RECENT AND PLANNED DEVELOPMENTS OF THE PROGRAM OXCAL

Christopher Bronk Ramsey[1] • Sharen Lee

Research Laboratory for Archaeology and the History of Art, University of Oxford, Dyson Perrins Building, South Parks Road, Oxford OX1 3QY, United Kingdom.

ABSTRACT. OxCal is a widely used software package for the calibration of radiocarbon dates and the statistical analysis of ^{14}C and other chronological information. The program aims to make statistical methods easily available to researchers and students working in a range of different disciplines. This paper will look at the recent and planned developments of the package. The recent additions to the statistical methods are primarily aimed at providing more robust models, in particular through model averaging for deposition models and through different multiphase models. The paper will look at how these new models have been implemented and explore the implications for researchers who might benefit from their use. In addition, a new approach to the evaluation of marine reservoir offsets will be presented. As the quantity and complexity of chronological data increase, it is also important to have efficient methods for the visualization of such extensive data sets and methods for the presentation of spatial and geographical data embedded within planned future versions of OxCal will also be discussed.

INTRODUCTION

OxCal as a computer program grew out of needs within the Oxford Radiocarbon Accelerator Unit for methods to calibrate and analyze ^{14}C dates. It was first made available for more general use in 1994 (Bronk Ramsey 1995). Subsequent developments (Bronk Ramsey 2001, 2008, 2009a,b) have focused on incorporating new statistical methods and changes to the user interface. The models are all defined using a Chronological Query Language (CQL: Bronk Ramsey 1998), which enables the models to be easily circulated and published. The program continues to be central to much of the research of the Oxford group (e.g. Bronk Ramsey et al. 2010, 2012; Higham et al. 2011), and further development tends to follow the needs for the projects in which they are involved. Given the widespread use of the program (with over 27,000 sessions by over 3000 users in the run-up to the Paris Radiocarbon conference), the intention is to continue the model of public free access to this resource.

The user interface is already open-source, which has enabled the production of a Japanese version of the program by Takayki Omori (http://www.nendai.nagoya-u.ac.jp/oxcal/index.html) and, in the long run, the intention is to make the entire source code open-access (GNU GPLv3) so that it can be compiled on different platforms and can be further developed by others.

The recent and planned developments reported here can be divided into 3 main categories: provision of more flexible prior probability definitions; provision of new, more robust statistical models; and the development of data presentation approaches.

MORE FLEXIBLE PRIOR PROBABILITY DEFINITIONS

The essence of the Bayesian method of analysis is the inclusion of prior probability. In OxCal, this is mostly performed by the definition of chronological models such as `Sequence`, `Phase`, etc. (Bronk Ramsey 2009a). The program also allows for the entry of likelihood information for individual parameters both in specific forms such as those for ^{14}C dating, and in more general ways that would be more appropriate for other dating methods:

```
R_Date("14C-1",3000,30);
Date("OSL-1",N(-1000,150));
```

[1]Corresponding author. Email: christopher.ramsey@rlaha.ox.ac.uk.

© 2013 by the Arizona Board of Regents on behalf of the University of Arizona
Proceedings of the 21st International Radiocarbon Conference edited by A J T Jull & C Hatté
RADIOCARBON, Vol 55, Nr 2–3, 2013, p 720–730

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

where `R_Date` calculates the likelihood distribution for the calibrated date as a function of ^{14}C concentration, while `Date` allows to provide the probability density function (PDF) of the date of an event directly.

"14C-1" and "OSL-1" are the names of events; the last 2 arguments of R_Date are the ^{14}C age and its uncertainty, while N(μ,σ) generates a Normal distribution PDF with mean μ and standard deviation σ. However, information to be included in a model might also include prior information on the values of specific parameters, and this has been possible in previous versions of OxCal for phase boundaries and intervals, so that you might for example have

```
Boundary("B1",U(1066,1230));
```

where a uniform prior has been applied extending from the start of AD 1066 to AD 1230, only allowing this boundary to lie between those specific dates, or:

```
Interval("I1",N(100,30));
```

specifying the prior PDF for this interval between 2 events with a mean of 100 and a standard deviation of 30.

