RADIOCARBON CALIBRATION AND ANALYSIS OF STRATIGRAPHY: THE OxCal PROGRAM

CHRISTOPHER BRONK RAMSEY

Oxford University Radiocarbon Accelerator Unit, 6 Keble Road, Oxford OX1 3QJ, England

ABSTRACT. People usually study the chronologies of archaeological sites and geological sequences using many different kinds of evidence, taking into account calibrated radiocarbon dates, other dating methods and stratigraphic information. Many individual case studies demonstrate the value of using statistical methods to combine these different types of information. I have developed a computer program, OxCal, running under Windows 3.1 (for IBM PCs), that will perform both ^{14}C calibration and calculate what extra information can be gained from stratigraphic evidence. The program can perform automatic wiggle matches and calculate probability distributions for samples in sequences and phases. The program is written in C++ and uses Bayesian statistics and Gibbs sampling for the calculations. The program is very easy to use, both for simple calibration and complex site analysis, and will produce graphical output from virtually any printer.

INTRODUCTION

The use of radiocarbon calibration programs (such as Stuiver and Reimer (1993) or van der Plicht (1993) and the program described here) to show the relation between ^{14}C and calendar dates is well established. Archaeologists are now familiar with using the probability distributions generated by these programs to show the probable age ranges of dated material. Straightforward calibration is all that is required if the objects dated have no other chronological or stratigraphic information associated with them. In most cases, however, the archaeologist must consider other information relevant to the chronology along with the evidence from ^{14}C dating. The evaluation of this information is a complex operation involving considerable archaeological skill and insight. Recently, some new mathematical methods have been applied to archaeological problems to demonstrate the use of statistical methods (Buck *et al.* 1991; Buck, Litton and Smith 1992; Buck, Litton and Scott 1994; Christen *et al.* 1996; Buck, Christen and Litton 1996). A chronological model for the site (based on stratigraphy and other similar evidence) is planned by the archaeologist and combined with the ^{14}C dating evidence. From this, one can infer much information, such as the probable date of an individual event, phase durations and sequencing.

OxCal uses the techniques described above, with some important extensions, to test the internal consistency and reliability of the model and to treat tree rings and other types of uniform deposit. In addition to providing a tool for post-excavation analysis, the program includes a facility for the simulated dating of an archaeological site. This helps the archaeologist or ^{14}C laboratory to determine 1) the number of ^{14}C dates needed to answer a particular question, 2) optimum sampling locations within the archaeological sequence, and 3) whether the calibration curve is likely to cause problems of interpretation.

RADIOCARBON CALIBRATION

The first stage to any analysis involving ^{14}C dates is calibration. OxCal can be used as a simple calibration program. Ranges can be evaluated using either the intercept or the probability method. ΔR reservoir corrections can also be made to marine curves (Stuiver and Braziunas 1993). The calibration data files from either the Groningen program (van der Plicht 1993) or CALIB (Stuiver and Reimer 1993) can be used (these data files are compiled from information in Stuiver and Kra (1986) and Stuiver, Long and Kra (1993)). The results generated by this program are very close to those generated by other programs when using the same data sets.

Proceedings of the 15th International ^{14}C Conference, edited by G. T. Cook, D. D. Harkness, B. F. Miller and E. M. Scott. RADIOCARBON, Vol. 37, No. 2, 1995, P. 425–430

Chronological Models

OxCal has been designed so that it is easy to build chronological models for combination with dating evidence. My intention was to provide a program that can be used by people with a good knowledge of the archaeological evidence who do not wish to get involved with the details of mathematical procedures. Archaeological judgment is very important in making deductions from stratigraphic information because objects in a stratigraphic sequence need not be in chronological order when there is residual or intrusive material present.

In addition to ^{14}C dates, other dating evidence (*e.g.*, coins, historical evidence, thermoluminescence and other radiometric dates) can be used. ^{14}C dates are automatically calibrated and all subsequent stages of the analysis are carried out on the calendar scale. Also built into the program are methods for combining dates (before or after calibration) and the associated χ^2 statistical tests.

The chronological model consists of relations among individual items (Harris 1989) which are, however, more easily expressed in terms of sequences and phases. In this context, a phase means a group of elements for which there are no fixed relations; a sequence is a group of elements in a given order. Within each of these, the individual elements can themselves be sequences or phases.

