

Supplementary Information
The IntCal20 approach to radiocarbon calibration curve
construction:
A new methodology using Bayesian splines and
errors-in-variables

Heaton TJ Blaauw M Blackwell PG Bronk Ramsey C
Reimer PJ Scott EM

April 29, 2020

S1 Mixing and Convergence

For the mainly dendrochronologically-dated time period from approximately 0-14 cal kBP, the curve is predominantly obtained via Gibbs sampling and so convergence is not a significant concern. We therefore concentrate on assessing convergence for the older period of the curve from 14-55 cal kBP whereby the MCMC updating is Metropolis-within-Gibbs. In this older period, all the data have uncertain calendar ages, i.e. have errors-in-variables, which are updated during curve construction within the MCMC according to a Metropolis-Hastings (MH) step. It is this calendar age updating step where we might be most concerned about MCMC mixing and convergence, both in terms of ensuring the chain has reached equilibrium and that it has fully explored the potential parameter space to overcome potential multi-modality in the posterior. Addressing these concerns is a significant factor in our change to Bayesian splines from the previous random walk approach.

In using splines, the MCMC updating is less constrained and so hopefully enables the chain to move around and explore the space more freely. With splines, given the current set of calendar ages, we can update the entire curve simultaneously in a single step. This is a key improvement on the previous random walk approach whereby the curve value was updated a single year at a time conditional upon its value at all other times, with the consequence it was very slow and cumbersome to move. Further, the faster updating possible with splines enable us to perform diagnostics and run multiple chains. This was not computationally feasible with the previous random walk approach. The use of tempering also aims to address convergence concerns, specifically potential multi-modality.

The large number of parameters in our MCMC mean that it is difficult to assess convergence using standard statistical tests. However, in order to assess whether the target MCMC chain has reached its equilibrium distribution, and whether it fully explores the potential parameter space to overcome potential multi-modality, we performed the following visual checks.

S1.1 Log-likelihood

For the older part of the curve, where we have errors-in-variables, we provide in Figure S1 a trace plot of the log-likelihood of the target (non-tempered) MCMC chain. This indicates that the likelihood has reached a settled value by the point we begin to sample from it and hence provides support that hopefully we have reached equilibrium in the chain itself.

S1.2 Tempering

Perhaps the most significant concern regarding mixing lies in the potential for calendar age posteriors of the constituent data to be multi-modal, i.e. have multiple potential values, with

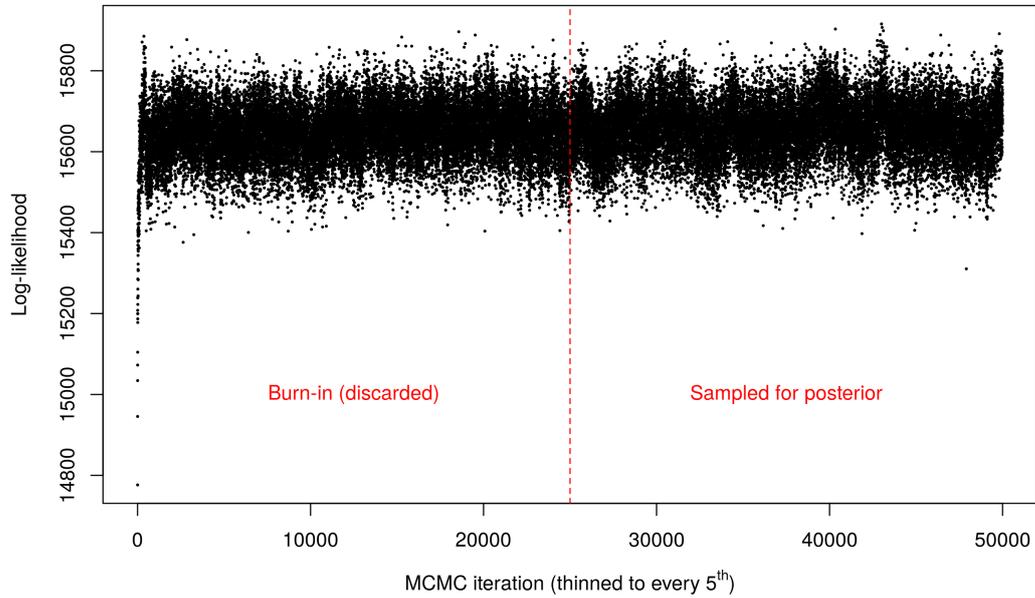


Figure S1: A trace-plot of the log-likelihood of the target MCMC chain, thinned to every 5th realisation i.e. iteration 10,000 equates to running the MCMC for 50,000 updates. The first 25,000 thinned iterations (i.e. 125,000 MCMC updates pre-thinning) of our sampler are discarded as burn-in. Our curve estimate is based upon the iterations after this burn-in period.

