Comment on “Radiocarbon calibration curve spanning 0 to 50,000 years BP based on paired $^{230}$Th/$^{234}$U/$^{238}$U and $^{14}$C dates on pristine corals” by R.G. Fairbanks et al. (Quaternary Science Reviews 24 (2005) 1781–1796)\(^5\) and “Extending the radiocarbon calibration beyond 26,000 years before present using fossil corals” by T.-C. Chiu et al. (Quaternary Science Reviews 24 (2005) 1797–1808)\(^5\,\,\,\text{\footnote{doi of original article: 10.1016/j.quascirev.2005.04.007}}\)

1. Introduction

Radiocarbon calibration curves are essential for converting radiocarbon-dated chronologies to the calendar timescale. Prior to the 1980s numerous differently derived calibration curves based on radiocarbon ages of known age material were in use, resulting in “apples and oranges” comparisons between various records (Klein et al., 1982), further complicated by until then unappreciated inter-laboratory variations (International Study Group, 1982). The solution was to produce an internationally agreed calibration curve based on carefully screened data with updates at 4–6 year intervals (Klein et al., 1982; Stuiver and Reimer, 1986, 1993; Stuiver et al., 1998). The IntCal working group has continued this tradition with the active participation of researchers who produced the records that were considered for incorporation into the current, internationally ratified calibration curves, IntCal04, SHCal04, and Marine04, for Northern Hemisphere terrestrial, Southern Hemisphere terrestrial, and marine samples, respectively (Hughen et al., 2004; McCormac et al., 2004; Reimer et al., 2004).

Fairbanks et al. (2005), accompanied by a more technical paper, Chiu et al. (2005), and an introductory comment, Adkins (2005), recently published a “calibration curve spanning 0–50,000 years”. Fairbanks et al. (2005) and Chiu et al. (2005) have made a significant contribution to the database on which the IntCal04 and Marine04 calibration curves are based. These authors have now taken the further step to derive their own radiocarbon calibration extending to 50,000 cal BP, which they claim is superior to that generated by the IntCal working group. In their papers, these authors are strongly critical of the IntCal calibration efforts for what they claim to be inadequate screening and sample pretreatment methods. While these criticisms may ultimately be helpful in identifying a better set of protocols, we feel that there are also several erroneous and misleading statements made by these authors which require a response by the IntCal working group. Furthermore, we would like to comment on the sample selection criteria, pretreatment methods, and statistical methods utilized by Fairbanks et al. (2005) in derivation of their own radiocarbon calibration.

2. Data quality and screening

In our opinion, the Fairbanks et al. (2005) paper is extremely misleading about the current status of radiocarbon calibration and the quality of data used to construct IntCal04 and Marine04 beyond the end of the tree-ring data sets, which correctly starts at 12,410 cal BP (Friedrich et al., 2004; Reimer et al., 2004). In Fig. 1 of the Fairbanks et al. (2005) paper, they present a variety of radiocarbon records and imply that this is the current status of data suitable for estimating the radiocarbon calibration curve, although many of the data sets shown in their plot did not meet the criteria established by the IntCal working group (Reimer et al., 2002). In Fig. 1 we show the data sets actually used to derive the IntCal04 or Marine04 calibration curves. In actual fact, for the period prior to 26,000 cal BP, the IntCal working group declined to make a recommendation for radiocarbon calibration due to the large scatter and offsets between available data sets (van der Plicht et al., 2004). These include records from foraminifera dated using oxygen isotope correlations with the GISP2 or GRIP oxygen record, speleothems dated using U/Th methods, corals also dated using U/Th, and lacustrine records dated using varve counting methods or U/Th of aragonite. While individually, some of these records are highly coherent, the offset between, and scatter within, some of these data sets prevented us from recommending calibration beyond 26,000 cal BP. The scatter may in some cases be due to inadequate screening or pretreatment methods, while the offsets are likely due to poorly determined calendar ages or reservoir corrections.

