Radiocarbon, Vol 00, Nr 00, 2023, p 1–8 DOI:10.1017/RDC.2023.64
Selected Papers from the 24th Radiocarbon and 10th Radiocarbon & Archaeology International Conferences, Zurich, Switzerland, 11–16 Sept. 2022

PRELIMINARY RESULTS FROM GLASGOW INTERNATIONAL RADIOCARBON INTERCOMPARISON

E M Scott[1]* • P Naysmith[2] • E Dunbar[2]

[1]School of Mathematics and Statistics, University of Glasgow, Glasgow, UK
[2]SUERC Radiocarbon Laboratory, University of Glasgow, East Kilbride, UK

ABSTRACT. GIRI (Glasgow International Radiocarbon Intercomparison) was designed to meet a number of objectives, including to provide an independent assessment of the analytical quality of the laboratory/measurement and an opportunity for a laboratory to participate and improve (if needed). The principles in the design of GIRI were to provide the following: (a) a series of unrelated individual samples, spanning the dating age range, (b) linked samples to earlier intercomparisons to allow traceability, (c) known age samples, to allow independent accuracy checks, (d) a small number of duplicates, to allow independent estimation of laboratory uncertainty, and (e) two categories of samples—bulk and individual—to support laboratory investigation of variability. All of the GIRI samples are natural (wood, peat, and grain), some are known age, and overall their age spans approx. >40,000 years BP to modern. The complete list of sample materials includes humic acid, whalebone, grain, single ring dendro-dated samples, dendro-dated wood samples spanning a number of rings (e.g., 10 rings), background and near background samples of bone and wood. We present an overview of the results received and preliminary consensus values for the samples supporting a more in-depth evaluation of laboratory performance and variability.

KEYWORDS: accuracy, consensus, intercomparison.

INTRODUCTION

Like every complex measurement process, radiocarbon (^{14}C) measurements have an uncertainty, often described as error, which reflects the variability that would be observed were we able to make repeated measurements on that sample (sometimes also described as precision). Calculation of the uncertainty includes contributions from a variety of different sources and may be done differently in individual laboratories. Accuracy of the measurement is another aspect of quality, referring to "closeness to the true age." Experimental verification of the accuracy and precision of measurements is an important aspect of routine quality assurance, and part of that assurance comes from participation in intercomparisons or proficiency trials. Given the complexity of the processes, the diversity of materials being dated, and ongoing technical developments, there has been a sustained effort by the radiocarbon community, based in part on a series of intercomparisons, to deliver experimental verification of measurement quality. Such intercomparisons have been influential in allowing full quantification of the uncertainties and accuracy on the reported ages, accounting for all laboratory processes as well as analysis exploring individual laboratory performance (Scott et al. 2017, 2018, 2022).

Quality assurance and quality control processes are critical and intertwined with the concept of measurement accuracy and precision and uncertainty quantification. Some of the key metrological concepts include bias, accuracy and precision, repeatability and reproducibility, most of which can be evaluated through participation in a carefully designed intercomparison. These concepts are critical in ensuring a well calibrated measurement system and part of that also comes from benchmarking of measurements, since it is highly likely that scientific studies will require demonstration of the comparability of results from different laboratories. The ^{14}C

*Corresponding author. Email: marian.scott@glasgow.ac.uk

community has undertaken a wide-scale, far reaching and evolving program of global intercomparisons, to the benefit of laboratories and users alike (Scott et al. 2018). Each intercomparison has been designed to meet a number of objectives, including the most fundamental one, to provide an independent assessment of the analytical quality of the laboratory/measurement and an opportunity for a laboratory to participate and improve (if needed).

At the core of any intercomparison lies the samples, and a further important objective of ^{14}C intercomparisons is the creation of a set of recognized reference materials which are well characterized. Reference materials can be used regularly and allow a laboratory to fully explore its own processes and procedures. Such reference materials, whose ^{14}C activity is known empirically, may be used on a daily basis and considered as working standards. Within the ^{14}C community, laboratories have often created their own working standards e.g., humic acids, cellulose or barley mash (Naysmith et al. 2019; Tripney et al. 2023 in this proceedings), but one advantage of *secondary* reference materials created as part of an intercomparison is that their activities have been verified in many laboratories, and includes a wide range of routinely dated materials.

