

© 2017 by the Arizona Board of Regents on behalf of the University of Arizona. This is an Open Access article, distributed under the terms of the Creative Commons Attribution-NonCommercial-NoDerivatives licence (<http://creativecommons.org/licenses/by-nc-nd/4.0/>), which permits non-commercial re-use, distribution, and reproduction in any medium, provided the original work is unaltered and is properly cited. The written permission of Cambridge University Press must be obtained for commercial re-use or in order to create a derivative work.

## REFINING THE HALLSTATT PLATEAU: SHORT-TERM $^{14}\text{C}$ VARIABILITY AND SMALL SCALE OFFSETS IN 50 CONSECUTIVE SINGLE TREE-RINGS FROM SOUTHWEST SCOTLAND DENDRO-DATED TO 510–460 BC

Piotr Jacobsson<sup>1\*</sup> • William Derek Hamilton<sup>1</sup> • Gordon Cook<sup>1</sup> • Anne Crone<sup>2</sup> • Elaine Dunbar<sup>1</sup> • Helen Kinch<sup>1</sup> • Philip Naysmith<sup>1</sup> • Brian Tripney<sup>1</sup> • Sheng Xu<sup>1</sup>

<sup>1</sup>Scottish Universities Environmental Research Centre (SUERC) Radiocarbon Dating Laboratory, Rankine Ave., East Kilbride, G75 0QF, Scotland, United Kingdom.

<sup>2</sup>AOC Archaeology Group, Edgefield Road Industrial Estate, Loanhead, Midlothian, EH20 9SY, Scotland, United Kingdom.

**ABSTRACT.** Radiocarbon ( $^{14}\text{C}$ ) wiggle-match dating is a technique with a substantial potential to improve the precision of dating timbers in situations where dendrochronology is not tenable. However, one of the key reasons why obtaining a dendrochronological determination might be difficult is the short-lived nature of timbers on a range of archaeological sites, something that also affects the efficiency of the wiggle-match dating technique. Combined with the potential for high expense that the technique presents, it is paramount that wiggle-match dating research design has a good empirical basis. To this end we dated 50 consecutive, individual rings from a timber that grew during the Hallstatt radiocarbon calibration plateau (ca. 750–400 cal BC) in southwest Scotland. The results indicate that (1) the precision and accuracy of wiggle-match dates carried out on short-lived sequences during the Hallstatt plateau may suffer due to insufficient resolution of the calibration data, (2) sampling time-frames roughly equivalent to the underpinning calibration data are recommended (for the period in question this means decadal blocks), and (3) short-lived sequences are at risk of losing accuracy if the actual past trend of radiocarbon diverges from the mean of the radiocarbon calibration curve.

**KEYWORDS:** calibration, secular  $^{14}\text{C}$  variations, single tree rings, wiggle-match dating.

### INTRODUCTION

The results presented in this paper stem from a specific problem of archaeological research design: how to obtain radiocarbon ( $^{14}\text{C}$ )-based dates for wetland structures, precise to a century or better, during a calibration plateau. This problem is of significant importance to wetland archaeology in Scotland, where the tradition of building dwellings on artificial islands (often referred to as crannogs) begins in earnest sometime during the Hallstatt plateau, a period between 750 and 400 cal BC, when most  $^{14}\text{C}$  determinations will calibrate to a period of several centuries (Crone 2012). In practice, this means that the earliest wetland sites cannot be distinguished from those that followed several hundred years later, or even whether this phenomenon began at the onset of the Iron Age (750 BC), or several hundred years into it. Much of the challenge comes from the limited number of dendrochronological samples. While some oak timbers from wetland sites in southwest Scotland can be dated through dendrochronology, thereby providing the required chronological precision (Cavers et al. 2011; Cavers and Crone forthcoming), this is unusual, as the majority of the structural components at these sites were made of alder (*Alnus glutinosa*), a species that cannot furnish absolute chronological determinations because it is generally too short-lived for master chronology construction (Crone 1988, 2014). Nevertheless, it is in principle possible to bypass this issue through the  $^{14}\text{C}$  wiggle-match dating technique. Within this technique, a series of measurements is taken throughout the lifespan of the timber, either on single tree rings or groups of rings. Because the cellulose within each ring is only associated with the year of ring formation, these measurements allow the reconstruction of a small section of the past  $^{14}\text{C}$  trend that can then be fitted to a

\*Corresponding author. Email: [pt.jacobsson@gmail.com](mailto:pt.jacobsson@gmail.com).

$^{14}\text{C}$  calibration curve (Bronk Ramsey et al. 2001). A more precise determination becomes possible because the position of strongest matching between the series of  $^{14}\text{C}$  values from the unknown wood and the  $^{14}\text{C}$  calibration curve will be limited. For the results of wiggle-matching to be consistent with those produced elsewhere, internationally ratified calibration curves ought to be used.

The concept of  $^{14}\text{C}$  wiggle-match dating was developed in the early decades of  $^{14}\text{C}$  dating (Fergusson et al. 1966), yet its application is not as frequent as might be expected given its potential. This can be traced to the costs of wiggle-match dating. Meaningful application, in most cases, will require dating several timbers to obtain a good appreciation for the chronology of a site as a whole. Given that each wiggle-match date is based on several  $^{14}\text{C}$  determinations, there is the continuing risk that the costs of such a project can exceed any financial figure that the various funding bodies can provide on a routine basis. Furthermore, the performance of the wiggle-match dating technique improves with increasing length of the dated sequence (Goslar and Madry 1998; Galimberti et al. 2004) and, as stated above, one of the problems facing Scottish wetland archaeology is the paucity of long-lived timbers. Therefore, before dating strategies for actual sites can be developed, it is of paramount importance to explore the performance of the technique and its intricacies, involving issues such as the selection of the appropriate incremental ages of constituent samples (single rings or multi-ring blocks), the evaluation of the precision that can be attained and the identification of unexpected limitations.

A number of previous studies present  $^{14}\text{C}$  determinations of known-age single tree rings. The best-known of these are the measurements reported by Stuiver et al. (1998), but a range of researchers have carried out further measurements (e.g. Miyake et al. 2012; Gütler et al. 2013). However, very few of these studies cover the Hallstatt plateau and hence might not be of direct use for research design in this period. The one exception is a collection of single rings measured by Suzuki et al. (2010), but this could not be applied in a Scottish context, as the  $^{14}\text{C}$  concentrations in these tree rings might have been affected by the Southern Hemisphere air masses Gütler (Suzuki et al. 2010). One recent development of potential future importance is the discovery of 1-yr excursions towards much higher atmospheric  $^{14}\text{C}$  concentrations in Japanese cedar from AD 775 (Miyake et al. 2012) and reproduced in a number of trees from a range of locations across the globe (Jull et al. 2014). In principle, these excursions enable wiggle-match dating to a single year (Wacker et al. 2014), but this is only possible if they are identified and this in turn requires high resolution and high precision data, which, for the time being, are not available for the Hallstatt plateau in the Northern Hemisphere.

Another issue is the lack of empirical studies concerned with the practice of wiggle-match dating of short sequences. The two papers that addressed the performance of the technique (Goslar and Madry 1998; Galimberti et al. 2004) were both based on simulated data and focused on long-lived material, for which they recommended increasing the length of the dated sequence as the means of improving precision, while maintaining control over the costs. While this advice cannot be disputed, it is not applicable in archaeological contexts where the timbers essential to understanding a site lived for less than 50 yr. Furthermore, because both Galimberti et al. (2004) and Goslar and Madry (1998) relied on data simulated from the calibration curve, their studies had no potential for identifying any of the issues that may arise in empirical data. The importance of this empirical evaluation becomes clear once we consider the very recent study of the performance of short-span wiggle-matches in the context of medieval buildings in England (Bayliss et al. 2017). Within this study three out of the five dated sequences failed to reproduce the known dendrochronological date for the felling of the timbers. It was on account of these limitations in both empirical and simulation studies that we needed to collect new data to guide future archaeological research design.