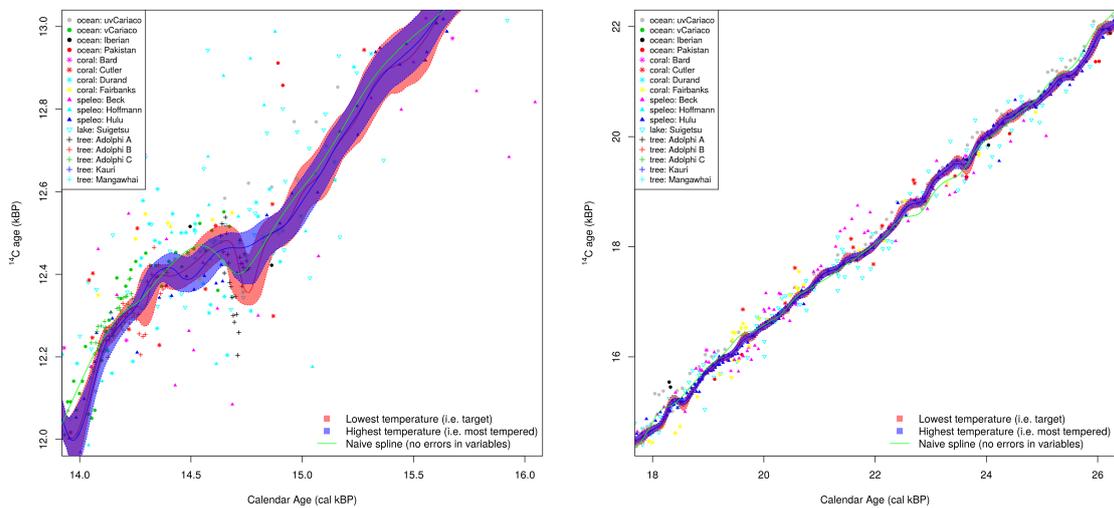


Figure S2: Effect of tempering. The left hand plot shows the curves in the region of the floating Bølling-Allerød tree ring sequences from 14-16 cal kBP, the right hand plot shows the curve over a longer period from 18-26 cal kBP. Note that the data are shown at their prior/observed calendar ages. Consequently assessing their fit to the estimated curve is not appropriate as they are relocated during our Bayesian errors-in-variables curve fitting approach. This is perhaps most apparent in the case of the Cariaco varved sediment (shown in green on the left hand plot) which demonstrates the same shape as the fitted curve yet has a prior calendar age estimate somewhat more recent. The calendar ages of these data will be jointly updated during curve construction to fit with the rest of the data.

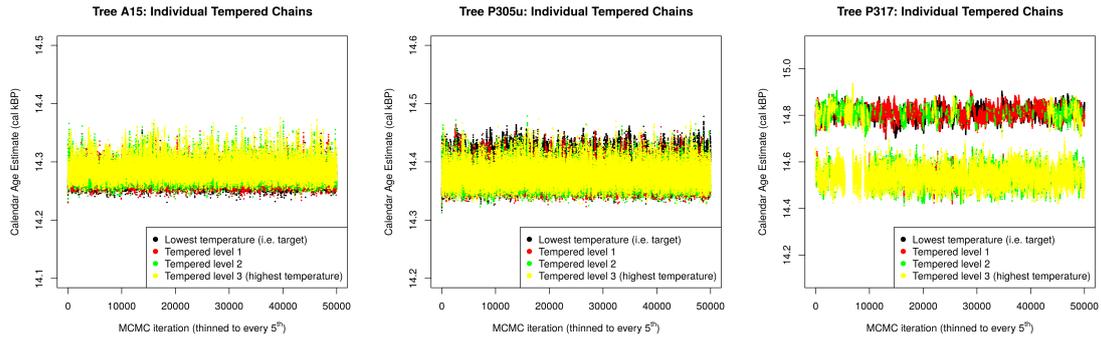


Figure S3: Trace plots of estimated location of floating trees for each tempered chain. The first 25,000 plotted values (corresponding to 125,000 MCMC updates pre-thinning) are considered burn-in and not used in our final curve estimate.

