

© 2020 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 licence (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted re-use, distribution, and reproduction in any medium, provided the original work is properly cited.

AN INDEPENDENT ASSESSMENT OF UNCERTAINTY FOR RADIOCARBON ANALYSIS WITH THE NEW GENERATION HIGH-YIELD ACCELERATOR MASS SPECTROMETERS

A T Aerts-Bijma*  • D Paul  • M W Dee  • S W L Palstra  • H A J Meijer 

Centre for Isotope Research (CIO), Energy and Sustainability Research Institute Groningen, Faculty of Science and Engineering, University of Groningen, Groningen, The Netherlands

ABSTRACT. The radiocarbon (^{14}C) dating facility at the Centre for Isotope Research, University of Groningen went through a major upgrade in 2017 and this included installation of a MICADAS accelerator mass spectrometer (AMS). In the first 18 months, we performed 4000 sample and 3000 reference measurements. A careful evaluation of those measurement results is presented, to characterize the various sources of uncertainty and to ultimately assign, for every sample measurement, a realistic expanded uncertainty. This analysis was performed on the measurements of secondary references and sample duplicates in various phases of their processing steps. The final expanded uncertainty includes both the ^{14}C measurement uncertainties and uncertainties originating from pretreatment steps. Where the ^{14}C measurement uncertainty includes straightforward uncertainties arising from Poisson statistics, background subtraction, calibration on Oxalic Acid II and $\delta^{13}\text{C}$ correction, the uncertainties originating from pretreatment steps are based on the spread of actual measurement results for secondary references and sample duplicates. We show that the ^{14}C measurement uncertainty requires expansion, depending on the number of processing steps involved prior to a ^{14}C measurement, by a maximum factor of 1.6 at our laboratory. By using these expansion (multiplication) factors, we make our reported uncertainty both more realistic and reliable.

KEYWORDS: contamination, MICADAS, Poisson statistics, quality assurance radiocarbon, sample duplicates, uncertainty analysis.

INTRODUCTION

Ever since the first determinations of the age of archeological objects by radiocarbon (^{14}C) in 1949 (Libby et al. 1949; Arnold and Libby 1949), researchers have striven for easier and more accurate measurements. One of the first breakthroughs was the gas proportional counting technique using CO_2 , developed and applied by de Vries and Barendsen (1952, 1954). In the 1980s, accelerator mass spectrometry (AMS) revolutionized the field: faster measurements and the capability to use orders of magnitude less sample made numerous new applications of ^{14}C possible (for example Damon et al. 1989; Linick et al. 1989; Cook et al. 2003; Meijer et al. 2006; Rasmussen et al. 2009; Dee et al. 2013). Since then, innovations to the AMS technique have led to incremental but important improvements, mostly concerning the precision and accuracy achievable.

Such innovations are illustrated by the changes in equipment used for ^{14}C measurement at the Centre for Isotope Research (CIO) of the University of Groningen. Proportional counters were in operation from the early 1950s until 2011. From 1994 until 2017 the CIO operated a 3 MV ^{14}C -dedicated accelerator mass spectrometer “tandetron” (AMS) (High Voltage Engineering Europa, the Netherlands) (Gott dang et al. 1995). Since September 2017, a MICADAS (MIni CARbon DAting System, Ionplus, Switzerland) has been in operation (Synal et al. 2007; Wacker et al. 2010). The MICADAS outperforms the previous AMS in several respects, the most important being increased efficiency of the source (typically 5% compared to $\approx 1\%$, resulting in more counts) and besides the measuring on graphite, the additional ability to measure CO_2 gas directly.

*Corresponding author. Email: a.t.aerts-bijma@rug.nl.

With the former generation of accelerator mass spectrometers, the number of individual counts acquired from a sample, nearly always determined the ^{14}C measurement uncertainty limitation (a product of the underlying Poisson statistics) and in fact also the final reported uncertainty. All other contributions to the ^{14}C measurement uncertainty (such as the background variability and the ^{13}C correction), were much smaller, which meant in practice that they were both negligible and impossible to determine. However, with the new MICADAS machine this is no longer the case, thanks to the much higher count rates from the source as a result of its increased efficiency. As the Poisson uncertainties have now decreased dramatically, other variables in the measurement contribute to the ^{14}C measurement uncertainty, and still this measurement uncertainty is much lower than before. In addition, contributions to the final reported uncertainty of the various preprocessing and processing steps are no longer negligible either. For a reliable and confident estimate of this final uncertainty, or with a better name the expanded uncertainty (as defined in JGCM 100 2008), we have to take all these contributions into account.

In this paper, we systematically evaluate those other contributions, such that in the end we obtain a reliable expanded uncertainty. With a total of approximately 7000 ^{14}C measurements in its first one and a half year of operation, we have gathered abundant information for this detailed uncertainty analysis.

We are, obviously, not the first to publish a manuscript discussing uncertainty analysis of ^{14}C measurements. Stuiver and Polach (1977) did address the issue in their much-cited paper, and (Hedges et al. 1989) describe which sources contribute to the uncertainty. Scott et al. (2007) give a nice general overview of uncertainty calculation in radiocarbon dating. None of these papers, however, treat the subject in all the details that we show and explain in the present work.

In the coming sections, we demonstrate quantification of uncertainty sources of the various steps in the process from sample to ^{14}C measurement as well as of the actual ^{14}C measurements. To determine and quantify each uncertainty contribution in this whole process, the long-term performance of our secondary references is key. However, as these materials are all pure, homogeneous substances, uncertainty based on their analysis alone might be systematically too small. Therefore, in addition we monitored known-age samples and unknown sample duplicates in various phases of the preparation process. In this way, we could establish whether homogeneity, complicated combustion conditions, or contaminations from the environment (soil-derived compounds, CO_2 from laboratory air, contamination through memory effect during combustion) play a significant role.

EXPERIMENTAL SETUP AND METHOD

The preparation of samples for ^{14}C analysis is dependent on the type of material. Archaeological samples (wood, bone, charcoal, seeds) usually need chemical pretreatment, followed by combustion, whereas CO_2 in air and in water only needs to be extracted in various ways (for example cryogenically or acidification of alkaline solutions). Still other samples, like carbonates, are treated with acid to convert them to CO_2 . The CO_2 produced, or isolated, is subsequently graphitized (reduced to elemental carbon) and pressed into AMS sample holders (usually called cathodes), with which the actual measurement can be performed.

During all the steps from sample preparation to measurement, utmost care and attention are required to keep the contamination accumulated to a minimum, as the ^{14}C variability will increase along with the amount of contamination. The accumulated contamination contributes to a greater expanded uncertainty.

In this paper, we restrict ourselves to samples with a regular graphite mass (2 mgC: mg of carbon). Gas measurements and small-sized (<0.6 mgC) graphite samples will be dealt with in a forthcoming publication. In the next section, we will briefly describe all the relevant preparation steps, with emphasis on the aspects that influence the ^{14}C signal in the samples, and thus contribute to the final uncertainty.

Processes from Sample to ^{14}C Measurement and Possible Contamination Sources

Sample

A ^{14}C measurement is performed on a small portion of the original sample. Inhomogeneity in the sample material is an issue in providing reliable dates and it is an important source in the final reported uncertainty, and at the same time hard to quantify (only with duplicate or even multiple sampling, as we will show).

Chemical Pretreatment

Sample-specific chemical pretreatment, in other words the best way to collect an isotopically reliable carbon fraction from each type of material, has been and continues to be the subject of many publications, discussions, and round-robin tests between laboratories. The routine chemical pretreatments used in our laboratory were originally summarized by Mook and Stuiver (1983) and more recently by Dee et al. (2020).

Chemical pretreatment can affect the final isotopic composition of a sample and hence contribute to the expanded uncertainty in several ways, such as through the introduction of contamination with “foreign” carbon during sample handling, or the incomplete removal of contamination that was originally present in the sample.

CO₂ Production

Different types of material need different techniques to produce CO₂. The techniques currently being used at the CIO are described in Dee et al. (2020). Combustion is performed with solid samples, like bones, seeds, charcoal, and wood, to produce CO₂. Carbonates are converted to CO₂ in an in-house glasswork manifold. It is described fully in Meijer (2009). The method of extracting CO₂ and measuring ^{14}C in cremated bones samples was first reported in Lanting et al. (2001). CO₂ in air is cryogenically extracted or chemically captured in a sodium hydroxide solution.

Possible contributions from the CO₂ production to the expanded uncertainty are incomplete combustion, memory from one sample to the next (either in the combustion oven, or in the CO₂ trap), contamination from the reagents (oven chemicals, helium, oxygen, and sodium hydroxide) and contamination during sample handling, for example not using ultraclean equipment and instruments. In addition, the exchange of CO₂ with CO₂ from a previous sample at the glass surface (wall absorption/desorption) and leakages can also cause contamination. The risk of contamination in various steps will lead to added uncertainty in the final result.

