RADIOCARBON AGE ASSESSMENT OF A NEW, NEAR BACKGROUND IAEA ^{14}C QUALITY ASSURANCE MATERIAL

A. G. HOGG,[1] THOMAS HIGHAM,[1] STEVE ROBERTSON,[2] ROELF BEUKENS,[3] TUOVI KANKAINEN,[4] F. G. McCORMAC,[5] JOHANNES VAN DER PLICHT[6] and MINZE STUIVER[7]

ABSTRACT. The ^{14}C Quality Assurance Programme coordinated by the IAEA (Rozanski *et al.* 1992) prepared a set of five new intercomparison materials, including 40–50 ka old subfossil wood excavated from New Zealand peat bogs (IAEA C-4 standard). Statistical analysis of 79 ^{14}C measurements made on the wood indicated considerable variation in the results, with a marked skewness toward more modern values. The wide range of results and the possibility of inhomogeneity within the standard prompted the recovery and analysis of replacement material. The new subfossil wood sample is kauri (*Agathis australis*), at least 50 ka old, excavated from a swamp in Northland. It is in the form of a single plank, 6 m long, weighing 80 kg. It will be forwarded to the IAEA in Vienna for milling and distribution. Subsamples were obtained from both ends of the plank and analyzed by six laboratories. We present here the results of these analyses and compare them with the previous IAEA intercalibration results for the C-4 standard.

INTRODUCTION

The ^{14}C Quality Assurance Programme coordinated by the IAEA in 1990 was designed to establish a bank of ^{14}C reference materials to assist laboratories engaged in various fields of scientific research in checking the quality of their work. Five new intercomparison materials and the ANU sucrose secondary standard were distributed to 137 laboratories worldwide, and percent modern carbon (pMC) consensus values obtained (Rozanski 1991; Rozanski *et al.* 1992).

The IAEA C-4 quality assurance standard is subfossil wood, kauri *(Agathis australis)*, at least 40 ka old, excavated from a peat swamp in Northland, New Zealand. The material consists of 70 kg of pulped, homogenized wood chips compressed into soft board. The 79 analyses submitted to IAEA showed a marked positive skewness (*i.e.*, strong skewness to more positive pMC values (Rozanski *et al.* 1992: Fig. 6)). Even after a severe statistical cleansing of the results, which resulted in *ca.* 54% being rejected, the accepted analyses showed too much variation for a consensus value to be calculated. Instead, a median value with its 95% confidence interval was given. Rozanski *et al.* (1992) suggest that inhomogeneity of the pulped kauri wood may be the reason for the skewness, with variable contamination by traces of modern carbon. In this context it is interesting to note that the background standard (C-1, Carrara marble) also showed positive skewness, and likewise required rejection of >50% of the analyses to achieve a consensus value. This suggests that factors other than sample inhomogeneity may be responsible for the spread of results.

Two of us (A.H. and T.H.) collected a new sample of subfossil kauri also from a Northland swamp (hereafter called the "new kauri standard") to provide an unquestionably homogeneous quality assurance material, representative of samples approaching the limits of ^{14}C dating. The standard is a 6 m plank (300 × 75 mm) weighing 80 kg. Subsamples (A and B) were obtained from both ends of the plank and ^{14}C-dated by each participating laboratory. Wood samples were pretreated by each

[1] Radiocarbon Laboratory, University of Waikato, Private Bag 3105, Hamilton, New Zealand
[2] Radiocarbon Dating Research, Australian National University, Canberra, ACT 2600 Australia
[3] IsoTrace Laboratory, University of Toronto, Toronto, Ontario M5S 1A7 Canada
[4] Geological Survey of Finland, FIN-02150 Espoo, Finland
[5] Radiocarbon Laboratory, Queen's University of Belfast, Belfast BT7 1NN Northern Ireland
[6] Centre for Isotope Research, University of Groningen, NL-9718 CM Groningen, The Netherlands
[7] Quaternary Research Center, University of Washington, Seattle, Washington 98195 USA

Proceedings of the 15th International ^{14}C Conference, edited by G. T. Cook, D. D. Harkness, B. F. Miller and E. M. Scott. RADIOCARBON, *Vol. 37, No. 2, 1995, P. 797–803*

laboratory, utilizing either cellulose or AAA pretreatment methods (see Table 1). We discuss here the results and compare them with those obtained for the IAEA C-4 standard.

