Miniature radiocarbon measurements (< 150 µg C) from sediments of Lake Żabińskie, Poland: effect of precision and dating density on age-depth models

5 Paul D. Zander[1], Sönke Szidat[2], Darrell S. Kaufman[3], Maurycy Żarczyński[4], Anna I. Poraj-Górska[4], Petra Boltshauser-Kaltenrieder[5], Martin Grosjean[1]

[1]Institute of Geography & Oeschger Centre for Climate Change Research, University of Bern, Bern, CH-3012, Switzerland
[2]Department of Chemistry and Biochemistry & Oeschger Centre for Climate Change Research, University of Bern, Bern, CH-3012, Switzerland
10 [3]School of Earth and Sustainability, Northern Arizona University, Flagstaff, AZ 86011, USA
[4]Faculty of Oceanography and Geography, University of Gdańsk, Gdańsk, 80-309, Poland
[5]Institute of Plant Sciences & Oeschger Centre for Climate Change Research, University of Bern, Bern, CH-3013, Switzerland

Correspondence to: Paul D. Zander (paul.zander@giub.unibe.ch)

Abstract

15 The recent development of the MIni CArbon DAting System (MICADAS) allows researchers to obtain radiocarbon (^{14}C) ages from a variety of samples with miniature amounts of carbon (< 150 µg C) by using a gas ion source input that bypasses the graphitization step used for conventional ^{14}C dating with accelerator mass spectrometry (AMS). The ability to measure smaller samples, at reduced cost compared with graphitized samples, allows for greater dating density of sediments with low macrofossil concentrations. In this study, we use a section of varved sediments from Lake Żabińskie, NE Poland, as a case
20 study to assess the usefulness of miniature samples from terrestrial plant macrofossils for dating lake sediments. Radiocarbon samples analyzed using gas-source techniques were measured from the same depths as larger graphitized samples to compare the reliability and precision of the two techniques directly. We find that the analytical precision of gas-source measurements decreases as sample mass decreases, but is comparable with graphitized samples of a similar size (approximately 150 µg C). The age uncertainty is consistently less than 150 years (uncalibrated 1σ) for samples larger than 40 µg C. The reliability of ^{14}C
25 ages from both techniques is assessed via comparison with a best-age estimate for the sediment sequence, which is the result of OxCal V-sequence that integrates varve counts with ^{14}C ages. No bias is evident in the ages produced by either gas-

source input or graphitization. None of the ^{14}C ages in our dataset are clear outliers; the 95% confidence intervals of all 48 calibrated ^{14}C ages overlap with the median best-age estimate. The effects of sample mass (which defines the expected analytical age uncertainty) and dating density on age-depth models are evaluated via simulated sets of ^{14}C ages that are used

30   as inputs for OxCal P-sequence age-depth models. Nine different sampling scenarios were simulated in which the mass of ^{14}C samples and the number of samples were manipulated. The simulated age-depth model results suggest that the lower analytical precision associated with miniature samples can be compensated for by increased dating density. The data presented in this paper can improve sampling strategies and can inform expectations of age uncertainty from miniature radiocarbon samples as well as age-depth model outcomes for lacustrine sediments.

35

Keywords: radiocarbon, MICADAS, lake sediments, OxCal, age-depth modeling

1 Introduction

Radiocarbon (^{14}C) dating is the most widely used technique to date sedimentary sequences that are less than 50,000 years old. The robustness of age-depth models can be limited by the availability of suitable material for dating; this is particularly a

40   problem for studies on sediments from alpine, polar, or arid regions where terrestrial biomass is scarce. Most accelerator mass spectrometry (AMS) labs recommend that samples contain 1 mg or more of carbon for reliable ^{14}C age estimations. It is well established that terrestrial plant macrofossils are the preferred material type for dating lake sediments because bulk sediments or macrofossils may have an aquatic source of carbon, which can bias ^{14}C ages (Groot et al., 2014; MacDonald et al., 1991; Tornqvist et al., 1992; Barnekow et al., 1998; Grimm et al., 2009). Furthermore, a high density of ^{14}C ages (i.e. one age per

45   500 years) is recommended to reduce the overall chronologic uncertainty of age-depth models (Blaauw et al., 2018). Researchers working on sediments with low abundances of terrestrial plant macrofossils face difficult choices about whether to date sub-optimal materials (e.g. bulk sediment or aquatic macrofossils), pool material from wide sample intervals, or rely on few ages for their chronologies. The problem of insufficient material can affect age estimates at all scales from an entire sedimentary sequence to a specific event layer which a researcher wishes to determine the age of as precisely as possible.

50

Recent advances have reduced the required sample mass for AMS ^{14}C analysis, opening new opportunities for researchers (Delqué-Količ et al., 2013; Freeman et al., 2016; Santos et al., 2007; Shah Walter et al., 2015). The recently developed MIni CArbon DAting System (MICADAS) has the capability to analyze samples with miniature masses via the input of samples in

a gaseous form, thus omitting sample graphitization (Ruff et al., 2007, 2010a, 2010b; Synal et al., 2007; Szidat et al., 2014;

55 Wacker et al., 2010a, 2013). Samples containing as little as a few µg C can be dated using the gas-source input of the MICADAS. The analysis of such small samples provides several potential benefits for dating lake sediments: 1) the possibility to date sediments that were previously not dateable using ^{14}C due to insufficient material, 2) the ability to date sedimentary profiles with a greater sampling density and lower costs per sample, and 3) the ability to be more selective when choosing what material will be analyzed for [14]C. The disadvantage of miniature samples is increased analytical uncertainty, which is

60 caused by lower counts of carbon isotopes and the greater impact of contamination on the measurement results. The goal of this study is to assess the potential benefits and limits of applying miniature ^{14}C measurements to dating lake sediments. We aim to answer the following questions in this study: 1) How reliable and how precise are gas-source ^{14}C ages compared with graphitized ages? 2) What is the variability of ^{14}C ages obtained from a single stratigraphic level? 3) How do analytical precision and dating density affect the accuracy and precision of age-depth models for lake sediments?

65

In this study, we use the sediments of Lake Żabińskie Poland, as a case study to investigate the application of gas-source ^{14}C measurements to lake sediments. We focus on a continuously varved segment of the core, which spans from roughly 2.1 to 6.8 ka. We report the results of 48 radiocarbon measurements (17 using graphitization and 31 using the gas-source input) in order to compare the precision and reliability of gas-source ^{14}C ages with graphitized samples. The core was sampled such that up

70 to five ages were obtained from 14 distinct stratigraphic depths. A floating varve chronology was integrated with the ^{14}C ages to produce a best-age estimate using the OxCal V-sequence routine (Bronk Ramsey, 2008). This best-age estimate is used as a benchmark for the ^{14}C results. The results of our ^{14}C measurements were used to constrain a statistical model designed to simulate sets of ^{14}C ages in order to test nine different hypothetical sampling scenarios in which we manipulate the number of ages and the mass of C per sample, which determines the analytical uncertainty of the simulated ages. By comparing the results

75 of the simulated age-depth model outputs from these simulated ^{14}C ages with the best-age estimate from which the simulated ages were derived, we can improve our understanding of how the number of ages and their analytical precision influence the accuracy and precision of radiocarbon-based age-depth models.

2 Materials and Methods

2.1 Core material and radiocarbon samples

80 Cores were obtained from Lake Żabińskie (coring site: 54.1318° N, 21.9836° E, 44 m water depth) in 2012 using an UWITEC piston corer (90 mm diameter). Lake Żabińskie is a small (41.6 ha), relatively deep (44.4 m) kettle-hole lake with a catchment of 24.8 km^2. A composite sediment profile was constructed from overlapping, 2-m-long cores by correlating distinctive stratigraphic features. The composite sequence spans 19.4 m. Varve-based chronologies and ^{14}C measurements have been published for the most recent 2000 years of the Lake Żabińskie sedimentary sequence (Bonk et al., 2015; Żarczyński et al.,

85 2018). Published downcore varve counts stop above a ~90-cm-thick slump/deformed unit that is dated to 1962-2071 cal yr BP (present = 1950 CE). This study focuses on a section of core (7.3-13.1 m depth in our composite sequence) directly below this slump unit, which was selected because it features well-preserved varves continuously throughout the section.

Samples of 1- to 2-cm-thick slices of sediment were taken from the core (sample locations and core images are found in

90 Supplementary File 1), then sieved with a 100 μm sieve. Macrofossil remains were identified and photographed (Supplementary File 2), and only identifiable terrestrial plant material was selected for ^{14}C measurements. Suitable macrofossils from a single stratigraphic level were divided into subsamples for analysis, with the goal of producing one graphitized ^{14}C age and 2-4 gas-source ages from each depth. When convenient, we grouped samples by the type of material (leaves, periderm, needles, seeds or woody scales), though 11 samples are a mixture of material types. In most cases,

95 subsamples within a stratigraphic level are assumed to be independent, meaning they may have different true ages. However, there are some subsamples that were taken from single macrofossil fragments (six subsamples taken from two fragments sampled from two different depths), thus these samples have the same true age. It is also possible that subsamples from a single depth may be from the same original material without our knowledge (i.e. a macrofossil could break into several pieces while sieving, and these pieces could be analyzed as separate subsamples).

100

Sample material was treated with an acid-base-acid (ABA) method at 40°C, using 0.5 mol/L HCl, 0.1 mol/L NaOH and 0.5 mol/L HCl for 3 h, 2 h and 3 h, respectively. After drying at room temperature, samples were weighed, and those less than 300 μg were input to the gas ion source via combustion in an Elementar Vario EL Cube elemental analyser (Salazar et al., 2015). Larger samples were graphitized following combustion using automated graphitization equipment (AGE) (Szidat et al., 2014).

