Double the dates and go for Bayes — impacts of model choice, dating density and quality on chronologies

Maarten Blaauw[a,*], J. Andrés Christen[b], K. D. Bennett[c,d], Paula J. Reimer[a]

[a] *School of Natural and Built Environment, Queen's University Belfast, Belfast BT7 1NN, Northern Ireland, United Kingdom*
[b] *Centro de Investigación en Matemáticas CIMAT, Guanajuato 36023, Guanajuato, Mexico*
[c] *School of Geography & Sustainable Development, University of St Andrews, St Andrews KY16 9AL, Scotland, United Kingdom*
[d] *Queen's Marine Laboratory, Queen's University Belfast, Portaferry BT22 1PF, Northern Ireland, United Kingdom*

Abstract

Reliable chronologies are essential for most Quaternary/geological studies, but little is known about how age-depth model choice as well as dating density and quality affect the precision and accuracy of chronologies. A meta-analysis suggests that most existing late-Quaternary studies contain fewer than one date per millennium, and thus provide millennial-scale precision at best. We use simulations to estimate what dating density and quality are required to obtain accurate chronologies at a certain precision. For many studies, a doubling in dating density would significantly improve chronologies and thus their value for reconstructing and interpreting past environmental changes. Commonly used basic age-depth models stop becoming more precise after a certain dating density is reached, but Bayesian age-depth models, which take advantage of chronological ordering, keep on improving with more dates. Moreover, Bayesian models produce more realistic errors for cores with few dates, and can reach multi-decadal precision at high resolution. Bayesian age-depth models are also much more robust against dating scatter and outliers. Our simulations show that basic age-depth models underestimate uncertainty and are inaccurate at low dating densities, and perform poorly at high dating densities. Bayesian age-depth models outperform basic ones at all tested dating densities, qualities and time-scales. We recommend that

*Corresponding author: maarten.blaauw@qub.ac.uk

chronologies should be based on a minimum of 2 dpm wherever realistically possible.

Keywords: age-depth model, radiocarbon dates, chronological uncertainties, Bayesian statistics

---

INTRODUCTION

Whenever an additional level of a sedimentary site, core or event is dated, our knowledge of its chronology increases (Bennett, 1994; Bennett and Fuller, 2002). However, dating is expensive and time-consuming, and it can prove challenging to collect sufficient reliable material for dating. A single radiocarbon (^{14}C) date often costs several hundred dollars, and waiting times can amount to months, or even years if several iterations of dating are required. Moreover, radiocarbon and other scientific dates have laboratory errors, the size of which reflects measurement uncertainties as well as the nature of laboratory sample treatment. Sites extending further back in time generally have larger absolute dating errors and thus lower precision age-depth models. Typical relative ^{14}C dating errors hover around 1% (0.5% for modern AMS systems), however lower sample sizes can result in larger errors although higher-precision dates can be obtained with longer counting times. Given the many ways through which ^{14}C or other absolute dates can be offset from their actual age, some degree of scatter is unavoidable in sequences of dates. Repeated measurements of single samples within and between laboratories sometimes show more scatter than can be accounted for by reported errors (Bronk Ramsey et al., 2004; Christen and Pérez E., 2009; Scott, 2013), perhaps owing to inhomogeneous sampling or laboratory-introduced offsets. At times high-resolution ^{14}C dating can reveal unexpectedly large scatter for some core sections (e.g. Lohne et al., 2013; Groot et al., 2014).

Data in the Neotoma Palaeoecology Database (neotomadb.org) show that since the 1960s the majority of late Quaternary sites have been dated using just a few ^{14}C dates (median 5 dates per core), equivalent to ca 1 date every

1400 years or 0.72 dates per millennium (dpm) (Figure 1). Only a very few sites reach much higher dating densities of ca 10–30 dpm (e.g. Kilian et al., 1995; Gulliksen et al., 1998; Lohne et al., 2013; Mauquoy et al., 2002; Blaauw et al., 2004; Southon et al., 2012). For the last 15 years, mean dating density is slightly higher, at 1.3 dpm. Only 14% of the sites have > 2 dpm; 2% have > 4 dpm.

Here we investigate: (i) whether current typical dating densities are sufficient; (ii) the degree to which higher dating densities enhance chronologies; (iii) whether certain types of age-depth models provide more realistic estimates of precision and accuracy (Telford et al., 2004; Parnell et al., 2011; Trachsel and Telford, 2017); and (iv) the extent to which chronologies are affected by dating error, scatter and outliers. Our analysis of existing cores enables estimates of chronological precision, but not accuracy because the 'true' sedimentation histories of the sites are unknown. Therefore we also use a three-staged simulation approach of (i) modelling the accumulation over time of a site (sim_{acc}); (ii) producing a range of basic and Bayesian age-depth models (sim_{age}) and comparing them to the true sim_{acc} timeline; and (iii) sequentially adding single dates (sim_{dat}) using a sampling strategy after Christen and Sansó (2011), followed by re-running the age-depth models (sim_{age}), repeating as necessary. To test the chronological impact of dating quality, we simulated a range of values for dating error and scatter, as well as outlying dates. Most simulated cores spanned several metres, but some were shorter, high-resolution sections.