OxCal adds another built-in distribution for such definitions: the Poisson distribution. This is defined by the mean, and can be given an optional scaling parameter. Thus:

```
Pois(100);
Pois(100,2);
```

The first will give a mean of 100 and a standard deviation close to 10 (being similar to a Normal distribution). The Poisson distribution is a discrete distribution with each occurrence or event given a value of 1 and is defined by the mean number of events (100 in the examples here). The scaling parameter (2 in the second example) assigns a different value to the discrete events (2 in this case, thus giving a mean of 200 and a standard deviation close to 20). Although the PDF generated is discrete, it will be interpolated. This prior may be more appropriate for intervals than the Normal distribution in cases where you know that the value cannot be negative. A significant difference to a normal distribution is only seen if the mean is less than 10. A command, `LnN`, has also been added to provide log-normal distributions that might be used in similar circumstances.

OxCal v 4.2 also extends the easy definition of priors to other queries and so expressions of prior probability can be added to any of the commands: `Difference`, `First`, `Last`, `Shift`, and `Span`. Examples are

```
First("F1",U(-1000,-900));
Last("L1",U(1066,1230));
Shift("C","A",B",U(1066,1230));
Difference("D","A","B",N(0,100));
```

It was possible to define these prior constraints or probabilities under previous versions using cross-referencing, but this should provide a more convenient formalism.

In previous versions of OxCal, `Delta_R` and `Mix_Curves` could only have normally distributed priors. For models where Markov chain Monte Carlo (MCMC) is used, it is now possible to define different distributions. To do this, you replace the mean and standard deviation with a distribution expression, so for example you can have

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

```
Delta_R("Local1",400,200);
Delta_R("Local2",N(400,200));
Delta_R("Local3",U(0,800));
```

The first 2 definitions are identical, simply expressed in a different way. The third (Local3) provides an uniform prior for ΔR within the limits of 0 and 800. This allows models to be constructed that extract estimates for ΔR from the data in the model. It is still necessary to have some overall constraints on the parameter, and for this purpose the uniform distribution can be used. Extensive use of this method is made in Bronk Ramsey et al. (2012). Similarly, for the command `Mix_Curve` you can have

```
Mix_Curve("Mixed1","IntCal09","Marine09",50,10);
Mix_Curve("Mixed2","IntCal09","Marine09",N(50,10));
Mix_Curve("Mixed3","IntCal09","Marine09",U(0,100));
```

The parameters at the end of this function define the percentage contribution from the second reservoir (constrained under MCMC to be in the range 0–100). The first 2 definitions here are identical with a normally distributed prior for the mixture with a mean of 50 and standard deviation of 10. The third allows the model complete freedom to have any mixture. Like the `Delta_R` example above, this now provides the opportunity to use an unbiased estimate for reservoir mixtures. In principle, this approach can be used to extract dietary information from a chronological model without making assumptions based, for example, on stable isotope values.

MORE ROBUST STATISTICAL MODELS

Trapezium Phase Modeling

For many applications, it is possible to construct chronological models from fairly simple building blocks such as multiple phases with abrupt transitions (Buck et al. 1992; Bronk Ramsey 2009a), or deposition models with a defined flexibility (Bronk Ramsey 2008). However, in some circumstances such underlying assumptions do not reflect the prior knowledge as well as one would like.

An important addition to OxCal v 4.2 is the addition of trapezium-based phase models (Karlsberg 2006; Lee and Bronk Ramsey 2012). These allow for a gradual start and end to single phases, as in the original conception of the single trapezium, or gradual transitions between phases in multiple phase models. This form of multiple phase model is likely to have many different applications where the real world phases are not really expected to have abrupt starts or ends (see examples in Lee and Bronk Ramsey 2012; Lee et al. 2013b, these proceedings).

To implement trapezium models in OxCal, existing multi- or single-phase models simply have to have the `Boundary` command modified. Conceptually, the only change to the model from the standard uniform phase model is that the boundaries between phases are no longer abrupt but have a transition time associated with them. To indicate this in the model, all that is required is a `Transition` command nested within the `Boundary`. However, it is also often useful to add `Start` and `End` queries that recover the start and end of the transition period. The command thus looks like:

```
Boundary("B")
{
 Transition("T");
 Start("S");
 End("E");
};
```


https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

When such a model is run, the `Boundary` "B" will recover the mid-point of the transition, the `Transition` "T" will recover the duration of the transition, and the optional `Start` "S" and `End` "E" will return the start and end dates of the transition, respectively. All 3 of the new commands (`Transition`, `Start`, and `End`) can be given prior distributions. For example, if one wishes to define that the transition must take place within 100 yr one can use

```
Boundary("B")
{
 Transition("T",U(0,100));
};
```

The simple phase model can be seen as a special case of the trapezium-based model but where the transition period is defined to be very short. The trapezium-based model is intrinsically more flexible.