In the course of the current study, we dated 50 consecutive single rings from an archaeological timber, whose outermost surviving ring was dated to 450 BC (Crone forthcoming, and

see below). These data were meant to provide an empirical basis for the development and verification of different approaches to wiggle-match dating tree-ring sequences of less than 50 yr. In principle, any wiggle-match constructed from these data, which would display both satisfactory precision and envelop the target date, could be considered as a basis for future research design. In other words, this would permit us to evaluate what combinations of rings make for precise and reliable wiggle-matches on the Hallstatt plateau. As so often happens though, the study uncovered a suite of very different issues that made us realize the need for rethinking the relationship between the wiggle-match data and the calibration curve.

## SAMPLE AND METHODS

To undertake the study, we measured the  $^{14}\text{C}$  activity in a series of tree rings that were dated through dendrochronology, so that the efficiency of different combinations of rings in developing a wiggle-match that agreed with the date for the timber felling could be evaluated. The timber in question had to date to the Hallstatt plateau so that the performance of the wiggle-match dating technique could be assessed under these more challenging conditions. Furthermore, it was considered ideal for the timber in question to have a clear ring pattern and derive from a site local to Scotland, so as to identify any taphonomic issues that could affect later studies.

One timber that fitted these criteria was the oak T-947, discovered at the remains of the lake dwelling at Cults Loch 3 in Galloway (southwest Scotland). T-947 was an oak stake that formed part of a possible fence line on the lake dwelling; it had 109 surviving growth rings, the outermost rings having decayed away. It produced good visual and statistical correlations with other oak sequences from the site and was incorporated into a site tree-ring master chronology, CLMNx9, 272 yr in length. CLMNx9 compared well with a number of Northern Irish chronologies, producing a t-value of 5.79 against the Belfast Long Chronology, BLC7000, over the chronological span, 709–438 BC. Consequently, because T-947 was a component of CLMNx9, a calendar date of 450 BC could be ascribed to its outermost surviving ring.

Once delivered to Scottish Universities Environmental Research Centre (SUERC), the timber was cut into transverse slices to ease sub-sampling for wiggle-match dating. The surfaces of the transverse slices were then prepared using a razor blade to enhance the visibility of ring-widths and individual rings were cut using razor blades or a scalpel. Because the outer rings of the timber were too decayed for effective sampling, ring 1 in the  $^{14}\text{C}$  sequence corresponds to ring 11 in the dendrochronological sequence, which was laid down in 460 BC. The sequence is counted towards the pith, so ring 50 was laid down in 509 BC.

Cellulose was extracted from the cut tree rings using a protocol modified from Hoper et al. (1998) and combusted at 850°C in evacuated sealed quartz tubes using copper oxide as the oxygen source and silver foil to remove halides and other contaminants (Vandeputte et al. 1996). The resulting carbon dioxide was cryogenically purified and a 3 mL subsample was reduced to elemental carbon (graphite) using the two-step iron and zinc process proposed by Slota et al. (1987). The graphite was then pressed into aluminium target holders and measured on the SUERC 5MV tandem AMS. During routine analysis, targets are measured in groups of seven unknowns and three QA samples. The QA materials comprise one oxalic acid II (SRM-4990C) standard, one humic acid standard of less than 1 half-life in age (sample T of the Third International Radiocarbon Intercomparison [TIRI]) (Scott et al. 1997) and either one background sample (cellulose extracted from Miocene-age wood provided by the Hohenheim dendrochronology laboratory) or one Glengoyne Distillery barley mash sample (sample A of

TIRI) (Scott et al. 1997). The CO<sub>2</sub> from each of the samples was also subsampled for an off-line  $\delta^{13}\text{C}$  analysis to provide a fractionation correction factor for the <sup>14</sup>C age determinations. Where indicated in the text and tables, routine precision determinations were combined to yield high-precision equivalents using the method of Ward and Wilson (1978). High-precision determinations with a standard error of less than 20 <sup>14</sup>C yr were obtained by increasing the measurement time and also increasing the number of QA samples (to two oxalic acid II and two TIRI humic acid samples, or two oxalic acid II, a geological age wood and a TIRI humic acid sample) per six unknown samples, to confirm the precision.

All wiggle matches were conducted within a Bayesian framework as defined by Bronk Ramsey et al. (2001) and implemented in OxCal 4.2 (Bronk Ramsey 2009) using the IntCal13 calibration curve (Reimer et al. 2013). A Bayesian wiggle-match calculates the posterior probability of the calibrated date range for the final determination in the series, given the relationships of all the measurements, thus providing the entire probability density for the final result. The reliability of the match is evaluated using an agreement index, which estimates how well the proposed model fits the curve relative to independent calibrations. In essence, if the model requires the distributions of individual measurements to be altered to a large degree, the agreement index will decline. The actual value of the index below which the reliability of the match ought to be questioned depends on the number of measurements included. OxCal also calculates the  $\chi^2$  estimate between the constituent measurements of the wiggle-match and the calibration curve. This estimate provides an alternative source of quality assurance.

## RESULTS

The measurements of the T-947 single rings were conducted in two batches in October 2013 and February 2014. Forty-seven measurements from the first batch, measured to high precision (approx. 2‰ <sup>14</sup>C, or 16 <sup>14</sup>C yr) were successful (Table 1), with results ranging between 2508 ± 14 and 2356 ± 15 BP. Because of the breadth of this range and from expected calibration curve values, three of the rings (13, 20, and 24) were remeasured (Table 2). Unlike the measurements from the first batch, these follow-up samples were measured three times to routine precision and their ages were combined to obtain the equivalent of a high-precision determination using the method of Ward and Wilson (1978). The results were in agreement with the outcome of the first batch of measurements (Figure 1), supporting the validity of the single-ring data.

The outcomes of the wiggle-match conducted on the available measurements from the 47 single-rings are far from satisfactory (Figure 2), due to the high  $\chi^2$  test statistic (95.548; 5% critical value = 61.869) and the low agreement index  $A_{\text{combine}} = 1.4\%$  (minimum recommended value,  $A_n = 10.3\%$ ). This is driven by the divergences between the <sup>14</sup>C ages within the measured series and the calibration curve (Figure 3). Over the period of interest (460–510 BC), the calibration data consist of decadal blocks of wood (Reimer et al. 2013) and the mean value of the curve produced from these data fluctuates between 2410 and 2420 BP, with only a slight rise to 2430 BP near the beginning of the series around 510 BC. The single-ring T-947 data, on the other hand, display a substantial degree of variability:

- At the beginning of the series, between 509 and 503 BC, there is a clear increase in the <sup>14</sup>C ages from 2416 to 2466 BP, followed by a sudden decrease to 2403 BP by 496 BC.
- Thereafter, the trend shifts towards greater ages until it peaks at 2500 BP in 483 BC.
- This is followed by a sudden decrease in ages to between 2420 and 2460 BP for the 13 yr until 470 BC.

Table 1 Results of the  $^{14}\text{C}$  age determinations for the first batch of T-947 single rings.