METHODS

We used a three-step process on cores with both low and high dating density cores for our analysis. First the accumulation rate (sim_{acc}) was simulated to obtain an estimated calendar date for each core (θ_d). Then we assigned a radiocarbon date based on the IntCal13 calibration curve with random variation within in assigned limits. Finally we used a variety of age-depth

models and calculated the difference between the model age and our assigned age to quantify which model worked better for the different data density cores.

Data

Age-depth models were applied to two datasets:

1. Cores from the entire Neotoma Palaeoecology Database (neotomadb.org). We analysed the dating density of all cores with at least two ^{14}C dates and spanning at least 500 yr.
2. The sequence at Kråkenes (western Norway, 61.48°N, 5.7°E: Gulliksen et al., 1998; Lohne et al., 2013), which has 118 accelerator mass spectrometry ^{14}C ages over the interval ca 14–8 kyr BP (so ca 20 dpm). In order to estimate the effect of changing dating density, we removed all but the topmost and bottommost dates, and then sequentially added single dates (using the method outlined below) until reaching 20 dpm.

Sedimentary sequences (sim_{acc})

Besides real data sets, we also simulated hypothetical cores. Sedimentation was simulated by modelling the deposition time represented within each depth section $ds_i = d_i - d_{i-1}$ of a core (default every cm between 0 and 500 cm). Unless stated otherwise, the deposition time at the topmost section was sampled from a gamma distribution as in Blaauw and Christen (2011) with $acc_1 \sim Gamma(acc.shape, acc.shape/acc.mean)$, defaults 50 yr cm^{-1} and 1.1 for mean and shape, respectively. Its top age, θ_0, was set as 0 cal BP by default. The deposition rate of each section ds_i was modelled to deviate from the preceding section ds_{i-1} by a random value sampled from a uniform distribution with width $2\,acc.var$ (default 3.0), where deposition times could not go below $acc.min$ (default 5.0 yr cm^{-1});
$acc_t \sim max(acc.min, acc_{t-1} + \text{Unif}(-acc.var, acc.var))$. These simulated deposition histories then provided the 'true' calendar age for each depth of the simulated cores, θ_d.

We then simulated the initial and sequential (see later) dating of these artificial cores, sim_{dat}. For the simulations here, calibrated ^{14}C dates were modelled, but other types of dates could also be used. For each depth to be dated (see below), its 'true' calendar age θ_d was known from the simulated deposition history sim_{acc}. Uncertainty in having sampled contemporaneous material was then simulated by adding random variation x_{scat} (default 10 yr) from a normal distribution $\theta' \sim N(\theta, {x_{scat}}^2)$. Additionally, with a probability p_{out} (default 5%), the calendar date was treated as an outlier and shifted by up to x_{shift} (default 1000) years:

$$\theta'' = \begin{cases} \text{Unif}(\theta' - x_{shift}, \theta' + x_{shift}) & p_{out} \\ \theta' & 1 - p_{out} \end{cases} \quad (1)$$

Then we used the IntCal13 ^{14}C calibration curve (Reimer et al., 2013), which provides estimates of the ^{14}C age μ_θ for each calendar age θ. We simulated a ^{14}C date with some scatter by taking the IntCal13 ^{14}C age of θ'', $\mu_{\theta''}$, and adding some scatter $y_{\theta''} \sim N(\mu_{\theta''}, \sigma^2)$ where the laboratory error $\sigma = max(\sigma_{min}, y_{scat} \times \epsilon \times \mu_{\theta''})$, with σ_{min} the minimum error (default 20 ^{14}C yr), y_{scat} an error multiplier (default 1.5) and ϵ the analytical uncertainty (default 1%)

The sim_{acc} simulations presented here aim to model what we consider to be realistic accumulation histories of commonly used sites such as Holocene lakes or bogs. We did not invoke more chronologically disruptive features such as hiatuses, extremely variable accumulation rates, large sections without datable material or systematic ^{14}C age offsets, but this could be investigated. However, the approach we present can be applied to individual sections of sequences, as well as to whole sequences.

Age-depth modelling (sim_{age})

We applied four types of age-depth models, which produced thousands to millions of random iterations to provide many calendar age estimates for each core depth. We first used the popular basic model of linear interpolation as

implemented in *psimpoll* (Bennett, 2007) and *clam* (Blaauw, 2010), which assumes that accumulation rates were constant between neighbouring dated depths and changed, potentially abruptly, exactly at the dated depths (Bennett, 1994). We then applied a basic model that varies more smoothly over time (smooth spline in *clam*). Since ages further down a core must be older, even if the dates or models suggest otherwise, software implementing the above approaches can be instructed to remove any iterations with age-depth model reversals after modelling. Finally, we tested two Bayesian piecewise linear models that use gamma distributions as prior information in order to ensure chronological ordering of each iteration. *Bchron* (Haslett and Parnell, 2008) simulates steps in time and depth sampled from gamma distributions, whereas *Bacon* (Blaauw and Christen, 2011) models the accumulation rates of many equally spaced depth sections based on a gamma distribution (here set at mean 50 and shape 1.1 to allow for many accumulation rates), and a beta distribution to invoke a degree of dependence in accumulation rate between neighbouring depths. Both *Bchron* and *Bacon* have routines to handle outliers, whereas for basic age-depth models outliers need to be removed manually. OxCal's P_Sequence (Bronk Ramsey, 2008) was also tried but individual runs and analyses interpolated to 500 1-cm intervals took days instead of minutes, rending it less suitable for these intensive simulation exercises. R code (R Core Team, 2017) is available on Figshare (doi:10.6084/m9.figshare.3808311).