105 Radiocarbon data was processed using the software BATS (Wacker et al., 2010a). Additional corrections were applied to the

data to account for cross contamination (carryover), and constant contamination (blanks) (Gottschalk et al., 2018; Salazar et al., 2015). The parameters for these corrections were calculated based on standard materials (the primary NIST standard oxalic acid II (SRM 4990C) and sodium acetate (Sigma-Aldrich, No. 71180) as ^{14}C-free material) run with the sample batches. We applied a constant contamination correction of 1.5 ± 0.2 µg C with 0.72 ± 0.11 F^{14}C and a cross contamination correction of

110 $(1.2 \pm 0.3\ \%)$ from the previously run sample. Radiocarbon age uncertainties were fully propagated for each correction. In total, 48 ages were obtained from 14 distinct stratigraphic levels (17 graphitized and 31 gas-source measurements).

2.2 Varve count

Varves in Lake Żabińskie are biogenic, with calcite-rich laminae deposited in spring and summer, and darker laminae containing organic detritus and fine clastic material deposited in winter (Żarczyński et al., 2018). We defined the boundary of

115 each varve year by the onset of calcite precipitation (i.e., the upper boundary of dark laminae and lower boundary of light-colored laminae). Varves were counted using CooRecorder software (Larsson, 2003) on core images obtained from a Specim PFD-CL-65-V10E linescan camera (Butz et al., 2015). Three people performed independent varve counts, and these three counts were synthesized, and uncertainties calculated according to the methodology recommended by Żarczyński et al. (2018).

120 Because of the slump deposit above our section of interest, the varve chronology is 'floating' and must be constrained by the ^{14}C ages. Several different approaches could be used to compare the varve count with the ^{14}C ages, all of which rely on some assumptions. One method would be to select a dated level within the core and tie the varve count to the age at this level. Such an approach assumes that the radiocarbon-based age at the tie point is correct. Instead, we used the OxCal V-sequence to integrate all available chronological information including varve counting and ^{14}C ages into a single model to determine a best-

125 age estimate for the sequence. The advantages of this approach are that all ages are considered equally likely to be correct, and the error estimate of the V-sequence is relatively consistent along the profile, whereas the error associated with the varve count is small at the top of the section, but increases downcore.

2.3 Age-depth modeling

Age-depth modeling was performed using OxCal 4.3, which integrates the IntCal13 calibration curve for ^{14}C ages with

130 statistical models that can be used to construct age-depth sequences (Bronk Ramsey, 2008, 2009; Bronk Ramsey and Lee, 2013; Reimer et al., 2013). As an initial test to compare the reliability of gas-source ages and graphitized ages, and their effect on age-depth models, we produced three sequence models: one using all obtained ^{14}C ages, one using only graphitized ages,

and one using only gas-source ages. For all OxCal models in this study, ages measured from the same depth were combined (using the function *Combine*) into a single ^{14}C age with uncertainty before calibration and integration into the age-depth

135 sequence. The OxCal P-sequence uses a Bayesian approach for modelling sediment deposition in which a parameter (k) determines the extent to which sedimentation rates are allowed to vary. For all P-sequence models in this study, we used a uniformly distributed prior for k such that $k_0 = 1$, and $\log_{10}(k/k_0) \sim U(-2, 2)$; this allows k to vary between 0.01 and 100.

The varve counts and all ^{14}C ages were incorporated into an OxCal V-sequence in an approach similar to that used by Rey et

140 al. (2019). We input the number of varves in 10 cm intervals to the V-sequence as an age 'Gap' with associated uncertainty. The OxCal V-sequence assumes normally distributed uncertainties for each gap, whereas our varve count method produces asymmetric uncertainty estimates. We used the mean of the positive and negative uncertainties as input to the V-sequence. However, OxCal sets the minimum uncertainty of each 'Gap' equal to 5 years, which in most cases is larger than the mean uncertainty in our varve count over a 10 cm interval. The V-sequence combines the varve information with the ^{14}C ages (both

145 graphitized and gas-source ages) to produce a more precise age-depth sequence.

2.4 Age-depth model simulation

In order to test the effects of analytical uncertainty and dating density (number of ages per time interval) on age-depth models, we designed an experiment in which nine different sampling scenarios were simulated for the Lake Żabińskie sedimentary sequence to determine the expected precision and accuracy of resulting age-depth models. Three different sampling densities

150 were simulated for the 5.8-m-long section: 5 ages, 10 ages, and 20 ages (equivalent to approximately 1, 2, and 4 ages per millennium, respectively). For each of these sampling densities three different sample-size scenarios were simulated: 35 µg C, 90 µg C, 500 µg C. These scenarios were designed to represent different sampling circumstances such as high or low abundances of suitable material for ^{14}C analysis, and different budgets for ^{14}C analysis. Radiocarbon ages were simulated using a technique similar to Trachsel and Telford (2017). In brief, we distributed the simulated samples evenly by depth across the

155 5.8-m-long section, and then used the median output of the OxCal V-sequence as the assumed true age for a given depth. This calibrated assumed true age was back-converted to ^{14}C years using IntCal13 (Reimer et al., 2013). A random error term was added to the ^{14}C age to simulate the analytical uncertainty. The error term was drawn from a normal distribution with mean zero and standard deviation equivalent to the based on the relationship between sample mass and precision found in the results of our ^{14}C measurements (1). The same expected analytical uncertainty was used for the age uncertainty for each simulated

160 age. These simulated ^{14}C ages were input into an OxCal P-sequence using the same uniform distribution for the k-parameter

as described in the previous section. This experiment was repeated 30 times for each scenario to assess the variability of possible age-model outcomes. We quantify the accuracy of the age-depth models as the deviation of the median modelled age from the true age at a given depth. We define precision as the width of the age-depth model confidence interval (CI).

3 Results

165   ### 3.1 Radiocarbon measurements

In total, 48 radiocarbon measurements on terrestrial plant macrofossils were obtained from the section of interest resulting in a range of ages from 8 to 5988 ^{14}C years (Table 1). Thirty-one ages were measured using the gas-source input; these samples contained between 11 and 168 µg C. Seventeen samples containing between 115 and 691 µg C were measured using graphitization. Analytical uncertainties for the ^{14}C measurements range from ± 41 to ± 328 ^{14}C years with higher values

170   associated with the smallest sample masses. The uncertainties for gas-source measurements and graphitized measurements are comparable for samples that contain a similar amount of carbon (Figure 1). Based on an assumed Poisson distribution of the counting statistics, one would expect age uncertainty to decrease as sample masses become larger following the relationship $N^{-0.5}$, where N is the number of the measured ^{14}C atoms in the sample. This relationship fits our data well for larger samples, however, as the mass of C is reduced, the uncertainty becomes greater than predicted by this relationship due to corrections

175   applied for cross-contamination and constant contamination (see Sect. 2.1; Gottschalk et al., 2018; Salazar et al., 2015), which have a greater effect on smaller samples. Samples containing less than 40 µg C (roughly equivalent to 80 µg of dry plant material) produce uncertainties greater than ± 150 years (1σ). We use a power-model fit with least-squares regression, to estimate the typical age uncertainty for a given sample mass ($r^2 = 0.90$, $p < 0.001$, Fig. 1).

180   When comparing measurements taken from within a single sediment slice we find good agreement for all ^{14}C ages, regardless of whether the samples were analyzed with the gas-source input or via a graphitized target, and no clear bias based on the type of macrofossil that was dated (Figure 2, Figure 3). One method to test whether the scatter of ages is consistent with the expectations of the analytical uncertainty is a reduced chi-squared statistical test, also known as Mean Square Weighted Deviation (MSWD) in geochronological studies (Reiners et al., 2017). If the spread of ages is exactly what would be expected

185   from the analytical uncertainty, the value of this statistic is 1. Lower values represent less scatter than expected, and larger values represent more scatter than expected. Of the 11 sampled depths with three or more ages, only one grouping of ages (811

cm, MSWD = 3.07) returned an MSWD that exceeds a 95% significance threshold for acceptable MSWD values that are consistent with the assumption that the age scatter is purely the result of analytical uncertainty.

3.2 Varve count and age-depth modeling

190   In total, 4644 (+155/- 176) varves were counted in the section of interest, with a mean varve thickness of 1.26 ± 0.58 mm (full varve count results are available at https://dx.doi.org/10.7892/boris.134606). Sedimentation rates averaged over 10 cm intervals range from 0.91 to 2.78 mm/year. All chronological data (^{14}C ages and varve counts) were integrated to generate a best-age estimate for the section of interest using an OxCal V-sequence (output of the Oxcal V-sequence is available at https://dx.doi.org/10.7892/boris.134606). This produced a well-constrained age-depth model with a 95% confidence interval

195   (CI) width that ranges from 69 to 114 years (mean 86 years). OxCal uses an agreement index to assess how well the posterior distributions produced by the model (modelled ages at the depth of ^{14}C ages) agree with the prior distributions (calibrated ^{14}C ages). The agreement index for our OxCal V-sequence is 66.8%, which is greater than the acceptable index of 60%. Three of the fourteen dated levels in the V-sequence had agreement indices less than the acceptable value of 60% (A = 22.8, 48.5, 52.6% for sample depths = 1283.0, 1176.1, 732.5 cm, respectively), nonetheless we find the model fit acceptable as all 48 ^{14}C

200   ages overlap with the median output of the V-sequence. We use the V-sequence as a best-age estimate for subsequent data comparisons and analyses.

