RADIOCARBON, Vol 45, Nr 2, 2003, p 159–174

SECTION 3: PRELIMINARY ANALYSIS OF THE RESULTS

3.1 INTRODUCTION

In this section, we present the exploratory analysis of the results submitted by the extended deadline of December 2000. We first deal with Samples C–J, before considering the near-background samples A and B (Kauri wood). The aims of the exploratory analysis are to discover the range of results reported for each sample and the initial evaluation of the effects of any factors that might be a source of variation in the results. For each sample, in turn, we consider the main summary statistics—the number of results reported (N), their mean or average, median, the standard deviation (StDev), the standard error of the mean (Sem), the quartiles (25th [Q1] and 75th [Q3] percentiles), and the minimum (Min) and maximum (Max)—before graphically studying the overall distribution of results in the form of a boxplot, with a view to identifying any extreme or outlying observations. The summary statistics and distribution of results for each laboratory type are also shown. Further details on the statistical methods used are contained in Appendix 3.

3.2 FIRI SAMPLE C: TURBIDITE

The sample was mainly coccolith calcite from a single distal turbidite emplaced on the Maderia Abyssal Plain. It was selected because of its provenance and age. Laboratories had been instructed not to pretreat the sample. This sample had also previously been used in TIRI.

Table 3.1 Descriptive statistics: all results (yr BP)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
93	17,945	18,140	693	72	14,600	18,640	17,900	18,260

Table 3.2 Descriptive statistics: all results (yr BP) by laboratory type

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	34	18,175	18,175	135	23	17,850	18,470	18,100	18,260
GPC	18	17,990	18,180	743	175	15,230	18,640	17,890	18,315
LSC	41	17,735	18,090	874	136	14,600	18,610	17,740	18,193

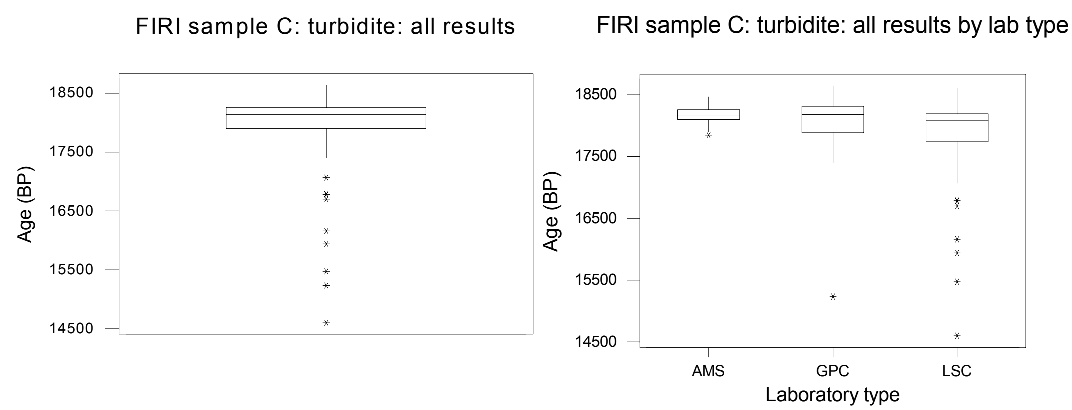

Figure 3.1 Distribution of results for Sample C by all results (left) and laboratory type (right)

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

3.2.1 Comments

From Table 3.1, we can see that the mean and median age are slightly different at 17,945 and 18,140 BP, suggesting that the distribution is skewed. There is a wide range of results (14,600–8640 BP), but 50% of the values lie between 17,900 and 18,260 BP (interquartile range, Q3 to Q1, of 360 yr).

Table 3.2 shows the results for the different laboratory types. There is little difference in the median age for the 3 laboratory types, but, interestingly, we see that the standard deviation for both LSC and GPC laboratories are considerably larger than that for AMS laboratories.

Figure 3.1 graphically shows the distribution of results, with any extreme values (or outliers) identified by an asterisk.

We can see that there is a long lower tail for the turbidite results. When we consider the distribution by laboratory type, we see that this tail is predominantly composed of results from LSC laboratories.

In the homogeneity testing (Section 1), significant differences had been identified between the results from the 2 laboratories, which could be explained by the effect of pretreatment. The mean non-pretreated result had been 18,157 BP.

The turbidite sample had also been used in TIRI (see Part II), where on the basis of 30 results, calculation of the TIRI consensus value gave a result of 18,155 BP with a 1 σ of 34 yr.

3.3 FIRI SAMPLE D: BELFAST DENDRO-DATED PINE

The sample was from a Scots pine tree from Garry Bog, Co. Antrim, Northern Ireland, and had 40 annual growth rings dating from 3239–3200 BC. This sample was distributed in duplicate as Samples D and F. Its ^{14}C age (from the master calibration curve) is approximately 4495 BP.

Table 3.3 Descriptive statistics: all results (yr BP)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
108	4494.4	4517.5	224.2	21.6	2990.0	5060.0	4471.5	4579.0

Table 3.4 Descriptive statistics: all results by laboratory type

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	41	4530.3	4520.0	52.0	8.1	4430.0	4670.0	4500.0	4550.0
GPC	20	4495.1	4504.5	75.9	17.0	4273.0	4600.0	4468.5	4522.5
LSC	47	4462.9	4535.0	331.7	48.4	2990.0	5060.0	4400.0	4590.0

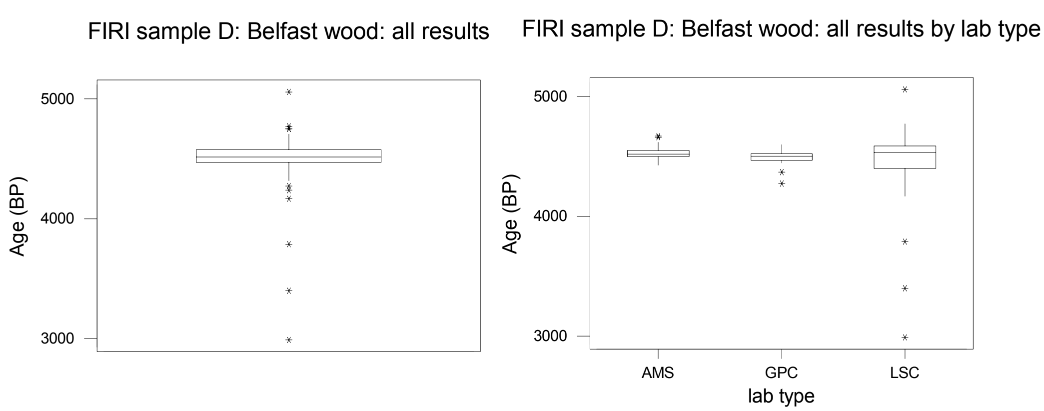

Figure 3.2 Distribution of results for Sample D by all results (left) and laboratory type (right)

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

3.3.1 Comments

We can see from Table 3.3 that the mean and median age are slightly different at 4494 and 4517 BP. We see a wide range (2990–5060 yr), but 50% of the values lie between 4471 and 4579 BP (i.e., just over 100 yr).

