Russell, N., Cook, G. , Ascough, P. , Scott, E.M. , and Dugmore, A.J. (2012) *Examining the inherent variability in ΔR: New methods of presenting ΔR values and implications for MRE studies.* Radiocarbon . ISSN 0033-8222

http://eprints.gla.ac.uk/49640/

Examining the inherent variability in ΔR: New methods of presenting ΔR values and implications for MRE studies.

N Russell[1], GT Cook[1], PL Ascough[1], EM Scott[2], AJ Dugmore[3]

[1]Scottish Universities Environmental Research Centre, Rankine Avenue, Scottish Enterprise Technology Park, East Kilbride G75 0QF, Scotland.

[2]Department of Statistics, University of Glasgow, Glasgow, G12 8QQ.

[3]Institute of Geography, School of Geosciences, University of Edinburgh, Old High School, Infirmary Street, Edinburgh, EH8 9XP, Scotland.

Abstract:

Currently, there is significant on-going research into the temporal and spatial variability of marine ^{14}C 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 without considering the further issue of uncertainty in the modelled marine age that is introduced through the use of a box diffusion model. 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 summarising 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 standardisation 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 (>100 m depth approx.) ^{14}C depletion relative to the contemporaneous atmosphere (Stuiver and Braziunas 1993). Therefore, as a result of the eventual upwelling of deep waters, the surface oceans (0-50 m depth approx.) 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. Ascough et al. 2004, 2005a, 2005b, 2006, 2007a, 2007b, 2009, Butler et al. 2009, Cage et al. 2006, Mangerud et al. 2006, Soares and Martins 2009, 2010, Olsen et al. 2009, Reimer et al. 2002, 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 $\Delta R = 0$) (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 therefore ΔR) values have in turn been used as proxies for changes in localised oceanic regimes (e.g. Kennett et al. 1997, Kovanen and Easterbrook 2002).

The potential uncertainties inherent in deriving ΔR values fall into three 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 sample ^{14}C ages and their associated errors, 3) the modelled marine ^{14}C ages used in ΔR calculation (see Figure 1 below) and the uncertainty arising from the use of a relatively simple marine model to generate these, and 4) 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. An important point is that the marine model uncertainties (point 3 above) are not further considered in the present paper. These model uncertainties are likely to be considerable and will add in quadrature to the variability discussed below, however, consideration of this uncertainty is outside the scope of this study. 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 that is demonstrated

Methods of calculating ΔR.

The concept of identifying localised ΔR variations is discussed in detail by Stuiver et al. (1986), and Stuiver and Braziunas (1993). A ΔR value is 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 modelled marine ^{14}C age is then derived for this sample, by converting the terrestrial/atmospheric ^{14}C age ± 1 sigma to a modelled 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 modelled marine ^{14}C age and the measured ^{14}C age of the marine carbon sample (Fig. 1). The 1σ error on the ΔR values is calculated by the propagation of errors as shown in Equation 1.

