1. Introduction

In the 1980's, with the growth in the number of radiocarbon laboratories, including new AMS laboratories, a proposal was made for a formal quality assurance program to be introduced (Long and Kalin, 1990)). This could take the form of a laboratory inter-comparison or proficiency trial as set out in Thompson (2006), where a selection of samples is chosen to be used in the inter-comparison and all working laboratories are invited to take part to check their own individual performance. Following from early work, a community programme of inter-comparisons began (Scott et al, 2018). The samples selected to be used in these programmes were natural and routinely dated materials, many of which had the potential to become internationally recognised reference materials. The main criteria for selecting samples were that they should:1) Be of archaeological and/or geological interest, 2) Cover the broad spectrum of laboratory experience (age, sample type, etc), 3) Satisfy rigorous homogeneity testing, 4) Be known age if possible. In later inter-comparisons and when the community became predominantly AMS dominated, we added the criterion that they should be available in sufficient quantity that they could be archived and thus available as reference materials.

In all the inter-comparison trials, when deciding on which samples should be included in the program, one of the main requirements is homogeneity. Homogeneity of the reference material refers to the variability observed in true replicates (sub-units). Therefore, for some materials, this may require the material to be physically and/or chemically homogenized. For many of the samples selected, homogenisation has been a major undertaking. In the early inter-comparisons, we required a large amount of material per sample due to the requirements of the radiometric laboratories. This has been reduced considerably over the years as the advances in radiocarbon dating has allowed for smaller samples to be dated. Most of the samples are checked for homogeneity before being sent out, with the exception of bone samples as they come from one source and should be homogeneous and tree rings dated using dendrochronology. Wood samples do not typically require homogeneity testing especially if they have been dendro dated (we often sought tree-rings sequences that lay on a plateau of the C-14 curve- see Scott et al, this volume), or where the blocks were relatively short (e.g. 10-40 ring blocks). Peat samples have also been used in the inter-comparisons, most commonly provided as the humic acid fraction which since precipitated from solution, will be homogeneous.

Over the last 30 years, we have used a number of different design structures for the inter-comparisons, to set new challenges for the laboratories in all aspects of radiocarbon dating, and to provide information concerning variation and the components of variation in the dates and their associated errors, e.g. due to different pre-treatment methods, laboratory backgrounds. We also have occasionally provided duplicate samples to check variation within a laboratory. Statistical analysis of the results focusses on summarising laboratory performance (bias, variability) and in providing consensus values for each sample. For complete descriptions of all the trials see Scott et al (2018).

In this short paper, we concentrate on the peat samples used in the inter-comparison studies. We will describe the pre-treatment method used to extract the humic acid fraction, and the connections between the different inter-comparisons where the same material has been used on several occasions (sometimes as raw peat or alternatively as humic acid). Where appropriate, updated consensus values will be provided. Finally, we provide an illustration of the benefits that an individual laboratory can gain from a well characterised inter-comparison sample.

2. Samples and studies

2.1 The different peat samples and their pre-treatment

There are 6 different peat samples that have been used since the International Collaborative Study (ICS) (Cook et al, 1990, Aitchison et al, 1990, Harkness et al, 1989), described below:

2.1.1 Hekla Peat sample was collected from Svinavatn, North Iceland in August 1991 by Professor A. Dugmore andDr A. Newton (University of Edinburgh). It is associated with a tephra layer corresponding to one of the largest eruptions of the Hekla volcano. The tephra layer corresponding to the eruption was exposed over approximately a 2 m length, at a depth of approximately 1 m below the overlying vegetation. The tephra layer was then removed and a 1 cm thick layer of peat lying beneath the tephra was extracted. This sample was dried and ground to a fine powder before being mixed thoroughly and was sent out for the laboratories to test their pre-treatment methods. This material was used in TIRI (TIRI D)

2.1.2 Ellanmore Peat During 1991 a bulk sample comprising ca 10 kg of peat was cut from a freshly cleaned exposure by Dr A.M. Hall and Dr D.D.Harkness. The Ellanmore peat occurs as a 50 cm thick horizon intercalated with glacial diamicts and is exposed in a stream bank section of the Reisgill Burn, Ellanmore, Caithness, Scotland. This sample was split and humic acid was extracted from one half of the sample and the remainder was air dried, ground and mixed thoroughly. Both fractions were dated in the TIRI study as TIRI E and H.

