DEVELOPMENT OF THE RADIOCARBON CALIBRATION PROGRAM

Christopher Bronk Ramsey

Oxford Radiocarbon Accelerator Unit, 6 Keble Road, Oxford OX2 6JB, UK. Email: christopher.ramsey@archaeology-research.oxford.ac.uk.

ABSTRACT. This paper highlights some of the main developments to the radiocarbon calibration program, OxCal. In addition to many cosmetic changes, the latest version of OxCal uses some different algorithms for the treatment of multiple phases. The theoretical framework behind these is discussed and some model calculations demonstrated. Significant changes have also been made to the sampling algorithms used which improve the convergence of the Bayesian analysis. The convergence itself is also reported in a more comprehensive way so that problems can be traced to specific parts of the model. The use of convergence data, and other techniques for testing the implications of particular models, are described.

INTRODUCTION

When OxCal was first released (1/8/1994, see Bronk Ramsey 1994, 1995) it provided, in addition to straight-forward calibration, various techniques for the analysis of groups of radiocarbon events in phases, sequences and (based largely on the work of Buck et al. 1991, 1992, 1994). New techniques were also provided to test the internal consistency of these Models based on "agreement indices" (Bronk Ramsey 1995), which are, in effect pseudo Bayes factors (see for example chapter 9 of Gilks et al. 1996). Also provided was a method of wiggle matching tree rings by factoring together the shifted probability distributions from each ^{14}C measurement (referred to in Bronk Ramsey 1995, details in the OxCal manual and Bronk Ramsey et al. (2001). Further development of the command line "language", CQL is described in Bronk Ramsey (1998), with the main aims of the program's development.

Since the first release of the program much work has been done to widen its applicability, make it easier to use with large data-sets and, of course, to deal with bugs. Details of all of this are given in the manual for the program and are not a suitable subject for publication here. This paper will cover instead the mathematical developments of the program (on which there is also further information in the program's manual).

CALIBRATION

Two changes have been made to the calibration algorithms used in OxCal. The first change is a more precise treatment of the error terms in the calibration. We will define here the calibration curve to be a function r(t) with a standard error of $\sigma(t)$, with the measured values for a sample being $r_m \pm \sigma_m$. Versions of OxCal up to and including 3.2, like many other first generation Calibration programs, used the probability calculation for calibration of:

$$p(t) \propto \exp(-(r_m - r(t))^2 / 2(\sigma_m^2 + \sigma^2(t))) \quad (1)$$

The later versions include a term, in line with current versions of CALIB (Stuiver and Reimer 1993), which takes account of variability in the calibration curve errors in the normalization of the probability distributions.

$$p(t) \propto \frac{\exp(-(r_m - r(t))^2 / 2(\sigma_m^2 + \sigma^2(t)))}{\sqrt{\sigma^2 + \sigma^2(t)}} \quad (2)$$

© 2001 by the Arizona Board of Regents on behalf of the University of Arizona
RADIOCARBON, Vol 43, Nr 2A, 2001, p 355–363
Proceedings of the 17th International ^{14}C Conference, edited by I Carmi and E Boaretto

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

This is only likely to make a significant difference when the calibration curve errors are large and variable—as they are at the start of the Holocene. This is particularly important with the new INTCAL98 calibration curve (Stuiver et al. 1998). This modification was suggested by Stephan Puchegger of Vienna (personal communication).