Fairbanks et al. (2005) and Chiu et al. (2005) argue that their new coral-based radiocarbon record is sufficiently well dated and robust that they can exclude all these other data sets to prepare a precise and accurate radiocarbon
However, it should be clarified that this problem mainly affects samples that did not stay below sea level, thereby exposing them to alteration by fresh water. It should be noted that virtually all corals that grew before the Last Glacial Maximum (LGM) have been exposed to some level of meteoric alteration. The only exception to this rule are corals from subsiding islands where subsidence rates are sufficient to compensate for the sea-level fall that took place right before the LGM (i.e., between 26,000 and 35,000 yr BP). To avoid the potential for meteoric alteration, the rate of subsidence in such locales would have to be extremely large, on the order 3–12 mm/yr (Lambeck et al., 2002; Siddall et al., 2003). No such reefs have yet been exploited for radiocarbon calibration purposes. We note that all of the pre-LGM corals used in the Fairbanks et al. (2005) study were collected either from Barbados or Araki (Vanuatu). Both these sites have substantial uplift (not subsidence) rates and Araki, being in the West Pacific Warm Pool also has very high rainfall rates. Thus the pre-LGM reefs at both these sites would certainly have been exposed to the potential for freshwater alteration prior to the LGM.

Because of the potential for coral alteration (especially of pre-LGM corals), the IntCal group has adopted a set of screening and pretreatment protocols which must be met for all coral data incorporated into the IntCal radiocarbon calibration, which Fairbanks et al. (2005) claim are inadequate. To support this assertion, both Fairbanks et al. (2005) and the companion paper by Chiu et al. (2005) show figures displaying data published by Bard et al. (1998), Yokoyama et al. (2000), and Cutler et al. (2004). In doing so, they neglected to clearly state that the data reported by Yokoyama et al. (2000) did not meet quality criteria set by the IntCal group (Reimer et al., 2002; Hughen et al., 2004) and were thus not selected for inclusion in the IntCal compilation (Hughen et al., 2004; Reimer et al., 2004). By including the Yokoyama et al. (2000) data set in these comparisons Fairbanks et al. (2005) give the erroneous impression that these data were used in the IntCal04 calibration, and imply that IntCal data selection criteria are haphazard.

Fairbanks et al. (2005) and Chiu et al. (2005) describe at length the techniques they used to select and pretreat their own samples, which they claim are better than those used by the IntCal group. We appreciate the fact that they have expended considerable effort to investigate these techniques, and feel that the discussion stimulated by their efforts may be helpful in designing better screening and pretreatment protocols. However, it is not at all clear that their techniques are superior to those advocated by the IntCal group.

### 3. Chemical pretreatment of corals

Fairbanks et al. (2005) state that the selection criteria used by the IntCal group are inadequate, and assert—using model calculations—that the partial-dissolution sample pretreatment methods advocated by the IntCal group may

![Figure 1. Radiocarbon calibration data used to construct the IntCal04 and Marine04 calibration curves.](image-url)
actually increase the level of recent contamination. Extensive studies by Burr et al. (1992) show quite the opposite; however, indicating that in many cases partial dissolution of the coral removes recent contamination. Indeed, Burr et al. (1992) and Yokoyama et al. (2000) have done careful stepwise experiments of coral leaching for which both leachates and residues were dated. They observed that the residue is usually older than the leachate. A similar conclusion was reached by Bard et al. (1990) who compared $^{14}$C ages of Barbados samples obtained with and without leaching. Collectively, these studies demonstrate that partial dissolution of the coral often leads to older $^{14}$C ages. It should be the opposite if the Fairbanks–Chiu model was valid.

Fairbanks et al. (2005) and Chiu et al. (2005) also claim that the $^{14}$C age of the secondary calcite is always younger than the original sample resulting in $^{14}$C ages too young for slightly altered samples. Chiu et al. (2005) even model this effect, suggesting that it affects data measured by other groups. This discussion is largely unfounded and the choice of model parameters is ad hoc. Indeed, the main assumption that the $^{14}$C age of the diagenetic calcite is always younger is in conflict with several lines of evidence.