All age-depth models were produced as outlined above. Each age-depth model sim_{age} was then compared to the known simulated sim_{acc} age θ_d for each depth d, calculating its accuracy as standardized offset, $z_d = |\bar{x}_d - \theta_d|/\sigma_d$, where $\bar{x}_d$ and σ_d are the mean and standard deviation, respectively, of the modelled ages. Standardizing ensures that offsets can be compared between core depths modelled at different precisions. Then the minimum, maximum and mean z over all depths was taken as the age-depth model's accuracy, whereas its precision was calculated as the minimum, maximum and mean of the modelled 95% confidence intervals. In this context, precision refers to the

degree of uncertainty in an estimate and accuracy refers to the actual error or offset of the estimate to its true value.

Sequential dating (sim_{dat})

After each age-depth model sim_{age} was run, we used that age-depth model to determine which depth to date next. This has been investigated by Buck and Christen (1998) and Christen and Buck (1998) using simulations that were computationally extremely time-consuming. Here we adopt a much faster sampling design score developed by Christen and Sansó (2011), which predicts which next data point among all available candidates is likely to provide the most new information. Only one candidate depth is selected at a time, although this not a realistic scenario for real-life ^{14}C dating. In future work we therefore plan to enable selecting multiple depths.

Let $s_1, s_2, \ldots, s_M$ be the depths at which we may take a sample to be dated (by radiocarbon or otherwise). Let $d_1, d_2, ..., d_m$ (a subset of the s_i) be the depths at which we already have dates $y_m = (y_1 \pm \sigma_1, y_2 \pm \sigma_2, \ldots, y_m \pm \sigma_m)$. Let $cov(d_i, d_j)$ be the covariance of depths d_i and d_j calculated from the joint posterior distribution of the age-depth model using the currently dated depths, that is $G(d|y_m)$. This covariance structure may be approximated using the Monte Carlo output to estimate the chronology (Blaauw, 2010; Haslett and Parnell, 2008; Blaauw and Christen, 2011). For all iterations $t = 1, 2, \ldots, T$, we calculate the covariance of the corresponding ages $G(s_i|\theta^{(t)}, x^{(t)})$ and $G(s_j|\theta^{(t)}, x^{(t)})$. Let also $V(s_i) = cov(s_i, s_i)$ be the variance at depth s_i. The score A for a new candidate depth d_{m+1} to be dated is (Christen and Sansó, 2011):

$$A(d_{m+1}) = (1 - \|r(d_{m+1})\|) \, \frac{1}{M} \sum_{j=1}^{M} \frac{cov(s_j, d_{m+1})^2}{V(s_j)V(d_{m+1})},$$

where

$$\|r(d_{m+1})\| = \sqrt{\sum_{k=1}^{m} \frac{cov(d_k, d_{m+1})^2}{V(d_k)V(d_{m+1})}}.$$

The score has a formal justification in terms of maximizing the reduction in predictive variance of the new sample point d_{m+1}, specifically it is an approximated and computationally simple version of the sequential Active Learning Cohn strategy used in robotics (Christen and Sansó, 2011). Intuitively, the score chooses a new sample point that is correlated with other locations given the term

$$\frac{1}{M}\sum_{j=1}^{M}\frac{cov(s_j, d_{m+1})^2}{V(s_j)V(d_{m+1})},$$

and consequently favours depths with high variance in their age estimates (large uncertainties), as well as depths which inform us much about the ages of other depths in the sequence (high covariance). However, the term $1 - \|r(d_{m+1})\|$ penalizes depths correlated with locations already sampled, thus separating the dated depths (see Christen and Sansó, 2011, for further intuitive and technical justifications of the score and some examples showing its performance).

The same methods and results apply to the dating density of entire cores or of specific sections of cores. We note that it often makes scientific and financial sense to only apply higher dating resolutions, and thus reach higher chronological precision, for specific core sections of interest (e.g. the 1-m long sections discussed further below).

RESULTS

Depending on their dating density and the chosen age-depth model type, chronologies for cores from the Neotoma database reach millennial to centennial-scale precision (Figure 1). At first sight, basic age-depth models based on linear interpolation or smooth splines (Bennett, 2007; Blaauw, 2010) appear to produce more precise chronologies than do Bayesian models (Haslett and Parnell, 2008; Blaauw and Christen, 2011). At below average dating densities, adding dates enhances the precision of basic age-depth models, but this effect levels off at average and higher dating densities. This precision, even

with low dating densities, is due to the implicit assumption that the age-depth model chosen is the true one, so the model has zero error for the choice of age-depth model. Adding in error for age-depth model choice substantially reduces the total precision (Bennett, 1994; Blaauw and Heegaard, 2012). Bayesian age-depth models on the other hand consistently become more precise as dating density increases. The same patterns appear when all but the top and bottom dates are removed from the Kråkenes sequence, followed by re-adding its dates one by one until reaching 20 dpm (Figure 2).

At the initial, lowest dating densities, most age-depth models fail, unsurprisingly, to capture the long-term shapes of the simulated age-depth trajectories (Telford et al., 2004; Trachsel and Telford, 2017) even though the 95% error ranges of the Bayesian models mostly overlap with the 'true' ages. As a few more strategically chosen dates are added, all models improve to follow a site's main features. More complicated histories require more dates (Figures 3–4, Supplementary information animations 1–4). However what happens at higher dating densities depends largely on the chosen age-depth model type.