Graphitization Systems

The CO₂ samples from all different sources will have to be reduced to elemental carbon (commonly referred to as graphitized) for ^{14}C measurements with higher precision. Our graphitization set-up is described in Aerts-Bijma et al. (1997); De Rooij et al. (2010); and Dee et al. (2020).

Possible contributions from the graphitization process to the expanded standard uncertainty are exchange with previous CO₂ at the glass surface and the presence of contamination in the glasswork manifold or in the iron powder. We store the elemental carbon samples (graphite) produced in the reaction tubes in which they were formed, at room temperature and with Argon added up to atmospheric pressure. These samples are pressed just before measurement, into aluminum sample holders with an automated home-built press at approximately 1 MPa (Aerts-Bijma et al. 1997).

During the graphitization process sample-to-sample contamination can also occur if the reactors are not cleaned sufficiently. Furthermore, experience has taught us that the pressed graphite-iron mixture is susceptible to carbon uptake from air, and even while at vacuum in the ionization chamber (Paul et al. 2016).

¹⁴C Measurements

We performed all the relevant ¹⁴C measurements on our MICADAS. In our routine operation, a regular batch consists of five Oxalic Acid II references, necessary for the calibration of the batch (the tuning of the machine also requires one additional Oxalic Acid II reference), four sample-specific background references (¹⁴C-free material, resembling the sample materials as much as possible), two secondary references and 28 samples (unknowns). Approximately forty minutes of measurement time per sample yields typically 750,000 ¹⁴C counts for the Oxalic Acid II calibration material.

Input Data for the Uncertainty Analysis

As part of our general quality control and assurance procedures, (secondary) reference materials and background materials are processed alongside the samples, and these materials are selected to resemble the various sample types as closely as possible. The results for these various materials are of course a valuable source of information for our uncertainty assessment. In addition, sample duplicates are also frequently analyzed.

Samples Analyzed as Duplicates

In our routine operation, we regularly prepare sample duplicates. The sample is divided into two portions and then the chemical pretreatment, the CO₂ extraction, the graphitization and the ¹⁴C measurement are performed on different days, as if they were two different unknowns. Thus, everything in the whole process is performed as independently as possible. These full duplicates yield a wealth of information about the expanded uncertainty. As they are pretreated and measured in different batches, their uncertainties are to a large extent independent.

To discriminate these “full” duplicates from other partial duplicates (see below) we call them pretreatment duplicates. In order to enhance the readability of this paper we divided the different kinds of duplicate also into categories from 1 to 4. These pretreatment duplicates are from now on called category (cat.) 4 duplicates, as all four steps from chemical pretreatment, CO₂ preparation, graphitization and ¹⁴C measurement are different for these duplicates.

A special cat. 4 duplicate is the VIRI F Horse bone. In some series of bones, which have to be pretreated, leftover material from the VIRI intercomparison (Scott et al. 2010), the VIRI F horse bone, is pretreated as well. This sample is considered a known-age cat. 4 duplicate, because many ¹⁴C laboratories have dated this material.

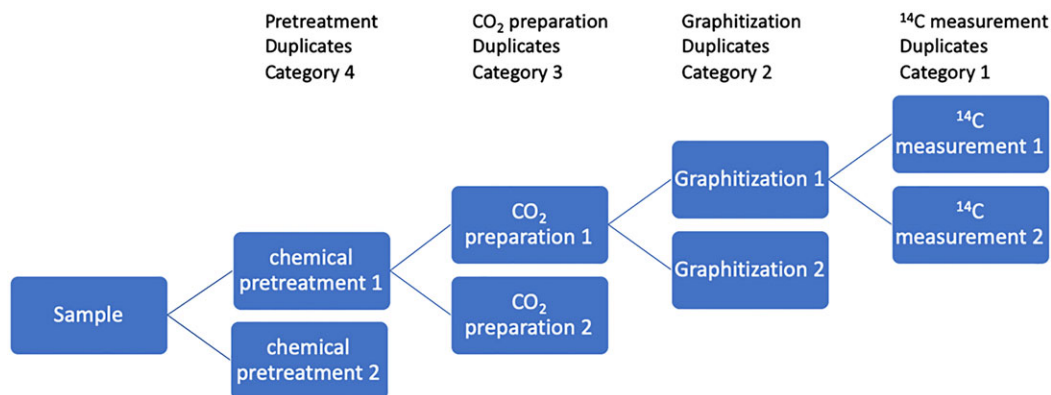


Figure 1 A schematic overview of different categories of duplicates. A higher category number refers to a higher number of independent steps in the total process.

CO_2 preparation duplicates are samples divided into different portions after the chemical pretreatment, which are then separately handled further, so these duplicates share the chemical pretreatment, but not the following three steps (CO_2 preparation, graphitization, and ^{14}C measurement). These duplicates are thus cat. 3 duplicates.

For the graphitization duplicates, CO_2 from one CO_2 preparation process is divided into two or three portions, and each portion is separately graphitized, making them cat. 2 duplicates.

A cat. 1 ^{14}C measurement duplicate means that the graphite formed is split into two portions (so all parts of the process before are common). Such duplicates rarely occur.

All duplicates contribute to a better understanding of the contributions of the different preparation steps to the expanded uncertainty. The different duplicates are further clarified in Figure 1.

Background Materials

Several sample-specific ^{14}C -free (“dead”) materials are available in our laboratory for identification of the “background”; that is, the modern carbon contamination, accumulated during the process from chemical pretreatment until ^{14}C measurement. Background wood (“bgw”, Kitzbuhel I, Tirol, Austria) and background collagen (“bgc”, Latton Quarry LQH 12) (Cook et al. 2012) have been selected for measuring the combustion background. Both materials are known to be far older than the detection limit of ^{14}C . The background material for carbonates (shells and cremated calcined bones) is named GS-35 (grained marble from a stonemasonry in Groningen). The graphitization background gas is Rommenh oller CO_2 , a fossil gas of geological origin (Linde gases).

Secondary References

Besides the sample duplicates and background materials, several secondary references are combusted and graphitized together with the samples and treated as “samples of known ^{14}C content”, to monitor the total process. As the secondary references are all pure substances, they form the basis for determining and quantifying each uncertainty contribution in the whole process. These secondary references are IAEA-C8, IAEA-C7 (oxalic acid, Le Clercq et al. 1998), and GS-51 (Groningen Standard, cane sugar). For

these references, extensive measurement records over more than 20 years have been compiled. This present study is, however, based on data measured with the MICADAS only.

These secondary reference materials are treated in two separate ways. The first involves them being combusted in large quantities, yielding five to ten liters of pure CO₂ collected in small cylinders (called “bulk”). Thanks to the large quantities of gas, we can use them to prepare many samples over the years, which then do not contain combustion-induced variability (in analogy to the different types of sample duplicates, we can call them cat. 2 secondary references). Therefore, these cylinder gas analyses can monitor the variability induced by the graphitization process and the subsequent measurement alone. However, we also use these secondary reference materials as individual samples, where they are combusted in small amounts (2 mgC) and serve as combustion references (cat. 3 secondary references).

CALCULATION OF OUR BEST ESTIMATE FOR THE ¹⁴C MEASUREMENT UNCERTAINTY

The ¹⁴C content of a sample is expressed as Fraction Modern (Reimer et al. 2004). Because the original batch of the calibration material, Oxalic Acid I is exhausted, Oxalic Acid II is the international calibration reference (cal) with assigned values for $F^{14}C_n \equiv 134.066\%$ and $\delta^{13}C_{VPDB} \equiv -17.8\%$.

Every measured sample is calibrated using:

$$F^{14}C_n = 134.066 * \frac{({}^{14}C/{}^{12}C)_{sample} - ({}^{14}C/{}^{12}C)_{bg}}{({}^{14}C/{}^{12}C)_{cal} - ({}^{14}C/{}^{12}C)_{bg}} * \frac{\left(\frac{0.975}{1+\delta^{13}C_{sample}}\right)^2}{\left(\frac{0.975}{1+\delta^{13}C_{cal}}\right)^2} \quad (1)$$

The subscript sample, bg and cal refers to sample, background and Oxalic Acid II respectively; $\delta^{13}C_{sample}$ is the value measured by the MICADAS; its $\delta^{13}C$ scale is calibrated using the assigned Oxalic Acid II value of $\delta^{13}C_{VPDB} = -17.8\%$.

The uncertainty in a ¹⁴C measurement is then derived from the partial derivatives of $F^{14}C_n$ with respect to each of the variables and is called $dF^{14}C_n$.