Date BC	SUERC-number	Age ( $^{14}\text{C}$ yr BP)	1- $\sigma$ error	$\delta^{13}\text{C}$ (‰)
460	48457	2424	14	-26.2
461	48458	2376	13	-24.1
462	48459	2389	15	-24.3
463	48460	2396	15	-24.7
464	48461	2413	15	-23.9
465	48462	2425	17	-24.8
466	48467	2457	15	-24.3
467	48468	2435	16	-25.0
468	48469	2356	15	-25.2
469	48470	2381	16	-24.8
470	48471	2445	17	-23.5
471	48472	2434	13	-24.6
472	48477	2475	16	-24.3
473	48478	2432	14	-23.9
474	48479	2440	15	-23.7
475	48480	2420	14	-24.4
476	48481	2442	12	-23.1
477	48482	2421	16	-23.9
478	48487	2450	16	-24.0
479	48488	2444	15	-23.6
480	48489	2433	14	-24.9
481	48490	2424	15	-24.7
482	48491	2441	15	-23.8
483	48492	2508	14	-24.5
484	48497	2402	16	-24.1
486	48499	2471	16	-23.7
487	48500	2449	15	-23.8
489	48502	2396	13	-25.5
490	48507	2466	19	-24.5
491	48508	2462	14	-24.9
492	48509	2426	5	-24.7
493	48510	2424	16	-24.0
494	48511	2437	16	-25.2
495	48512	2431	20	-24.5
496	48517	2403	16	-24.4
497	48518	2443	15	-24.2
498	48519	2416	16	-24.0
499	48520	2463	15	-24.2
500	48521	2422	16	-24.7
502	48527	2459	16	-24.1
503	48528	2466	16	-24.3
504	48529	2437	15	-24.1
505	48530	2428	16	-25.3
506	48531	2424	14	-24.1
507	48532	2416	15	-25.8
508	48537	2415	14	-26.8
509	48538	2416	15	-25.8

Table 2 Results of the  $^{14}\text{C}$  age determinations for the second batch of T-947 single rings. The 5% critical value for a  $\chi^2$  test at 2 degrees of freedom is 5.99.

Date BC	SUERC-nr	Age ( $^{14}\text{C}$ yr BP)	1- $\sigma$ error	$\delta^{13}\text{C}$ (‰)	$\chi^2$	Combined age	Combined error (1 $\sigma$ )
472	50863	2447	29	-24.4	0.1	2446	16
	50864	2440	24	-24.4			
	50865	2453	29	-24.3			
479	50866	2481	29	-23.8	4.2	2444	17
	50867	2452	29	-23.6			
	50868	2398	29	-23.6			
483	50869	2481	29	-25.0	0.2	2489	17
	50873	2488	29	-25.5			
	50874	2497	29	-25.4			

- Next is another decrease in  $^{14}\text{C}$  ages, sustained for 2 yr, reaching the lowest value of 2356 BP, before rebounding up to 2435 BP at 467 BC.
- From that point on, the  $^{14}\text{C}$  ages decline until they reach the lowest value of 2376 BP in 461 BC, before a subsequent renewed age increase to 2435 BP in 467 BC.

In general, the results of the T-947 measurements might be displaying the irregular cyclical trend expected of the Schwabe cycle, also known as the solar cycle, that has a duration of between 9 and 14 yr and which can be expected to be the greatest contributor to short-term variability in cosmogenic radionuclide activities, such as  $^{14}\text{C}$  (Beer et al. 2012). However, given the shortness of our series, the statistical verification of this observation is difficult and so we remain uncertain as to whether we have discovered the Schwabe cycles on the Hallstatt plateau in actuality. Güttler et al. (2013) identified the solar cycle in their data spanning 52 biannual samples from cal AD 1010–1113, although it was clearer in their data. This difference can be explained through our samples coming from the period of a calibration plateau, when the short-term variability driven by the Schwabe cycle would have been occurring at the same time as whatever complex process was driving the formation of the unusual shape of the calibration curve.

### Discussion 1: Sample Spans

The T-947 single-ring data displayed much greater short-term variability than observed in the shape of the calibration curve, IntCal13 (Reimer et al. 2013). This result is not unexpected as the two studies mentioned previously (Suzuki et al. 2010; Güttler et al. 2013), also observed this and had to smooth their data before fitting it back to the calibration curve. This divergence can be traced to the data underpinning the calibration curve: in the period of interest it consists of decadal or bidecadal blocks of wood dated through dendrochronology (Pearson et al. 1986; Stuiver and Becker 1986; Taylor and Southon 2013). These sample spans mean that 65–100% of the information on the short-term variability driven by the Schwabe cycle, with its 9–14-yr periodicity, becomes attenuated (Beer et al. 2012) and so the calibration data contain insufficient short-term variability to allow for good wiggle-matches. This is further supported by the study of Tyers et al. (2009), who showed that on the parts of the calibration curve based upon single-ring data, single-ring wiggle-matches appear to have a good chance of success. Therefore, the need to avoid single rings is specific to the parts of the calibration curve underpinned by

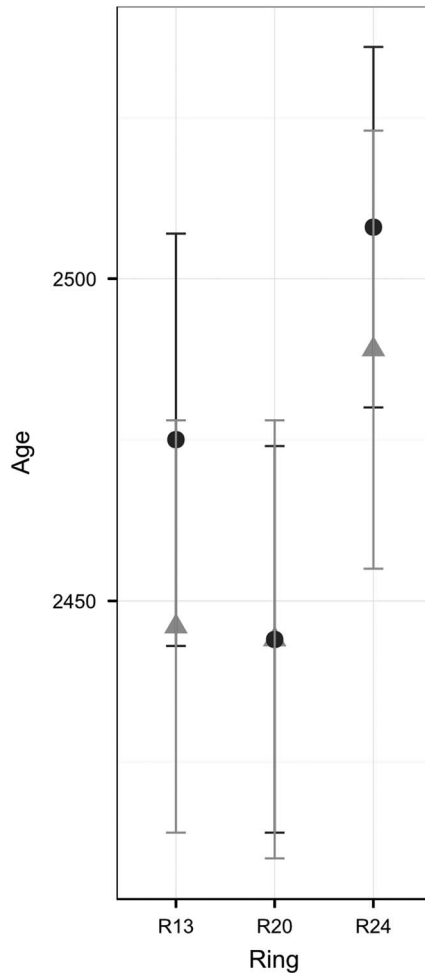


Figure 1 Comparison between the  $^{14}\text{C}$  results of the overlapping rings from the two batches of T-947 single ring determinations. Error bars represent  $2\sigma$ . Circles represent the first and triangles the second batch of measurements.

multi-ring blocks and the discrepancy between the T-947 data and IntCal can be traced to this fact.

However, this leaves open the question of what ring-spans of individual samples are preferable during the Hallstatt plateau. An intuitive answer can be attained from the basic principles of the wiggle-match dating technique. In general, the shorter the sample span, the more short-term variability will be preserved and hence the more limited range of possible positions-of-best-match for the wiggle-match. Having said that, any variability in excess of what is within the calibration curve will, in the best case, become redundant information, and, in other cases, may lead to erroneous results. Hence, the optimal choice of sample spans for wiggle-match dating is one that matches the underpinning calibration data sets and, for this period of interest, that would be decadal blocks of wood.

