

SYSTEMATIC BIASES IN RESULTS OF THE INTERNATIONAL COLLABORATIVE STUDY AND THEIR PROBABLE SOURCES

MIECZYŚLAW F PAZDUR, ROMUALD AWSIUK, TOMASZ GOSLAR and ANNA PAZDUR

Radiocarbon Laboratory, Institute of Physics, Silesian Technical University
Krzywoustego 2, PL-44-100 Gliwice, Poland

ABSTRACT. Results of the International Collaborative Study show an unexpectedly large scatter of individual dates as well as systematic biases. Very high values of linear correlation coefficients are observed for all results of Stage 2 and for benzene samples of Stage 1. We observed moderate correlations for carbonate samples and the lowest for natural samples of wood and peat of Stage 3. The correlation is practically negligible among results obtained in different stages. The probable reasons for such effects are seen in medium-term changes in the calibration of the counting systems.

INTRODUCTION

The current interlaboratory cross-check project of ^{14}C dating, initiated by participants of the 12th International Radiocarbon Conference in Trondheim, 1985 (Scott *et al* 1986) and supported by more than 50 laboratories, ended in 1988. Preliminary reports, presented by organizers of this study (Scott *et al*, in press; 1989) and detailed lists of coded results available to all participating laboratories lead us to the conclusion that the present status of the ^{14}C dating method is unsatisfactory. The observed scatter in five sets of duplicate samples significantly exceeds 1000 yr, and in the other, 500 yr. Values of correlation coefficients of ^{14}C dates obtained on duplicate samples are greater than 0.9 in four cases, indicating significant systematic biases.

These observations indicate that permanent and well-prepared future projects for improving the accuracy of ^{14}C dating are necessary and urgently needed. To start with such a project, it seems necessary to understand the origins of the errors shown by results of the present study. This report is intended to give some global characteristics of ICS results and to explain observed correlations in terms of laboratory practice.

METHODS AND RESULTS

Coded results obtained in all stages and distributed by organizers of the ICS were used for calculations of linear correlation coefficients among results of duplicate samples and among mean values of duplicates of different samples. With the information provided, we could not distinguish between gas counting (GC), accelerator mass spectrometry (AMS) and liquid scintillation (LS) labs. However, we presume that the LS labs are those using benzene samples in Stage 1. Thus, this type of analysis gives only a general statistical description of the performance of the whole group of participating laboratories. A high correlation coefficient indicates significant bias, *ie*, on average, both compared dates are shifted either towards older or younger values.

Table 1 shows results of calculations of correlations among dates obtained on duplicate samples of the same age. The table contains results for the total group and two subgroups of LS and non-LS labs. We made systematic calculations of correlation coefficients for all possible combinations of mean values obtained on duplicate samples measured in all stages of this study. Table 2 lists correlations among results of the same stage. The highest correlations are for all samples of Stage 2 and for benzene of Stage 1, whereas most correlations in Stage 3 are insignificant. Interstage correlations are slight, and are somewhat more significant for presumed LS laboratories. Table 3 lists the highest values.

TABLE 1
Correlation coefficients of dates obtained on duplicate samples

	Sample	Labs	No.	Correlation	Range (yr)
1	Benzene #1	LS	20	0.926	1300
	Benzene #2	LS	20	0.953	1700
	Carbonate #1	GC & AMS	32	0.738	500
2	Carbonate #2	GC & AMS	32	0.722	850
	Cellulose	All	40	0.910	1900
	Cellulose	LS	16	0.953	1900
	Cellulose	GC & AMS	24	0.763	800
	Algal lithothamnion	All	28	0.858	1800
	Algal lithothamnion	LS	13	0.900	1800
	Algal lithothamnion	GC & AMS	15	0.807	700
	Humic acid	All	17	0.918	550
	Humic acid	LS	6	0.905	250
	Humic acid	GC & AMS	11	0.942	550
	3	Wood #1	All	36	0.168
Wood #1		LS	10	-0.013	500
Wood #1		GC & AMS	26	0.429	400
Wood #2/3		All	35	0.581	700/900
Wood #2/3		LS	10	0.522	
Wood #2/3		GC & AMS	25	0.681	
Shell		All	34	0.718	1500
Shell		LS	11	0.771	1500
Shell		GC & AMS	23	0.625	500
Peat		All	31	0.498	650
Peat		LS	11	0.602	650
Peat	GC & AMS	20	0.370	400	