Every measured quantity in Eq. (1) has its own uncertainty. The uncertainty in the $({}^{14}C/{}^{12}C)_{sample}$ is the statistical uncertainty (Poisson counting statistics). The uncertainty in $({}^{14}C/{}^{12}C)_{cal}$, is the uncertainty in the mean of the calibration reference (Oxalic Acid II). The uncertainty in the mean value for the calibration material, and not the standard deviation, is the right choice, as this uncertainty in the mean is relevant for the accuracy of the calibrated scale. Instead of dealing with this uncertainty on a per batch basis, we use the average of the uncertainty in the mean over a considerable number of batches (over the preceding 4 months, typically, under normal circumstances, 50 batches) to avoid the statistical fluctuations in our estimate of the ¹⁴C measurement uncertainty. In the first phase of operation, we did not do that, which led to an underestimation of this uncertainty (see Appendix 1). The relevant uncertainty in the next variable, $({}^{14}C/{}^{12}C)_{bg}$, is the spread (standard deviation) in the background, also over the preceding 4 months. This spread is relevant for the variability in the individual backgrounds and thus also for the samples. The 4-monthly values are closely monitored for sudden or gradual changes on a monthly basis. The uncertainty in the variable $\delta^{13}C_{sample}$ (the $\delta^{13}C$ from the sample measured by AMS) is the uncertainty derived from the raw measurements. The standard error of the mean of the independent raw measurements (in normal routine, 8 independent measurements)

for each graphite sample, is calculated and serves as the uncertainty in $\delta^{13}\text{C}_{\text{sample}}$. For the last variable, $\delta^{13}\text{C}_{\text{cal}}$, we again need the uncertainty in the mean, and the typical value is $\pm 0.1\%$ (which makes this uncertainty source negligible in practice).

The quadratic sum of the above-mentioned components times their partial derivatives results in the ^{14}C measurement uncertainty ($d\mathbf{F}^{14}\text{C}_n$).

$$\begin{aligned}
 d\mathbf{F}^{14}\text{C}_n = & \left\{ \left(\frac{\sigma_{\text{poisson}}}{({}^{14}\text{C}/{}^{12}\text{C})_{\text{sample}} - ({}^{14}\text{C}/{}^{12}\text{C})_{\text{bg}}} \right)^2 + \left(\frac{\bar{\sigma}_{(\text{OxII},4 \text{ months})}}{({}^{14}\text{C}/{}^{12}\text{C})_{\text{cal}} - ({}^{14}\text{C}/{}^{12}\text{C})_{\text{bg}}} \right)^2 \right. \\
 & + \left(\frac{\sigma_{(\text{bg},4 \text{ months})} \cdot (({}^{14}\text{C}/{}^{12}\text{C})_{\text{sample}} - ({}^{14}\text{C}/{}^{12}\text{C})_{\text{cal}})}{(({}^{14}\text{C}/{}^{12}\text{C})_{\text{sample}} - ({}^{14}\text{C}/{}^{12}\text{C})_{\text{bg}}) \cdot (({}^{14}\text{C}/{}^{12}\text{C})_{\text{cal}} - ({}^{14}\text{C}/{}^{12}\text{C})_{\text{bg}})} \right)^2 \\
 & \left. + \left(\frac{2 \cdot \sigma_{\delta^{13}\text{C}_{\text{sample}}}}{1 + \delta^{13}\text{C}_{\text{sample}}} \right)^2 + \left(\frac{2 \cdot \sigma_{\delta^{13}\text{C}_{\text{cal}}}}{1 + \delta^{13}\text{C}_{\text{cal}}} \right)^2 \right\}^{0.5} * \mathbf{F}^{14}\text{C}_n \quad (2)
 \end{aligned}$$

This calculation, using the partial derivatives-approach, is a classical, linearized approximation of the real value of the ^{14}C measurement uncertainty. Correlations between the uncertainties in the different variables are ignored. We compared the outcome of this calculation to a Monte Carlo approach using the NIST Uncertainty Machine (NIST 2019). This is a web-based application for evaluating the measurement uncertainty associated with an output quantity defined by a measurement model of the form $y = f(x_0, \dots, x_n)$ (Lafarge and Possolo 2015). The Uncertainty Machine provides a numerically calculated probabilistic estimate of the uncertainty. The differences between the linearized approximation via the partial derivatives and the calculation via Monte Carlo turned out to be negligible (this is basically caused by the small size of the uncertainties relative to the values, making the linear approximation a very good one). Therefore, we preferred the ease of using the analytical method of the partial derivatives.

Researchers often report uncertainties in their results that do not contain all relevant sources of uncertainty. Therefore, their uncertainty estimates are usually too low. This is mostly caused by the fact that some of the uncertainty sources are very hard to estimate properly. This approach holds also for the analysis software from the MICADAS for data reduction, called BATS (Wacker et al. 2010a). This data analysis package is provided with the MICADAS, and it is a very powerful and versatile tool, so most groups operating a MICADAS use this package, including us. The ^{14}C measurement uncertainty the BATS package provides is based on the Poisson statistics, the so-called molecular correction ($^{13}\text{C}^+$ resulting from broken-up molecules), and the scatter of the blank samples. As the true ^{14}C measurement uncertainty is underestimated by this combination (and the programmers realize that, of course), in BATS an additional, arbitrary size error can be added by the user. This approach is in line with the “dark uncertainty” philosophy (see below). However, we prefer to explicitly account for all the uncertainty contributions as explained above, such that we produce the most reliable estimate for the expanded uncertainty in our ^{14}C measurement results.

For our approach, it is of course essential to be able to check if the ^{14}C measurement uncertainty ($\mathbf{dF}^{14}\text{C}_n$) that we calculate is indeed a good measure of that uncertainty. Therefore, we monitor the relationship between those calculated uncertainties and the realized uncertainties, where the latter are determined from the spread in (long) time series of various reference materials. Ideally, their ratio should be around 1.

This approach dates back to Birge (1932). The ^{14}C measurement uncertainties, as we calculate them along with the measurands, are called “internal errors”, whereas the uncertainties observable from the spread in measurands, are called “external errors”. The internal errors are then the expectation; the external errors are the realization of the uncertainty (Birge 1932) calls these the “prediction” and “answer to the prediction”, respectively). Their ratio is the reduced χ_{red}^2 (“chi-squared”). If the predicted internal error is correct, the value of χ_{red}^2 will be 1 within a certain statistical variability. However, if certain sources of uncertainty have not been accounted for in the internal error, χ_{red}^2 will be larger than 1. This is often the case in interlaboratory intercomparisons. The apparent extra source of uncertainty is called “dark uncertainty,” and there are different approaches for its calculation. The option in BATS to add extra uncertainty is in fact a possibility to account for this dark uncertainty. The “error multiplier” that has been used in the radiocarbon world can be interpreted along the same lines (Scott et al. 2007).

Birge’s (1932) original work has been taken up and extended by statisticians since then, for recent developments see for example (Rukhin 2009, Koepke et al. 2017, Merktas et al. 2019). For the sake of completeness, we give the expressions for the necessary quantities (weighted means, internal and external errors) in the Appendix 2.

In our attempt to account for all sources of uncertainty, we strive for the absence of “dark uncertainty”. Nevertheless, it is very possible, even likely, that such uncertainty still exists, as we cannot account quantitatively for variability in the chemical preparation and combustion, even though some of this variability is contained in the standard deviation of the background material, and in the calibration error.

We have two sources available with which we can check the completeness of our uncertainties. The secondary references provide long records, the spreads of which deliver the external errors. Sample duplicates on the other hand provide only two independent ^{14}C measurements, $\mathbf{F}^{14}\text{C}_n^1$ and $\mathbf{F}^{14}\text{C}_n^2$, each with their individual ^{14}C measurement uncertainty $\mathbf{dF}^{14}\text{C}_n^1$ and $\mathbf{dF}^{14}\text{C}_n^2$. The quadratic sum of the individual measurement uncertainties gives the uncertainty $\mathbf{dF}^{14}\text{C}_{n(\text{duplicates})}$, the difference between those two measurements gives $\Delta_{duplicates}$.

The ratio between $\Delta_{duplicates}$ and $\mathbf{dF}^{14}\text{C}_{n(\text{duplicates})}$

$$f_\sigma = \frac{\Delta_{duplicates}}{\mathbf{dF}^{14}\text{C}_{n_{duplicates}}} \quad (3)$$

is a value that should scale according to Gaussian expectations: for a large number of duplicates the average value of f_σ should be ≈ 0 , and the $\sigma(f_\sigma)$ should be ≈ 1 , and thus in 68% of the cases, the value should be between -1 and 1 . If the standard deviation of this distribution of f_σ , is in general too large, this would imply that the calculated ^{14}C measurement uncertainties are too low, and some “dark uncertainty” is present. This f_σ , is calculated for all duplicates, and the spread of this distribution, $\sigma(f_\sigma)$ is a measure for the expanded uncertainty for each of the various types of duplicates.

RESULTS FOR OUR UNCERTAINTIES

For our previous HV AMS, the calculated uncertainty according to the propagation of uncertainties from Eq. (1) has proven to be an adequate estimate for the expanded uncertainty from all different sources in the total process from chemical pretreatment until measurement. This was no surprise, however, as all contributions that could not be accounted for (chemical pretreatment variability) were overshadowed by the Poisson statistics contribution to the calculated ^{14}C measurement uncertainty.

For the MICADAS, however, this Poisson contribution is much smaller than for the HV AMS, due to the increased overall efficiency. When this pure measurement uncertainty decreases, further investigation of the other contributions to the expanded uncertainty is possible, and in fact necessary.