Table 3.4 shows the results for the different laboratory types. There is little difference in the median for the 3 laboratory types. Interestingly, as with Sample C, we see that the standard deviation for the results from the LSC laboratories is considerably larger than that for GPC and AMS laboratories.

From Figure 3.2, we can see that there is a lower tail for the results. When we consider the distribution by laboratory type, we see that this tail is predominantly composed of LSC results.

3.4 FIRI SAMPLE F: BELFAST DENDRO-DATED PINE

Table 3.5 Descriptive statistics: all results (yr BP)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
103	4521.4	4504.0	195.8	19.3	4100.0	5870.0	4460.0	4560.0

Table 3.6 Descriptive statistics: all results by laboratory type

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	37	4534.2	4534.0	62.0	10.2	4420.0	4710.0	4489.0	4570.0
GPC	21	4485.0	4470.0	120.1	26.2	4250.0	4740.0	4439.5	4528.5
LSC	45	4527.8	4500.0	279.9	41.7	4100.0	5870.0	4420.0	4555.0

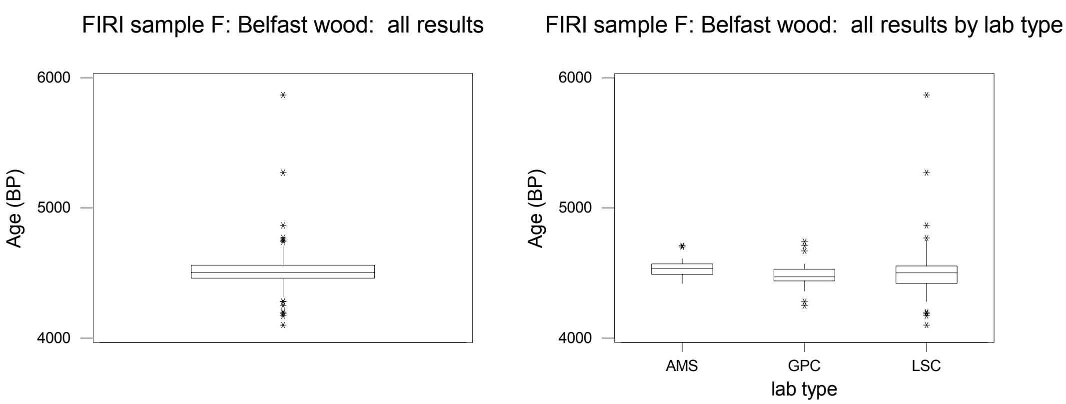

Figure 3.3 Distribution of results for Sample F by all results (left) and laboratory type (right)

3.4.1 Comments

From Table 3.5, we can see that the mean and median age are only slightly different at 4521 and 4504 BP. We also see a narrower range (4100–5870) than for Sample D and that 50% of the values lie between 4460 and 4560 BP (i.e., exactly 100 yr).

Table 3.6 shows the results for the different laboratory types. There is little difference in the median for the 3 laboratory types. Again, we see that the standard deviation for LSC laboratories is considerably larger than that for GPC and AMS.

The median and the middle 50% range for Sample F is almost identical to the results for Sample D.

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

From Figure 3.3, it is clear that there is both a lower and upper tail for the results. When we consider the distribution by laboratory type, we see that this tail is predominantly composed of results from LSC laboratories.

3.5 FIRI SAMPLE E: HUMIC ACID

Table 3.7 Descriptive statistics: all results (yr BP)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
139	11,781	11,780	545	46	7700	15,150	11,670	11,872

Table 3.8 Descriptive statistics: all results by laboratory type (yr BP)

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	65	11,822	11,800	188	23	11,430	13,000	11,765	11,870
GPC	26	11,768	11,734	240	47	11,300	12,314	11,617	11,920
LSC	48	11,731	11,726	888	128	7700	15,150	11,591	11,878

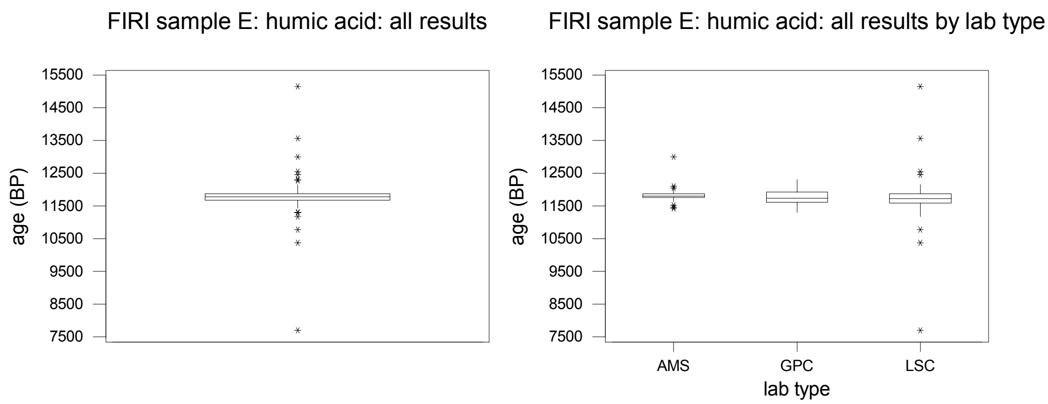

Figure 3.4 Distribution of results for Sample E by all results (left) and laboratory type (right)

3.5.1 Comments

For the humic acid, the mean and median are again in excellent agreement at 11,822 and 11,800 BP, respectively. Again, there is a wide range (7700–15,150 BP), but the interquartile range (IQR) is much narrower (11,670–11,872 BP). We see the same features (Figure 3.4) as before when we look at the summary statistics for each laboratory type with broadly similar mean/median values, but LSC laboratory results have a much larger standard deviation. The distribution of results shows the presence of some extreme values, again predominantly, but not exclusively, reported by LSC laboratories.