TABLE 2
Correlations between mean dates of duplicate samples of same stage

Stage	Samples	No.	Correlation
1	Benzene 1/benzene 2	20	0.771
1	Carbonate 1/carbonate 2	32	0.375
2	Cellulose/algal lithothamnion	27	0.825
2	Cellulose/humic acid	16	0.821
2	Algal lithothamnion/humic acid	4	0.799
3	Wood 1/wood 2/3	35	0.635
3	Wood 1/shell	32	0.285
3	Wood 1/peat	29	0.484
3	Wood 2/3/shell	32	0.666
3	Wood 2/3/peat	28	0.545
3	Shell/peat	29	0.195

TABLE 3
Highest values of interstage correlation coefficients

Samples	No.	Correlation
Benzene 1/humic acid	6	0.58
Benzene 2/algal lithothamnion	13	0.50
Benzene 1/wood 1	10	0.43
Benzene 1/wood 2/3	10	0.43
Benzene 1/shell	11	0.48
Carbonate 1/peat	22	0.40
Carbonate 2/peat	22	0.41
Algal lithothamnion/wood 1	21	0.40

DISCUSSION

The correlation coefficients show some regularities, possibly indicating laboratory practice as a source of some systematic errors. First, the highest correlations occur between duplicates (Table 1) and, slightly lower, between mean values of the same stage, whereas correlations between different stages are, in most cases, insignificant. The differences between r values listed in Tables 1 and 2 are not so distinct. Thus, we conclude that systematic biases that cause observed correlations have two main sources. First, and predominant, are undetected medium-term changes of the fundamental physical parameters of counting systems (counting rates of background and modern reference standards) which occur over time intervals of a few months (2 - 4, as deduced from the schedule for completing measurements of a single stage). Second, a less obvious source is chemical composition of sample material, shown by systematic differences between values of correlation coefficients $r(\text{duplicates}) > r(\text{mean dates})$. Although it is not statistically justified procedure, mean values of correlation coefficients between duplicates and between mean dates of duplicates are compared in Table 4. Physically, duplicate samples were taken from the same batch and it seems reasonable to assume that different batches of the same material (benzene or carbonate) contain slightly different admixtures. Higher correlations between duplicates can be associated with failures in two steps of the dating process - preparation and purification of the counting medium and correcting for quenching effects.

TABLE 4
Averaged values of correlation coefficients between duplicates and mean dates of duplicate samples

Stage	Mean correlation coefficients		Relative difference
	Duplicates	Mean dates	%
1 (benzene)	0.94	0.77	18
1 (carbonate)	0.73	0.37	48
2	0.90	0.81	10
3	0.50	0.47	6

Very high correlations observed in results of Stage 2 are associated with larger dispersions of both individual dates and mean dates of duplicate samples of cellulose and algal lithothamnion (Tables 1 and 2). The correlations among results obtained on different samples of this stage, noted also by Scott *et al* (1989), are much higher than in Stages 1 and 3 (Tables 3 and 4). The results of Stage 2 suggest that observed systematic biases are associated with difficulties of obtaining a

