RADIOCARBON, Vol 53, Nr 2, 2011, p 277–288

EXAMINING THE INHERENT VARIABILITY IN ΔR: NEW METHODS OF PRESENTING ΔR VALUES AND IMPLICATIONS FOR MRE STUDIES

N Russell[1] • G T Cook[1,2] • P L Ascough[1] • E M Scott[3] • A J Dugmore[4]

ABSTRACT. Currently, there is significant ongoing research into the temporal and spatial variability of marine radiocarbon reservoir effects (MREs) through quantification of ΔR values. In turn, MRE studies often use large changes in ΔR values as proxies for changes in ocean circulation. ΔR values are published in a variety of formats with variations in how the errors on these values are calculated, making it difficult to identify trends or to compare values, unless the method of calculating the ΔR is explicitly described or all of the data are made available in the publication. This paper demonstrates the large range in ΔR values (+34 to –122) that can be obtained from a single, secure archaeological context when using the multiple paired sample approach, despite the fact that the terrestrial entities were of statistically indistinguishable ^{14}C ages, as were the marine samples. This demonstrates the inherent variability in the ΔR calculations themselves and we propose that, together with calculation of mean ΔR, the distribution of ΔR values should be displayed, e.g. as histograms in order to illustrate the full data range. This spread is only apparent when employing a multiple paired sample approach as the uncertainty derived on a single pair of samples, taking account only of the errors on the individual ^{14}C ages, will never truly represent the overall variability in ΔR that results from the intrinsic variability in the population of ^{14}C ages in samples that might have been used. Consequently, ΔR values and the associated uncertainty calculated from single pairs should be treated with some caution. We propose that, where possible, when using paired archaeological samples, that a multiple paired approach should be employed as it will test the context security of the material used in the ΔR calculations. When summarizing the values by the weighted average, we also propose that the standard error for predicted values should be employed as this will fully encompass the uncertainty of a future ΔR calculation, using different samples for a similar time and location. Finally, we encourage future publishing of ΔR values using the histogram format, making all of the data available. This will help ensure that ΔR values are comparable across the literature and should provide a framework for standardization of publication methods.

INTRODUCTION

The marine radiocarbon reservoir effect (MRE) manifests itself as a ^{14}C age offset at any point in time between samples formed in the terrestrial biosphere (which is in equilibrium with the atmosphere) and samples formed in the marine environment (Stuiver et al. 1986). This offset is variable on both a temporal and spatial basis (Stuiver et al. 1986; Stuiver and Braziunas 1993) and exists because of the extended mean residence time of ^{14}C in the oceans, particularly in the deep oceans. During circulation within deep waters that are separated from contact with atmospheric CO_2, radioactive decay of ^{14}C atoms results in deep-ocean (about >100 m depth) depletion relative to the contemporaneous atmosphere (Stuiver and Braziunas 1993). Therefore, as a result of the eventual upwelling of deep waters, the surface oceans (about 0–50 m depth) are also depleted in ^{14}C relative to the atmosphere, although to a lesser extent than the deep ocean. Because of the known variability in the MRE, current research themes in the Northern Hemisphere (e.g. Reimer et al. 2002; Ascough et al. 2004, 2005a,b, 2006, 2007a,b, 2009; Cage et al. 2006; Mangerud et al. 2006; Butler et al. 2009; Olsen et al. 2009; Soares and Martins 2009, 2010; Russell et al. 2010) have focused on refining MRE values for specific locations and periods in time. The most common approach to quantifying these variations is the determination of ΔR values, where a ΔR value represents a regional offset from the global average surface water MRE (for which ΔR = 0) (Stuiver et al. 1986; Stuiver and Braziunas 1993). If the ΔR is positive, this represents an increased MRE for the region compared with the global average, and vice versa for negative values. The generation of site-specific MRE (and

[1]Scottish Universities Environmental Research Centre, Rankine Avenue, Scottish Enterprise Technology Park, East Kilbride G750QF, Scotland.
[2]Corresponding author. Email: g.cook@suerc.gla.ac.uk.
[3]Department of Statistics, University of Glasgow, Glasgow G128QQ, Scotland.
[4]Institute of Geography, School of Geosciences, University of Edinburgh, Old High School, Infirmary Street, Edinburgh EH89XP, Scotland

therefore ΔR) values have in turn been used as proxies for changes in localized oceanic regimes (e.g. Kennett et al. 1997; Kovanen and Easterbrook 2002).