To test the reliability of gas-source ages versus graphitized ages we created three OxCal P-sequences using: 1) all ^{14}C ages, 2) only graphitized ages, and 3) only gas-source ages. The results of all three of these age-depth models agree well with the best-

205   age estimate of the V-sequence, although with larger 95% CIs (Figure 2). The agreement index was greater than the acceptable value of 60 for all three models overall, and for each dated depth within all three models. The P-sequence using all ^{14}C ages spans 4838 ± 235 years, which is slightly greater than, but overlapping with, the total number of varves counted (the V-sequence estimates 4681 ± 79 years in the section). There is no clear bias observed in the age-depth models produced using either the gas-source or graphitized samples. The age-depth model outputs clearly show that a very precise age can narrowly

210   constrain the age-model uncertainty at the depth of that sample, however, if dating density is low, the uncertainty related to interpolation between ages becomes large. Despite the lower precision of the gas-source ages, the model based on only gas-source ages actually has a lower mean CI width than the model with graphitized ages (mean 95% CI width: 373 years for the gas-source model, 438 years for the graphitized model). However, the direct comparison between the gas-source-only and the graphitized-only age models is confounded by differences in the number and spacing of samples. Specifically, there are no

215   graphitized ages between the top of the section (724 cm) and 811 cm, and between 1082 and 1200 cm, which results in wide
       CI in these sections. On the other hand, uncertainty is reduced compared to the gas-source model in the depths adjacent to the
       graphitized ages due to higher precision such that 40% of the section (in terms of depth) has lower age uncertainty in the
       graphitized model.

3.3 Age-depth model simulations

220   Nine different sampling scenarios (described in Sect. 2.3) were simulated to test the effects of dating density and analytical
       precision on age-depth model confidence intervals. For each of the nine scenarios, sets of ^{14}C ages were simulated 30 times to
       create an ensemble of age-depth models for each scenario. One set of these simulated age-depth models is shown in Figure 4,
       and an animation of the full set of simulated models is available online (Supplementary File 3). The age-depth models were
       evaluated for their precision (mean width of the 95% CI) and accuracy (the mean absolute deviation from the best-age estimate;
225   summarized in Figure 5 and Table 2). As expected, we find that increased dating density and increased sample masses improve
       both the accuracy and precision of the age-depth models. It is notable that increasing the number of ages can compensate for
       the greater uncertainty associated with smaller sample sizes. For instance, the mean CI of age-depth models based on ten, 90
       μg C samples is narrower than age-depth models with five, 500 μg C samples (Table 2). However, the effect of analytical
       precision is greater on the mean absolute deviation from the best-age estimate. Increased dating density does tend to reduce
230   the deviation from the best-age estimate (especially if the ages are imprecise), but the three scenarios that use 500 μg samples
       perform better than all other scenarios, as applied to our study site, in terms of deviation from the best-age estimate, regardless
       of the sampling density. Additionally, increased dating density does not improve the deviation from the best-age estimate for
       the 500 μg sample scenarios. This result may be due to the relatively constant sedimentation rates in our sedimentary sequence,
       which reduces errors caused by interpolation in scenarios with low dating density. Another prominent pattern visible in the
235   simulations is the large spread of performance for models with relatively few and imprecise ages (Figure 5). Increasing the
       number of samples and, especially, the mass of samples has a large impact on the agreement among the model ensembles.

       An additional measure of age-model quality is the Chron Score rating system (Sundqvist et al., 2014), which uses three criteria
       to assess the reliability of age-depth models: 1) delineation of downcore trend (D), 2) quality of dated materials (Q), and 3)
240   precision of calibrated ages (P). These metrics are combined using a reproducible formula to provide a Chron Score (G) in
       which higher values represent more reliable age-depth models:

$$G = -w_D D + w_Q Q + w_P P$$

245   We used the default weighting parameters (w_D, w_Q, and wP = 0.001, 1 and 200) for each component of the Chron Score
formula as described in Sundqvist et al. (2014). The Q parameter depends on two factors – the proportion of ages which are
not rejected or reversed (i.e. an older age stratigraphically above a younger age), and a qualitative classification scheme for
material types. We modified the threshold for determining if an age is considered a reversal such that if a ^{14}C age is older than a
stratigraphically higher age by more than the age uncertainty (1σ), the age is considered to be stratigraphically reversed. This
250   is different from the default setting, which is 100 years. For the material type classification (m), the simulated age models were
assigned the value 4, which is the value assigned to chronologies based on terrestrial macrofossils. For more details on the
Chron Score calculation see Sundqvist et al. (2014). The mean Chron Scores for the simulated age models (Table 2) show that
doubling dating density substantially improves the Chron Score, but the effect is greater when moving from 5 to 10 ages than
from 10 to 20 ages. The effect of increased precision on the Chron Score is also substantial; it is essentially defined by the
255   Chron Score formula, in which precision is assessed as $P = s^{-1}$ where s is the mean 95% range of all calibrated ^{14}C ages. The
effect of precision on the Chron Score is also determined by the weighting factors mentioned above.

4 Discussion

4.1 Radiocarbon measurements

The results of our ^{14}C measurements from repeated sampling of single stratigraphic levels provide useful information for other
260   researchers working with miniature ^{14}C analyses, or any ^{14}C samples from lake sediments. We show that there is an exponential
relationship between sample mass and the resulting analytical uncertainty. The exact parameters of this relationship will
depend on several factors that are not considered here, such as the laboratory conditions, and the age of the material (Gottschalk
et al., 2018), however, the general shape of the relationship should hold. These data can inform researchers about the expected
range of uncertainty for ^{14}C ages from samples of a given size. We find that samples larger than 40 µg C yield ages that are
265   precise enough to be useful for dating Holocene lake sediments in most applications, and even smaller samples can provide
useful ages if no other material is available.

It is well documented that ^{14}C ages can be susceptible to sources of error that are not included within the analytical uncertainty
of the measurements. Such errors can be due to lab contamination, sample material which is subject to reservoir effects (i.e.

270   bulk sediments or aquatic organic matter; Groot et al., 2014; MacDonald et al., 1991; Tornqvist et al., 1992), or from depositional lags (terrestrial organic material which is older than the sediments surrounding it; Bonk et al., 2015; Howarth et al., 2013; Krawiec et al., 2013). Errors related to reservoir effects can be avoided by selecting only terrestrial plant material for dating (Oswald et al., 2005). Dating fragile material such as leaves (as opposed to wood) may reduce the chances of dating reworked material with a depositional lag, but generally this source of error is challenging to predict and dependent on the

275   characteristics of each lake's depositional system. To identify ages affected by depositional lags, it is necessary to compare with other age information. Consequently, the identification of outlying ages is facilitated by increased dating density.

In our dataset, multiple ^{14}C measurements were performed on material taken from a single layer, which enables outlier detection. We find that the scatter of ^{14}C ages obtained from the same depths is generally consistent with what would be

280   expected based on the analytical uncertainties of the ages. There are no clear outliers in the data, and every single ^{14}C age has a calibrated 95% CI that overlaps with the median of our best-age estimate OxCal V-sequence. This can be explained in part by the fact that the V-sequence is fit to the ^{14}C ages, but it is also evidence that no age in this dataset is incongruent with the other available chronological information (other ^{14}C ages and varve counts). This notion is further demonstrated by the fact that 10 of 11 sampled levels from which we obtained three or more ages returned an MSWD within the 95% confidence

285   threshold for testing age scatter (see Sect. 3.1; Reiners et al., 2017). This test is typically used for repeated measurements on the same sample material, however, in our study, many of the measurements from within a single sediment slice are from material that has different true ages. The MSWD test indicates that the variability in ages among samples from within a single sediment slice can reasonably be expected given the analytical uncertainty. However, in this study, no more than five samples were measured per depth, and thus the range of acceptable values for the MSWD is relatively wide due to the small number

290   of degrees of freedom. Additionally, the analytical uncertainties are relatively large for the gas-source samples, allowing for wide scatter in the data without exceeding the MSWD critical value. Despite these caveats, the consistency between the variability among ages from one level and the analytical uncertainties allows us to make two important conclusions. 1) The analytical precision estimates are reasonable, even for miniature gas-source samples. 2) When material is carefully selected and taxonomically identified for dating, the sources of error that are not considered in the analytical uncertainty (e.g.

295   contamination or depositional lags) are relatively minor in our case study. However, this second conclusion is highly dependent on the sediment transport and depositional processes, which are site specific. Depositional lags still likely have some impact on our chronology. Six ^{14}C ages from plant material collected from the Lake Żabińskie catchment in 2015 yielded a range of ages from 1978-2014 CE (Bonk et al., 2015) suggesting that the assumption that ^{14}C ages represent the age of the sediments

11

surrounding macrofossils is often invalid. The scale of these age offsets is likely on the scale of a few decades for Lake
300   Żabińskie sediments, which is inconsequential for many radiocarbon-based chronologies, but is the same order of magnitude
as the uncertainty of our best-age estimate from the OxCal V-sequence, and should be considered when reporting or
interpreting radiocarbon-based age determinations with very high precision.

The lack of outliers in our dataset is an apparent contrast with the findings of Bonk et al. (2015), who report that 17 of 32
305   radiocarbon samples taken from the uppermost 1000 years of the Lake Żabińskie core were outliers. The outlying ages were
older than expected based on the varve chronology, and this offset was attributed to reworking of terrestrial plant material. The
identification of outliers did not take into account uncertainties of the radiocarbon calibration curve and varve counts, which
could explain some of the differences between the ^{14}C and the varve ages. Still, 8 of 32 ages reported by Bonk et al. (2015)
have calibrated 2σ age ranges that do not overlap with varve count age (including the varve count uncertainty). The higher
310   outlier frequency in the Bonk et al. (2015) data might be explained by their generally more precise ages and the fact that their
varve count is truly independent from the ^{14}C ages.