3.6 FIRI SAMPLE G: BARLEY MASH

This sample was provided as a duplicate sample with Sample J and reflected current atmospheric levels.

Table 3.9 Descriptive statistics: all results (pMC)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
99	110.08	110.50	2.86	0.29	94.47	121.00	109.71	111.08

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

Table 3.10 Descriptive statistics: all results by laboratory type (pMC)

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	34	110.3	110.3	0.68	0.12	109.0	111.9	109.8	110.8
GPC	19	110.6	111.0	1.36	0.31	107.0	112.6	110.0	111.4
LSC	46	109.6	110.4	4.04	0.60	94.2	121.0	108.8	111.3

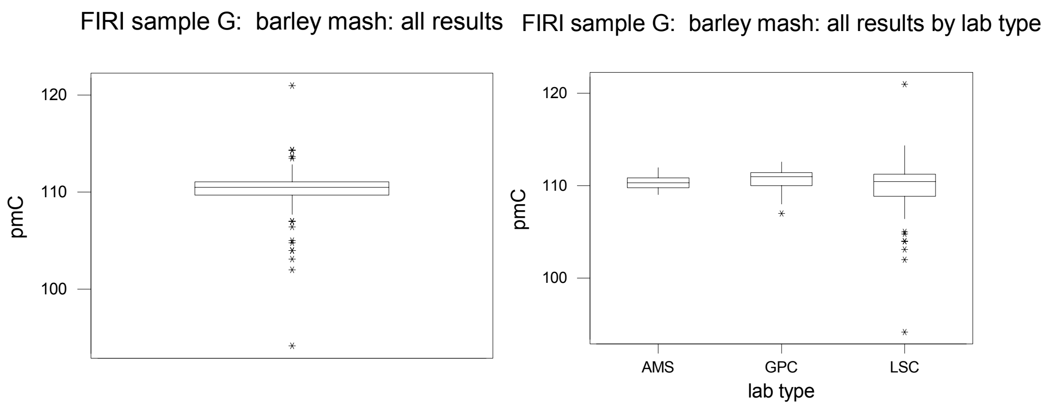

Figure 3.5 Distribution of results for Sample G by all results (left) and laboratory type (right)

3.6.1 Comments

The mean pMC value is estimated at 110.1 and 50% of the data lie in the range 109.7–111.1 (Table 3.9). It is clear, however, from the summary statistics and the graphs that again there are a number of extreme values and that these are reported predominantly by LSC laboratories (Table 3.10).

3.7 FIRI SAMPLE J: BARLEY MASH

Table 3.11 Descriptive statistics: all results (pMC)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
99	110.4	110.6	2.73	0.27	97.1	122.0	110.0	111.3

Table 3.12 Descriptive statistics: all results by laboratory type (pMC)

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	99	110.4	110.6	2.73	0.27	97.1	122.0	110.0	111.3
GPC	19	110.8	110.7	1.19	0.27	108.3	114.4	110.4	111.3
LSC	45	110.0	110.8	3.93	0.59	97.1	122.0	109.0	111.6

3.7.1 Comments

The mean pMC value is estimated at 110.4 and 50% of the data lie in the range 110.0–111.3. However, it is clear from the summary statistics and the graphs that again there are a number of substantial outliers and that these are reported by LSC laboratories. The distribution of results is very similar to that observed for FIRI G.

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

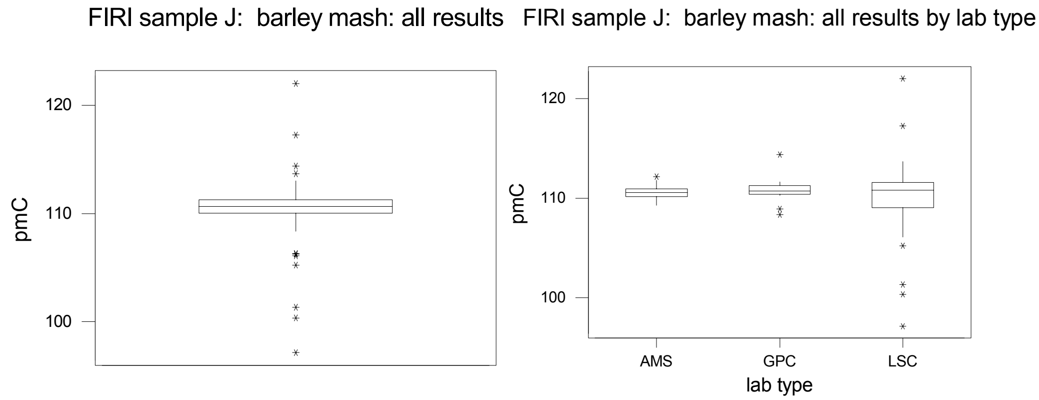

Figure 3.6 Distribution of results for Sample J by all results (left) and laboratory type (right)

3.8 FIRI SAMPLE H: HOHENHEIM DENDRO-DATED OAK

The sample had 20 annual growth rings dating from 313–294 BC, which corresponds to a ^{14}C age of 2215 BP.

Table 3.13 Descriptive statistics: all results (yr BP)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
99	2240.9	2230.0	165.4	16.6	1530.0	2980.0	2180.0	2290.0

Table 3.14 Descriptive statistics: all results by laboratory type (yr BP)

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	36	2228.7	2230.0	48.2	8.0	2135.0	2318.0	2202.3	2260.0
GPC	20	2259.7	2204.0	193.3	43.2	2093.0	2980.0	2180.0	2267.5
LSC	43	2242.4	2232.0	211.3	32.2	1530.0	2690.0	2160.0	2340.0