The potential uncertainties inherent in deriving ΔR values fall into 3 main categories: 1) the samples used to generate the ^{14}C ages from which the ΔR values will be calculated; 2) the generation of the ^{14}C ages and their associated errors; and 3) the actual calculation of the ΔR value, and the number of ^{14}C ages used in its calculation. This paper assesses the degree to which apparent shifts in ΔR values can be explained by examining the degree of variability inherent in the production of single (mean) ΔR values, even when based upon multiple paired samples. In so doing, this work challenges the reproducibility of ΔR values that are derived using single pairs of terrestrial and marine ^{14}C ages in other methodological approaches. The paper first discusses our own calculation methods before proposing a best-practice method of publishing ΔR determinations and associated errors in order to incorporate the variability demonstrated.

METHODS OF CALCULATING ΔR

The concept of identifying localized ΔR variations is well discussed by Stuiver et al. (1986) and Stuiver and Braziunas (1993). A ΔR value is essentially calculated using a sample of marine carbon for which the terrestrial/atmospheric ^{14}C age is known, or can be established with a high degree of confidence. A modeled marine ^{14}C age is then derived for this sample, which is achieved by converting the terrestrial/atmospheric ^{14}C age ±1 σ to a modeled marine age via interpolation between the IntCal04 atmospheric curve and the Marine04 curve (Reimer et al. 2004; Hughen et al. 2004). ΔR is the difference between this modeled marine ^{14}C age and the measured ^{14}C age of the marine carbon sample (Figure 1). The 1-σ error on the ΔR values is calculated by the propagation of errors shown in Equation 1.

$$\sigma_{\Delta R} = \sqrt{(\sigma_w + \sigma_m)^2} \quad (1)$$

where ($\sigma_{\Delta R}$) is the 1-σ error for the ΔR determination, σ_w is the error on the measured marine age, and σ_m is the error on the modeled marine age.

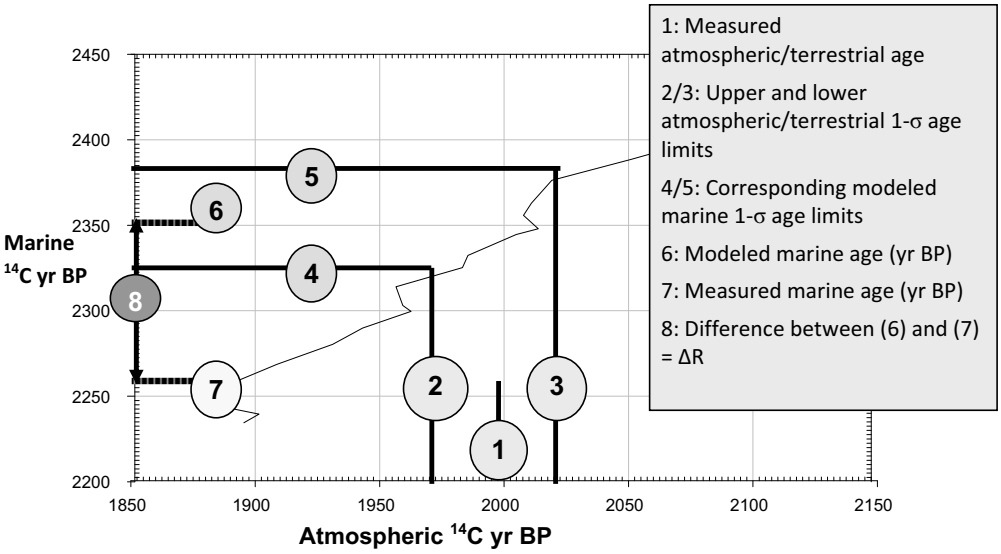


Figure 1 Graphical representation of the determination of a ΔR value showing interpolation of atmospheric and marine ages

https://doi.org/10.1017/S003382220005654X Published online by Cambridge University Press

A variety of methodological approaches are used to obtain suitable ^{14}C ages for calculation of ΔR. These include measurement of: 1) known-age samples from museum collections; 2) samples associated with onshore/offshore tephra isochrones; and 3) paired samples from secure archaeological contexts. These methods are all discussed in detail by Ascough et al. (2005a). More recently, Butler et al. (2009) have used samples of *Arctica islandica* from their "annually resolved multi-centennial (489-year), absolutely aged" master chronology. While this is potentially extremely useful in providing a continuous record of ΔR values, it is currently limited in time to a 489-yr period (late- and Post-Medieval periods) and in the future will be limited to locations where *Arctica islandica* shells will be found in numbers sufficient to duplicate the chronological work. Ascough et al. (2005a) supported an approach involving multiple paired samples, where the terrestrial and marine ^{14}C age used to calculate ΔR is based upon multiple samples of both material types, using short-lived species from secure archaeological contexts (i.e. where there is a high degree of confidence that all organisms within the deposit have the same time of death). Again, this technique is temporally limited, only providing snapshots in time of ΔR values, but these snapshots are available for time periods of importance in archaeology. Secure archaeological contexts are selected through close consultations with site excavators and excavation reports to identify contexts containing marine material (generally mollusk shell) and terrestrial entities (carbonized grains, herbivore bone, etc.) that have been relatively unaffected by postdepositional disturbance (e.g. Ascough et al. 2007a, 2009). Ideally, the contexts should contain a high volume of sample material and have well-defined boundaries to ensure the samples were deposited at the same time. Selecting several entities of each sample type helps reinforce context security by producing ^{14}C ages that can be subjected to chi-squared (χ^2) testing to demonstrate that they are statistically indistinguishable from each other. The χ^2 test determines whether each sample within a group is statistically indistinguishable at 95% confidence from the remainder and therefore can be considered contemporary. The critical value for the χ^2 test differs according to the number of measurements within a group and this value is compared to the *T* statistic for each group to determine whether the samples are statistically indistinguishable (Ward and Wilson 1978). The calculation of the *T* statistic is shown in Equation 2:

$$T = \sum \frac{(t_i - t)^2}{\sigma_i^2} \tag{2}$$

where t is the weighted mean of the ^{14}C age group, t_i is the individual ^{14}C measurement, and σ_i is the error on the individual measurement.

Where the *T* statistic for the group is less than the critical value, the samples are considered to be contemporaneous, whereas when the *T* statistic is greater than the critical value the samples are not considered to be internally coherent and consequently the ages are subjected to more intense scrutiny (see Ascough et al. 2007a, 2009). The method of calculating the *T* statistic means that samples contributing significantly to *T*, which therefore are non-contemporaneous with the remainder of the multiple samples, can be identified and excluded from ΔR calculations as appropriate.

^{14}C ages that pass the χ^2 test are then used to calculate ΔR. This is achieved by converting the terrestrial ^{14}C ages to modeled marine ^{14}C ages, allowing direct comparison with the measured marine ^{14}C ages from the contemporaneous marine samples. In cases where samples do not pass the χ^2 test, a judgement call has to be made on whether the samples from this context are in fact suitable for determining a ΔR value. Using the multiple paired sample approach, it is possible to formulate the problem of determining the variability in the ΔR value, in terms of a resampling strategy by which we mean a procedure that draws many samples from some (pseudo-)population (i.e. bootstrapping). For each draw, we compute a test statistic, in this case ΔR and the resulting set of ΔR values consti-

https://doi.org/10.1017/S003382220005654X Published online by Cambridge University Press

tutes the sampling distribution (often called a reference distribution) of that statistic, and we can use that sampling (reference) distribution to draw inferences about ΔR.

By using every possible pairing when all samples pass the χ^2 test, 16 estimates of ΔR can be calculated for a context containing 4 terrestrial and 4 marine entities. A weighted mean is then calculated to allow the publication of a single representative value that places more weight on the values with lower associated errors. The ΔR values are then typically published using the mean value and the associated error on the mean. This paper proposes that the associated error on the mean is not always fully representative of the inherent variability within the set of ΔR values produced using the multiple paired sample approach.

SOURCES OF UNCERTAINTY IN THE ΔR CALCULATION

In order to address the issues in the production of an appropriate error term for ΔR calculations, sources of error and uncertainty associated with the determination of a ΔR value have been identified as follows:

1) Uncertainty associated with the identification of suitable samples: These are well discussed by Ascough et al. (2005a).

2) Errors associated with the ^{14}C analysis procedures: These include (i) contamination. This is an unquantifiable error that can derive from contamination at any stage throughout the entire laboratory process and incorporates any human error in the sample preparation. As far as possible, this can be identified by reference to known-age standards measured in the same batch as the unknown samples, although 100% elimination of contamination can never be guaranteed. (ii) Inappropriate errors placed on the age measurements: This has to be a realistic estimate of the error and should not be based solely on counting statistics. At SUERC, the counting error is based on overall statistics of approximately 3‰ or better, but the final quoted error associated with a measurement is limited by the standard deviation on a series of standards of known activity, of which there are typically 13 in a batch. We use a Scots pine sample collected from the Garry Bog, Northern Ireland, as the secondary "known-age" standard. This has an in-house laboratory code of BC and has been dendrodated to 3299–3257 BC, with an average ^{14}C age of 4471 BP. This sample was used in the Fourth International Radiocarbon Intercomparison Study where its code was FIRI I. The results from the study gave a consensus value of 4485 ± 5 BP (Scott 2003). The standards data for the batch that we use to illustrate the problems in defining a ΔR and a representative error are given in Table 1. The site for which we are defining the ΔR is Archerfield, which is situated on the east coast of Scotland.

If the error on a sample measurement is an underestimate, this could lead to samples being falsely identified by the χ^2 test as non-contemporaneous, while overestimation of the error could have the opposite effect. Using the data in Table 1, the standard deviation on the 13 measurements would be the limiting factor on the error associated with sample measurements, i.e. unknown samples measured to 3‰ counting statistics would be assigned an error of 32 yr. (iii) Rounding of ages and errors: The convention at SUERC and generally in the ^{14}C community has been to round ages (up or down) to the nearest multiple of 5 yr and round errors up to the next multiple of 5 yr. Sample measurements from the batch in Table 1 would therefore be reported with an error of ±35 yr.