Additionally, our dataset allows us to compare the results of ^{14}C ages obtained from different types of macrofossil materials,
which we grouped into the following categories: leaves (including associated twigs), needles, seeds, periderm, woody scales,
315   and samples containing mixed material types (Figure 3). When comparing the calibrated median age of each sample to the
median of our best-age estimate, we find that the difference between the age offsets of the different material types is not
significant at the $\alpha = 0.05$ level (ANOVA, F = 2.127, p = 0.08). This is likely due to our selective screening of sample material,
which only includes terrestrial plant material while avoiding aquatic insect remains or possible aquatic plant material, as well
as the relatively small number of samples within each material type. There does appear to be a tendency for seeds to produce
320   younger ages, and two of the three woody scale samples yielded ages that are approximately 300 years older than the best-age
estimate. This could be due to the superior durability of woody materials compared with other macrofossil materials, which
enables wood to be stored on the landscape prior to being deposited in the lake sediments. A larger number of samples would
allow for more robust conclusions about the likelihood of certain material types to produce biased ages.

4.2 The OxCal V-sequence best-age estimate

325   Prominent varves in the sediments of Lake Żabińskie provide additional chronological information that we use to inform our
assessment of ^{14}C ages. This approach to integrating varve counts with ^{14}C ages can provide more precise and more reliable

age estimates than either technique alone. The resulting age-depth relation has a relatively narrow CI (mean 95% CI is 86 yr). Extremely precise age estimates were also produced using this method for Moossee, Switzerland by Rey et al. (2019). A combination of varve counts and ^{14}C ages from the Moossee sediments generated a V-sequence output with a mean 95% CI

330 of 38 years. The higher precision in the Moossee study compared to our V-sequence output is primarily attributed to the higher dating density in Moosse with 27 radiocarbon ages over ~3000 years (3.9-7.1 ka) versus our study, which used 48 ages, but from only 14 unique depths, over ~4700 years. This comparison shows that repeated measurements from the same depth are less useful than analyses from additional depths. This approach to integrating varve counts and ^{14}C ages could potentially be improved by a better integration of varve count uncertainties into the OxCal program. Currently the uncertainties on age 'Gaps'

335 in OxCal are assumed to be normally distributed and cannot be less than 5 years. Nevertheless, the result of the OxCal V-sequence is an age-depth model that is much more precise than those constructed only using ^{14}C ages and provides a useful reference to compare with the ^{14}C ages. It is important to note that the best-age estimate is not independent of the ^{14}C ages; it is directly informed by the ^{14}C ages.

4.3 Age-depth model simulations

340 The simulated age-depth modelling experiment allows us to assess the effects of dating density and sample mass (expected precision) on the outputs of age-depth models constructed for the section of interest in the Lake Żabińskie sediment core. Models based on relatively few, but very precise ages, are tightly constrained at the sample depths, but the CI widens further away from these depths (Figure 4, Supplementary File 3). In contrast, models based on a greater sampling density produce confidence intervals with relatively constant width. If models are built using a high density of imprecise ages, the CI of the

345 model output can actually be narrower than the CI of the individual ages. Bayesian age-depth models in particular can take advantage of the stratigraphic order of samples to constrain age-depth models to be more precise than the individual ages that make up the model (Blaauw et al., 2018), however this is only achievable when dating density is high enough. The results from this experiment suggest that, in the case of the Lake Żabińskie sequence, doubling the number of ages can approximately compensate for an increased analytical uncertainty of 50 years.

350

The Chron Score results provide a succinct summary of the reliability of the chronologies produced in the different simulated sampling scenarios. The Chron Score becomes more sensitive to changes in precision as precision increases, so the difference in the Chron Scores between the 500 µg and 90 µg scenarios (1σ uncertainty of ± 39 and 92 years, respectively) is greater than the difference between the 90 µg and 35 µg scenarios (1σ uncertainty of ± 92 and 148 years, respectively). Increased dating

13

355　density consistently improves the Chron Score results, with a stronger impact seen when shifting from 5 to 10 ages compared to shifting from 10 to 20 ages. The improvement of the Chron Score due to increased dating density is generally consistent for each of the different sample mass scenarios This differs from the age-depth model statistics where increased dating density has a greater impact on mean age-depth model precision in the larger sample mass scenarios (more precise ages). The opposite effect is seen in the mean absolute deviation results, where mean absolute deviation is reduced substantially as dating density

360　increases for the smaller sample scenarios, and not at all for the 500 µg scenario. For all measures of chronologic performance, we find a greater improvement when increasing the number of ages from 5 to 10 ages compared to increasing from 10 to 20 ages, suggesting there are some diminishing returns from increased dating density. This result is in accordance with the results of Blaauw et al. (2018). While the Chron Score results are strictly dependent on the parameters chosen for the calculation, they intuitively make sense. Because Chron Score results use only the simulated 14C ages as input and are unaffected by the age

365　modelling routine, the patterns exhibited in the scores may be more applicable to a variety of sedimentary records.

　In real-world applications, there are additional advantages from increasing dating density. Many lacustrine sequences have greater variability in sedimentation rates than the sequence modelled here. More fluctuations in sedimentation rate require a greater number of ages to delineate the changes in sedimentation. Additionally, outlying ages and age scatter beyond analytical

370　uncertainty are not considered in this modelling experiment. In most cases, detecting outlying ages becomes easier as dating density increases. Because this experiment is only applied to a single sedimentary sequence, the results may not be directly applicable for other sedimentary records with different depositional conditions. In the future, this type of age model simulation could be applied to a range of sedimentary sequences with a variety of depositional conditions.

4.4 Recommendations for radiocarbon sampling strategy

375　Radiocarbon sampling strategies will always be highly dependent on project-specific considerations such as how the chronology will affect the scientific goals of the project, budget and labor constraints, the nature of the sedimentary record in question, and the availability of suitable materials. A goal of this study is to provide data that can inform sampling strategies for building robust chronologies, particularly in cases where suitable material may be limited. Firstly, an iterative approach to ^{14}C measurements is preferred. An initial batch of measurements should target a low dating density of perhaps one date per

380　2000 years. Subsequent samples should aim to fill in gaps where age uncertainty remains highest (Blaauw et al., 2018), or where preliminary age-depth trends appear to be non-linear. In accordance with many previous studies (e.g. Howarth et al., 2013; Oswald et al., 2005), we advocate for careful selection of material identified as terrestrial in origin. If the mass of such

14

material is limited, the MICADAS gas-source is useful for dating miniature samples, and we are convinced that miniature samples of terrestrial material are preferable to dating questionable material or bulk sediments. Samples as small as a few μg

385 C can be measured using the MICADAS, though samples larger than 40 μg C are recommended for more precise results (Holocene samples with 40 μg C are expected to have analytical uncertainty of ~138 years). Dating small amounts of material from single depths is also preferable to pooling material from depth segments that may represent long time intervals. A general rule of thumb is to avoid taking samples with depth intervals representing more time than the expected uncertainty of a ^{14}C age. To improve the accuracy of age-depth models, a higher priority should be placed on achieving sufficiently high dating

390 density (ideally greater than one age per 500 years; Blaauw et al., 2018) using narrow sample-depth intervals. In most cases, this goal should be prioritized over the goal of gathering larger sample masses in order to reduce analytical uncertainties. Multiple measurements from within a single stratigraphic depth, as we have done in this study, can be useful in sediments where age scatter (possibly from reworked material) is expected. In such cases, multiple measurements from a single depth could allow for identification of certain types of material that should be avoided, and if age results do not agree well, the

395 youngest age is most likely to be correct (assuming no contamination by modern carbon). If age scatter is not expected, single measures of pooled macrofossils are more cost-effective than repeat measurements from a single depth. Although increased dating density does incur greater cost, gas-source ages have reduced costs substantially compared to graphitized ages allowing for greater dating density at similar cost. Analytical costs for gas-source analyses are laboratory specific but there is a substantial reduction in cost for gas-source ages compared to graphitized sample measurements. Use of smaller samples can

400 reduce the labor time required to isolate suitable material from the sediment, however handling and cleaning miniature samples can add additional challenges.

5 Conclusions

- AMS ^{14}C analysis of Holocene terrestrial plant macrofossils using the MICADAS gas-ion source produces unbiased ages with similar precision compared to graphitized samples that contain similar mass of carbon (approximately 120-160 μg

405 C).

- The precision of a ^{14}C age can be approximately estimated based on the amount of carbon within a sample. Holocene samples containing greater than 40 μg C produce ages with analytical uncertainty expected to be less than 150 years. Uncertainty increases exponentially as samples get smaller so 10 μg C samples are expected to have 277 years of uncertainty.

15

410 • The variability among ages obtained from 1- or 2-cm-thick samples in the Lake Żabińskie sediment core is compatible with the variability expected due to analytical uncertainty alone.

• We find no clear evidence in our dataset for age bias based on the type of macrofossil material dated, which we limited to terrestrial plant material.

• Judging from the output of age-depth models, the lower precision of miniature gas-source ages can be compensated for

415 by increasing sampling density. Based on sets of simulated ^{14}C ages that mimic the ^{14}C ages of our study core, together with age-depth models generated using OxCal, doubling dating density roughly compensates for a decrease in analytical precision of 50 years.

• The effect of ^{14}C age precision is among several factors that influence chronological precision. The thickness of the depth interval used to obtain samples, the ability to select identifiable terrestrial materials or to analyze more than one type of

420 material, the reliability of detecting age outliers, and the amount of variability in sedimentation rate all determine the accuracy and precision of an age-depth model, which are both improved by increasing the number of ages.