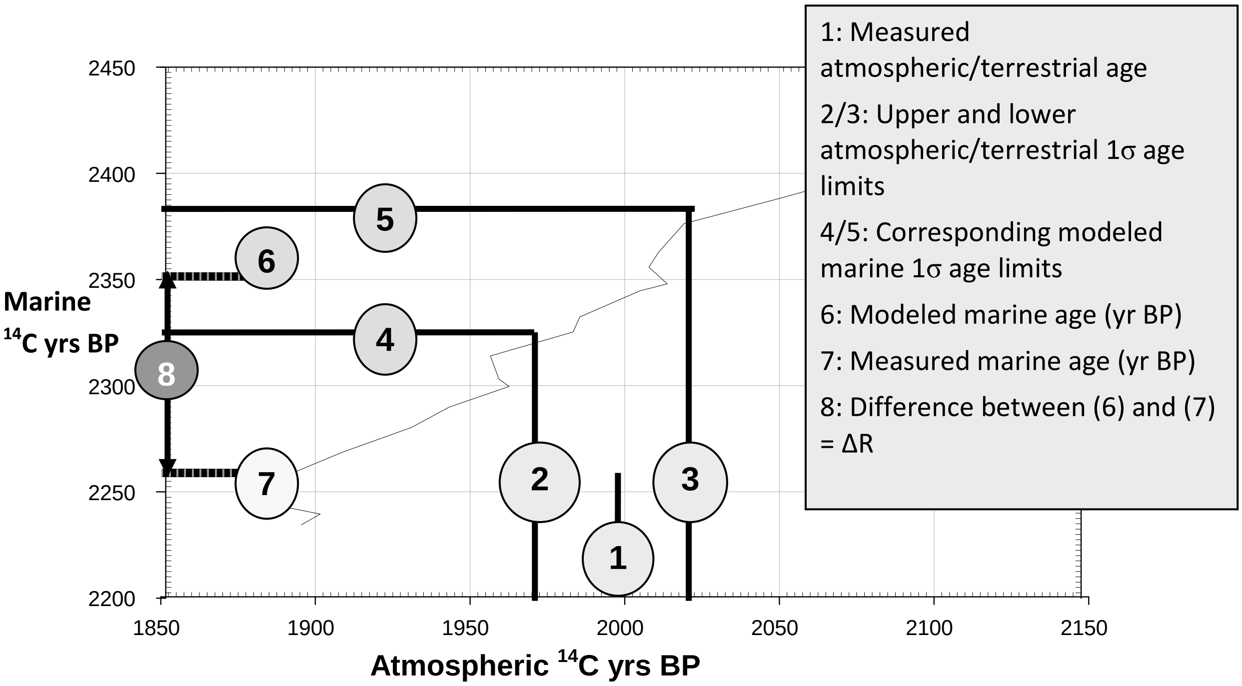


Figure 1: Graphical representation of the determination of a ΔR value showing interpolation of atmospheric and marine ages.

Equation 1. Propagation of ΔR errors:

$$\sigma_{\Delta R} = \surd\ (\sigma_w + \sigma_m)^2$$

Where ($\sigma_{\Delta R}$) = the 1σ error for the ΔR determination

(σ_w) = the error on the measured marine age

(σ_m) = the error on the modelled marine age.

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 year 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 mollusc shell) and terrestrial entities (carbonised grains, herbivore bone, etc.) which have been relatively unaffected by post-depositional 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.

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

Where: t = the weighted mean of the ^{14}C age group
t_i = the individual ^{14}C measurement
σ_i = the error on the individual measurement

Equation 2: *T*-statistic calculation.

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 the *T*-statistic, 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 modelled 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 re-sampling strategy. By this 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 constitutes the sampling distribution (often called a reference distribution) of that statistic. We can then 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) Underpinning the ΔR calculation lies a marine (box diffusion) model and the uncertainty on this has not been considered here. As discussed above, to quantify the model uncertainty is beyond the scope of this paper; nevertheless, it is clear that the effect of this uncertainty would be to increase the variability in the ΔR values.
2) Uncertainty regarding the contemporaneity of terrestrial and marine ^{14}C ages used to generate ΔR values. These uncertainties and recommendations for sample selection criteria that minimize such uncertainties are discussed in detail by Ascough et al. (2005a).
3) 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 estimate of the error has to be a realistic 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 (Naysmith et al 2010). 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 dendro-dated to 3299–3257 BC, with an average ^{14}C age of 4471 BP (Scott 2003) 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 in this example is Archerfield, which is situated on the east coast of Scotland.

Sample Code*	Age (Years BP)	Counting Statistics Error (1σ)
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 1. Standards data for the relevant batch of ^{14}C measurements that included samples from Archerfield (Archerfield sample ^{14}C measurements are given in Table 2).

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 years. However, 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 years and round errors up to the next multiple of 5 years. Sample measurements from the batch in Table 1 would therefore be reported with an error of ± 35 years. It is conceivable that some ΔR values could be calculated with ^{14}C ages that are unrounded, or rounded differently than to the nearest 5 years. This has potential to introduce a source of uncertainty in ΔR calculation, as the number of individual sample ^{14}C ages in a group identified as contemporaneous by the χ^2 test is affected by the size of the error on each ^{14}C age. Underestimation of the sample ^{14}C errors can lead to fewer ^{14}C ages passing the χ^2 test for contemporaneity. Conversely, overestimation of sample ^{14}C errors may lead to a larger number of the tested ^{14}C ages passing the χ^2 test. The effect that rounding of radiocarbon ages and their errors can have on ΔR values can be demonstrated by a worked example (Table 2) using previous data from the site of Archerfield on the North Sea coast of Scotland (Russell et al. 2010).

