between the ice-core and radiocarbon timescales these calendar age estimates are not used for IntCal20.

### Improvements in Output

The statistical innovations also mean that we can now provide several new facets to our output:

1. Annualized output—we are able to provide curve estimates on an annual grid enabling more detailed calibration. This will be needed in light of the increased demand to calibrate annual radiocarbon determinations. While we do not discuss the implications for calibration itself in this paper, such a detailed annual calibration curve will likely increase the potential for multimodal calendar age estimates which require significant care in interpretation, especially during periods of plateaus in the calibration curve. See the IntCal20 companion paper (van der Plicht et al. 2020 in this issue) for more details and an illustrative example with the calibration of the Minoan Santorini/Thera eruption.
2. Predictive intervals—if the IntCal20  $^{14}\text{C}$  data contain additional sources of variation beyond their laboratory quantified uncertainties (due to potential regional, species or growing season differences), i.e. they appear more widely spread around the calibration curve than can be explained by their quoted uncertainties, we will need to take this into account for calibration of new determinations too. Specifically, any  $^{14}\text{C}$  determination a user calibrates against the curve is likely to contain similar levels of *unseen* additional variation. If we can assess the level of additional variation seen within the IntCal20 data, then we can incorporate this into calibration for a user by providing a predictive interval on the curve. Intuitively this predictive interval aims to incorporate both uncertainty in the value of the curve itself and potential additional uncertainty sources such as regional, species or growing season effects beyond the laboratory quoted uncertainty. These predictive intervals are therefore more relevant for calibration than curve intervals which do not incorporate or adapt to potential additional sources of variability.
3. Posterior information arising from curve construction—the Bayesian implementation allows us to provide posterior estimates for many aspects of interest. For example, all of the records with uncertainties in their calendar timescales (the various floating tree-ring sequences, marine sediments, Lake Suigetsu and the speleothems) will be calibrated during the curve's construction. We can provide these posterior calibrated age estimates along with other information such as the level of over-dispersion seen in the data, and posterior estimates of marine reservoir ages and dead carbon fractions.
4. Complete realizations of the curve from 0–55 cal kBP—historically IntCal output has consisted of pointwise posterior means and variances. However, the Bayesian approach provides a set of underlying curve realizations which provide covariance information on the value of the curve at any two calendar ages. When calibrating single determinations, this covariance does not affect the calendar age estimates obtained, i.e. calibrating against the pointwise posterior means and variances provides the same inference as calibrating against the set of individual curve realizations. However, when calibrating multiple determinations simultaneously, for example if we are seeking to determine the length of time elapsed between two determinations or fit a more complex model, use of these complete realizations in consequent calibration offers potential to improve insight. Work is planned by the group to explore how this may be best incorporated into existing calibration software.

We discuss some of these aspects in more detail, and present some of the output available, in Section 5 and the Supplementary Information.

### 3 A BRIEF SUMMARY OF BAYESIAN SPLINES, ERRORS-IN-VARIABLES AND PREDICTIVE INTERVALS

For a typical calibration user the key elements of the new methodology consist of the change to Bayesian splines with the accompanying shift in the modeling and fitting domains; the continued recognition that many of the data have uncertainties in their calendar ages and the significant effect that has on curve estimation; and the use of predictive intervals on the published curve for calibration. We therefore commence with an intuitive explanation of these three elements. The technical details can be found later in Section 4. We note that, in our intuitive descriptions, the observational model may be further complicated by  $^{14}\text{C}$  determinations representing multiple, as opposed to single, calendar years and the inclusion of a reservoir age or dead carbon fraction for those observations which are not directly atmospheric. However, for clarity of exposition, we do not consider either of these factors here, and refer to Section 4 for details on how these additions can be incorporated.

#### 3.1 Bayesian Splines and Choice of Modeling Domain

##### 3.1.1 Frequentist Ideas

Suppose that we observe a function  $f$ , subject to noise, at a series of times  $\theta_i$ ,  $i = 1, \dots, n$ ,

$$Y_i = f(\theta_i) + \varepsilon_i \quad i = 1, \dots, n,$$

where, for the time being, we assume that the times  $\theta_i$  are known absolutely. To obtain a spline estimate for the unknown function we seek to find a function that provides a satisfactory compromise between going close to the data but yet does not overfit. This can be done by choosing the estimate  $\hat{f}$  that minimizes, over a suitable set of functions,

$$S(f) = \text{FIT}(f, Y) + \lambda \text{PEN}(f),$$

where  $\text{FIT}(f, Y)$  measures the lack of agreement between a potential  $f$  and the observed  $Y$ ; and  $\text{PEN}(f)$  represents a penalty for functions that might be overfitting. Typically,  $\text{FIT}(f, Y)$  consists of the sum-of-squares difference between  $f(\theta_i)$  and  $Y_i$ , while  $\text{PEN}(f)$  assesses the roughness of a proposed  $f$  ensuring that the more variable  $f$  the larger the penalty given to it with the aim of preventing the spline estimate from overfitting the data. The parameter  $\lambda$  determines the relative trade-off between how one values fidelity to the observed data, i.e.  $\text{FIT}(f, Y)$ , compared with function roughness,  $\text{PEN}(f)$ . A large  $\lambda$  will heavily penalize roughness and typically results in a smooth curve that is less close to the data; while a small  $\lambda$  will mean the spline goes closer to the data but at the expense of being more variable.

##### 3.1.2 Selection of Appropriate Fitting and Modeling Domains for Radiocarbon

Within the radiocarbon community there are three commonly used domains<sup>2</sup>:  $\Delta^{14}\text{C}$ ,  $F^{14}\text{C}$  and the radiocarbon age. Given  $g(\theta)$ , the historical level of  $\Delta^{14}\text{C}$  in year  $\theta$  cal BP, we can freely convert between the domains as follows:

- $F^{14}\text{C}$  domain:  $f(\theta) = \left(\frac{1}{1000}g(\theta) + 1\right)e^{-\theta/8267}$ . The value of  $F^{14}\text{C}$ , or fraction modern (Reimer et al. 2004b), denotes the relative proportion of radiocarbon to stable carbon in a sample compared to that of a sample with a conventional radiocarbon age of 0  $^{14}\text{C}$  yrs BP (i.e. mid-1950 AD) after accounting for isotopic fractionation. It is a largely linear calculation from the instrumental measurements meaning its uncertainties are approximately normal.

<sup>2</sup>We use the age-corrected  $\Delta^{14}\text{C}$  as described in Stuiver and Polach (1977).

- Radiocarbon age domain:  $h(\theta) = -8033 \ln f(\theta)$ . The radiocarbon age, in  $^{14}\text{C}$  yrs BP, is obtained from the fraction modern using Libby's original half-life and without any calibration. Since this is a non-linear mapping of  $F^{14}\text{C}$ , the uncertainties are no longer even approximately normally distributed as we approach the limit of the technique.

When fitting a smoothing spline, we have some flexibility in the precise choice of  $\text{FIT}(f, Y)$ , our assessment of fit, and our roughness penalty  $\text{PEN}(f)$ . For radiocarbon purposes, the natural domain to assess the quality of fit of a proposed calibration curve to observations is the  $F^{14}\text{C}$  domain, the raw scale on which the determinations are obtained. In this  $F^{14}\text{C}$  domain, our measurement uncertainties are symmetric. Conversely, in the radiocarbon age domain these uncertainties become asymmetric as we progress back in time. We therefore choose

$$\text{FIT}(f, F) = \sum_{i=1}^n \frac{1}{\sigma_i^2} (f(\theta_i) - F_i)^2,$$

where  $F_i$  are the observed  $F^{14}\text{C}$  values of the determinations and  $\sigma_i$  their associated uncertainties in the  $F^{14}\text{C}$  domain.

$F^{14}\text{C}$  is not however the most appropriate domain in which to assess roughness since it exhibits exponential decay over time making equitable penalization of potential calibration curves more difficult. Instead a more natural choice is the  $\Delta^{14}\text{C}$  domain where, a priori, one might expect a calibration curve to display approximately equal roughness over its length. For our penalty function we therefore choose

$$\text{PEN}(f) = \int \{g''(\theta)\}^2 d\theta,$$

where  $g''(\theta)$  is the second derivative of the proposed level of  $\Delta^{14}\text{C}$ , a standard penalty for function roughness.

We summarize this by saying that we perform modeling in the  $\Delta^{14}\text{C}$  domain as it is here we penalize roughness; while we perform fitting in the  $F^{14}\text{C}$  domain since it is here we assess fidelity of the curve to the raw observations. Since, given  $\theta$ , the transformation from  $F^{14}\text{C}$  to  $\Delta^{14}\text{C}$  is affine we can utilize these different modeling and fitting domains while still maintaining a practical spline estimation approach.

### 3.1.3 Bayesian Reframing

To reinterpret the above idea in a Bayesian framework we can split our functional

$$S(f) = \sum_{i=1}^n \frac{1}{\sigma_i^2} (f(\theta_i) - F_i)^2 + \int \{g''(\theta)\}^2 d\theta$$

into two distinct parts. The penalty component  $\int \{g''(\theta)\}^2 d\theta$  can be considered as specifying the prior distribution on the space of calibration curves; more precisely, it gives the negative log density of the prior. Intuitively this summarizes our prior beliefs about the form of the calibration curve before we observe any actual data. We then wish to update this prior belief, in light of the observed data, to obtain our posterior. This updating is achieved via the fitting component  $\sum_{i=1}^n \frac{1}{\sigma_i^2} (f(\theta_i) - F_i)^2$  which represents the negative log-likelihood of the observed data under an assumption that each  $F_i \sim N(f(\theta_i), \sigma_i^2)$ . The value of  $S(f)$  then represents the negative posterior log-density for a potential calibration curve  $f$ .

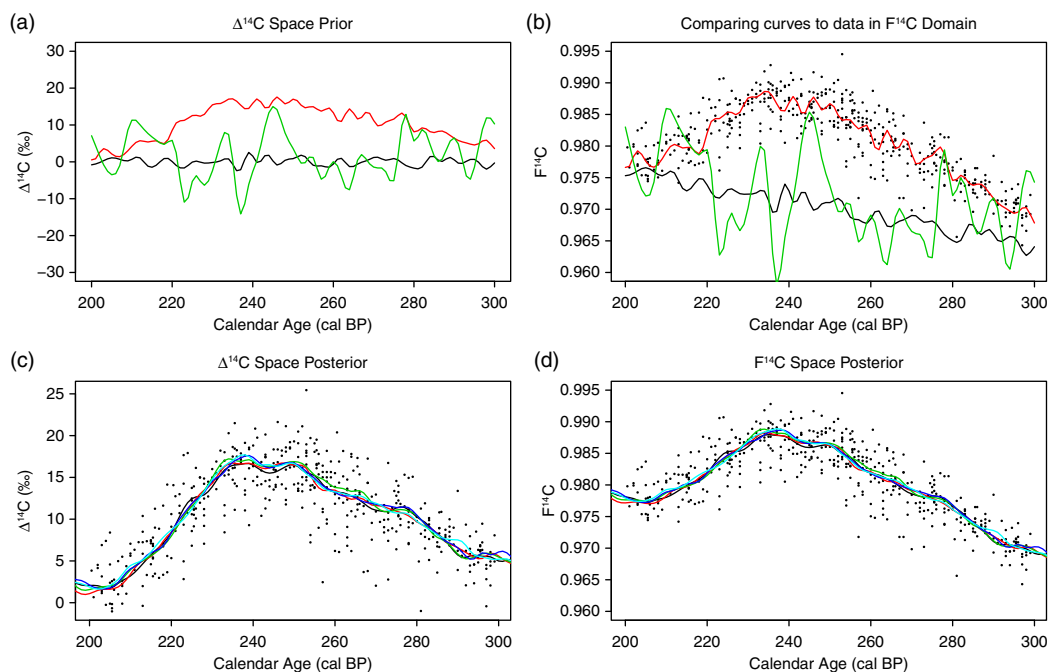


Figure 1 An illustration of Bayesian splines. Panel (a) shows some potential calibration curves in the  $\Delta^{14}\text{C}$  domain drawn from the prior. These are then compared with the observed data in the  $F^{14}\text{C}$  domain as shown in panel (b) to form our Bayesian posterior. The bottom two panels (c) and (d) show posterior realizations of potential curves (shown in  $\Delta^{14}\text{C}$  and  $F^{14}\text{C}$  space respectively) obtained via MCMC that provide a satisfactory trade-off between agreement with our prior penalizing over-roughness and the fit to our observed data.

This idea is illustrated in Figure 1. In panel (a) we present three potential calibration curves in the  $\Delta^{14}\text{C}$  domain. Using just our prior, we give each curve an initial weight corresponding to their roughness. The black curve is the least rough and so would be given highest prior weight as a plausible potential calibration curve. The green curve is the roughest and so has least weight according to our prior. Each of these  $\Delta^{14}\text{C}$  curves is then converted into the symmetric  $F^{14}\text{C}$  space and compared with our observations as shown in panel (b). The plausibility of each curve is then updated to incorporate the fit to this data. Based upon this, the red curve would have a higher posterior density as it fits relatively closely while not being overly rough. Both the green and black curves would have a low posterior density and so be considered highly implausible as they do not pass near the data. Use of MCMC allows us to generate any number of plausible calibration curves drawn from the posterior that provide a satisfactory trade-off between the roughness prior and fidelity to the observations. Some such realizations are shown in panels (c) and (d) in both the  $\Delta^{14}\text{C}$  and  $F^{14}\text{C}$  domains respectively. These realizations are then summarized to produce the final IntCal20 curve.

### 3.1.4 Variable Smoothing and Knot Selection

A further choice to be made when spline smoothing is the set of functions (or potential calibration curves) to search over for our functional  $S(f)$ . This set is called a basis. We use cubic splines where this basis is determined by specifying what are known as knots at distinct calendar times. The more knots one has in a particular time period, the more the

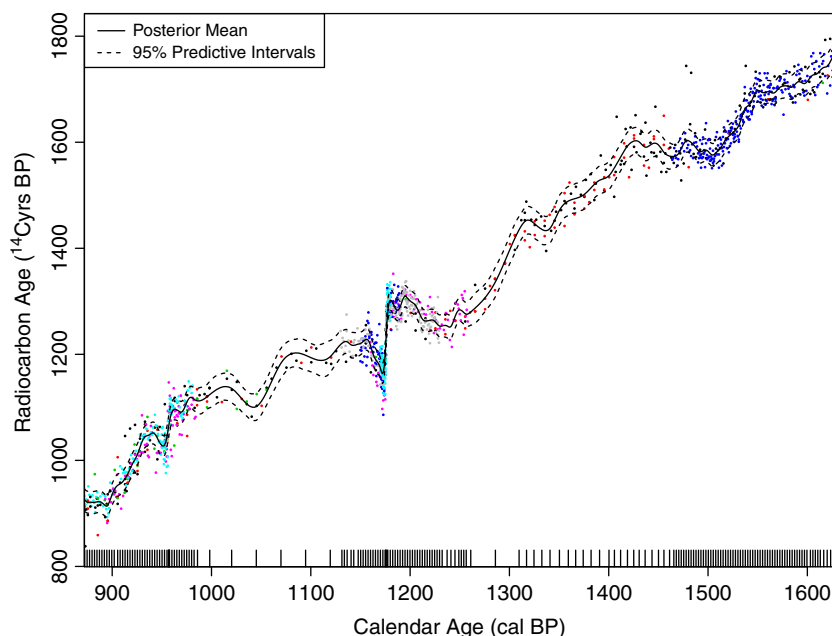


Figure 2 Variable smoothing and knot selection. Shown as a rug of tick marks along the bottom are the locations of the knots for the cubic spline. These are placed at quantiles of the observed calendar ages to provide variable smoothing. In dense regions, we can identify more detail in the calibration curve; while where the underlying data is less dense we perform more smoothing. In particular, note the additional knots placed around the two Miyake-type events (i.e. 957 and 1176 cal BP) which allow the curve to vary much more rapidly at these times. The points from different datasets within the IntCal database are shown in different colors to distinguish them.

spline can vary. One common option, known as a smoothing spline, is to place a knot at every calendar age  $\theta_i$  for  $i = 1, \dots, n$ . However, since we have a very large number of observations  $n$  this is computationally impractical. Instead we use P-splines (see Berry et al. 2002, for details) where we select a smaller number of knots; this is equivalent to restricting the potential calibration curves to a somewhat smaller subspace of functions.

In implementing P-splines we need to make sure that the curves we consider are still able to provide sufficient detail for a calibration user and identify fine scale features such as solar cycles where we have the ability to do so. The data on which we base our curve have highly variable density. In some periods, such as recent dendrodated trees, we have a great density of determinations while in others the underlying data is more sparse. To adapt to this and keep required detail, we choose a large number of knots and place them at calendar age quantiles of the data to provide variable smoothing. This approach of locating the knots at observed quantiles is standard in the regression literature (e.g. Harrell 2001). Where our data is dense, we can pick out the required fine detail but where the data is more sparse we smooth more significantly. An example can be seen in Figure 2. This figure also highlights how we pack additional knots around known Miyake-type events, narrow spikes

(sub-annual) of increased  $^{14}\text{C}$  production, to enable us to better retain these events in the final curve. For more details on choice of knots and placement see Sections 4.3.4 and 4.4.3.

## 3.2 Errors-in-Variables Regression

### 3.2.1 Calendar Age Uncertainty

Within the IntCal database (<http://www.intcal.org>), many of the determinations have calendar ages which are not known absolutely. This is particularly the case as we progress further back in time and calendar age estimates are constructed from uranium thorium (U-Th) dating, e.g. speleothems (Southon et al. 2012; Cheng et al. 2018; Beck et al. 2001; Hoffmann et al. 2010) and corals (Bard et al. 1990, 1998, 2004; Cutler et al. 2004; Durand et al. 2013; Fairbanks et al. 2005; Edwards et al. 1993; Burr et al. 1998, 2004); varve counting, e.g. parts of Cariaco Basin (Hughen et al. 2004) and Lake Suigetsu (Bronk Ramsey et al. 2020 in this issue); and palaeoclimate tuning/tie-pointing, e.g. other parts of Cariaco Basin (Hughen and Heaton 2020 in this issue), and the Pakistan and Iberian Margins (Bard et al. 2013). Furthermore, in the case of several of the floating tree-ring sequences, we have only a relative set of calendar ages and no absolute age estimate. Regression in this situation is called errors-in-variables since we have errors/uncertainties in both the calendar age and radiocarbon determination variables. For these observations we therefore observe pairs  $\{(F_i, T_i)\}_{i \in \mathcal{I}}$  where:

$$\begin{aligned} F_i &= f(\theta_i) + \varepsilon_i, \\ T_i &= \theta_i + \psi_i. \end{aligned}$$

Here  $f(\cdot)$  is our calibration curve of interest in the  $F^{14}\text{C}$  domain;  $\varepsilon_i \sim N(0, \sigma_i^2)$  independently; and  $\psi_i$  describes the uncertainty in our calendar age estimate. The form of this calendar age uncertainty varies between the datasets. While there is no restriction on the distribution of  $\psi_i$  for our Bayesian spline approach, we model all these calendar age uncertainties as normally distributed but with appropriate covariances. For some sets they are considered independent, e.g. corals dated by U-Th; while for others, e.g. floating tree-ring sequences and those records dated by varve counting or palaeoclimate tie-pointing (Heaton et al. 2013), there is considerable dependence between the observations. For more detail on the various covariance structures in the calendar age uncertainties see Niu et al. (2013).