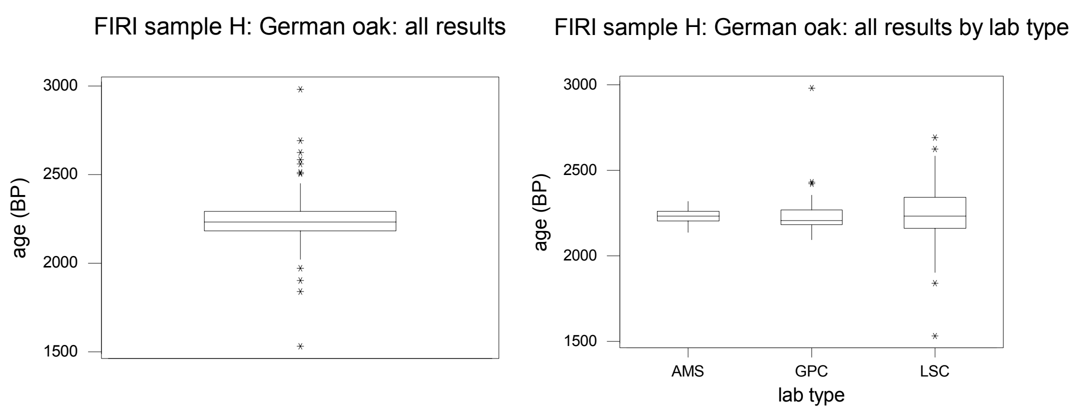

Figure 3.7 Distribution of results for Sample H by all results (left) and laboratory type (right)

3.8.1 Comments

The mean ^{14}C age is estimated as 2241 yr BP and the IQR is 2180–2290 BP (90 yr), but the full range of the data is again extended due to the presence of outliers. The mean and the median age correspond well to the master calibration value ascribed to this sample.

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

3.9 FIRI SAMPLE I: BELFAST CELLULOSE

Table 3.15 Descriptive statistics: all results (yr BP)

N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
96	4484.6	4490.0	218.8	22.3	3780.0	5650.0	4420.0	4560.0

Table 3.16 Descriptive statistics: all results by laboratory type (yr BP)

Type	N	Mean	Median	StDev	Sem	Min	Max	Q1	Q3
AMS	35	4499.1	4490.0	74.1	12.5	4400.0	4710.0	4450.0	4550.0
GPC	18	4498.8	4463.0	192.4	45.3	4290.0	5100.0	4399.0	4493.8
LSC	43	4466.9	4500.0	297.2	45.3	3780.0	5650.0	4380.0	4580.0

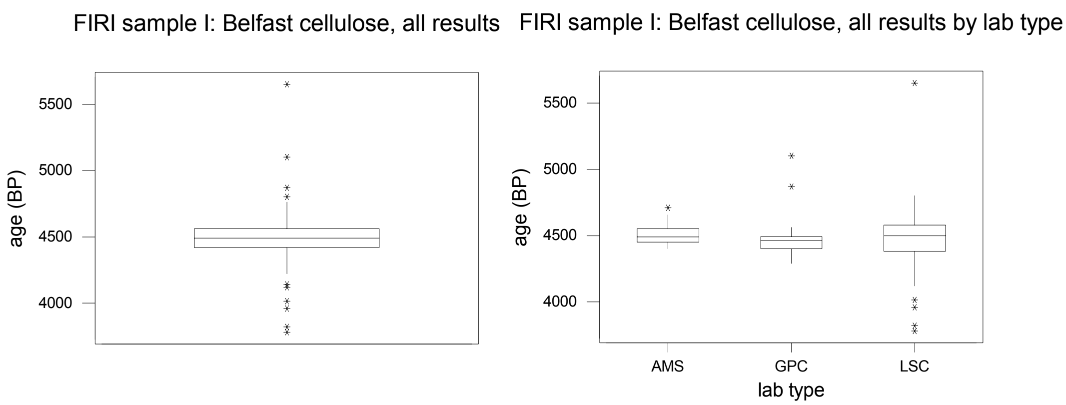

Figure 3.8 Distribution of results for Sample I by all results (left) and laboratory type (right)

The second Belfast sample spanned a contiguous set of rings to FIRI D and F. The sample, which had a finite 40-yr ring span, had a dendrochronologically-determined age span of 3299–3257 BC. This corresponds roughly to a ^{14}C age of 4471 BP.

3.9.1 Comments

The mean and median are very close together at 4485 yr BP, and approximately 15 yr younger than linked samples D and F. The IQR is 140 yr. The graphs show the presence of outliers, again predominantly from LSC laboratories.

3.10 CONCLUSIONS FROM THE PRELIMINARY DISTRIBUTION OF RESULTS

The preliminary analysis of the results for FIRI Samples C–J has shown a consistent pattern, with a reasonably tight IQR (the mid-50% of the results) but with a large range (usually determined by a small number of extreme values). The IQR is reasonably constant at around 100 yr, extending to 300 yr for the oldest sample (Sample C). In the main, although not solely, the extreme results have been reported by liquid scintillation laboratories. From the tables of summary statistics, it is also apparent that the standard deviation in all samples is much larger for LSC laboratories than for GPC or AMS laboratories. Figures A1.a to A1.j in Appendix 1 show the full distribution of results for each sample as well as the $\pm 2\ \sigma$ range for the individual results. These figures also show the same overall pattern as observed in the boxplots, but now the effect of, and relationship to, the quoted error is also

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

apparent. In these figures, a steeply sloping section indicates that there are a large number of laboratories with very similar results; such a feature is very striking in Figure A1.e, and to a lesser extent in A1.d and A1.f. It is also clear that the size of the quoted error does vary quite substantially amongst laboratories. This preliminary analysis has not formally used the associated laboratory quoted error and in the next section, the quoted errors are further explored. For this purpose, all results in Section 3.11 have been quoted in % modern carbon (pMC) for comparability purposes.