TABLE 1. Summary of pMC Data for the New Kauri Standard

Lab	Lab type*	Background standard	Pretreatment method	Sample A (pMC)	Sample B (pMC)	Pooled mean T†($\chi^2_{1,0.05}$)	IAEA C-4 (pMC), [pretreatment]
QL	GPC	Anthracite	Cellulose	0.212 ± 0.014	−0.015 ± 0.022	0.147 ± 0.012 75.8 (3.8)	0.065 ± 0.028 [none]
UB	LSC	Combusted benzene	Cellulose	0.150 ± 0.020	0.240 ± 0.015	0.208 ± 0.012 13.0 (3.8)	0.198 ± 0.016 [AAA]
GrN	GPC	Anthracite	AAA	0.090 ± 0.024	0.109 ± 0.025	0.099 ± 0.017 0.30 (3.8)	0.070 ± 0.060 [AAA]‡
Su	GPC	Anthracite	Cellulose	0.231 ± 0.032	0.171 ± 0.027	0.196 ± 0.021 2.05 (3.8)	0.212 ± 0.025 [cellulose]
To	AMS	Anthracite & others§	Cellulose AAA	0.222 ± 0.025 0.412 ± 0.020	0.212 ± 0.019	0.216 ± 0.015 0.10 (3.8)	0.208 ± 0.017 [AAA]#
Wk	LSC	Combusted benzene	Cellulose	0.233 ± 0.017	0.209 ± 0.021	0.224 ± 0.013 0.79 (3.8)	0.200 ± 0.050 [none]

*GPC = gas proportional counting; LSC = liquid scintillation counting; AMS = accelerator mass spectrometry

†T and ($\chi^2_{1,0.05}$) as per Wilson and Ward (1981)

‡New C-4 result utilizing "strong" AAA pretreatment (van der Plicht, personal communication 1994); published C-4 pMC = 0.33 ± 0.03

§Background standards include Groningen anthracite, Carrara marble and CO_2 from natural gas

#New C-4 result using cellulose fraction (Beukens, personal communication 1994); published C-4 (AAA) pMC = 0.501 ± 0.016.

INTERPRETATION OF RESULTS

Assessment of pMC and Homogeneity

The principal aims of this research were twofold: 1) we wanted to investigate the homogeneity of the new kauri standard; 2) we hoped to assign a definitive age to the material based on the results. Table 1 shows the results. Four laboratories, the Centrum voor Isotopen Onderzoek in Groningen (GrN), Geological Survey of Finland (Su), IsoTrace Laboratory at the University of Toronto (To) and University of Waikato (Wk), obtained statistically indistinguishable results for both samples (Table 1, column 7 and Fig. 1). Although the other two laboratories, Quaternary Isotope Laboratory at Seattle (QL) and the Queen's University of Belfast Laboratory (UB), obtained statistically dissimilar results, they did not agree upon which end was the older and which the younger: QL found sample A younger, whereas UB found sample B younger.

Neither Group A nor Group B samples are homogeneous populations—Sample Group A: T=33.1; $\chi^2_{5:0.05} = 11.1$; Sample Group B: T=106.7; $\chi^2_{5:0.05} = 11.1$ (for explanation of χ^2, see Appendix). For this reason, we evaluate further the pMC and homogeneity of this new kauri standard. We have used two different methods to accomplish this. The first follows the approach used in the IAEA intercomparison exercise (Tables 2A and 2B) (Rozanski *et al.* 1992), and the second involves cluster analysis as described by Wilson and Ward (1981) (Table 3).

Evaluation Using the IAEA Method (after Rozanski et al. 1992)

The IAEA method attempted to achieve a consensus value by applying a series of statistical tests to identify outliers. Tables 2A and 2B summarize data evaluation as determined by the IAEA method. First, we examined each group of results (*i.e.*, Group A samples, then Group B samples), then exam-

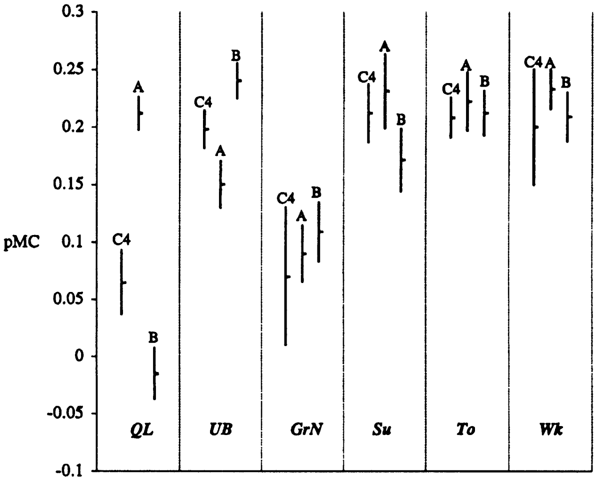

Fig. 1. pMC data showing ±1 σ uncertainties for the new kauri standard and the IAEA C-4 standard. Samples A and B were obtained from opposite ends of one 6-m-long kauri plank.