The simplest way to demonstrate the effect that rounding of ^{14}C ages and their errors can have on ΔR values is to use a worked example. Previous MRE studies on the North Sea coast of Scotland (Russell et al. 2010) produced 8 terrestrial and 8 marine samples from the site of Archerfield (context 90) in East Lothian, Scotland. The results of this worked example are shown in Table 2.

https://doi.org/10.1017/S003382220005654X Published online by Cambridge University Press

Table 1 Standards data for the relevant batch.

Sample code*	Age (yr BP)	Counting statistics error (1 σ)
BC1226	4551	24
BC1227	4461	24
BC1228	4490	25
BC1229	4522	25
BC1230	4470	24
BC1231	4514	25
BC1232	4477	26
BC1233	4501	24
BC1234	4462	24
BC1235	4488	24
BC1236	4535	24
BC1237	4439	21
BC1238	4474	24
Mean ±1 std dev	*4491 ± 32*	

Table 2 ^{14}C and $\delta^{13}C$ results for marine and terrestrial samples (with and without rounding) from Archerfield 90 (data from Russell et al. 2010).

SUERC nr	Sample material	^{14}C age (BP) ±1 σ (no rounding)	^{14}C age (BP) ±1 σ (conventional publication with rounding)	$\delta^{13}C$ (‰) relative to PDB ±0.1‰
19669	Limpet (*Patella vulgata*)	823 ± 32	825 ± 35	0.1
19670	Limpet (*Patella vulgata*)	830 ± 32	830 ± 35	–2.4
19671	Limpet (*Patella vulgata*)	912 ± 32	910 ± 35	0.7
19675	Limpet (*Patella vulgata*)	897 ± 32	895 ± 35	–1.8
19676	Winkle (*Littorina littorea*)	910 ± 32	910 ± 35	1.9
19677	Winkle (*Littorina littorea*)	840 ± 32	840 ± 35	1.2
19678	Winkle (*Littorina littorea*)	932 ± 32	930 ± 35	0.5
19679	Winkle (*Littorina littorea*)	940 ± 32	940 ± 35	1.0
	Mean ±1 std dev	*886 ± 47*	*885 ± 46*	
19680	Barley (*Hordeum vulgare*)	497 ± 32	495 ± 35	–22.4
19681	Barley (*Hordeum vulgare*)	471 ± 32	470 ± 35	–23.1
19685	Barley (*Hordeum vulgare*)	502 ± 32	500 ± 35	–24.0
19686	Barley (*Hordeum vulgare*)	493 ± 32	495 ± 35	–24.1
19687	Oat (*Avena* sp.)	485 ± 32	485 ± 35	–25.3
19688	Oat (*Avena* sp.)	502 ± 32	500 ± 35	–24.9
19689	Oat (*Avena* sp.)	455 ± 32	455 ± 35	–25.0
19690	Oat (*Avena* sp.)	527 ± 32	525 ± 35	–24.1
	Mean ±1 std dev	*492 ± 22*	*491 ± 21*	

In this data set, using the unrounded ages and errors leads to the exclusion of 1 marine ^{14}C age (SUERC-19669) from the ages used for ΔR calculation, on the basis that this age leads to an unacceptably high *T* statistic within the χ^2 test. If the rounded ages and errors were used, this age would be included in the group used for ΔR calculation, as the rounded ages would all pass the χ^2 test.

The ΔR values calculated from the various pairing of terrestrial/marine ^{14}C ages ranged from ΔR = +34 ± 40 to ΔR = –122 ± 42. Weighted mean values and associated errors were calculated using the rounded and unrounded data sets, producing ΔR values of –33 ± 6 (unrounded data) and –42 ± 6

https://doi.org/10.1017/S003382220005654X Published online by Cambridge University Press