Figure 1 shows how the program displays such information. Figure 1A shows the simple calibrated ^{14}C dates along with their relations. Figure 1B shows the results of the analysis on which the chronological relations are imposed. The definition of a model is normally performed through the program's graphical user interface but it is actually stored as a command file as shown in Figure 2; command files can be edited as text or generated by other programs.

Besides the nesting of phases and sequences, the program allows for more complex relations, *e.g.*, *termini ante quem*, *termini post quem*, cross-linking between sequences and minimum gaps between sequence elements. Together, these should allow the program to cope with almost any situation.

Sometimes a more specific model is required, such as wiggle matching measurements on tree rings. This type of information can be analyzed with OxCal using age separations that are known either exactly or approximately. These methods are also applicable to certain deposits, such as varved lake sediments.

The treatment of archaeological phases also requires more specific models. The most widely applicable model operates on the assumption that the dated samples were uniformly sampled from the phase in question; this alters the way in which the elements of the phase are treated and enables time boundaries for the phase to be estimated. This method should be used with caution as, frequently, the samples derive from one construction or destruction level and are not evenly distributed.

Information from the Model

The methods described above can do much more than simply re-evaluate the chronology of the dated events. Several different types of question can be answered during the analysis.

From a series of dated events, one can extract valuable information, such as the probability distribution for the dates of the first and last events and a determination of their chronological span. One can also generate a probability distribution for the time difference between any two events. The likely order of dated events can be found where the individual probability distributions overlap (Buck, Litton and Scott 1994). OxCal will generate a list of possible sequences along with their probabilities.