The second change allows mixed calibration curves to be defined using the program. The main application of this is for instances of human bone where there is evidence for a mix of marine and terrestrial sources of diet. There is also the opportunity to mix geological age carbon, which might be relevant, for example, in river environments. If the two curves r_1 and r_2 are to be mixed to give r_m where the proportion of the second curve is $p_2 \pm \delta_2$. The mixed curve is given by:

$$r_m = (1-p_2)r_1 + p_2 r_2 \tag{3}$$

$$\sigma_m = \sqrt{((1-p_2)\sigma_1)^2 + (p_2\sigma_2)^2 + (\delta_2(r_1-r_2))^2} \tag{4}$$

The user interface should help you through the stages needed to do this but essentially to define a mixed curve the two curves to be mixed must first be specified along with any ΔR corrections. For example one might have:

```
Plot
{
 Curve "intcal98" "C:/Program Files/OxCal3/intcal98.14c";
 Curve "local_marine" "C:/Program Files/OxCal3/marine98.14c";
   Delta_R 100 30;
 Mix_Curves "mixed" "intcal98" "local_marine" 20 5;
 R_Date 660 35;
};
```

This will define the sources for the two curves to mix and the local ΔR correction to be made to the marine curve; in this case, we have specified a marine component of 20±5%. A file dead.14c is provided with a ^{14}C age of 100,000 BP so that geological age carbon can be mixed in using the same technique.

In principle any number of curves can be mixed since a mixed curve can itself be mixed with another once it has been defined.

ANALYSIS OF SEQUENCES AND PHASES

Background to Original Implementation

The main Bayesian analysis procedures of OxCal have not altered significantly since its first release, except in the detailed treatment of *boundaries*. To understand how these are used it is necessary to understand the underlying assumptions in any analysis. For an overview of the application of Markov Chain Monte Carlo to this type of problem see Gilks et al. (1996); this paper will deal with the aspects of this that are specific to this particular application.

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

The main underlying assumptions are that events are independent and are equally likely to occur at any point on the time scale. These assumptions are identical to those made for the classical Poisson distribution. These, apparently neutral, assumptions do, however, have important consequences when applied to a discrete group of events rather than a infinitely long distribution. For such a group the number of possible combinations of dates is a strong function of the span of the group. If the span of the dates is given by *s* and the number of events under consideration is *n*, the number of possible combinations is, in fact, proportional to s^{n-2}. This is true whether or not the events are constrained to be in a series. This would be fair enough if the events were truly independent, but in practice events under investigation are usually related in some way and are uniformly distributed only over a limited time range rather than throughout the whole time scale.

For this reason a method of grouping events into coherent groups, using *boundaries* was included in the first version of OxCal. This was based on the mathematical formulations of Buck et al. (1992). In this method the events in a model can be divided up between *boundaries*. The events between these *boundaries* are assumed to be uniformly distributed (with Poisson distributed intervals), but only over a limited time span, rather than the infinite time span of the initial model. In order to overcome the statistical weight of greater *boundary* separations, the analysis is deliberately biased towards shorter *boundary* gaps in order to exactly balance this effect. If we consider two boundaries in the model, b_i and b_i+1 between which there are n_i events, we apply a bias of:

$$(b_{i+1} - b_i)^{-n_i} \tag{5}$$

The net effect of this is that the effective prior for the span of any group of events is now uniform, that is any span is treated as equally likely.

These methods have been applied satisfactorily in many archaeological contexts (see for example Bronk Ramsey and Allen 1995 or Needham et al. 1998). Models without *boundaries* will tend to bias to spans which are too long and so, although for small numbers of events it usually makes little difference, they should really be included in almost any model. Version 3.4 of OxCal will warn if no *boundaries* have been used and allows *boundaries* to be put in automatically.

Multiple Boundaries

Even with this model, there is still a bias to longer overall spans if the dates are split up into a number of groups. The more groups the dates are split into, the stronger this bias is. Effectively the phenomenon seen above still applies to the *boundaries* themselves. This makes very little difference in most situations as the number of related groups is usually small. However, to overcome this an additional bias can be added to the model: if there are *m boundaries* in a sequence, we can apply a bias of:

