RADIOCARBON, Vol 46, Nr 1, 2004, p 17–24

TOWARDS HIGH-PRECISION AMS: PROGRESS AND LIMITATIONS

Christopher Bronk Ramsey[1] • Thomas Higham • Philip Leach

Oxford Radiocarbon Accelerator Unit, University of Oxford, United Kingdom.

ABSTRACT. Precision and accuracy in accelerator mass spectrometry (AMS) dating relies on the systematic reduction of errors at all stages of the dating process, from sampling to AMS measurement. With new AMS systems providing much better precision and accuracy for the final stage of the process, we need to review the process as a whole to test the accuracy of reported results. A new High Voltage Engineering Europa (HVEE) AMS system was accepted at Oxford in September 2002. Since then, the system has been in routine use for AMS dating and here we report on our experiences during the first year. The AMS system itself is known to be capable of making measurements on single targets to a precision of better than 0.2% for the $^{14}C/^{13}C$ ratio and better than 0.1% for the $^{13}C/^{12}C$ ratio. In routine operation, we measure known-age wood to a precision of just above 0.3%, including uncertainties in background and pretreatment. At these levels, the scatter in results is no higher than reported errors, suggesting that uncertainties of ±25 to ±30 ^{14}C yr can be reliably reported on single target measurements. This provides a test of all parts of the process for a particular material in a particular state of preservation. More generally, sample pretreatment should remove as much contamination as feasible from the sample while adding as little laboratory contamination as possible. For more complex materials, such as bone, there is clearly more work needed to prove good reproducibility and insignificant offsets in all circumstances. Strategies for testing accuracy and precision on unknown material are discussed here, as well as the possibilities of one day reaching precisions equivalent to errors of <±20 ^{14}C yr.

INTRODUCTION

At the last Radiocarbon Conference in Jerusalem and the last Accelerator Mass Spectrometry (AMS) Conference in Nagoya, there was some debate about the precision that can be obtained from a single AMS determination. In September 2002, the HVEE AMS system at Oxford was accepted (Bronk Ramsey et al. 2002; Gottdang et al. 2001) and showed the capability for very high-precision measurements on the AMS system itself. It was not clear at that time, however, whether the precision for routine operation on real samples would show similar improvements or be limited by other factors. We now have a year of experience of operating the new AMS and this paper draws on that experience to look at the potential for high-precision measurements by AMS.

Most very high-precision (defined here as <0.25% or <±20 ^{14}C yr) ^{14}C measurements have been undertaken using conventional counting methods (see, for example, Pearson 1979; McCormac 1992; Kalin et al. 1995). This is because of a number of possible factors, one of which may simply be cost. On the more scientific level, however, the 2 methods have different advantages and disadvantages.

When it comes to precision, the key theoretical disadvantages of AMS are the following:

- The difficulty of achieving isotope ratio stability over the measurement period; stability of about 0.1% is needed for high precision;
- The small sample size makes the method vulnerable to low levels of contamination either from the sample itself or more critically (since the ^{14}C isotope composition is often more radically different) from the pretreatment process;
- The danger of selecting unrepresentative portions of inhomogeneous samples.

The theoretical advantages are the following:

- The possibility of performing more rigorous sample pretreatment (even to the compound-specific level);
- The ability to select less contaminated fractions of inhomogeneous samples;

[1]Corresponding author. Email: christopher.ramsey@archaeology-research.oxford.ac.uk.

Proceedings of the 18th International Radiocarbon Conference, edited by N Beavan Athfield and R J Sparks
RADIOCARBON, Vol 46, Nr 1, 2004, p 17–24

- The ability to sub-sample and repeat measurements.

Each stage of the dating process (sampling, pretreatment, combustion, graphitization, and AMS measurement) play an important role in determining precision (Bryant et al. 2001). The complex subjects of taphonomy and sampling clearly depend very strongly on the specific application (e.g. Bayliss 2000; van Strydonck et al. 2000) and will not be discussed further here beyond pointing out the problems of homogeneity and the need to exclude any contamination.