GIRI Design and Samples

The core principles in the design of GIRI were to use natural samples, which spanned the typical dating age range, to use duplicates, and to include samples with known age and to have samples that tie the results to previous intercomparisons. Additional criteria for selecting samples were that they should be available in sufficient quantity to allow an archive to be kept and be homogeneous in ^{14}C. A significant practical challenge comes from the need to provide sufficient material for 80+ laboratories and ideally to have sufficient material remaining for future use. These considerations mean that material must be sourced in bulk, which raises concerns about sample homogeneity, especially when routine measurements are now being made on a few milligrams of material.

SAMPLES AND THEIR DESCRIPTIONS

Building on our previous work and experiences, all of the GIRI samples are natural (wood, peat and grain), some are known age and overall, their ages span approx. >40,000 years BP to modern. In the case of peat, the samples have been pretreated to humic acid, and we also include a cellulose sample, but other samples require pretreatment. The complete list of sample materials for AMS laboratories are presented in Table 1. A smaller set of samples were prepared for radiometric laboratories and presented in Table 2. Both tables provide relevant information about the samples including their previous use and published consensus values or dendro-dates. We are immensely grateful to all the sample providers.

Additional Sample Information

For each of samples I and J, identical wood slivers were sampled from blocks of sub fossil kauri (*Agathis australis*). Each of these blocks comes from trees that were excavated from Waipu (W603) and Kai Iwi Lakes (FIN09) in Northland, New Zealand. The W603 block (W603_1101-1120) covers 20-years while the FIN09 block (FIN09_4681-4720) covers 40-years. The 20-year and 40-year blocks of time were chosen from tree-ring constrained ^{14}C sequences to lie on a ^{14}C plateau. The ^{14}C ages lie in the range 20–30 ka BP (W603) and

https://doi.org/10.1017/RDC.2023.64 Published online by Cambridge University Press

Table 1 Sample descriptions for AMS laboratories.

Sample	Previous use	Previous consensus value
Sample A: barley mash	Sample A: also used as TIRI A	Sample A: 116.35 pMC
Sample B: humic acid from St Bees	Sample B: also used as VIRI U	Sample B:11778 BP
Sample C: barley mash	Sample C: from 2017	
Sample D: humic acid	Sample D: new preparation	
Sample E: dendro-dated wood	Sample E: Belfast Q11473 AD1476-1478 (4th of a set of rings used in SIRI)	
Sample F: barley mash	Sample F: from 2019	
Sample G: dendro-dated wood	Sample G: 10 rings Q7780 (3220-3211BC) linked to TIRI B, FIRI D	Sample G: 4503BP. 4508BP
Sample H and HB: dendro-dated wood; single ring	Sample H: 315BC Sample HB: 318BC	
Sample I: Kauri wood	Sample I: 20-30ka BP	
Sample J: Kauri wood	Sample J: 35-45ka BP	
Sample K: whalebone	Sample K: TIRI L	Sample K: 12788 BP
Sample L and LB: dendro-dated wood; single ring	Sample L: 250BC Sample LB: 251BC	
Sample M: dendro-dated cellulose	Sample M: FIRI optional and VIRI O, dendro-dated to 1880-1820AD	Sample M: 125BP
Sample N: Kauri wood	Sample N: background	
Sample O: humic acid	Sample O: St Bees, VIRI U, FIRI E	Sample O: 11780BP, 11778BP
Sample P: dendro-dated wood	Sample P: FIRI H	Sample P: 2232 BP
Sample Q: dendro-dated wood	Sample Q: AD 1586	Sample Q: AD 1586

Table 2 Sample description for radiometric laboratories.