The expanded uncertainty in the final result is composed of four major contributions. These contributions, consecutively from latest to earliest in the sample handling process, are as follows: the contribution from the actual ^{14}C measurement (Eq. 1), from the graphitization, from the CO_2 preparation and from the chemical pretreatment. The first contribution, the ^{14}C measurement uncertainty ($\mathbf{dF}^{14}\text{C}_n$, Eq. 2) is the minimum uncertainty and is present in all the ^{14}C determinations. As this study is restricted to measurements on graphite cathodes, the extra uncertainty of the graphitization step, the second contribution, is automatically also incorporated in all measurements from background materials, secondary references (cat. 2) and sample duplicates (cat. 2). The third contribution from the CO_2 extraction is visible in the combustion background materials wood (bgw) and collagen (bgc), in the individually combusted secondary references (cat. 3), and in the CO_2 preparation sample duplicates (cat. 3, same chemical pretreatment, three different following steps). Finally, the fourth contribution, the uncertainty added in the chemical pretreatment, can be investigated with the pretreatment sample duplicates (cat. 4, where everything in the total process is different, see Figure 1). These four major contributions to the expanded uncertainty will be treated in the following texts.

As mentioned before, the first major contribution, the uncertainty in a ^{14}C measurement, is derived from the partial derivatives of $\mathbf{F}^{14}\text{C}_n$ with respect to each of the variables ($\mathbf{dF}^{14}\text{C}_n$). Figure 2 shows the typical contribution of the uncertainty in each variable from a representative measurement batch to this calculated ^{14}C measurement uncertainty. The quadratic sum of those components results in the ^{14}C measurement uncertainty (line f, black).

The uncertainty in the $(^{14}\text{C}/^{12}\text{C})_{\text{sample}}$ is the statistical uncertainty (Poisson counting statistics) (line a, gray). The Poisson counting statistics is still the largest contribution to $\mathbf{dF}^{14}\text{C}_n$. For low ^{14}C activities, the uncertainty is dominated by the spread in the background materials (line c, green).

This calculated ^{14}C measurement uncertainty needs to be put to the test. We expect it to be a valid uncertainty for pure gas samples, but for samples requiring pretreatment some extra “dark” uncertainty probably plays a role.

The first thorough check on our calculated uncertainties is given by the long-term spread of our secondary references. Table 1 provides the summary statistics for those secondary references. The references with a graphitization step only (cat. 2), are Rommenhöller (background), IAEA-C8 (bulk), IAEA-C7 (bulk), GS-51 (bulk), and Oxalic Acid II (bulk). Table 1

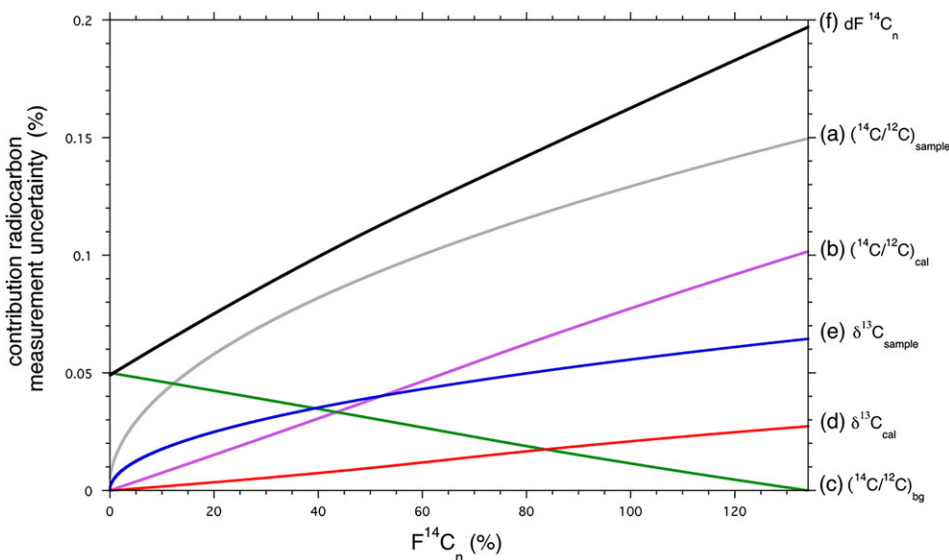


Figure 2 ^{14}C measurement uncertainty contributions (slightly smoothed) due to the partial derivatives of the variables in Eq. (1) for a representative measurement batch. The quadratic sum of those components results in the ^{14}C measurement uncertainty ($dF^{14}\text{C}_n$, line f, black). The uncertainty in $(^{14}\text{C}/^{12}\text{C})_{\text{sample}}$ is the statistical uncertainty (Poisson counting statistics) (line a, grey). The Poisson counting statistics is still the largest contribution to $dF^{14}\text{C}_n$. The uncertainty in $(^{14}\text{C}/^{12}\text{C})_{\text{cal}}$ (line b, magenta, calibration material is Oxalic Acid II) has practically no influence on samples with a low $(^{14}\text{C}/^{12}\text{C})_{\text{sample}}$, but the contribution increases for samples with a higher $(^{14}\text{C}/^{12}\text{C})_{\text{sample}}$. For samples with a low $(^{14}\text{C}/^{12}\text{C})_{\text{sample}}$, the uncertainty is dominated by the spread in $(^{14}\text{C}/^{12}\text{C})_{\text{bg}}$ (line c, green). Line d (red) and line e (blue) are the contribution due to the partial derivatives of, respectively, $\delta^{13}\text{C}_{\text{sample}}$ (measured by MICADAS) and $\delta^{13}\text{C}_{\text{cal}}$ (Oxalic Acid II). The latter one is practically negligible. (Please see electronic version for color figures.)

contains both the external and internal standard deviations (for the calculation equations see Appendix 2), and also χ_{red}^2 . The last column of Table 1 gives the probability that the difference between both standard deviations is significant (based on the statistics of the χ_{red}^2 distribution).

For four of the five graphitization references (cat. 2), χ_{red}^2 is <1 , implying (with on average $\approx 85\%$ probability) that the realized, external measurement uncertainty is somewhat smaller than the calculated, (internal) uncertainty. In other words, our calculated uncertainty ($dF^{14}\text{C}_n$) might be slightly overestimated.

The contribution from combustion is quantified by the secondary references that are individually combusted (cat. 3). Those are also listed in Table 1, and they have χ_{red}^2 values somewhat larger than 1, indicating that the calculated $dF^{14}\text{C}_n$ is a slight underestimation, and that there is some “dark” uncertainty present. Oxalic acid has a χ_{red}^2 much smaller than 1. However, the spread of the Oxalic Acid II is not representative, because this is used as calibration material and therefore for every batch the mean value is calibrated to become the assigned value of 134.066%.

The background wood (bgw) and background collagen (bgc) samples were chemically pretreated in large quantities, but individually combusted. Therefore, this pretreatment cannot influence their spread and those background references can be used as CO_2

Table 1 Long-term data of cat. 2 and cat. 3 secondary reference materials, from 1-7-2018 until 1-4-2019. N represents the number of measurements. The measured Fraction Modern $F^{14}C_n$ is an averaged result weighted by the individual uncertainties ($dF^{14}C_n$). The calculated ^{14}C measurement uncertainty $dF^{14}C_n$ is averaged. The squared external standard deviation (σ_{ext}) divided by the squared $dF^{14}C_n$ leads to the reduced Chi square (χ_{red}^2 , for equations see Appendix 2). Cat. 2 references show a χ_{red}^2 smaller than 1, implicating that $dF^{14}C_n$ is slightly overestimated. Cat. 3 references do have a χ_{red}^2 larger than 1, implicating that the combustion process contributes to a higher spread in the data. The last column gives the probability that the difference between both standard deviations is significant (based on the statistics of the χ_{red}^2 distribution).

Name	Material	Cat.	N	$F^{14}C_n$ (%)	$dF^{14}C_n$ (%)	(σ_{ext}) (%)	χ_{red}^2	Significance of difference
RH	CO ₂ gas	2	116	0.17	0.035*	0.032*	0.8	94%
IAEA-C8	Oxalic acid	2	37	15.13**	0.06	0.07	1.2	81%
IAEA-C7	Oxalic acid	2	37	49.75**	0.11	0.10	0.8	80%
GS-51	Cane sugar	2	26	108.76**	0.16	0.14	0.7	86%
OX II	Oxalic acid	2	213	134.07	0.18	0.18	0.9***	85%
Bgw	Wood	3	136	0.23	0.035*	0.039*	1.2	95%
Bgc	Collagen	3	12	0.25	0.034*	0.05	1.7	94%
GS-35	Marble	3	32	0.38	0.06	0.07	1.3	88%
IAEA-C8	Oxalic acid	3	22	15.06	0.06	0.06	1.0	—
IAEA-C7	Oxalic acid	3	31	49.49	0.11	0.12	1.2	79%
GS-51	Cane sugar	3	28	108.81	0.16	0.20	1.5	96%
OX II	Oxalic acid	3	124	134.07	0.19	0.15	0.6***	>>99%

*When the significant digit is between 1 and 4 an extra digit is shown.