Chemical pretreatment is almost certainly the key element in precision dating. It is much more difficult to check for reproducibility and accuracy at this stage than in the remaining stages where pre-prepared standards can be used. There are essentially two important questions to be asked of any pretreatment method:

1. Does the method remove contamination present in the sample (acceptable levels will depend on the relative age of the contaminants but are likely to be <1–5%)?
2. Does the method add significant levels of contamination (acceptable levels are <0.1% since ^{14}C concentration is usually radically different)?

To some extent, these work against one another in that more complex pretreatment methods are more likely to add contaminants of their own. For many materials, AMS should have the advantage that more rigorous pretreatment is possible while still retaining sufficient sample to date. However, the very small size of AMS samples makes them particularly susceptible to added contamination (for example, a sample containing 10 mg C needs to keep any added contaminants below 10 μg).

In the combustion and graphitization stages, it is important that any contaminants introduced (and there are always some) are well characterized. By this stage for AMS, the sample size is likely to be of the order of 1 mg C, so the levels of any contaminants need to be well understood at the sub-microgram level. Graphitization reproducibility is also important in that variations in the form of the graphite can give variability in beam characteristics in the AMS, which might, in turn, affect the accuracy of the AMS measurement (depending on the AMS system).

The AMS measurement itself is, of course, very important in determining precision and accuracy. However, the ability of an AMS to make precise and accurate measurements is also relatively easily tested by repeat measurements on standards of various kinds.

AMS SYSTEM AND METHODOLOGY

All of the measurements described in this paper have been performed using our routine procedures. After sampling and pretreatment (depending on type), the samples are

- Loaded into tin capsules; these capsules are pre-cleaned in cyclohexane and acetone, then a sub-sample combusted to check the blank contribution is <2 μg C;
- Combusted in a CHN analyzer furnace system; each sample combustion is preceded by the combustion of an empty tin capsule to purge the system;
- Passed through a gas chromatograph containing *Carbosieve*™;
- Measured for stable isotope values ($\delta^{13}C$ and $\delta^{15}N$); the $\delta^{13}C$ ratio is not used in the date calculation (see below);
- Collected as CO_2 from the He stream;
- A fixed aliquot of CO_2, equivalent to 1.5–2.0 mg C, is taken; a small proportion of samples and standards are made at half this size and then measured on the AMS in specific batches; samples smaller than this are measured as gas and are not discussed here;
- Mixed with hydrogen in a ratio 2.2:1 H_2:CO_2;

- Reduced to graphite on an iron catalyst at 560 °C;
- Pressed into a target (2-mm-diameter hole) for the HVEE 846 ion source.

The targets are then collected into batches for AMS measurement. A batch of up to 56 targets will typically comprise the following:

- 5–6 oxalic acid II samples;
- 2 anthracite backgrounds (graphitized in the same way as all other samples but combusted in bulk for us by the CIO Groningen laboratory);
- 2 known-age samples (tree rings and other standards);
- Other quality assurance and background samples (Mary Rose bone, Jurassic charcoal, etc.), depending on applications;
- Unknown samples making up the total.

Thus, we get several independent measurements on HOXII for each run which gives one measure of internal consistency.

The AMS used is an HVEE Tandetron with 2 recombinators (one used for the graphite measurements as outlined here on an 846 ion source, the other is used for gas or graphite measurements with a SO110 ion source).

To calculate a date, we do the following:

- Subtract the AMS/graphitization background from the measurements on anthracite;
- Calculate the $\delta^{13}C$ for each target using the $^{13}C/^{12}C$ ratio relative to the average value for the HOXII targets;
- Calculate the stable isotope corrected ^{14}C date from the $^{14}C/^{13}C$ ratio of the samples relative to the average value from the HOXII samples with the AMS derived $\delta^{13}C$ value; note that this is mathematically identical to calculating the date from the $^{14}C/^{12}C$ ratio but with a squared stable isotope correction.