the consequence of a multi-modal posterior for the curve itself. Our MCMC sampler alternates between updating the curve given the current set of calendar age estimates via Gibbs; and then proposing new calendar ages, given the current curve estimate, in a Metropolis-Hastings step. Enabling mixing of calendar age estimates is a particular challenge since, within the Metropolis-Hastings step, the curve which helps determine whether a proposed new set of calendar ages should be accepted or not has been created based upon a fit to the previous set of calendar ages. This means the probability of large changes to calendar age estimates being accepted in any single update is low. Hence, we have the potential for calendar age estimates to get stuck in a single mode of the posterior as switching to another mode would require a significant change to the curve.

The potential for such an issue is most evident when considering the floating tree-ring sequences. Here, proposed calendar age changes result in moving the complete tree-ring sequence as a whole while rigidly maintaining its internal structure. In the case such a floating sequence has multiple potential posterior calendar age modes, a change from one mode to another in a single update step is challenging since it would be highly probable any such proposed move would be rejected by the MH — placing the complete tree-ring sequence at the alternate location would be unlikely to fit with the current curve estimate which had been obtained based upon the sequence’s previous location.

Through tempering we hope to reduce this risk. Specifically, we aim to create higher temperature chains in which the calendar ages of the data can move much more freely within the update steps and cover the full range of potential different modes. In proposing exchanges between these freer chains at high temperatures, which explore more of the space, and those that are less flexible as we reduce the temperatures, we can hopefully enable our sampler to switch between potential modes in our target untempered chain.

To demonstrate the effect, and benefit, of tempering we present in Figure S2 the posterior curve estimates obtained for the target chain (i.e. not tempered) and the chain run at the highest temperature (i.e. increasing up the observational noise). In these plots, the blue curve is the estimate obtained at the highest temperature while the red curve is the target. For interest, we also show, in green, the naive spline estimate obtained were we to ignore the errors-in-variables. As expected, increasing the temperature provides a smoother curve compared with the target. This is particularly evident in the left-hand plot where we might be most concerned about the mixing of the three floating tree-ring sequences which are present in this time period. The higher temperature curve is much smoother and does not show the same significant inversion around 14.75 cal kBP indicating that, at these higher temperatures, the fitted locations of the floating tree-ring sequences are more easily able to move in calendar age. Hence, by proposing state swaps between different temperature chains, we provide the target chain with the potential to switch between multiple modes if it is appropriate. While this is perhaps most important for our tree-ring sequences, it also enables improved mixing for all our uncertain calendar age data.

Relatedly, Figure S3 provides the trace of the calendar age estimates for the three floating Bølling-Allerød tree ring sequences (Adolphi et al., 2017) obtained with each tempered chain. Since