$$(b_m - b_1)^{m-2} \tag{6}$$

which exactly counteracts this effect. We are now assuming that, not only are the events grouped between two *boundaries* uniformly distributed over a limited range of time, but that the *boundaries* themselves are also uniformly distributed over a limited range. Where the overall model is unconstrained in date this makes the effective prior for the span of all of the dates considered to be uniform. Where there is an upper and lower limit on the possible dates, however, this will tend to favour shorter spans slightly, something which can be overcome with a bias of:

$$\frac{1}{(b_{ulim} - b_{llim}) - (b_m - b_1)} \tag{7}$$

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

These two additional factors were suggested by Goeff Nichols of Aukland University, New Zealand (private communication) as was the notion of treating the *boundaries* themselves as uniformly distributed events. These factors have been incorporated into v3.2 of OxCal and onwards.

To summarize: the underlying prior assumption made is that the span of the whole timescale of the analysis is equally likely to be any length (the prior for the overall span is uniform); within this, we assume that the *boundaries* that separate chronological stages of the archaeological or geological process are uniformly distributed throughout the overall span (with a Poisson interval distribution); between the *boundaries*, we assume that the dated events within any chronological stage are also themselves uniformly distributed.

Nesting Boundaries

A complicated model might have *boundaries* at various different hierarchical levels. In a site context, for example the construction and destruction of a building might be classified as such. On a more general scale the transition from early to middle and middle to late bronze age might likewise be classified as temporal *boundaries*. These are clearly different in nature and need to be treated as such in any analysis. For this reason, in OxCal only *boundaries* at the same level of nesting within the model structure are treated as a coherent group. For example part of a model might include:

```
Sequence
{
 Boundary "Start of EBA";
 Phase "EBA" {...events of EBA phase...};
 Boundary "EBA-MBA transition";
 Phase "MBA"
 {
  Sequence "House MBA-I"
  {
   Boundary "House construction";
   Phase "House Use" {...events relating to house use...};
   Boundary "House destruction";
  }
  ...other events of MBA phase...
 }
 Boundary "MBA-LBA transition";
 Phase "LBA" {...events of LBA phase...}
};
```

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

In such a case, as far as the outer *boundaries* are concerned, the House Construction and Destruction are merely two events within the MBA phase. As far as the inner *boundaries* are concerned, the transitions to and from the MBA are simply treated as constraints.

MARKOV CHAIN MONTE CARLO

Several improvements have been made to the Markov Chain Monte Carlo (MCMC) sampling used in this program. For details of this technique see Gilks et al. (1996).

The first improvement has been to go from integer arithmetic (originally used to improve speed and reduce memory use) to a fully floating point algorithm (all versions after 2.18). This greatly helped with normalization problems as phase lengths get close to zero. There are still some compromises which have had to be made in order to provide reasonable convergence. This affects phases which are less than half the length of the resolution (i.e. 2 years for the default resolution of 4).

The second main change has been to the sampling algorithm itself. The original program used only the Gibbs sampler. In this, each pass provides a new possible state for the system. It was implemented in such a way that only one event could change at a time. An alternative approach is the Metropolis-Hastings algorithm. In this, each pass does not necessarily produce a new state, but as each pass is faster to calculate, the model converges faster. Furthermore, it is easier to implement state transitions involving several of the modelled events. The newer versions of OxCal (3.3 and later) use a mixture of Gibbs and Metropolis-Hastings and allow groups of events (defined within *boundaries*) to be shifted, expanded and contracted as a whole. The user of the program need not really be aware of all of this. The key thing is that all of these algorithms are designed to give a representative selection of possible states for the model as a whole. These new methods provide much faster convergence for complex models.

The third development is the provision of extra tools for dealing with convergence. The original program reported poor convergence (as defined in Bronk Ramsey 1995). The program now automatically increases the number of iterations until the convergence is reasonable. In some cases, this may never happen and the program may run almost indefinitely. In such cases the program must be aborted and it is useful to have some methods for investigating what is happening. A new option in the program allows convergence data to be included in the plots. This results in the convergence for each element of the model being displayed on multi-plots (in square brackets) and individual plots show the details of the MCMC sampling process for a portion of the run. Figure 1 shows one such plot where the convergence is rapid and Figure 2 a case where the convergence is slower.