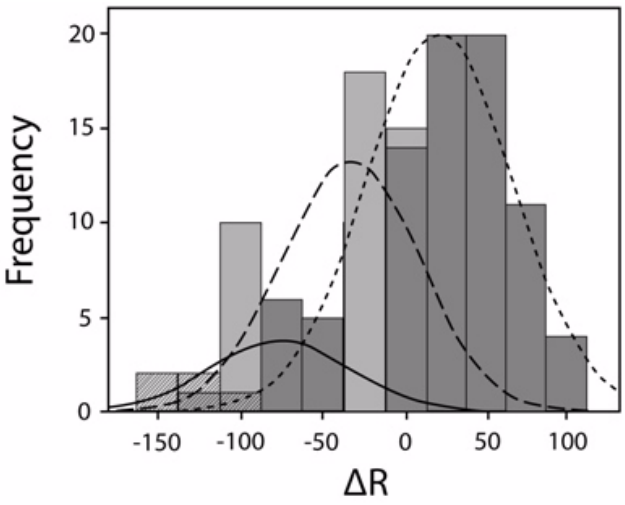
(rounded data), which in this instance are statistically indistinguishable at 2 σ. Therefore, in this example, although the use of rounded versus unrounded ^{14}C ages results in different groups of marine ^{14}C ages included for ΔR calculation, the ΔR values actually calculated based on rounded versus unrounded data are not significantly different. However, this will not be the case for all data sets, and the key point is that it is possible that under some circumstances, statistically different ΔR values could arise depending upon whether rounded or unrounded ^{14}C ages were used in calculation.

3) Uncertainties associated with the ΔR value: Two important points emerge from the above: 1) Is the standard error on the mean sufficient to encompass any future individual measurements made on samples from the same context? If not, then the quoted error is not sufficiently robust. For example, the unrounded data produce errors in ΔR values (calculated as per Equation 1) in the range 37–40 yr. ΔR values at the extremes of the ranges such as ΔR = –118 ± 40 when compared to the mean ΔR of –33 ± 6 would not pass the χ^2 test. 2) We limit the error on a measurement in accordance with the variability on a set of standards, which, for this batch, had a standard deviation of 32 ^{14}C yr (Table 1). In addition, we are assuming that samples within a context are inherently of the same age. This can be justified for the terrestrial samples as the standard deviation is 21 ^{14}C yr for both the unrounded and rounded data. However, for the marine data, the standard deviations are 43 ^{14}C yr for unrounded data and 47 ^{14}C yr for rounded. Therefore, there is additional variability here that is either associated with the age of the samples or the integrity of the context. We would propose a conservative approach of using the standard deviation on the 8 marine samples as the limiting factor on the error on the marine ages.

NEW METHODS

Publishing the mean value from ΔR calculations for each context is commonplace (Reimer et al. 2002; Ascough et al. 2004, 2005, 2006, 2007, 2007a, 2009; Weisler et al. 2009; Soares and Martins 2010) and provides a concise method of presenting the values. However, in order to understand the true spread of values as a more appropriate measure of variability, a useful method is to employ a histogram to display the variability in ΔR values derived from multiple pairs of terrestrial and marine samples (i.e. the range of 16 ΔR values calculated from individual pairings of 4 terrestrial and 4 marine sample ^{14}C ages). The histogram should be illustrated alongside the mean value (Figure 2). For the purposes of this paper, histograms were constructed using Minitab® 16 using the Normal curves to allow assessment of indeed whether the distribution of ΔR values is Normal. To demonstrate this, 3 sites were chosen from a previous publication on ΔR variability (Russell et al. 2010). The ΔR values were recalculated using the method of limiting errors described above and the spread of values as displayed in Table 3 were plotted in the histogram in Figure 2. The mean ΔR values with small associated errors at 2 σ (Archerfield 90: ΔR = –42 ± 10; Arbroath Abbey: ΔR = 7 ± 14; and 16-18 Netherkirkgate: ΔR = –95 ± 28) were previously interpreted as indicating water bodies of different ^{14}C specific activities (Russell et al. 2010).

Publishing ΔR values in this manner allows for a better understanding of the population that the mean value relates to, and the possible variability in the ΔR value. This method allows all of the data from the multiple calculations in a multiple paired sample approach to be laid bare and interpreted with appropriate caution. Using the data from the 3 sites in Figure 2, it can be seen that although the mean values for the sites vary from ΔR = +15 to ΔR = –76 using the Normal probability density curves (and histograms), there is considerable overlap, suggesting that the populations are not as distinguishable as the previously published mean values and associated errors had suggested.



Variable	ΔR Arithmetical Mean	StDev	N
Archerfield	-34	42	56
16-18 Netherkirkgate	-76	42	16
Arbroath Abbey	+20	45	90

Figure 2 Direct comparison of the distribution of ΔR values from 3 sites

The standard error on the mean represents how precisely we "know" the population mean value, but if instead we actually wish to make a statement about a future (hypothetical ΔR value) calculated from this population, then we also need to include a measure of the variability within that population (which would be the standard deviation). This point was illustrated using the case study at Archerfield where the error on the weighted mean was only ± 10, giving false security in the refinement available of ΔR values from this context, given that the values ranged from ΔR = +34 to ΔR = –122. We therefore propose the use of the standard error for predicted values (Equation 3) in order to represent the true variability inherent in ΔR calculations from a multiple paired sample approach:

$$\sigma = \sqrt{(x^2 + y^2)} \quad (3)$$

where x is the error on the weighted mean and y is the standard deviation on the ΔR values.