2.1.3 Icelandic Peat This sample was collected in August 1991 from Solheimajokull, South Iceland by Professor A Dugmore and Dr A Newton. The peat was taken from a thin section between two tephra layers, at approximately 1 m below the underlying vegetation layer. The whole peat was dried, ground to a fine powder and thoroughly mixed. This sample was used in TIRI as sample M.

2.1.4 St Bees Head Peat. This sample is from a costal cliff deposit at St Bees head in Cumbria, north-western England (NGR NX 9472 1196), which had been exposed to erosion. The apparently well humified felted peat deposit is approximately 0.5 m thick and is overlain by several meters of lacustrine material of Holocene age that is largely mineral in nature. Approximately 20 kg of peat were collected and taken back to the laboratory for testing but on the basis of discrepancies between humic and humin results the sample was not used. The site was subsequently re-sampled and approximately 30 kg from a slightly different elevation was collected. The second sample was used in the FIRI and VIRI studies as FIRI E and VIRI U.

2.1.5 Siberian Peat. This peat sample comes from a peat deposit in Siberia, the sample was provided by Professor Kh Arslanov from St Petersburg. This sample is close to background and was used in VIRI (VIRI J).

2.1.6 Letham Moss Peat Central Scotland. The peat sample came from Letham Moss, Central Scotland, where a well-humified sample was collected from freshly cut exposures (about 20 cm depth to provide limited age variation). This sample was first used in the International Collaborative Study (ICS) (1988), where both the raw peat and humic acid were used. The humic samples were then subsequently used as VIRI T and SIRI N.

2.2 Peat Pre-treatment Method

A brief description of the humic acid extraction method used is given below.

Whole Peat/Humic acid extraction. Well-humified peat samples were collected from freshly cut exposures (about 20 cm depth to provide limited age variation). The raw samples were 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. If whole peat and humic acid were required then half of the product was retained in this form and mixed further. To obtain the humic acid fraction, the remainder 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 to constant weight. The final product was again subjected to physical mixing. The alkali-insoluble (humin) residues from the extraction were also recovered and retained for future reference (Harkness et al., 1989).

2.3 The inter-comparison studies

A brief summary of the studies where peat and humic acid were used is given below (full details can be found in Scott et al, 2018).

ICS (Cook et al, 1990, Harkness et al, 1989, Scott et al, 1989, 1990, 1991): In this three stage trial, one of the goals was the quantitative assessment of variability. In Stage 1, laboratory prepared carbonate and benzene samples were distributed. In Stage 2 we provided homogenised, pre-treated humic acid samples (in duplicate) and in Stage 3, the raw peat was also provided in duplicate. In this study only 8 of the 50+ laboratories were AMS.

Following the ICS study, **TIRI** (the Third International Radiocarbon Inter-comparison) (Scott et al., 1992, Scott 2003) was organised and commenced in 1991. TIRI was a single stage study, with a core set of samples and optional samples also provided (Ellanmore humic acid was a core sample, the whole peat was optional). The next study in the sequence was **FIRI** (the Fourth International Radiocarbon Inter-comparison) which was completed in 2000. This again was a single stage study, but again providing core (humic acid) and optional (whole peat) samples, while some samples were provided in duplicate (this fact was blind to participants). The Fifth International Radiocarbon Inter-comparison (**VIRI)** commenced in 2004 with the third and final suite of samples including humic acid. More than 70 laboratories participated, of which over half were AMS. The most recently completed exercise is **SIRI** (the Sixth International Radiocarbon Inter-comparison), which commenced in 2013 and was completed in 2016. Again, this was a single stage trial, designed predominantly for AMS facilities.

2.3 Statistical analysis

Our approach has been first to assess the distribution of results, identifying any outliers, before proceeding to evaluate laboratory performance (in terms of bias and error multipliers - both internal and external (Aitchison et al, 1989) and to quantify the consensus value for each material (including uncertainty) (Rozanski et al, 1992) and the updated consensus value approach tailored to AMS measurement (Scott et al, 2018). Differences, as a result of

pre-treatment, have been examined using relatively simple methods including ANOVA (Scott, 2003).