SUERC- No.	**Sample material**	**^{14}C age (BP) ± 1σ (no rounding)**	**^{14}C age (BP) ± 1σ (conventional publication with rounding)**	**δ^{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**	

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

"In the example above, the χ^2 test statistic (T) for the unrounded group of marine ages is T= 15.26 (χ^2:$_{0.05}$ = 14.07), compared to a χ^2 test statistic of T= 12.29 (χ^2:$_{0.05}$ = 14.07)for the rounded group of marine ages. Use of the unrounded ages would require that 1 marine ^{14}C age (SUERC-19669) is excluded from the sample group, after which the remaining sample ^{14}C ages pass the χ^2 test. The use of unrounded ages for these samples therefore results in use of a different set of samples (i.e. excluding SUERC-19669) for ΔR calculation compared to the use of rounded ages (when all samples would pass the χ^2 test and SUERC-19669 would be included).

The ΔR values calculated from the various pairing of terrestrial/marine ^{14}C ages in the table above ranged from ΔR = +34 ± 40 to ΔR = -122± 42. Weighted mean values and associated errors were calculated using the rounded and unrounded datasets, producing ΔR values of -33 ± 6 (unrounded data) and -42 ± 6 (rounded data) which in this instance are statistically indistinguishable at 2σ. In this instance therefore, the use of rounded versus unrounded data does not ultimately produce significantly different ΔR values. However, it is possible that under some circumstances, statistically different values could arise from the use of unrounded versus rounded data, meaning this consideration is not a trivial one for ΔR calculations. Acknowledgement must be given to the fact that unrounded ages may not be available to all researchers carrying out ΔR investigations, and we therefore recommend the use of unrounded dates as best practice, but admit that it may be applicable only under ideal circumstances.

Two important points emerge from the above discussion: 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 ^{14}C years. If we consider all the possible pairings of ΔR values from the samples from a single context, and subject these to the χ^2 test for contemporaneity, 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 years (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 years for both the unrounded and rounded data. However, for the marine data, the standard deviations are 43 ^{14}C years for unrounded data and 47 ^{14}C years 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 ages.

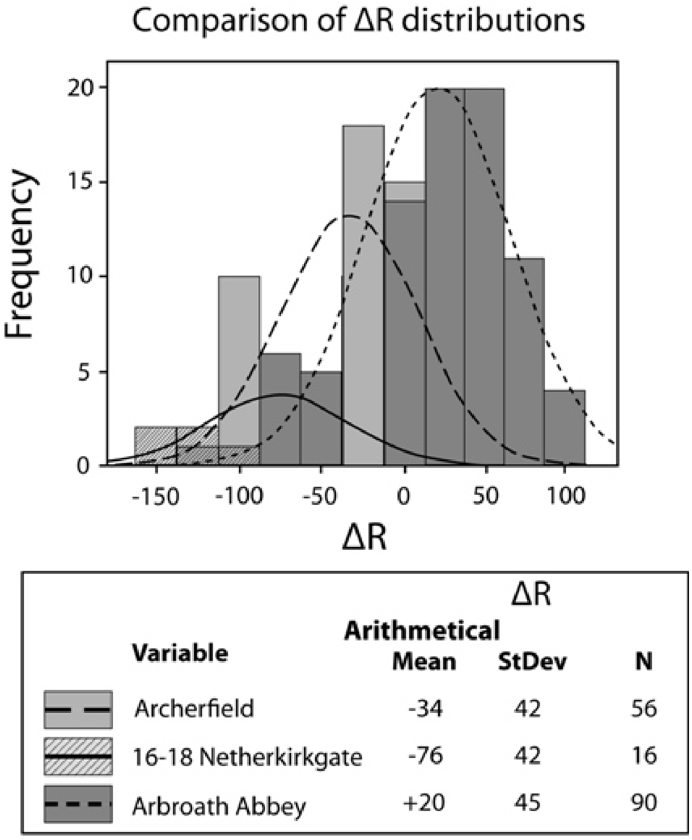
New Methods:

Publishing the mean value from ΔR calculations for each context is commonplace (Ascough et al 2004, 2005, 2006, 2007a, 2007b, 2009, Reimer et al 2002 Russell et al 2010, Soares and Martins 2010, Weisler et al 2009) 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 (Fig. 2). For the purposes of this paper, histograms were constructed using Minitab® 16 using the Normal curves to allow assessment of whether the distribution of ΔR values is indeed Normal. To demonstrate this, three 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 2σ (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).