• This study can inform sampling strategies and provide expectations about radiocarbon-based age-depth model outcomes.

Data Availability

The key datasets associated with this manuscript (Varve count results and the best-age estimate OxCal V-sequence output) are

425 available at the Bern Open Repository and Information System, https://dx.doi.org/10.7892/boris.134606

Author Contributions

PZ prepared samples, designed and performed the age modelling experiment, analyzed results, and prepared the manuscript with contributions from all authors. MG, DK, SS, and PZ designed the strategy and goals of the study. MZ, APG, and PZ performed varve counting. PBK identified and selected suitable macrofossils. SS oversaw ^{14}C analyses. DK assisted with

430 laboratory work.

Competing interests

The authors declare no competing interests

16

Acknowledgements

17

References

440

Blaauw, M., Christen, J. A., Bennett, K. D. and Reimer, P. J.: Double the dates and go for Bayes — Impacts of model choice, dating density and quality on chronologies, Quat. Sci. Rev., 188, 58–66, doi:10.1016/j.quascirev.2018.03.032, 2018.

Bonk, A., Tylmann, W., Goslar, T., Wacnik, A. and Grosjean, M.: Comparing varve counting and ^{14}C-Ams chronologies in the sediments of Lake Żabińskie, Northeastern Poland: Implications for accurate ^{14}C dating of lake sediments, Geochronometria, 42(1), 157–171, doi:10.1515/geochr-2015-0019, 2015.

Bronk Ramsey, C.: Deposition models for chronological records, Quat. Sci. Rev., 27(1–2), 42–60, doi:10.1016/j.quascirev.2007.01.019, 2008.

Bronk Ramsey, C.: Bayesian Analysis of Radiocarbon Dates, Radiocarbon, 51(1), 337–360, doi:10.1017/s0033822200033865, 2009.

Bronk Ramsey, C. and Lee, S.: Recent and Planned Developments of the Program OxCal, Radiocarbon, 55(2), 720–730, doi:10.1017/s0033822200057878, 2013.

Butz, C., Grosjean, M., Fischer, D., Wunderle, S., Tylmann, W. and Rein, B.: Hyperspectral imaging spectroscopy: a promising method for the biogeochemical analysis of lake sediments, J. Appl. Remote Sens., 9(1), 096031, doi:10.1117/1.jrs.9.096031, 2015.

Delqué-Količ, E., Comby-Zerbino, C., Ferkane, S., Moreau, C., Dumoulin, J. P., Caffy, I., Souprayen, C., Quilès, A., Bavay, D., Hain, S. and Setti, V.: Preparing and measuring ultra-small radiocarbon samples with the ARTEMIS AMS facility in Saclay, France, in Nuclear Instruments and Methods in Physics Research, Section B: Beam Interactions with Materials and Atoms., 189–193, doi:10.1016/j.nimb.2012.08.048, 2013.

Freeman, E., Skinner, L. C., Reimer, R., Scrivner, A. and Fallon, S.: Graphitization of small carbonate samples for paleoceanographic research at the godwin radiocarbon laboratory, University of Cambridge, Radiocarbon, 58(1), 89–97, doi:10.1017/RDC.2015.8, 2016.

Gottschalk, J., Szidat, S., Michel, E., Mazaud, A., Salazar, G., Battaglia, M., Lippold, J. and Jaccard, S. L.: Radiocarbon Measurements of Small-Size Foraminiferal Samples with the Mini Carbon Dating System (MICADAS) at the University of Bern: Implications for Paleoclimate Reconstructions, Radiocarbon, 60(2), 469–491, doi:10.1017/RDC.2018.3, 2018.

Groot, M. H. M., van der Plicht, J., Hooghiemstra, H., Lourens, L. J. and Rowe, H. D.: Age modelling for Pleistocene lake sediments: A comparison of methods from the Andean Fúquene Basin (Colombia) case study, Quat. Geochronol., 22,

144–154, doi:10.1016/j.quageo.2014.01.002, 2014.

Howarth, J. D., Fitzsimons, S. J., Jacobsen, G. E., Vandergoes, M. J. and Norris, R. J.: Identifying a reliable target fraction for

470   radiocarbon dating sedimentary records from lakes, Quat. Geochronol., 17, 68–80, doi:10.1016/j.quageo.2013.02.001, 2013.

Krawiec, A. C. L., Kaufman, D. S. and Vaillencourt, D. A.: Age models and tephrostratigraphy from two lakes on Adak Island, Alaska, Quat. Geochronol., 18, 41–53, doi:10.1016/j.quageo.2013.07.002, 2013.

Larsson, L.: COORECORDER v2. 3.13: Image Co-ordinate Recording Program, 2003.

475 MacDonald, G. M., Beukens, R. P. and Kieser, W. E.: Radiocarbon Dating of Limnic Sediments: A Comparative Analysis and Discussion, Ecology, 72(3), 1150–1155, doi:10.2307/1940612, 1991.

Oswald, W. W., Anderson, P. M., Brown, T. A., Brubaker, L. B., Feng, S. H., Lozhkin, A. V., Tinner, W. and Kaltenrieder, P.: Effects of sample mass and macrofossil type on radiocarbon dating of arctic and boreal lake sediments, Holocene, 15(5), 758–767, doi:10.1191/0959683605hl849rr, 2005.

480 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. and van der Plicht, J.: 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,

485   2013.

Reiners, P. W., Carlson, R. W., Renne, P. R., Cooper, K. M., Granger, D. E., McLean, N. M. and Schoene, B. (Eds.): Interpretational approaches: making sense of data, in: Geochronology and Thermochronology, John Wiley & Sons, Ltd., Hoboken, NJ, USA, 65–82, 2017.

Rey, F., Gobet, E., Szidat, S., Lotter, A. F., Gilli, A., Hafner, A. and Tinner, W.: Radiocarbon wiggle matching on laminated

490   sediments delivers high-precision chronologies, Radiocarbon, 61(1), 265–285, doi:10.1017/RDC.2018.47, 2019.

Ruff, M., Wacker, L., Gäggeler, H. W., Suter, M., Synal, H. A. and Szidat, S.: A gas ion source for radiocarbon measurements at 200 kv, Radiocarbon, 49(2), 307–314, doi:10.1017/S0033822200042235, 2007.

Ruff, M., Szidat, S., Gäggeler, H. W., Suter, M., Synal, H. A. and Wacker, L.: Gaseous radiocarbon measurements of small samples, Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms, 268(7–8), 790–794,

495   doi:10.1016/j.nimb.2009.10.032, 2010a.

Ruff, M., Fahrni, S., Gäggeler, H. W., Hajdas, I., Suter, M., Synal, H. A., Szidat, S. and Wacker, L.: On-line radiocarbon

measurements of small samples using elemental analyzer and MICADAS gas ion source, Radiocarbon, 52(4), 1645–1656, doi:10.1017/S003382220005637X, 2010b.

Salazar, G., Zhang, Y. L., Agrios, K. and Szidat, S.: Development of a method for fast and automatic radiocarbon measurement

500   of aerosol samples by online coupling of an elemental analyzer with a MICADAS AMS, Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms, 361, 163–167, doi:10.1016/j.nimb.2015.03.051, 2015.

Santos, G. M., Southon, J. R., Griffin, S., Beaupre, S. R. and Druffel, E. R. M.: Ultra small-mass AMS ^{14}C sample preparation and analyses at KCCAMS/UCI Facility, 259, 293–302, Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms, doi:10.1016/j.nimb.2007.01.172, 2007.

505   Shah Walter, S. R., Gagnon, A. R., Roberts, M. L., McNichol, A. P., Gaylord, M. C. L. and Klein, E.: Ultra-Small Graphitization Reactors for Ultra-Microscale [14]C Analysis at the National Ocean Sciences Accelerator Mass Spectrometry (NOSAMS) Facility, Radiocarbon, 57(1), 109–122, doi:10.2458/azu_rc.57.18118, 2015.

Sundqvist, H. S., Kaufman, D. S., McKay, N. P., Balascio, N. L., Briner, J. P., Cwynar, L. C., Sejrup, H. P., Seppä, H., Subetto, D. A., Andrews, J. T., Axford, Y., Bakke, J., Birks, H. J. B., Brooks, S. J., De Vernal, A., Jennings, A. E., Ljungqvist,

510   F. C., Rühland, K. M., Saenger, C., Smol, J. P. and Viau, A. E.: Arctic Holocene proxy climate database - New approaches to assessing geochronological accuracy and encoding climate variables, Clim. Past, 10(4), 1605–1631, doi:10.5194/cp-10-1605-2014, 2014.

Synal, H. A., Stocker, M. and Suter, M.: MICADAS: A new compact radiocarbon AMS system, Nucl. Instruments Methods Phys. Res. Sect. B Beam Interact. with Mater. Atoms, 259(1), 7–13, doi:10.1016/j.nimb.2007.01.138, 2007.

515   Szidat, S., Salazar, G. A., Vogel, E., Battaglia, M., Wacker, L., Synal, H.-A. and Türler, A.: [14]C analysis and sample preparation at the new Bern Laboratory for the Analysis of Radiocarbon with AMS (LARA), Radiocarbon, 56(2), 561–566, doi:10.2458/56.17457, 2014.

Tornqvist, T. E., De Jong, A. F. M., Oosterbaan, W. A. and Van Der Borg, K.: Accurate dating of organic deposits by AMS ^{14}C measurement of macrofossils, Radiocarbon, 34(3), 566–577, doi:10.1017/S0033822200063840, 1992.