### 3.2.2 Importance of Recognizing Calendar Age Uncertainty

Incorporation of calendar age uncertainty in the construction of the IntCal calibration curves has occurred since IntCal04 (Reimer et al. 2004a) and is key for reliable curve construction. It was therefore crucial to retain within the new methodology. A range of statistical methods have been developed to deal with errors-in-variables regression. From the frequentist, non-Bayesian, perspective Fan and Truong (1993) introduced a kernel-based approach which achieves global consistency; Cook and Stefanski (1994) proposed a more general approach known as SIMEX. These approaches were not however considered suitable for our application due to the genuine prior information we have on certain aspects of the data; the wider, almost universal, use of Bayesian methodology within radiocarbon calibration; and the independent interest in several aspects of the calibration data which could be provided more simply through a Bayesian method. We therefore began the development of our method using the Bayesian splines of Berry et al. (2002).

The effect of not recognizing calendar age uncertainty, where it is present, in regression is quite varied. In the case that the calendar age uncertainties are entirely independent of one another,

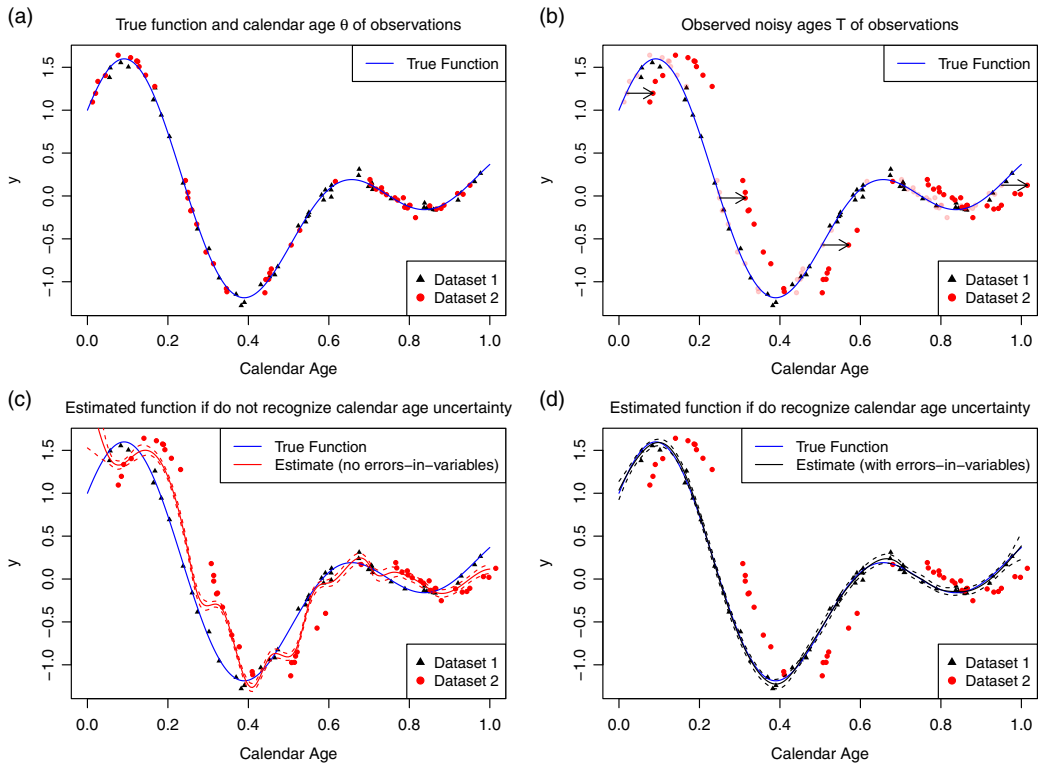


Figure 3 The importance of recognizing calendar age uncertainty (and correctly representing covariance within that uncertainty) when constructing a curve based on data arising from records with different observed timescales. Panel (a) shows the true, underlying, calendar ages of the data in two different records; while panel (b) shows a joint shift in the observed calendar ages within record 2. In such a case that observed timescales in two records are offset from one another, if we ignore this calendar age uncertainty then our spline estimate will introduce spurious variation as it flips between the records as in panel (c). Conversely if we incorporate such calendar age uncertainty and accurately represent it then we can still recover the underlying function accurately, as shown in (d).

such as U-Th dating, not recognizing calendar age uncertainty will typically result in curve estimates that are overly smooth and do not attain the peaks and troughs of the true underlying function. See Samworth and Poore (2005) for an illustrative case study. However, in the case of IntCal, the situation is made more complicated by the shared dependence of the calendar age estimates within particular datasets. Specifically, the entire timescale for, e.g. Hulu Cave, is not the same as the Suigetsu or Cariaco timescale. As a consequence, all of these individual datasets may show the same overall features but at slightly different times. We require our method to recognize that these different timescales may need to be aligned, within their respective uncertainties, to keep these shared features. This requires us to shift multiple ages jointly by stretching/squashing the timescales accordingly. A failure to adapt to this joint calendar age uncertainty can give curve estimates which either introduce spurious wiggles as the curve flips between the different datasets or lose major features entirely.

### Illustration

We provide in Figure 3 an illustrative example of the need to incorporate calendar age uncertainty. For simplicity, in this example, we fit and model our spline in the same

domain. We consider a straightforward underlying function which we wish to reconstruct from 100 noisy observations  $\{(Y_i, T_i)\}_{i=1}^{100}$

$$f(\theta) = e^{-\theta} \sin(4\pi\theta) + \cos(2\pi\theta),$$

where our underlying  $\theta_i \sim \text{Beta}(1.1, 1.1)$  and observed  $Y_i \sim N(f(\theta_i), 0.05^2)$  for  $i = 1, \dots, 100$ . Further, let us assume that these observations arise from two different sediment cores, 50 from core 1 (shown as black triangles) and 50 from core 2 (shown as red dots) as seen in panel (a). Within core 1, the calendar ages are known absolutely. However, within core 2 the observed timescale is somewhat shifted/biased so that when we observe the calendar ages  $T$  in this core they all share the same joint shift from their true values, i.e.

$$T_i = \theta_i \quad \text{if } i \text{ is in core 1,}$$

$$T_i = \theta_i + \psi \quad \text{if } i \text{ is in core 2,}$$

where  $\psi \sim N(0, 0.1^2)$ . The observed pairs  $\{(Y_i, T_i)\}_{i=1}^{100}$  are presented in panel (b) showing the shift in timescale within core 2. If we attempt to reconstruct the function using splines based upon our observed  $\{(Y_i, T_i)\}_{i=1}^{100}$  without recognizing the calendar age uncertainty, i.e. assuming both cores are on the same timescale, then we obtain the estimate in the panel (c). The estimate is spuriously variable as the spline will try to pass near all of the data on their non-comparable timescales. Conversely, if we recognize that the timescale in core 2 may need to be shifted onto the timescale of core 1, we obtain the estimate shown in panel (d). If we inform our Bayesian method that the calendar age observations in core 2 are subject to uncertainty, it will estimate the size of the joint shift needed to register core 2 onto the true timescale of core 1 and simultaneously reconstruct the true underlying function.

### Implications for Visually Assessing Curve Fit

While it is important to incorporate errors-in-variables in curve construction, and accurately represent any dependence in the calendar age uncertainties, this does cause some difficulty in visually assessing the quality of fit between the raw constituent data and the final IntCal curve. Since the proposed method can shift the calendar ages of the observed data left and right within their uncertainties as described, one cannot estimate the final curve's goodness-of-fit by eye using only the radiocarbon age axis if the raw data are only plotted at their initial, observed calendar ages. The method will have tried to align shared features even if they occur at somewhat different observed times within the different sets. This should be taken into consideration when viewing the final curve against the raw data. We provide the posterior calendar age estimates for the true calendar ages  $\theta_i$  for all the data. The difficulty of assessing the fit of the curve by eye is further compounded by the offsets in  $^{14}\text{C}$  described in Section 4.4, in particular the marine reservoir ages which vary significantly over time and so will change as the MCMC updates the calendar ages of the data.

## 3.3 Predictive Intervals and Over-Dispersion

### 3.3.1 Background and Motivation

As well as creating a curve which has the correct posterior mean, it is key to make sure we provide appropriate intervals on the curve to ensure that, when used for calibration, the calendar age estimates produced are not over- or under-precise. This becomes particularly relevant for IntCal20 due to the large increase in data available to create the curve, and also the increased measurement precision provided by current laboratories. In addition to the laboratory reported uncertainty, there are a wide range of possible further sources of variation in the recorded  $^{14}\text{C}$  within objects of the same age, for example determinations come from different

locations; have different local environments; tree rings may be of different species; and have different periods of growth due to local weather and so may have differing elements of wood from late/early growth. Even when the same sample is measured in different laboratories, we have evidence of a greater level of observation spread (i.e. over-dispersion) in the  $^{14}\text{C}$  measurements within objects from the same calendar year than the laboratory reported uncertainties would support (see e.g. Scott et al. 2017). Recognizing this potential over-dispersion, whatever its cause, is important both for curve estimation and resultant calibration for IntCal20. Specifically, if there is more variation in determinations from the same calendar year than the uncertainties reported by the laboratory, we need to make sure we incorporate that in our modeling, and also recognize its existence in the objects users will then calibrate against the IntCal20 curve.

We achieve this through the inclusion of a term to quantify the level of over-dispersion which we adaptively estimate, based on the high-quality and dense IntCal data, within our curve construction. This ensures that our IntCal20 curve does not become over-precise as an estimate of the Northern Hemispheric average. However, this alone is not sufficient for calibration since any additional sources of variation seen within the IntCal data are also likely to be present in uncalibrated determinations. We must therefore propagate this over-dispersion through to the calibration process itself. This is achieved using predictive curve intervals which incorporate not only uncertainty in the Northern Hemispheric average atmospheric curve but also the potential additional sources of variation seen in individual  $^{14}\text{C}$  determinations. Importantly, if no such additional variability exists then our model will estimate this appropriately, however if such additional sources of variability do seem present in the calibration data, our method will also recognize this and adapt accordingly. Without such an over-dispersion term then, as we get more data, the calibration curve will have intervals that become narrower and narrower even if the underlying data suggests considerably higher variability. Eventually, this would mean that the data entering the curve itself would potentially not calibrate to their known ages.

Such an idea of irreducible uncertainty was introduced in Niu et al. (2013) but here we extend the idea further and generate predictive curve intervals which are more relevant for calibration users. Importantly, this does not however mean that we produce a calibration tool that any measurement (no matter its quality) can be calibrated against—we wish to maintain a calibration curve which has basic minima for data quality for both what goes into its construction and also what can be reliably calibrated using it.

### 3.3.2 *The Over-Dispersion Model: Additive Errors Scaling with $\sqrt{F^{14}\text{C}}$*

The basic tree-ring model for an annual measurement assumes that a determination of calendar age  $\theta$  arises from a hemispherically uniform atmospheric level of  $^{14}\text{C}$  shared by all determinations of the same calendar age. Under this model, the only uncertainty is that reported by the laboratory so that, in the  $F^{14}\text{C}$  domain, the observed value is

$$F_i = f(\theta_i) + \varepsilon_i,$$

where  $\theta_i$  is the year the determination represents; and  $\varepsilon_i \sim N(0, \sigma_i^2)$  with  $\sigma_i$  the uncertainty reported by the laboratory. The function  $f(\theta)$  is our unique estimate of the hemispheric level of  $F^{14}\text{C}$  present in the atmosphere at calendar age  $\theta$ .

Currently it is not feasible to identify the specific sources that may potentially contribute extra variability beyond that reported by the laboratory (e.g. regional, species, growing season

differences) and so we aim to cover them all in a unified approach. We therefore modify the above model to allow any determination to have an additional and independent source of potential variability  $\eta_i$  beyond that reported by the laboratory, i.e.

$$F_i = f(\theta_i) + \varepsilon_i + \eta_i \quad \text{for } i = 1, \dots, n$$

where  $\eta_i \sim N(0, \tau_i^2)$  are independent random effects with unknown variance  $\tau_i^2$ . To create the curve we simultaneously estimate both  $\tau_i$ , the level of over-dispersion, and the calibration curve  $f(\cdot)$ .

After investigation of several options, see Section 5.1 and the Supplementary Information, we model the level of over-dispersion (i.e. the level of any additional variability) to scale proportionally with the underlying value of  $\sqrt{f(\theta)}$  i.e.  $\tau_i = \tau \sqrt{f(\theta_i)}$  so that  $\eta_i \sim N(0, \tau^2 f(\theta))$ . We choose a prior for the constant of proportionality  $\tau$  based upon data taken from the SIRI inter-comparison project (Scott et al. 2017), and update this prior within curve construction based upon the IntCal data. The information provided by the very large volume of IntCal data dominates the SIRI prior and so the posterior estimate for the level of over-dispersion is primarily based upon the high-quality IntCal data.

### 3.3.3 Implications for Calibration

A user seeking to calibrate a high-quality <sup>14</sup>C determination against the IntCal20 curve is likely to have a measurement that has been subject to similar additional sources of potential variation to whatever is seen in the IntCal data, and so have a similar level of over-dispersion. However, such a user typically has no way of assessing this over-dispersion themselves. Consequently, they should calibrate their determination using predictive curve estimates which additionally account for potential over-dispersion in their own determination. These predictive estimates are based upon the posterior of

$$f(\theta) + \eta(\theta),$$

where  $\eta(\theta) \sim N(0, \tau^2 f(\theta))$  and  $\tau$  is the posterior estimate of over-dispersion based upon the IntCal data. For IntCal20 we therefore report this predictive interval since it is more relevant for calibration. It is slightly wider than the corresponding credible interval for  $f$  but more likely to give accurate calibrated dates.

Note that the level of over-dispersion, and hence the predictive intervals, we incorporate into the IntCal20 curve is based upon the high-quality, and screened, IntCal database. If a new uncalibrated <sup>14</sup>C determination has additional sources of variability beyond those present in the IntCal tree-ring database (e.g. tree-ring species or locations not represented in IntCal, or a non-tree-ring sample) then this may mean that the level of over-dispersion for that uncalibrated determination is higher than that incorporated within the IntCal prediction intervals. This would result in a potentially over-precise calibrated age estimate. Users are advised to be cautious in such circumstances.

## 4 CREATING THE CURVE

As in previous versions of IntCal, the IntCal20 curve itself is created in two linked sections. Firstly we create the more recent part of the curve (extending from 0 cal kBP back to approximately 14 cal kBP) which is predominantly based upon dendrodated tree-ring determinations. Secondly we create the older part of the curve (from approximately 14 cal kBP back to 55 cal kBP) which is based upon a wider range of material, e.g. corals, macrofossils, forams, speleothems as well as five floating tree-ring chronologies, that are of

uncertain calendar ages. Furthermore, these older  $^{14}\text{C}$  determinations are often not direct atmospheric measurements and hence have marine reservoir ages or dead carbon fractions. We therefore split our technical description of the approach accordingly. The two curve sections are stitched together to ensure smoothness and continuity through appropriate design of spline basis and conditioning the older part of the curve on the, already estimated, more recent section. This means that we can still produce sets of complete curve realizations from 0–55 cal kBP. Future work will aim to adapt the approach into a single step, updating the entire 0–55 cal kBP range as one.

#### 4.1 Notation

##### *Calibration Curve by Domain*

We can represent the atmospheric history of  $^{14}\text{C}$ , as a function of calendar age  $\theta$ , in three domains equivalently:  $\Delta^{14}\text{C}$ ,  $F^{14}\text{C}$ , and radiocarbon age. For any proposed history, i.e. calibration curve, we switch between these domains as appropriate within our statistical methodology. Let us denote:

- $g(\theta)$ —the  $^{14}\text{C}$  calibration curve represented in  $\Delta^{14}\text{C}$  space, i.e. the level of  $\Delta^{14}\text{C}$  at  $\theta$  cal BP. We model  $g(\theta)$  in our spline basis and penalize roughness of the curve in this domain.
- $f(\theta)$ —the  $^{14}\text{C}$  calibration curve represented in  $F^{14}\text{C}$  space. Given  $g(\theta)$  then

$$f(\theta) = \left( \frac{1}{1000}g(\theta) + 1 \right) e^{-\theta/8267}.$$

It is this  $F^{14}\text{C}$  domain that is used to assess goodness-of-fit to the observed data.

- $h(\theta)$ —the  $^{14}\text{C}$  calibration curve represented in radiocarbon age:

$$h(\theta) = -8033 \ln f(\theta).$$

This is the domain in which the calibration curve is plotted in the main IntCal20 paper (Reimer et al. 2020 in this issue).

##### *Observed Data*

We consider all of our observations in  $F^{14}\text{C}$  since, as explained in Section 3.1, in this domain our  $^{14}\text{C}$  measurement uncertainties are symmetric. We define:

- $F_i$ —the observed  $F^{14}\text{C}$  value of the data used to create the IntCal20 curve.
- $\sigma_i$ —the laboratory-reported uncertainty on the observed  $F^{14}\text{C}$ .
- $m_i$ —the number of annual years of metabolization that a determination represents. The determination is considered to represent the mean of this block.
- $T_i$ —the observed calendar age of a determination, which could either be the true calendar age (in the case of absolutely dendrodated determinations) or an estimate with uncertainty (in the case of e.g. corals, varves, speleothems).

##### *Model Parameters*

Within the model we update:

- $\beta_j$ —the spline coefficients which describe the calibration curve.
- $\lambda$ —the smoothing parameter for the spline estimate.

- $\theta_i$ —the true calendar age of a determination or, in the case of a block-average determination, the most recent year of metabolization included in the block.
- $\tau$ —the level of over-dispersion in the observed  $F^{14}\text{C}$  under a model whereby the potential additional variability for determination  $F_i$  is  $\eta_i \sim N(0, \tau^2 f(\theta_i))$ .
- $r_{\mathcal{K}}(\theta)$ —the offset measured in terms of radiocarbon age, either due to dead carbon fraction (dcf) or marine reservoir age (MRA), between a determination from set  $\mathcal{K}$  and the atmosphere at time  $\theta$  cal BP. These will be specified either by  $\nu_{\mathcal{K}}$ , the mean dcf offset/coastal MRA shift for a particular dataset; or further spline coefficients  $\beta_{\mathcal{C}}$  in the case of the Cariaco unvarved record. These are only needed when estimating the older part of the curve.
- $\kappa_i$  and  $\varrho_{\mathcal{K}}$ —adaptive error multipliers for data arising from the older time period to enable heavier tailed errors and the down-weighting of outliers. Also only included when estimating the older part of the curve.