	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error	Sample pairing		ΔR	Error
a	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																
b	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																
c	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
		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																

Table 3: All possible pairings of ΔR for the 3 sites and weighted mean values for ΔR alongside standard errors for predicted values at 1σ, a) Archerfield 90; b) 16–18 Netherkirkgate and c) Arbroath Abbey.



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.

Publishing ΔR values in this manner allows for a better understanding of the population to which the mean value relates, 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 three 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.

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:

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

Equation 3. Standard error for predicted values where x = the error on the weighted mean and y = 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 2σ. Weighting the mean ΔR values rather than displaying the arithmetical means from the normalised histograms can vary the ΔR value. For example, at the site of 16 -18 Netherkirkgate the mean value in Fig. 2 is -76; whereas the weighted mean value as shown in Fig. 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.

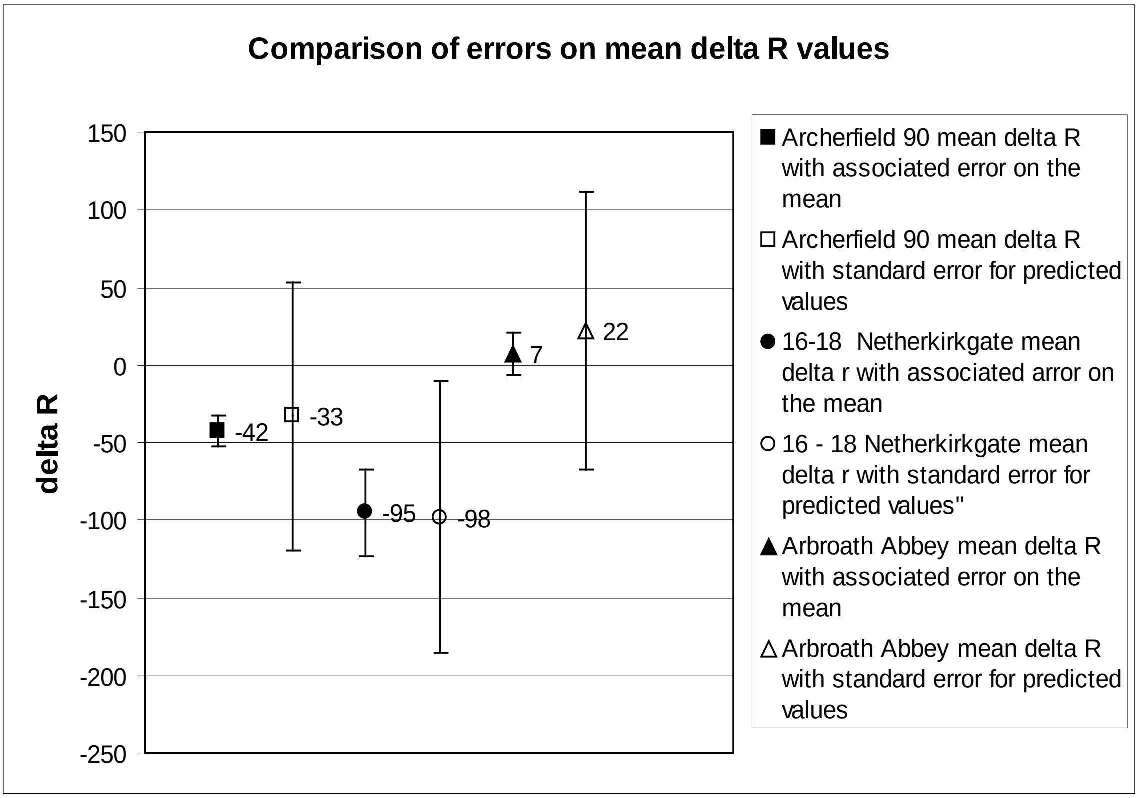


Figure 3. Comparison of ΔR values showing error on the mean (filled symbols) (Russell et al. 2010) and standard error for predicted values (empty symbols). All errors are presented at 2σ.