Our simulations reveal several important implications of different approaches to age-depth modelling. As with the analyses of cores from the Neotoma database and Kråkenes, different model types produce very different precision estimates (Figure 4a, b, e, f). The commonly used basic models of linear interpolation and smooth spline appear at first sight to be pleasingly precise (average 95% ranges mostly under ca 500 yr), due to the implicit zero error for choice of age-depth model (see above), but the Bayesian age-depth models are more realistic, reconstructing much larger uncertainties especially at low to average dating densities (up to 1 dpm). However, the supposedly higher precision of basic age-depth models comes at a severe cost; they are inaccurate especially at low to average dating densities. Indeed, at those dating densities basic age-depth models are offset from the 'true' ages by many standard deviations; Figure 4b, d). The Bayesian models on the other hand are consistently accurate and produce realistic estimates of precision, the true ages

lying well within two standard deviations (95%) at most depths and dating densities (Figure 4f, h). Even so, in all of our simulations (basic and Bayesian) and at almost all dating densities, the 95% age ranges of some depths lie outside their 'true' ages (envelopes extending above the 2 sd limit in Figure 4). Above ca 1 dpm, linear interpolation age-depth models do not become more precise at increasing dating densities, whereas the smooth-spline models show some improvement after reaching ca 5 dpm. However, the Bayesian models keep on improving. This is because Bayesian models take advantage of chronological ordering, causing ever-increasing precision (yet remaining accurate) as more and more dates start to overlap. At dating densities high enough to match the multi-decadal wiggles in the ^{14}C calibration curve (Kilian et al., 1995; Gulliksen et al., 1998; Lohne et al., 2013; Mauquoy et al., 2002; Blaauw et al., 2004; Southon et al., 2012) (Figure 5), some sections of Bayesian models calculated with *Bacon* can reach multi-decadal precision. Above 30 dpm even basic models gain precision again as repeated dating of individual depths enhances their age estimates.

The curves relating dating density to model precision and, especially, accuracy are not entirely smooth. Sometimes adding a few extra dates will provide an extra piece of information that suddenly results in much more precise and, or, accurate chronologies. However the opposite can also happen when, for example, adding an extra date causes an age reversal with basic age-depth models, or an outlying date produces a less accurate model (particularly for linear interpolation, which is very sensitive to outliers). Bayesian models calculated with *Bchron* seem to consistently lose accuracy at higher dating densities (Figure 4h).

Our simulations show a clear impact of error size on model precision but not accuracy (Figure 6). Dating scatter (Christen and Pérez E., 2009; Scott, 2013) on the other hand appears to have little impact on age-depth model precision (no impact for linear interpolation and a minor one for *Bacon*), but it severely impacts accuracy. Model offsets increase along with increasing scatter, although *Bacon*'s offsets are always considerably lower than those of linear

interpolation. Similarly to dating scatter, outliers have little to no impact on model precision while severely affecting accuracy (Figure 6i–l). Most basic age-depth models are offset by two or more standard deviations once more than 20% of the dates are outlying, while *Bacon*, remarkably, remains reliable until over 50% of dates are outlying.

IMPLICATIONS

In the early, pre-AMS, days of ^{14}C dating, dating density of cores was necessarily low because slices covering many centimetres or even decimetres (and thus centuries of sedimentation) had to be submitted to obtain sufficient datable ^{14}C for the conventional decay counting method (e.g. Bennett et al., 1992; Haberle and Lumley, 1998). With the advent of AMS dating in the 1990s this limitation has largely been lifted, and prices of single dates have come down in real terms. However median dating density remains below 1 dpm (Figure 1), perhaps because the research community still considers 1 date per millennium to be a reasonable rule-of-thumb in order to establish chronologies, or because funding for chronologies has not increased (with the exception of special cases where chronology is the emphasis of the study). Our simulations show, however, that current typical dating densities are insufficient. Commonly used basic age-depth models may fail to capture the main features of a site's accumulation history and produce highly over-optimistic precision estimates (Figure 3). Increasing the dating density to ca 2 dpm, and using Bayesian methods, produces age-depth models that give reasonable confidence for centennial-scale precision estimates. If sub-centennial chronological precision is needed (at least for selected core sections; Figure 5), dating densities over 50 dpm are required, together with age-depth models that take advantage of chronological ordering. Thus, chronologies should be built starting with a skeleton chronology of, say, 1 date every 2 millennia (0.5 dpm), after which small batches of strategically sampled depths should be dated sequentially until reaching ca 2 dpm, even though this is time-consuming.

Basic age-depth modelling approaches such as linear interpolation remain widely accepted and used within the community of past climate/environmental research. For example, Shurtliff et al. (2017) argue for 'linear interpolation to be as good an approach as any', since 'all age-depth models contain considerable uncertainty that is difficult to fully quantify'. However, as a community we should ask ourselves whether these basic approaches remain suited to their task (Bennett, 1994; Telford et al., 2004; Trachsel and Telford, 2017). The method of linear interpolation suggests higher precision in-between dated depths, and becomes more precise with larger gaps between dated depths (Bennett, 1994). This unintuitive result arises because this model implicitly assumes (i) that this is the true model, although we know that it is not; and (ii) that ages between dated points lie along a straight line, which is rarely, if ever, going to be true. Relaxing the assumption (e.g., with Bayesian methods such as *Bacon*) produces the more intuitive result that sections with fewer dates have higher uncertainties. Bayesian models excel since they simulate many different alternative 'routes' by which a site could have accumulated in-between dated depths, diverging more from the linearly-interpolated relationship if less 'guidance' is present. Thus at low to average dating densities, Bayesian models such as *Bchron* or *Bacon* are preferable to basic models, since their model assumptions produce more realistic reconstructions and confidence intervals. At higher dating densities, the dates start to steer the models more directly and the accuracy of basic and Bayesian models is comparable — though Bayesian models that enforce chronological ordering can become much more precise.