sufficiently pure counting medium from partly pretreated samples. This conclusion is supported by the experience of the Gliwice Radiocarbon Laboratory (ICS code 11). In Stage 2, we produced much more deviant results than in other stages of the ICS. We repurified CO₂ from combustion of cellulose and humic acid several times before achieving acceptable purity according to empirical criteria. However, the gas purity parameters were close to the lower limits of acceptability and the counting efficiency corrections (Pazdur, Walanus & Mościcki 1978) were enormously large. The results on cellulose are discordant. The disparity, defined by Scott *et al* (1989), is equal to 2.68 with one date (2388 ± 47 BP) close to the quartile range (2160 - 2370 BP; Scott *et al* 1989) and the second (2603 ± 65 BP) by 350 yr older than the mean value of 36 results of all participating laboratories. Duplicate samples of humic acid yielded consistent results (3678 ± 44 BP, 3662 ± 45 BP, disparity = 0.25) but deviating by ca 280 yr, *ie*, by > 6σ towards older ages. The results of Stage 2, compared with much better results from Stages 1 and 3, suggest that our purification line is effective enough for routine natural samples (carbonate, wood, peat). However, the purification line does not produce sufficiently purified gas from impure samples that are not encountered in routine laboratory practice (chlorine is suspected as the main impurity in samples of Stage 2). Moreover, we realized that the criteria for accepting the purity of the counting gas are too weak, and finally, that the methods used to introduce corrections for decreased counting efficiency do not work correctly for large deviations from the ideal purity of the counting gas.

The validity of the last conclusion seems to be of general concern for both GC and LS labs. Particularly, the highly correlated and scattered results of benzene samples from Stage 1 may be, at least in part, owing to inappropriate quenching corrections and unprecedented shifts of pulse height spectra of laboratory synthesized benzene.

Polach (1989) discusses the enormously large scatter of the results of the benzene sample from Stage 1 and suggests several possible reasons for observed aberrations. He concludes that the results show no reasonable grouping and statistical methods are not sufficient to evaluate such data. The same conclusion applies to results on the second benzene sample of Stage 1 and the humic acid of Stage 2. However, we found similar large scatter and high correlation for cellulose and algal lithothamnion of Stage 2 and natural shell of Stage 3. Figures 1-3 illustrate the dates on duplicates of these samples, indicating presumed LS and non-LS labs. Three outlying laboratories (A, B, C) are clear outliers; two (A, B) are responsible for the large ranges listed in Table 1. After we reject obvious outliers, the remaining dates show reasonable scatter in the interval 400 - 700 yr, *ie*, similar to that found for other samples. Some grouping of results of Stage 2 is visible; no satisfactory clustering occurs for shell results. Figure 1 also shows some shift of LS results towards older ages with respect to non-LS results. The group of results remaining after the rejection of outliers shows no evidence for discrimination between LS and non-LS laboratories.

SUMMARY AND CONCLUSIONS

Our analysis revealed significant correlations among dating results of the same samples or different samples of the same stage, *ie*, among samples processed during relatively short periods of 3-6 months. The main sources of such behavior of the analyzed sets of dates are undetected medium-term changes in the performance of counting systems, which occur for both LS and GC laboratories. The results on natural samples are less correlated and show more satisfactory grouping than those obtained on partly pretreated or laboratory synthesized samples. The LS labs seem to be more prone to producing outlying results.

The results of the ICS, if evaluated on the basis of ranges listed in Table 1, may lead to extreme opinions that after 40 years of improving the radiocarbon dating method, we are still at the starting point of the Libby's solid carbon counter with an accuracy of 300 - 500 yr (Arnold