Description	Previous use	Consensus value
Sample A: Kauri wood	Sample A: GIRI N above	Sample A: background
Sample B: mammoth bone	Sample B:	Sample B:
Sample C: barley mash	Sample C: GIRI A	Sample Q:
Sample D: barley mash	Sample D: GIRI F	Sample A:
Sample E: charcoal	Sample E: VIRI P	Sample E: 1747BP
Sample F: whole peat	Sample F: FIRI E	Sample Q: 11780BP

35–45 ka BP (FIN09) (Hogg et al. 2021). For the AMS intercomparison samples, rings were initially cut out of radial strips located parallel to original tree ring measurements. The extracted blocks were dimensionalized further to retain a precise number of rings and to ensure the AMS samples aligned to known locations within both the W603 and FIN09 series, allowing traceability back to the original high precision dating. Wafers 2–3 mm thick covering the entire 20-year and 40-year ring sequences were then cut from the dimensionalized blocks and broken down into matchstick-sized material using a scalpel. Each matchstick contained the entire 20-year/40-year sequence from W603/FIN09, respectively. The samples were provided by Drew Lorrey and Alan Hogg (Hogg, personal communications).

For samples H (HB) and L (LB) ca. 2 g samples of single years (early and late wood) from a sequoiadendron giganteum tree, sample LIN 01 were provided. To extract this material, the entire period from 345 BC to 246 BC was dissected, and individual ring samples created. The samples were provided by Charlotte Pearson. Four rings were selected and described in Table 1.

For sample E, this was the 4th of a set of 4 single rings taken from a floor joist from a house (Medieval Period) provided by Queens University Belfast. Three single rings samples had been used previously in SIRI (F, G, and H with dendro dates of AD 1487, 1479, and 1475).

Sample Q (MAG-C63) is a single ring from a beam removed from the Great Tower at St Mary Magdalen College, Oxford (51.75°N, 1.24°W) during repair works in the 1960s. The entire beam is 6.1 m long, by 0.3 m square and weighs over a metric tonne. Whole rings (earlywood and latewood) have been dissected by professional dendrochronologists, each sample being split across the ring so that it contains roughly equal amounts of earlywood and latewood. Each sample weighs approximately 50 mg. The samples were provided by Alex Bayliss.

Samples B and O were humic samples that had previously been used in VIRI, FIRI. Sample D was a new peat sample collected in 2020. Well-humified peat samples were and air dried and sieved through a 3-mm mesh to remove large root fragments, oven dried and mixed by several passages through a grinding mill. To obtain the humic acid fraction, the peat was subjected to successive digestions in 2M potassium hydroxide and the alkali-soluble humic acid extracts were removed by filtration and combined. The humic acid was then precipitated from the bulk solution by adjusting to pH3 with sulphuric acid. The resulting humic acid slurry was separated by centrifugation, re-bulked, washed several times with distilled water and oven dried at 70°C. The resultant granules were washed with warm distilled water, filtered, and dried. The final product was again subjected to physical mixing (Naysmith et al. 2019).

https://doi.org/10.1017/RDC.2023.64 Published online by Cambridge University Press

Table 3 Summary statistics for barley mash samples.

Sample	n	Mean	Stdev	Q1	Median	Q3
A	99	1.1643	0.0075	1.1619	1.1652	1.1679
C	98	1.0227	0.0072	1.0199	1.0225	1.0247
F	96	1.0162	0.0117	1.0133	1.0156	1.0176

Table 4 Summary statistics for humic acid samples.

Sample	n	Mean	Stdev	Q1	Median	Q3
D	98	3826	70.5	3796	3818	3847
B	98	11813	110	11775	11810	11840
O	106	11826	153	11770	11818	11847

PRELIMINARY RESULTS

GIRI was delayed from 2019 and samples were finally dispatched in 2021, with results received in 2022. More than 70 laboratories received samples, the vast majority being AMS facilities. By the deadline, 55 AMS laboratories returned results, and for some samples, we have more than 100 measurements, made up of sometimes more than 8 replicate measurements per laboratory. A very small number of radiometric laboratories participated with a small set of samples. Preliminary results are presented for the different sample groups—barley mash, humic acid, dendro-dated wood and for previously used samples with published consensus values.