### 4.2 Basic Model for Observed Data and Curve

As described in Section 3.1, we fit our curve to the data in the  $F^{14}\text{C}$  domain while the curve modeling is performed in the  $\Delta^{14}\text{C}$  domain. Specifically, for a single-year determination of the direct atmosphere with no offset and no over-dispersion, we observe pairs  $(F_i, T_i)_{i=1}^n$  where

$$\begin{aligned} F_i &= f(\theta_i) + \varepsilon_i, \\ &= \left( \frac{g(\theta_i)}{1000} + 1 \right) e^{-\theta_i/8267} + \varepsilon_i; \\ T_i &= \theta_i + \psi_i. \end{aligned}$$

Here  $g(\theta)$ , the value of  $\Delta^{14}\text{C}$  over time, is modeled as

$$\begin{aligned} g(\theta) &= \sum_{j=1}^K \beta_j B_j(\theta) \\ &= \mathbf{B}(\theta)^T \boldsymbol{\beta}, \end{aligned}$$

where  $\mathbf{B}(\theta) = (B_1(\theta), \dots, B_K(\theta))^T$  and the  $B_j(\cdot)$  are cubic B-splines (Green and Silverman 1993) at a chosen fixed set of knots. To maintain computational feasibility, we consider  $K \ll n$  so that the number of splines in our basis is considerably smaller than the total number of observations in the IntCal database. Knot number and placement is discussed in Sections 4.3.4 and 4.4.3.

#### Prior on $\boldsymbol{\beta}$

The Bayesian spline approach, equivalent to penalizing roughness in  $\Delta^{14}\text{C}$  by  $\lambda \int_a^b \{g''(\theta)\}^2 d\theta$ , is obtained by placing a “partially improper” Gaussian process on the spline coefficients  $\boldsymbol{\beta}$ :

$$\pi(\boldsymbol{\beta}|\lambda) = \left( \frac{\lambda}{2\pi} \right)^{\text{rank}(\mathbf{D})/2} (|\mathbf{D}|^*)^{1/2} \exp \left\{ -\frac{\lambda}{2} \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \right\},$$

where  $\mathbf{D}$  is a penalty matrix which penalizes the integrated squared second derivative and  $|\mathbf{D}|^*$  its generalized determinant (see Green and Silverman 1993, for details). For splines of degree  $m$  (cubic splines have degree 3) then  $\text{rank}(\mathbf{D}) = K - (m - 1)$ .

*Prior on  $\lambda$* 

As standard within Bayesian splines (e.g. Berry et al. 2002), we place a hyperprior on the smoothing parameter  $\lambda \sim Ga(A, B)$ :

$$\pi(\lambda) = \frac{1}{\Gamma(A)B^A} \lambda^{A-1} \exp\left(-\frac{\lambda}{B}\right) \quad (\text{for } \lambda > 0).$$

We select an uninformative prior on  $\lambda$  with  $A = 1$  and  $B = 50000$ .

### 4.3 Creating the Predominantly Dendrodated Part of the Curve Back to Approximately 14 cal kBP

Back to approximately 14 cal kBP, we have a sufficient density of tree-ring determinations that we can estimate the curve based solely upon these direct atmospheric observations<sup>3</sup>. The main challenges in creating this more recent section of the curve are:

- High density and volume of data—there are 10,713 <sup>14</sup>C measurements in this, predominantly dendrodated, section. Such a large volume of data makes a computationally efficient algorithm essential.
- Consequent demand for high level of detail in calibration curve—this, more recent, period of the calibration curve is the most heavily interrogated by archaeologists who require high precision, not just for individual calibration but also for modeling of multiple dates. This increases the user demand for fine detail in the calibration curve, e.g. incorporating solar cycles.
- Blocking within the data—many of the <sup>14</sup>C determinations do not relate to the measurement of the atmospheric <sup>14</sup>C in a single year but rather averages of multiple tree rings and so represent multiple years.
- Floating tree-ring sequences around 12.7 cal kBP—several late glacial trees have their chronologies estimated by wiggle matching and so while their internal chronology is known absolutely, their absolute ages are not. Otherwise, the true calendar ages of all determinations are known absolutely for this section of curve.

#### 4.3.1 Modifications to Basic Model

##### **Blocking and Additive Errors**

Our model for the tree-ring determinations, taking account of blocking, considers the observed  $F_i$  as

$$F_i = \frac{1}{m_i} \left\{ \sum_{j=0}^{m_i-1} f(\theta_i + j) \right\} + \varepsilon_i + \eta_i \quad \text{for } i = 1, \dots, n$$

where  $m_i$  is the number of annual rings in the block for that determination;  $\theta_i$  is the most recent year (start ring) of the block; and  $\varepsilon_i \sim N(0, \sigma_i^2)$  where  $\sigma_i$  is the uncertainty reported by the laboratory. This implicitly assumes that the tree rings over the section of  $m_i$  rings are approximately equal in width, so that approximately equal annual amounts of wood were deposited and have ended up in

<sup>3</sup>The IntCal database contains tree-ring determinations extending back to 14,189 cal BP all of which are used to create a predominantly dendrodated tree-ring-only curve from 0–14,189 cal BP. However, beyond 13,913 cal BP the density of this data was not sufficient to estimate the curve precisely. Hence, we merge the predominantly dendrodated curve section with the older section at 13,913 cal BP.

the final measured sample. The  $\eta_i \sim N(0, \tau_i^2)$  are independent random effects, with  $\tau_i$  unknown, that represent potential over-dispersion in the <sup>14</sup>C determinations.

**Additive Errors Model**

As described in Section 5.1 we model  $\tau_i$ , the standard deviation of the additive over-dispersion on the *i*th determination, as proportional to the square-root of the underlying <sup>14</sup>C, i.e.

$$\tau_i^2 = \tau^2 \frac{1}{m_i} \sum_{j=0}^{m_i-1} f(\theta_i + j).$$

The individual values of  $\eta_i$  are not of interest and so we integrate them out, updating only the overall constant of proportionality  $\tau$ . Formally, a structure where  $\tau_i$  depends upon the underlying curve means this term should be incorporated in the update to  $\beta$  within our sampler—a change to the curve (via its spline coefficients  $\beta$ ) would mean a change to the additive uncertainty  $\tau_i$  on all observations. However, this would make the  $\beta$  update no longer a Gibbs step. Since, given the overall multiplier  $\tau$ , the change in  $\tau_i$  between two potential curves is minimal, when updating  $\beta$  we consider  $\tau_i$  as fixed; see Section 4.3.3 for more details.

**Floating Tree-Ring Sequences**

Several late-glacial tree sequences measured by Aix, ETH and Heidelberg (Reinig et al. 2018; Capano et al. 2020 in this issue) were dated by wiggle-matching so, while their internal relative chronologies were known precisely, their absolute calendar ages were somewhat uncertain. For such a floating tree-ring sequence, the true calendar ages are:

$$\theta_j^{\text{float}} = \xi + \delta_j \text{ for } j \text{ in floating sequence.}$$

where  $\xi$  is the unknown start date (i.e. most recent year) of the floating sequence and  $\delta_j$  is the precisely-known internal ring-count chronology. We place a discretized (integer) normal prior on  $\xi \sim N(M, N^2)$ . Here, the prior mean *M* and variance *N*<sup>2</sup> for the start date of each such tree-ring sequence were provided. All of the other dendrodated determinations were assumed to have their calendar age  $\theta_i$  known absolutely.

**4.3.2 Efficient Incorporation of Blocking**

To incorporate blocking exactly, at each iteration of our MCMC we need to evaluate the level of <sup>14</sup>C at every year represented within a block, and then average over them appropriately for each of our *n* determinations. One might think this additional calculation will make the method much slower as it requires the evaluation of the curve at many more than *n* calendar years. However, in the predominantly dendrodated part of the curve, it is possible to exactly incorporate blocking without any compromise on the speed of the estimation procedure. Since the vast majority of the data in this period have exactly known calendar ages (only the floating tree-ring sequences have uncertain ages) we can calculate an initial matrix that relates, for each blocked determination, the necessary averaged <sup>14</sup>C to the underlying spline coefficients. This matrix then remains fixed throughout estimation. Further it has the same dimensions as if blocking had been ignored and hence does not significantly alter the speed of the MCMC updates.

**Finding Values of <sup>14</sup>C at Each Year Represented in a Block**

We begin by describing how we can relate each blocked determination to the spline coefficients  $\beta$ . Let  $\theta^A = (\theta_1, \theta_1 + 1, \dots, \theta_1 + m_1 - 1, \theta_2, \dots, \theta_2 + m_2 - 1, \dots, \theta_n + m_n - 1)^T$  be a vector concatenating every calendar year, including any potential blocking, represented in all *n* determinations. Also, define vectors  $c_\theta^A$  and  $e_\theta^A$  so that

$$c_{\theta,l}^A = \frac{1}{1000} e^{-\theta_l^A/8267} \quad \text{and} \quad e_{\theta,l}^A = e^{-\theta_l^A/8267}.$$

Then, create the matrix  $\mathcal{B}_\theta^A$  containing the spline bases evaluated at each value in  $\theta^A$ :

$$\mathcal{B}_\theta^A = \begin{pmatrix} B_1(\theta_1) & \dots & B_K(\theta_1) \\ B_1(\theta_1 + 1) & \dots & B_K(\theta_1 + 1) \\ \vdots & & \vdots \\ B_1(\theta_n + m_n - 1) & \dots & B_K(\theta_n + m_n - 1) \end{pmatrix},$$

so that, at the calendar years  $\theta^A$ , the modeled value of  $\Delta^{14}\text{C}$  is  $\mathbf{g}^A = \mathcal{B}_\theta^A \boldsymbol{\beta}$  where  $\boldsymbol{\beta}$  represents the spline coefficients. The modeled value of  $F^{14}\text{C}$  at an individual calendar age  $\theta_l^A$  then becomes  $f_l^A = c_{\theta,l}^A g_l^A + e_{\theta,l}^A$  and at the full  $\theta^A$ ,  $\mathbf{f}^A = \mathcal{B}_\theta^A \boldsymbol{\beta} + \mathbf{e}_\theta$  where  $\mathcal{B}_\theta^A$  is formed by multiplying each row in  $\mathcal{B}_\theta^A$  by the corresponding element in  $\mathbf{c}_\theta^A$ .

**A Blocking Matrix**

For each blocked determination, we now wish to average the values of  $F^{14}\text{C}$  according to the multiple years represented in that block. Consider an averaging matrix  $\mathbf{M}$  corresponding to averaging the individual values in  $\mathbf{f}^A$ :

$$\mathbf{M} = \begin{pmatrix} 1/m_1 & \dots & 1/m_1 & 0 & \dots & 0 & \dots & \dots \\ 0 & 0 & 0 & 1/m_2 & \dots & 1/m_2 & 0 & \dots \\ \vdots & & & & & & & \end{pmatrix}.$$

Our observed vector of  $F^{14}\text{C}$  measurements  $\mathbf{F}$  thus becomes

$$\begin{aligned} \mathbf{F} &= \mathbf{M} \mathcal{B}_\theta^A \boldsymbol{\beta} + \mathbf{M} \mathbf{e}_\theta + \boldsymbol{\varepsilon} + \boldsymbol{\eta}, \\ &= \mathcal{B}_\theta^* \boldsymbol{\beta} + \mathbf{e}_\theta^* + \boldsymbol{\varepsilon} + \boldsymbol{\eta}. \end{aligned}$$

Note that both the matrix  $\mathcal{B}_\theta^*$  and  $\mathbf{e}_\theta^*$  depend upon the true calendar ages. However, since for the vast majority of the tree-rings, the calendar ages are known exactly, one only has to calculate  $\mathcal{B}_\theta^*$  and  $\mathbf{e}_\theta^*$  once as it is then fixed. The final  $\mathcal{B}_\theta^*$  has dimension  $K \times n$  and hence the Gibbs updating of  $\boldsymbol{\beta}$  is the same speed as if one had no blocking. To incorporate the floating tree-ring sequences, for any potential sequence start date  $\xi$  we only need update a small submatrix of  $\mathcal{B}_\theta^*$  and  $\mathbf{e}_\theta^*$  corresponding to these observations. The required updates for all sequence start dates  $\xi$  covering the prior range can be calculated and stored before commencing curve estimation.

**4.3.3 Details of MCMC Algorithm**

**Posterior**

Using the above hierarchical structure, the joint posterior is proportional to

$$\begin{aligned} &[\xi, \boldsymbol{\beta}, \lambda, \tau | \mathbf{F}, \boldsymbol{\theta}] \propto [\mathbf{F} | \boldsymbol{\beta}, \boldsymbol{\theta}, \xi, \tau][\boldsymbol{\beta} | \lambda][\lambda | \tau][\tau | \xi] \\ &\propto \det(\mathbf{W}_\tau)^{\frac{1}{2}} \exp \left\{ -\frac{1}{2} \left[ \{(\mathbf{F} - \mathbf{e}_\theta^* - \mathcal{B}_\theta^* \boldsymbol{\beta})^T \mathbf{W}_\tau (\mathbf{F} - \mathbf{e}_\theta^* - \mathcal{B}_\theta^* \boldsymbol{\beta})\} + \frac{(\xi - M)^2}{N^2} + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} \right] \right\} \\ &\times \lambda^{A + \text{rank}(\mathbf{D})/2 - 1} \times \exp \left( -\frac{\lambda}{B} \right), \end{aligned}$$

where  $\mathbf{W}_\tau = \text{diag}(\sigma_1^2 + \tau_1^2, \dots, \sigma_n^2 + \tau_n^2)^{-1}$  includes the additive error representing the over-dispersion. We can sample from this using Metropolis-within-Gibbs but where most of the updates can be performed directly via Gibbs and only a few require Metropolis-Hastings (MH).

**Gibbs Updating  $\beta|\mathbf{F}, \lambda, \xi, \tau$**

$$\pi(\beta|\mathbf{F}, \lambda, \tau) \propto \exp\left\{-\frac{1}{2}\left[\{(\mathbf{F} - \mathbf{e}_\theta^* - \mathcal{B}_\theta^*\beta)^T \mathbf{W}_\tau (\mathbf{F} - \mathbf{e}_\theta^* - \mathcal{B}_\theta^*\beta)\} + \lambda\beta^T \mathbf{D}\beta\right]\right\}$$

For the purposes of maintaining a Gibbs update here, we treat  $\mathbf{W}_\tau = \text{diag}(\sigma_1^2 + \tau_1^2, \dots, \sigma_n^2 + \tau_n^2)^{-1}$  as fixed and independent of  $\beta$ , i.e. we ignore the dependence of the individual over-dispersion  $\tau_i$  on the curve. With this minor approximation, algebra gives a direct complete conditional

$$\beta|\mathbf{F}, \lambda \sim \text{MVN}(\mathbf{Q}\mathcal{B}_\theta^{*T}\mathbf{W}_\tau(\mathbf{F} - \mathbf{e}_\theta^*), \mathbf{Q}),$$

where

$$\mathbf{Q} = (\mathcal{B}_\theta^{*T}\mathbf{W}_\tau\mathcal{B}_\theta^* + \lambda\mathbf{D})^{-1}.$$

Here large computational savings can be made if we calculate  $\mathbf{Q}(\mathcal{B}_\theta^{*T}\mathbf{W}_\tau\mathbf{F})$  rather than  $(\mathbf{Q}\mathcal{B}_\theta^{*T})\mathbf{W}_\tau\mathbf{F}$  since  $\mathbf{Q}$  is  $K \times K$  while  $\mathcal{B}_\theta^{*T}$  is  $K \times n$  (and  $n \gg K$ ).

**Gibbs Updating  $\lambda|\beta, A, B$**

Due to conjugacy of the  $\lambda$  prior:

$$\pi(\lambda|\beta) \propto \lambda^{A+\text{rank}(\mathbf{D})/2-1} \exp\left\{-\lambda\left(\frac{\beta^T \mathbf{D}\beta}{2} + \frac{1}{B}\right)\right\},$$

i.e.  $\lambda|\beta \sim \text{Ga}(A', B')$  where  $A' = A + \text{rank}(\mathbf{D})/2$  and  $B' = \left(\frac{\beta^T \mathbf{D}\beta}{2} + \frac{1}{B}\right)^{-1}$ .

**MH Updating  $\xi|\mathbf{F}, \lambda, \beta, \tau$**

To update the start of our floating tree-ring sequences, we use an MH step:

- Propose  $\xi' \sim \text{MVN}(\xi, \sigma_{\text{prop}})$ ;
- Calculate  $\mathcal{B}_\theta^*$  and  $\mathbf{e}_\theta^*$  where  $\theta'$  are the calendar ages of the floating sequence with the proposed new start date  $\xi'$ . Similarly calculate  $\mathbf{e}_\theta^*$  and  $\mathbf{e}_\theta^*$ . Accept  $\xi'$  according to Hastings Ratio:

$$\text{HR} = \min\left\{1, \frac{\pi(\xi'|\mathbf{F}, \beta, \tau)}{\pi(\xi|\mathbf{F}, \beta, \tau)}\right\},$$

where

$$\pi(\xi|\mathbf{F}, \beta, \tau) = \exp\left\{-\frac{1}{2}\left[\{(\mathbf{F} - \mathbf{e}_\theta^* - \mathcal{B}_\theta^*\beta)^T \mathbf{W}_\tau (\mathbf{F} - \mathbf{e}_\theta^* - \mathcal{B}_\theta^*\beta)\} + \frac{(\xi - M)^2}{N^2}\right]\right\}.$$

**MH Updating  $\tau|\mathbf{F}, \lambda, \beta$**

This also requires an MH update step. Let the residuals between the observed  $\mathbf{F}$  and fitted values  $\hat{\mathbf{F}} = \mathcal{B}_\theta^*\beta + \mathbf{e}_\theta^*$  be

$$\mathbf{R} = \mathbf{F} - \mathcal{B}_\theta^*\beta - \mathbf{e}_\theta^*.$$

Then, we have independently  $R_i \sim N(0, \sigma_i^2 + \tau_i^2)$  where  $\tau_i = \tau\sqrt{\hat{F}_i}$ . With our prior  $\tau \sim \pi(\tau)$  we update:

- Sample  $\tau' \sim N_+(\tau, \eta^2)$ ;
- Accept according to Hastings Ratio:

$$HR = \min \left\{ 1, \frac{f(\mathbf{R}; \tau') \pi(\tau') \Phi(\tau/\eta)}{f(\mathbf{R}; \tau) \pi(\tau) \Phi(\tau'/\eta)} \right\},$$

where  $f(\mathbf{R}; \tau) = \prod_{i=1}^n \phi(R_i, 0, \sigma_i^2 + \tau_i^2)$  and  $\tau_i = \tau \sqrt{\hat{F}_i}$ . Note the asymmetric proposal adjustment.