3. Results and Discussion

Table 1 presents the combined reference information for all peat samples including their codes and published consensus values. For those samples used as optional in TIRI and FIRI, we have simply reported the mean and standard error since typically there were insufficient numbers of results to confirm a consensus value. Subsequently, we focus on sub-analysis, possible only due to the study designs we have introduced.

Table 1: Summary values for all peat and humic acid samples

Study	Sample Code	Sample Type	Consensus age ((^{14}C BP) with 1 sigma uncertainty
ICS	Stage 2	Letham Moss peat - humic acid fraction (provided in duplicate)	3368 ± 114*
ICS	Stage 3	Letham Moss - whole peat (provided in duplicate)	3379 ± 94*
TIRI	Sample D	Hekla peat, Iceland	3810 ± 7
	Sample E	Ellanmore peat - humic acid fraction	11129 ± 12
	Sample H	Ellanmore whole peat (*optional)*	11152 ± 23
	Sample M	Icelandic whole peat	1682 ± 15
FIRI	Sample E	St Bees peat - humic acid fraction	11780 ± 7
	Sample M	St Bees - whole peat (*optional*)	11139±49*
VIRI	Sample J	Siberian peat - humic acid fraction	43231 ± 141
	Sample T	Letham Moss peat - humic acid fraction (**ICS 2 and 3**)	3360 ± 16
	Sample U	St Bees peat - humic acid fraction **(FIRI E)**	11778 ± 18
SIRI	Sample N	Letham Moss peat - humic acid fraction (**VIRI T and ICS 2 and 3**)	3369 ± 4

* This value is not the uncertainty on the consensus value but rather the standard deviation.

3.1 Humic and whole peat paired samples

Starting from the ICS, a hierarchical structure using both the whole peat and humic acid, as well as duplicate samples, allowed the analysis to explore any differences between the two materials and any increased variability due to the introduction of individual laboratory pre-treatments, as well as the individual laboratory reproducibility. The results from ICS (stages

2 and 3) showed that there was a small, but significant increase in the variability of results between stages 2 and 3 (Table 2). In addition, in both stages 2 and 3, the samples were provided in duplicate which allowed the investigation of the 'internal error multiplier' (Scott et al, 1991). There was generally good agreement between the duplicate pairs, indicating that the results were reproducible. The consensus values are in good agreement between the two stages, but perhaps not surprising, given the time period, the variability in the results (expressed as a standard deviation) is more in keeping with the typical laboratory quoted errors which were on average 50+ years.