**Recently a memory problem in the bulk combustion line was discovered to which IAEA-C7 and IAEA-C8 were vulnerable. This is the reason why the measured ^{14}C values of those cat. 2 secondary references are slightly different from the assigned values. For the purpose of this paper this has no further consequences.

***For every batch the mean value of the Oxalic Acid II references is calibrated to become the assigned value of 134.066%. Therefore, its overall spread is not representative.

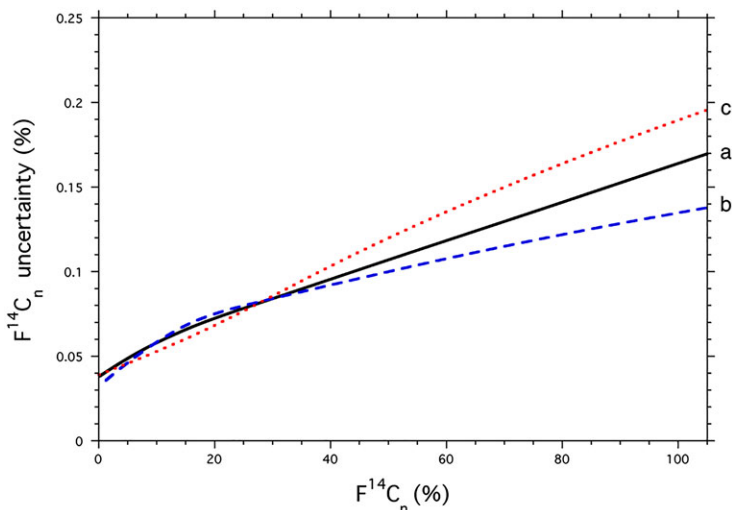


Figure 3 Long-term (≈ 1 year) standard deviation of cat. 2 (b) and cat. 3 (c) secondary references versus the Fraction Modern (^{14}C content). For comparison, the calculated ^{14}C measurement uncertainty is shown (averaged $\mathbf{dF}^{14}\text{C}_n$ for last one and a half year) (a, black solid line). b. Spread of category (cat.) 2 secondary references (blue dashed line). These secondary references are Rommenh oller gas, IAEA-C8, IAEA-C7 and GS-51 (Groningen Standard, cane sugar). c. Spread of cat. 3 secondary references (red dotted line). These secondary references are background wood, IAEA-C8, IAEA-C7 and GS-51 (Groningen Standard, cane sugar).

preparation duplicates (cat. 3). For bgw, we indeed get a result comparable to the other cat. 3 materials. For background collagen, however, we observe the highest χ_{red}^2 of all materials: 1.7. The reason for this high, and significant value is not well understood and more data are needed as this value is calculated from only 12 data points.

Uncertainty estimates leads to maximum ages measurable in a system. The standard deviation of background collagen implies that the minimum $\mathbf{F}^{14}\text{C}_n$ distinguishable from background values, is two times 0.05%, so 0.1% on graphite samples (Stuiver and Polach 1977; van der Plicht and Hogg 2006). This corresponds to 55,000 years BP. However, as the absolute $\mathbf{F}^{14}\text{C}_n$ values for the background wood and the background collagen are 0.23–0.25%, even though these materials are known to be of infinite age, we never report ages older than corresponding to these activities (48,000 years BP) (van der Plicht and Palstra 2016).

Figure 3 visualizes the average results over the first full year of measurements from the secondary references. The calculated ^{14}C measurement uncertainty (from Eq. 2, $\mathbf{dF}^{14}\text{C}_n$, averaged data of last one and a half year), is shown in black again (line a). The long-term spread of our secondary CO_2 references (cat. 2), contains contributions from the actual ^{14}C measurement and the graphitization, but no further variability due to individual combustion, and $\mathbf{dF}^{14}\text{C}_n$ is expected to be a good estimate of their uncertainty (see above). The standard deviation of the cat. 2 references is shown in blue (line b). The realized spread in the long-term measurements is lower than line a, and, for higher $\mathbf{F}^{14}\text{C}_n$ values, even approaching the Poisson statistics uncertainty (which is shown as the gray line (a) in Figure 2).

The standard deviation of cat. 3 secondary references is shown by line c (red) in Figure 3. Six combustion references from Table 1 (cat. 3, not oxalic acid) are used to construct this line. The displayed spread at zero percent F^{14}C_n , is the spread of bgw. In practice there are many more bgw than bgc measurements, therefore bgc measurements are disregarded. The GS-35 measurements are disregarded in Figure 3 as well, as this material is a carbonate and therefore not combusted.

The differences between the external standard deviation of the secondary references, that have only a graphitization step (cat. 2), and the secondary references, that have both a graphitization and a combustion step (cat. 3), are significant (line b and c in Figure 3) and amount to $\approx 30\%$. As expected, the combustion process contributes to a greater spread in the data. χ_{red}^2 is the ratio of the squared external standard deviation and the squared ^{14}C measurement uncertainty ($\text{dF}^{14}\text{C}_n^2$). The average χ_{red}^2 for cat. 3 references is 1.3. To represent the uncertainty in these combusted references, our calculated (internal errors) dF^{14}C_n (line a) need to be multiplied by ≈ 1.15 (which is the square root of the average χ_{red}^2 of the appropriate samples in Table 1).

The next source of information about contributions to the expanded uncertainty comes from sample duplicates in various phases of the process (see Figure 1). Data from air samples from our atmospheric station at Lutjewad (the station is described in Van der Laan-Luijkx et al. 2010) provide information about quantification of the graphitization uncertainty (cat. 2 duplicates). CO_2 is dissolved in an alkaline solution during the sampling of atmospheric air and for ^{14}C measurement, this CO_2 is released again by using acid. The released CO_2 fraction is divided into three equal portions, which makes these samples graphitization triplicates. The average spread in these triplicates is 0.18%, which compares favorably to the calculated ^{14}C measurement uncertainty of 0.16%.

The unknown sample CO_2 preparation duplicates (cat. 3) provide information about the third major contribution to the expanded uncertainty, the contribution from combustion. For the CO_2 preparation duplicates of unknown samples, the standard deviation of the distribution of f_σ , $\sigma(f_\sigma)$, is on average 1.4, meaning the dF^{14}C_n had to be increased by 40% to match the spread (See Table 2). As mentioned before, when comparing secondary references from cat. 3 (c from graph 3) dF^{14}C_n should be enlarged by approximately 15%. We attribute this large difference between sample duplicates and secondary references to inhomogeneity in the samples and connected to this inhomogeneity the chance of success in homogeneously removing exogenous contaminants in the samples. It illustrates that the CO_2 preparation from inhomogeneous samples will contribute much more to the expanded uncertainty than the CO_2 preparation from pure materials. If unaccounted for, this would lead to “dark uncertainty” in our results, which would come to light for example in intercomparisons. Our attempt to quantify this extra uncertainty is thus by randomly re-measuring samples on a regular basis.

All described secondary references do not require a chemical pretreatment, because the references are pure materials. Therefore, the fourth major contribution to the expanded uncertainty can only be quantified by sample pretreatment duplicates (cat. 4) (among which the known sample VIRI F, horse bone). Monitoring the pretreatment duplicates (cat. 4) revealed that the expanded uncertainty for unknown random samples (bones, charcoal, wood) is dF^{14}C_n , increased by a factor of 1.6 (see Table 2). Splitting the pretreatment duplicates into different materials would be desirable but is impeded by the low number of measurements.

Table 2 Comparison of the observed differences of two ^{14}C measurements for various duplicates from unknown samples, with the expected uncertainty. (The expected uncertainty is the quadratic sum of the individual measurement uncertainties.) The spread of the ratio f_σ , (Eq. 3), $\sigma(f_\sigma)$, indicates in how far the observed uncertainty deviates from our calculated one. If $\sigma(f_\sigma)$ is larger than 1, the calculated ^{14}C measurement uncertainties ($\mathbf{dF}^{14}\text{C}_n$) are too low, and some “dark uncertainty” is present. For random solid materials, like bone and charcoal and wood samples, $\sigma(f_\sigma)$ is on average 1.6, meaning $\mathbf{dF}^{14}\text{C}_n$ had to be increased by 60% to match the spread. $\mathbf{dF}^{14}\text{C}_n$ from more homogeneous materials like α -cellulose had to be increased by 40% to match the spread. For a cat. 2 duplicate this increase is 10%.