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 2σ, 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). If the variability shown by the multiple paired sample approach for statistically indistinguishable dates is considered alongside the use of the larger standard error for predicted values at 2 σ when comparing mean ΔR values, (or MRE values), considerable changes in the significance of reservoir offsets, both temporally and spatially may be apparent. This may be of importance to studies using MRE variability as a proxy for oceanographic changes that have identified large scale and rapid fluctuations in ΔR or MRE values over relatively short timescales in various regions (Burr et al 2009, Fontugne et al 2004). 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 errors on the measurements of ^{14}C data used in the calculation of ΔR values 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 when determining ΔR values, (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 dataset of pairings used to calculate ΔR and/ or using histograms can help give a better representation of the 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 standardising 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. The paper has not dealt with the topic of the marine model uncertainty which in itself would deserve a separate discussion. This does not however weaken the argument concerning the presentation of the ΔR variability.

Acknowledgements

NR thanks NERC (Grant No. 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:

http://www.aocarchaeology.com/field-archaeology/archerfield.htm last visited 07/06/10.

Ascough, P.L., Cook, G.T., Dugmore, A.J., Barber, J., Higney, E., and Scott, E.M. 2004. Holocene variations in the Scottish marine radiocarbon reservoir effect. *Radiocarbon* **46**, 611-620.

Ascough, P., Cook, G.T**.** and Dugmore, A.J. 2005a. Methodological approaches to determining the marine radiocarbon reservoir effect. *Progress in Physical Geography* **29**, 532-547.

Ascough, P.L., Cook, G.T., Dugmore, A.J., Scott, E.M. and Freeman, S.P.H.T. 2005b Influence of mollusc species on marine DELTA R determinations. *Radiocarbon* **47**(3) 433-440.

Ascough P., Cook G., Church M. J., Dugmore A. J., Arge S. V. McGovern T. H. 2006. Variability in North Atlantic marine radiocarbon reservoir effects at c.1000 AD. *The Holocene* **16**(1), 131-136.

Ascough P. L., Cook G. T., Dugmore A. J. and Scott E. M. 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) 438-447.

Ascough P. L., Cook G. T., Church M. J., Dugmore A. J., McGovern T. G., Dunbar E., Einarsson Á., Friðriksson A. and Gestsdóttir H. 2007b. Reservoirs and Radiocarbon: ^{14}C dating problems in Myvatnssveit, Northern Iceland. *Radiocarbon* **49**(2) 947-961.

Ascough P., Cook G. T. and Dugmore A. J. 2009. North Atlantic Marine ^{14}C Reservoir Effects: implications for late-Holocene chronological studies. *Quaternary Geochronology* **4**(3), 171-180.

Butler P. G., Scourse J. D., Richardson C. A., Wanamaker A. D., Bryant C. L. and Bennell J. D. 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), 230-241.

Burr, G. S., Beck, J. W., Corrège, T., Cabioch, G., Taylor, F. W., Donahue, D. J. (2009) Modern and Pleistocene reservoir ages inferred from South Pacific corals. *Radiocarbon* 51(1): 319-335.

Cage A. G., Heinemeier J. and Austin, W. E. N. 2006. Marine radiocarbon reservoir ages in Scottish coastal and fjordic waters. *Radiocarbon* **48**(1), 31-43

Fontugne, M., M. Carré, I. Bentaleb, M. Julien, D. Lavallée. 2004. Radiocarbon reservoir age variations in the South Peruvian upwelling during the Holocene. *Radiocarbon* 46(2), 531-537.