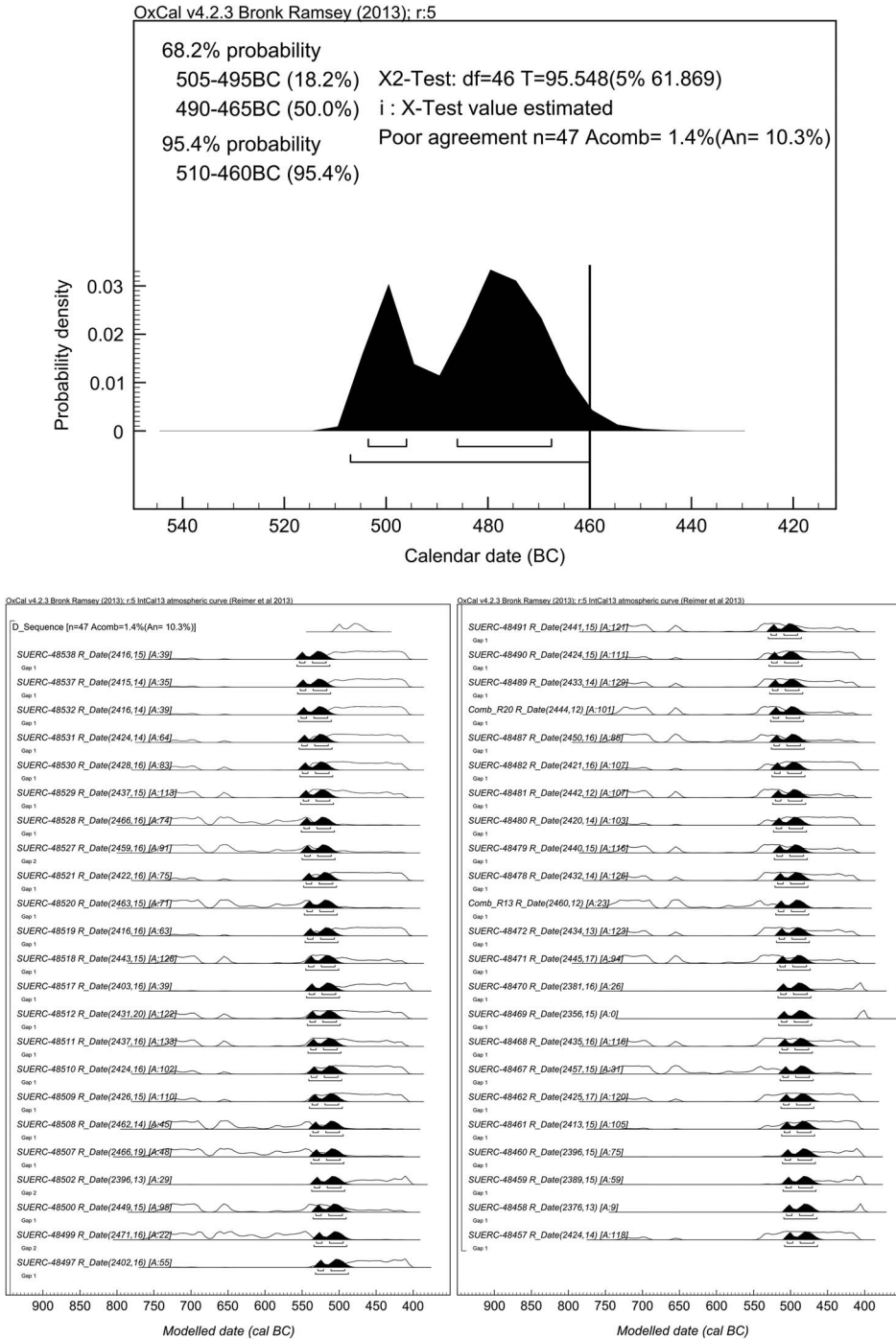


Figure 2 Results of the wiggle-match on the single ring determinations of T-947 data: the summary result (above) and the detailed results (below). The vertical line in the upper panel marks the target date of 460 cal BC.



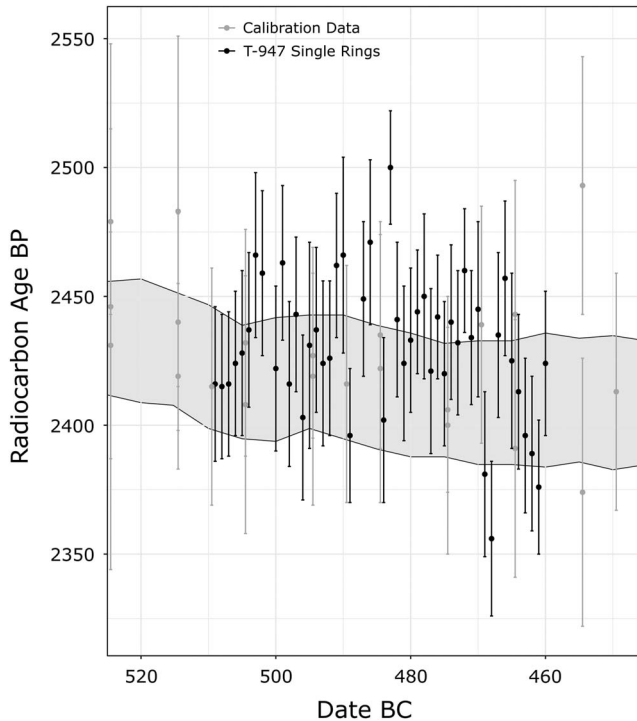


Figure 3 T-947 single-ring data and the relevant calibration data plotted against the 95.4% probability interval of the IntCal13 calibration curve. The old value at 483 cal BC has been re-measured multiple times and hence is considered valid. Error bars are  $2\sigma$ .

Alternative approaches are possible. One intuitive way of addressing the issue is to re-model the calibration data in a way that allows for the short-term variability. In general, the algorithm upon which IntCal13 relies establishes the variability of past  $^{14}\text{C}$  based on the covariance between the surrounding calibration data (Niu et al. 2013). It thus stands to reason that if this was modified to allow for greater short-term variability, a curve suitable for single-ring dating could become viable. In reality, this approach has three major shortcomings: (1) the greater variability within the curve, not backed by higher-resolution data, would lead to a loss of precision that would negate the gains from dating single rings; (2) the  $^{14}\text{C}$  activity of any decadal sample block need not correspond to the  $^{14}\text{C}$  content of its central rings, meaning that decadal blocks cannot be treated as attenuated single-ring data; and (3) the curve thus produced would give results on a different scale from the ratified “cal BC,” leading to potential issues when comparing to other studies. A more viable alternative, put forward by Millard (2008), is to use the single-ring data from the wiggle-match to model a section of the curve, pass the output through a filter simulating the attenuation of decadal sampling and match the results of this procedure to the calibration curve. In principle, this approach is superior as the largest amount of information is retrieved; however, modeling the section of calibration curve would require a large enough number of high-precision measurements to be successful and so this method is not suitable for large projects requiring large numbers of wiggle-match dates on a routine basis. Hence, the approach proposed in the preceding paragraph, based on matching the sample spans to the underlying calibration curve, was considered most suitable for the archaeological problem at hand.

Nevertheless, intuition as to the behavior of the decadal sample blocks is not a sufficient basis for recommendations on dating strategies as this task requires assessment of the data. The issue was approached in two ways. First, the single-ring data were combined into consecutive decadal averages using the method of Ward and Wilson (1978), modified to allow for the differences in ring-widths, as thicker rings would have contributed more to the  $^{14}\text{C}$  value of the decade had it been sampled and measured as a block of 10 rings. The second move was to conduct routine measurements on decadal blocks from T-947, beginning at ring 5 and ending near the pith, with 5-yr overlaps between the samples.

The results of the wiggle-match, based on the decadal averages of single-ring data show that measuring 10-yr sample blocks ought to produce models that agree with the calibration curve. However, the analysis also indicated the possibility of systematic offsets between the actual trend of past  $^{14}\text{C}$  values in the tree-ring archive and the calibration curve. Moreover, it became apparent that when working with very fine-scale models, the choices as to the resolution at which the algorithm updates the model from the prior to the posterior distributions have significant effects on the results.