3.11 SUMMARY OF THE DISTRIBUTION OF QUOTED ERRORS

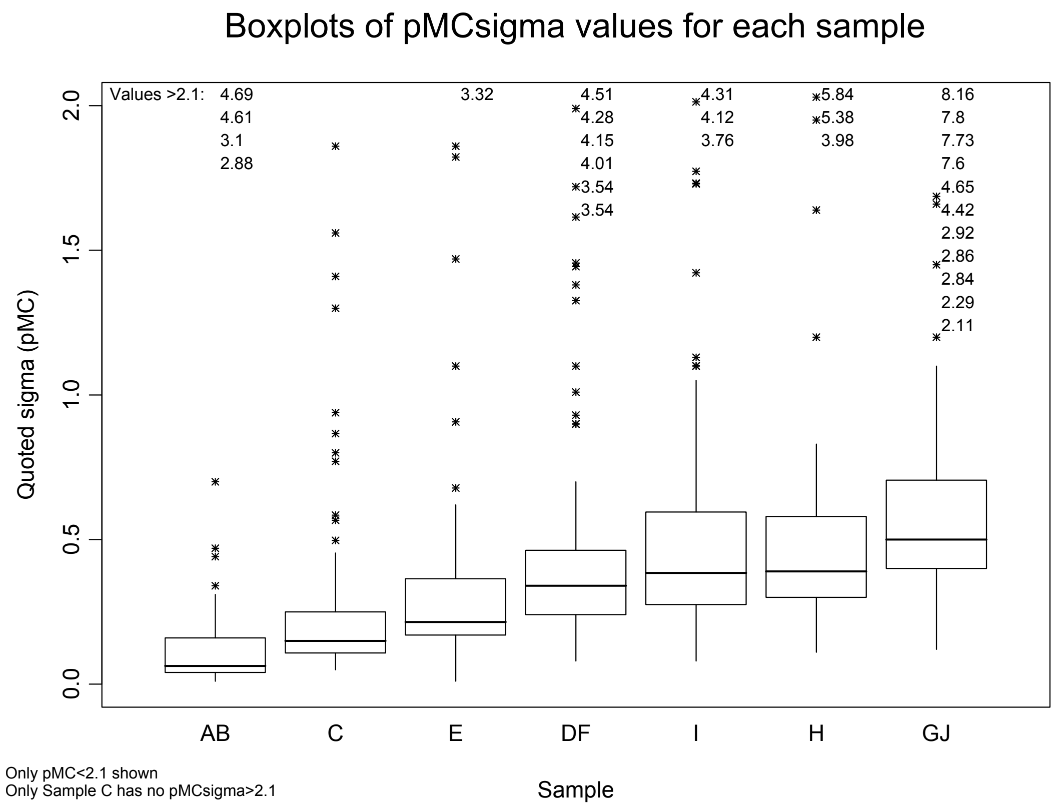

Figure 3.9 Distribution of laboratory quoted errors

Figure 3.9 shows the distribution of quoted errors (all results are given in terms of pMC) for all samples (now including Samples A and B). Extreme values (outliers) are clearly marked by the asterisks. It is clear from the figure that there is a relationship between the pMC and the quoted error, with the quoted error slowly increasing as the sample pMC increases. Similarly, from figures for the different laboratory types, it was quite clear that the quoted errors tend to be larger and more variable for LSC laboratories than for the other laboratory types, and that the AMS laboratories quoted errors tend to be smaller and for there to be much less scatter in their magnitude.

3.12 SUMMARY OF THE $\delta^{13}C$

Laboratories were asked to provide $\delta^{13}C$ values for each sample and to indicate whether these values were measured or estimated. Table 3.17 summarizes the number of laboratories providing this information. In the reporting questionnaire, laboratories were also asked to indicate the stage of the dating process to which the fractionation measure best referred.

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

Table 3.17 Summary of $\delta^{13}C$ reporting

Lab type	Estimated	Estimated and measured	Measured	Missing	**Total nr of labs**
AMS	0	2	22	1	**25**
GPC	2	1	14	11	**28**
LSC	8	2	29	10	**49**
All	10	5	65	12	**92**

The different parts of the process where $\delta^{13}C$ was measured were classified as:

1. The raw material;
2. The material after pretreatment;
3. The actual sample measured.

The $\delta^{13}C$ values for each sample are summarized first for all results in Table 3.18, and then by the stage of the process in Table 3.19.

Table 3.18 Summary table for $\delta^{13}C$ (all results)

Sample	N	Mean	Median	StDev	Min	Max	Q1	Q3
AB	170	–23.9	–24	1.48	–31	–20.1	–24.7	–23.3
C	82	0.51	1.1	2.84	–22.6	3.864	0.8	1.2
DF	188	–24.8	–25.0	1.36	–32.2	–21.6	–25.3	–24
E	119	–28.7	–29.1	2.2	–34.3	–12.3	–29.5	–28.4
GJ	172	–28.9	–29.1	1.34	–34.1	–24.5	–29.5	–28.6
H	87	–25.0	–24.9	1.34	–31.1	–21.1	–25.5	–24.4
I	86	–23.8	–23.7	0.85	–25.5	–20	–24.3	–23.4

Table 3.19 Summary statistics for $\delta^{13}C$ by process stage

Sample/(Stage)	**AB**	(1)	(2)	(3)	**C**	(1)	(2)	(3)	**DF**	(1)	(2)	(3)
N	153	10	62	66	72	12	22	32	165	9	70	68
Mean	–23.8	–24.2	–23.7	–23.9	0.6	0.9	0.9	1.0	–24.7	–25.1	–24.5	–25.0
Median	–23.9	–24.1	–23.8	–23.9	1.1	1.0	1.1	1.1	–24.9	–25.0	–24.3	–25.0
StDev	1.5	0.5	1.4	1.8	3.0	0.5	0.8	1.3	1.4	0.6	1.3	1.6
Min	–31.0	–25.0	–31.0	–31.0	–22.6	–0.7	–2.4	–3.4	–32.2	–26.0	–32.2	–32.2
Q1	–24.4	–24.5	–24.2	–24.6	0.9	0.8	0.9	1.0	–25.3	–25.5	–25.1	–25.5
Q3	–23.2	–24.0	–23.2	–22.9	1.3	1.2	1.1	1.4	–23.9	–25.0	–23.7	–24.1
Max	–20.1	–23.4	–20.9	–20.1	3.9	1.3	1.6	3.9	–21.6	–24.0	–21.7	–21.6
Sample/(Stage)	**E**	(1)	(2)	(3)	**GJ**	(1)	(2)	(3)	**H**	(1)	(2)	(3)
N	69	20	17	27	155	37	37	67	79	4	34	34
Mean	–29.0	–28.8	–29.1	–29.3	–29.1	–28.8	–29.0	–29.4	–25.0	–25.4	–24.7	–25.2
Median	–29.1	–28.8	–29.1	–29.3	–29.1	–29.0	–28.9	–29.3	–24.8	–25.3	–24.8	–25.0
StDev	1.3	1.5	1.1	0.6	0.9	0.9	0.6	1.1	1.4	0.6	1.5	1.4
Min	–32.9	–29.9	–32.9	–30.2	–34.1	–29.8	–30.8	–34.1	–31.1	–26.0	–31.1	–31.1
Q1	–29.5	–29.5	–29.2	–29.6	–29.5	–29.5	–29.2	–29.7	–25.5	–25.9	–25.2	–25.5
Q3	–28.8	–28.8	–28.9	–29.1	–28.7	–28.6	–28.5	–28.9	–24.3	–24.8	–24.1	–24.4
Max	–23.0	–23.2	–27.7	–27.6	–25.9	–26.0	–28.0	–25.9	–21.1	–24.8	–21.1	–23.0
Sample/(Stage)	**I**	(1)	(2)	(3)								
N	77	18	20	33								
Mean	–23.7	–23.7	–23.5	–23.8								
Median	–23.7	–23.7	–23.5	–23.9								
StDev	0.7	0.6	0.6	0.8								
Min	–25.5	–25.1	–25.0	–25.5								
Q1	–24.0	–24.0	–23.7	–24.4								
Q3	–23.4	–23.2	–23.3	–23.6								
Max	–21.7	–22.3	–21.7	–21.7								