TESTING PERFORMANCE

Obviously, analysis software of this sort needs to be tested, both to ensure that it is as free of "bugs" as possible, and to see if the overall approach is achieving its objectives. Here, I will concentrate on the latter, since testing of this kind should sometimes be performed by the user of the program too, where complex models are being used. We will give two example tests here which show possible approaches.

Simulated Dating

In this approach we consider events that conform to the designed model, we simulate the ^{14}C dating of these events and then we see if the analysis can reveal the true characteristics of the event distribution.

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

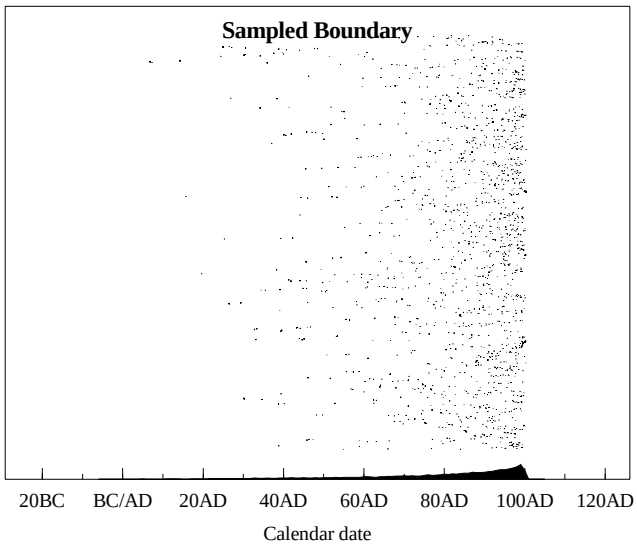

Figure 1 This shows an example of good, rapid convergence.

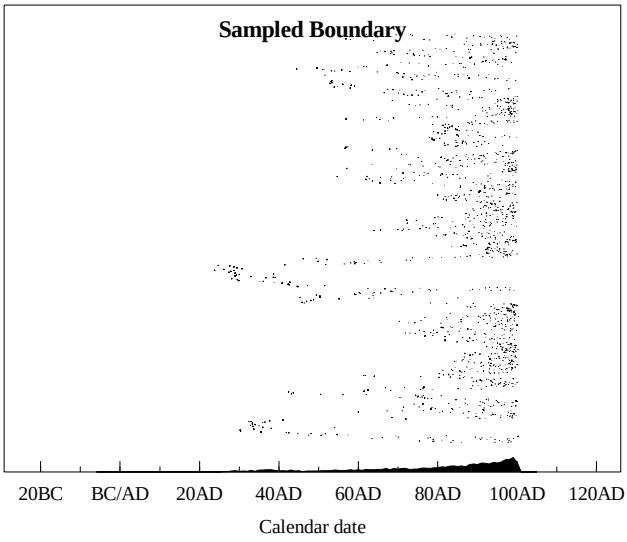

Figure 2 An example of slower convergence—this model would take more iterations to achieve reasonable convergence.

As an example, we consider a simple scenario with four phases: I=1–200 AD, II=201–500 AD, III=501–600 AD and IV=601–800 AD. We assume that from each phase we have a number of evenly distributed events which have been ^{14}C dated. To complicate matters the density of dates is different in each phase (50/century in phase I, 17 in II, 100 in III, and 50 in IV).

The simulated ^{14}C dates are generated using OxCal's R_Simulate function which generates the kind of date you would expect for an object given the true age and the precision obtainable. The model is then constructed with five *boundaries* and four phases.