Dating scatter seems more disruptive to age-depth model accuracy than error size (Figure 6), so it makes more sense to re-date single depths (e.g. to assess the reliability of dating different components) or date multiple depths (constraining the ages of individually dated depths through enforcing chronological ordering), rather than obtaining single high-precision dates. Over past decades, the palaeoenvironmental community has repeatedly been warned to take uncertainties into account (Maher, 1972; Bennett, 1994;

Blaauw et al., 2007; Blaauw, 2010; Jackson, 2012). Dating uncertainties should not be abused by uncritically linking events between sites (Blaauw, 2012), nor camouflaged by plotting fossil-based reconstructions against calendar time as single curves without any error visualisation. Basic age-depth models, especially those based on the hugely popular linear interpolation, are highly sensitive to outliers and severely overestimate precision at low to average dating densities. Moreover, because they do not improve beyond ca 1–2 dpm, they under-perform at above-average dating densities. In contrast, Bayesian approaches provide the most accurate age-depth models, with reliable precision estimates throughout an impressive range of dating density and quality, continually improving as dating density improves. We strongly recommend aiming for dating densities ≥ 2 dpm (e.g. 20 or more dates along a Holocene sequence) with realistic and stated error estimates. If this is not realistically achievable for an entire sequence, it should be achieved for any shorter sections where the most precise chronology is needed to meet the objectives of the investigation. This will require a modest increase in funds for dating and higher usage of available Bayesian age-depth models, but will provide the accuracy and precision needed to interpret and correlate chronologies at the level required for most palaeoenvironmental questions.

ACKNOWLEDGEMENTS

This work was partly funded by a Banco Santander travel grant to MB. Thanks to Andrew Parnell for help with *Bchron* and to Simon Goring for releasing R code to extract sites from the Neotoma Paleoecology database (http://www.neotomadb.org); the work of the data contributors and the Neotoma community is gratefully acknowledged. We are also grateful to the Past Earth Network (http://www.pastearth.net/) for writing support (grant number EP/M008363/1).

References

Bennett, K. D., 1994. Confidence intervals for age estimates and deposition times in late-Quaternary sediment sequences. The Holocene 4, 337–348.

Bennett, K. D., 2007. *psimpoll* and *pscomb* programs for plotting and analysis. URL `http://www.chrono.qub.ac.uk/psimpoll/psimpoll.html`

Bennett, K. D., Boreham, S., Sharp, M. J., Switsur, V. R., 1992. Holocene history of environment, vegetation and human settlement on Catta Ness, Lunnasting, Shetland. Journal of Ecology 80, 241–273.

Bennett, K. D., Fuller, J. L., 2002. Determining the age of the mid-Holocene *Tsuga canadensis* (hemlock) decline, eastern North America. The Holocene 12, 421–429.

Blaauw, M., 2010. Methods and code for classical age-modelling of radiocarbon sequences. Quaternary Geochronology 5, 512–518.

Blaauw, M., 2012. Out of tune: the dangers of aligning proxy archives. Quaternary Science Reviews 36, 38–48.

Blaauw, M., Christen, J. A., 2011. Flexible paleoclimate age–depth models using an autoregressive gamma process. Bayesian Analysis 3, 457–474.

Blaauw, M., Christen, J. A., Mauquoy, D., van der Plicht, J., Bennett, K. D., 2007. Testing the timing of radiocarbon-dated events between proxy archives. The Holocene 17, 283–288.

Blaauw, M., Heegaard, E., 2012. Estimation of age-depth relationships. In: Birks, H. J. B., Juggins, S., Lotter, A., Smol, J. P. (Eds.), Tracking Environmental Change Using Lake Sediments, Developments in Paleoenvironmental Research 5. Springer, Dordrecht, pp. 379–413.

Blaauw, M., van Geel, B., van der Plicht, J., 2004. Solar forcing of climatic change during the mid-Holocene: indications from raised bogs in The Netherlands. The Holocene 14, 35–44.

Bronk Ramsey, C., 2008. Deposition models for chronological records. Quaternary Science Reviews 27, 42–60.

Bronk Ramsey, C., Higham, T., Leach, P., 2004. Towards high-precision AMS: Progress and limitations. Radiocarbon 46, 17–24.

Buck, C. E., Christen, J. A., 1998. A novel approach to selecting samples for radiocarbon dating. Journal of Archaeological Science 25, 303–310.

Christen, J. A., Buck, C. E., 1998. Sample selection in radiocarbon dating. Applied Statistics 47, 543–557.

Christen, J. A., Pérez E., S., 2009. A new robust statistical model for radiocarbon data. Radiocarbon 51, 1047–1059.

Christen, J. A., Sansó, B., 2011. Advances in the sequential design of computer experiments based on active learning. Communications in Statistics — Theory and Methods, 4467–4483.