Fairbanks et al. (2005) and Chiu et al. (2005) emphasize that the data by Yokoyama et al. (2000) are affected by diagenetic calcite because their samples were collected in uplifted terraces exposed to meteoric alteration. The scatter of these data is taken as a signature of diagenetic alteration. However, the most puzzling anomaly is that these $^{14}$C ages are often very much older than implied by alteration. Indeed, it is unlikely that this would occur, so there is some doubt that this pretreatment method would produce any positive effect on the radiocarbon measurement. That this could have some effect on the U/Th measurements is another story however, as both U and Th have high affinity for organic matter. It is not clear, however, whether Chiu et al. (2005) applied this technique to coral samples used for U/Th dating, in an effort to remove secondary U or Th.

4. Detection of secondary calcite in corals

Fairbanks et al. (2005) and Chiu et al. (2005) also describe their work aimed at detecting secondary calcite in corals using XRD in order to eliminate samples contaminated by secondary calcite. This is actually an important topic; however it is also not a new idea. This screening technique is routinely used, and all coral samples used in the IntCal calibration had the requirement that they have less than 1% quantifiable calcite using XRD analysis. As described in the previous section, IntCal protocols furthermore require that the samples have undergone partial dissolution to rid them of secondary overgrowths (including secondary aragonite, which is not detected by the technique proposed by Fairbanks et al. (2005) and Chiu et al. (2005)).

As a test the CEREGE group has analyzed a *Porites* coral sample that is partly recrystallised to calcite. The $^{14}$C age of the calcite is $15,940 \pm 100$ yr BP, whereas the pristine aragonite is $13,160 \pm 140$ yr BP. Most probably, part of the carbon included in the secondary calcite comes from dissolution of older carbonates, which form the substratum of the recent reef. Dissolution and reprecipitation of old carbonate is widely known in reef environments. This further shows that assuming that the calcite $^{14}$C age is always younger than the true age is an invalid and ad hoc assumption.

Chiu et al. (2005) discuss at great lengths the measures they have taken to pretreat their samples with $\mathrm{H}_2\mathrm{O}_2$ in order to remove organic contamination from their corals prior to radiocarbon analysis. Their data in Table 3 do not really indicate that this pretreatment method had done any perceptible good, however. It is not at all clear how the accidental organic mater inclusions found in corals would react to form CO$_2$ during phosphoric acid hydrolysis of the aragonite. Indeed, it is unlikely that this would occur, so there is some doubt that this pretreatment method would produce any positive effect on the radiocarbon measurement. That this could have some effect on the U/Th measurements is another story however, as both U and Th have high affinity for organic matter. It is not clear, however, whether Chiu et al. (2005) applied this technique to coral samples used for U/Th dating, in an effort to remove secondary U or Th.
leading to well-known shifts in $2\theta$ of the XRD peaks. Indeed, this shift was even used previously to estimate the Mg content in calcite (Goldsmith and Graf, 1958). The secondary calcite is often heterogeneous and polygenetic, which makes the calcite peaks much thicker than if they were from pure CaCO$_3$. Consequently, a calibration based on peak height (as done in the Chiu et al. (2005) study) seriously underestimates the calcite content in altered corals (Fig. 2). Again, the best way is to use peak areas to build such a quantitative calibration, which is the usual procedure in quantitative XRD (Bish and Post, 1989).

5. Marine reservoir issue

The conversion of marine $^{14}$C data to atmospheric equivalents requires a correction for the effect of “old” carbon in the large oceanic carbon reservoir. This is usually achieved via estimation of the aptly named reservoir or apparent age (Mangerud, 1972). The reservoir age is variable over time, due to ocean circulation and carbon cycle changes, and space, due to local/regional ocean context and dynamics (Broecker et al., 1960; Monge Soares, 1993). As in the IntCal04 work, Fairbanks et al. (2005) estimate this correction from the offset between Holocene tree-ring data and the overlapping coral data sets. This provides a minimum estimate of the reservoir age error and perhaps a false impression of constancy through time—indeed the data reported by Fairbanks et al. (2005) and Chiu et al. (2005) indicate significant underlying high-frequency variability masked by an “average”. For example, comparison of the varved Cariaco Basin record with an extension of tree-ring radiocarbon data (not yet dendrochronologically linked) indicates that the reservoir age in the western tropical Atlantic may have varied by hundreds of years during deglaciation (Kromer et al., 2004). This offset appears in the continuous decadally resolved Cariaco Basin foraminiferal data set as well as the less dense Barbados coral data. Fairbanks et al. (2005) ascribe all of the reservoir variability to the atmosphere despite the fact that the coral data are on approximately annual samples and the tree-rings are decadal averages. It is instead more likely that the observed reservoir variability is the consequence of ocean dynamics. This is of particular relevance with regards to the much more oceanographically complex southwest Pacific where Araki is located: on seasonal to inter-annual timescales this region has modern pre-bomb reservoir age variations of $\sim$150 $^{14}$C yrs (Guilderson et al., 2004; Schmidt et al., 2004) (Fig. 3). Similar variability is inferred from fossil corals (Burr et al., 2004).