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

The results of the analysis are shown in Figure 3 (450,000 iterations). This shows that the *boundaries* deduced by the analysis, do fall across the correct dates (confirming that the biases for number of events in the phase has operated correctly). The Sum function, which provides a frequency modulated date distribution has yielded a function which is a good approximation to the true frequency distribution of events. This is generally the case if the dating resolution is good in comparison to the overall scale of the Sum distribution, although there will be random fluctuations from the true distribution which do not reflect the original process (as commented by Steier et al. 2001, though these fluctuations are random and not a reproducible function of the calibration curve).

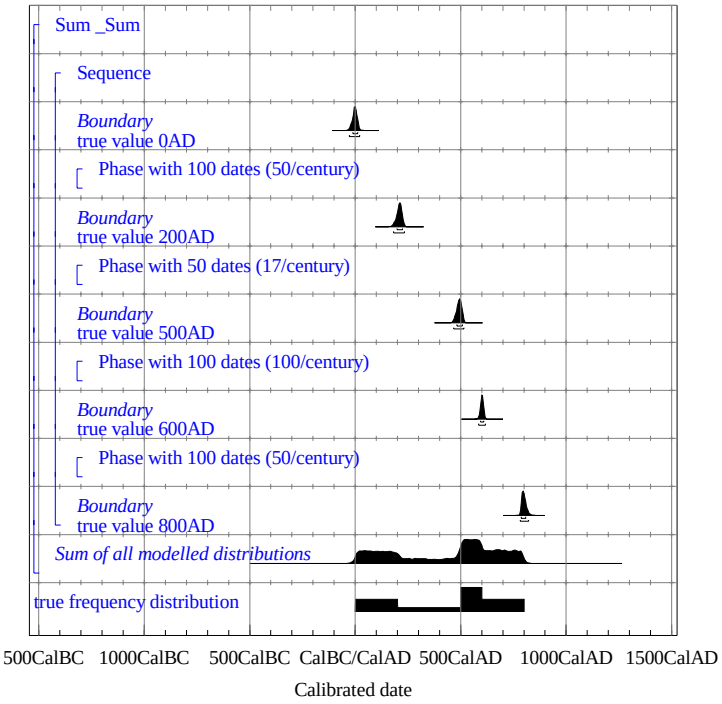

Figure 3 This shows how the position of boundaries can be reconstructed by the simulated radiocarbon dating of events within the phases they divide; the Sum, which provides a frequency modulated age distribution, provides a good approximation to the true frequency distribution of the dates.

Effective Priors

Another, slightly more theoretical approach is to look at what result is obtained from a model when virtually no dating information is given at all. This then shows us what the effective prior for the model is. This method can be used to see if the results of an analysis are largely artifacts of a particular model rather than arising from the ^{14}C dates and other chronological information.

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

As an example of this we will look at a very simple model like:

```
Sequence
{
 C_Date 0;
 Sequence "with six events and three boundaries"
 {
  Boundary;
  Event; Event; Event;
  Boundary;
  Event; Event; Event;
  Boundary;
  Span;
 };
 C_Date 100;
};
```

Here we use the Event function to denote an undated event; the span function will calculate the total span of the sequence. The C_Date terms constrain the whole sequence to within the first century AD.

Figure 4 shows the generated distributions for the spans of various sequences of this kind. The algorithm is designed to provide a uniform prior for the span and it can be seen that it does this satisfactorily, regardless of the number of events. Because of details of the algorithm (see section on MCMC above), however, there is still a bias against very short overall spans if the sequence is divided up with a large number of *boundaries*. This fact should be taken into account when modeling very short duration events. In such a circumstance a model with fewer *boundaries* (ideally just two) should be tried