ined the combined results (*i.e.*, Group A samples plus Group B samples). We included the combined results for completeness because of the small number of samples (6) in each of the 2 groups.

The tests can be summarized as follows (Rozanski *et al.* 1992: 509):

Stage 1. Values outside the range HL - 3(HU - HL) and HU + 3(HU - HL) were rejected, providing the *preliminary consensus value* (HL = lower quartile, HU = upper quartile and HU–HL = the interquartile range (see Table 2A)).

Stage 2. A subgroup of results was identified by accepting the result x,s if $|(x-m)/s| < 2$, where x is the pMC, s the quoted error and m, the preliminary consensus value found in Stage 1 (see Table 2B).

Stage 3. A *final consensus value* was obtained from a weighted average of the stage 2 results (see Table 2B).

As may be seen in Table 2A, Stage 1 of the characterization process does not result in rejection of any samples from either Groups A or B. By applying Stage 2, the 2 samples with the lowest pMC values are removed from each group, leaving 4 samples in each (Table 2B). The same four samples are removed when Stage 2 is applied to the whole data set. The consensus values for Groups A and B are statistically indistinguishable with a pooled mean of 0.220 ± 0.006 (T=6.6; $\chi^2_{7:0.05} = 14.1$) (Table 2B).

TABLE 2A. Statistical Analysis by the IAEA Method: Preliminary Summary of Results

Samples	Total no. of analyses	No. of analyses after outlier removal	Median	Interquartile range	Lower quartile (HL)	Upper quartile (HU)
Group A	6	6	0.217	0.111	0.120	0.231
Group B	6	6	0.190	0.179	0.047	0.226
Group A + B	12	12	0.211	0.072	0.150	0.222

TABLE 2B. Statistical Analysis by the IAEA Method: Final Consensus Values

Samples	No. of analyses*	Consensus value† (pMC)	Estimated standard deviation‡ (pMC)
Group A	4 (6)	0.222	0.005
Group B	4 (6)	0.218	0.011
Group A + B	8 (12)	0.220	0.006

*Number of accepted analyses; total number of analyses submitted is indicated in parentheses

†Calculated as a weighted average (Appendix)

‡Estimated standard error (Appendix)

Evaluation by Cluster Analysis (after Wilson and Ward 1981)

We also used cluster analysis, as outlined in Wilson and Ward (1981), to determine if distinct groups can be recognized in the data. For each group, the results were taken as an instance of Case 1 (Wilson and Ward 1981: 20), *i.e.*,

> one can assume that all determinations have the same corresponding real age value since the determinations are made on a single object and that differences in the determination values have occurred due to changes in the circumstances (often uncontrollable) under which each determination was made.

The individual results are shown in Table 3. Subsets within Groups A or B are 1 to k and k+1 to *n*, where k is indicated by the maximum Λ.

Table 3. Cluster Analysis*

Subset	Sample	pMC	Uncertainty (Wi)	Λ
Group A Samples				
1	GrN/A	0.090	0.024	21.625
	UB/A	0.150	0.020	28.411
2	QL/A	0.212	0.014	11.930
	To/A	0.222	0.025	8.744
	Su/A	0.231	0.032	6.408
	Wk/A	0.233	0.017	
Group B Samples				
0	QL/B	−0.015	0.022	84.758
1	GrN/B	0.109	0.025	87.356
2	Su/B	0.171	0.027	73.412
	Wk/B	0.209	0.021	46.053
	To/B	0.212	0.019	29.115
	UB/B	0.240	0.015	

* After Wilson and Ward (1981)

Group A samples fall into 2 subsets (1 and 2 in Table 3), each of which is homogeneous, giving pooled pMC values of 0.125 ± 0.015 and 0.222 ± 0.014. Group B samples can also be subdivided into 2 subsets, but the first is not homogeneous and needs to be further subdivided, giving 3 subsets (0, 1 and 2), with pMC values of −0.015 ± 0.022 (QL/B), 0.109 ± 0.025 (GrN/B) and the pooled result for the remaining 4 samples of 0.218 ± 0.010. Group A and B samples are arranged according to subset in Table 4.