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

The boxplots in Figure 3.10 show the pattern of measured $\delta^{13}C$ values for the samples, except Sample C. The barley and humic samples are comparable and lighter than the wood samples. There may be some suggestion that Sample I (cellulose) is heavier than Samples D and F. It is also of interest to consider the differences in the $\delta^{13}C$ values at the different stages and this is shown graphically in Figures 3.11 and 3.12. It should be remembered that the $\delta^{13}C$ values should not be used as the reference isotopic ratio for these samples; rather, it may prove a useful marker for the variation in measurement. The results have shown small differences in the different process stages. There is little evidence for any of the samples that there is significant variation in the fractionation incurred at the different stages. There is some variation in the $\delta^{13}C$ values quoted, but these effects are likely to be small in the overall variation of the results.

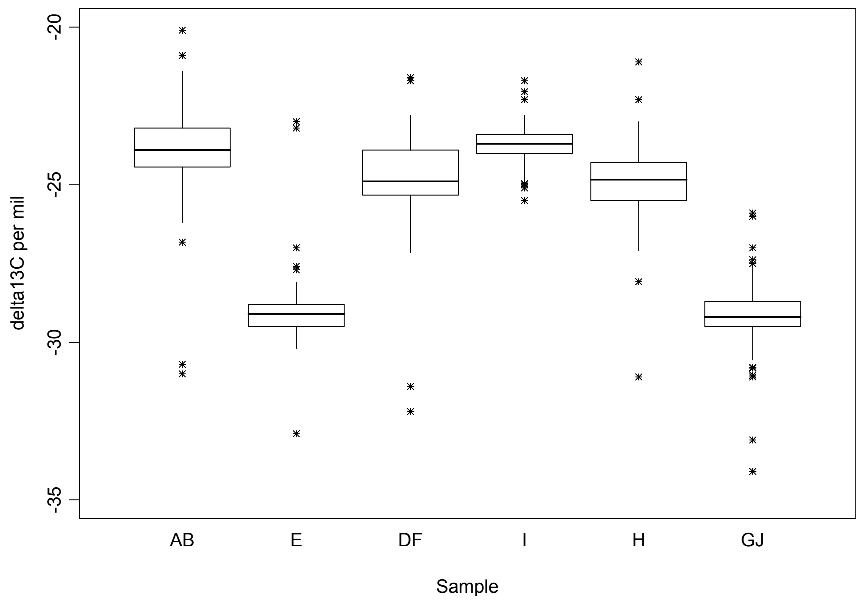

Figure 3.10 $\delta^{13}C$ for all samples (except Sample C, turbidite)

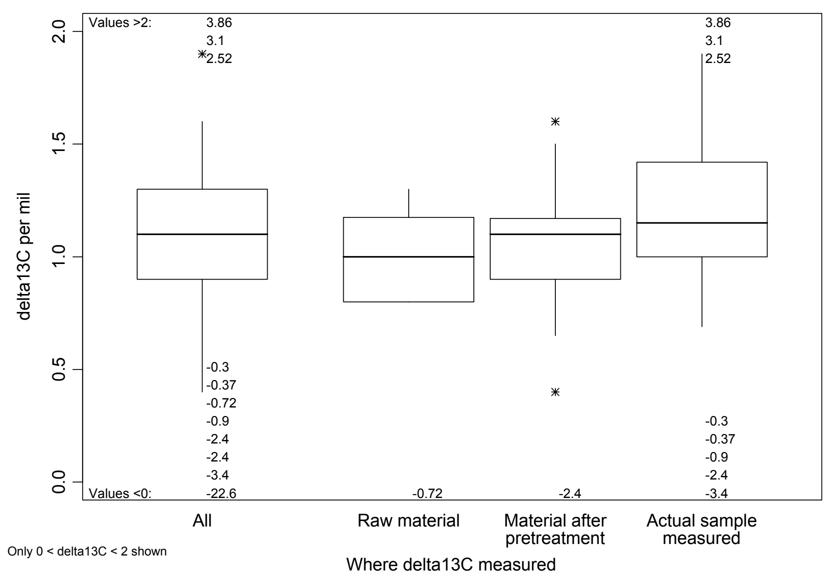

Figure 3.11 $\delta^{13}C$ for Sample C (turbidite) in process for different point of measure categories