#### 4.3.4 Additional Considerations

##### **Choice of Splines**

For the portion of the curve from 0 cal kBP back to approximately 14 cal kBP, we placed 2000 knots at the unique calendar age quantiles of the constituent determinations (each blocked determination was represented by its midpoint) to permit variable smoothing dependent upon data density. In regions where there were more data this allowed us to pick out finer scale variation. Such a selection enabled close to annual resolution in the detail of the final curve while still maintaining computational feasibility in curve construction. We also placed additional knots in the vicinity of the known Miyake-type events at 774–5 AD, 993–4 AD and ~660 BC (Miyake et al. 2012, 2013; O'Hare et al. 2019) to enable us to capture more rapid variation at these times. Around the calendar age of each such event, we added an extra 4 jittered knots, i.e. after addition of small amounts of random noise—for the ~660 BC event we used a slightly larger jitter due to its slightly less certain timing. If there is no event at these times then the curve will not introduce one but, if there is, then these additional knots will allow it to be picked up more clearly. All knot locations then remained fixed throughout the sampler.

##### **Outlier Screening**

In order to identify potential outliers, after fitting a preliminary curve, scaled residuals (incorporating potential blocking) were calculated for each datum, i.e. for an annual measurement

$$r_i = \frac{F_i - \hat{f}(\theta_i)}{\sqrt{\sigma_i^2 + v(\theta_i)^2}},$$

where  $v(\theta_i)$  is the posterior standard deviation on the calibration curve at  $\theta_i$  cal BP, the calendar age of that datum. These residuals were then combined for each dataset  $\mathcal{K}$  into a scaled deviance,  $Z_{\mathcal{K}} = \sum_{i \in \mathcal{K}} r_i^2$ , and compared with a  $\chi_{n_{\mathcal{K}}}^2$  where  $n_{\mathcal{K}}$  are the number of determinations in that set. The mean offset of each dataset from the preliminary curve was also calculated  $\nu_{\mathcal{K}} = \frac{1}{n_{\mathcal{K}}} \sum_{i \in \mathcal{K}} \{F_i - \hat{f}(\theta_i)\}$ . Sets which had low  $p$ -values for their deviance and high mean offsets were then discussed with the data set providers as to whether they should be included in the final IntCal or not (see Bayliss et al. 2020 in this issue, for more details).

##### **Run Length**

The MCMC was run for 50,000 iterations. The first 25,000 of these iterations were discarded as burn-in. The remaining 25,000 were used to create the final curve (thinned to every 10th) and passed to the older part of the curve to create a seamless merging between the curve sections as discussed in Section 4.4.1. Since the main update steps (i.e.  $\beta$  and  $\lambda$ ) are Gibbs this was felt to be sufficient although convergence was further assessed by initializing the sampler at different values and comparing the resultant curve estimates visually.

#### 4.4 Creating the Older Part of the Curve

Beyond approximately 14 cal kBP, the number of tree-ring determinations decreases and they are not sufficient to create a precise calibration curve. We therefore merge the predominantly dendrodated curve with the older section of curve, which is estimated from a wider range of  $^{14}\text{C}$  material, at 13,913 cal BP. In this older period, the underlying data used to construct the curve incorporates corals, foraminifera, macrofossils and speleothems, in addition to a small number of floating tree-ring sequences. These alternative sources of data are typically not direct measurements of the atmosphere but instead are offset due to marine reservoir effects and dead carbon fractions. Further, their true calendar ages are quite uncertain and can only be estimated. They therefore present several further challenges for construction of an atmospheric curve.

##### 4.4.1 Modifications to Basic Model

###### **Uncertain Calendar Ages**

All our data in the older part of the curve have estimates (either in the form of noisy observations or priors) for their calendar ages rather than absolutely known values. For the corals and speleothems these estimates are obtained via U-Th dating and are considered independent; for the Cariaco Basin they are either provided by varve counting or elastic palaeoclimate tie-pointing (Heaton et al. 2013); similar elastic palaeoclimate tie-pointing provides the ages for the Pakistan and Iberian margins; for Lake Suigetsu they are provided by wiggle matching; while the floating tree-ring sequences have internally known chronologies but no absolute age estimates. We assume all our calendar age estimates are (potentially multivariate) normal. For a dataset  $\mathcal{K}$  on which we have noisy observations  $\mathbf{T}_{\mathcal{K}}$  of its calendar ages e.g. U-Th dated corals and the varve counted section of Cariaco, we model:

$$\mathbf{T}_{\mathcal{K}} \sim \text{MVN}(\boldsymbol{\theta}_{\mathcal{K}}, \Psi_{\mathcal{K},T}) \text{ so that } [\mathbf{T}_{\mathcal{K}}|\boldsymbol{\theta}_{\mathcal{K}}] \propto \exp\left\{-\frac{1}{2}\left\{(\mathbf{T}_{\mathcal{K}} - \boldsymbol{\theta}_{\mathcal{K}})^T \Psi_{\mathcal{K},T}^{-1}(\mathbf{T}_{\mathcal{K}} - \boldsymbol{\theta}_{\mathcal{K}})\right\}\right\}.$$

Here  $\Psi_{\mathcal{K},T}$  is the pre-specified and fixed covariance matrix that encodes the dependence within the calendar age observations for that dataset. For such datasets we place an uninformative prior on  $\boldsymbol{\theta}_{\mathcal{K}}$ . For Lake Suigetsu and the datasets where calendar age estimates are obtained via palaeoclimate tie-pointing, no actual calendar ages were observed but rather we have a prior on their true values:

$$[\boldsymbol{\theta}_{\mathcal{K}}] \propto \exp\left\{-\frac{1}{2}\left\{(\mathbf{T}_{\mathcal{K}} - \boldsymbol{\theta}_{\mathcal{K}})^T \Psi_{\mathcal{K},T}^{-1}(\mathbf{T}_{\mathcal{K}} - \boldsymbol{\theta}_{\mathcal{K}})\right\}\right\},$$

and the  $\mathbf{T}_{\mathcal{K}}$  and  $\Psi_{\mathcal{K},T}$  now represent our prior mean and covariance for the calendar ages. As we have no datasets on which we have both priors and noisy calendar age observations, these two types of calendar age estimate become equivalent for the purposes of updating. Details of how to construct the various covariance matrices for the case of varve counting and wiggle matching can be found in Niu et al. (2013); for the datasets with priors obtained by tie-pointing in Heaton et al. (2013); and for Lake Suigetsu in Bronk Ramsey et al. (2020 in this issue).

###### **Offsets: Reservoir Ages and Dead Carbon Fractions**

Several of our datasets do not provide direct measurements of the atmosphere but are instead offset. The speleothem records contain carbon obtained from dripwater which has passed through old limestone and so is a mixture of atmospheric  $\text{CO}_2$  and dissolved old (dead) carbon that has no  $^{14}\text{C}$  activity. The offset in radiocarbon age this creates is called a dead

carbon fraction (dcf) and is specific to the speleothem. Similarly the marine records have an atmospheric offset called a marine reservoir age (MRA) that arises due to both the limited gas exchange with the atmosphere and ocean circulation drawing up deep old carbon. These MRAs are location specific and vary over time. For both of these types of data we can incorporate the offset (either dcf or MRA) in the radiocarbon age domain as:

$$h_{\text{offset}}(\theta) = h_{\text{atmos}}(\theta) + r_{\mathcal{K}}(\theta),$$

where  $h_{\text{offset}}(\theta)$  is the radiocarbon age at time  $\theta$  cal BP in the offset environment;  $h_{\text{atmos}}(\theta)$  our atmospheric calibration curve in the radiocarbon age domain; and  $r_{\mathcal{K}}(\theta)$  our offset. This alters our observational model so that, in the  $F^{14}\text{C}$  domain, for determinations from dataset  $\mathcal{K}$  and with  $F^{14}\text{C}$  domain calibration curve  $f(\theta)$ , the offset becomes a multiplier,

$$F_i = e^{-r_{\mathcal{K}}(\theta_i)/8033} f(\theta_i) + \varepsilon_i.$$

These offsets must be estimated within curve construction to adaptively synchronize the different records. For IntCal20, speleothem dcf's were considered to be approximately constant over time but with an unknown level. MRAs were modeled as time-varying, providing a step forward from IntCal09 and IntCal13. Initial MRA estimates for each of our locations were obtained by creating a preliminary atmospheric calibration curve using the same Bayesian spline methodology but based only on the Hulu Cave record (Southon et al. 2012; Cheng et al. 2018). This Hulu-based curve was then used as a forcing for an enhanced Hamburg Large Scale Geostrophic (LSG) ocean general circulation model to provide estimates  $\delta_{\mathcal{K}}(\theta)$  of MRA at each given location (see Butzin et al. 2020 in this issue, for details). Due to coastal effects, these LSG estimates were considered to define the shape of the MRA, i.e. relative changes over time, but subject to a constant potential coastal shift. Ideally we might cycle the process of building the curve and re-estimating MRAs several times but the LSG model had a run time of several weeks and so this was not feasible. Notwithstanding the above attempts to accurately model dcf and MRA, we recognized that there was further variation in the offsets over time we were not able to fully capture. This was incorporated through the addition of further independent variability in the offsets from one calendar age to the next. Consequently our model for the offsets is:

$$\begin{aligned} \text{Speleothems(dcf's)} : r_{\mathcal{K}}(\theta_i) &\sim N(\nu_{\mathcal{K}}, \zeta_{\mathcal{K}}^2), \\ \text{Marine records(MRAs)} : r_{\mathcal{K}}(\theta_i) &\sim N(\nu_{\mathcal{K}} + \delta_{\mathcal{K}}(\theta_i), \zeta_{\mathcal{K}}^2), \end{aligned}$$

where  $\zeta_{\mathcal{K}}$  is the further independent variability in offsets from year-to-year around our LSG/constant dcf model and added to data before curve construction. We place priors on the constant dcf mean/coastal shifts  $\nu_{\mathcal{K}}$ :

$$\pi(\nu_{\mathcal{K}}) \sim N(\rho_{\mathcal{K}}, \omega_{\mathcal{K}}^2).$$

We do not aim to estimate the precise values of  $r_{\mathcal{K}}(\theta_i)$  during curve construction. However, the parameter  $\mathbf{v}$ , containing the values of  $\nu_{\mathcal{K}}$  which determine the mean dcf/MRA offset for each dataset, is updated within our MCMC sampler.

To fully specify our offset model, values of the prior mean dcf/coastal shift,  $\rho_{\mathcal{K}}$ , for each dataset; and  $\omega_{\mathcal{K}}$ , our uncertainty on the level of this shift, were estimated from the overlap between additional data available for each speleothem/marine location and the dendrodated curve from 0–14 cal kBP. Similarly, these overlaps provided an estimate for each  $\zeta_{\mathcal{K}}$ . See Section 5.2 for more details. Finally, for the two floating Southern Hemisphere kauri trees we assumed an offset of  $43 \pm 23$   $^{14}\text{C}$  yrs ( $1\sigma$ ) based upon the North-South hemispheric

offset estimated in SHCal13 (Hogg et al. 2013). Being direct measurements of the NH atmosphere, both Suigetsu and the three Bølling-Allerød floating tree-ring sequences are not offset, i.e. for all  $\theta$ ,  $r_{\mathcal{K}}(\theta) = 0$ .

**Offsets: Cariaco Basin**

The LSG model was not able to adequately resolve the MRA within the geographically unique Cariaco Basin which has a shallow sill that potentially limits exchange with the wider ocean. A further model for the MRAs in this location was therefore needed. As it covered a short time period, the MRA for the Cariaco varved record (Hughen et al. 2004) was modeled as for the speleothems, i.e. independently varying around a constant level. For the Cariaco unvarved record (Hughen and Heaton 2020 in this issue), we modeled the  $F^{14}\text{C}$  domain multiplier corresponding to any offset as

$$\gamma_{\mathcal{K}}(\theta) = e^{-r_{\mathcal{K}}(\theta)/8033} = \sum_{k=1}^{30} \beta_{\mathcal{C},k} B_{\mathcal{C},k}(\theta),$$

a further Bayesian spline with 30 knots placed at jittered quantiles of the Cariaco prior calendar ages. The value of  $\beta_{\mathcal{C}} = (\beta_{\mathcal{C},1}, \dots, \beta_{\mathcal{C},30})^T$  determining this spline was also estimated during curve construction but with a fixed and large smoothing parameter  $\lambda_{\mathcal{C}}$ . This approach meant rapid changes in the  $^{14}\text{C}$  determinations within Cariaco were considered as atmospheric signal while smoother, longer term drifts away from the other data that might otherwise introduce spurious features into the curve, would be resolved as time-varying reservoir effects.

**Heavier Tailed Errors**

Due to the diversity of the datasets in this older time period, as well as the difficulty in incorporating all their potential complexities, we wished to create a curve that was not overly influenced by single determinations. We therefore permit the possibility of heavier tailed errors in the  $F^{14}\text{C}$  determinations for the data in this older section of curve extending beyond 13,913 cal BP. This is incorporated by introducing an error multiplier where for each observation we maintain

$$F_i = e^{-r_{\mathcal{K}}(\theta_i)/8033} f(\theta_i) + \varepsilon_i,$$

but with each determination’s reported uncertainty  $\sigma_i$  scaled according to a further parameter  $\kappa_i$

$$\varepsilon_i | \kappa_i \sim N\left(0, \frac{\sigma_i^2}{\kappa_i}\right), \quad \text{where } \kappa_i \sim Ga(\text{shape} = \frac{\varrho_{\mathcal{K}(i)}}{2}, \text{rate} = \frac{\varrho_{\mathcal{K}(i)}}{2}),$$

and  $\mathcal{K}(i)$  is the dataset of observation  $i$ . All observations  $i$  belonging to the same dataset  $\mathcal{K}$  therefore have the same shape and rate in the Gamma prior for their individual  $\kappa_i$ . This model, integrating out  $\kappa_i$ , equates to a Student’s  $t$ -distribution for an individual observation  $i$ ,

$$\varepsilon_i \sim t(\text{location} = 0, \text{scale} = \sigma_i, \text{df} = \varrho_{\mathcal{K}(i)}).$$

Further, for each dataset  $\mathcal{K}$ , we place an independent hyperprior on the value of its particular  $\varrho_{\mathcal{K}}$  parameter so that the observations belonging to that dataset are grouped,

$$\varrho_{\mathcal{K}} \sim Ga(\text{shape} = A_{\varrho}, \text{rate} = B_{\varrho}),$$

to capture potential differences between the various records. Consequently, if a dataset is seen to contain several outliers, the other determinations in that same set will also be treated with

more caution. We choose a subjective hyperprior encapsulating an expectation of a low level of heavy-tailed behavior by selecting, for each dataset,  $A_\varrho = 100$  and  $B_\varrho = \frac{1}{2}$ . We update both  $\boldsymbol{\kappa} = (\kappa_1, \dots, \kappa_n)^T$  and  $\boldsymbol{\varrho}$ , the values of the various  $\varrho_{\boldsymbol{\kappa}}$ , within our sampler.

### **Parallel Tempering**

A significant concern with the previous random walk approach was how well the MCMC sampler mixed. With this random walk approach, we only updated the calibration curve one calendar year at a time, conditional on the value at all other times, which restricted the ability of the curve estimate to move significantly. Furthermore, this update was performed by MH meaning we would frequently reject proposed updates. By switching to Bayesian splines, we are instead able to update, conditional on the current calendar ages, the entire curve through its spline coefficients simultaneously via Gibbs. These dual changes, to update the entire curve at once and to do so via Gibbs sampling as opposed to MH, hopefully begin to address mixing concerns. However, due to the uncertain calendar ages it is likely the posterior for the curve (and the true calendar ages) will remain multi-modal. In an attempt to overcome any remaining concerns over mixing, we also incorporate parallel tempering. In tempering, we run multiple chains concurrently. One of these MCMC chains has as its target our posterior of interest while the others have modified, higher temperature, targets. These higher temperature targets are typically flattened versions of the posterior of interest designed so that, as the temperature increases, the corresponding MCMC chain mixes more easily. We then propose swaps between the states of the multiple chains so that the chains that mix well (i.e. those that run at higher temperatures) improve the mixing of the chains which do not (i.e. our posterior of interest). We are free to choose the elements of the likelihood we wish to temper and by how much. We only temper the likelihood of the observed data, i.e.  $[\mathbf{F}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\beta}_C, \boldsymbol{\kappa}]$ ; and  $[\mathbf{T}|\boldsymbol{\theta}]$  (or equivalent prior  $[\boldsymbol{\theta}]$ ) since these were thought to be the main elements restricting mixing and doing so keeps the MCMC updates straightforward. Our coupled chains run with a pair of temperatures  $\alpha_1, \alpha_2 > 1$  with each chain sampling from a modified posterior,

$$\begin{aligned} [\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\nu}, \boldsymbol{\beta}_C, \boldsymbol{\kappa}, \boldsymbol{\varrho} | \mathbf{F}, \mathbf{T}]_{\alpha_1, \alpha_2} &\propto L_{\alpha_1, \alpha_2}(\boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \boldsymbol{\nu}, \boldsymbol{\beta}_C, \boldsymbol{\kappa}, \boldsymbol{\varrho} | \mathbf{F}, \mathbf{T}) \\ &= [\mathbf{T}|\boldsymbol{\theta}]^{1/\alpha_1} [\mathbf{F}|\boldsymbol{\theta}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\beta}_C, \boldsymbol{\kappa}]^{1/\alpha_2} [\boldsymbol{\beta}|\lambda][\boldsymbol{\nu}][\boldsymbol{\beta}_C][\lambda][\boldsymbol{\kappa}][\boldsymbol{\varrho}][\boldsymbol{\theta}]^{1/\alpha_1}, \end{aligned}$$

so higher temperatures give flatter posteriors. We show in Section 4.4.3 how these modifications are simply equivalent to increasing our observational noise and so we can maintain straightforward and fast updates for each chain. We run the chain at four temperatures, including the unmodified posterior of actual interest where  $\alpha_1 = \alpha_2 = 1$ , in parallel. We discuss the effect of tempering, and how it aids mixing, in more detail within the Supplementary Information.

### **Merging with Already Created Tree-Ring-Only Section of Curve**

We need to ensure we smoothly merge our older section of calibration curve with the independently created tree-ring-only curve described in Section 4.3. This is achieved by selection of an appropriate overlapping set of knots for the two sections. We first identify, in the tree-ring-only curve, the calendar age at which to create the join. This is determined to be the calendar age of the knot at which, due to the ending of the tree-ring determinations, the curve uncertainty on the tree-ring-only curve begins to increase significantly. For IntCal20, this was selected to be at 13,913 cal BP. This knot and the two knots either side that were used to create the tree-ring-only curve (five tree-ring-only knots in total) are then copied into the knot basis for the older portion of curve. On each update

of the spline coefficients for the older section of curve, we pass the three spline coefficients that relate to the value of a particular realization of the tree-ring-only section of curve at the join. The Gibbs update for the older spline coefficients is then performed conditional on these three tree-ring-only spline coefficients. This ensures that when the posterior realizations of the two curves are connected together for the final curve, each realization is itself a cubic spline with not only continuity but also smoothness over the transition between the two sections.