Barley Mash

Table 3 provides the summary statistics for the three barley mash samples. GIRI A had been used previously in TIRI and the published consensus value was 116.35 pmC. Each table provides the number of measurements (n), the mean, median and standard deviation (stdev) and the lower and upper quartile (Q_1 and Q_3).

Humic Acid

Three humic acid samples were provided, two (B and O) of which were duplicates and used previously as VIRI U and FIRI E with published consensus values of 11780 BP. Table 4 provides the summaries for these 3 samples.

Dendro-Dated Wood Samples

A total of 6 dendro-dated wood samples were provided, 3 were single ring (H (HB), L (LB) and Q, while the others were 2, 10 and 20 rings respectively. H (and HB) and L (and LB) are contiguous single rings since there was insufficient material to provide every laboratory with a unique sample. The dendro dates for each sample are shown in Table 5 below.

The summary statistics for the dendro-dated wood samples are given in Table 6.

https://doi.org/10.1017/RDC.2023.64 Published online by Cambridge University Press

Table 5 Dendro dates for GIRI wood.

Description	Dendro dates
GIRI E	AD 1476–AD 1478
GIRI G	3220–3211 BC
GIRI H and HB	315 BC and 318 BC
GIRI L and LB	250 BC and 251 BC
GIRI P	313–294 BC
GIRI Q	AD 1586

Table 6 Summary statistics for dendro-dated samples.

Sample	n	Mean	Stdev	Q1	Median	Q3
E	113	378	48.56	364.5	380	395
G	112	4523	48.5	4506	4526	4553
H	50	2208	44	2187	2212	2232
HB	43	2200	27	2187	2201	2214
L	48	2241	58	2214	2235	2251
LB	47	2239	35	2218	2237	2255
P	112	2227	62	2215	2235	2255
Q	83	336	46	321	336	355

Table 7 Summary statistics for remaining samples.

Sample	n	Mean	Stdev	Q1	Median	Q3
I	95	23644	168	23551	23660	23750
J	99	38571	886	38179	38640	39062
K	94	12780	114	12745	12792	12844
M	91	132	32	115	131	151
N	104	0.002146	0.00181	0.0013	0.0018	0.0025

Remaining Samples

For the remaining samples of Kauri wood, bone and cellulose, summaries are provided in Table 7. Sample N is a close to background sample so results are given in Fm rather than age.

Relationship to Previous Consensus Values

For those samples used in previous intercomparisons, we have published consensus values, and so are able to consider the relationship of these new results to the previous consensus. The summaries of the differences are shown in Table 8. The mean differences from previous consensus value, when tested, show small but statistically significant ($p<0.05$) results for samples B, G, M, and O. In these cases, the new results are significantly different from the previously published consensus value (although the differences are small, less than 50 years). Figure 1 examines the z-scores for the same samples, where the z-score is defined as the difference between the reported age and the consensus value standardized for the quoted uncertainty. Values lying between +/– 3 are generally considered as acceptable

https://doi.org/10.1017/RDC.2023.64 Published online by Cambridge University Press

Table 8 Preliminary consensus values and summaries of the differences from the consensus values. Confidence intervals (CI) are provided for those differences which are statistically significantly different to previous consensus values.

Sample	Consensus	Mean	stdev	Q1	Median	Q3	95% CI
A*	1.1635	0.00081	0.00755	–0.0016	0.0017	0.0044	
B	11778	35.4	110	–2.8	32	62	13.4–57.4
G	4503	19.7	48	3	23	50	10.6–28.7
K	12780	–8	114	–42.8	3.6	56	
M	125	7.3	32	–10	6.2	26	0.7–13.9
O	11778	47.9	153	–8.3	–40	69.2	18.4–77.3
P	2232	–4.7	62	–17	3.5	23	

*A is reported in pMC, all other ages in years BP.