6. Statistical methodologies

In addition to offering new data, Fairbanks et al. (2005) outline a range of numerical methods that they have used for estimating the radiocarbon calibration curve from raw data. In what follows we offer a criticism of both the generalities and the specifics of the numerical/statistical methods chosen by Fairbanks et al. (2005). Before we do so, however, we note that in describing previous methods for estimating calibration curves in the presence of uncertainty on both $x$- and $y$-axes, Fairbanks et al. (2005) fail to acknowledge the full scope of the most recent work in this area. Buck and Blackwell (2004) have developed a tailored framework for this specific problem.
that was ratified by the international radiocarbon community at its 18th Conference in Wellington, New Zealand in September 2003, and was used to estimate all of the new internationally agreed-upon calibration curves (Hughen et al., 2004; McCormac et al., 2004; Reimer et al., 2004).

The approach taken leads to a completely explicit model-based Bayesian framework which is described in detail in Buck and Blackwell (2004). The framework is capable of incorporating observations with independent errors in both dimensions (precisely the problem that Fairbanks et al. (2005) wish to solve), as well as observations with much more complex or asymmetric error structures like those that arise from wiggle matched and/or counted sequences (avoided by Fairbanks et al. (2005)). Fairbanks et al. (2005) imply such error structures cannot be quantified at all, but on the contrary, the Buck and Blackwell model provides objective and reasonable error estimates for all IntCal04 data sets (Reimer et al., 2004).

In their paper, Fairbanks et al. (2005) tell us that their methods provide a calibration procedure (with error estimation) that is based on hierarchical Bayesian statistical model. Despite a paucity of detail on statistical methods, it is clear that what these authors are doing is neither Bayesian nor hierarchical in the sense that most statisticians or philosophers of science would use the terms.

The Bayesian approach to inference requires careful definition of available a priori knowledge and a clear statement (usually in the form of a statistical likelihood) of the link between the available data (including uncertainty terms) and the events or concepts that we wish to learn about. Fairbanks et al. (2005) offer neither a clearly stated prior nor a sufficiently detailed likelihood for their methods to be described as Bayesian. Hierarchical statistical modeling usually refers not to the nature of implementation as these authors suggest, but to the nature of the model needed to represent the problem under investigation. As such, an hierarchical statistical model represents features of the problem at different scales or levels that are linked together via a conceptual hierarchy. Such an approach could be taken to modeling the radiocarbon calibration problem, but there is no evidence in this paper that such an approach was used.

The authors say that they used a “least-squares estimation approach”. Such approaches are usually adopted in situations where we wish to avoid having to fully specify a parametric statistical model. This is most puzzling, because they also say that they are deriving conditional distributions of the form $p(y|x)$. If this were the case they should be working with a fully parametric model, the parameters of which relate directly to the features of the calibration curve about which they wish to learn. This is exactly the kind of approach taken by Gómez Portugal Aguilar et al. (2002) whom Fairbanks et al. (2005) cite, suggesting that their “least-squares estimation approach” is similar. If they are using the term least-squares estimation as it is usually used, there is no theoretical and little practical similarity between it and the model-based ("random-walk") method of Gómez Portugal Aguilar et al. (2002).