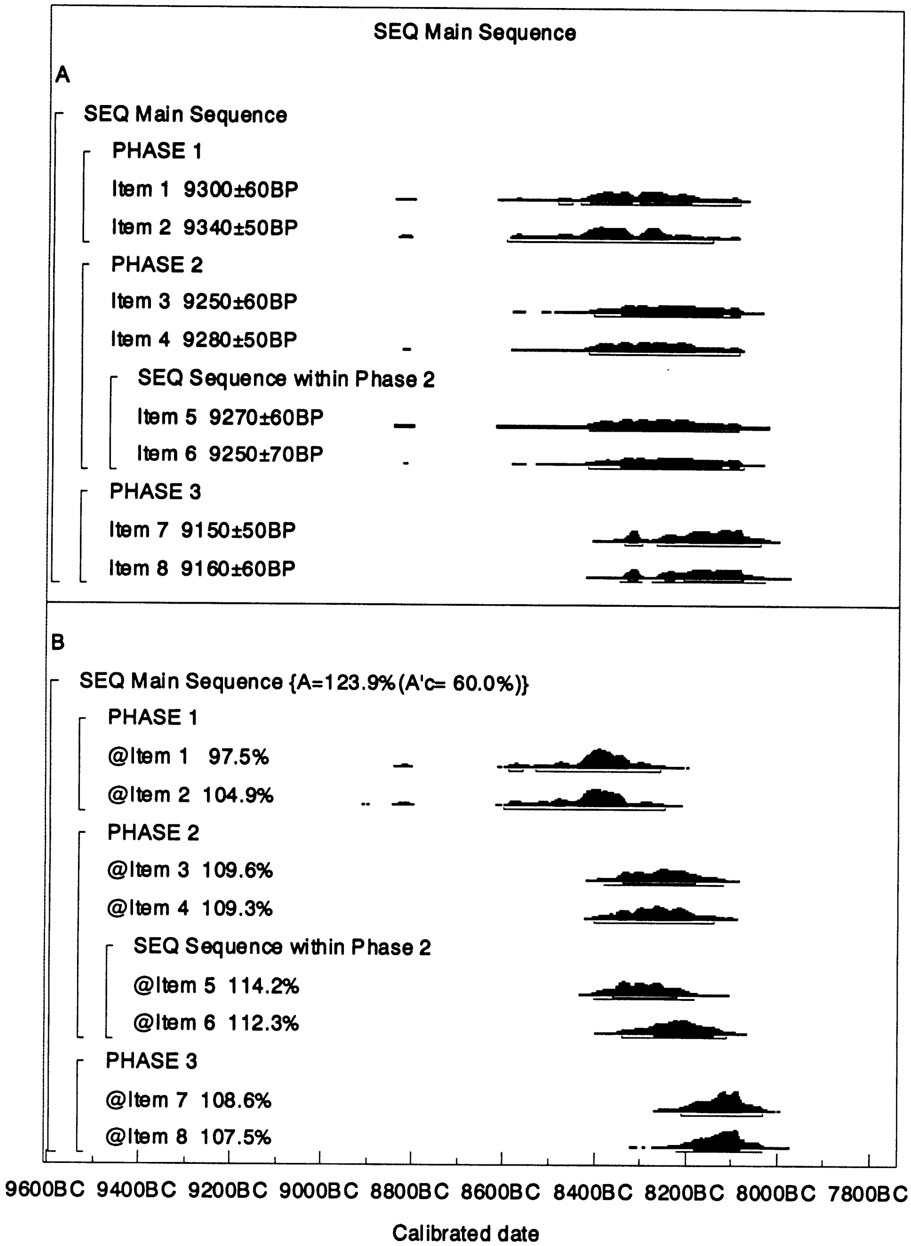

Fig 1. Results of OxCal analysis showing, A, the simple calibrated results and, B, the effect of imposing the chronological model

If samples are uniformly distributed, time boundaries and intervals between boundaries of archaeological phases can be estimated.

In using this type of analysis, one risks imposing a model inconsistent with the dating evidence. Thus, checks must be made both on individual items and on the model as a whole to ensure a reasonable level of consistency between the dating evidence and the other information. For each dated item,

```
SEQ "Main Sequence"
{
 PHASE "1"
 {
  DATE "Item 1" 9300 60;
  DATE "Item 2" 9340 50;
 };
 PHASE "2"
 {
  DATE "Item 3" 9250 60;
  DATE "Item 4" 9280 50;
  SEQ "Sequence within Phase 2"
  {
   DATE "Item 5" 9270 60;
   DATE "Item 6" 9250 70;
  };
 };
 PHASE "3"
 {
  DATE "Item 7" 9150 50;
  DATE "Item 8" 9160 60;
 };
};
```

Fig 2. Example of a model-definition command file

an agreement index A is calculated, which has a value of *ca.* 100% but which is sometimes higher and might fall as low as 60%. An agreement index $A_{overall}$ is also calculated for the model as a whole, which again is typically *ca.* 100% and should not fall below 60%. The thresholds for these indices are chosen so that they are close to the 5% confidence level of the χ^2 test for simple combinations.

The application of these agreement indices allows the user to test for intrusive dates or unreliable chronological models. If an individual item does not agree well with a model that otherwise seems reasonable, one can query its position. The program will then calculate the probability that the sample's position is, indeed, correct.

Formulation of Dating Strategy

Probably the most important stage in any dating program is the formulation of a coherent dating strategy. OxCal has features designed to help the archaeologist with this. In particular, the program can show how many ^{14}C dates are needed to answer particular questions, whether ^{14}C dating is capable of answering a question at all given the calibration curve, and which samples are best selected to gain most information. This can be achieved by using a function that randomly generates simulated ^{14}C dates. If, for example, one has an object believed to date from 2000 BC and the ^{14}C laboratory reports dates for this period with a precision of 50 yr, the program will generate dates such as

$$3697 \pm 50 \quad 3707 \pm 50 \quad 3634 \pm 50 \quad 3672 \pm 50 \, .$$

https://doi.org/10.1017/S0033822200030903 Published online by Cambridge University Press

Using these with ^{14}C calibration, one can see the number of dates needed to obtain suitably short ranges in calendar years or, with models of the type described above, which samples should be dated to obtain a sufficiently accurate estimate for the duration of a phase.

PROGRAM DETAILS

Full details of the program along with the mathematical methods used are given in the manual, which is available, along with the program, from the author. The calibration is performed in much the same way as in other programs (*e.g.*, Stuiver and Reimer 1993; van der Plicht 1993). In OxCal, the uncertainties associated with the calibration curve are taken into account and a cubic interpolation method is used between the data points. Combinations prior to calibration are performed using the usual weighted average and χ^2 test (Shennan 1988: 65). After calibration, combinations of different types of dates are made by multiplying the probability distributions; a very similar method is also employed for wiggle matching by offsetting the probability distributions on the calendar scale.