Variable Rigidity for Poisson-Process Modeling

Another development that allows more flexibility into the modeling is in the case of the Poisson process model (Bronk Ramsey 2008). In this model, the rigidity (defined by the k parameter) is usually defined exactly. However, an extension to this model was implemented but not documented in v 4.1 of OxCal. The reasonable values for the k value might vary by orders of magnitude, depending on the sedimentary context. For this reason, the program allows for the specification of a variable k, in place of a fixed k and the prior for this is expressed in terms of the $v = \log_{10}(k/k_0)$ to allow for variability over a wide range. Mathematically, we now have:

$$k = k_0\, 10^v$$

$$v \sim D$$

where D is some prior distribution, typically uniform over a reasonable range. The specification of the variability in k is given by the addition of another parameter to the `P_Sequence` command. The parameters are

```
P_Sequence(name,k0,p,D)
```

where *name* is the label for the sequence, k_0 is the base k parameter, p is the interpolation rate, and D is the prior distribution for v (see above). Thus, if we have the depth of a sedimentary sequence defined in terms of depth in cm, for example:

```
P_Sequence("P",1,3,U(-2,2))
```

would define $k_0 = 1$ cm^{-1}, interpolation rate = 3 cm^{-1} (that is, output from the model given every 3.3 mm), and variability in k allowed between a factor of 10^{-2} and 10^2. In other words, this model no longer has a fixed k but will average over values of k between 0.01 and 100 cm^{-1}. This should provide a robust model for any sedimentary sequence. If the depth is given in terms of meters instead, a reasonable $k_0 = 100$ m^{-1} and the model can be modified to:

```
P_Sequence("P",100,300,U(-2,2))
```

In practice, the most probable values of k are defined by the data, and in many cases it turns out that the marginal posterior for $\log_{10}(k/k_0)$ is Normal—that is, that k is defined by a log-normal distribution. If you have several similar sedimentary sequences, it may be possible to learn from this the most appropriate distribution for the multiplier. This can be entered in place of the uniform distribution in the model definition.

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

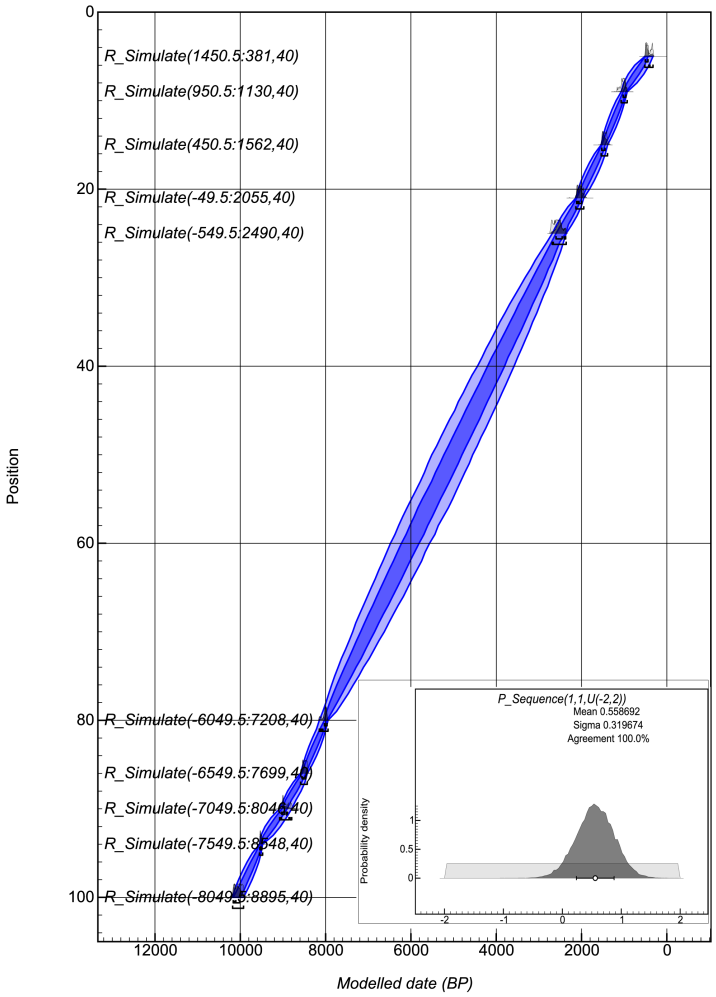


Figure 1 Use of the variable k parameter for Poisson process modeling. This simulated data set is chosen to represent fairly linear deposition. The inset graph shows the inferred posterior distribution for $\log_{10}(k/k_0)$, in this case suggesting that the most realistic values for k are well above k_0.