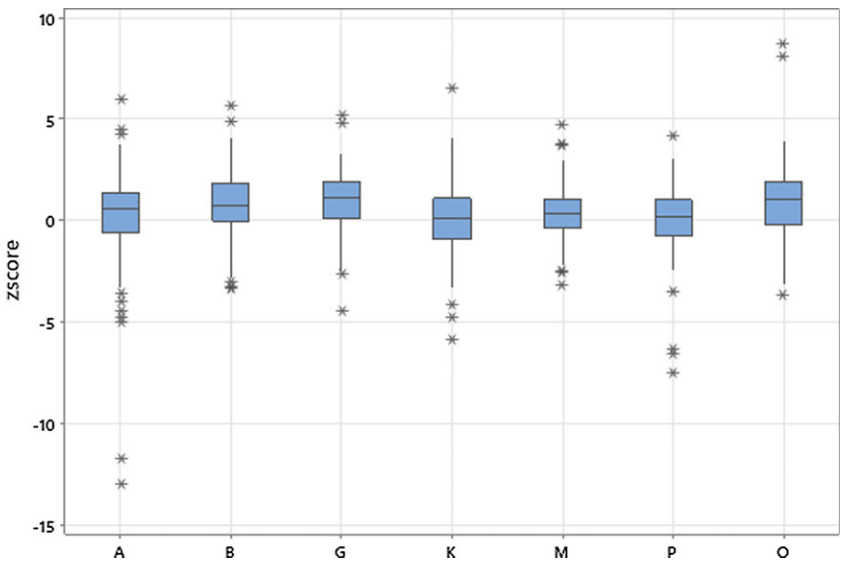

Figure 1 Boxplots of z-scores.

(Thompson 2022). Figure 3 shows that the majority of z-scores lie within acceptable bounds, but that there are still some results which lie outside them, that the boxplots are not perfectly centred on zero, with some indication of small offsets. Further analysis will examine these results in more detail and refine as appropriate the consensus values.

CONCLUSIONS AND DISCUSSION

These first preliminary results from GIRI provide an overview of the samples, rather than individual laboratory performance. This is intentional since it will allow laboratories to examine their results in depth. The provenance and known ages of the samples are also provided for reference. The results have shown that these new results are predominantly consistent with the previously reported consensus values, and where there are differences, they are small in magnitude. They have shown that performance across the modern to 15,000 years range is broadly consistent based on z-score analysis. A small number of outlying results can be

https://doi.org/10.1017/RDC.2023.64 Published online by Cambridge University Press

seen in the results (again not unexpected and in keeping with previous studies). Ongoing analysis is now exploring the laboratory-based performance and exploring also the full age range including the older samples (>25,000 BP).

REFERENCES

Hogg A, Lorrey AM, Turney CS, Palmer JG, Boswijk G, Fenwick P. 2021. Advances and limitations in establishing a contiguous high-resolution atmospheric radiocarbon record derived from subfossil kauri tree rings for the interval 60–27 cal kyr BP. Quaternary Geochronology. doi: 10.1016/j.quageo.2021.101251

Naysmith P, Scott EM, Dunbar E, Cook GT. 2019. Humics—their history in the radiocarbon inter-comparisons studies. Radiocarbon 61(5): 1413–1422.

Scott EM, Naysmith P, Cook GT. 2018. Why do we need ^{14}C inter-comparisons?: The Glasgow ^{14}C inter-comparison series, a reflection over 30 years. Quaternary Geochronology 43:72–82.

Scott EM, Naysmith P, Cook G. 2022. What lies behind radiocarbon intercomparisons and the design of the new intercomparison, GIRI? Nuclear Instruments and Methods in Physics Research Section B 525:62–66. doi: 10.1016/j.nimb.2022.06.015

Scott EM, Naysmith P, Cook GT. 2017. Should archaeologists care about ^{14}C inter-comparisons? Why? A summary report on SIRI. Radiocarbon 59(5):1589–1596. doi: 10.1017/RDC.2017.12

Thompson M. 2022. Assigned values in the GeoPT Proficiency Testing Scheme. Geostandards and Geoanalytical Research 46(1):37–41.

Tripney B, Dunbar E, Naysmith P, Scott EM. 2023. Routine quality assurance in the SUERC radiocarbon laboratory. Radiocarbon. doi: 10.1017/RDC.2023.45.

https://doi.org/10.1017/RDC.2023.64 Published online by Cambridge University Press