	Pretreatment duplicates (Cat. 4)	Pretreatment duplicates (Cat. 4)	Pretreatment duplicates (Cat. 4)	CO ₂ preparation duplicates (Cat. 3)	Graphitization duplicates (Cat. 2)
Material	Solids (wood, bones, charcoal)	α -Cellulose	VIRI F Horse bone	CO ₂ combustion and carbonates	Lutjewad (CO ₂)
Number of pairs	25	38	9	27	130
$\sigma(f_\sigma)$	1.6	1.4	1.2	1.4	1.1

Interestingly, all the VIRI F Horse bone pretreatment duplicates (cat. 4) (paired in two) show a smaller $\sigma(f_\sigma)$ of 1.2, suggesting a lower expanded uncertainty. The VIRI F measurements were paired in duplicates to allow us a direct comparison with duplicates we do randomly on unknown samples. Since a data set of nine pairs is small, we also calculate f_σ by comparing the standard deviation of all VIRI F measurements with the average calculated $\mathbf{dF}^{14}\text{C}_n$. The result is similar to the paired method (as it should). The reason for this smaller $\sigma(f_\sigma)$ is unknown; maybe this bone sample was less contaminated compared to other samples. The sample duplicates from tree-rings where the fraction α -cellulose was extracted (cat. 4), also showed a smaller $\sigma(f_\sigma)$ of 1.4 compared to various other unknown samples (of which $\sigma(f_\sigma)$ is 1.6 as mentioned earlier). The reason for this is probably that during pretreatment most of the contaminants and other naturally occurring compounds were removed, as the extracted α -cellulose is a more uniform biopolymer. As $\sigma(f_\sigma)$ of cat. 4 from more homogeneous materials hardly differs with the $\sigma(f_\sigma)$ of cat. 3 duplicates, it indicates that the additional sample handling from the chemical pretreatment does not contribute to the increase of the expanded uncertainty. The increase of $\sigma(f_\sigma)$ for cat. 4 pretreatment duplicate samples of 1.6 is, therefore probably merely due to inhomogeneity of the sample material and, perhaps related to that, the variability of the success rate of the chemical pretreatment. The expanded uncertainties for various processes and samples, using the calculated $\mathbf{dF}^{14}\text{C}_n$ and the $\sigma(f_\sigma)$ factors are shown in Table 3. The use of $\sigma(f_\sigma)$ factors is in fact an error multiplier approach (Scott et al. 2007). Of course, it would be preferable to also quantify and include all other uncertainty sources, but as these are next to impossible to determine, this pragmatic solution is acceptable. Still, multiplication factors should be as close to unity as possible, otherwise the uncertainty analysis apparently fails to include the major sources of uncertainty.

Table 3 Minimum final reported uncertainties (expanded uncertainties) for various processes and samples, using the calculated $dF^{14}\text{C}_n$ and the multiplication factors from Table 2. The results are shown in Fraction Modern (%) and in ^{14}C years (years BP). These uncertainties are valid for single measurements. The last column shows samples, which undergo the full pretreatment (chemical preparation, combustion, graphitization and ^{14}C measurement). Columns to the left represent fewer steps in the sample handling process. As an example, when a bone sample is pretreated, combusted, graphitized, has its radiocarbon activity measured, and the date is calculated to be 1800 years BP, the minimum achievable uncertainty is 23 ^{14}C years BP. On the other hand, a contemporary atmospheric CO_2 sample ($F^{14}\text{C}_n = 100\%$) is reported with an uncertainty of 0.18%, which is the equivalent of 14 ^{14}C years BP.

$F^{14}\text{C}_n$ % (years BP)	$dF^{14}\text{C}_n$ % (years BP)	Graphitization % (years BP) ($1.1 \times dF^{14}\text{C}_n$ (%))	+ Combustion % (years BP) ($1.4 \times dF^{14}\text{C}_n$ (%))	+ Chemical pretreatment* % (years BP) ($1.6 \times dF^{14}\text{C}_n$ (%))
0.1 (>50,000)	0.04 (2600)	0.04 (2800)	0.05 (3400)	0.06 (3800)
15 (15,000)	0.07 (36)	0.07 (39)	0.09 (50)	0.11 (57)
30 (9,700)	0.08 (23)	0.09 (25)	0.12 (31)	0.13 (36)
50 (5,600)	0.11 (17)	0.12 (19)	0.15 (24)	0.17 (27)
80 (1,800)	0.14 (14)	0.16 (16)	0.20 (20)	0.23 (23)
100 (0)	0.16 (13)	0.18 (14)	0.23 (18)	0.26 (21)
134 (-2,351)	0.20 (12)	0.22 (13)	0.28 (17)	0.32 (19)

* α -cellulose pretreatment is an exception (see Table 2), for which we can use the data in the 4th column.

During the early phases of measuring with the MICADAS, a realistic determination of the expanded uncertainty was not possible due to limited available data and so to encompass uncertainties arising from various sources, we used a provisional multiplication factor of 1.5 for the ^{14}C measurement uncertainty for every sample. Our present assessment showed that this “educated guess” was quite appropriate.

Attempts to Reduce the Expanded Uncertainty

The main goal of this study was to determine the expanded uncertainty of a ^{14}C measurement and to quantify the contributions to this uncertainty. As an extension of the project while performing and analyzing all the measurements that were described in this paper, we also tried to reduce this expanded uncertainty. One obvious possibility is to improve the counting statistics by increasing the measurement time. The Poisson uncertainty will of course gradually decrease by collecting more counts during a longer measurement time, but at the same time other uncertainty sources in the measurement (calibration stability, ^{13}C signal stability) might increase, and after a certain point will outbalance the Poisson gain.

We tried this out by performing an experiment where we measured a batch with a net measurement time exceeding 10,000 seconds per sample. The batch contained (among others) seven Oxalic Acid II references and eight IAEA-C8 references, all produced from CO_2 from their respective bulk materials (thus cat. 2).

In this experiment with very long measurement times, we calculated the measurement standard deviation after 1700, 2300, 3500 until 10,000 seconds measurement time per sample. These

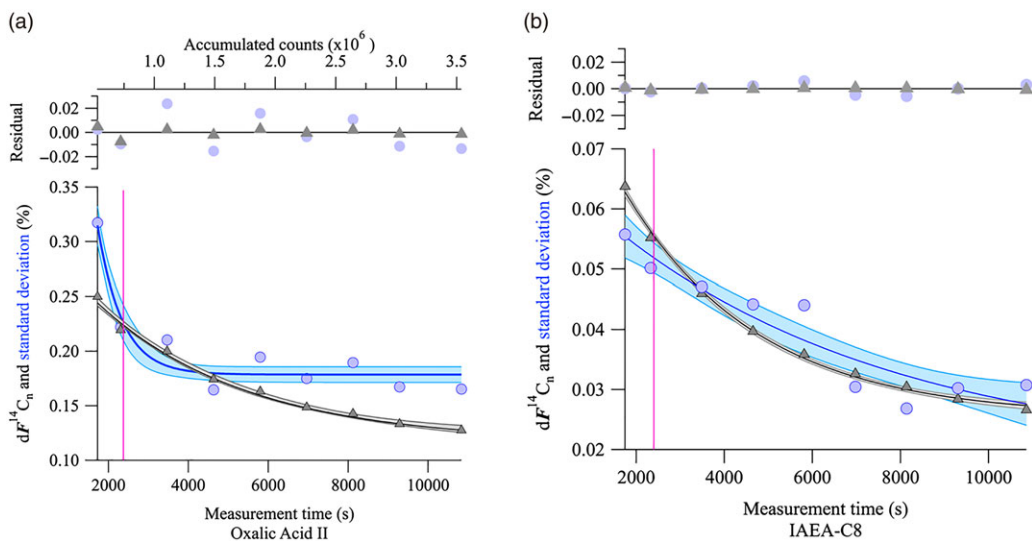


Figure 4 References measured with a very long ^{14}C measurement time in order to determine the optimal measurement time for a sample. The standard deviation from seven Oxalic Acid II references (a) and eight IAEA-C8 (b) (all cat. 2, blue) versus a measurement time of more than 10,000 seconds (3.5×10^6 accumulated counts for Oxalic Acid II). The calculated ^{14}C measurement uncertainty ($dF^{14}\text{C}_n$) is displayed in black. The shaded area around the standard deviation (blue) and $dF^{14}\text{C}_n$ (black) is the confidence band (1σ , 68%). The pink line shows the routine measurement time of 2400 seconds.

results provided important insights into the optimum measurement time, but as the same cathodes were analyzed for the comparison, the data are not independent of each other.

The main conclusion is that in contrast to the Poisson uncertainty from individual references that obviously decreased with measurement time, the observed spread in the Oxalic Acid II references (a) did not significantly improve for measurement times over 4000 seconds. The random contribution to the spread due to the different graphitization reactions, causing spread in the ^{13}C stability is a plausible reason. In all cases, a longer measurement time obviously leads to a decrease in (the calculated) $dF^{14}\text{C}_n$. On the other hand, the standard deviation of IAEA-C8 (b) still does decrease with time. However, the gain in years from 2400 sec measurement time (magenta line) to 4000 sec is only 5 years (BP) for IAEA-C8 (for Oxalic Acid II only 2 years BP). This improvement is hardly ever worth the investment of doubling of the measurement time. Therefore, our routine measurement time of 2400 seconds is optimally chosen.

We conducted several independent batches with various measurement times as well. The calculated ^{14}C measurement uncertainty ($dF^{14}\text{C}_n$) for those independent measurements also revealed no significant improvement after 2400 seconds. All measurements in this paper were conducted at 2400 seconds.