Figure 3 shows the previously published weighted mean ΔR values and associated errors compared with the new method using unrounded ages and the standard error for predicted values. Errors on the mean are represented at 2 σ. Weighting the mean ΔR values rather than displaying the arithmetical means from the normalized histograms leads to a large shift in the ΔR value from the site of 16-18 Netherkirkgate. The mean value is shown in Figure 2 as –76, whereas the weighted mean value as shown in Figure 3 is –98. Using the weighted mean takes into account the very small errors associated with the lower ΔR values calculated from sample T4 (Table 3), thus weighting the mean towards a more negative value.

https://doi.org/10.1017/S003382220005654X Published online by Cambridge University Press

Table 3 All possible pairings of ΔR for the 3 sites and weighted mean values for ΔR alongside standard error for predicted values at 1 σ.

Sample pairing		ΔR	Error	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error
Archerfield 90															
T1	M1	–101	49	T2	M1	–86	48	T3	M1	–104	49	T4	M1	–99	49
	M2	–19	49		M2	–4	48		M2	–22	49		M2	–17	49
	M3	–34	49		M3	–19	48		M3	–37	49		M3	–32	49
	M4	–21	49		M4	–6	48		M4	–24	49		M4	–19	49
	M5	–91	49		M5	–76	48		M5	–94	49		M5	–89	49
	M6	1	49		M6	16	48		M6	–2	49		M6	3	49
	M7	9	49		M7	24	48		M7	6	49		M7	11	49
T6	M1	–104	49	T7	M1	–75	47	T8	M1	–118	48				
	M2	–22	49		M2	7	47		M2	–36	48				
	M3	–37	49		M3	–8	47		M3	–51	48				
	M4	–24	49		M4	5	47		M4	–38	48				
	M5	–94	49		M5	–65	47		M5	–108	48				
	M6	–2	49		M6	27	47		M6	–16	48				
	M7	6	49		M7	35	47		M7	–8	48				
Weighted mean ΔR = –33. Standard error for predicted values = 43.															
16-18 Netherkirkgate															
T1	M1	–77	62	T2	M1	–57	65	T3	M1	–87	64	T4	M1	–150	36
	M2	–40	62		M2	–20	65		M2	–50	64		M2	–113	36
	M3	–86	62		M3	–66	66		M3	–96	65		M3	–159	35
	M4	–40	62		M4	–20	66		M4	–50	65		M4	–113	35
Weighted mean ΔR = –98. Standard error for predicted values = 44.															
Arbroath Abbey															
T1	M1	3	81	T2	M1	10	78	T3	M1	5	80	T4	M1	5	80
	M2	24	81		M2	31	78		M2	26	80		M2	26	80
	M3	40	81		M3	47	78		M3	42	80		M3	42	80
	M4	10	81		M4	17	78		M4	12	80		M4	12	80

Table 3 All possible pairings of ΔR for the 3 sites and weighted mean values for ΔR alongside standard error for predicted values at 1 σ. *(Continued)*

Sample pairing		ΔR	Error	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error
	M5	54	81		M5	61	78		M5	56	80		M5	56	80
	M6	28	81		M6	35	78		M6	30	80		M6	30	80
	M7	–42	81		M7	–35	78		M7	–40	80		M7	–40	80
	M8	17	81		M8	–10	78		M8	–15	80		M8	–15	80
	M9	–86	81		M9	–79	78		M9	–84	80		M9	–84	80
	M10	–22	81		M10	–15	78		M10	–20	80		M10	–20	80
T5	M1	19	75	T6	M1	45	72	T7	M1	23	74	T8	M1	42	74
	M2	40	75		M2	66	72		M2	44	74		M2	63	74
	M3	56	75		M3	82	72		M3	60	74		M3	79	74
	M4	26	75		M4	52	72		M4	30	74		M4	49	74
	M5	70	75		M5	96	72		M5	74	74		M5	93	74
	M6	44	75		M6	70	72		M6	48	74		M6	67	74
	M7	–26	75		M7	0	72		M7	–22	74		M7	–3	74
	M8	–1	75		M8	25	72		M8	3	74		M8	22	74
	M9	–70	75		M9	–44	72		M9	–66	74		M9	–47	74
	M10	–6	75		M10	20	72		M10	–2	74		M10	17	74
T9	M1	61	69												
	M2	82	69												
	M3	98	69												
	M4	68	69												
	M5	112	69												
	M6	86	69												
	M7	16	69												
	M8	41	69												
	M9	–28	69												
	M10	36	69												

Weighted mean ΔR = 22. Standard error for predicted values = 45.