Goring, S., Dawson, A., Simpson, G. L., Ram, K., Graham, R. W., Grimm, E. C., Williams, J. W., 2015. neotoma: A programmatic interface to the Neotoma Paleoecological Database. Open Quaternary 1 (2).

Groot, M. H. M., van der Plicht, J., Hooghiemstra, H., Lourens, L. J., Rowe, H. D., 2014. Age modelling for Pleistocene lake sediments: A comparison ofmethods from the Andean Fúquene Basin (Colombia) case study. Quaternary Geochronology 22, 144–154.

Gulliksen, S., Birks, H. H., Possnert, G., Mangerud, J., 1998. A calendar age estimate of the Younger Dryas–Holocene boundary at Kråkenes, western Norway. The Holocene 8, 249–259.

Haberle, S. G., Lumley, S. H., 1998. Age and origin of tephras recorded in postglacial lake sediments to the west of the southern Andes, 44°S to 47°S. Journal of Volcanology and Geothermal Research 84, 239–256.

Haslett, J., Parnell, A. C., 2008. A simple monotone process with application to radiocarbon-dated depth chronologies. Journal of the Royal Statistical Society, Series C 57, 399–418.

Jackson, S. T., 2012. Representation of flora and vegetation in Quaternary fossil assemblages: known and unknown knowns and unknowns. Quaternary Science Reviews 49, 1–15.

Kilian, M. R., van der Plicht, J., van Geel, B., 1995. Dating raised bogs: new aspects of AMS ^{14}C wiggle matching, a reservoir effect and climatic change. Quaternary Science Reviews 14, 959–966.

Lohne, Ø. S., Mangerud, J., Birks, H. H., 2013. Precise ^{14}C ages of the Vedde and Saksunarvatn ashes and the Younger Dryas boundaries from western Norway and their comparison with the Greenland Ice Core (GICC05) chronology. Journal of Quaternary Science 28, 490–500.

Maher, Jr, L. J., 1972. Absolute pollen diagram of Redrock Lake, Boulder County, Colorado. Quaternary Research 2, 531–553.

Mauquoy, D., van Geel, B., Blaauw, M., van der Plicht, J., 2002. Evidence from northwest European bogs shows 'Little Ice Age' climatic changes driven by variations in solar activity. The Holocene 12, 1–6.

Parnell, A. C., Buck, C. E., Doan, T. K., 2011. A review of statistical chronology models for high-resolution, proxy-based holocene palaeoenvironmental reconstruction. Quaternary Science Reviews 30, 2948–2960.

R Core Team, 2017. R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria.
URL `https://www.R-project.org`

Reimer, P. J., Bard, E., Bayliss, A., Beck, J. W., Blackwell, P. G., Ramsey, C. B., Buck, C. E., Cheng, H., Edwards, R. L., Friedrich, M., Grootes,

P. M., Guilderson, T. P., Haflidason, H., Hajdas, I., Hatté, C., Heaton, T. J., Hoffmann, D. L., Hogg, A. G., Hughen, K. A., Kaiser, K. F., Kromer, B., Manning, S. W., Niu, M., Reimer, R. W., Richards, D. A., Scott, E. M., Southon, J. R., Staff, R. A., Turney, C. S. M., van der Plicht, J., 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0-50,000 years cal BP. Radiocarbon 55, 1869–1887.

Scott, E. M., 2013. Radiocarbon dating — sources of error. In: Mock, C. J., Elias, S. A. (Eds.), Encyclopedia of Quaternary Science. Elsevier, Amsterdam, pp. 324–328.

Shurtliff, R. A., Nelson, S. T., McBride, J. H., Rey, K. A., Tucker, J. C., Godwin, S. B., Tingey, D. G., 2017. A 13 000 year multi-proxy climate record from central Utah (western USA), emphasizing conditions leading to large mass movements. Boreas 46, 308–324.

Southon, J., Noronha, A. L., Cheng, H., Edwards, R. L., Wang, Y., 2012. A high-resolution record of atmospheric ^{14}C based on Hulu Cave speleothem H82. Quaternary Science Reviews 33, 32–41.

Telford, R. J., Heegaard, E., Birks, H. J. B., 2004. All age–depth models are wrong: but how badly? Quaternary Science Reviews 23, 1–5.

Trachsel, M., Telford, R. J., 2017. All agedepth models are wrong, but are getting better. The Holocene 27, 860–869.

Figure captions

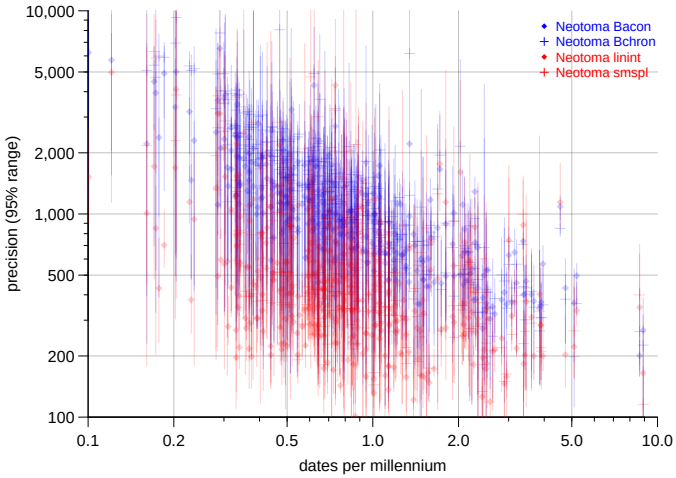

Figure 1: Dating density (dates per millennium) and age-depth model precision (95% error ranges) of all 356 ^{14}C-dated cores spanning at least 500 yr and having at least 2 ^{14}C dates, extracted from the Neotoma Database using R code (Goring et al., 2015). Vertical lines and symbols indicate the minimum to maximum mean age-depth model precision of each core. Red shows basic age-depth models (diamonds linear interpolation, crosses smooth spline), blue Bayesian (diamonds *Bacon*, crosses *Bchron*). Note logarithmic axes.