Table 1 (comparison of secondary references from cat. 2 and 3) and Figure 3 (line b and c) showed the influence from the combustion process to the expanded uncertainty. A possible, although unlikely, cause might be a memory effect from one combustion to another in the combustion set up and cryogenic collection system. Experiments with blank (background material) combustions after Oxalic Acid II references did not show a significant memory

effect in the combustion set up. Still, to be absolutely sure, we recently started to combust an empty tin capsule before every individual combustion to see whether we could more definitively understand and even determine the size of this potential memory effect. The extra oxygen pulse should reduce possible leftover material that was not completely converted into CO₂. Further data collection is needed to determine if the additional blank combustions improve the final reported uncertainty. The other contamination source could be the cryogenic collection system. This system is more than 20 years old, and the constant freezing and heating of the glass may have introduced micro cracks that function as active adsorption spots for the exchange of CO₂ and thus increase the memory from sample to sample. Regular refreshing of the glass system may reduce this contamination and hence, the contribution to the expanded uncertainty. The replacement of all the glass components of the cryogenic collection system will be a major operation; therefore, further research will be needed to see, whether this will indeed reduce the contribution to the expanded uncertainty.

The contribution from chemical pretreatment is visible in Table 2 for cat. 4 sample duplicates, but also for cat. 3 sample duplicates, where the multiplication factor is considerably larger than for pure substances (Table 1, ~1.15). A large part is apparently due to the inhomogeneity or intractable contamination of the sample and therefore it is not easily possible to reduce this contribution. Obviously, running duplicate (or multiple) samples would help to some extent, as inhomogeneities and the varying success of removing contaminants would average out. However, for the vast majority of samples this is not an option due to the increased costs involved (and sometimes the need for more material). An experiment on the automation of pretreatment, as a means of standardizing the process, and reducing random errors will be investigated in the near future.

DISCUSSIONS AND RECOMMENDATIONS

The new generation of high-yield accelerator mass spectrometers delivers very small measurement uncertainties, due to higher count rates. The reported uncertainty in the final outcome, however, must be the expanded uncertainty, a firm measure of the spread that can be expected in case of multiple analyses of the same material, by one or more laboratories.

When performing sample duplicates in the same laboratory, and monitoring the spread in the data, it is already clearly apparent that the ¹⁴C measurement uncertainty is too small to serve as the reported uncertainty. Therefore, in the field of ¹⁴C measurements it is very common to use a multiplying factor of the ¹⁴C measurement error for the uncertainty in the final outcome (Scott et al. 2007). This is in line with the “dark uncertainty” concept, well known from intercomparison of results between different laboratories (Koepke et al. 2017; Merktas et al. 2019).

However, we deemed it necessary to achieve a better and more thorough understanding of the build-up of uncertainty in the whole chain from chemical pretreatment to the final measurement, such that we can report a reliable expanded uncertainty in our publications, and to our customers. To report such a reliable uncertainty is obviously very important for participation in round robin tests and other intercomparisons.

For laboratories in the field of ¹⁴C, we recommend measuring full duplicates on all the kinds of samples that are normally measured; noting, of course, that most laboratories already have such protocols implemented. A nice example of such a practice was described in a recently published work by Sookdeo et al. (2019), where in addition to process duplicates, the

authors also emphasize including process backgrounds for high-quality measurements. The results of the measured duplicates give a very good insight in the quality of the measurements. This protocol of measuring duplicates is especially useful when participating in intercomparisons. In the optimal case, where all participants estimate their expanded uncertainty well, the “dark uncertainty” would be minimal. Weighting the data for averaging with the expanded uncertainty would then make sense. Therefore, we recommend that in future intercomparisons a report should be added on the basis used for the stated uncertainty.

For quality improvement and thus reduced expanded uncertainty in the final outcome, it is recommended to measure secondary references in the various steps of the process from CO₂ preparation up to the actual ¹⁴C measurement. This gives insights into where improvements in quality can be achieved and which steps are limiting the further reduction of the expanded uncertainty. Using homogeneous materials has the advantage that the effects are clear. On the other hand, one should not claim the results of such homogeneous secondary references as valid for the real samples, as our work has shown that “real samples” show larger spread, most likely due to variability in the success of removing contamination in the pretreatment process.

CONCLUSIONS

After detailed uncertainty analysis using measurements from secondary references and sample duplicates during the first year and a half of MICADAS operation in Groningen, we are confident that we can report an expanded uncertainty that is representative of the real uncertainty in our final ¹⁴C measurements. This expanded uncertainty incorporates contributions from the chemical pretreatment, the CO₂ preparation, the graphitization and the ¹⁴C measurement. We systematically evaluate the contributions to the ¹⁴C measurement uncertainty. This uncertainty is the basis for the expanded (final) uncertainty. As our work has shown, for samples, like bone, wood or charcoal, which undergo chemical pretreatment, combustion, graphitization, and ¹⁴C measurement, the calculated ¹⁴C measurement uncertainty must be multiplied by factor 1.6 to get the expanded uncertainty.

For more homogeneous samples, like a one-year tree ring sample where α -cellulose is collected, this multiplication factor is 1.4. Similarly, for CO₂ samples collected from air, this factor is 1.1.

The achievement of this present work is twofold: first that we have checked our carefully calculated ¹⁴C measurement uncertainty and shown that it is a reliable basis for reporting the final uncertainty, and second that we have established evidence-based multiplication factors for the various sample types. Future ring tests will benefit from this method of uncertainty estimate.

ACKNOWLEDGMENTS

We would like to thank the staff of the Centre for Isotope Research in Groningen, Dicky van Zonneveld, Henk Been, Marc Bleeker, Fsaha Ghebru, Berthe Verstappen-Dumoulin, Sven de Bruijn, Regina Linker, Henk Jansen, Margot Kuitens, Janette Spriensma, and Patricia Wietzes for their work. Without them we could not have performed this study (or run a ¹⁴C laboratory at all). We would like to thank Antonio Possolo (NIST) for pointing us to papers providing statistical background. We would also like to thank the two anonymous referees for their careful review and valuable suggestions for improvement.