As part of the implementation of their method, the authors describe an “ensemble” technique for handing the uncertainty on the calendar scale. The explanation they give here is unclear and lacks detail. However, presuming that they assumed normal, and hence continuous, errors on the calendar scale, it is unclear from the paper how they can justify their discretized implementation. Furthermore, it is not clear why such discretization is necessary (except perhaps at the final computational stage). Even if it were appropriate, a 10 year grid is certainly too coarse given some of the smaller standard deviations on errors in both the $x$ and $y$ directions with which the authors are working.

The authors use an optimal averaging procedure which is not clearly described. In particular, it is unclear how or why one should do a “chi-squared test procedure” when there is known to be an error in both dimensions. Even more worrying is the “inflation of measurement errors by factors of 1.8 and 1.3...”. This is an important issue relating to data quality and reliability, not just an implementation detail. One of the major rationales for their paper is the claim, early on, that the authors “can compute a rigorous error estimate”. If arbitrary inflation of laboratory estimates of uncertainty is part of the process, without explanation as to how or why they were derived, then any error estimates that arise are far from rigorously derived. Here the authors entirely ignore the significant positive (quality and robustness) developments in radiocarbon calibration treatment as described and implemented in Buck and Blackwell (2004), Reimer et al. (2002, 2004), Hughen et al. (2004) and McCormac et al. (2004).

In the final part of the section on “Implementation details”, the authors tell us that they derived “...a priori estimates necessary for our procedure...”. In practice, the only specifically stated prior assumption relates to what they call the prior on the slope which they say has a “standardized deviation” (sic) of 0.8 yr/yr. If by this statement the authors mean that they are taking 0.8 yr/yr as the standard deviation of the prior on the slope of radiocarbon calibration curve, it is extremely small when compared to other recent studies (Christen and Nicholls, 2000; Buck and Blackwell, 2004). There are two further “a priori estimates” that they say they made; neither are justified or explained. We are not told what was assumed about the “expected slope changes from point to point” and it is also unclear what is meant by “age variability inside 10 year grids”. This lack of clarity with regard to assumptions made about inputs to their method is worrying, particularly in the light of the fact that there is no discussion about the sensitivity of their results to changes in assumptions made.

7. Conclusions

Fairbanks et al. (2005) have not made a convincing argument for a stand alone coral calibration curve based
on their data only. Their sample screening criteria are no more stringent than that used for the coral data included in the IntCal04 and Marine04 calibration curves, their chemical pretreatment methods do not show any improvement over the partial dissolution method, and their curve was produced using statistical methods that are incompletely described and seemingly ad hoc.

Multiple, independently derived data sets are required to provide confirmation and validation of all $^{14}$C to calendar age calibration data. The Fairbanks et al. (2005) data now becomes one more—obviously important—record to add to this database. It does not, however, resolve any of the issues of conflict between data sets that led the IntCal group to decline to recommend calibration beyond 26,000 cal BP. Use of the term “comparison curve” instead of “calibration curve” has been advocated until these issues are resolved (van der Plicht, 2000; Beck et al., 2001).

In an ideal and utopian world continuous or annual updates to calibration data sets and calibration programs would be useful. In the real world, however, independently producing calibration (as opposed to comparison) curves causes confusion for the user communities (Blockley et al., 2001; Housley et al., 2001). This is particularly problematic given the peer-review process which leads to considerable lags between submission and publication. Frequent updates and/or confusion as to which is the current “best” estimate of the calibration curve can result in researchers “chasing their tail” even after a publication has been accepted and in press. Fairbanks et al. (2005) are not the first to produce an independent calibration curve for routine use by the user community. The CalPal program has been used to produce compilations from varying quality $^{14}$C records for calibration purposes (Jöris and Weninger, 2000; Weninger and Jöris, 2004). The IntCal working group has not formally commented on CalPal but requested a disclaimer be associated with the XRD spectra presented in Fig. 2. We also wish to thank A. Mix for a thorough and critical review which resulted in improvements to this comment. A portion of the work on the IntCal04 project was funded by a grant from the Leverhulme Trust.

Acknowledgments

The authors wish to thank S. Sépulcre for measuring the XRD spectra presented in Fig. 2. We also wish to thank A. Mix for a thorough and critical review which resulted in improvements to this comment. A portion of the work on the IntCal04 project was funded by a grant from the Leverhulme Trust.

References