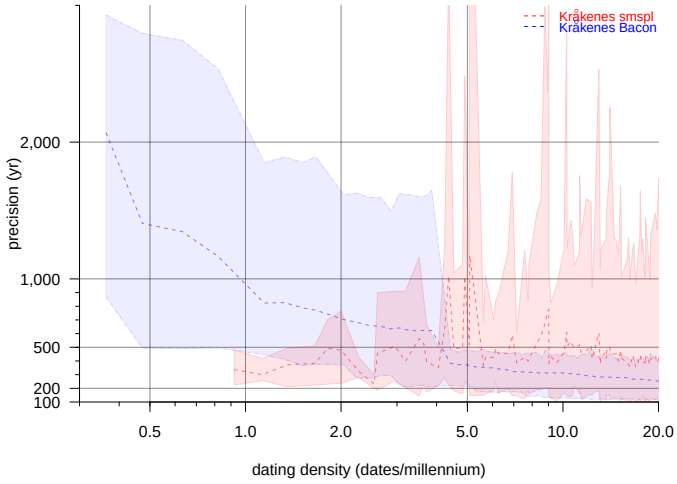

Figure 2: Dating density (dates per millennium) and age-depth model precision (95% error ranges) of the high-resolution dated Kråkenes record (Gulliksen et al., 1998; Lohne et al., 2013). Shaded envelopes and dashed curves show minimum to maximum resp mean age-depth model precision upon sequential re-sampling of the record (red smooth spline (smspl), blue *Bacon*). Note logarithmic axes.

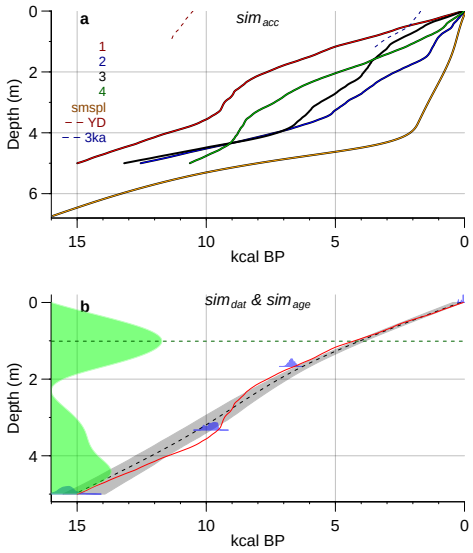

Figure 3: The three-stage simulation of sedimentation, sequential dating, and age-depth modelling. The upper panel shows seven sim_{acc} simulations using a range of random seeds (1: 5113, 2: 2995, 3: 5993, 4: 6993), a smooth spline (smspl) through the "Example" core provided with *clam* (Blaauw, 2010) (orange), simulation 11136 (Younger Dryas) and simulation 1102 (at 3 kcal BP Hallstatt Plateau). The lower panel shows the sequential process of selecting which depth to date next, producing the resulting age-depth model, and comparing the model to the known ages. In this example, so far four depths have been dated (blue silhouettes; sim_{dat}) from the simulated core (red curve; sim_{acc}). A smooth-spline age-depth model is drawn through these dates (grey envelope and dashed black line; sim_{age}). Its age estimates can be compared to the 'true' history (red). From the sample spacing and the age-depth model's variance and covariance at each depth a sampling score is calculated (green distribution). The depth of 1 m has the highest score (dashed green line) and will thus be dated next.