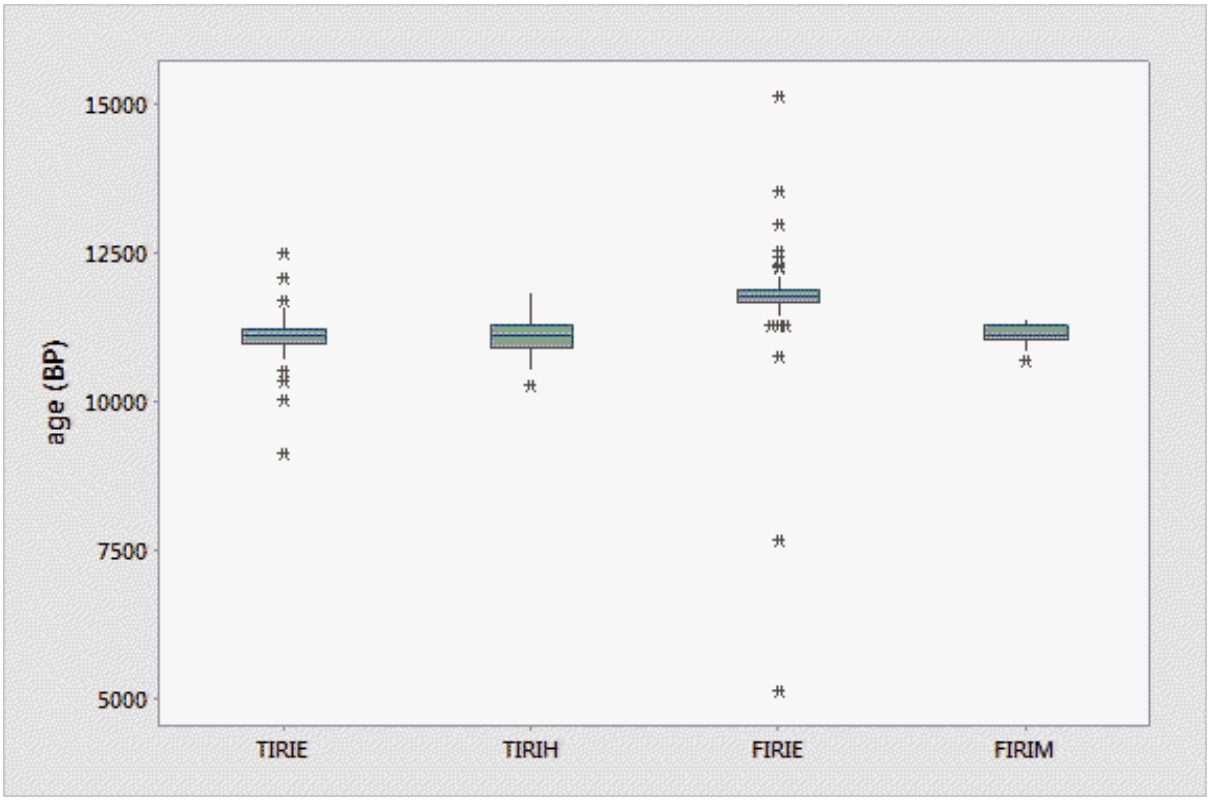

Figure 1 Boxplot of linked whole peat and humic acid samples

Table 2 Summary statistics for the radiocarbon ages for whole peat and humic acid samples

	number	mean	St dev	min	median	max
ICS2 (Letham Moss humic)	42	3366	167	2683	3391	3678
ICS3 (Letham Moss whole peat)	77	3384	104	3140	3375	3780
TIRI E (Ellanmore humic)	72	11076	461	9153	11115	12510
TIRI H	35	11115	311	10280	11300	11860

(Ellanmore whole peat)						
FIRI E (St Bees humic)	136	11746	781	5160	11780	15150
FIRI M (St Bees whole peat)	15	11139	191	10710	11120	11413

In TIRI, we made use of whole peat and humic acid from Ellanmore to investigate the contribution of laboratory pre-treatment methods to the overall variability in the results. The results are shown on Figure 1 and Table 2. The mean values in Table 2 and consensus values for whole peat and humic acid in Table 1 are the same within the uncertainty on the values and there is little evidence of increased variability in the whole peat results which might have been expected due to different laboratory pre-treatments. Similarly, in FIRI, we made use of whole peat and humic acid from St Bees. Results can be seen in Table 2 and Figure 1. Figure 1 also shows evidence of a relatively small number of outliers which have been removed before the robust calculation of the consensus values.

3.2 Humic acid samples used in multiple inter-comparisons

Consistently, we have designed linking samples across the inter-comparisons, one of which is humic acid.

3.2.1 St Bees peat

Humic acid from the St Bees peat sample has been used in both FIRI and VIRI, as well as Letham Moss used in ICS, VIRI and SIRI. This allows us to study the reproducibility of the results over time. Table 3 show the consistency of the consensus values (the same within their uncertainty) and of the distribution of results. The combined set of results can then be used to refine the published consensus value, for the St Bees Head humic acid of 11779 ± 10 BP.