REFERENCES

- Aerts-Bijma AT, Meijer HAJ, van der Plicht J. 1997. AMS sample handling in Groningen. *Nuclear Instruments and Methods in Physics Research B* 123(1–4):221–225.
- Arnold JR, Libby WF. 1949. Age determinations by radiocarbon content: checks with samples of known age. *Science* 110(2869):678–680.
- Birge RT. 1932. The calculation of errors by the method of least squares. *Physical Review* 40(2):207–227.
- Cook AC, Wadsworth J, Southon JR, van der Merwe NJ. 2003. AMS Radiocarbon dating of rusty iron. *Journal of Archaeological Science* 30(1):95–101.
- Cook GT, Higham TFG, Naysmith P, Brock F, Freeman SPHT, Bayliss A. 2012. Assessment of infinite-age bones from the Upper Thames Valley, UK, as ^{14}C background standards. *Radiocarbon* 54(3–4):845–853.
- Damon PE, Donahue DJ, Gore BH, Hatheway AL, Jull AJT, Linick, TW, et al. 1989. Radiocarbon dating of the Shroud of Turin. *Nature* 337(6208):611–615.
- Dee M, Wengrow D, Shortland A, Stevenson A, Brock F, Girdland Flink L, Bronk Ramsey C. 2013. An absolute chronology for early Egypt using radiocarbon dating and Bayesian statistical modelling. *Proceedings of the Royal Society A* 469:0395.
- Dee M, Palstra SWL, Aerts-Bijma A Th, Bleeker MO, Brijn de S, Ghebru F, Jansen HG, Kuitens M, Paul D, Richie RR, Spriensma JJ, Scifo A, Zonneveld van D, Verstappen-Dumoulin BMAA, Wietzes-Land P, Meijer HAJ. 2020. Radiocarbon dating at Groningen: new and updated chemical pretreatment procedures. *Radiocarbon* 62(1):63–74.
- De Rooij M, van der Plicht J, Meijer HAJ. 2010. Porous iron pellets for AMS ^{14}C analysis of small samples down to ultra-microscale size (10–25 μgC). *Nuclear Instruments and Methods in Physics Research B* 268(7–8):947–951.
- De Vries HL, Barendsen GW. 1952. A new technique for the measurement of age by radiocarbon. *Physica XVIII* (8–9):652.
- De Vries HL, Barendsen GW. 1954. Measurements of age by the carbon-14 technique. *Nature* 174:1138–1141.
- Gott dang A, Mous DJW, van der Plicht J. 1995. The HVEE ^{14}C system at Groningen. *Radiocarbon* 37(2):649–656.
- Hedges REM, Law IA, Bronk CR, Housley RA. 1989. The Oxford accelerator mass spectrometry facility: technical developments in routine dating. *Archaeometry* 31(2):99–113.
- Joint Committee for Guides in Metrology 2008. Evaluation of measurement data—guide to the expression of uncertainty in measurement (Gum 1995 with minor corrections). *JCGM* 100:2008.
- Koepke A, Lafarge T, Possolo A, Toman B. 2017. Consensus building for interlaboratory studies, key comparisons, and meta-analysis. *Metrologia* 54(3):S34–S62.
- Lafarge T, Possolo A. 2015. The NIST uncertainty machine. *NCSLI Measure Journal of Measurement Science* 10(3):20–27.
- Lanting JN, Aerts-Bijma AT, van der Plicht J. 2001. Dating of cremated bones. *Radiocarbon* 43(2A):249–254.
- Le Clercq M, van der Plicht J, Groning M. 1998. New ^{14}C reference materials with activities of 15 and 50 pMC. *Radiocarbon* 40(1):295–297.
- Libby WF, Anderson EC, Arnold JR. 1949. Age determination by radiocarbon content: World-wide assay of natural radiocarbon. *Science* 109(2827):227–228.
- Linick TW, Damon PE, Donahue DJ, Jull AJT. 1989. Accelerator mass spectrometry: the new revolution in radiocarbon dating. *Quaternary International* 1:1–6.
- Merkatas C, Toman B, Possolo A, Schlamming S. 2019. Shades of dark uncertainty and consensus value for Newtonian constant of gravitation. *Metrologia* 56:054001:1–16.
- Meijer HAJ, Pertuisot MH, van der Plicht J. 2006. High-accuracy ^{14}C measurements for atmospheric CO_2 samples by AMS. *Radiocarbon* 48(3):355–372.
- Meijer HAJ. 2009. Stable isotope quality assurance using the “Calibrated IRMS” strategy. *Isotopes in Environmental and Health Studies* 45(2):150–163.
- Mook WG, Streurman HJ. 1983. Physical and chemical aspects of radiocarbon dating. *Proceedings of the First International Symposium ^{14}C and archaeology, Groningen* 1(8):31–55.
- NIST. 2019. The NIST uncertainty machine. <https://uncertainty.nist.gov>, last accessed 20/03/09.
- Paul D, Been HA, Aerts-Bijma AT, Meijer HAJ. 2016. Contamination on AMS sample targets by modern carbon is inevitable. *Radiocarbon* 58(02):407–418.
- Rasmussen KL, van der Plicht J, Doudna G, Nielsen F, Hojrup P, Stenby EH, Pedersen CT. 2009. The effects of possible contamination on the radiocarbon dating of the Dead Sea Scrolls II: empirical methods to remove castor oil and suggestions for redating. *Radiocarbon* 51(3):1005–1022.
- Reimer PJ, Brown TA, Reimer RW. 2004. Discussion: reporting and calibration of post-bomb ^{14}C data. *Radiocarbon* 46(3):1299–1304.
- Rukhin AL. 2009. Weighted means statistics in interlaboratory studies. *Metrologia* 46(3):323–331.
- Scott EM, Cook GT, Naysmith P. 2007. Error and uncertainty in radiocarbon measurements. *Radiocarbon* 49(2):427–440.
- Scott EM, Cook GT, Naysmith P. 2010. A Report on Phase 2 of the Fifth International Radiocarbon

- Intercomparison (VIRI). *Radiocarbon* 52(3): 846–858.
- Sookdeo A, Kromer B, Büntgen U, Friedrich M, Friedrich R, Helle G, et al. 2019. Quality dating: a well-defined protocol implemented at ETH for high-precision ^{14}C -dates tested on late glacial wood. *Radiocarbon* 33:1–9.
- Stuiver M, Polach HA. 1977. Discussion: reporting of ^{14}C data. *Radiocarbon* 19(3):355–363.
- Synal H-A, Stocker M, Suter M. 2007. MICADAS: a new compact radiocarbon AMS system. *Nuclear Instruments and Methods in Physics Research B* 259(1):7–13.
- van der Laan-Luijkx IT, Karstens U, Steinbach J, Gerbig C, Sirignano C, Neubert REM, van der Laan S, Meijer HAJ. 2010. CO_2 , $\delta\text{O}_2/\text{N}_2$ and APO: observations from the Lutjewad, Mace Head and F3 platform flask sampling network. *Atmospheric Chemistry and Physics* 10(21): 10691–10704.
- van der Plicht J, Hogg A. 2006. A note on reporting radiocarbon. *Quaternary Geochronology* 1:237–240.
- van der Plicht J, Palstra SWL. 2016. Radiocarbon and mammoth bones: What's in a date. *Quaternary International* 406:246–251.
- Wacker L, Christl M, Synal HA. 2010a. Bats: a new tool for AMS data reduction. *Nuclear Instruments and Methods in Physics Research B* 268(7–8): 976–979.
- Wacker L, Bonani G, Friedrich M, Hajdas I, Kromer B, Němec, M, et al. 2010. MICADAS: routine and high-precision radiocarbon dating. *Radiocarbon* 52(2):252–262.

APPENDIX 1

Table 4 Comparison of the calculation of the ^{14}C measurement uncertainty (Eq. 2, $\mathbf{dF}^{14}\text{C}_n$) using daily values and four-monthly-averaged values for the error in the mean of the Oxalic Acid II references and for the spread in the backgrounds.

Name	Material	Cat.	N	$\mathbf{F}^{14}\text{C}_n$ (%)	$\mathbf{dF}^{14}\text{C}_n$ (%)	σ_{ext} (%)	χ_{red}^2
RH (before July 18) Batch values	CO ₂ gas	2	90	0.15	0.007	0.026	8.2
RH (after July 18) 4-monthly average	CO ₂ gas	2	116	0.17	0.035	0.032	0.8
Bgw (before July 18) Batch values	Wood	3	149	0.22	0.011	0.036	6.4
Bgw (after July 18) 4-monthly average	Wood	3	136	0.23	0.035	0.039	1.2

Comparison of ^{14}C uncertainty calculation with MICADAS batch averages and calculations with four-monthly averages.

During the first months of operation of the MICADAS, from September 2017 until July 2018, the calculation of the ^{14}C measurement uncertainty (Eq. 2, $\mathbf{dF}^{14}\text{C}_n$) was performed using the same methodology as with our former HV accelerator, namely based on daily or batch values for the error in the mean of the Oxalic Acid II references and the daily or batch values for the spread in the backgrounds. Table 4 provides the summary statistics for two background materials, namely background wood and Rommenhöller gas, for both situations (the data for the period after July 2018 is also shown in Table 1). From the data, it is apparent that the calculation of the ^{14}C measurement uncertainty before July 2018 for background materials gives an uncertainty that underestimates the spread of the data. The internal error is too small (by a factor of 3). After July 2018 the uncertainty calculation is based on four-monthly averaged values, instead of the daily (batch) values. The four-monthly average value of the background is more reliable as a typical background for the samples, and the spread in these data of the last four months, is more realistic in comparison with the spread from only four backgrounds in one batch. In other words, the variability between batches is larger than the variability within one batch.

The actual external standard deviations of the measurements are approximately the same for both periods, as they should be, as the actual measurements were performed in the same way before and after that date.

APPENDIX 2

Calculations of weighted means, internal and external errors and χ_{red}^2 .

To calculate the weighted average (\bar{x}_w) of data from the secondary references, the weighting factor (w_i is given by the reciprocal of the variance (that is the square of the calculated ^{14}C measurement uncertainty, $\mathbf{dF}^{14}\text{C}_n$ of the ^{14}C measurement value (n is number of measurements and i represents a single measurement).

$$w_i = \frac{1}{\sigma_i^2} \quad (4)$$

The weighted mean in this case is (where x_i is a single measurement):

$$\bar{x}_w = \frac{\sum_{i=1}^n w_i \cdot x_i}{\sum_{i=1}^n w_i} \quad (5)$$

The external standard deviation is calculated according to:

$$\sigma_{ext} = \sqrt{\left(\left(\frac{\sum_{i=1}^n w_i \cdot x_i^2}{\sum_{i=1}^n w_i} - \bar{x}_w^2 \right) * \left(\frac{n}{n-1} \right) \right)} \quad (6)$$

The external error in the mean $\sigma_{m,ext}$ is calculated by dividing the external standard deviation by the square root of the number of measurements:

$$\sigma_{m,ext} = \sqrt{\left(\left(\frac{\sum_{i=1}^n w_i \cdot x_i^2}{\sum_{i=1}^n w_i} - \bar{x}_w^2 \right) * \left(\frac{1}{n-1} \right) \right)}, \quad \text{or} \quad \sigma_{m,ext} = \frac{\sigma_{ext}}{\sqrt{n}} \quad (7)$$

The internal error in the mean $\sigma_{m,int}$, (also called standard error of the weighted mean (SEWM)) is given by:

$$\sigma_{m,int} = \sqrt{\frac{1}{\sum_{i=1}^n w_i}} \quad (8)$$

Their ratio:

$$\chi_{red}^2 = \left(\frac{\sigma_{m,ext}}{\sigma_{m,int}} \right)^2 = \sum_{i=1}^n (w_i * (x_i - \bar{x}_w)^2) * \frac{1}{n-1} \quad (9)$$