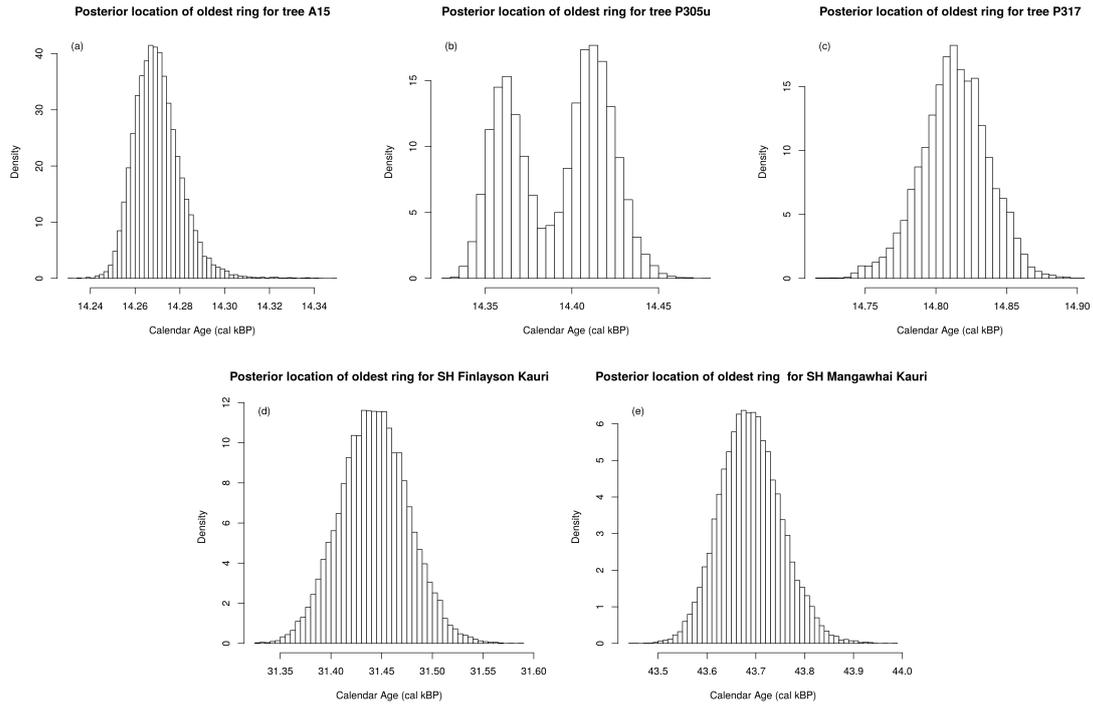


Figure S4: Posterior calendar age estimates for three Bølling-Allerød floating tree ring sequences (Adolphi et al., 2017); and two older Southern Hemisphere kauri trees (Turney et al., 2010, 2017).

these trees shown the most significant internal ^{14}C variation, they have the potential to have the most difficult mixing. As described in the main manuscript we increase the amount of tempering on tree P305u and P317 further, in comparison to the other uncertain age data used for IntCal20, to address this. At the higher temperatures (e.g. the yellow, most tempered chain) the calendar ages can move relatively freely, seen most clearly in the case of P305u and P317. In the case of P305u, the highest temperature chain moves evenly from approximately 14.35 to 14.45 cal kBP; however as seen in Figure S4, the untempered target has a bimodal posterior with peaks around 14.36 cal kBP and 14.41 cal kBP. If we were not able to propose swaps from a chain which moved more freely and covered both these modes we might not be able to satisfactorily explore them both in the target but would rather remain solely in one or the other. Conversely, in the case of tree P317, the highest temperature chain appears to alternate between two distinctly different modes, one around 14.5 cal kBP and another around 14.8 cal kBP. As the temperature reduces towards the target, the 14.5 cal kBP mode becomes less likely with probability concentrating around a single mode located around 14.8 cal kBP for the target posterior. The rejection of proposed swaps to the more recent 14.5 cal kBP mode suggests that, for this tree and its quoted ^{14}C uncertainties the older mode around 14.8 cal kBP is the appropriate fitted age.

S2 Posterior information

S2.1 Fitting locations of floating tree-ring sequences

We highlight some of the posterior information provided by the MCMC. Firstly we present, in Figure S4, the posterior calendar ages of the five floating tree ring sequences — three from the Bølling-Allerød (Adolphi et al., 2017) and the two older SH kauri trees (Turney et al., 2010, 2017). An uninformative prior was placed on the calendar ages of these five sequences and hence the posteriors are determined according to their fit with the rest of the constituent IntCal20 data on which some calendar age information is available. In using such uninformative priors, as opposed