TABLE 4. Summary of Cluster Analysis of the New Kauri Standard

Subset	Group	pMC	Pooled pMC	T	$\chi^2_{1,0.05}$	No. of samples in subset
0	B	−0.015 ± 0.022	−0.015 ± 0.022			1
1	A	0.125 ± 0.015	0.121 ± 0.013	0.31	3.8	2
	B	0.109 ± 0.025				1
2	A	0.222 ± 0.014	0.220 ± 0.008	0.05	3.8	4
	B	0.218 ± 0.010				4

On the basis of Table 4, we suggest that there are two principal homogeneous subsets with pooled pMC values of 0.121 ± 0.013 and 0.220 ± 0.008. Analysis of the data by the IAEA approach differs from the clustering method only in that it grouped subsets 0 and 1, emphasizing subset 2. On the basis of the cluster analysis, we consider this is not justified and that both principal subsets of data should be retained. It also reinforces the view that samples from both ends of the plank (*i.e.*, A and B samples) are indeed homogeneous. Figure 2 shows these subsets graphically.

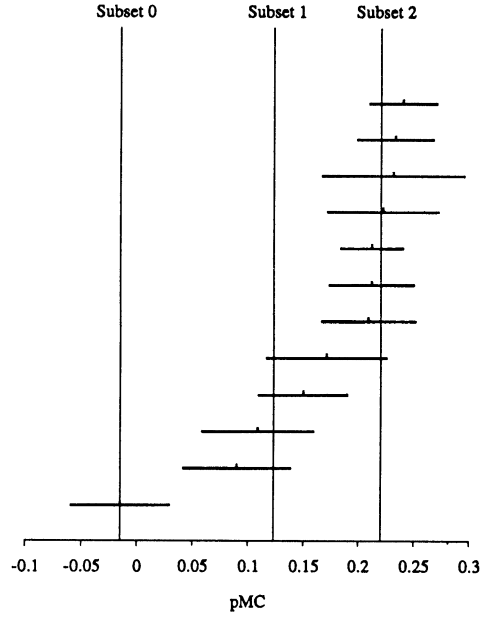

Fig. 2. Plot of pMC ± 2σ uncertainties for all results. Subsets refer to differing pMC levels (see text for details).

The results of both the IAEA and the cluster analysis methods indicate that a pooled mean of all 12 results is not justified, with the data falling into two principal subsets. It is equally apparent that both ends of the plank are statistically indistinguishable in their pMC contents. The variation within the data must therefore be attributable to other factors, such as variable sample pretreatment or inherent laboratory problems.

Pretreatment

No formal investigation was made of the effect of different pretreatments upon the measured pMC results. The IsoTrace Laboratory (To), however, measured both the cellulose fraction and the fraction remaining after AAA pretreatment (Table 1). The latter resulted in a significantly higher pMC result. Both IsoTrace and Groningen applied more rigorous pretreatments for the new kauri standard than they had for the IAEA C-4 standard, and obtained significantly lower pMC values. Clearly, pretreatment appears to be an important variable. We could not, however, identify the most effective method, and indeed, results varied among laboratories using identical pretreatments. Thus, it is likely that pretreatment is undoubtedly an important factor in explaining *some* of the variation in the results, but is not solely responsible. This conclusion appears justified when the results of the IAEA C-4 standard and the pretreatments applied are considered (Table 1, col. 8). Two laboratories dated untreated material, obtaining a pMC range of 0.065 to 0.20 pMC; 3 other labs used AAA treatment with results ranging from 0.07 to 0.208 pMC, and 1 used cellulose extraction (0.212 pMC).

Laboratory Analysis

No correlation could be found between laboratory type and background material, which would explain the two pMC principal subsets outlined above. Either the Subset 1 measurements represent an overestimation of background levels, or the higher Subset 2 measurements result from modern contamination in the sample preparation procedures or underestimate of the background signal. At present, we cannot be confident which hypothesis is correct, or assign a "true" pMC value to the new kauri standard. It may be wiser to apply a narrow range (0.12–0.21 pMC) to the standard until these problems are resolved.