520   Trachsel, M. and Telford, R. J.: All age–depth models are wrong, but are getting better, Holocene, 27(6), 860–869, doi:10.1177/0959683616675939, 2017.

Wacker, L., Bonani, G., Friedrich, M., Hajdas, I., Kromer, B., Němec, M., Ruff, M., Suter, M., Synal, H. A. and Vockenhuber, C.: MICADAS: Routine and high-precision radiocarbon dating, Radiocarbon, 52(2), 252–262, doi:10.1017/S0033822200045288, 2010a.

525   Wacker, L., Christl, M. and Synal, H. A.: Bats: A new tool for AMS data reduction, Nucl. Instruments Methods Phys. Res.

Sect. B Beam Interact. with Mater. Atoms, 268(7–8), 976–979, doi:10.1016/j.nimb.2009.10.078, 2010b.

Wacker, L., Fahrni, S. M., Hajdas, I., Molnar, M., Synal, H. A., Szidat, S. and Zhang, Y. L.: A versatile gas interface for routine radiocarbon analysis with a gas ion source, Nuclear Instr. Meth. Phys. Res. Sect. B, 294, 315–319, doi:10.1016/j.nimb.2012.02.009, 2013.

530 Żarczyński, M., Tylmann, W. and Goslar, T.: Multiple varve chronologies for the last 2000 years from the sediments of Lake Żabińskie (northeastern Poland) – Comparison of strategies for varve counting and uncertainty estimations, Quat. Geochronol., 47, 107–119, doi:10.1016/j.quageo.2018.06.001, 2018.

535 **Figure and Table Captions**

Figure 1: Analytical precision of AMS radiocarbon ages (without calibration) versus the mass of carbon in the sample. Note the logarithmic scale on the x-axis. The dashed line represents the theoretical expected relationship between sample mass and age uncertainty based on the Poisson distribution of counting statistics. The solid line represents the best-fit power model for our dataset.

Figure 2: A) Comparison of age-depth model outputs from OxCal (Bronk Ramsey, 2008, 2009; Bronk Ramsey and Lee, 2013; Reimer
540 **et al., 2013). From left to right: OxCal V-sequence using all ^{14}C ages as well as varve counts as inputs; OxCal P-sequence using all ^{14}C ages as inputs; OxCal P-sequence using only gas-source ^{14}C ages; OxCal P-sequence using only graphitized ^{14}C ages. The median age of the V-sequence is considered the best-age estimate and is repeated in all four panels as a red line. Gray lines represent the upper and lower limits of the 95% confidence interval of each model. Black lines represent the median ages of the P-sequences. B) Radiocarbon calibrated age probability density functions for each measured age, grouped by composite depth. The best-age**
545 **estimates from the OxCal V-sequence are plotted as red lines for comparison. The = symbol adjacent to some probability density functions indicates that these ages (within a single depth) came from the same specimen and have the same true age.**

Figure 3: Offsets between median calibrated ^{14}C ages and the best age estimate from the OxCal V-sequence. Data are grouped by material type. Higher values indicate that the sample age is older than the best-age estimate.

Figure 4: Results of age-model simulations to test the effects of sampling density and sample mass on age-model results. Each panel
550 **shows the output of an OxCal P-sequence using simulated ^{14}C ages as inputs compared with the best-age estimate from the V-sequence (shown in red). Simulated ^{14}C ages are based on the decalibrated best-age estimate of a given depth and the expected uncertainty associated with the mass C in the simulated ^{14}C age, which defines not only the age uncertainty, but also a random error term added to each simulated age. Plots show one ensemble member out of 30 simulations. An animation of all 30 simulations can be found in Supplementary File 3.**

555 **Figure 5: A) Boxplots showing the distribution of the mean 95% confidence interval widths produced by simulated age-depth models. Results are grouped by dating density along the x-axis, and by sample mass (smaller mass = greater uncertainty) using different**

22

colors. Each boxplot represents the distribution of results produced for 30 unique sets of simulated ^{14}C samples. Data points that are greater (less) than the 75th (25th) percentile plus (minus) 1.5 times the interquartile range are plotted as single points beyond the extent of the whiskers. B) Same as A, but showing the mean absolute deviation from the best-age estimate (median output of OxCal V-sequence).

560

Table 1: Results of the 48 ^{14}C analyses obtained for this study. Uncertainties of ^{14}C ages refer to 68% probabilities (1σ) whereas ranges of calibrated and modelled ages represent 95% probabilities.

Table 2: Table summarizing the effect of dating density (number of ages) and analytical precision (sample mass) on the accuracy, precision and reliability of OxCal P-sequence models generated from simulated ^{14}C ages. Each of the nine scenarios was simulated 565 30 times; presented values are the mean of the 30-member ensemble. Precision is assessed by the mean width of the age-depth model 95% confidence interval. Accuracy is measured by the mean absolute deviation from the OxCal V-sequence best-age estimate, which is the reference from which ^{14}C ages were simulated. Chron Score is a metric designed to assessing the reliability of age-depth models where higher numbers represent greater reliability (Sundqvist et al., 2014).

Supplementary Files

570 Supplementary File 1: Core images and location of ^{14}C ages.
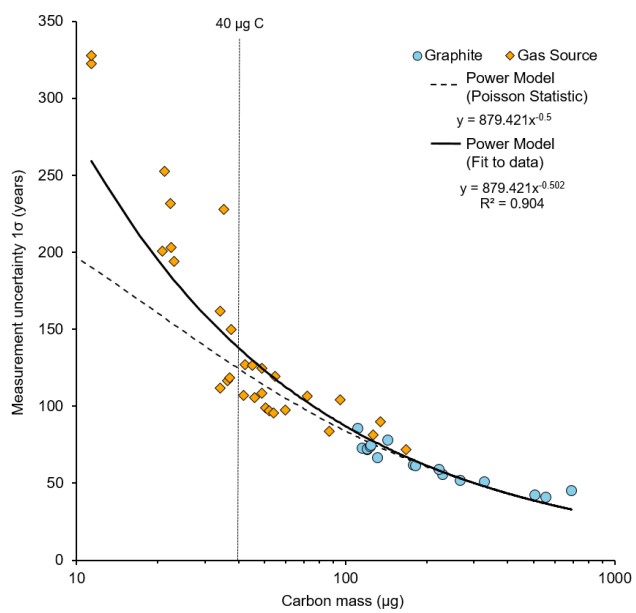Supplementary File 2: Microscope images of macrofossils used for ^{14}C dating.
Supplementary File 3: Animation of OxCal P-sequence age-depth models for all 30 iterations of simulated sampling scenarios (animated version of Figure 4).

575

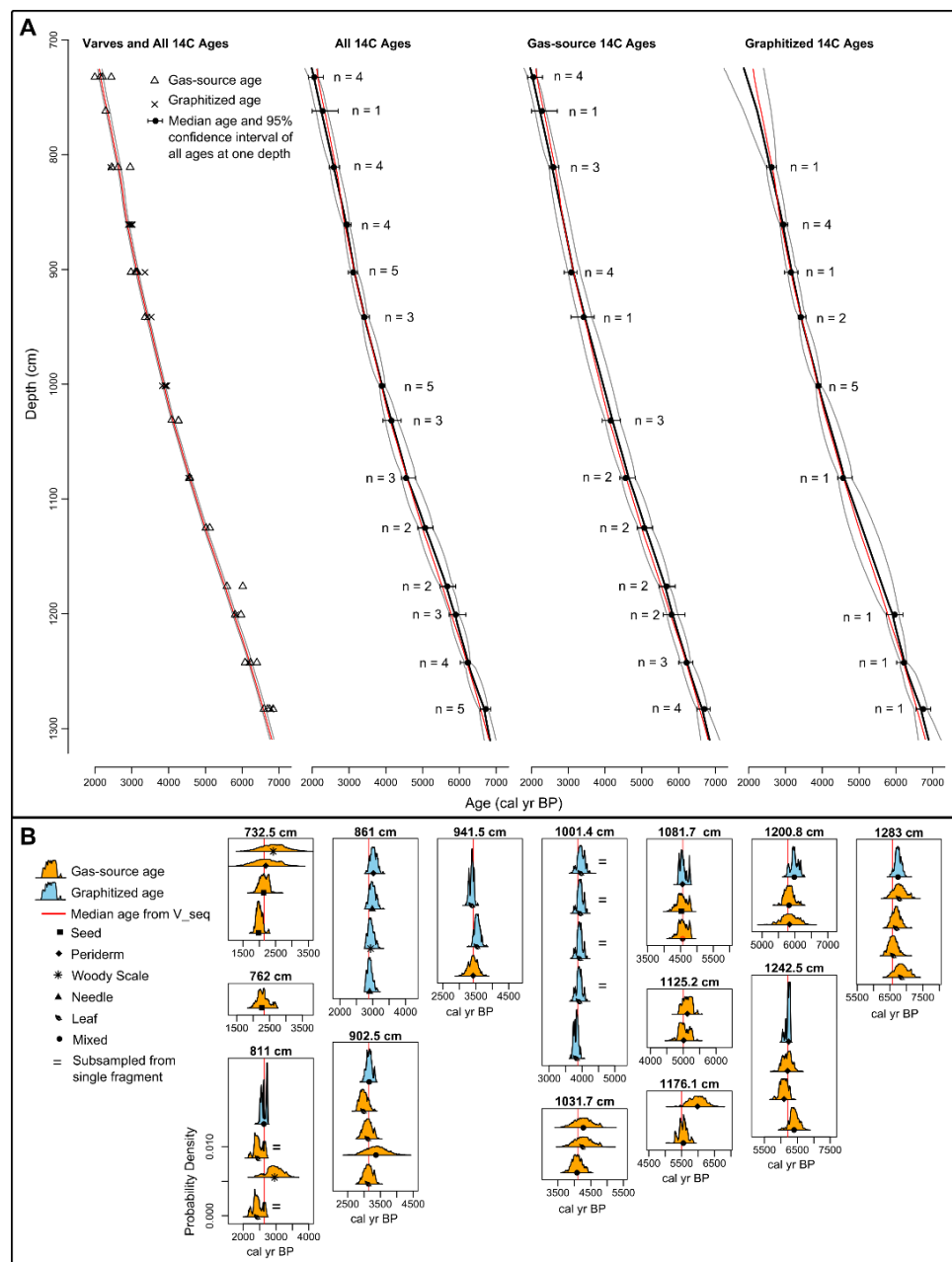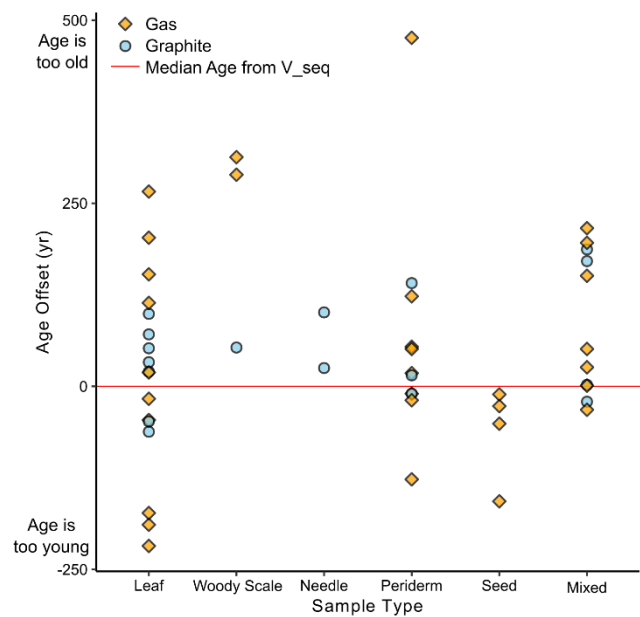**Figure and Table Captions**