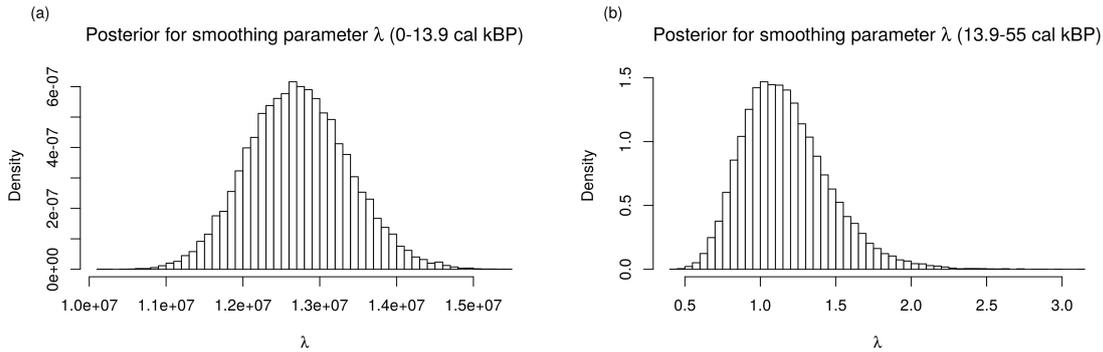


Figure S5: Posterior estimate of the spline smoothing parameter λ in a) the predominantly dendrochronologically-dated based section of the curve from 0-13.9 cal kBP; and b) the older time period from 13.9-55 cal kBP where the curve is based on a wider range of material. Note that due to the differing density of knots in each section, we would not expect the posterior estimates to be the same in each region.

to for example the ice-core based calendar age estimate provided by Adolphi et al. (2017), we maintain independence between the radiocarbon and ice-core timescales. This allows consideration of potential transfer functions between the radiocarbon and ice-core timescales without concerns over circularity. Note in particular the bi-modality for tree P305u indicating there are two potential fitting calendar age locations for this sequence, as already highlighted in Section S1.2. Since this ring sequence exhibits significant internal ^{14}C variability (see Adolphi et al., 2017), this will also lead to some bi-modality in the posterior for the curve itself. Approximately half the curve realisations will correspond to ^{14}C history where this tree-ring sequence begins around 14.36 cal kBP and other half of the curve realisations will correspond to ^{14}C history where this feature begins around 14.41 cal kBP.

S2.2 Posterior on smoothing parameters

In Figure S5 we present the posterior estimate of λ , the spline smoothing parameter, for both sections of the curve separately — firstly the predominantly dendrochronologically-dated section extending from 0-13,910 cal BP; and secondly the older section of curve from 13,910-55,000 cal BP based upon a wider variety of material including speleothems, lake and marine sediment cores, corals and floating tree-ring sequences. Since the number, and density, of knots in the two separate sections of curve are not equivalent the two posteriors are not directly comparable with one another.

S3 Choice of additive error model

We introduce an additive error term to recognise and adapt to potential additional variability in ^{14}C determinations beyond the lab-reported uncertainty. Specifically, in the $F^{14}\text{C}$ domain, we model the observed values as

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

where θ_i is the calendar year of determination i , and the function $f(\theta_i)$ is our estimate of the atmospheric level of $F^{14}\text{C}$ in that calendar year. Under this model, the observational uncertainty is decomposed into two terms: $\epsilon_i \sim N(0, \sigma_i^2)$, with σ_i known, being the independent uncertainty reported by the laboratory; and $\zeta_i \sim N(0, \tau_i^2)$ the additive error which models the potential over-dispersion present in our ^{14}C observations and takes the form of independent random effects with unknown variance τ_i^2 . We considered three potential models for the form of τ_i (the standard deviation of the additive error):

1. *Constant additive error model* — A constant standard deviation model i.e.