The main point to note here is that we use total integrated charge ratios (less any time for cleaning up targets) and use the AMS-derived $\delta^{13}C$ rather than that from the stable isotope mass spectrometer. This corrects for any linear mass fractionation taking place in the graphitization, ion source, or AMS system. AMS labs argue whether this is necessary or not and it may depend on the AMS system itself. In our case, we have good evidence to suggest it is important. During commissioning our AMS, we measured 35 HOXII samples with a variety of graphitization and pressing methods; this gave rise to a range of $^{13}C/^{12}C$ ratios even within 1 run (Figure 1). However, if linear fractionation was corrected for as outlined above, none were more than 1.7 σ from the expected value; the average precision was 0.19% (equivalent to 15 yr) with a standard deviation of 0.17% (equivalent to 14 yr). In other words, with AMS stable isotope correction, the results are within their statistical errors.

More evidence that fractionation is generally linear can be seen in Figure 2 where absolute isotope ratios are considered.

Ultimately, whatever the arguments about methodology might be, the only real test is the reproducibility and accuracy of measurements using the method, and the following sections concentrate on aspects of this.

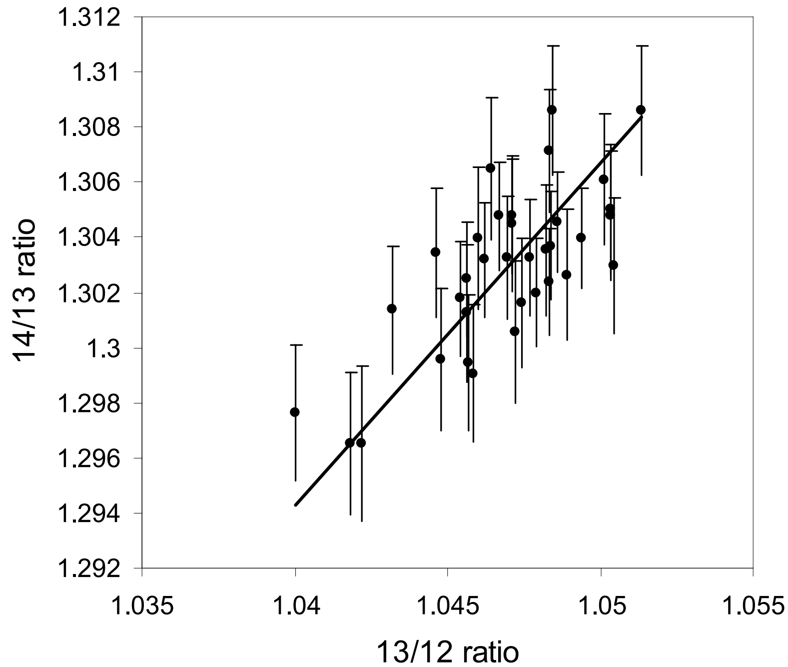

Figure 1 This figure shows a plot of $^{14}C/^{13}C$ ratio ($\times 10^{-12}$) against $^{13}C/^{12}C$ ratio ($\times 10^{-2}$) during a series of runs made throughout acceptance tests for the Oxford AMS; the stable isotope ratios have been normalized so that the average for each batch is the same as the overall average; thus, only internal run variation is displayed; the fractionation observed is linear; refer to main text for details of conclusions.