COMPARISON WITH IAEA C-4 QUALITY ASSURANCE MATERIAL

The IAEA C-4 measurements for each of the six participating laboratories are shown in the last column in Table 1 and are included in Figure 1. Figure 2 gives the distribution of results for the new kauri standard. We do not see the marked positive skewness shown by the IAEA C-4 results (see Fig. 6, Rozanski *et al.* 1992) in this study. As mentioned above, C-4 results from Groningen and Toronto differ from those previously released in the 1990 IAEA report (Rozanski 1991). In both cases, more intensive pretreatment of the new kauri standard resulted in significantly lower pMC values (see footnotes in Table 1). It is apparent from Figure 1 that the new kauri standard determinations are very similar to the IAEA C-4 results. The GrN, Su, To and Wk laboratories give indistinguishable results for the two standards, whereas the QL and UB C-4 values lie within the range of the new kauri standard determinations.

Once again, there is clearly a combination of reasons for the widespread of results obtained for the IAEA C-4 reference material. The positive skewness probably results from either modern contamination in the sample preparation procedures, incorrect background assessment or inadequate and/or variable sample pretreatment.

CONCLUSION

A number of observations can be made from this study:

1. An 80-kg plank of homogeneous subfossil wood, at least 50 ka old, is available for use as an IAEA quality assurance material if the ^{14}C community decides that the present C-4 standard is of doubtful purity and a replacement standard is needed. We were unable to assign a definitive pMC value, and instead, suggest a range, 0.12 to 0.21 pMC. Further work will be needed to achieve a more precise value.
2. These results suggest that much of the skewness in the IAEA C-4 data is due to factors other than sample inhomogeneity. Probable causes are incorrect background assessment or inadequate sample pretreatment.
3. This study has illuminated difficulties associated with the measurement of very old materials. A number of possible sources of error have been identified that need to be investigated further.

ACKNOWLEDGMENTS

We thank Mr. A. R. Watson for generously supplying the kauri timber. Thanks also to Henry Polach for his constructive comments in reviewing the manuscript.

REFERENCES

Rozanski, K. 1991 Report of the International Atomic Energy Agency Consultants' Group Meeting on C-14 Reference Materials for Radiocarbon Laboratories, Feb. 18–20, 1991. Section of Isotope Hydrology, IAEA, Vienna.

Rozanski, K., Stichler, W., Rozanski, R., Scott, E. M., Beukens, R. P., Kromer, B. and van der Plicht, J. 1992 The IAEA ^{14}C Intercomparison Exercise 1990. *In* Long, A. and Kra, R. S., eds., Proceedings of the 14th International ^{14}C Conference. *Radiocarbon* 34(3): 506–519.

Wilson, S. R and Ward, G. K. 1981 Evaluation and clustering of radiocarbon age determinations: Procedures and paradigms. *Archaeometry* 23(1): 19–39.

APPENDIX

Weighted Estimation

Calculations follow Rozanski *et al.* (1992). The "ese($\hat{\mu}$)" term is incorrect in Rozanski *et al.* (1992: 519) and should read:

$$\mathrm{ese}(\hat{\mu}) = \hat{\sigma}\sqrt{(\Sigma(1/W_i^2))^{-1}} \quad .$$

Test of Homogeneity

To test the hypothesis that a series of determinations are consistent, *i.e.*, the same age, determine the test statistic "T", where

$$T_{[1,n]} = \sum_1^n (A_i - A_p)^2/E_i^2$$

where the pooled mean, A_p, for age estimates, $A_i \pm E_i$, is given by

$$A_p = \left\{\sum_1^n A_i/E_i^2\right\} / \left\{\sum_1^n 1/E_i^2\right\}$$

and $T_{[1,n]}$ has a chi-square distribution with (n−1) degrees of freedom. For the series to be consistent, T must be less than $\chi^2_{n-1,0.05}$.

Cluster Analysis

For the clustering exercise, the data has been taken as an instance of Case I (Wilson and Ward 1981). Assuming the data has failed the T test for homogeneity, one determines the optimal split as follows:

1. Order the data points, A_1 .. A_n, by their given mean ages A_i, such that

$$A_1 \le A_2 \le \ldots \le A_n \ .$$

2. Calculate the statistic

$$\Lambda_k = T_{[1,n]} - T_{[1,k]} - T_{[k+1,n]}$$

where k = 1, ..., n−1 and T is as previously defined. The optimal split is given by the value of k that maximizes Λ_k. Subgroups may be further split while $\Lambda \ge 3.84$.