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

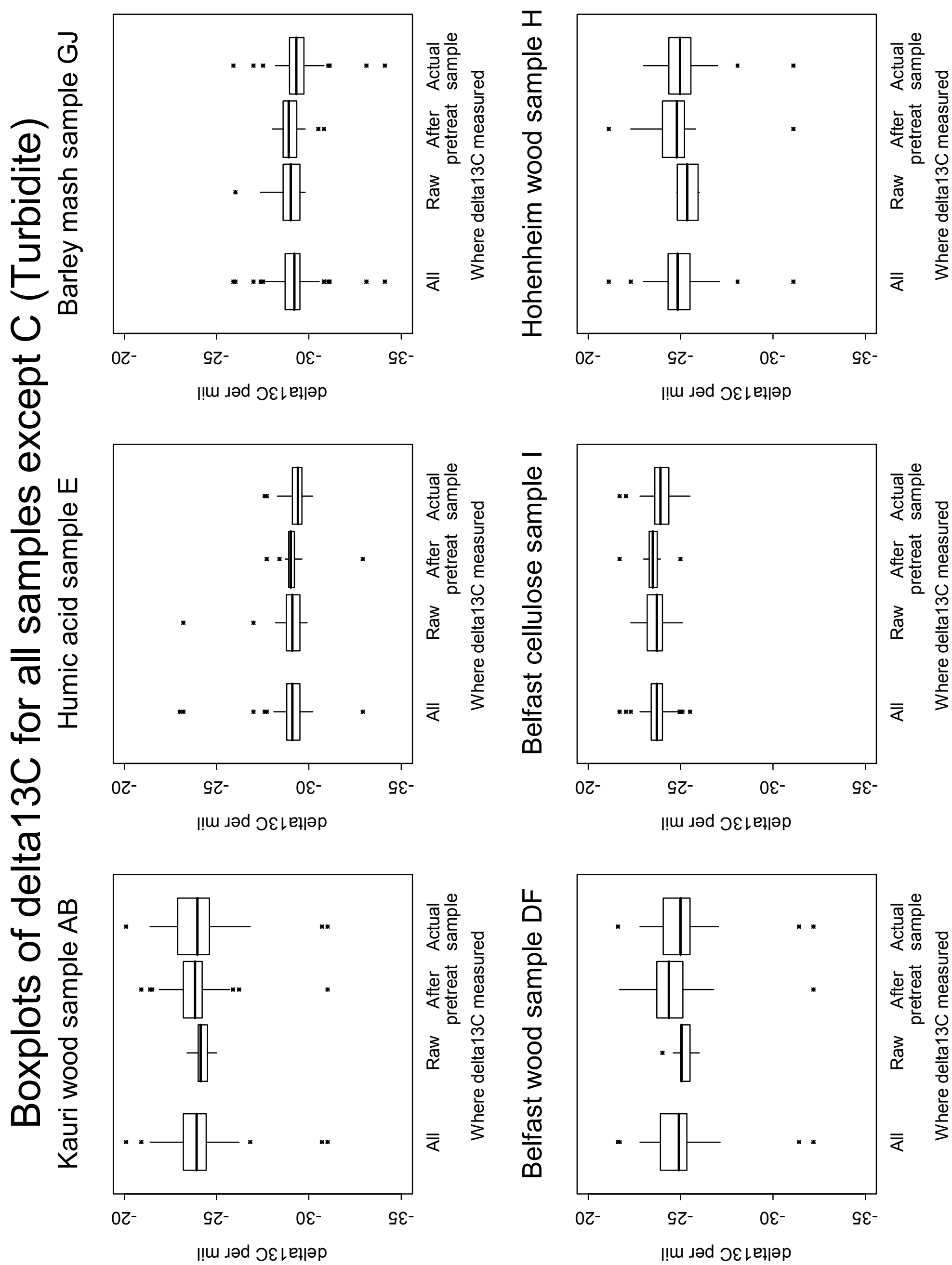

Figure 3.12 $\delta^{13}C$ for all samples by stage in process

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

3.13 OUTLIERS OR EXTREME VALUES

3.13.1 Outlier Definitions

There are many ways for defining outliers and no universal statistical definition. In this report, we have used a conventional definition that is widely used in the statistical software, and in particular, is used to identify outliers when producing boxplots.

For the purposes of this investigation of outliers and the similarities they exhibited, outliers were defined as all results greater or less than 1.5× IQR from the middle 50% of the results or result < Q1 – 1.5(Q1–Q3) or result > QU + 1.5(QL–QU), where *QL* and *QU* are the upper and lower quartiles, respectively.

In a previous analysis of ^{14}C results from an intercomparison, the standard consensus (Rozanski et al. 1992) calculations were used at the first stage of the calculation, a definition such that outliers were classed as those results that were more than 3 IQR from the middle 50% of the results (i.e., result < QL – 3(QL–QU) or result > QU + 3(QL–QU), where *QL* and *QU* are the upper and lower quartiles, respectively.

Using the 1.5 IQR definition, the outlier boundaries are defined below:

Table 3.20 Outlier boundaries (in pMC for Samples AB and GJ)

Limit	AB	C	DF	E	GJ	H	I
Lower	–0.5	17,362	4313	11,358	108	2004	4210
Upper	1.3	18,796	4723	12,168	113	2457	4770

3.13.2 Outlier Description

A total of 122 observations from 1056 (i.e., slightly over 10%) were identified as outliers using these definitions and here we explore the nature of these outlier observations.

Table 3.21 Percentage distribution of outliers amongst laboratory types

Laboratory type	Number of identified outliers	%
AMS	1	0.82
GPC	15	12.30
LSC	106	86.89
All	**122**	**100.00**

Thus, of the 122 outliers, 87% came from LSC laboratories.

We can also consider whether there was any association with the outlier results and the modern standard material or background material used.

Table 3.22 Number of outliers reported where laboratory used the given standard material

Modern standard material	Number of outliers	%
ASUC	15	14.02
Benz	17	15.89
NBS1	23	21.50
NBS2	45	42.06
other	7	6.54

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

Over half of the outliers were submitted by laboratories using NBS Ox1 and NBS Ox2.

Table 3.23 Number of outliers reported where laboratory used the given background material

Background material	Number of outliers	%
Anth	17	17.17
Benz	39	39.39
Coal	5	5.05
Graph	4	4.04
Marble	17	17.17
None	2	2.02
Other	15	15.15
All	**99**	**100.00**

The distribution of outliers is uniform over the sample; thus, no single sample contributes the majority of the outliers if we consider the joint distribution of laboratory type and standard used for those outlier results. The distribution is shown in the table below.

In terms of background material, the most common background material is benzene (scintillation-grade benzene) and over 39% of the outliers are associated with the use of benzene as the background material.

Table 3.24 Number of outliers reported where laboratory used the given background material

Sample	Number of outliers	%
A	11	9.02
B	7	5.74
C	11	9.02
D	12	9.84
E	13	10.66
F	16	13.11
G	16	13.11
H	13	10.66
I	13	10.66
J	10	8.20

Table 3.25 Numbers of outliers for given laboratory type and modern standard material

	ASUC	Benz	NBS1	NBS2	Other	**All**
AMS	0	0	0	1	0	**1**
GPC	2	0	2	11	0	**15**
LSC	13	17	21	33	7	**91**
All	**15**	**17**	**23**	**45**	**7**	**107**

There appears to be no statistical association between laboratory type and modern standard used for the outlier results.