**Required Simplifications—Blocking and Over-Dispersion**

To make the creation of the older section of the curve computationally practical, we do not incorporate exact modeling of blocking or over-dispersion. The large number of uncertain calendar ages mean that including blocking is not feasible. Each update to  $\theta$  requires recalculation of the spline basis matrix  $B_\theta$ . Doing this at every year in  $\theta^A$ , i.e. every year represented in all the blocks, would be extremely slow. Instead, we consider each determination to represent the single year in its centre. This significantly reduces the number of calendar ages  $\theta$  at which we have to recalculate our spline basis matrix. This is likely to have little effect on the overall curve since the density of data in this older section is not sufficient to identify very fine-scaled features. Over-dispersion was also not included in modeling the older curve as we lack the required data density to reliably estimate it. Further, our previous estimate of over-dispersion related to tree-ring determinations only, as opposed to the broader range of <sup>14</sup>C sources used within the older section of curve. Instead our modification to permit heavier tailed uncertainties aims to deal with potential over-dispersion in curve estimation. Note however, that when calibrating atmospheric <sup>14</sup>C determinations against the curve in this older time period, we would still expect them to have similar potential additional sources of variation (beyond the laboratory quoted uncertainty) as seen in the predominantly dendrodated section. We therefore maintain predictive intervals for our final, older curve estimate by adding back in the over-dispersion estimated within the 0–14 cal kBP tree rings to the posterior curve estimate.

**4.4.2 Updating the MCMC**

We present the update scheme for the MCMC sampler using notation that refers to the full parameter space. In practice, several of the update steps are done dataset-by-dataset, for example calendar ages or offsets. The calculations required to perform such dataset-specific updates often only require consideration of the likelihood restricted to that particular dataset and so remain fast.

**Posterior**

Let us define, where subscripts denote the unknown (and updated) variables on which the elements depend:

$$\begin{aligned} \gamma_{\mathcal{K}}(\theta) &= e^{-r_{\mathcal{K}}(\theta)/8033} & \Gamma_{\nu, \beta_C, \theta} &= \text{diag}(\gamma_{\mathcal{K}(1)}(\theta_1), \dots, \gamma_{\mathcal{K}(n)}(\theta_n)), \\ C(\theta) &= \frac{1}{1000} e^{-\theta/8267} & C_\theta &= \text{diag}(C(\theta_1), \dots, C(\theta_n)), \\ E(\theta) &= e^{-\theta/8267} & \mathbf{e}_\theta &= (E(\theta_1), \dots, E(\theta_n))^T. \end{aligned}$$

Note that our atmospheric-offset multipliers  $\gamma_{\mathcal{K}}(\theta)$  are determined by either  $\nu_{\mathcal{K}}$ , the values of the corresponding mean dcf or coastal MRA shift, or  $\beta_C$  in the case of the Cariaco unvarved record. Then, considering each datum as representing a single year and letting  $\mathbf{F} = (F_1, \dots, F_n)^T$ , we can calculate for each dataset  $\mathcal{K}$  the offset-adjusted calibration curve:

$$f_{\mathcal{K}}(\theta) = \gamma_{\mathcal{K}}(\theta)[C(\theta)\mathbf{B}(\theta)^T\boldsymbol{\beta} + E(\theta)], \text{ and so}$$

$$\mathbf{F}|\boldsymbol{\theta}, \boldsymbol{\beta}, \nu, \boldsymbol{\beta}_C, \boldsymbol{\kappa} \sim MVN\left(\Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}}[C_{\boldsymbol{\theta}}\mathcal{B}_{\boldsymbol{\theta}}\boldsymbol{\beta} + \mathbf{e}_{\boldsymbol{\theta}}], \text{diag}\left(\frac{\sigma_i^2}{\kappa_i}\right)\right).$$

Our joint posterior then becomes

$$\begin{aligned} & \boldsymbol{\theta}, \boldsymbol{\beta}, \lambda, \nu, \boldsymbol{\beta}_C, \boldsymbol{\kappa}, \boldsymbol{\rho}|\mathbf{F}, \mathbf{T} \propto [\mathbf{T}|\boldsymbol{\theta}][\mathbf{F}|\boldsymbol{\theta}, \boldsymbol{\beta}, \nu, \boldsymbol{\beta}_C, \boldsymbol{\kappa}][\boldsymbol{\beta}|\lambda][\nu][\boldsymbol{\beta}_C][\lambda][\boldsymbol{\kappa}][\boldsymbol{\rho}][\boldsymbol{\theta}] \\ & \propto \prod_{\mathcal{K}} \left( \prod_{j \in \mathcal{K}} \frac{\sqrt{\kappa_j}}{\sigma_j} \right) \times \exp\left\{-\frac{1}{2}[(\mathbf{T} - \boldsymbol{\theta})^T \Psi_T^{-1}(\mathbf{T} - \boldsymbol{\theta})] + \right. \\ & \quad \left. \{(\mathbf{F} - \Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}}[C_{\boldsymbol{\theta}}\mathcal{B}_{\boldsymbol{\theta}}\boldsymbol{\beta} + \mathbf{e}_{\boldsymbol{\theta}}])^T \mathbf{W}_{\mathcal{K}}(\mathbf{F} - \Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}}[C_{\boldsymbol{\theta}}\mathcal{B}_{\boldsymbol{\theta}}\boldsymbol{\beta} + \mathbf{e}_{\boldsymbol{\theta}}])\} + \sum_{\mathcal{K}} \frac{(\nu_{\mathcal{K}} - \rho_{\mathcal{K}})^2}{\omega_{\mathcal{K}}^2} \right. \\ & \quad \left. + \lambda \boldsymbol{\beta}^T \mathbf{D} \boldsymbol{\beta} + \lambda_C \boldsymbol{\beta}_C^T \mathbf{D}_C \boldsymbol{\beta}_C\right\} \times \lambda^{A + \text{rank}(\mathbf{D})/2 - 1} \times \exp\left(-\frac{\lambda}{B}\right) \\ & \quad \times \prod_{\mathcal{K}} \left( \prod_{j \in \mathcal{K}} \frac{\frac{\theta_{\mathcal{K}}}{2}}{\Gamma(\frac{\theta_{\mathcal{K}}}{2})} \kappa_j^{\frac{\theta_{\mathcal{K}}}{2} - 1} e^{-\frac{\theta_{\mathcal{K}}}{2} \kappa_j} \right) \times \prod_{\mathcal{K}} \frac{B_{\theta}^{A_{\theta}}}{\Gamma(A_{\theta})} \varrho_{\mathcal{K}}^{A_{\theta} - 1} e^{-B_{\theta} \varrho_{\mathcal{K}}}, \end{aligned}$$

where  $\mathbf{W}_{\boldsymbol{\kappa}} = \text{diag}\left(\frac{\sigma_1^2}{\kappa_1}, \dots, \frac{\sigma_n^2}{\kappa_n}\right)^{-1}$  includes the precision multiplier representing the heavy tails, and  $\Psi_T$  is the combined covariance matrix for all the calendar ages. As for the predominantly dendrodated section of curve, we can sample from this using a Metropolis-within-Gibbs algorithm, and again much of the updating can be performed directly via Gibbs sampling.

**Gibbs updating the calibration curve  $\boldsymbol{\beta}|\mathbf{F}, \boldsymbol{\theta}, \lambda, \nu, \boldsymbol{\beta}_C, \boldsymbol{\kappa}$ .**

This update is very similar to the equivalent update in the tree-ring-based section of curve. Letting  $\mathbf{F}^* = \mathbf{F} - \Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}}\mathbf{e}_{\boldsymbol{\theta}}$  then, were it not for our need to create a smooth transition from the trees, the posterior

$$\boldsymbol{\beta}|\mathbf{F}^*, \lambda, \boldsymbol{\theta}, \nu, \boldsymbol{\kappa} \sim MVN(\boldsymbol{\mu}_{\boldsymbol{\beta}} = \mathbf{Q}\mathcal{B}_{\boldsymbol{\theta}}^T C_{\boldsymbol{\theta}} \Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}} \mathbf{W}_{\boldsymbol{\kappa}} \mathbf{F}^*, \mathbf{Q}),$$

where  $\mathbf{Q} = (\mathcal{B}_{\boldsymbol{\theta}}^T [C_{\boldsymbol{\theta}} \Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}} \mathbf{W}_{\boldsymbol{\kappa}} \Gamma_{\nu, \boldsymbol{\beta}_C, \boldsymbol{\theta}} C_{\boldsymbol{\theta}}] \mathcal{B}_{\boldsymbol{\theta}} + \lambda \mathbf{D})^{-1}$  and, due to our dropping of blocking, all the matrices in the square brackets are diagonal so can be multiplied rapidly. To create the smooth transition, we draw a realization from the, already estimated, tree-ring-based section of curve and condition on  $\boldsymbol{\beta}^R$ , the value of the three basis coefficients affecting the value of the spline at the joining knot. Then

$$\boldsymbol{\beta}^{-R}|\boldsymbol{\beta}^R = \mathbf{a} \sim MVN(\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}}, \bar{\mathbf{Q}}),$$

where  $\bar{\boldsymbol{\mu}}_{\boldsymbol{\beta}} = \boldsymbol{\mu}_{\boldsymbol{\beta}}^{(-R)} - \mathbf{Q}_{-R,R} \mathbf{Q}_{R,R}^{-1}(\mathbf{a} - \boldsymbol{\mu}_{\boldsymbol{\beta}}^R)$ , and  $\bar{\mathbf{Q}}$  is the Schur complement of  $\mathbf{Q}_{R,R}$  in  $\mathbf{Q}$ . Here, for example  $\mathbf{Q}_{R,R}$  corresponds to the elements of the covariance that relate to the realized tree-ring based spline coefficients we condition on.

**MH updating the calendar ages  $\boldsymbol{\theta}|\mathbf{T}, \mathbf{F}, \boldsymbol{\beta}, \nu, \boldsymbol{\beta}_C, \boldsymbol{\kappa}$**

This step is as for the update of the wiggle matched trees. Whether a dataset has a prior on their calendar ages or noisy observations does not affect the update. For each dataset  $\mathcal{K}$  in turn we:

- Propose  $\theta_{\mathcal{K}'} \sim MVN(\theta_{\mathcal{K}}, \Sigma_{\text{prop}})$ ;
- Calculate  $\mathcal{B}_{\theta}$  and  $\mathcal{B}_{\theta'}$  (and similarly recalculate  $\Gamma_{\nu, \beta_C, \theta'}$ ,  $C_{\theta'}$  and  $E_{\theta'}$  at  $\theta_{\mathcal{K}'}$ ). Accept according to Hastings Ratio:

$$HR = \min \left\{ 1, \frac{\pi(\theta_{\mathcal{K}'} | \mathbf{T}_{\mathcal{K}}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\kappa})}{\pi(\theta_{\mathcal{K}} | \mathbf{T}_{\mathcal{K}}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\kappa})} \right\},$$

where

$$\pi(\theta_{\mathcal{K}} | \mathbf{T}_{\mathcal{K}}, \mathbf{F}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\kappa}) = \exp \left\{ -\frac{1}{2} \left[ \left\{ (\mathbf{T}_{\mathcal{K}} - \theta_{\mathcal{K}})^T \Psi_{\mathcal{K}, T}^{-1} (\mathbf{T}_{\mathcal{K}} - \theta_{\mathcal{K}}) \right\} + \left\{ (\mathbf{F} - \Gamma_{\nu, \beta_C, \theta} [C_{\theta} \mathcal{B}_{\theta} \boldsymbol{\beta} + \mathbf{e}_{\theta}])^T \mathbf{W}_{\boldsymbol{\kappa}} (\mathbf{F} - \Gamma_{\nu, \beta_C, \theta} [C_{\theta} \mathcal{B}_{\theta} \boldsymbol{\beta} + \mathbf{e}_{\theta}]) \right\} \right] \right\}.$$

Note that we only need to recalculate the elements in the likelihood of  $\mathbf{F}$  (e.g. the elements of  $\mathcal{B}_{\theta}$ ) that relate to  $\theta_{\mathcal{K}}$ . In the case of independent error, i.e.  $T_j \sim N(\theta_j, \sigma_j^2)$ , we use a proposal standard deviation  $\sigma_{\text{prop}} = 2\sigma_j$  and we can accept/reject each proposal  $j = 1, \dots, n_{\mathcal{K}}$  independently. In the case of dependent error with a covariance matrix, i.e.  $\mathbf{T}_{\mathcal{K}} \sim MVN(\theta_{\mathcal{K}}, \Psi_{\mathcal{K}, T})$ , or an equivalent prior, we update the whole of  $\theta_{\mathcal{K}}$  simultaneously by sampling a proposal  $\theta' \sim MVN(\theta, \Sigma_{\text{prop}})$ , where  $\Sigma_{\text{prop}} \propto \Psi_{\mathcal{K}, T}$ , to obtain higher acceptance rates.

**MH updating the offsets  $\nu_{\mathcal{K}} | \mathbf{F}, \theta, \boldsymbol{\beta}, \lambda, \xi, \tau$**

Similarly, for each offset dataset:

- Propose a new constant dcf/MRA shift  $\nu_{\mathcal{K}'} \sim N(\nu_{\mathcal{K}}, \sigma_{\text{prop}})$ ;
- Calculate resultant  $F^{14}\text{C}$  multiplier  $\Gamma_{\nu', \beta_C, \theta}$ . Accept according to Hastings Ratio:

$$HR = \min \left\{ 1, \frac{\pi(\nu_{\mathcal{K}'} | \mathbf{F}, \theta, \boldsymbol{\beta}, \boldsymbol{\nu}^{-\mathcal{K}}, \boldsymbol{\kappa}) \varphi(\nu_{\mathcal{K}'}; \rho_{\mathcal{K}}, \omega_{\mathcal{K}}^2)}{\pi(\nu_{\mathcal{K}} | \mathbf{F}, \theta, \boldsymbol{\beta}, \boldsymbol{\nu}^{-\mathcal{K}}, \boldsymbol{\kappa}) \varphi(\nu_{\mathcal{K}}; \rho_{\mathcal{K}}, \omega_{\mathcal{K}}^2)} \right\},$$

where

$$\pi(\nu_{\mathcal{K}} | \mathbf{F}, \theta, \boldsymbol{\beta}, \boldsymbol{\nu}^{-\mathcal{K}}, \boldsymbol{\kappa}) = \exp \left\{ -\frac{1}{2} [(\mathbf{F} - \Gamma_{\nu, \beta_C, \theta} [C_{\theta} \mathcal{B}_{\theta} \boldsymbol{\beta} + \mathbf{e}_{\theta}])^T \mathbf{W}_{\boldsymbol{\kappa}} (\mathbf{F} - \Gamma_{\nu, \beta_C, \theta} [C_{\theta} \mathcal{B}_{\theta} \boldsymbol{\beta} + \mathbf{e}_{\theta}])] \right\}$$

and  $\varphi(\cdot; \rho_{\mathcal{K}}, \omega_{\mathcal{K}}^2)$  is a normal density with mean  $\rho_{\mathcal{K}}$  and variance  $\omega_{\mathcal{K}}^2$ .

As when updating the calendar ages for each dataset in turn, when calculating this Hastings Ratio we can restrict our attention to the dataset in question, as elements relating to other datasets will cancel, making this step considerably faster.

**Gibbs updating the Cariaco Basin unvarved MRA  $\beta_C | \mathbf{F}, \theta, \boldsymbol{\beta}, \lambda, \nu, \boldsymbol{\kappa}$**

Updating the spline MRA coefficients  $\beta_C$  for the Cariaco Basin is analogous to the update of the calibration curve coefficients  $\boldsymbol{\beta}$ . This can intuitively be seen by considering how these determinations are represented and swapping the offset multiplier  $\gamma_{\mathcal{K}}(\theta_{\mathcal{K}})$  with the calibration curve estimate  $f(\theta)$ , i.e. for determinations  $F_i$  belonging to the Cariaco Basin unvarved record,

$$F_i = \gamma_{\mathcal{K}(i)}(\theta_i) f(\theta_i) + \varepsilon_i = f(\theta_i) \gamma_{\mathcal{K}(i)}(\theta_i) + \varepsilon_i,$$

and noting that, given  $\beta$  and  $\theta$ , the curve estimates  $f(\theta_i)$  are known and  $\gamma_{\mathcal{K}}(\theta_i) = \sum_{k=1}^{30} \beta_{C,k} B_{C,k}(\theta_i)$ , a linear combination of the MRA spline coefficients  $\beta_C$ . More rigorously, this can be seen if we restrict our likelihood to solely the Cariaco determinations, and change the ordering of  $\Gamma_{\nu, \beta_C, \theta}$  and  $[C_{\theta} B_{\theta} \beta + \mathbf{e}_{\theta}]$ . Then, letting  $\mathcal{F}_{\mathcal{K}} = \text{diag}(C_{\theta_{\mathcal{K}}} B_{\theta_{\mathcal{K}}} \beta + \mathbf{e}_{\theta_{\mathcal{K}}})$  be a diagonal matrix representing the current atmospheric calibration curve at the Cariaco calendar ages, our posterior for the MRA spline  $\beta_C$  becomes

$$[\beta_C | F_{\mathcal{K}}] | \beta_C \propto \exp \left\{ -\frac{1}{2} [(\mathbf{F}_{\mathcal{K}} - \mathcal{F}_{\mathcal{K}} B_{C, \theta_{\mathcal{K}}} \beta_C)^T W_{\mathbf{K}} (\mathbf{F}_{\mathcal{K}} - \mathcal{F}_{\mathcal{K}} B_{C, \theta_{\mathcal{K}}} \beta_C) + \lambda_C \beta_C^T \mathbf{D}_C \beta_C] \right\}.$$

**Gibbs updating the smoothing parameter  $\lambda | \beta$**

As before,  $\lambda | \beta \sim Ga(A', B')$  where  $A' = A + \text{rank}(\mathbf{D})/2$  and  $B' = \left( \frac{\beta^T \mathbf{D} \beta}{2} + \frac{1}{B} \right)^{-1}$ .

**Gibbs updating the precision multipliers  $\kappa | \mathbf{F}, \beta, \theta, \nu, \beta_C, \varrho$**

For each datum,

$$\kappa_i | F_i, \beta, \theta_i, \gamma_{\mathcal{K}(i)} \sim Ga\left(\frac{\varrho_{\mathcal{K}(i)}}{2} + \frac{1}{2}, \frac{\varrho_{\mathcal{K}(i)}}{2} + \frac{R_i^2}{2\sigma_i^2}\right),$$

where  $R_i = F_i - \gamma_{\mathcal{K}(i)}(\theta_i) C(\theta_i) B(\theta_i)^T \beta - E(\theta_i)$  is the current residual, and  $\frac{\varrho_{\mathcal{K}(i)}}{2}$  is the current shape/rate prior for the dataset  $\mathcal{K}$  to which datum  $i$  belongs.