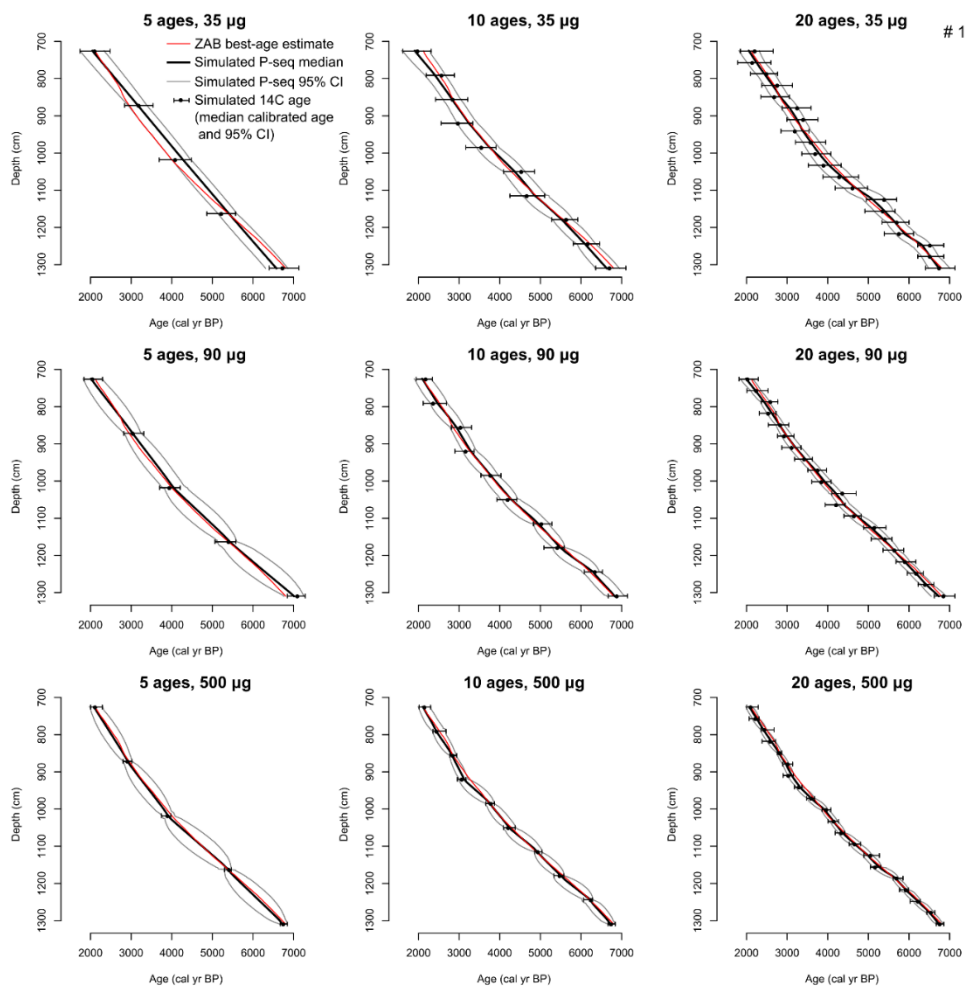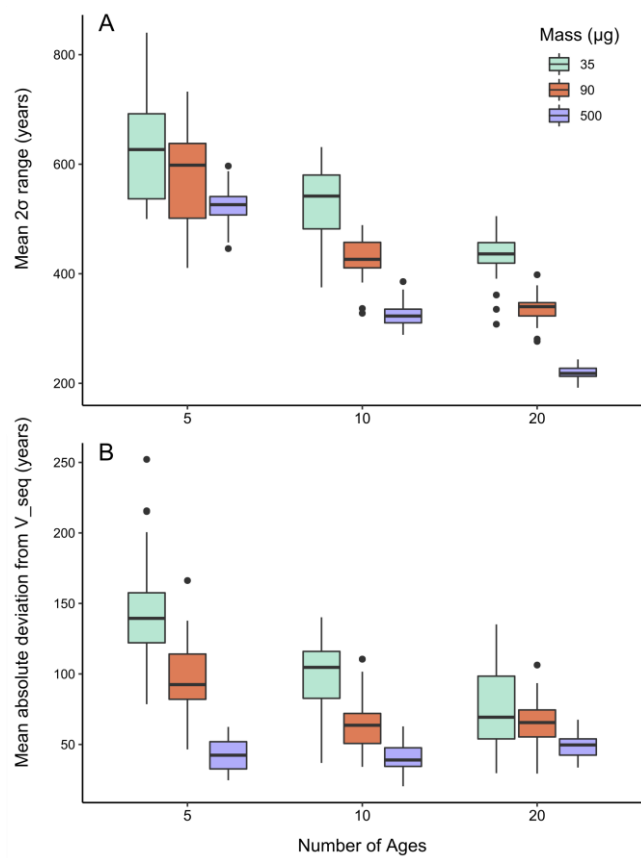Figure 1

580

Figure 2

Figure 3

585

Figure 4

590

Figure 5

Table 1

Core ID	Top Core Depth (cm)	Bottom Core Depth (cm)	Centered Composite Depth (cm)	Material	Carbon Mass (μg)	Gas/ Graphite	Lab ID	^{14}C age (BP)	Calibrated Age (Cal yr BP)[1]	Modelled Age from OxCal V-sequence (Cal yr BP)[2]
ZAB-12-4-3-2	75	77	732.5	*Pinus sylvestris* seed fragments (seed wing, and fragments of seed)	168	Gas	BE-9791.1.1	2028 ± 72	1823-2154 (94.2%); 2273-2293 (1.2%)	2106-2218
ZAB-12-4-3-2	75	77	732.5	Terrestrial seed fragment	34	Gas	BE-9793.1.1	2149 ± 112	1867-2361 (95.4%)	2106-2218
ZAB-12-4-3-2	75	77	732.5	Periderm (coniferous)	11	Gas	BE-9792.1.1	2190 ± 322	1416-1463 (0.7%); 1480-1499 (0.3%); 1516-2968 (94.4%)	2106-2218
ZAB-12-4-3-2	75	77	732.5	Woody scale	11	Gas	BE-9794.1.1	2386 ± 328	1636-1648 (0.2%); 1697-3253 (94.8%); 3297-3325 (0.4%)	2106-2218
ZAB-12-3-4-2	36	37	762	*Alnus* seed fragments	36	Gas	BE-9503.1.1	2273 ± 117	1998-2543 (86.3%); 2559-2619 (3.6%); 2631-2702 (5.5%)	2297-2402
ZAB-12-3-4-2	85	86	811	Dicotyledonous leaf fragment[3]	87	Gas	BE-9502.1.2	2358 ± 84	2159-2254 (10.1%); 2300-2715 (85.3%)	2611-2703
ZAB-12-3-4-2	85	86	811	Dicotyledonous leaf fragment[3]	127	Gas	BE-9502.1.1	2379 ± 82	2183-2235 (3.7%); 2305-2722 (91.7%)	2611-2703
ZAB-12-3-4-2	85	86	811	Deciduous tree/shrub woody scales	21	Gas	BE-9501.1.1	2809 ± 201	2437-3447 (95.4%)	2611-2703