Table 3 Summary statistics for humic acid and whole peat samples from St Bees and Letham Moss used in more than 1 inter-comparison

	number	mean	St dev	min	median	max
St Bees						
FIRI E	136	11746	781	5160	11780	15150
VIRI U	65	11758	168	11010	11774	12220
Letham Moss						
ICS2	42	3366	167	2683	3391	3678
ICS3	77	3384	104	3140	3375	3780

VIRI T	34	3343	50	3244	3349	3465
SIRI N	66	3371	51	3215	3368	3536

However, the peat sample that has been used most is the humic acid sample from Letham Moss peat which has now been used in three inter-comparison studies. A large quantity of peat was sampled from Letham moss and then split into two subsamples. The first subsample was used in the ICS study and the second was extracted a few years later and then used in the VIRI and SIRI intercomparisons

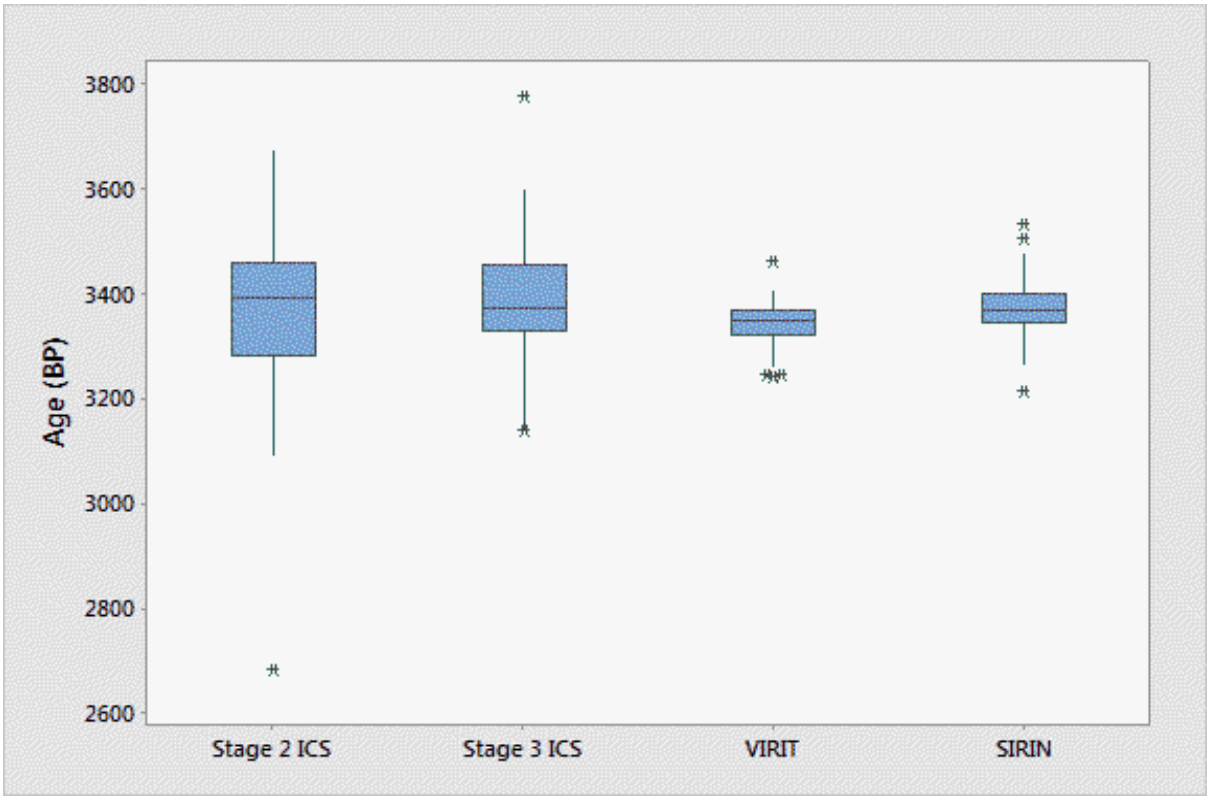

Figure 2 Boxplots for humic acid results from Letham Moss peat

Consensus values have been recalculated for the Letham Moss humic acid sample using the published procedures, resulting in a new consensus value of 3369 ± 5.

These two new consensus values are based on a very large number of results from individual laboratories submitted during ICS, TIRI, VIRI and SIRI, thus allowing us to improve the precision with which the consensus value is characterised where appropriate.