To incorporate the other information, OxCal uses a technique called *Gibbs Sampling*, which involves a large number of iterations. Buck *et al.* (1991), Buck, Litton and Smith (1992) and Buck, Litton and Scott (1994) describe this in some detail. I have made extensions to improve model consistency and convergence testing, which are particularly important for a program intended for general release. The test for model consistency is based on an agreement index A, which, for a prior distribution p(t) (the probability distribution before the chronological model has been taken into account), and a posterior distribution p′(t) (the distribution given the model), is defined as

$$A = \frac{\int p(t)\, p'(t)\, dt}{\int p(t)\, p(t)\, dt} \quad . \tag{1}$$

A suitable acceptance threshold for this is *ca.* 60% for a similar level of discrimination as the χ^2 test at 5%. An overall agreement index is also defined for the model as a whole, given by

$$A_{overall} = \left[\prod_{i=1}^{n} A_i\right]^{1/(\sqrt{n})} \tag{2}$$

where A_i is the agreement index for the i'th of n events. This has a similar threshold of 60%. Convergence tests are also needed to see if the Gibbs Sampling procedure is giving truly representative results. Such tests are performed by this program using an overlap integral C between the accumulated probability distribution P(t) and that for the last set of iterations p(t)

$$C = \frac{\left(\int p(t)\, P(t)\, dt\right)^2}{\int P^2(t)\, dt \int p^2(t)\, dt} \quad . \tag{3}$$

Experience suggests that, if this integral is ever <95%, the algorithm is probably unstable and the results of the analysis should not be used. Another extension to the application of Gibbs Sampling to this type of problem is the inclusion of a method for wiggle matching in cases where the age gaps among items are known only approximately.

CONCLUSION

The computer program OxCal allows the user to perform many different kinds of analysis on ^{14}C dates, from simple calibration to analysis of entire archaeological sites including information from stratigraphy and other dating methods. Complex chronological models can be built and imposed on the dating evidence to provide a coherent picture. The graphical output from the program shows both the details of the model imposed and the results of the analysis in such a way that others should be able to reproduce the results easily. The program should be useful for archaeologists and ^{14}C labs both in the planning of dating programs and in the analysis of the dating evidence from archaeological sites.

OxCal runs under Windows 3.1 and is available over the Internet on the World Wide Web page: `http://www.ox.ac.uk/depts/rlaha/`, or from the author.

REFERENCES

Buck, C. E., Christen, J. A., and Litton, C. D. 1996 Archaeology, modeling and radiocarbon. *Radiocarbon*, in press.

Buck, C. E., Kenworthy, J. B., Litton, C. D. and Smith, A. F. M. 1991 Combining archaeological and radiocarbon information: A Bayesian approach to calibration. *Antiquity* 65: 808–821.

Buck, C. E., Litton, C. D. and Smith, A. F. M. 1992 Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19: 497–512.

Buck, C. E., Litton, C. D. and Scott, E. M. 1994 Making the most of radiocarbon dating: Some statistical considerations. *Antiquity* 68: 252–263.

Christen, J. A., Kenworthy, J. B., Ottaway, B. S., Buck, C. E. and Litton, C. D. 1996 Archaeological interpretation of a suite of radiocarbon determinations: A Bayesian case study. *Radiocarbon*, in press.

Harris, E. C. 1989 *Principles of Archaeological Stratigraphy*, 2nd edition. London, Academic Press: 170 p.

Shennan, S. 1988 *Quantifying Archaeology*. Edinburgh, Edinburgh University Press: 364 p.

Stuiver, M. and Braziunas, T. F. 1993 Modeling atmospheric ^{14}C influences and ^{14}C ages of marine samples to 10,000 BC. *In* Stuiver, M., Long, A. and Kra, R. S., eds., Calibration 1993. *Radiocarbon* 35(1): 137–189.

Stuiver, M. and Kra, R.S., eds. 1986 Calibration Issue, Proceedings of the 12th International ^{14}C Conference. *Radiocarbon* 28(2B): 805–1030.

Stuiver, M. and Reimer, P. J. 1986 A computer program for radiocarbon age calculation. *In* Stuiver, M. and Kra, R.S. eds., Calibration Issue, Proceedings of the 12th International ^{14}C Conference. *Radiocarbon* 28 (2B): 1022–1030.

Stuiver, M., Long, A., and Kra, R. S., eds. 1993 Calibration 1993. *Radiocarbon* 35(1): 1–244.

Stuiver, M. and Reimer, P. J. 1993 Extended ^{14}C data base and revised CALIB 3.0 ^{14}C age calibration program. *In* Stuiver, M., Long, A., and Kra, R. S., eds., Calibration 1993. *Radiocarbon* 35(1): 215–230.

van der Plicht, J. 1993 The Groningen radiocarbon calibration program. *In* Stuiver, M., Long, A., and Kra, R. S., eds., Calibration 1993. *Radiocarbon* 35(1): 231–237.