The resolution issue emerges from the frequency of points at which the models are updated. Using the default settings, OxCal updates Bayesian models at 5-yr intervals, matching the resolution at which the statistical curve model behind IntCal13 for the period in question was constructed (note that this is applied to the calibration curve model and not to the underpinning calibration data which are decadal). In most analyses, changing this frequency of updating points has little effect on the final results, but the high frequency and precision of the T-947 decadal averages mean that changes in this domain lead to significant changes to the wiggle-match results. To be more specific, as the resolution of updating increases, so does the number of possible points of fit, leading to noticeable gains in precision. Hence, the posterior distribution of the wiggle-match on the decadal averages is also dependent on updating frequency (Figure 4):

- Updating every single year (resolution = 1), the calibrated date range lies between 498 and 490 cal BC (95.4%), with a good fit to the calibration curve ( $\chi^2 = 20.762$ ; 5% critical value at 40 degrees of freedom is 54.927;  $A_{\text{combine}} = 310.7\%$ , with the recommended  $A_n = 11.0\%$ ).
- Updating every 5 yr (resolution = 5), the calibrated date range lies between 510 and 475 cal BC (95.4%), with a good fit to the calibration curve ( $\chi^2 = 19.437$ ; 5% critical value at 40 degrees of freedom is 54.927;  $A_{\text{combine}} = 184.1\%$ , with the recommended  $A_n = 11.0\%$ ).
- Updating every 10 yr (resolution = 10), the calibrated date range lies between 505 and 455 cal BC (95.4%), with a good fit to the calibration curve ( $\chi^2 = 24.168$ ; 5% critical value at 40 degrees of freedom is 54.927;  $A_{\text{combine}} = 64.3\%$ , with the recommended  $A_n = 11.0\%$ ).

The changes in the agreement indices and  $\chi^2$  test statistics emerge due to the changing frequency of points of fit between the decadal averages and the calibration curve: as the resolution decreases, so does the number of points of fit, leading to greater divergences between the model and the data. Of the three wiggle-match results only the one constructed at decadal resolution includes the target date. While at first this might seem to confirm the accuracy of the wiggle-match, on reflection this is not the case, for the consecutive decadal averages are spaced at 1-yr intervals and hence matching at single-year resolution is most appropriate. This indicates that the calibrated date range when updating every single year (498–490 cal BC) is at least 30 yr earlier than the target date (460 BC). The origin of this offset is discussed in the subsequent section.

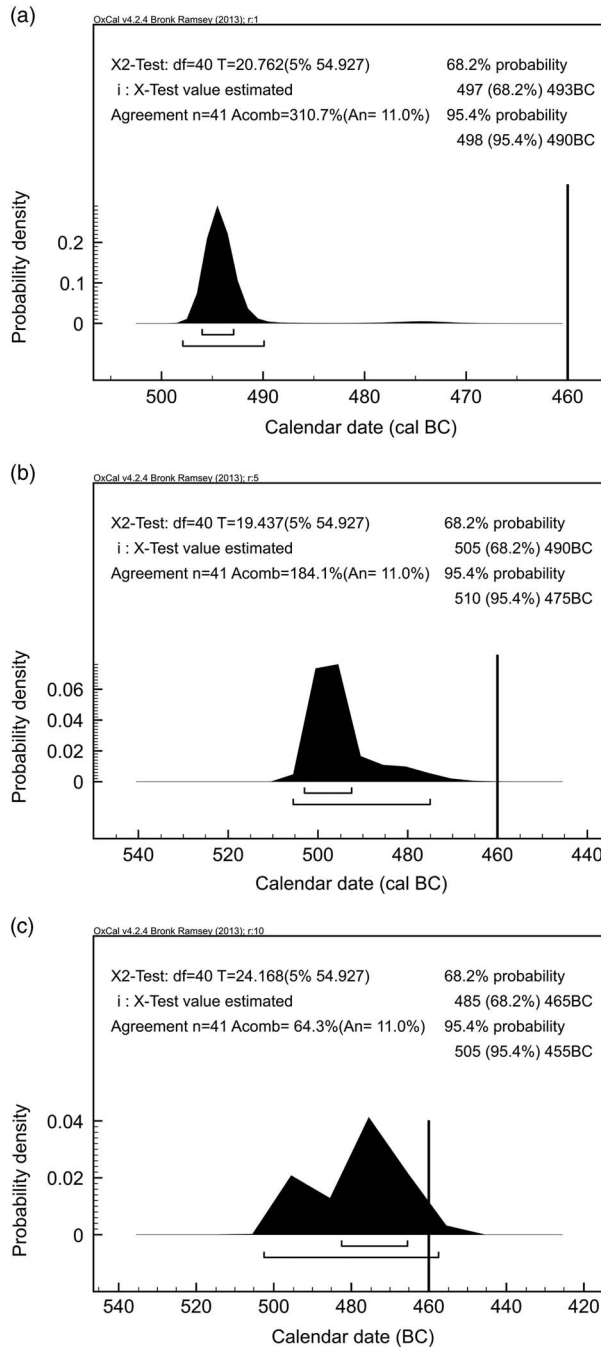


Figure 4 Summary results of the wiggle-match on the decadal averages of the T-947 single-ring data with: (a) 1-yr resolution, (b) 5-yr resolution, and (c) 10-yr resolution. The vertical line marks the target date of 460 cal BC. Note that for single-year resolution rounding was not implemented.

Table 3 Results of the  $^{14}\text{C}$  age determinations on decadal blocks of T-947 wood.

Rings	SUERC-nr	Age ( $^{14}\text{C}$ yr BP)	1- $\sigma$ error	$\delta^{13}\text{C}$ (‰)
5–14	52772	2431	29	–24.1
15–24	52773	2397	29	–24.1
25–34	52774	2497	29	–24.1
35–44	52775	2454	28	–24.5
45–54	52776	2395	29	–25.1
55–64	52777	2486	28	–24.5
65–74	52778	2489	29	–25
75–84	52782	2508	29	–24.2
85–94	52783	2414	29	–24.1
10–19	52784	2440	28	–24.8
20–29	52785	2424	29	–28.8
30–39	52786	2461	28	–24.9
40–49	52787	2492	28	–24.9
50–59	52788	2435	28	–24.3
60–69	52792	2486	28	–23.7
70–79	52793	2418	28	–24.3
80–89	52794	2463	28	–23.9

The results of the decadal measurements (Table 3), confirm the results of the decadal averages. The statistics that indicate the quality of fit to the calibration curve are good ( $\chi^2 = 24.356$ ; 5% critical value at 16 degrees of freedom is 26.296;  $A_{\text{combine}} = 18.5\%$ , with  $A_n = 17.1\%$ ), showing that the decadal measurements do follow the calibration curve. What is more interesting, however, is the calibrated date range of the wiggle-match, lying at 645–640 cal BC (0.5% probability) or 505–450 cal BC (94.9%), with the greatest probability in the range 495–475 cal BC (68.2%) (Figure 5). Even though the 95.4% range of this new wiggle-match covers the target date of 460 BC, the majority of the probability is located over the preceding decades, as was the case with the results of the wiggle-matches based on the decadal averages. Hence, although the decadal sample blocks fit the calibration curve well, in line with the intuition on matching sample spans to the underpinning calibration data, other factors seem to also be in play.

### Discussion 2: Effects of Small-Scale Offsets on Wiggle-Match Dating

The apparent systematic offset between the results of the T-947 measurements and the target date can be traced to the relationship between the actual past trend in  $^{14}\text{C}$  activity and its estimate in the mean of the calibration curve. This becomes clear when the decadal averages (discussed above) are plotted against the calibration curve (Figure 6), for although they all at least overlap within 2  $\sigma$  of the calibration curve, they also cluster above the mean of the calibration curve for a substantial part of the wiggle-match. The result is that the individual calibrations of the decadal averages are all shifted closer to the shape of the calibration curve around 500 cal BC, leading to the observed systematic offset.