ZAB-12-3-4-2	85	86	811	Dicotyledonous leaf fragments, woody scales	553	Graphite	BE-9500.1.1	2544 ± 41	2490-2644 (54.6%); 2653-2669 (1.6%); 2676-2754 (39.2%)	2611-2703
ZAB-12-4-4-2	20	21	861	*Pinus sylvestris* needle	131	Graphite	BE-9497.1.1	2799 ± 67	2760-3076 (95.4%)	2850-2929
ZAB-12-4-4-2	20	21	861	Woody scale	120	Graphite	BE-9498.1.1	2820 ± 72	2774-3082 (90.3%); 3092-3143 (5.1%)	2850-2929
ZAB-12-4-4-2	20	21	861	*Pinus sylvestris* needle	115	Graphite	BE-9496.1.1	2857 ± 73	2790-3174 (95.4%)	2850-2929
ZAB-12-4-4-2	20	21	861	Periderm (deciduous)	120	Graphite	BE-9499.1.1	2885 ± 72	2807-2813 (0.3%); 2844-3229 (95.1%)	2850-2929
ZAB-12-4-4-2	61.5	62.5	902.5	Periderm, Dicotyledonous leaf fragments, woody scales	21	Gas	BE-9495.1.1	3158 ± 252	2764-3984 (95.4%)	3113-3187
ZAB-12-4-4-2	61.5	62.5	902.5	Dicotyledonous leaf fragment	54	Gas	BE-9494.1.1	2845 ± 96	2761-3215 (95.4%)	3113-3187
ZAB-12-4-4-2	61.5	62.5	902.5	Dicotyledonous leaf fragment	50	Gas	BE-9494.1.2	2968 ± 99	2876-3374 (95.4%)	3113-3187
ZAB-12-4-4-2	61.5	62.5	902.5	Dicotyledonous leaf fragment	52	Gas	BE-9494.1.3	2944 ± 97	2866-3358 (95.4%)	3113-3187
ZAB-12-4-4-2	61.5	62.5	902.5	Dicotyledonous leaf fragments, periderm fragments	230	Graphite	BE-9493.1.1	2980 ± 56	2979-3271 (86.8%); 3286-3340 (8.6%)	3113-3187
ZAB-12-4-4-2	100.5	101.5	941.5	Periderm	37	Gas	BE-9491.1.1	3197 ± 119	3078-3095 (0.6%); 3107-3131 (0.8%); 3137-3700 (93.9%)	3391-3462
ZAB-12-4-4-2	100.5	101.5	941.5	Dicotyledonous leaf fragments	123	Graphite	BE-9490.1.2	3296 ± 74	3375-3696 (95.4%)	3391-3462
ZAB-12-4-4-2	100.5	101.5	941.5	Dicotyledonous leaf fragments	328	Graphite	BE-9490.1.1	3145 ± 51	3226-3466 (95.4%)	3391-3462
ZAB-12-3-5-2	44	45	1001.4	Dicotyledonous leaf fragments	691	Graphite	BE-9489.1.1	3542 ± 45	3697-3930 (92.5%); 3945-3965 (2.9%)	3845-3915
ZAB-12-3-5-2	44	45	1001.4	Dicotyledonous leaf fragment3	179	Graphite	BE-9489.1.2	3593 ± 62	3717-4084 (95.4%)	3845-3915

ZAB-12-3-5-2	44	45	1001.4	Dicotyledonous leaf fragment3	222	Graphite	BE-9489.1.4	3603 ± 59	3724-3796 (8.7%); 3817-4086 (86.7%)	3845-3915
ZAB-12-3-5-2	44	45	1001.4	Dicotyledonous leaf fragment3	182	Graphite	BE-9489.1.3	3616 ± 62	3725-3795 (6.1%); 3819-4092 (88.2%); 4128-4141 (0.9%)	3845-3915
ZAB-12-3-5-2	44	45	1001.4	Dicotyledonous leaf fragment3	124	Graphite	BE-9489.1.5	3631 ± 75	3721-3802 (6.8%); 3810-4153 (88.6%)	3845-3915
ZAB-12-4-5-1	24	26	1031.2	Betula seed fragments, terrestrial woody material, woody scale, periderm fragments	42	Gas	BE-9795.1.1	3724 ± 107	3829-4417 (95.4%)	4084-4155
ZAB-12-4-5-1	25	26	1031.7	Leaf fragments	23	Gas	BE-9487.1.1	3856 ± 194	3731-3745 (0.4%); 3770-3790 (0.6%); 3825-4832 (94.5%)	4084-4155
ZAB-12-4-5-1	25	26	1031.7	Wood fragment, Periderm fragments	22	Gas	BE-9488.1.1	3856 ± 203	3725-3753 (0.8%); 3760-3795 (1.0%); 3820-4836 (93.6%)	4084-4155
ZAB-12-4-5-1	75	76	1081.7	Periderm, woody scales	60	Gas	BE-9485.1.1	4062 ± 97	4296-4332 (2.2%); 4348-4837 (93.2%)	4540-4616
ZAB-12-4-5-1	75	76	1081.7	Betula alba seed	46	Gas	BE-9486.1.1	4042 ± 105	4249-4275 (1.6%); 4282-4832 (93.8%)	4540-4616
ZAB-12-4-5-1	75	76	1081.7	Periderm fragments	266	Graphite	BE-9484.1.1	4065 ± 52	4421-4660 (75.0%); 4666-4709 (6.7%); 4755-4813 (13.7%)	4540-4616

ZAB-12-4-5-1	118.5	119.5	1125.2	Periderm fragments	49	Gas	BE-9483.1.2	4387 ± 108	4655-4668 (0.4%); 4757-4706 (2.0%); 4810-5318 (93.0%)	4960-5042
ZAB-12-4-5-1	118.5	119.5	1125.2	Periderm fragments	135	Gas	BE-9483.1.1	4475 ± 90	4860-5320 (94.8%); 5424-5434 (0.6%)	4960-5042
ZAB-12-5-6-1	54	55.5	1176.1	Woody seed fragments, leaf fragments, woody scales	95	Gas	BE-9481.1.1	4850 ± 104	5321-5422 (11.5%); 5436-5761 (79.1%); 5820-5887 (4.8%)	5500-5591
ZAB-12-5-6-1	54	55.5	1176.1	Periderm fragments	22	Gas	BE-9482.1.1	5246 ± 232	5485-5513 (0.5%); 5580-6536 (94.9%)	5500-5591
ZAB-12-5-6-1	79	80	1200.8	Periderm fragments	35	Gas	BE-9480.1.1	5081 ± 228	5320-5425 (3.4%); 5433-6315 (92.0%)	5745-5832
ZAB-12-5-6-1	79	80	1200.8	Periderm, woody scale	42	Gas	BE-9479.1.1	5063 ± 127	5586-6121 (93.9%); 6148-6178 (1.5%)	5745-5832
ZAB-12-5-6-1	79	80	1200.8	Periderm fragments and woody scales	111	Graphite	BE-9478.1.1	5197 ± 86	5745-6190 (95.4%)	5745-5832
ZAB-12-5-6-2	5	6	1242.5	Periderm and woody scale	49	Gas	BE-9476.1.1	5601 ± 125	6032-6038 (0.1%); 6120-6148 (0.9%); 6177-6694 (93.9%); 6702-6718 (0.5%)	6175-6267
ZAB-12-5-6-2	5	6	1242.5	Periderm	72	Gas	BE-9475.1.1	5294 ± 107	5768-5806 (1.6%); 5890-6300 (93.8%)	6175-6267
ZAB-12-5-6-2	5	6	1242.5	Periderm fragments	45	Gas	BE-9477.1.1	5410 ± 127	5920-6439 (95.4%)	6175-6267

ZAB-12-5-6-2	5	6	1242.5	Pinus periderm fragments	504	Graphite	BE-9474.1.1	5402 ± 43	6020-6080 (7.6%); 6109-6155 (10.2%); 6173-6294 (77.6%)	6175-6267
ZAB-12-5-6-2	45.5	46.5	1283	Dicotyledonous leaf fragments	34	Gas	BE-9473.1.3	5988 ± 162	6479-7250 (95.4%)	6531-6643
ZAB-12-5-6-2	45.5	46.5	1283	Dicotyledonous leaf fragments	55	Gas	BE-9473.1.2	5787 ± 119	6317-6860 (94.9%); 6871-6880 (0.5%)	6531-6643
ZAB-12-5-6-2	45.5	46.5	1283	Dicotyledonous leaf fragments	74	Gas	BE-9473.1.1	5868 ± 107	6415-6418 (0.1%); 6435-6949 (95.3%)	6531-6643
ZAB-12-5-6-2	45.5	46.5	1283	Dicotyledonous leaf fragments	38	Gas	BE-9473.1.4	5936 ± 150	6436-7165 (95.4%)	6531-6643
ZAB-12-5-6-2	45.5	46.5	1283	Dicotyledonous leaf fragments, periderm fragment	143	Graphite	BE-9472.1.1	5916 ± 78	6547-6946 (95.4%)	6531-6643

600 [1] Ages calibrated using OxCal 4.3 with the IntCal13 calibration curve (Bronk Ramsey, 2009; Reimer et al., 2013). The probability of each listed age range is noted in parenthesis.
[2] Range represents 95% confidence interval.
[3] These samples were subsampled from a single fragment prior to analysis, thus samples within the same depth with this symbol have the same true age.
605

Table 2

Sample Mass (µg)	Expected Precision (yr)	Number of ages in model		
		5 ages (1.07 per kyr)	10 ages (2.14 per kyr)	20 ages (4.27 per kyr)
		Mean 95% CI width (yr)		
35	± 148	633	527	433
90	± 92	577	430	335
500	± 39	524	325	219
		Mean absolute deviation from OxCal V-sequence (yr)		
35	± 148	144	99	78
90	± 92	98	64	65
500	± 39	42	40	49
		Chron Score		
35	± 148	2.46	3.14	3.48
90	± 92	2.87	3.64	4.09
500	± 39	3.92	4.74	5.18