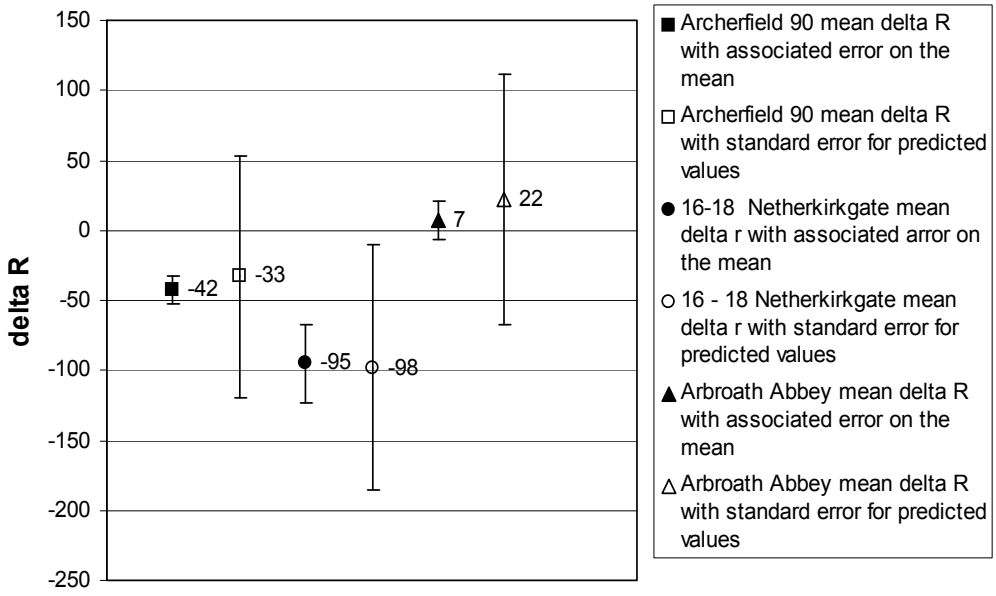


Figure 3 Comparison of ΔR values showing error on the mean (filled symbols) (Russell et al. 2010) and standard error for predicted values (empty symbols).

It can be seen that when using the error on the mean, there is no overlap even at 2 σ and therefore the values could be interpreted as significantly different. However, using the standard error for predicted values results in significant overlap at 2 σ, suggesting that these values are indistinguishable at this level of confidence. Using a much larger error on the mean values may not be desirable but offers a more realistic estimate of the range in which future calculations of ΔR values for these sites may lie. Using the standard error for predicted values represents the true variability inherent within the ΔR calculation itself, as well as providing better information on the prediction and comparability of future values. This is important when considering that ΔR values are often used as proxy indicators for specific ocean ^{14}C activity and shifts in oceanic regimes that may force such a change (e.g. Kennett et al. 1997; Kovanen and Easterbrook 2002). Using a larger error term such as the standard error for predicted values may result in an increased overlap between ΔR values, meaning that the values are no longer significantly different and therefore conclusions on oceanic or climatic proxies cannot be drawn. This may lead to the reinterpretation of currently available ΔR values for global ocean waters.

CONCLUSIONS

It is our opinion that all ^{14}C data used in the calculation of ΔR values should be in the raw form with no rounding introduced and that the errors on the measurements must be realistic and based on replicate measurements of "in-house" standards or a similar regime. It is also our suggestion that using multiple paired samples is the best approach, (a) because each group of marine and terrestrial samples is subjected to a χ^2 test to demonstrate that they are contemporary and this will give confidence that the samples used to calculate ΔR are from secure contexts and that the terrestrial and marine samples are therefore contemporary in age and (b) because this will give the best indication of the likely variability in ΔR values that could be expected from the context. Publishing the full data set of pairings used to calculate ΔR and/or using histograms can help give a better representation of the

https://doi.org/10.1017/S003382220005654X Published online by Cambridge University Press

variability inherent in the calculation and the level of refinement realistically achievable. Of course, a mean ΔR value and an associated error are required when calibrating unknown samples. We suggest that the weighted mean should be employed and that the most appropriate error to use is the standard error for predicted values, which encompasses both the standard deviation of the distribution of ΔR values as well as the associated error on the mean. By standardizing publication methods, ΔR values can be used more accurately by all, and the appropriate conclusions of what significant shifts in ΔR may or may not signify.

ACKNOWLEDGMENTS

NR thanks NERC (Grant nr NE/F002211/1) and Historic Scotland for studentship support. Thanks are also given to the staff of the SUERC Radiocarbon Dating and AMS laboratories for ^{14}C measurements.

REFERENCES

Ascough PL, Cook GT, Dugmore AJ, Barber J, Higney E, Scott EM. 2004. Holocene variations in the Scottish marine radiocarbon reservoir effect. *Radiocarbon* 46(2):611–20.

Ascough P, Cook GT, Dugmore AJ. 2005a. Methodological approaches to determining the marine radiocarbon reservoir effect. *Progress in Physical Geography* 29: 532–47.