The offset observed in the wiggle-matches in all probability originates from the small systematic difference between the decadal averages of the single-ring measurements and the estimated position of the mean of the calibration curve. The occasional presence of such offsets at different points in time can be expected as the past trend in  $^{14}\text{C}$  activity was continuous and so the values of preceding years would have had an effect on the values of the succeeding years, i.e. they autocorrelate. Also, for short periods of time over the 50,000 yr of the calibration curve, the

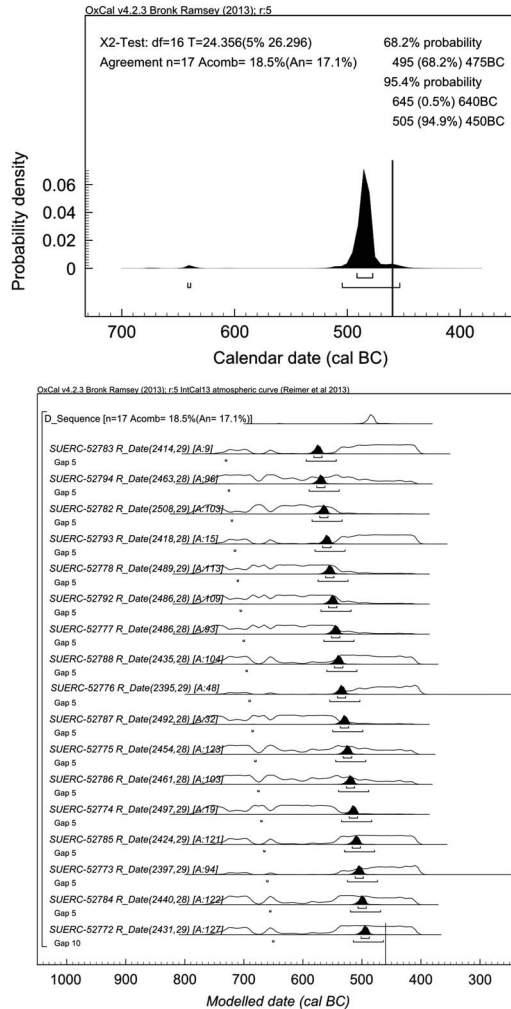


Figure 5 Results of the wiggle-match on the decadal determinations of the T-947 data: the summary result (above) and the detailed results (below). The vertical line marks the target date of 460 cal BC.

calibration measurements could have clustered above or below the past trend of  $^{14}\text{C}$  on a stochastic basis alone. In these periods, the mean of the calibration curve can be expected to deviate from the actual past trend in  $^{14}\text{C}$ , even though it remains within the estimated error boundaries—as was the case with the T-947 decadal averages. This has little conceivable effect on most applications of the  $^{14}\text{C}$  dating technique, but in the case of short sequence wiggle-matches, such transient offsets can lead to the development of a small-scale systematic error, even though the actual trend in past  $^{14}\text{C}$  activities is within the uncertainty margin of the calibration curve. There is an alternative explanation though, as the fifth century cal BC witnesses substantial divergence within the underpinning calibration data (Reimer et al. 2013). Over this period, some of the calibration data produced at Belfast display a systematic offset towards ages younger than those implied by the ratified calibration curve and much of the Seattle calibration data witnesses a deflection towards older ages. This could imply that there

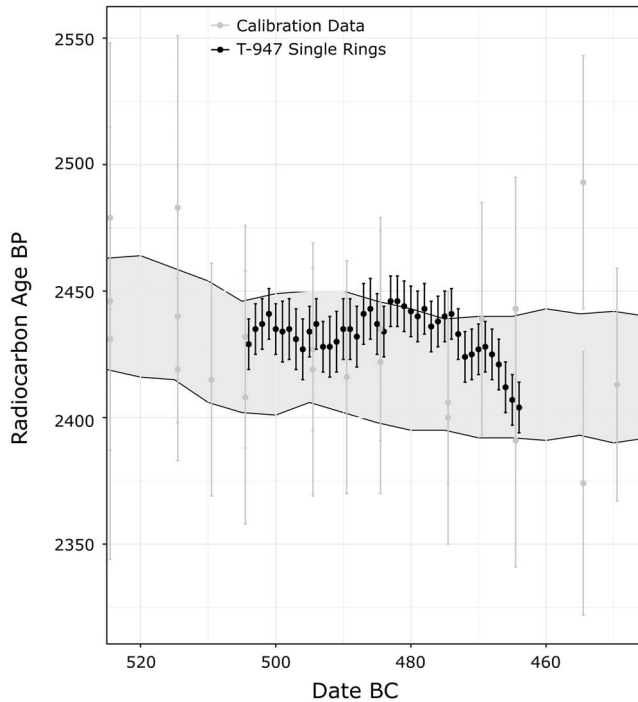


Figure 6 Decadal averages for T-947 single-ring data and the relevant calibration data plotted against the 95.4% probability interval of the IntCal13 calibration curve. Notice how even though the errors on the decadal averages overlap with the probability interval at all points, the means of these averages diverge from the mean of the calibration curve and would need to be placed towards older dates for a closer fit. Error bars are  $2\sigma$ .

have either been divergences in underpinning  $^{14}\text{C}$  trends, or technical issues with the underpinning calibration measurements, which could have induced the offset we witness here. At the same time, however, until we have more detailed calibration data over this period (the new UC Irvine data makes an important step in that direction; Taylor and Southon 2013), it is difficult to assess what caused the calibration discrepancies in the fifth century cal BC and we do not pursue this discrepancy further.

The impact of the autocorrelation-induced error is relative to the span of the sequence to be dated, the precision of the individual measurements and the number of measurements put into the wiggle-match. As the length of the wiggle-matched sequence increases, the probable effects of a small offset over its duration decrease. Hence, the recommendations of Galimberti et al. (2004) and Goslar and Madry (1998) that extending the lengths of the wiggle-matches whenever possible ought to be reiterated, not only for the sake of modeled date precision, but also for accuracy. However, where the available material is short-lived, this approach cannot be pursued. Instead, accuracy has to be gained from the relationship between the scale of the offsets in question and measurement precision and intensity; in essence, the better the precision of the individual measurements and the greater their frequency, the better the description of the past trend in  $^{14}\text{C}$  activity and the greater the probability of a systematic offset in short-sequence wiggle-matches. In the case of T-947, the presence of the autocorrelation-induced error could only be identified due to the very high precision of the decadal averages. If we only had routine

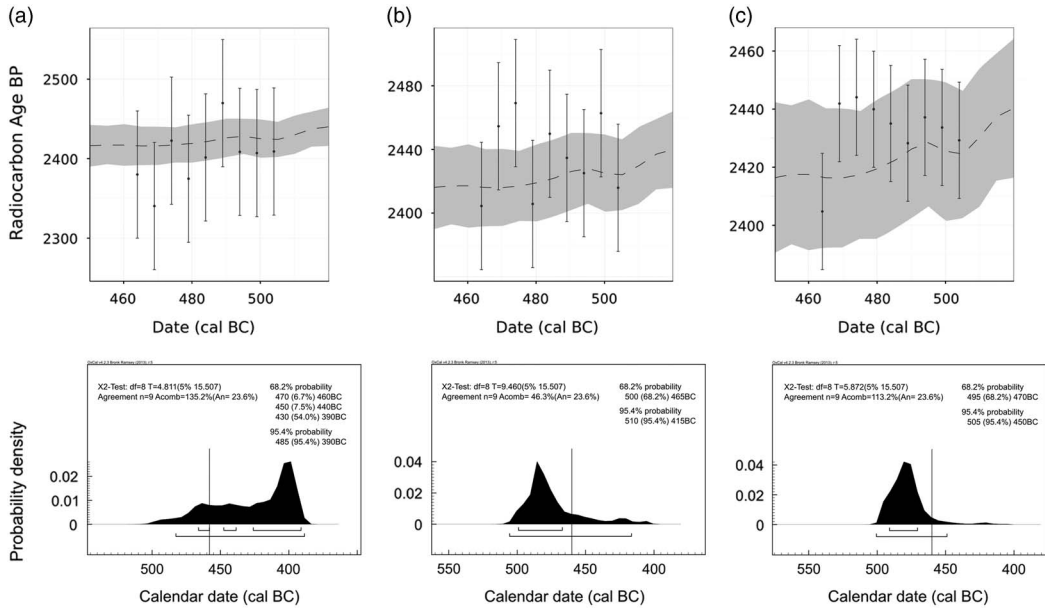


Figure 7 Simulations from the means of the T-947 decadal averages with altered standard deviations to represent different measurement precisions (a: 40, b: 20, and c:  $10^{14}\text{C}$  yr at  $1\sigma$ ). These and other simulation results back the intuition that the effect of the small-scale offsets depends on the relationship between the precision of the measurements on the unknown sample, their frequency and the precision of the calibration curve. Error bars are  $2\sigma$ .

precision decadal measurements to work with, the presence of the target date within the 95.4% modeled date range would have meant that the possibility of autocorrelation would never be considered.