**MH updating the expected tail behavior  $\varrho_{\mathcal{K}} | \kappa$**

For each dataset  $\mathcal{K}$ :

- Propose new parameter  $\varrho_{\mathcal{K}'} \sim N(\varrho_{\mathcal{K}}, \sigma_{\text{prop}})$ ;
- Accept according to Hastings Ratio:

$$HR = \min \left\{ 1, \frac{\pi(\varrho_{\mathcal{K}'} | \kappa) \pi(\varrho_{\mathcal{K}'})}{\pi(\varrho_{\mathcal{K}} | \kappa) \pi(\varrho_{\mathcal{K}})} \right\},$$

where

$$\pi(\varrho_{\mathcal{K}} | \kappa) = \prod_{j \in \mathcal{K}} \frac{\frac{\varrho_{\mathcal{K}}}{2}}{\Gamma(\frac{\varrho_{\mathcal{K}}}{2})} \kappa_j^{\frac{\varrho_{\mathcal{K}}}{2} - 1} e^{-\frac{\varrho_{\mathcal{K}}}{2} \kappa_j} \quad \text{and} \quad \pi(\varrho_{\mathcal{K}}) = \frac{B_{\varrho}^{A_{\varrho}}}{\Gamma(A_{\varrho})} \varrho_{\mathcal{K}}^{A_{\varrho} - 1} e^{-B_{\varrho} \varrho_{\mathcal{K}}}.$$

**4.4.3 Additional Considerations**

**Choice of Splines**

The approach to creating the older part of the calibration curve is considerably more computationally intensive than for the predominantly dendrodated section of curve. In this older section, we have to update the calendar ages of the constituent determinations requiring recalculation of the value of the curve each time. This becomes less feasible as the number of knots increases. Consequently, for this older section, we are not able to choose such a large number of knots for our spline basis as used in the dendrodated section. To select our basis for the older section of curve, we therefore began by placing 400 knots at quantiles (with the addition of a small amount of random jitter) of the observed/prior calendar ages  $T_i$ . This meant that a knot was placed approximately every 5th observed calendar age, similar to the dendrodated section of curve. However, in order to make the

fitted placement of the three Bølling-Allerød floating tree-ring sequences more equitable, we overrode this knot choice in the period from 14.2–15.2 cal kBP. In this period, 50 knots were placed evenly (i.e. one every 20 calendar years) and all other knots were removed. Tree-ring sequences P305u and P317 in particular show considerable variability in <sup>14</sup>C over their course; see Adolphi et al. (2017) and Figure 7. If the knots used to model the calibration curve over the potential fitted range of these trees lie unevenly in calendar age, as would be the case if we maintained a knot placement based upon the observed quantiles of the other IntCal20 data, the highly variable <sup>14</sup>C nature of these tree-ring sequences may have meant that the method had a preference/bias to locate them where there were more knots as opposed to where they best fitted the other data. Conversely placing knots at even calendar age spacings aims to remove this potential bias. One knot every 20 calendar years is still sufficient to pick out required curve detail in this period. Finally, we added into the spline bases the 5 knots required for merging with the tree-ring section of curve and removed other knots in this period. In total, after these modifications, we were left with 444 knots meaning that the overall method still maintained computational feasibility. These knot locations then remained fixed throughout the sampler, i.e. did not change as the calendar ages  $\theta_i$  were updated.

**Updating with Parallel Tempering**

Tempering our target likelihood for the calendar age terms  $[\mathbf{T}|\boldsymbol{\theta}]$  (or alternatively the prior  $[\boldsymbol{\theta}]$ ) concerning a particular dataset  $\mathcal{K}$ , we choose

$$[\mathbf{T}_{\mathcal{K}}|\boldsymbol{\theta}_{\mathcal{K}}]^{1/\alpha_1} \propto \exp\left\{-\frac{1}{2\alpha_1} \left[ \{(\mathbf{T}_{\mathcal{K}} - \boldsymbol{\theta}_{\mathcal{K}})^T \Psi_{\mathcal{K},T}^{-1} (\mathbf{T}_{\mathcal{K}} - \boldsymbol{\theta}_{\mathcal{K}})\} \right]\right\}$$

This is equivalent to the untempered version but with an adjusted covariance  $\Psi_{\mathcal{K},T}(\alpha_1) = \alpha_1 \Psi_{\mathcal{K},T}$ . To run the chain at this temperature we simply adjust the calendar age covariance (for observed  $\mathbf{T}$  or prior) accordingly. We can then perform all updates as described in Section 4.4.2 but with this modified calendar age covariance.

In tempering the likelihood for the observed  $\mathbf{F}_{\mathcal{K}}$  for a particular dataset  $\mathcal{K}$ , we select

$$[\mathbf{F}_{\mathcal{K}}|\boldsymbol{\theta}_{\mathcal{K}}, \boldsymbol{\beta}, \boldsymbol{\nu}, \boldsymbol{\beta}_C, \boldsymbol{\kappa}]^{1/\alpha_2} \propto \left( \prod_{j \in \mathcal{K}} \frac{\sqrt{\kappa_j}}{\sqrt{\alpha_2} \sigma_j^2} \right) \times \exp\left\{-\frac{1}{2\alpha_2} (\mathbf{F}_{\mathcal{K}} - \Gamma_{\nu_{\mathcal{K}}, \boldsymbol{\theta}_{\mathcal{K}}} [C_{\boldsymbol{\theta}_{\mathcal{K}}} \mathcal{B}_{\boldsymbol{\theta}_{\mathcal{K}}} \boldsymbol{\beta} + \mathbf{e}_{\boldsymbol{\theta}_{\mathcal{K}}}] )^T \mathbf{W}_{\boldsymbol{\kappa}_{\mathcal{K}}} (\mathbf{F}_{\mathcal{K}} - \Gamma_{\nu_{\mathcal{K}}, \boldsymbol{\theta}_{\mathcal{K}}} [C_{\boldsymbol{\theta}_{\mathcal{K}}} \mathcal{B}_{\boldsymbol{\theta}_{\mathcal{K}}} \boldsymbol{\beta} + \mathbf{e}_{\boldsymbol{\theta}_{\mathcal{K}}}] ) \right\},$$

where for example, in a slight abuse of notation,  $\Gamma_{\nu_{\mathcal{K}}, \boldsymbol{\theta}_{\mathcal{K}}}$  denotes the diagonal offset multiplier matrix for the specific dataset  $\mathcal{K}$ , and will depend upon only  $\nu_{\mathcal{K}}$ . In the case of the Cariaco Basin unvarved record, we would instead have  $\Gamma_{\boldsymbol{\beta}_C, \boldsymbol{\theta}_{\mathcal{K}}}$ . This is also equivalent to the untempered version but modifying all the <sup>14</sup>C variances within dataset  $\mathcal{K}$  to be  $\sigma_j^2(\alpha_2) = \alpha_2 \sigma_j^2$  for all  $j \in \mathcal{K}$ . To perform MCMC on the chain with this temperature, we make this modification and then all update steps remain as in Section 4.4.2 but with the temperature-adjusted <sup>14</sup>C variances.

The amount of tempering applied is dependent upon the datasets. We apply most tempering to the two floating Bølling-Allerød tree-ring sequences P305u and P317 since their internal <sup>14</sup>C determinations are highly variable and we wish to fully explore their potential fitting location;

and also the Cariaco Basin unvarved record which has a large number of observations with high calendar age covariances where again we might otherwise be concerned about mixing. For these three datasets, we chose values of  $\alpha_2$  that were equivalent, at the highest temperature of the four parallel chains, to increasing the uncertainty  $\sigma_i$  on the  $F^{14}\text{C}$  determinations by  $\approx 36\%$  from the quoted laboratory values. For the other data, the highest  $\alpha_2$  temperature was equivalent to increasing  $\sigma_i$  by  $\approx 5\%$ . We performed the same, small amount of tempering for the calendar ages as otherwise the acceptance rate for swapping between chains became very low. For all the data, the highest  $\alpha_1$  temperature of the four chains was equivalent to increasing the calendar age covariance by  $\approx 10\%$ .

### **Run Time and Convergence**

We ran the four tempered chains in parallel, each for 250,000 iterations. Four independent swaps between the tempered chains were proposed every 5th iteration. Convergence was assessed by initializing the chains at several different starting points and then visually observing the robustness in the curve estimates; along with assessment of other outputs such as the posterior estimates for the calendar ages of the floating tree-ring sequences and the spline based MRA of the Cariaco Basin; and the overall model likelihood. For a more thorough discussion of convergence and the effect of tempering, see the Supplementary Information. The level of tempering chosen was based upon a trade-off between maintaining a reasonable acceptance rate of chain swaps, and a highest temperature that both allowed the Bølling-Allerød trees P305u and P317 to move and gave a “highest-temperature” curve estimate that had removed most fine detail; again see the Supplementary Information. In total, there were 4407 accepted chain swaps, with 559 between the two lowest temperatures, i.e. swaps that involved the untempered target of true interest. To create the final curve, we removed the first half of the target chain as burn-in and then thinned to every 50th realization leaving us with 2500 posterior realizations covering the older time-period from which to create the final complete curve.

### **4.5 Final Curve Output and Forming Predictive Intervals**

Due to our merging process, the posterior realizations from the two sections of the curve occurred in pairs with one another. Each realization from the older section had a matching realization from the tree-ring-only section. When combined, such a pair gives a complete curve realization extending from 0–55 cal kBP that is still, in itself, a cubic spline and hence smooth over the join between the two sections. After pairing these realizations, our MCMC provided 2500 plausible posterior realizations of complete  $\Delta^{14}\text{C}$  calibration curves, i.e.  $g_j(\theta)$  for  $j = 1, \dots, 2500$ , from 0–55 cal kBP.

Additionally, each of these complete curves has an accompanying posterior estimate for the constant of proportionality  $\tau$  in our (square-root) model for the over-dispersion/additional uncertainty linked with the specific tree-ring realization. To create the posterior mean and predictive variance IntCal summaries, each of these  $\Delta^{14}\text{C}$  curves,  $g_j(\theta)$  for  $j = 1, \dots, 2500$ , were transformed into the  $F^{14}\text{C}$  domain. For each corresponding  $f_j(\theta)$ , and each calendar year  $\theta$  on which output was desired, 50 predictive values of  $F^{14}\text{C}$  were created by adding back in the estimated over-dispersion corresponding to that particular curve realization. These predictive values therefore incorporate the potential additional variability (beyond the reported lab uncertainty) one might expect to see in a new tree-ring  $^{14}\text{C}$  determination to be calibrated:

$$f_{jk}(\theta) \sim N(f_j(\theta), \tau_{(j)}^2 f_j(\theta)) \quad \text{for } k = 1, \dots, 50.$$

Here  $\tau_{(j)}$  is the estimated constant of proportionality corresponding to curve realization  $j$ . These 125,000 predictive  $F^{14}\text{C}$  values (50 for each of the 2500 curve realizations) were transformed into radiocarbon ages and finally summarized by pointwise means and variances to form the published IntCal20. The difference between using the thinned set of 2500 realizations rather than the entire post-burn-in sample was negligible. However, this thinning allows us to retain a practical set of complete curve realizations for later independent use which would give the same calibrated ages of individual  $^{14}\text{C}$  determinations as the published IntCal20 summaries. Each of these 2500 realizations has been stored for potential future extraction of covariance information and to enable more precise age modeling should it be desired.

## 5 ADDITIONAL INFORMATION PROVIDED BY A BAYESIAN APPROACH

### 5.1 Over-Dispersion

#### 5.1.1 Choice of Model for Tree-Ring Over-Dispersion and Prior

To select both the appropriate scaling model for the level of over-dispersion present within tree-ring  $^{14}\text{C}$  determinations and a suitable prior on its size, we investigated the over-dispersion seen in comparisons of same-tree-ring determinations made across laboratories taken from SIRI (the sixth radiocarbon laboratory intercomparison exercise of Scott et al. 2017). This SIRI data consisted of repeat measurements of trees from two distinct time periods—3 trees from a period about 300  $^{14}\text{C}$  yrs BP; and 2 trees from a period about 10,000  $^{14}\text{C}$  yrs BP. Each of the trees had about 80 such repeat measurements. For each tree we fitted a Bayesian model

$$F_i = f(\theta_i) + \varepsilon_i + \eta_i,$$

where  $\varepsilon_i \sim N(0, \sigma_i^2)$  is the laboratory reported uncertainty and  $\eta_i \sim N(0, \tau_i^2)$  is the additive error which models the potential over-dispersion present in our  $^{14}\text{C}$  observations. We considered three potential scaling models for the form of  $\tau_i$ , the standard deviation of the additional uncertainty on  $F_i$ : a constant model where  $\tau_i = \tau$  for all  $i$ ; a model where it scales linearly with the underlying value of  $F^{14}\text{C}$ , i.e.  $\tau_i = \tau f(\theta_i)$ ; and a model where it scales with the square-root of  $F^{14}\text{C}$ , i.e.  $\tau_i = \tau \sqrt{f(\theta_i)}$ .

To investigate the suitability of each potential scaling model, we fitted them to the recent and the older sets of SIRI trees separately and compared the posterior estimates for the constant of proportionality  $\tau$  obtained in each. The most suitable scaling model was taken as the one that provides the same posterior for this parameter  $\tau$  when the model is fitted to the older trees as for when it is fitted to the recent trees. The results of fitting all three scaling models can be seen in the Supplementary Information. The optimal model was clearly seen to be the square-root scaling. Figure 4 provides the posterior  $\tau$  estimates under this square-root scaling model on each of the two time periods separately; as we can see the estimates of the proportionality parameter  $\tau$  we get under this scaling model are fairly similar and show significant levels of over-dispersion in the SIRI data.

This square-root scaling model for the additional uncertainty, i.e.  $\eta_i \sim N(0, \tau^2 f(\theta_i))$  was therefore incorporated into curve construction. To provide a prior on  $\tau$ , the constant of proportionality, to take forward into IntCal the SIRI data were combined before the over-dispersion model was refitted to them. The posterior mean and variance of the resultant joint estimate for  $\tau$  were then used to provide our prior:

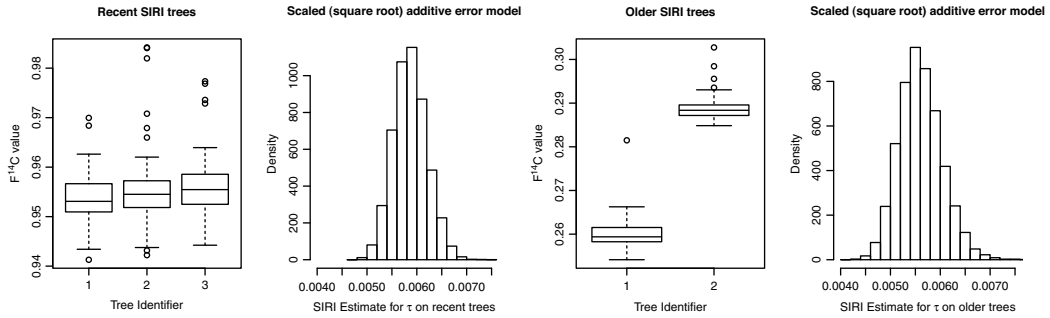


Figure 4 Fitting scaled over-dispersion model to SIRI (with the standard deviation of additive uncertainty scaling proportionally to  $\sqrt{f(\theta)}$ ) on each time period separately—Boxplots of observed  $F_i$  values in each time period and posterior histogram of  $\tau$  under model where additive error  $\eta_i \sim N(0, \tau^2 f(\theta_i))$ . The left pair of plots show the value of  $\tau$  when run on the three younger trees (from around 300  $^{14}C$  yrs BP); and the right two plots the posterior value of  $\tau$  when run on the two older trees (from around 10,000  $^{14}C$  yrs BP). If the posterior estimates of  $\tau$  in the two periods are similar then this suggests the model is appropriate.

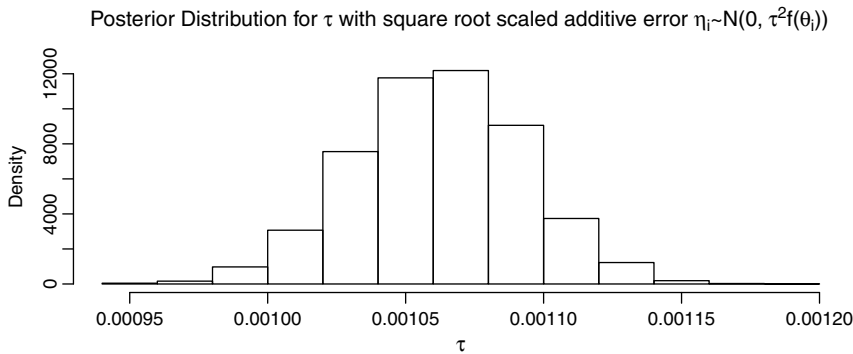


Figure 5 The posterior for the level of over-dispersion in the tree-ring determinations within the IntCal data. We model the additional variation, in the  $F^{14}C$  domain, as  $\eta_i \sim N(0, \tau^2 f(\theta_i))$ .

$$\pi(\tau) \sim N(0.0056, 0.00045^2).$$

Note that this SIRI data does not itself enter into IntCal and so provides a genuinely independent prior on  $\tau$ .

### 5.1.2 Posterior

In Figure 5 we plot the posterior estimate for the constant of proportionality  $\tau$  after curve construction. Due to the large volume of IntCal data, this posterior is largely determined by the IntCal tree-ring determinations which dominates the SIRI-informed prior. This posterior estimate incorporates both additional variation between the IntCal laboratories and other potential sources of variation in the  $^{14}C$  measurement of tree rings such as location and species. This IntCal-based posterior for the level of over-dispersion is considerably smaller than the SIRI prior, approximately one fifth of the size. This is to be expected since the IntCal data is screened and the SIRI data covered a large number of laboratories. The difference between the SIRI-informed prior and the IntCal-based

posterior does however give some indication as to the much greater range of over-dispersion that might exist within uncalibrated determinations and emphasizes the need for laboratories to engage with intercomparison exercises and provide appropriate assessments of measurement uncertainty before determinations are calibrated against IntCal20 in order to avoid over-confident dating.

## 5.2 MRAs and Dead Carbon Fractions

### 5.2.1 Priors on Marine and Dead Carbon Fraction Offsets

For each marine and speleothem dataset, the offset to the Northern Hemisphere atmosphere in terms of radiocarbon age was modeled as

$$\begin{aligned} \text{Speleothems (dcfs)} : r_{\mathcal{K}}(\theta_i) &\sim N(\nu_{\mathcal{K}}, \zeta_{\mathcal{K}}^2), \\ \text{Marine records (MRAs)} : r_{\mathcal{K}}(\theta_i) &\sim N(\nu_{\mathcal{K}} + \delta_{\mathcal{K}}(\theta_i), \zeta_{\mathcal{K}}^2), \\ &\text{with a prior } \pi(\nu_{\mathcal{K}}) \sim N(\rho_{\mathcal{K}}, \omega_{\mathcal{K}}^2). \end{aligned}$$

Here  $\delta_{\mathcal{K}}(\theta)$  are our location-specific MRA estimates obtained from the LSG model in its ‘‘GS’’ scenario (Butzin et al. 2020 in this issue) forced by a preliminary estimate of atmospheric  $\Delta^{14}\text{C}$  using the same Bayesian spline and errors-in-variables approach as IntCal20 but based only upon data from Hulu Cave. To complete our offset model, we are required to specify  $\zeta_{\mathcal{K}}$ , the variability in MRA/def offset from one calendar year to the next beyond that already present in the LSG/constant dcf model. We must also provide suitable values for  $\rho_{\mathcal{K}}$  and  $\omega_{\mathcal{K}}$  which determine our prior on  $\nu_{\mathcal{K}}$ , the mean dcf/coastal shift. Note we only place a prior on this shift  $\nu_{\mathcal{K}}$ , its precise value will be updated during curve construction.