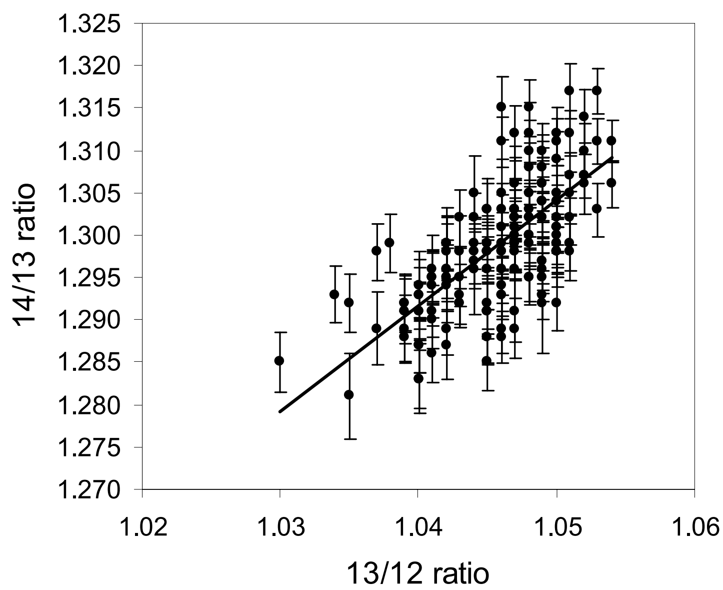

Figure 2 This plot shows the same quantities as Figure 1 but for all HOXII samples run since commissioning and, in this case, the absolute ratio has been plotted; thus, much of the variation seen here is run-run variation depending on AMS setup, stripper gas pressure, etc.; despite this, it can clearly be seen that most of the fractionation taking place is linearly mass dependent and, therefore, best corrected for by using the AMS stable isotope value.

AMS PERFORMANCE

In order to test the performance of the AMS system alone, it is really necessary to exclude some other aspects of the process. For the AMS acceptance tests, therefore, we used batches of homogenized HOXII graphite produced for HVEE by the Leibniz-Labor AMS Facility (Kiel) and pressed into targets by them. In the test of precision, 12 such targets were measured in 2 batches over 2 days. Each of them was measured to a precision of 0.2% (equivalent to 16 yr). The scatter in $^{14}C/^{13}C$ ratios was within the statistical uncertainty and the scatter in $^{13}C/^{12}C$ was 0.05% (or 0.03% if 1 outlier is removed). Thus, the intrinsic reproducibility of the system on essentially identical material seems to be better than 0.1%.

The results quoted in the previous section showed at the time that even when the graphite was of more variable consistency, the stable isotope-corrected ^{14}C concentration was still stable to well within 0.2%, implying a similar underlying instrument precision.

The machine background (no ^{14}C injected) is about 0.017% (70 ka BP). For our own graphite samples which have gone through the standard graphitization process (with about 1.5–2.0 mg C), the background is about 0.15% (52 ka BP) and in the best conditions can be as low as 0.1% (55 ka BP). For electrode graphite (no processing but unknown composition), we get a background just slightly lower than this (about 57 ka BP). This paper is really concentrated at high precision at the recent end of the ^{14}C spectrum (where precision as such is most usually an issue). Clearly, AMS backgrounds are low enough to be insignificant in this context.

Acceptance tests are, of course, not the same as routine measurements. Since the commissioning of the AMS, however, we have been quoting a precision on samples less than 2000 yr old, which averages to 27 ^{14}C yr (c.f. 38 ^{14}C yr with our old AMS) and have measured our HOXII samples to an average precision of 0.29% (equivalent to 23 ^{14}C yr, c.f. 0.43% on our old AMS). Clearly, we need to justify this in terms of accuracy and repeatability.

The first thing we can do is to look at how the HOXII values scatter around the mean value. For each HOXII target, we calculate how far away from the central value it is as a proportion of its statistical uncertainty. To correct for the loss of 1 degree of freedom, we then multiply this figure by $\sqrt{[n/(n-1)]}$, where n is the number of standards in that batch. This is to take account of the fact that the mean is defined from the values themselves. The results are shown in Table 1 and seem to be within the range expected.

Table 1 Proportion of measurements on HOXII, tree rings, and duplicate samples, lying within 1, 2, and 3 σ of expected values; in the case of the duplicate measurements and those on HOXII, the ranges have been corrected for the loss of 1 degree of freedom, since the comparisons are against an average rather than a known value.