Nevertheless, the routine precision decadal data demonstrate that with lower measurement precision and fewer determinations, the wiggle-matches can be accurate. This is an expected result, as with increasing measurement precision and frequency the description of the past  $^{14}\text{C}$  trend also improves, thus increasing the effects on any small-scale offsets. This intuition is further supported through limited simulation studies (Figure 7, Table 4). To carry out these we have simulated wiggle-matches based on nine measurements of varying precision (10, 20, and  $40^{14}\text{C}$  yr at  $1\sigma$ ), with a target date of 460 cal BC. The simulated measurements were drawn based on our decadal averages and repeated 20 times for each precision value, with the results scored for both precision (length of the posterior distribution) and accuracy of the 68.2% and 95.4% highest posterior density intervals. Although more extensive studies are called for before the risk introduced by the offsets can be quantified, the evidence suggests that, with less intensive sampling patterns and routine precision measurements ( $3\text{--}4\text{‰}$   $\Delta^{14}\text{C}$ , or  $24\text{--}32^{14}\text{C}$  yr at  $1\sigma$ ), it ought to be possible to attain results precise to less than a century (95.4% probability), while retaining accuracy. In archaeological contexts such results can be further improved upon by inclusion in Bayesian site models (Hamilton et al. 2007).

The approach suggested above is also advantageous from the point of view of the archaeological research design. As mentioned earlier, one of the original concerns of the study was not so much the question of attaining accurate results, but rather the issue of obtaining them on a routine basis at an acceptable cost. The outcomes of the wiggle-match conducted on decadal sample blocks measured to routine precision and the results of simulations based on the decadal

Table 4 Results of the simulations on effects of the autocorrelation offset upon the precision and accuracy of the T-947 data, for a target date of 460 cal BC. Precision refers to the total length of the calibrated date ranges and Accuracy refers to whether the given simulated wiggle-match included the target date within the respective date range (68.2% or 95.4%). The simulations were based on the means of overlapping T-947 decadal averages, spaced at 5-yr intervals, and standard deviations indicated in the table (10, 20, 40 yr).

1- $\sigma$ meas. error ( $^{14}\text{C}$ yr)	Simulation nr	68.2% precision	68.2% accurate	95.4% precision	95.4% accurate
40	1	60	Yes	100	Yes
	2	80	Yes	140	Yes
	3	65	No	145	No
	4	70	Yes	100	Yes
	5	80	Yes	115	Yes
	6	85	Yes	115	Yes
	7	80	No	110	Yes
	8	75	Yes	125	Yes
	9	80	Yes	110	Yes
	10	80	Yes	150	Yes
	11	70	Yes	100	Yes
	12	75	Yes	110	Yes
	13	80	Yes	150	Yes
	14	80	No	165	Yes
	15	85	Yes	100	Yes
	16	55	Yes	120	Yes
	17	85	Yes	150	Yes
	18	75	Yes	120	Yes
	19	70	Yes	105	Yes
	20	70	Yes	125	Yes
20	1	40	No	75	Yes
	2	70	Yes	90	Yes
	3	60	Yes	95	Yes
	4	45	Yes	95	Yes
	5	35	No	65	Yes
	6	45	Yes	100	Yes
	7	35	No	85	Yes
	8	45	No	95	Yes
	9	45	Yes	95	Yes
	10	25	No	75	Yes
	11	45	Yes	95	Yes
	12	65	Yes	90	Yes
	13	30	No	65	No
	14	55	Yes	95	Yes
20	15	25	No	85	Yes
	16	30	No	75	Yes
	17	70	Yes	95	Yes
	18	60	Yes	100	Yes
	19	45	Yes	100	Yes
	20	25	No	65	Yes
10	1	25	No	70	Yes
	2	30	No	70	Yes



Table 4 (Continued)

1- $\sigma$ meas. error ( $^{14}\text{C}$ yr)	Simulation nr	68.2% precision	68.2% accurate	95.4% precision	95.4% accurate
	3	35	No	70	Yes
	4	30	No	70	Yes
	5	30	No	55	Yes
	6	30	No	50	Yes
	7	35	Yes	80	Yes
	8	60	Yes	75	Yes
	9	20	No	80	No
	10	35	Yes	85	Yes
	11	20	No	75	Yes
	12	30	No	70	Yes
	13	30	No	40	No
	14	25	No	40	No
	15	35	No	85	Yes
	16	40	Yes	90	Yes
	17	30	Yes	90	Yes
	18	20	No	40	No
	19	20	No	50	Yes
	20	40	Yes	90	Yes

averages of single ring data show that the necessary parameters can be attained for sequences of less than 50 yr within four or five measurements. Furthermore, once multiple timbers dated in such a fashion are combined within even a simple model, such as the bounded uniform phase (Buck et al. 1992; Bronk Ramsey 2009), the precision of the final calibrated and modeled estimates can reach about 100 yr, or less, even during the Hallstatt plateau (Jacobsson 2017).

## CONCLUSION

This research set out to evaluate the viability of wiggle-matching for routine dating of short-span (<50 yr) timbers from the Hallstatt calibration plateau. The main motivation for this work comes from the need to improve the dating of wetland settlements from the earlier phases of the Iron Age in Scotland, although the conclusions are applicable to a wide range of contexts where short-lived timbers survive in the archaeological record.

The two main recommendations that emerged during the course of the study are:

1. The need to match the number of rings within any particular measured subsample to the underpinning calibration data.
2. The need to be aware of the risk of autocorrelation-induced errors.

The issue of matching the sample spans derives from the smoothing effect of multi-year sampling that was necessary during the original development of the calibration data sets and it means that any sampling to a higher resolution will only introduce information redundancy, which, in some cases, may lead to biases. The autocorrelation-induced errors derive from the short-term offsets between the actual past trend of  $^{14}\text{C}$  activity and the mean of the calibration curve. These kinds of errors emerge due to the application of the wiggle-match model over those parts of the calibration curve where the past trend in  $^{14}\text{C}$  activity diverges from the mean of the

curve for several consecutive decades. Although this divergence ought to be within the uncertainty of the curve and would have little impact on the interpretation of single calibrations, as well as most models, it can affect wiggle-match dating. The issue can be overcome in most cases by avoiding the construction of very precise short-span wiggle-matches. Instead, the approach put forward is to build groups of coarser wiggle-matches and place them within suitable Bayesian site models. In this way, the accuracy of the results is more assured and the allocation of measurements is better suited to the possible challenges of actual archaeological sites, where it is often important to obtain multiple dates from a variety of structural features.

The results presented in this paper rest to a large extent on the state of the calibration program. However, with an increasing number of calibration data sets, improving precision of the calibration curve and the introduction of a larger number of single-ring data sets, the main recommendations suggested herein will begin to change. Basing wiggle-match dating on single rings may become viable for a greater range of periods than just the most recent half millennium and the increasing precision of the calibration curve will mean that there will be less possibility of auto-correlation induced errors. Nevertheless, until this happens, matching the sample spans and accounting for the risks of small-scale offsets appears to be a necessity in routine archaeological applications of the wiggle-match dating technique.