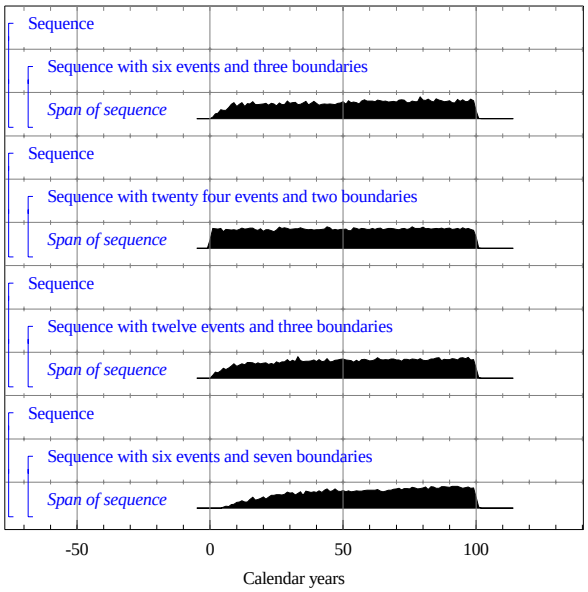

Figure 4 This shows the effective priors for spans of sequences with different numbers of events and boundaries; you can see from this that the prior for the span is essentially uniform but that if the number of boundaries gets high, the model does still bias against very short overall spans—this is due to details of the MCMC sampling algorithm

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press

to see if this gives radically different results. More generally this kind of sensitivity testing, to see how model-dependent the results are, is very useful.

CONCLUSIONS

There has been considerable development of the OxCal program over the last six years. The main developments have been in the user interface (not covered here) and in the operation of the Markov Chain Monte Carlo algorithm. The fundamental aspects of the mathematics underlying the software have not changed. However, the notion of *boundaries* has been further developed to make the models independent of the number of subdivisions (in practice this is not quite perfect) and external constraints. A new treatment of nested sequences and phases has also been developed. Tools are now provided to help with convergence and to ensure that the program continues to run until convergence is satisfactory.

For simple calibration, the algorithm has been modified to take better account of variation in the calibration curve uncertainty and provision has been made for the mixing of calibration curves.

REFERENCES

Bronk Ramsey C, MJ Allen. 1999. Analysis of the radiocarbon dates and their archaeological significance. In: Cleal RMJ, Walker KE, Montague R, editors. *Stonehenge in its landscape: twentieth century excavations.* London, English Heritage. p 526–35.

Bronk Ramsey C. 1994. Analysis of chronological information and radiocarbon calibration: The program OxCal. *Archaeological Computing Newsletter* 41:11–6.

Bronk Ramsey C. 1995. Radiocarbon calibration and the analysis of stratigraphy. *Radiocarbon* 37(2):425–30.

Bronk Ramsey C. 1998. Probability and dating. *Radiocarbon* 40(1):461–74.

Buck CE, Kenworthy JB, Litton CD, Smith AFM. 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65: 808–21.

Buck CE, Litton CD, Smith AFM. 1992. Calibration of radiocarbon dates pertaining to related archaeological events. *Journal of Archaeological Science* 19:497–512.

Buck CE, Litton CD, Scott EM. 1994. Making the most of radiocarbon dating: some statistical considerations. *Antiquity* 68:252–63.

Gilks WR, Richardson S, Spiegelhalter DJ. 1996. *Markov Chain Monte Carlo in practice*. Chapman and Hall, London.

Needham S, Bronk Ramsey C, Coombs D, Cartwright C, Pettitt PB. 1998. An independent chronology for British Bronze Age metalwork: the results of the Oxford Radiocarbon Accelerator Programme. *Archaeological Journal* 154:55–107.

Steier P, Rom W, Puchegger S. 2001. Critical aspects of Bayesian mathematics in ^{14}C calibration. *Radiocarbon.* This issue.

Stuiver M, Reimer PJ. 1993. Extended 14C database and revised CALIB radiocarbon calibration program. *Radiocarbon* 35:215–30.

Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac G, van der Plicht J, Spurk M. 1998. INTCAL98 radiocarbon age calibration, 24000–0 cal BP. *Radiocarbon* 40(3):1041–83.

https://doi.org/10.1017/S0033822200038212 Published online by Cambridge University Press