Most of the offset datasets have additional younger observations in the IntCal database which overlap with the tree-ring-only part of the calibration curve from 0–14 cal kBP that has already been created. These, more recent, offset observations are not themselves used to create the IntCal curve but, by comparison with these overlapping atmospheric tree-rings, can independently inform us on suitable values for these three parameters. We therefore fitted statistical models for the offset of the form above and estimated the various components for each set. In Figure 6 we show the results for both the Hulu Cave H82 speleothem (Southon et al. 2012) and the Kiritimati corals (Fairbanks et al. 2005). For H82, we see that the estimated mean dcf value in this period of overlap is approximately  $480 \pm 8$   $^{14}\text{C}$  yrs ( $1\sigma$ ) and that we need to add a further independent uncertainty of  $\approx 50$   $^{14}\text{C}$  yrs to the determinations to make them consistent with the dendrodated trees. For the H82 speleothem data used for the older part of the curve, we therefore select  $\rho_{\mathcal{K}} = 480$ ,  $\omega_{\mathcal{K}} = 8$  and  $\zeta_{\mathcal{K}} \approx 50$   $^{14}\text{C}$  yrs. For the Kiritimati corals, where we have data which overlap with atmospheric trees, we need to shift the LSG estimate down by approximately  $280 \pm 11$   $^{14}\text{C}$  yrs ( $1\sigma$ ) and add a similar independent uncertainty of  $\approx 50$   $^{14}\text{C}$  yrs to the determinations, i.e. we would select  $\rho_{\mathcal{K}} = -280$ ,  $\omega_{\mathcal{K}} = 11$  and  $\zeta_{\mathcal{K}} \approx 50$   $^{14}\text{C}$  yrs. Comparing the plotted red curves, representing the constant dcf model for Hulu and the shifted LSG output for Kiritimati, we see reasonable agreement with the raw data suggesting our approach to modeling the offsets is appropriate. These comparisons of overlap also give some insight into how different types of indirect records might perform as recorders of the atmospheric  $^{14}\text{C}$ , and where more modeling might be needed to understand the processes they undergo.

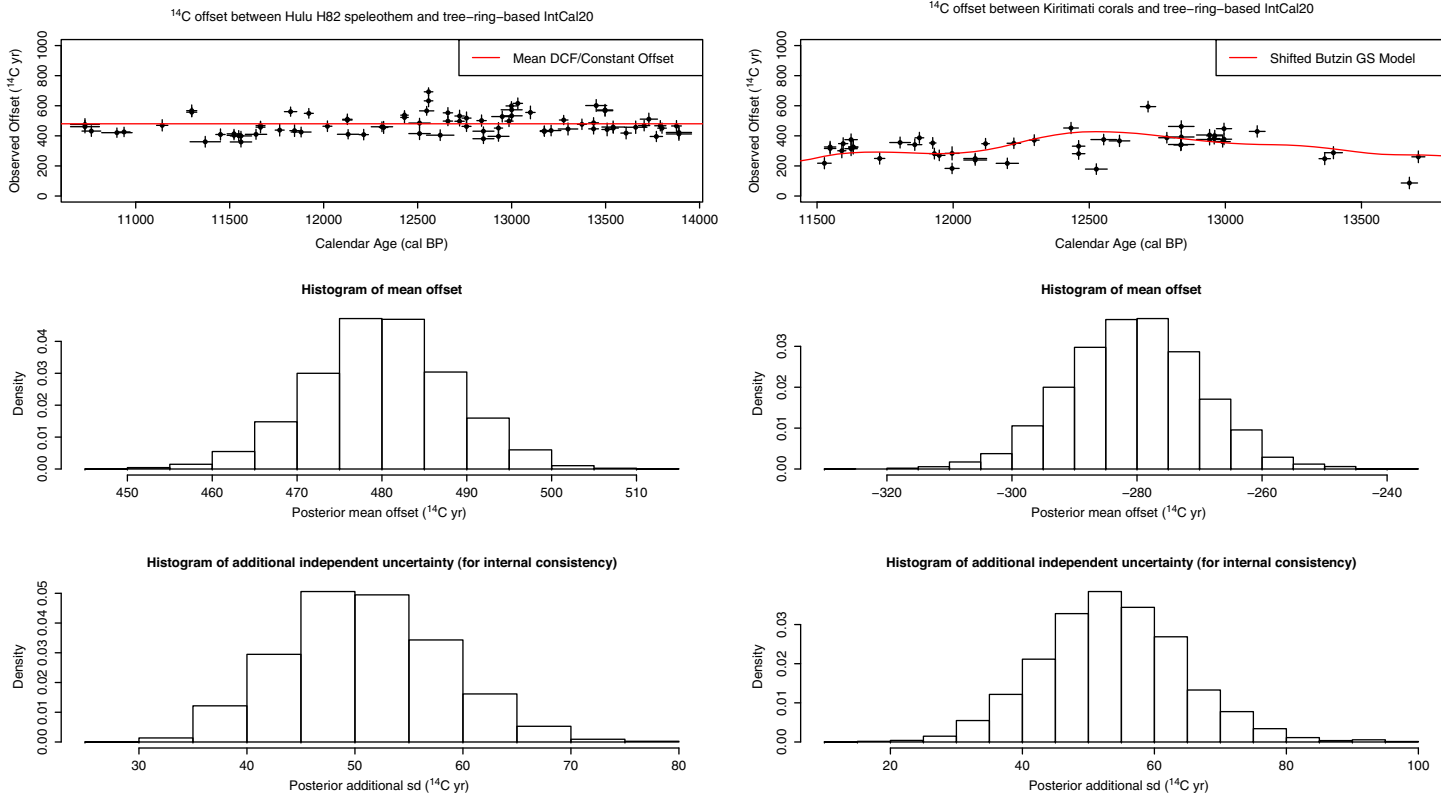


Figure 6 Estimating the offsets of Hulu speleothem H82 and Kiritimati corals to the IntCal20 tree-ring-based curve—The top plots show the observed offsets to the IntCal20 curve where it is based upon tree rings only. The middle plots show the estimate for  $\nu_{\mathcal{K}}$ , the mean dcf/coastal shift, based on these data. The mean and variance of these values,  $\rho_{\mathcal{K}}$  and  $\omega_{\mathcal{K}}^2$ , determine our prior on  $\nu_{\mathcal{K}}$ . The bottom plots show an estimate of the additional variability needed for the determinations to be consistent with the tree rings. We set  $\zeta_{\mathcal{K}}$  to be the mean of these values.

### **Datasets with No Overlapping Data**

Some datasets did not have any  $^{14}\text{C}$  determinations that were more recent than 14 cal kBP and so no information was available on overlap with the tree rings. For the Mururoa corals (Bard et al. 1998) and the Iberian and Pakistan Margins (Bard et al. 2013), we considered that the additional uncertainty  $\zeta_{\mathcal{K}}$ , to account for MRA variability beyond that seen in LSG, was 150  $^{14}\text{C}$  yrs to match coral data on which we did have overlap. We also placed an uninformative prior on the LSG coastal shift  $\nu_{\mathcal{K}}$ . This keeps the internal shape of the dataset but allows it to jointly move up/down together to fit the other records. The other two Hulu Cave speleothems (MSD and MSL in Cheng et al. 2018) were modeled as having separate and independent dcfs from the H82 speleothem but each having the same prior on its mean and the same additional variability  $\zeta_{\mathcal{K}}$ .

### **Splitting the Hoffman Speleothem**

The Hoffman speleothem (Hoffmann et al. 2010) consists of two separated sections—a set of relatively young data for which there was some overlap with the trees, and then a set of much older measurements. These were split in curve creation and a separate dcf was applied to the older measurements. For the more recent section, the prior for the mean dcf  $\nu_{\mathcal{K}}$  was based on the overlapping determinations, while an uninformative prior was placed on the mean dcf for the older measurements. Again this keeps the internal shape/structure but allows the dataset to jointly move up/down to fit best with the other material.

## **5.3 Curve Realizations**

Each iteration of our MCMC sampler generates a curve realization, i.e. what it believes is a plausible record of atmospheric  $^{14}\text{C}$  history. To create the published IntCal20 posterior mean and variance for the calibration curve a large set of these individual realizations are summarized/averaged on a calendar year by calendar year (pointwise) basis. In creating these pointwise summaries we lose potentially valuable curve information, in particular on the dependence in the value of the curve from one calendar age to the next. When we calibrate against IntCal20 using its pointwise means and variances we assume that the value of the curve is independent from one calendar year to the next, i.e. that the curve could potentially switch from being towards the top of its probability interval in one calendar year, to being towards the bottom in the next. For the calibration of single  $^{14}\text{C}$  determinations, losing such covariance information has no effect upon the calibrated age estimate one obtains. However, covariance may matter when one is jointly calibrating multiple determinations, for example when wiggle matching or incorporating them into a more complex model. In such instances, one may therefore wish to consider the individual realizations since they do contain covariance information. Furthermore, and importantly, the pointwise IntCal20 posterior mean should not be viewed in itself as a realization, i.e. it is not itself a potential  $^{14}\text{C}$  history, but rather an average of many.

We illustrate the difference between an individual posterior realization and the pointwise summarized values in Figure 7. Here we show, in color, four individual curve realizations compared to the summarized pointwise mean values (typically what is reported as the IntCal curve) with their 95% predictive intervals plotted in black. As can be seen the individual curve realizations are considerably more variable than the summarized IntCal mean. Individual realizations are always likely to exhibit significantly more short-term variability than the pointwise mean IntCal20 curve due to the averaging process by which the latter is obtained. This is most evident in periods where the underlying data is sparse

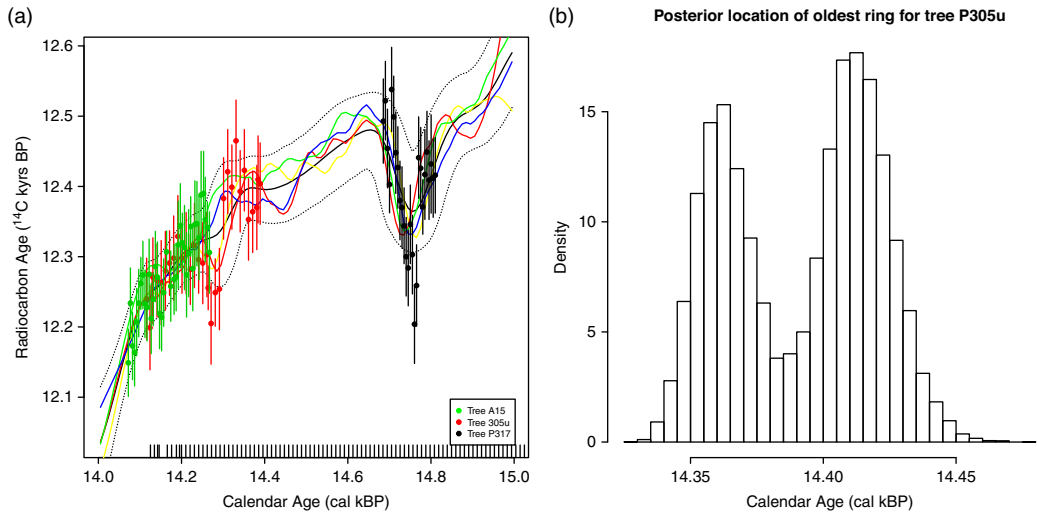


Figure 7 A sample of posterior information generated alongside curve production—Panel (a) illustrates the difference between sample realizations of the curve, shown in color, and the summarized pointwise IntCal20 mean in black; the rug shows the knot locations in this time period, note the even knot spacing from 14.2–15.2 cal kBP to allow equitable placement of the P305u and P317 tree-ring sequences. Panel (b) plots the posterior estimate of the calibrated age of floating tree-ring sequence P305u.

and consequently we have little information on the true value of the curve. In these periods each individual posterior realization will retain similar levels of internal variation/wiggleness as seen in other periods of the curve, as it has been informed that this is the kind of variation expected, but will place these wiggles at slightly different times due to the lack of data. However, once these individual realizations have been averaged over, the pointwise mean summary will tend towards interpolation and hence appear smooth.

In addition to this reduction in short-term variability, where the constituent IntCal data have uncertain calendar ages the posterior IntCal pointwise mean may also appear to smooth out more significant features. Such an effect can be seen in Figure 7 where the inversion in the mean curve around 14.75 cal kBP (corresponding to floating tree-ring sequence P317) is reduced compared with both the observed data and the individual realizations. In such instances where underlying data have uncertain calendar ages, each iteration of the sampler will have posited a particular set of potential true calendar ages for these determinations, and the resultant realization will be a plausible <sup>14</sup>C history on that basis. Each individual realization may therefore retain the significant feature but, once the differing calendar ages at which that feature may occur are averaged over, it may appear reduced in the pointwise mean summary. This can also be seen in Figure 7 where all the individual realizations show a more significant inversion than the pointwise IntCal20 mean, but located at slightly different times to account for uncertainty in the true calendar age of tree P317.

It is important to stress that, for calibration of single determinations, users do not need to consider more than the pointwise mean and variance summaries for accurate calibration. These pointwise values are sufficient and provide the simplest way for a calibration user with a single <sup>14</sup>C determination to obtain the correct calendar dates within current calibration software. We therefore have retained these summary values as our published

output for IntCal20. However, for those wishing to obtain ultimate accuracy in jointly modeling multiple  $^{14}\text{C}$  determinations, or if interest is in atmospheric  $^{14}\text{C}$  variability, individual curve realizations may provide more insight. The IntCal group is currently discussing how best to provide this additional information.

#### 5.4 Posterior Calibrated Calendar Ages

Our Bayesian approach, in addition to creating a calibration curve, also provides posterior estimates for all of the other parameters entered into the model, many of which are of potential interest in their own right. These parameters include the level of over-dispersion seen within tree-ring  $^{14}\text{C}$  determinations of the same calendar age as already discussed; the posterior MRA for the Cariaco Basin, which is shown in Hughen and Heaton (2020 in this issue); and the calendar ages  $\theta_i$  of all our uncertain age data. Figure 7 provides an illustration of the type of posterior estimates for the calendar ages  $\theta_i$  we obtain. The points overlaid in panel (a), shown with  $1\sigma$   $^{14}\text{C}$  uncertainty bars, represent the three floating Bølling-Allerød tree-ring sequences. While these trees enter the curve with no information as to their true calendar age  $\theta$ , they are themselves calibrated simultaneously to creating the curve according to how they best fit with the rest of the data. After curve construction we therefore obtain a posterior estimate for their true calendar ages. The plotted locations of the trees in panel (a) represent their posterior mean calendar ages. However, we obtain a complete posterior distribution for their calibrated ages. Panel (b) presents a histogram of the posterior calibrated age estimate for specific tree P305u. As we can see, this posterior estimate is bimodal suggesting two potential fitting locations—this potential for multimodality motivates our use of MCMC tempering. We generate posterior calibrated age estimates  $\theta_i$  for all our data in a similar way. See the Supplementary Information for posterior calendar age estimates for all our floating tree-ring sequences.

## 6 CONCLUSIONS AND FURTHER WORK

Our revised approach to the construction of the IntCal calibration curve offers several distinct advantages over the previous random walk approach used for IntCal09 and IntCal13 (Heaton et al. 2009; Niu et al. 2013). The use of Bayesian splines maintains a method of equal theoretical rigour and still fits within the desired Bayesian framework that is now universal for  $^{14}\text{C}$  calibration. However the increased speed with which the Bayesian spline curve can be constructed enables much greater practical flexibility in our modeling choices, and more detailed investigation of the effect of key elements on the final curve. The main reason for this increase in speed is a change to the MCMC. While, as in the random walk approach, the Bayesian spline fits using Metropolis-within-Gibbs, the critical step of updating the curve itself is performed via Gibbs and updates the entire curve simultaneously. Conversely, the random walk approach updated the curve one calendar year at a time and via Metropolis-Hastings which led to mixing issues and extremely slow convergence.

For users, the main differences in using the IntCal20 curves will be found through an increase in the level of annual detail in the curve; increased wiggleness in the curve as we extend back in time; and the introduction of prediction intervals which recognize potential additional sources of variation in observed  $^{14}\text{C}$  determinations beyond that occurring in laboratory measurement. For the first time, our new approach is also able to generate complete sets of plausible calibration curves that span from 0–55 cal kBP allowing access to covariance information. For joint calibration of multiple determinations, such as when analysing the time elapsed

between events, this covariance information has the potential to offer more detailed insight. The IntCal working group is currently in discussion about how to best disseminate this covariance information and is planning to provide a guide as to how it could be used within calibration. Including such covariance is likely to require modifications to current calibration software. These realizations, with their covariance information, have however already been used in the creation of Marine20 (Heaton et al. 2020 in this issue) to rigorously propagate our atmospheric uncertainties through to the marine calibration curve.

Further work is needed in several areas. Currently the methodology splits curve construction into two sections. Firstly, we create the curve back to approximately 14 cal kBP based upon only tree-ring determinations. The older section, based upon a wider range of  $^{14}\text{C}$  material, is then created conditional on this more recent section. Future work aims to combine these two steps into a single integrated approach. Our current approach to screening potential data outliers, based on comparison to a curve created using all determinations, is also somewhat influenced by data volume. A large, or dense, dataset will dominate the curve and so is less likely to be identified as an outlier than a shorter, or sparser, dataset. An alternative approach might consider the observed residuals for each set when compared against a curve estimate that excludes the particular set under consideration. This would however be considerably more computationally intensive. We also suggest valuable work would be to investigate potential causes of the over-dispersion seen in the data, e.g. whether this variability is shared by data coming from the same location, tree species, or laboratory. Some preliminary work in this area has already been done by Hogg et al. (2019). Further insight would improve our ability to calibrate multiple determinations such as when wiggle matching a series of radiocarbon determinations from the same tree and laboratory. Finally further work is also needed to resolve the MRA within the Cariaco Basin.

## ACKNOWLEDGMENTS

We would like to thank two reviewers for helpful comments which have improved the quality of this paper. T.J. Heaton is supported by Leverhulme Trust Fellowship RF-2019-140\9, “Improving the Measurement of Time Using Radiocarbon.”

## SUPPLEMENTARY MATERIAL

To view supplementary material for this article, please visit <https://doi.org/10.1017/RDC.2020.46>

## REFERENCES

- Adolphi F, Muscheler R, Friedrich M, Güttler D, Wacker L, Talamo S, Kromer B. 2017. Radiocarbon calibration uncertainties during the last deglaciation: Insights from new floating tree-ring chronologies. *Quaternary Science Reviews* 170:98–108. <http://www.sciencedirect.com/science/article/pii/S0277379117300641>. doi: 10.1016/j.quascirev.2017.06.026.
- Bard E, Arnold M, Hamelin B, Tisnerat-Laborde N, Cabioch G. 1998. Radiocarbon calibration by means of mass spectrometric  $^{230}\text{Th}/^{234}\text{U}$  and  $^{14}\text{C}$  ages of corals: An updated database including samples from Barbados, Mururoa and Tahiti. *Radiocarbon* 40(3):1085–1092. doi: 10.1017/S0033822200019135.
- Bard E, Hamelin B, Fairbanks RG, Zindler A. 1990. Calibration of the  $^{14}\text{C}$  timescale over the past 30,000 years using mass spectrometric U-Th ages from Barbados corals. *Nature* 345: 405–410.
- Bard E, Ménot G, Rostek F, Licari L, Böning P, Edwards RL, Cheng H, Wang Y, Heaton TJ. 2013. Radiocarbon calibration/comparison records based on marine sediments from the Pakistan and Iberian margins. *Radiocarbon* 55(4):1999–2019. doi: 10.2458/azu\_js\_rc.55.17114.