## ACKNOWLEDGMENTS

We would like to thank Historic Scotland for the funding for the analytical work. PJ would like to thank the Caledonian Research Fund for the Caledonian Scholarship that enabled this work. We would also like to thank Historic Environment Scotland for Supporting the Open Access publication of this paper.

## REFERENCES

- Bayliss A, Marshall P, Tyers C, Bronk Ramsey C, Cook G, Freeman SPHT, Griffiths S. 2017. Informing conservation: towards  $^{14}\text{C}$  wiggle-matching of short tree-ring sequences from medieval buildings in England. *Radiocarbon* 59(3):985–1007.
- Beer J, McCracken K, von Steiger R. 2012. *Cosmogenic Radionuclides: Theory and Applications in the Terrestrial and Space Environments*. Heidelberg: Springer.
- Bronk Ramsey C. 2009. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51(1):337–60.
- Bronk Ramsey C, van der Plicht J, Weninger B. 2001. “Wiggle matching” radiocarbon dates. *Radiocarbon* 43(2):381–9.
- Buck CE, Litton CD, Smith AFM. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19(5):497–512.
- Cavers G, Crone BA. forthcoming. *A Lake Dwelling in its landscape; Iron Age settlement at Cults Loch, Castle Kennedy, Dumfries Galloway*. Oxford: Oxbow Books.
- Cavers G, Crone BA, Engl R, Fouracre L, Hunter F, Robertson J, Thoms J. 2011. Refining chronological resolution of Iron Age Scotland: Excavations at Dorman’s Island Crannog, Dumfries and Galloway. *Journal of Wetland Archaeology* 10:71–108.
- Crone BA. 1988. Dendrochronology and the study of crannogs [unpublished PhD thesis]. University of Sheffield.
- Crone BA. 2012. Forging a chronological framework for Scottish crannogs; the radiocarbon and dendrochronological evidence. In: Midgley MS, Sanders J, editors. *Lake Dwellings after Robert Munro. Proceedings from the Munro International Seminar: The Lake Dwellings of Europe 22nd and 23rd October 2010*. University of Edinburgh. Leiden: Sidestone Press. p 139–68.
- Crone A. 2014. Dendrochronological studies of alder (*Alnus glutinosa*) on Scottish crannogs. *Journal of Wetland Archaeology* 14:22–33.
- Crone A. Forthcoming. Chronology. In: Crone A, Cavers G. Forthcoming.
- Ferguson CW, Huber B, Suess HE. 1966. Determination of the age of Swiss lake dwellings as an example of dendrochronologically-calibrated radiocarbon dating. *Zeitschrift Für Naturforschung* 19(A): 1173–7.
- Galimberti M, Bronk Ramsey C, Manning SW. 2004. Wiggle-match dating of tree-ring sequences. *Radiocarbon* 46(2):917–24.

- Goslar T, Madry W. 1998. Using the Bayesian method to study the precision of dating by wiggle-matching. *Radiocarbon* 40(1):551–60.
- Güttler D, Wacker L, Kromer B, Friedrich M, Synal H-A. 2013. Evidence for 11-year solar cycle in tree rings from 1010 to 1110 AD—progress on high precision AMS measurements. *Nuclear Instruments and Methods in Physics Research B* 294:459–63.
- Hamilton WD, Bayliss A, Menuge A, Bronk Ramsey C, Cook GT. 2007. “Rev Thomas Bayes: get ready to wiggle” Bayesian modelling, radiocarbon wiggle-matching, and the north wing of Baguley Hall. *Vernacular Architecture* 38(1): 87–97.
- Hoper ST, McCormac FG, Hogg AG, Higham TFG, Head MJ. 1998. Evaluation of wood pretreatments on oak and cedar. *Radiocarbon* 40(1): 45–50.
- Jacobsson P. 2017. Wiggle-match dating and the structural sequence at Cults Loch 3 in Cavers and Crone. Forthcoming.
- Jull AJT, Panyushkina IP, Lange TE, Kukarskih VV, Myglan VS, Clark KJ, Salzer MW, Burr GS, Leavitt SW. 2014. Excursions in the  $^{14}\text{C}$  record at A.D. 774–775 in tree rings from Russia and America. *Geophysical Research Letters* 41(8): 3004–10.
- Millard AR. 2008. Comment on article by Blackwell and Buck. *Bayesian Analysis* 3(2):255–61.
- Miyake F, Nagaya K, Masuda K, Nakamura T. 2012. A signature of cosmic-ray increase in AD 774–775 from tree rings in Japan. *Nature* 486(7402): 240–42.
- Niu M, Heaton TJ, Blackwell PG, Buck CE. 2013. The Bayesian approach to radiocarbon curve estimation: IntCal13, Marine13 and SHCal13 methodologies. *Radiocarbon* 55(4):1905–22.
- Pearson GW, Pilcher JR, Baillie MGL, Corbett DM, Qua F. 1986. High-precision  $^{14}\text{C}$  measurement of Irish oaks to show the natural  $^{14}\text{C}$  variations from AD 1840 to 5210 BC. *Radiocarbon* 28(2B): 911–34.
- Reimer PJ, Bard E, Bayliss A, Beck JW, Blackwell PG, Bronk Ramsey C, Buck CE, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hafliðason H, Hajdas I, Hatte C, Heaton TJ, Hoffmann DL, Hogg AG, Kaiser KF, Kromer B, Manning SW, Niu M, Reimer RW, Richards DA, Scott EM, Southon JR, Staff RA, Turney C SM, van der Plicht J. 2013. IntCal13 and Marine13 radiocarbon age calibration curves 0–50,000 years cal BP. *Radiocarbon* 55(4): 1869–87.
- Scott EM, Harkness DD, Cook GT. 1997. *T.I.R.I. Report on Third International Radiocarbon Comparison*. Glasgow: Glasgow University.
- Slota PJ, Jull AJT, Linick TW, Toolin LJ. 1987. Preparation of small samples for  $^{14}\text{C}$  accelerator targets by catalytic reduction of CO. *Radiocarbon* 29(2):303–6.
- Stuiver M, Becker B. 1986. High-precision decadal calibration of the radiocarbon time scale, AD 1950–2500 BC. *Radiocarbon* 28(2B):863–910.
- Stuiver M, Reimer PJ, Braziunas TF. 1998. High-precision radiocarbon age calibration for terrestrial and marine samples. *Radiocarbon* 40(3):1127–51.
- Suzuki K, Sakurai H, Takashi Y, Sato T, Gunji S, Tokanai F, Matsuzaki H, Tsuchiya Y. 2010. Precise comparison of  $^{14}\text{C}$  ages from Choukai Jindai cedar with IntCal04 raw data. *Radiocarbon* 52(4):1599–609.
- Taylor RE, Southon J. 2013. Reviewing the mid-first millennium BC  $^{14}\text{C}$  “warp” using  $^{14}\text{C}$  bristlecone pine data. *Nuclear Instruments and Methods in Physics Research B* 294:440–3.
- Tyers C, Sidell J, van der Plicht J, Marshall P, Cook G, Bronk Ramsey C, Bayliss A. 2009. Wiggle-matching using known-age pine from Jermyn Street, London. *Radiocarbon* 51(2): 385–96.
- Vandeputte K, Moens L, Dams R. 1996. Improved sealed-tube combustion of organic samples to  $\text{CO}_2$  for stable carbon isotope analysis. *Radiocarbon Dating and Percent Carbon Determinations. Analytical Letters* 29(15):2761–73.
- Wacker L, Güttler D, Goll J, Hurni JP, Synal H-A, Walti N. 2014. Radiocarbon dating to a single year by means of rapid atmospheric  $^{14}\text{C}$  changes. *Radiocarbon* 56(2):573–9.
- Ward GK, Wilson SR. 1978. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20(1):19–31.