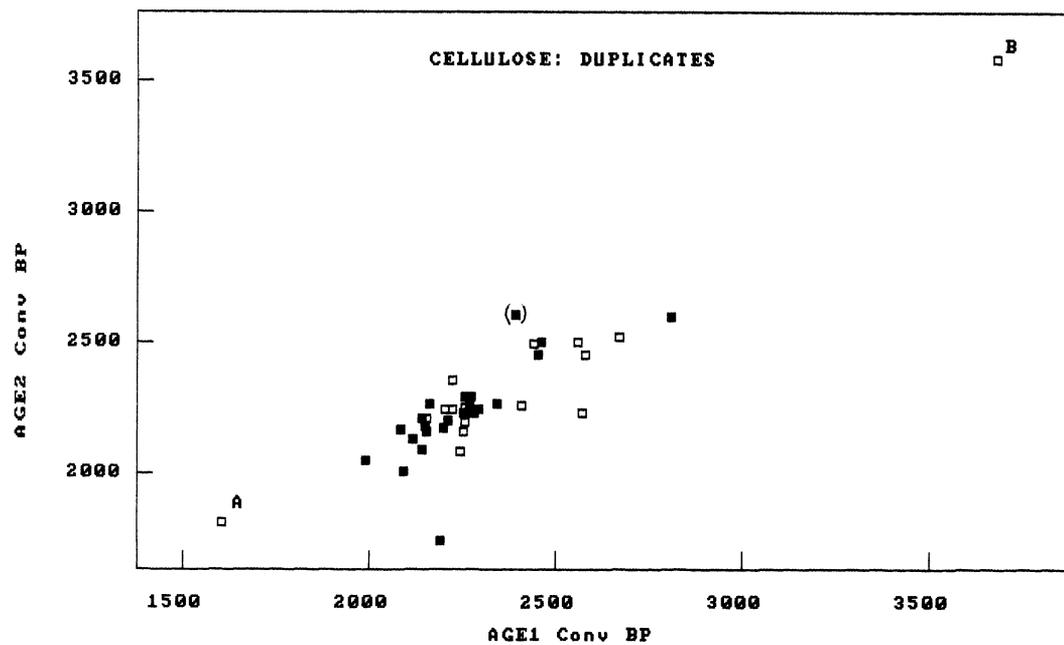


Fig 1. Results obtained on duplicate samples of cellulose, Stage 2. □ = LS labs, ■ = GC & AMS labs. Two distinctly outlying results are marked by A and B; the result of the Gliwice Lab is in parentheses.

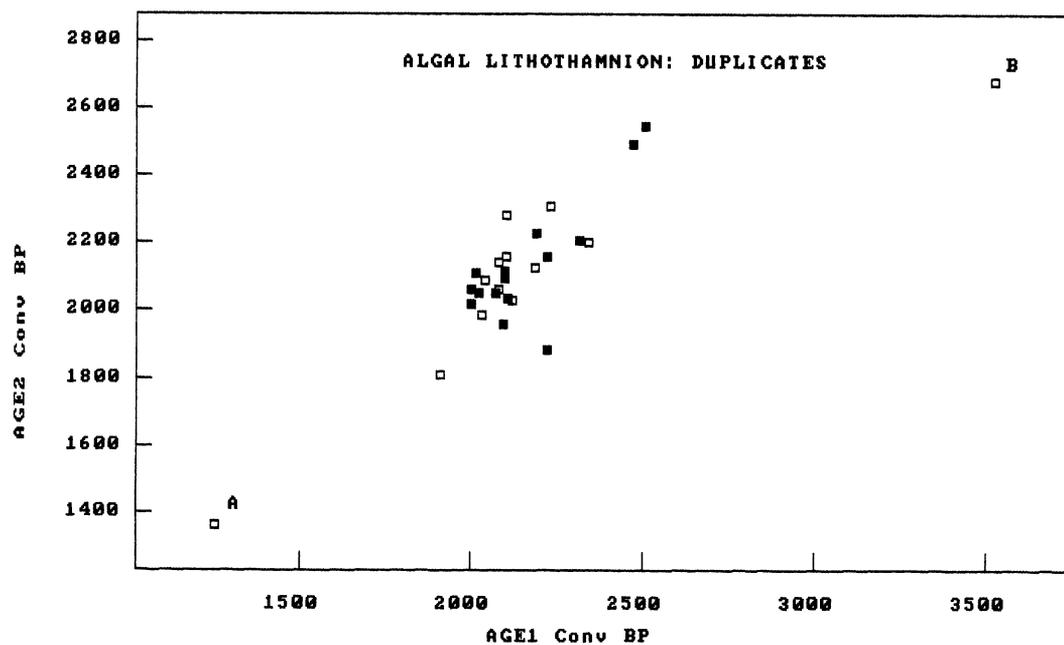


Fig 2. Results obtained on duplicate samples of algal lithothamnion, Stage 2. Two distinctly outlying results (A and B) are produced by the same labs as in Fig 1.