3.3 Laboratory benefits of a well characterised reference value

Reliable, precise and accurate ^{14}C age measurements are essential. Such measurements require traceability to international standards (such as Ox1 and Ox2) and to reference materials whose activities are estimated but accompanied by associated uncertainty statements. A reference material is typically a natural material and its activity (age) must therefore be characterised on the basis of measurements from many laboratories. Such characterisation is only possible when agreement between laboratories performing the measurements can be demonstrated, usually in an inter-laboratory comparison.

While inter-comparisons are only snap-shots in time, one significant benefit from a well-designed study, using appropriate materials (available in sufficient quantities), is to allow individual laboratories, in the future, to use well characterised materials as routine reference materials or secondary standards. The Letham Moss sample is one such reference material which has been used in the SUERC laboratory since 2010. In every routine AMS wheel there are 13 humic acid targets, used to calculate the minimum error associated with the wheel, which is then applied to all the measured unknowns in that wheel. This humic acid is used as a quality control tool to spot any small problems in sample combustion and graphitization. Table 3 shows the SUERC data summary since 2010. Figure 3 shows boxplots of all the results, including the inter-comparison as well as the SUERC results over time.

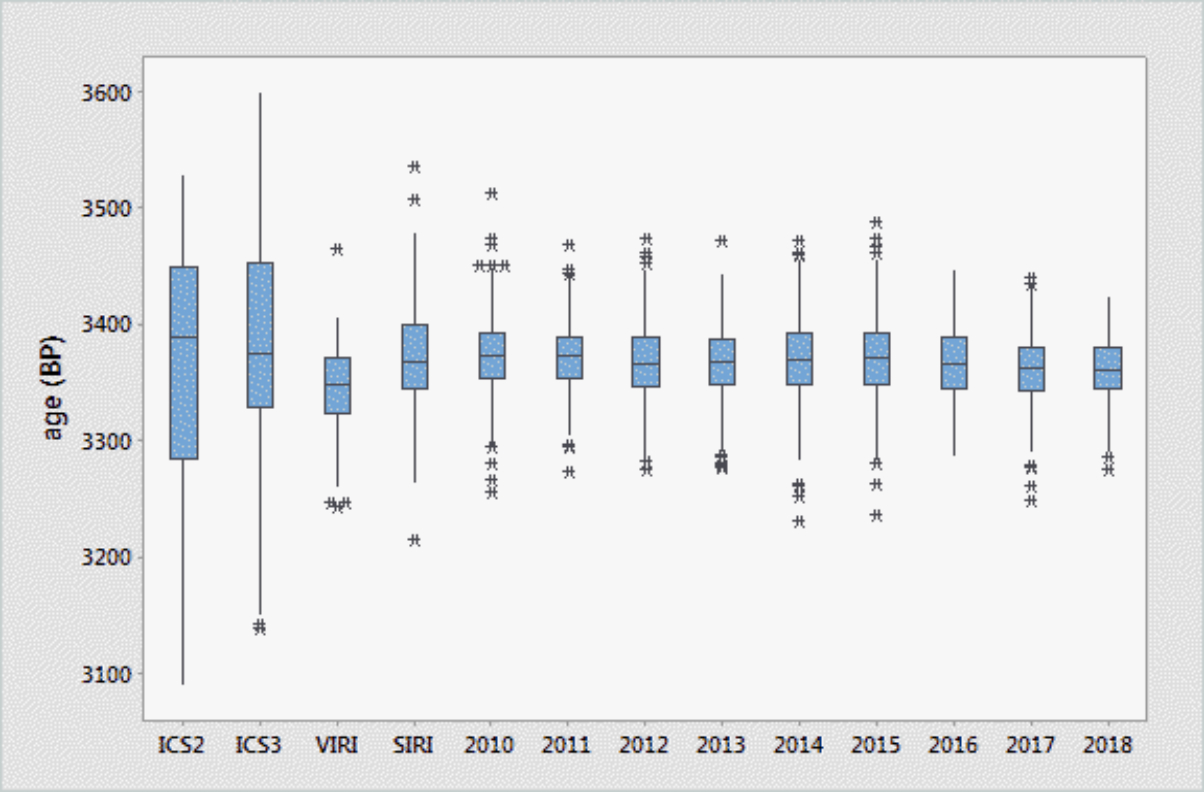

Figure 3. Boxplot of all Letham Moss results

year	2010	2011	2012	2013	2014	2015	2016	2017	2018
n	340	402	463	459	502	480	419	550	198
mean	3374	3372	3369	3368	3371	3372	3366	3362	3362
St dev	33	29	32	33	34	35	31	28	26
median	3374	3373	3367	3368	3370	3371	3366	3362	3361