Range offsets from true value as a factor of uncertainty (σ)	Proportion of 203 HOXII measurements lying within range	Proportion of 96 tree-ring measurements lying within range	Proportion of 66 duplicate AMS measurements on 30 samples lying within range	Proportion of 55 duplicate measurements (including pretreatment) on 27 samples lying within range
–3 to 3	100%	100%	100%	100%
–2 to 2	97.5%	95.8%	95.5%	94.6%
–1 to 1	75.9%	75.0%	69.7%	65.5%

These measurements essentially show the repeatability of the combustion, graphitization, target pressing, and AMS measurement. To see the implications of this on more normal samples, we need to include the stages of sample pretreatment and look at ratios different from those of the standard (in case of non-linearity).

Accuracy of Measurements on Wood

The most frequent other test we make on the AMS system is measurements on known-age wood, supplied to us by the Queen's University of Belfast and calibrated to the INTCAL98 calibration curve (Stuiver et al. 1998). Most of this wood has a ^{14}C concentration roughly half that of the HOXII standard, so it should provide a good measure of linearity. The wood undergoes an acid/base/acid treatment followed by chlorite bleach before the standard method described above (Hedges et al. 1989). We date decadal sections of wood similar to those used for the calibration curve. On the new AMS, we have so far measured 96 samples with an overall bias (to the older side) of 8.9 ± 3.3 yr. The proportions lying within 1, 2, and 3 σ are shown in Table 1. The implication of this is that we are not underestimating our errors.

Accuracy of Measurements on Bone

The pretreatment of bone is much more complex than any other routinely dated material (see Bronk Ramsey et al. 2004 for more details on current method and for details of known-age material). One problem here is that bone of really well-known composition is much more difficult to find because of the complexity of dietary and reservoir effects. We have chosen to work on a bone from the wreck of the Mary Rose. This ship sank in AD 1545 and the bone we are using is from pigs which were onboard the ship. From the calibration curve, we would expect the bone to have a ^{14}C date of 309 ± 4.6 BP. If we take samples in the range normally accepted within our procedures (>7 mg collagen), we have 13 measurements by our current method (some of which have a solvent extraction and some of which do not). The average uncertainty on these measurements is 23.7 ^{14}C yr and the standard deviation is 29.9 ^{14}C yr. This scatter is just low enough to pass a χ^2 test at 95% confidence. The average value is 321 ± 6.5 BP, slightly older than expected by about 10 ^{14}C yr, but calibrates to the correct age at 95% confidence.

Overall, we conclude from this that it is reasonable to quote errors lower than 30 yr on individual measurements of similar material, though clearly we would not be confident in using multiple measurements to provide an overall uncertainty of 6.5 ^{14}C yr on an unknown sample. It should also be borne in mind that in the case of bone, turnover rates and, more critically, dietary reservoir offsets can compromise accuracy of calendrical dating even if a precise and accurate measurement of the ^{14}C concentration of the bone is possible (e.g. Bonsall et al. 2004). We think that more research will be needed to improve precision much beyond current levels on bone.

Accuracy of Measurements on Charcoal

We are in the process of evaluating known-age charcoal material for routine testing. However, the pretreatment methods applied to charcoal are a subset of those applied to wood, so the tree-ring data is pertinent here. In addition, some measurements do have internal checks as in the case of wiggle-matched dating of charred wood from Miletos (Galimberti et al. 2004). In this instance, we have 7 decades of charcoal from a wiggle-matched piece of wood. The 7 measurements have an average precision of 21 yr (achieved by measuring each sample twice). The fit to the calibration curve is excellent and shows no obvious offsets, indicating that in this instance, anyway, we can measure charcoal to a precision close to 20 yr.