$$\zeta_i \sim N(0, \tau^2).$$

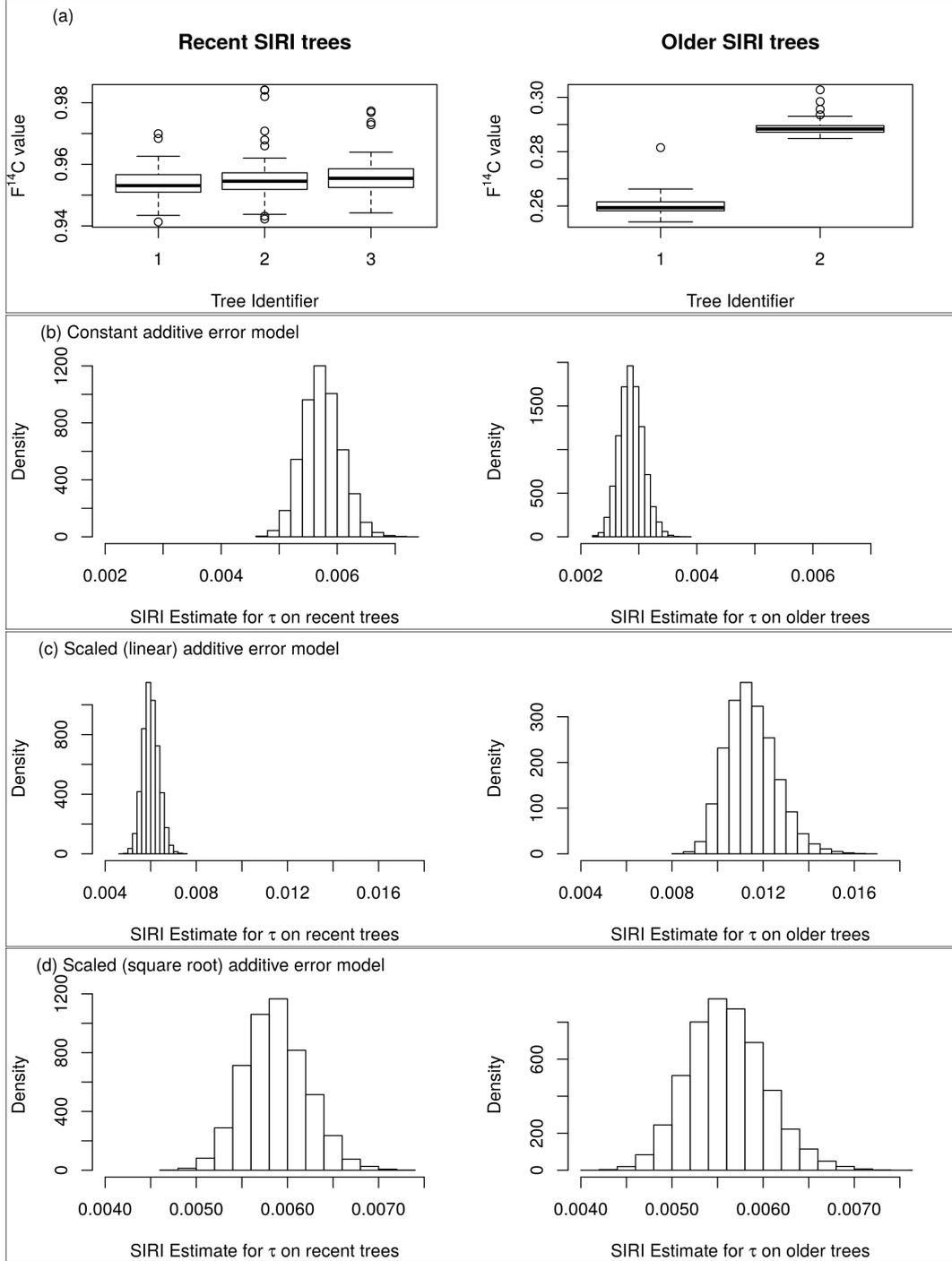


Figure S6: Comparing different additive error models. An appropriate model would provide the same/similar estimates of τ when fitted to both the more recent and the older trees taken from SIRI. We model the $F^{14}C$ on tree i in laboratory j as $F_{ij} = f(\theta_i) + \epsilon_{ij} + \zeta_{ij}$. Here $\epsilon_{ij} \sim N(0, \sigma_{ij}^2)$ are independent with σ_{ij} the laboratory-reported uncertainty; and ζ_{ij} denotes the over-dispersion. Panel a) shows box plots of the measured $F^{14}C$ laboratory values for each of the three recent SIRI tree-rings and the two older SIRI tree-rings. Panel b) shows the posterior estimates obtained for τ , on both the recent and older sets of trees separately, if we model the over-dispersion seen within any tree-ring as independent of the value of $f(\theta)$ i.e. $\zeta_{ij} \sim N(0, \tau^2)$. Panel c) if we model the over-dispersion seen within any tree-ring as scaling linearly with the value of $f(\theta)$ i.e. $\zeta_{ij} \sim N(0, \tau^2 f(\theta_i)^2)$. Panel d) if we model it as scaling linearly with $\sqrt{f(\theta)}$ i.e. $\zeta_{ij} \sim N(0, \tau^2 f(\theta_i))$.