- Bard E, Ménot-Combes G, Rostek F. 2004. Present status of radiocarbon calibration and comparison records based on Polynesian corals and Iberian margin sediments. *Radiocarbon* 46(3):1189–1202. doi: [10.1017/S0033822200033087](https://doi.org/10.1017/S0033822200033087).
- Bates D, Maechler M. 2019. Matrix: Sparse and dense matrix classes and methods. <https://CRAN.R-project.org/package=Matrix>. R package version 1.2-17.
- Bayliss A, Marshall P, Dee M, Friedrich M, Heaton TJ, Wacker L. 2020. IntCal20 tree-rings: An archaeological SWOT analysis. *Radiocarbon* 62. This issue.
- Beck, JW, Richards, DA, Edwards, LR, Silverman, BW, Smart, PL, Donahue, DJ, Herrera-Osterheld, S, Burr, GS, Calsoyas, L, Jull, TAJ, Biddulph, D. 2001. Extremely Large Variations of Atmospheric  $^{14}\text{C}$  Concentration During the Last Glacial Period. *Science* 292(5526):2453–2458. <https://science.sciencemag.org/content/292/5526/2453>. doi: [10.1126/science.1056649](https://doi.org/10.1126/science.1056649). arXiv: <https://science.sciencemag.org/content/292/5526/2453.full.pdf>.
- Berry SM, Carroll RJ, Ruppert D. 2002. Bayesian smoothing and regression splines for measurement error problems. *Journal of the American Statistical Association* 97(457):160–169. arXiv: <https://doi.org/10.1198/016214502753479301>. doi: [10.1198/016214502753479301](https://doi.org/10.1198/016214502753479301).
- Blackwell PG, Buck CE. 2008. Estimating radiocarbon calibration curves. *Bayesian Analysis* 3(2):225–442.
- Bronk Ramsey C, Heaton TJ, Schlolaut G, Staff RA, Bryant CL, Brauer A, Lamb HF, Marshall MH, Nakagawa T. 2020. Reanalysis of the atmospheric radiocarbon calibration record from Lake Suigetsu, Japan. *Radiocarbon* 62. This issue.
- Buck CE, Blackwell PG. 2004. Formal statistical models for estimating radiocarbon calibration curves. *Radiocarbon* 46(3):1093–1102. doi: [10.1017/S0033822200033026](https://doi.org/10.1017/S0033822200033026).
- Burr G, Galang C, Taylor F, Gallup C, Edwards RL, Cutler K, Quirk B. 2004. Radiocarbon results from a 13-kyr BP coral from the Huon Peninsula, Papua New Guinea. *Radiocarbon* 46(3):1211–1224. doi: [10.1017/S0033822200033105](https://doi.org/10.1017/S0033822200033105).
- Burr GS, Beck JW, Taylor FW, Recy J, Edwards RL, Cabioch G, Corregge T, Donahue DJ, O'Malley JM. 1998. A high-resolution radiocarbon calibration between 11,700 and 12,400 calendar years BP derived from  $^{230}\text{Th}$  ages of corals from Espiritu Santo Island, Vanuatu. *Radiocarbon* 40(3):1093–1105. doi: [10.1017/S0033822200019147](https://doi.org/10.1017/S0033822200019147).
- Butzin M, Heaton TJ, Köhler P, Lohmann G. 2020. A short note on marine reservoir age simulations used in IntCal20. *Radiocarbon* 62. This issue.
- Capano M, Miramont C, Shindo L, Guibal F, Marschal C, Kromer B, Tuna T, Bard E. 2020. Onset of the Younger Dryas recorded with  $^{14}\text{C}$  at annual resolution in French subfossil trees. *Radiocarbon* 62. This issue.
- Cheng H, Edwards RL, Southon J, Matsumoto K, Feinberg JM, Sinha A, Zhou W, Li H, Li X, Xu Y, Chen S, Tan M, Wang Q, Wang Y, Ning Y. 2018. Atmospheric  $^{14}\text{C}/^{12}\text{C}$  changes during the last glacial period from Hulu Cave. *Science* 362(6420):1293–1297. <https://science.sciencemag.org/content/362/6420/1293>. arXiv: <https://science.sciencemag.org/content/362/6420/1293.full.pdf>. doi: [10.1126/science.aau0747](https://doi.org/10.1126/science.aau0747).
- Cook JR, Stefanski LA. 1994. Simulation-extrapolation estimation in parametric measurement error models. *Journal of the American Statistical Association* 89(428):1314–1328. arXiv: <https://doi.org/10.1080/01621459.1994.10476871>. doi: [10.1080/01621459.1994.10476871](https://doi.org/10.1080/01621459.1994.10476871).
- Corporation M, Weston S. 2019. doParallel: Foreach parallel adaptor for the “parallel” package. <https://CRAN.R-project.org/package=doParallel>. R package version 1.0.15.
- Cutler KB, Gray SC, Burr GS, Edwards RL, Taylor FW, Cabioch G, Beck JW, Cheng H, Moore J. 2004. Radiocarbon calibration to 50 kyr BP with paired  $^{14}\text{C}$  and  $^{230}\text{Th}$  dating of corals from Vanuatu and Papua New Guinea. *Radiocarbon* 46(3):1127–1160. doi: [10.1017/S0033822200033063](https://doi.org/10.1017/S0033822200033063).
- Durand N, Deschamps P, Bard E, Hamelin B, Camoin G, Thomas AL, Henderson GM, Yokoyama Y, Matsuzaki H. 2013. Comparison of  $^{14}\text{C}$  and U-Th ages in corals from IODP #310 Cores Offshore Tahiti. *Radiocarbon* 55(4):1947–1974. doi: [10.2458/azu\\_js\\_rc.v55i2.16134](https://doi.org/10.2458/azu_js_rc.v55i2.16134).
- Edwards RL, Beck JW, Burr GS, Donahue DJ, Chappell JMA, Druffel ERM, Taylor FW. 1993. A large drop in atmospheric  $^{14}\text{C}/^{12}\text{C}$  and reduced melting in the Younger Dryas with  $^{230}\text{Th}$  ages of corals. *Science* 260(5110):962–968.
- Fairbanks RG, Mortlock RA, Chiu TC, Cao L, Kaplan A, Guilderson TP, Fairbanks TW, Bloom A, Grootes PM, Nadeau MJ. 2005. Radiocarbon calibration curve spanning 0 to 50,000 years BP based on paired Th-230/U-234/U-238 and C-14 dates on pristine corals. *Quaternary Science Reviews* 24(16–17):1781–1796.
- Fan, J, Truong YK. 1993. Nonparametric regression with errors in variables. *Ann. Statist.* 21(4):1900–1925. doi: [10.1214/aos/1176349402](https://doi.org/10.1214/aos/1176349402).
- Fasiolo M. 2016. An introduction to mvnfast. R package version 0.1.6. <https://CRAN.R-project.org/package=mvnfast>.
- Genz A, Bretz F, Miwa T, Mi X, Leisch F, Scheipl F, Hothorn T. 2019. mvtnorm: multivariate normal and t distributions. R package version 1.0-11. <https://CRAN.R-project.org/package=mvtnorm>.
- Green P, Silverman BW. 1993. Nonparametric regression and generalized linear models. New

- York: Chapman and Hall/CRC. doi: [10.1201/b15710](https://doi.org/10.1201/b15710).
- Harrell FE. 2001. Regression modeling strategies: with applications to linear models, logistic regression, and survival analysis. In: Springer series in statistics. New York: Springer.
- Heaton TJ, Bard E, Hughen K. 2013. Elastic tie-pointing—transferring chronologies between records via a Gaussian process. *Radiocarbon* 55(4):1975–1997. doi: [10.2458/azu\\_js\\_rc.55.17777](https://doi.org/10.2458/azu_js_rc.55.17777).
- Heaton TJ, Blackwell PG, Buck CE. 2009. A Bayesian approach to the estimation of radiocarbon calibration curves: the IntCal09 methodology. *Radiocarbon* 51(4):1151–1164. doi: [10.1017/S0033822200034214](https://doi.org/10.1017/S0033822200034214).
- Heaton TJ, Köhler P, Butzin M, Bard E, Reimer RW, Austin WEN, Bronk Ramsey C, Grootes PM, Hughen KA, Kromer B, Reimer PJ, Adkins J, Burke A, Cook MS, Olsen J, Skinner LC. 2020. Marine20—the marine radiocarbon age calibration curve (0–55,000 cal BP). *Radiocarbon* 62. This issue. doi: [10.1017/RDC.2020.68](https://doi.org/10.1017/RDC.2020.68).
- Hoffmann D, Beck J, Richards D, Smart P, Singarayer J, Ketchmark T, Hawkesworth C. 2010. Towards radiocarbon calibration beyond 28 ka using speleothems from the Bahamas. *Earth and Planetary Science Letters* 289(1–2): 110. doi: [10.1016/j.epsl.2009.10.004](https://doi.org/10.1016/j.epsl.2009.10.004).
- Hogg A, Hua Q, Blackwell P, Niu M, Buck C, Guilderson T, Heaton TJ, Palmer J, Reimer P, Reimer R, Turney C, Zimmerman S. 2013. SHCal13 Southern Hemisphere calibration, 0–50,000 years cal BP. *Radiocarbon* 55(4): 1889–1903. doi: [10.2458/azu\\_js\\_rc.55.16783](https://doi.org/10.2458/azu_js_rc.55.16783).
- Hogg AG, Heaton TJ, Bronk Ramsey C, Boswijk G, Palmer JG, Turney CSM, Southon J, Gumbley W. 2019. The influence of calibration curve construction and composition on the accuracy and precision of radiocarbon wiggle-matching of tree rings, illustrated by Southern Hemisphere atmospheric data sets from AD 1500–1950. *Radiocarbon* 61(5):1265–1291. doi: [10.1017/rdc.2019.42](https://doi.org/10.1017/rdc.2019.42).
- Hogg AG, Heaton TJ, Hua Q, Palmer JG, Turney CSM, Southon J, Bayliss A, Blackwell PG, Boswijk G, Bronk Ramsey C, Pearson C, Petchey F, Reimer P, Reimer R, Wacker L. 2020. SHCal20 Southern Hemisphere calibration, 0–55,000 years cal BP. *Radiocarbon* 62. This issue. doi: [10.1017/RDC.2020.59](https://doi.org/10.1017/RDC.2020.59).
- Hughen K, Heaton TJ. 2020. Updated Cariaco Basin <sup>14</sup>C calibration dataset from 0–60 cal kyr BP. *Radiocarbon* 62. This issue. doi: [10.1017/RDC.2020.53](https://doi.org/10.1017/RDC.2020.53).
- Hughen KA, Southon JR, Bertrand CJH, Frantz B, Zerbeño P. 2004. Cariaco Basin calibration update: Revisions to calendar and <sup>14</sup>C chronologies for Core P107-58Pc. *Radiocarbon* 46(3):1161–1187. doi: [10.1017/S0033822200033075](https://doi.org/10.1017/S0033822200033075).
- Miyake F, Masuda K, Nakamura T. 2013. Another rapid event in the carbon-14 content of tree rings. *Nature Communications* 4:1748. doi: [10.1038/ncomms2783](https://doi.org/10.1038/ncomms2783).
- Miyake F, Nagaya K, Masuda K, Nakamura T. 2012. A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan. *Nature* 486:240–242. doi: [10.1038/nature11123](https://doi.org/10.1038/nature11123).
- Niu M, Heaton TJ, Blackwell PG, Buck CE. 2013. The Bayesian approach to radiocarbon calibration curve estimation: The IntCal13, Marine13, and SHCal13 methodologies. *Radiocarbon* 55(4):1905–1922. doi: [10.2458/azu\\_js\\_rc.55.17222](https://doi.org/10.2458/azu_js_rc.55.17222).
- O'Hare P, Mekhaldi F, Adolphi F, Raisbeck G, Aldahan A, Anderberg E, Beer J, Christl M, Fahrni S, Svalb HA, Park J, Possnert G, Southon J, Bard E, ASTER Team, Muscheler R. 2019. Multiradionuclide evidence for an extreme solar proton event around 2,610 B.P. (~660 BC). *Proceedings of the National Academy of Sciences* 116(13):5961–5966. [www.pnas.org/content/116/13/5961.full.pdf](https://www.pnas.org/content/116/13/5961.full.pdf). doi: [10.1073/pnas.1815725116](https://doi.org/10.1073/pnas.1815725116).
- Pearson CL, Brewer PW, Brown D, Heaton TJ, Hodgins GWL, Jull AJT, Lange T, Salzer MW. 2018. Annual radiocarbon record indicates 16th century BCE date for the Thera eruption. *Science Advances* 4(8). <https://advances.sciencemag.org/content/4/8/eaar8241.full.pdf>. arXiv:[https://advances.sciencemag.org/content/4/8/eaar8241.full.pdf](https://arxiv.org/abs/https://advances.sciencemag.org/content/4/8/eaar8241.full.pdf). doi: [10.1126/sciadv.aar8241](https://doi.org/10.1126/sciadv.aar8241).
- R Core Team. 2019. R: A language and environment for statistical computing. R Foundation for Statistical Computing, Vienna, Austria. <https://www.R-project.org/>.
- Ramsay JO, Wickham H, Graves S, Hooker G. 2018. fda: Functional Data Analysis. <https://CRAN.R-project.org/package=fda>. R package version 2.4.8.
- Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Hogg AG, Hughen KA, Kromer BGM, Manning S, Bronk Ramsey C, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE. 2004a. IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–1058. doi: [10.1017/S0033822200032999](https://doi.org/10.1017/S0033822200032999).
- Reimer PJ, Brown TA, Reimer RW. 2004b. Discussion: Reporting and calibration of post-bomb <sup>14</sup>C data. *Radiocarbon* 46(3):1299–1304. doi: [10.1017/S0033822200033154](https://doi.org/10.1017/S0033822200033154).
- 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 T, 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):1111–1150. doi: [10.1017/s0033822200034202](https://doi.org/10.1017/s0033822200034202).
- Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck C, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hafliadason H, Hajdas I, Hatté C, Heaton TJ, Hoffmann 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):1869–1887. doi: [10.2458/azu\\_js\\_rc.55.16947](https://doi.org/10.2458/azu_js_rc.55.16947).
- Reimer P, Austin WEN, Bard E, Bayliss A, Blackwell PG, Bronk Ramsey C, Butzin M, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kromer B, Manning SW, Muscheler R, Palmer JG, Pearson C, van der Plicht J, Reimer RW, Richards DA, Scott EM, Southon JR, Turney CSM, Wacker L, Adolphi F, Büntgen U, Capano M, Fahrni S, Fogtmann-Schulz A, Friedrich R, Köhler P, Kudsk S, Miyake F, Olsen J, Reinig F, Sakamoto M, Sookdeo A, Talamo S. 2020. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* 62. This issue. doi: [10.1017/RDC.2020.41](https://doi.org/10.1017/RDC.2020.41).
- Reinig F, Nievergelt D, Esper J, Friedrich M, Helle G, Hellmann L, Kromer B, Morganti S, Pauly M, Sookdeo A, Tegel W, Treydte K, Verstege A, Wacker L, Büntgen U. 2018. New tree-ring evidence for the late glacial period from the northern pre-alps in eastern Switzerland. *Quaternary Science Reviews* 186:215–224. <http://www.sciencedirect.com/science/article/pii/S0277379117305802>. doi: [10.1016/j.quascirev.2018.02.019](https://doi.org/10.1016/j.quascirev.2018.02.019).
- Samworth R, Poore H. 2005. Understanding past ocean circulations: a nonparametric regression case study. *Statistical Modelling* 5(4):289–307. [arXiv:https://doi.org/10.1191/1471082X05st1020a](https://arxiv.org/abs/https://doi.org/10.1191/1471082X05st1020a). doi: [10.1191/1471082X05st1020a](https://doi.org/10.1191/1471082X05st1020a).
- Scott EM, Naysmith P, Cook GT. 2017. Should archaeologists care about <sup>14</sup>C intercomparisons? Why? A summary report on SIRI. *Radiocarbon* 59(5):1589–1596. doi: [10.1017/RDC.2017.12](https://doi.org/10.1017/RDC.2017.12).
- Southon J, Noronha AL, Cheng H, Edwards RL, Wang Y. 2012. A high-resolution record of atmospheric <sup>14</sup>C based on Hulu Cave speleothem H82. *Quaternary Science Reviews* 33:32–41. <http://www.sciencedirect.com/science/article/pii/S0277379111003817>. doi: [10.1016/j.quascirev.2011.11.022](https://doi.org/10.1016/j.quascirev.2011.11.022).
- Stuiver M, Polach HA. 1977. Discussion: Reporting of <sup>14</sup>C data. *Radiocarbon* 19(3):355–363. doi: [10.1017/S0033822200003672](https://doi.org/10.1017/S0033822200003672).
- Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac G, van der Plicht J, Spurk M. 1998. INTCAL98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon* 40(3):1041–1083. doi: [10.1017/S0033822200019123](https://doi.org/10.1017/S0033822200019123).
- Turney CS, Fifield LK, Hogg AG, Palmer JG, Hughen K, Baillie MG, Galbraith R, Ogden J, Lorrey A, Tims SG, et al. 2010. The potential of New Zealand kauri (*Agathis australis*) for testing the synchronicity of abrupt climate change during the Last Glacial Interval (60,000–11,700 years ago). *Quaternary Science Reviews* 29(27–28):3677–3682. <http://www.sciencedirect.com/science/article/pii/S0277379110003197>. doi: [10.1016/j.quascirev.2010.08.017](https://doi.org/10.1016/j.quascirev.2010.08.017).
- Turney CSM, Jones RT, Phipps SJ, Thomas Z, Hogg AG, Kershaw AP, Fogwill CJ, Palmer J, Bronk Ramsey C, Adolphi F, et al. 2017. Rapid global ocean-atmosphere response to Southern Ocean freshening during the last glacial. *Nature Communications* 8(520). <https://hdl.handle.net/10289/11469>. doi: [10.1038/s41467-017-00577-6](https://doi.org/10.1038/s41467-017-00577-6).
- van der Plicht J, Bronk Ramsey C, Heaton TJ, Scott EM, Talamo S. 2020. Recent developments in calibration for archaeological and environmental samples. *Radiocarbon* 62. This issue. doi: [10.1017/RDC.2020.22](https://doi.org/10.1017/RDC.2020.22).