Table 3: Summary ^{14}C age data from the SUERC laboratory for the Letham Moss humic acid sample

The SUERC laboratory, routinely prepared two targets from each combustion and then ran them randomly through the 32 graphite units in the laboratory. However in 2018 it was decided to run just one target from each combustion, to explore what effect this had on the scatter in the results. Table 3 shows that the standard deviation for 2018 so far is slightly reduced but is this due to the change to single combustions? The laboratory will continue to monitor the results for the rest of the year before making any firm conclusions regarding this change in working practice.

Figure 4 shows the boxplots of z-scores (Scott et al, 2018), which are the scaled (to the quoted error) deviations from the updated consensus value for each of the inter-comparisons and the SUERC results. Z-scores are used frequently to flag up measurements which are more than 2(3) quoted errors away from the consensus value, for further investigation but can also be used to monitor long-term performance. Figure 4 shows the stability of the results since ICS, and the proportionally small number of measurements that lie outwith the ± 2 limits.

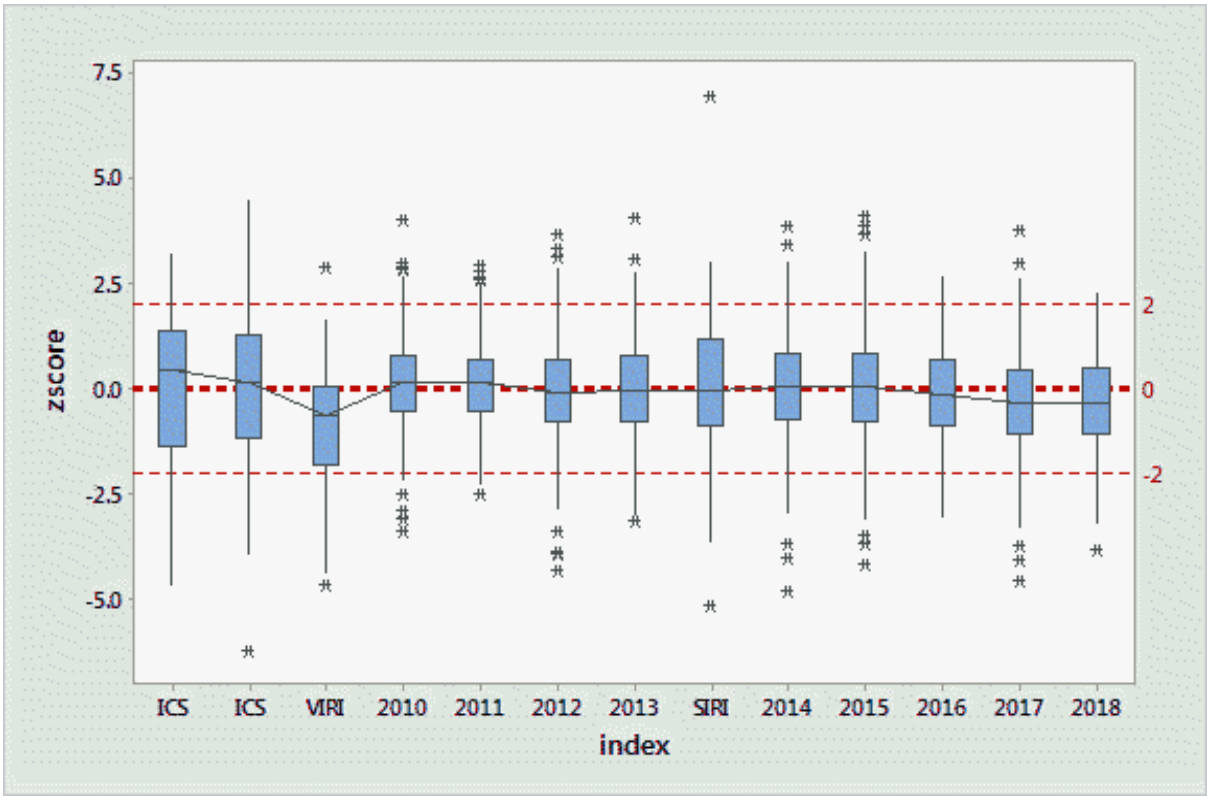

Figure 4; z-scores of all Letham moss results

4. Conclusions

Considerable care and attention has been given to the design of ^{14}C inter-comparisons, specifically in terms of the samples being used, and one more recent development (now a practical one) concerns the creation and testing of natural reference materials spanning the ^{14}C age span. This paper has drawn together our experiences when using peat and more specifically the humic acid fraction and has provided updated consensus values for two humic acid samples that have been widely used. The Letham Moss humic acid sample now has a consensus value of 3369 ± 5 y BP while the value for the St Bees humic acid sample is 11779 ± 10 y BP. The former has been used to great effect in the SUERC laboratory to monitor both accuracy and precision of the analyses over almost a decade and demonstrates the usefulness of such samples.

Acknowledgements

The extensive programmes of work have been funded from a number of sources including UK research councils (EPSRC and NERC), EU (FP4), NATO, Historic England and Historic