Testing Accuracy on Real Samples

Tests on known-age material are ideal for testing laboratory procedures. They are particularly useful for testing whether the methods themselves introduce any significant bias. However, in the case of actual samples from an archaeological or environmental context, the question of removal of sample-specific contamination becomes more important. To some extent, all of the 3 pretreatment cases above do address this issue but of course only in one specific type of context. More generally, it is only by testing for reproducibility of duplicate measurements (including duplicate pretreatments) that we can begin to address this issue. To investigate this, we randomly select samples (about 1 in 20) for duplicate pretreatment and dating and we sometimes pretreat samples more than once for other reasons. We also perform duplicate AMS measurements on some samples.

Over a period of 1 yr (from 1 September 2002), we have 55 duplicated pretreatments on 27 individual samples. We also have 66 duplicated AMS measurements on 30 samples with 1 pretreatment. We can look at the offset in these measurements from the mean value as a proportion of the quoted uncertainty in the same way that we did for the standards in Table 1 (taking into account the fact that there are n-1 degrees of freedom). The results are given in Table 1 and Figure 3.

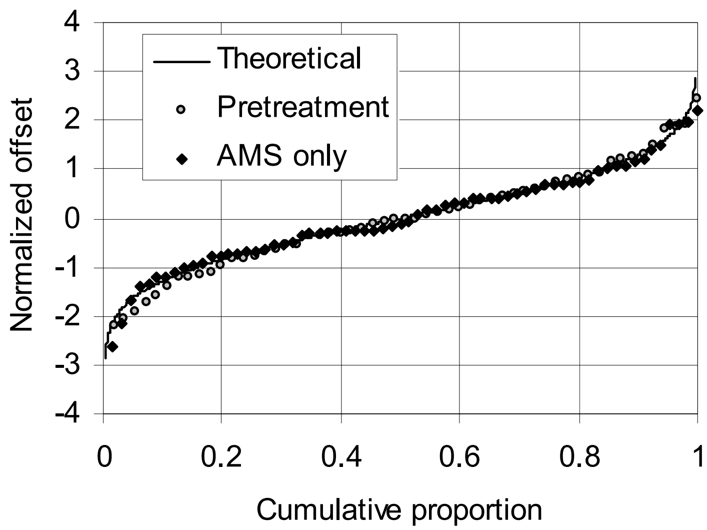

Figure 3 This figure shows the cumulative proportion of duplicate measurements lying within their quoted uncertainties. The normalized offset is the offset from the mean divided by the uncertainty and multiplied by $\sqrt{[n/(n-1)]}$, where n is the number of duplicate measurements.

In this analysis, we have included all duplicate measurements made in this period with 3 exceptions: one is a sample (P13468) which gave a very low yield (3.3 mg collagen) using our old bone pretreatment method, which was subsequently re-dated using our new method (i.e. not a true duplicate); the other two are targets (P14321 and P13951) which gave very low currents (less than one-third of the average) and for which the AMS measurement was subsequently repeated.

These data show that the reproducibility of measurements is in almost exact accord with the quoted uncertainties. As this is on normal samples rather than standards, this is a useful measure of accuracy, although it does not test for any systematic effects.

CONCLUSIONS

With the latest AMS instrumentation, it is possible to make AMS measurements to a higher precision than previously possible. At Oxford, we have been testing whether this improved AMS precision can be carried through into improved routine precision while retaining accuracy. All of the indications are that this is the case and that it is possible, for example, to quote routine precisions averaging 27 ^{14}C yr on single AMS measurements for samples from the last 2000 yr with demonstrable accuracy. Given that there is no scatter above the uncertainties quoted at this level and no measurable systematic offsets greater than about 10 yr, we would consider it valid to combine two such measurements (ideally with duplicate pretreatment) to provide an uncertainty just below 20 yr. Beyond that, we do not think we yet have sufficient evidence to draw any conclusions.

Although measurements on duplicates look encouraging, we cannot be sure that the accuracy of measurements on real samples from different contexts will always be this good, since we do not have the data to prove that contamination present in samples can always be removed sufficiently not to compromise the results at the 0.1% level. Ultimately, site-dependent issues assume importance, and we think that only by continuing to improve quoted precision and by evaluating internal consistency at specific sites will this issue really be tackled properly.