Hughen K. A., Baillie M. G. L., Bard E., Beck J. W., Bertrand C. J. H., Blackwell P. G., Buck C. E., Burr G. S., Cutler K. B., Damon P. E., Edwards R. L., Fairbanks R. G., Friedrich M., Guilderson T. P., Kromer B., McCormac G., Manning S., Bronk Ramsey C., Reimer P. J., Reimer R. W., Remmele S., Southon J. R., Stuiver M., Talamo S., Taylor F. W., van der Plicht J. and Weyenmeyer, C. E. 2004. MARINE04 Marine radiocarbon age calibration, 0-26 cal kyr BP. *Radiocarbon* **46**(3), 1059-1086.

Kennett D. J., Ingram L., Erlandson J.M. and Walker P. 1997. Evidence for temporal fluctuations in Marine Radiocarbon Reservoir Ages in the Santa Barbara Channel, Southern California. *Journal of Archaeological Science* **24**, 1051-1059.

Kovanen D.J. and Easterbrook D.J., 2002. Paleodeviations of radiocarbon marine reservoir values for the northeast Pacific. *Geology* **30**(3), 243-246

Mangerud J., Bondevik S., Gulliksen S., Karin Hufthammer A.and Hoisaeter T. 2006. Marine 14C reservoir ages for 19th century whales and molluscs from the North Atlantic. *Quaternary Science Reviews* **25** (23-24), 3228-3245.

Naysmith, P; Cook, G T; Freeman, S P H T; Scott, E M; Anderson, R; Xu, S; Dunbar, E; Muir, G K P; Dougans, A; Wilcken, K; Schnabel, C; Russell, N; Ascough, P L; Maden, C. 2010. 14C AMS at SUERC: Improving QA Data with the 5MV Tandem and 250kV SSAMS. *Radiocarbon* **52** 263 – 271.

Olsen, J., Rasmussen, P. and Heinemeier, J. 2009. Holocene temporal and spatial variations in the radiocarbon reservoir age of three Danish fjords. *Boreas* **38**, 458-470.

Reimer P.J., McCormac F.G., Moore J., McCormick F. and Murray E.V. 2002. Marine radiocarbon reservoir corrections for the mid- to late Holocene in the eastern subpolar North Atlantic. *The Holocene* **12**(2), 129-135.

Reimer P. J., Baillie M. G. L., Bard E., Bayliss A., Beck J. W., Bertrand C. J. H., Blackwell P. G., Buck C. E., Burr G. S., Cutler K. B., Damon P. E., Edwards R. L., Fairbanks R. G., Friedrich M., Guilderson T. P., Hogg A. G., Hughen K. A., Kromer B., McCormac G., Manning S., Bronk Ramsey C., Reimer R. W., Remmele S., Southon J. R., Stuiver M., Talamo S., Taylor F. W., van der Plicht J. and Weyhenmeyer C. E. 2004. INTCAL04 Terrestrial radiocarbon age calibration, 0-26 cal kyr BP. *Radiocarbon* **46**(3), 1029–1058.

Russell, N., Cook G. T, Ascough P. L, Dugmore A. J. 2010. Spatial variation in the Marine Radiocarbon Reservoir Effect throughout the Scottish Post-Roman to Late Medieval period: North Sea values (500 – 1350BP.) *Radiocarbon* **52** (2-3), 1166–1181.

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

Soares A.M.M. and Martins J.M.M. 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), 2875-2881.

Soares A.M.M. and Martins J.M.M. 2010. Radiocarbon dating of marine samples from Gulf of Cadiz: The reservoir effect. *Quaternary International* **221**(1-2), 9-12.

Stuiver, M. Pearson, G.W. and Braziunas, T. 1986. Radiocarbon age calibration of marine samples back to 9000 CAL YR BP. *Radiocarbon* **28**(2), 980 – 1021.

Stuiver M. and Braziunas T.F. 1993. Modelling atmospheric ^{14}C influences and ^{14}C ages of marine samples to 10,000BC. *Radiocarbon* **35**(1), 137 – 189.

Ward G. K. and Wilson S. R. 1978. Procedures for comparing and combining radiocarbon age determinations: A critique. *Archaeometry* **20**, 19–31.

Weisler, M., Hua, Q. and Zhao, J. 2009. Late Holocene C marine reservoir corrections for Hawaii derived from U-series dated archaeological coral. *Radiocarbon*, **51**(3), 955-968.