Table 3.26 Numbers of outliers for given laboratory type and background material

	Anth	Benz	Coal	Graph	Marble	Other	**All**
AMS	0	0	1	0	0	0	**1**
GPC	4	0	4	4	1	2	**15**
LSC	13	39	0	0	16	15	**83**
All	**17**	**39**	**5**	**4**	**17**	**17**	**99**

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

It seems that there is a statistical association between laboratory type, background material, and outlier results.

3.13.3 Distribution of Outliers Across Labs

Of the 92 laboratories in the intercomparison, there were 39 (42%) which had at least 1 result classed as an outlier. Information about each of these is given in the following tables.

Of the 39 laboratories that had 1 or more outliers, almost 60% (23) of these had more than 1 of their results thus classed and over one-fifth (9) had 5 or more such results (see Table 3.27).

Table 3.27 Count of laboratories in different number-of-outlier-results groups

Number of outliers	0	1	2	3	4	5	6	7	9	11	**Total**
Number of laboratories	53	16	6	4	4	1	3	2	2	1	**92**

From Table 3.28, over 75% (30) of the laboratories with outliers used LSC, while all but one of the rest used GPC. Thus, a larger proportion of the outlier laboratories used LSC, compared to the LSC representation in the overall set of results, where 53% of the laboratories used LSC.

Table 3.28 Count of laboratories in different measurement method groups

Measurement method	AMS	GPC	LSC	**Total**
Number of outlier laboratories	1	8	30	**39**
All laboratories	25	18	49	**92**

From Table 3.29, we can see that just over 50% (20) of these 39 laboratories did not state that they measured the $\delta^{13}C$ for all their samples. Nine of these 20 laboratories definitely estimated $\delta^{13}C$, 3 used both measured and estimated values, while the other 8 did not specifying whether or not they did. In the overall case, only 29% (27) of the 92 did not state that they measured the $\delta^{13}C$ for all their samples.

Table 3.29 Counts of outlier and all laboratories' $\delta^{13}C$ categories

$\delta^{13}C$ measured or estimated	Estimated	Estimated and measured	Measured	Missing	**Total**
Number of outlier laboratories	9	3	19	8	**39**
All laboratories	10	5	65	12	**92**

Table 3.30 shows the types of background and modern standard materials used by laboratories with outliers and all laboratories. From this table, we can see that benzene was a far more commonly used background material in the outlier group (38% of the time) than overall (21%). This was also the case with the modern standards, where 6 out of the 7 laboratories using benzene were in the outlier group. It should be noted that the types of benzene used varied from laboratory to laboratory, unlike the other modern standards.

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

Table 3.30 Types and numbers of laboratories using backgrounds and modern standards

Background material			Modern standard material		
Category	Outlier laboratories	All	Category	Outlier laboratories	All
Benzene	15	19	ANU Sucrose	3	8
CO_2	1	3	Benzene	6	7
Coal	9	17	NBS OXI	6	21
Graphite	1	4	NBS OXII	17	32
Marble	2	6	NBS OXI/OXII	0	5
Natural Gas	1	3	1 NBS & 1 other	3	3
Others	4	10	Other	2	7
More than 1	1	13	Missing	2	9
Missing	5	17	**Total**	**39**	**92**
Total	**39**	**92**			

3.13.4 Conclusions

A total of 122 observations out of 1056 (i.e., slightly over 10%) were identified as anomalous (i.e., outliers). From the statistical definition of an outlier, around 5% of the results would have been expected to have been classed as outliers. Thus, approximately twice as many outliers were identified as would be expected if they were occurring purely by chance. Of the 122 outliers, 87% came from LSC laboratories. The distribution of outliers was uniform over the 10 samples; thus, no single sample contributed the majority of the outliers. Thirty-nine laboratories (42%) had at least 1 result classed as an outlier. Of the 39, almost 60% (23) of these had more than 1 of their results thus classed, and over one-fifth (9) had 5 or more such results.

Table 3.31 Operational information concerning laboratories with at least 1 outlier

Lab nr	$\delta^{13}C$ measured (M) or estimated (E)	Background material	Modern standard material	Nr of outlier results
5	E	IAEA C1	NBS OXI	7
10	—	Benzene	Benzene	6
11	M	Anthracite	OXII / ANU	4
13	M	Benzene	NBS OXII	3
15	M	Anthracite	NBS OXII	1
16	—	—	Benzene	6
17	E	—	NBS OXI	1
18	M (E & M)	Anthracite	NBS OXI	1
19	E (E & M)	Methanol	NBS OXII	1
21	E	Benzene	Benzene	2
23	M	Anthracite	NBS OXII	2
26	—	—	—	6
28	—	—	NBS OXII	2
30	M	Benzene	Benzene	1
31	M	TIRI-G CO_2	ANU Sucrose	1
32	M	Marble	NBS OXII	1
39	M	Benzene	NBS OXII	5
42	E	Benzene	Benzene	2
43	M	Anthracite	ANU Sucrose	3
44	M	Graphite	NBS OXII	4

https://doi.org/10.1017/S0033822200032598 Published online by Cambridge University Press

Table 3.31 Operational information concerning laboratories with at least 1 outlier *(Continued)*

Lab nr	$\delta^{13}C$ measured (M) or estimated (E)	Background material	Modern standard material	Nr of outlier results
53	E	Marble	ANU Sucrose	9
56	M	Anthracite	NBS OXII	2
57	E	Natural Gas	OXII / C3	1
59	M	Anthracite	NBS OXII	1
63	E	Benzene	NBS OXII	2
66	M	Limestone	NBS OXI	1
67	E	Benzene	GIN	3
68	M (E & M)	Benzene	NBS OXII	1
59	M	Benzene	NBS OXII	7
70	—	IAEA C4	NBS OXI	11
71	—	Benzene	Other	3
75	—	Benzene	Other	1
76	M	Benzene	Other	1
78	—	—	—	9
80	M	Benzene	NBS OXI	1
81	E	Anthracite	NBS OXII	4
89	M	Benz/Anthracite	NBS OXII	1
90	M	Anthracite	NBS OXII	4
92	M	Benzene	NBS OXII	1

Clearly, a relatively small number of laboratories (14%) generated more than 60% of the outlying observations. The majority of these laboratories use liquid scintillation techniques (including direct absorption). However, it should be noted that there remains a substantial number of liquid scintillation laboratories with none or only 1 outlier.

Further analysis indicated that the presence of outliers was linked to the modern standard used, with some laboratories having no access to the primary standards of NIST OxI and OxII.