The evidence seems to suggest that AMS precision and accuracy can be improved even further. However, we recognize that it is crucial to continually monitor the accuracy and reproducibility, both using known-age material and duplicates of unknown samples to ensure that quoted errors do not become unrealistically precise. Clearly, pretreatment is the critical link in the measurement chain, both in terms of the removal of contaminants and the bias introduced by the processes themselves. The best approach would seem to be to use the most rigorous method available (to tackle contaminants in the sample) that can be demonstrated to give accurate and unbiased results on known-age material. However, it is likely that it is in pretreatment that the accuracy is ultimately limited.

REFERENCES

Bayliss A. 2000. On the taphonomy of charcoal samples for radiocarbon dating. In: Acte du colloque "C14 and Archaeology." *Revue d'Archéometrie*. 1998. GMCPA. p 51–6.

Bonsall C, Cook GT, Hedges REM, Higham TFG, Pickard C, Radovanovic I. 2004. Radiocarbon and stable isotope evidence of dietary change from the Mesolithic to the Middle Ages in the Iron Gates: new results from Lepenski Vir. *Radiocarbon*, these proceedings.

Bronk Ramsey C, Hedges REM, Mous D, Gottdang A, van den Broek R. 2002. *Status of the new AMS at Oxford*. Talk given at the 9th AMS Conference, Nagoya, Japan, September 2002.

Bronk Ramsey C, Higham TFG, Bowles A, Hedges REM. 2004. Improvements to the pretreatment of bone at Oxford. *Radiocarbon*, these proceedings.

Bryant C, Carmi I, Cook GT, Gulliksen S, Harkness DD, Heinemeier J, McGee E, Naysmith P, Possnert G, Scott EM, van der Plicht J, van Strydonck M. 2001. Is comparability of ^{14}C dates an issue?: A status report on the Fourth International Radiocarbon Intercomparison. *Radiocarbon* 43(2A):321–4.

Cook GT, Bonsall C, Hedges REM, McSweeney K, Boroneant V, Pettitt PB. 2001. A freshwater diet-derived ^{14}C reservoir effect at the Stone Age sites in the Iron Gates gorge. *Radiocarbon* 43(2A):453–60.

Galimberti M, Bronk Ramsey C, Manning SW. 2004. Wiggle-match dating of tree-ring sequences. *Radiocarbon,* these proceedings.

Gottdang A, Klein M, Mous DJW. 2001. Accelerator mass spectrometry at High Voltage Engineering Europa (HVEE). *Radiocarbon* 43(2A):149–56.

Hedges REM, Law IA, Bronk CR, Housley RA. 1989. The Oxford accelerator mass spectrometry facility: technical developments in routine dating. *Archaeometry* 31:99–113.

Kalin RM, McCormac FG, Damon PE, Eastoe CJ, Long A. 1995. Intercomparison of high-precision ^{14}C measurements at the University of Arizona and the Queen's University of Belfast Radiocarbon Laboratories. *Radiocarbon* 37(1):33–8.

McCormac FG. 1992. Liquid scintillation counter characterization, optimization and benzene purity correction. *Radiocarbon* 34(1):37–45.

Pearson GW. 1979. Precise ^{14}C measurement by LS counting. *Radiocarbon* 21(1):1–22.

Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac FG, van der Plicht J, Spurk M. 1998. INTCAL98 radiocarbon age calibration, 24,000–0 cal AD. *Radiocarbon* 40(3):1041–83.

van Strydonck M, Nelson DE, Crombé P, Bronk Ramsey C, Scott EM, van der Plicht J, Hedges REM. 2000. Rapport du groupe de travail: les limites de methode du carbone 14 applicquée a l'archéologie, What's in a ^{14}C date? In: Acte du colloque "C14 and Archaeology." *Revue d'Archéometrie*. 1998. GMCPA. p 433–48.