Ascough PL, Cook GT, Dugmore AJ, Scott EM, Freeman SPHT. 2005b. Influence of mollusc species on marine ΔR determinations. *Radiocarbon* 47(3):433–40.

Ascough P, Cook G, Church MJ. Dugmore AJ, Arge SV, McGovern TH. 2006. Variability in North Atlantic marine radiocarbon reservoir effects at c.1000 AD. *The Holocene* 16(1):131–6.

Ascough PL, Cook GT, Dugmore AJ, Scott EM. 2007a. The North Atlantic marine reservoir effect in the Early Holocene: implications for defining and understanding MRE values. *Nuclear Instruments and Methods in Physics B* 259(1):438–47.

Ascough PL, Cook GT, Church MJ, Dugmore AJ, McGovern TG, Dunbar E, Einarsson Á, Friðriksson A, Gestsdóttir H. 2007b. Reservoirs and radiocarbon: ^{14}C dating problems in Mývatnssveit, northern Iceland. *Radiocarbon* 49(2):947–61.

Ascough P, Cook GT, Dugmore AJ. 2009. North Atlantic marine ^{14}C reservoir effects: implications for late-Holocene chronological studies. *Quaternary Geochronology* 4(3):171–80.

Butler PG, Scourse JD, Richardson CA, Wanamaker AD, Bryant CL, Bennell JD. 2009. Continuous marine radiocarbon reservoir calibration and the ^{13}C Suess effect in the Irish Sea: results from the first multi-centennial shell-based marine master chronology. *Earth and Planetary Science Letters* 279(3–4):230–41.

Cage AG, Heinemeier J, Austin WEN. 2006. Marine radiocarbon reservoir ages in Scottish coastal and fjordic waters. *Radiocarbon* 48(1):31–43.

Hughen KA, Baillie MGL, Bard E, Beck JW, Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Kromer B, McCormac G, Manning S, Bronk Ramsey C, Reimer PJ, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE. 2004b. Marine04 marine radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1059–86.

Kennett DJ, Ingram L, Erlandson JM, Walker P. 1997. Evidence for temporal fluctuations in marine radiocarbon reservoir ages in the Santa Barbara Channel, southern California. *Journal of Archaeological Science* 24(11):1051–9.

Kovanen DJ, Easterbrook DJ. 2002. Paleodeviations of radiocarbon marine reservoir values for the northeast Pacific. *Geology* 30(3):243–6.

Mangerud J, Bondevik S, Gulliksen S, Hufthammer KA, Hoisaeter T. 2006. Marine ^{14}C reservoir ages for 19th century whales and molluscs from the North Atlantic. *Quaternary Science Reviews* 25(23–24):3228–45.

Olsen J, Rasmussen P, Heinemeier J. 2009. Holocene temporal and spatial variations in the radiocarbon reservoir age of three Danish fjords. *Boreas* 38:458–70.

Reimer PJ, McCormac FG, Moore J, McCormick F, Murray EV. 2002. Marine radiocarbon reservoir corrections for the mid- to late Holocene in the eastern subpolar North Atlantic. *The Holocene* 12(2):129–35.

Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Hogg AG, Hughen KA, Kromer B, McCormac G, Manning S, Bronk Ramsey C, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE. 2004. IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–58.

Russell N, Coo GT, Ascough PL, Dugmore AJ. 2010. Spatial variation in the marine radiocarbon reservoir effect throughout the Scottish post-Roman to Late Medieval period: North Sea values (500–1350 BP). *Radiocarbon* 52(2–3):1166–81.

Scott EM. 2003. The Third International Radiocarbon Intercomparison (TIRI) and The Fourth International Intercomparison (FIRI). *Radiocarbon* 45(2):135–328.

Soares AMM, Martins JMM. 2009. Radiocarbon dating of marine shell samples. The marine radiocarbon reservoir effect of coastal waters off Atlantic Iberia during Late Neolithic and Chalcolithic periods. *Journal of Archaeological Science* 36(12):2875–81.

Soares AMM, Martins JMM. 2010. Radiocarbon dating of marine samples from Gulf of Cadiz: the reservoir effect. *Quaternary International* 221(1–2):9–12.

Stuiver M, Braziunas TF. 1993. Modeling atmospheric ^{14}C influences and ^{14}C ages of marine samples to 10,000 BC. *Radiocarbon* 35(1):137–89.

Stuiver M, Pearson GW, Braziunas T. 1986. Radiocarbon age calibration of marine samples back to 9000 cal yr BP. *Radiocarbon* 28(2):980–1021.

Ward GK, Wilson SR. 1978. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20(1):19–31.

Weisler M, Hua Q, Zhao J-X. 2009. Late Holocene ^{14}C marine reservoir corrections for Hawaii derived from U-series dated archaeological coral. *Radiocarbon* 51(3):955–68.