Under this model, the level of over-dispersion on F_i is not dependent upon the level of ^{14}C in the object.

2. *Scaled (linear) additive error model* — A model where the standard deviation of the additive error scales linearly with $f(\theta)$, i.e.

$$\zeta_i \sim N(0, \tau^2 f(\theta_i)^2)$$

3. *Scaled (square root) additive error model* — A model where the standard deviation of the additive error scales linearly with $\sqrt{f(\theta)}$, i.e.

$$\zeta_i \sim N(0, \tau^2 f(\theta_i))$$

In order to assess which of these models was most suitable we considered data from the SIRI intercomparison exercise (Scott et al., 2017) consisting of repeat measurements of trees from two distinct time periods — 3 trees with radiocarbon determinations of around 300 ^{14}C yrs BP; and 2 trees with radiocarbon determinations of about 10,000 ^{14}C yrs BP. Approximately 80 repeat measurements were available on each of the younger group of trees; and 76 repeat measurements on each of the older trees. We implemented separate Bayesian analyses fitting the three above models on each distinct time period independently. Uninformative priors were placed on the values of $f(\theta)$ for each tree and the value of τ . MCMC was then performed for 100,000 iterations, with the first 50,000 discarded as burn-in.

For a given additive error model, this provided posterior estimates of τ for both the recent and older sets of trees independently. To assess model suitability, the consistency of these two posterior estimates over the differing time periods were considered. An appropriate model would be expected to provide similar posterior estimates for τ when fitted to the older trees as when fitted to the more recent trees. The relative estimates are shown in Fig S6. As can be seen, both the model where the standard deviation of the additive error is constant (option 1) and the model it scales linearly with $f(\theta)$ (option 2), provide posterior τ estimates that are considerably different when fitted to the more recent trees as compared to the older trees. This suggests these additive error models are not appropriate. Conversely, the model whereby the additive error scales linearly with $\sqrt{f(\theta)}$ provides similar estimates for τ in both the recent and older time periods. It was therefore decided to implement this square-root additive error model (option 3) to adapt to over-dispersion within the IntCal20 data. To provide a prior on the value of τ to use for curve construction, the SIRI data were combined into a single dataset and another Bayesian analysis performed. The mean and variance of the resultant joint estimate for τ were calculated and used to determine our prior for IntCal20 as described in the main manuscript.

References

- Adolphi, F., Muscheler, R., Friedrich, M., Gütler, 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. URL: <http://www.sciencedirect.com/science/article/pii/S0277379117300641>, doi:<https://doi.org/10.1016/j.quascirev.2017.06.026>.
- Scott, E.M., Naysmith, P., Cook, G.T., 2017. Should Archaeologists Care about ^{14}C Intercomparisons? Why? A Summary Report on SIRI. *Radiocarbon* 59, 1589–1596. doi:10.1017/RDC.2017.12.
- Turney, C.S., Fifield, L.K., Hogg, A.G., Palmer, J.G., Hughen, K., Baillie, M.G., Galbraith, R., Ogden, J., Lorrey, A., Tims, S.G., Jones, R.T., 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, 3677 – 3682. URL: <http://www.sciencedirect.com/science/article/pii/S0277379110003197>, doi:<https://doi.org/10.1016/j.quascirev.2010.08.017>.

Turney, C.S.M., Jones, R.T., Phipps, S.J., Thomas, Z., Hogg, A.G., Kershaw, A.P., Fogwill, C.J., Palmer, J., Bronk Ramsey, C., Adolphi, F., Muscheler, R., Hughen, K.A., Staff, R.A., Grosvenor, M., Golledge, N.R., Olander Rasmussen, S., Hutchinson, D.K., Haberle, S., Lorrey, A., Boswijk, G., Cooper, A., 2017. Rapid global ocean-atmosphere response to Southern Ocean freshening during the last glacial. *Nature Communications* 8. URL: <https://hdl.handle.net/10289/11469>, doi:10.1038/s41467-017-00577-6.