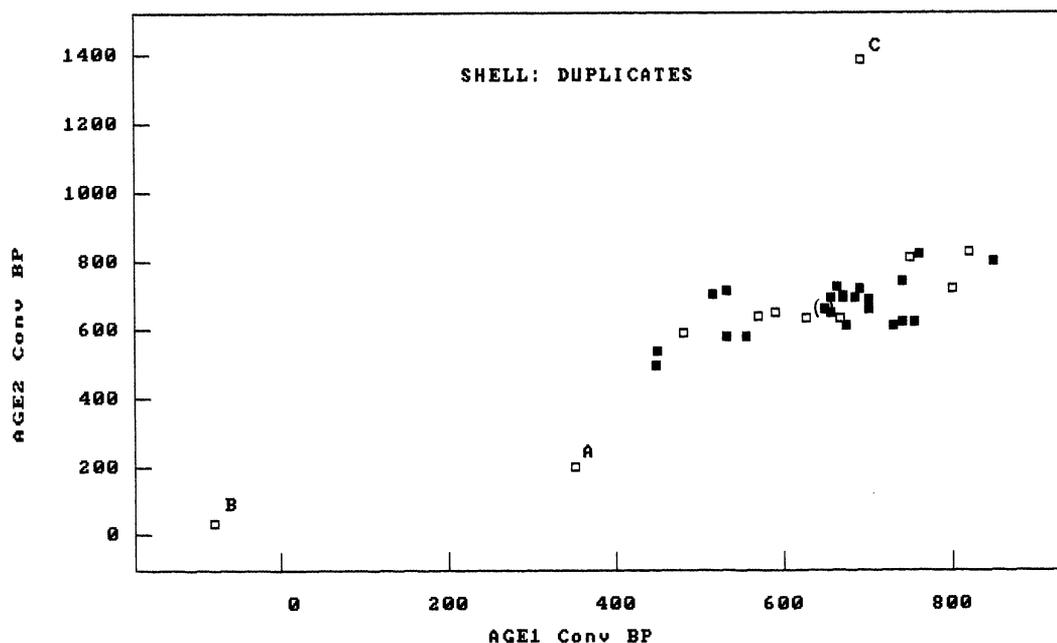


Fig 3. Results obtained on duplicate samples of natural shell, Stage 3. Explanations as in Figs 1 and 2.

& Libby 1951). In any case, future systematic projects are urgently needed, which in our opinion, should be supported by wider participation of laboratories. Very important for all participating labs will be shorter response time (ca 3 months) for 5 - 6 samples dated per stage, combining known-age natural samples (which may be used as secondary laboratory standards) and blind tests for checking for an effect of dating "standard" samples on the quality of results. Results of such a study should not be anonymous and should be regularly published in *RADIOCARBON*.

ACKNOWLEDGMENTS

This study was supported by Central Research Project CPBP 01.06.

REFERENCES

- Arnold, JR and Libby, WF 1951 Radiocarbon dates. *Science* 113(2927): 111-120.
- International Study Group 1982 An inter-laboratory comparison of radiocarbon measurements in tree rings. *Nature* 298: 619-623.
- Pazdur, MF, Walanus, A and Mościcki, W 1978 A method of continuous examination of counting efficiency during measurements of natural radiocarbon by CO₂ filled proportional counter. *Nuclear Instruments & Methods* 151: 541-547.
- Polach, HA 1989 Liquid scintillation ¹⁴C spectrometry: Errors and assurances. In Long, A and Kra, RS, eds, Internatl ¹⁴C conf, 13th, Proc. *Radiocarbon* 31(3): 327-331.
- Scott, EM, Aitchison, TC, Harkness, DD, Baxter, MS and Cook, GT 1989 An interim progress report on Stages 1 and 2 of the International Collaborative Program. In Long, A and Kra, RS, eds, Internatl ¹⁴C conf, 13th, Proc. *Radiocarbon* 31(3): 414-421.
- Scott, EM, Baxter, MS, Aitchison, TC, Harkness, DD and Cook, GT 1986 Announcement of a new collaborative study for intercalibration of ¹⁴C dating laboratories. *Radiocarbon* 28(1): 167-169.
- Scott, EM, Baxter, MS, Harkness, DD, Aitchison, TC and Cook, GT, in press, Recent progress of the International Calibration of Radiocarbon Laboratories. In Mook, WG and Waterbolk, HT, eds, Internatl symposium Archaeology and ¹⁴C, 2nd, Proc. *PACT*.