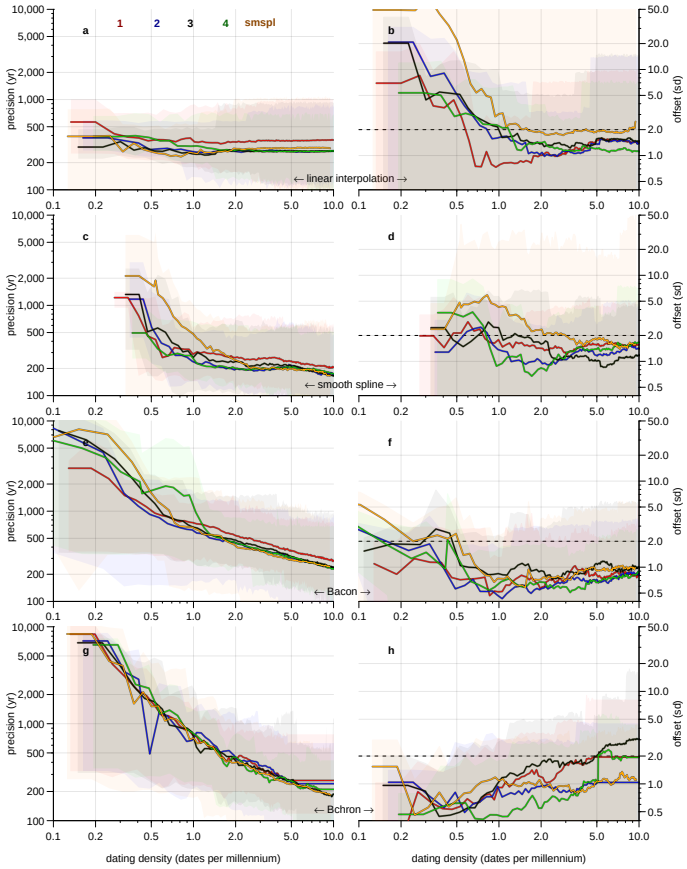

Figure 4: Impact of dating density on chronological precision and accuracy using basic (a-b, linear interpolation; c-d, smooth spline) and Bayesian (e-f, *Bacon* (Blaauw and Christen, 2011); g-h, *Bchron* (Haslett and Parnell, 2008)) age-depth models. For each of 5 simulated cores (see Figure 3 for key to colours), sim_{acc}, dates were added sequentially (sim_{dat}, up to 150 dates) and age-depth models constructed (sim_{age}). Curves show mean values and envelopes show minimum to maximum values for age-depth model precision (95% error ranges, left panels) and accuracy (standardized offset from 'true' ages, right panels; dashed curve shows 2 standard deviation offset). Note logarithmic axes.

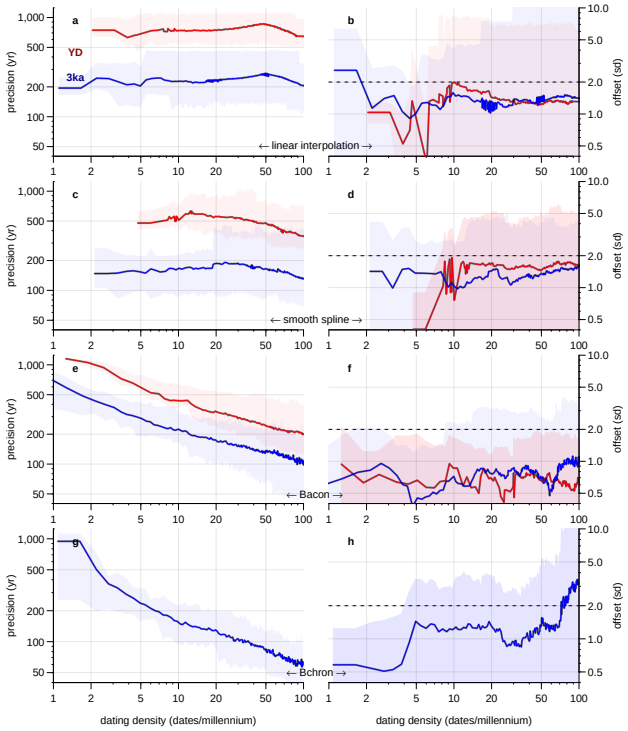

Figure 5: Impact on model reliability of high dating densities at periods with major wiggles in the IntCal13 ^{14}C calibration curve (Reimer et al., 2013) for a range of age-depth model types (a-b linear interpolation, c-d smooth spline, e-f *Bacon*, g-h *Bchron*). Red curves indicates simulation 11136 (see Figure 2) focusing on the Younger Dryas Period; blue curves shows simulation 1102 around the 3 kcal BP Hallstatt Plateau. Precision and accuracy are shown in left and right panels, respectively. At high dating densities, single depths are dated several times, causing conflicting age estimates and resulting in unsuccessful *Bchron* runs for the YD simulation. Note logarithmic axes.

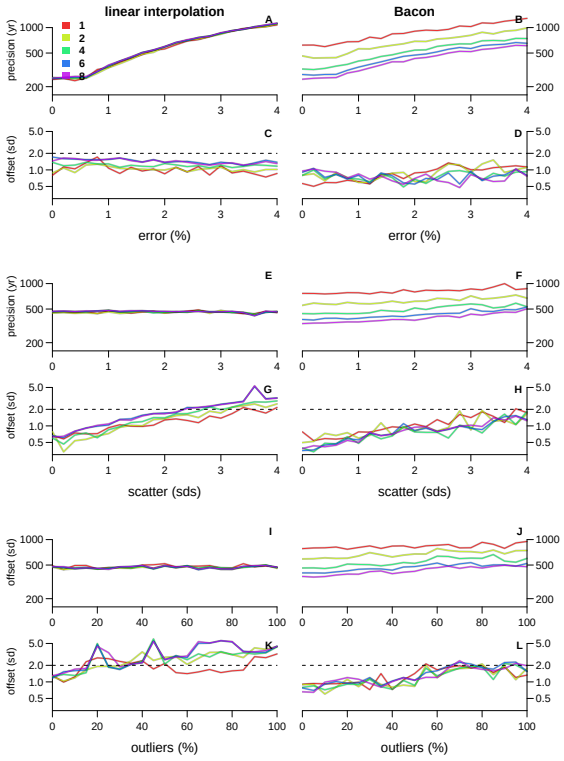

Figure 6: Impact of laboratory error (a-d), dating scatter (e-h) and outliers (i-l) on model reliability. Left panels show linear interpolation, right panels *Bacon*. Curves show mean precision and accuracy at a range of dpm (see legend in panel a). Given the time-consuming nature of these simulations, results are available only for linear interpolation and *Bacon*.