Environment Scotland. Colleagues have been immensely generous in their contributions of the materials which are such an important part of the inter-comparison. Finally, a huge thanks to the participating laboratories that have contributed several thousand ^{14}C age measurements over the years.

References

Aitchison, TC, Scott, EM, Harkness, DD, Baxter, MS and Cook, GT (1990). Report on Stage 3 of the International Collaborative Program. Radiocarbon, 32(3) , 271-278.

Cook G T, Harkness D D, Miller B F, Scott E M, Baxter M S, Aitchison T C (1990) International collaborative study: structuring and sample preparation. Radiocarbon, 32(3), 267-270

Harkness, D.D., Cook, G.T., Miller, B.F., Scott, E.M. and Baxter, M.S. (1989). Design and preparation of samples for the international collaborative study. Radiocarbon 31(3): 407-413.

Long A, Kalin R M, 1990. A suggested quality assurance protocol for radiocarbon dating laboratories. Radiocarbon 32(3), 329-334.

Rozanski K, Stichler W, Gonfiantini R, Scott E M, Beukens R P, Kromer B, Van der Plicht J, 1992. The IAEA ^{14}C intercomparison exercise 1990. *Radiocarbon* 34(3), 506-519.

Scott, E.M., Aitchison, T.C., Harkness, D.D., Baxter, M.S., Cook, G.T. (1989) An interim progress report on stages 1 and 2 of the international collaborative programs. Radiocarbon, 31, 414-421.

Scott E M, Aitchison T C, Harkness D D, Cook G T, Baxter M S, 1990. An overview of all three stages of the international radiocarbon intercomparison. *Radiocarbon* 32(3), 309-319.

Scott, E.M., Harkness, D.D., Cook, G.T., Aitchison, T.C., Baxter, M.S. (1991) Future quality assurance in 14C dating, Quaternary Proceedings, 1, 1-4.

Scott E M, Harkness D D, Miller B F, Cook G T, Baxter M S, 1992. Announcement of a further international intercomparison exercise. Radiocarbon 34(3), 528-532.

Scott, E M (ed), (2003). The third international radiocarbon inter-comparison (TIRI) and the fourth international radiocarbon inter-comparison (FIRI) 1990-2002: results, analyses, and conclusions, *Radiocarbon*, 45, 135-408.

Scott, E.M., Naysmith, P. and Cook, G.T. (2018). Why do we need ^{14}C inter-comparisons? The Glasgow ^{14}C inter-comparison series, a reflection over 30 years. Quaternary Geochronology 43, PP. 72-82.

Scott E M, Naysmith P, Cook G T (this volume). Life after SIRI, where next?

Thompson M, Ellison S R, Wood R. (2006) The international harmonized protocol for the proficiency testing of analytical chemistry laboratories. Pure Appl. Chem., 78(1), 145–196.