Figure 1 shows how this works with a simulated data set constructed so that the underlying deposition only has small deviations from what you would expect for a constant deposition rate. The model used is defined by the CQL code:

```
P_Sequence("",1, 1,U(-2,2))
{
 Boundary();
 R_Simulate("",calBP(10000),40){z=100;};
 R_Simulate("",calBP(9500),40){z=94;};
 R_Simulate("",calBP(9000),40){z=90;};
 R_Simulate("",calBP(8500),40){z=86;};
 R_Simulate("",calBP(8000),40){z=80;};
 R_Simulate("",calBP(2500),40){z=25;};
```


https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

```
 R_Simulate("",calBP(2000),40){z=21;};
 R_Simulate("",calBP(1500),40){z=15;};
 R_Simulate("",calBP(1000),40){z=9;};
 R_Simulate("",calBP(500),40){z=5;};
 Boundary();
};
```

What you can see in Figure 1 is that the interpolation in the large undated middle region of the deposition sequence is fairly well defined. As a consequence, the modeling finds that a k value higher than 1 is most appropriate (the mean distribution for $\log_{10}(k)$ is greater than zero). In Figures 2 and 3, the data are chosen to reflect increasingly variable deposition rates. For this reason, the estimated posterior value for k falls and the uncertainty in the interpolated age-depth model increases. This ability of the model rigidity to respond to the data is an advantage of this approach for many situations where there is no direct information on the linearity of the deposition rate.

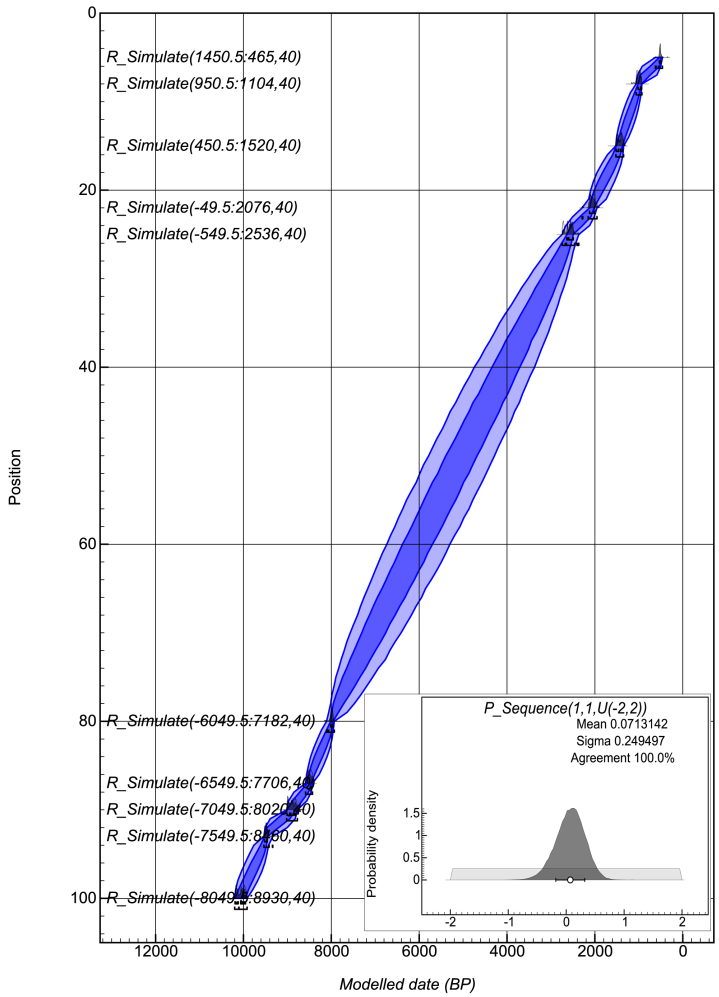


Figure 2 Compared to Figure 1, the deposition rate is now more variable and the posterior for $\log_{10}(k/k_0)$ is close to zero, suggesting that k_0 is a reasonable estimate for k.

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

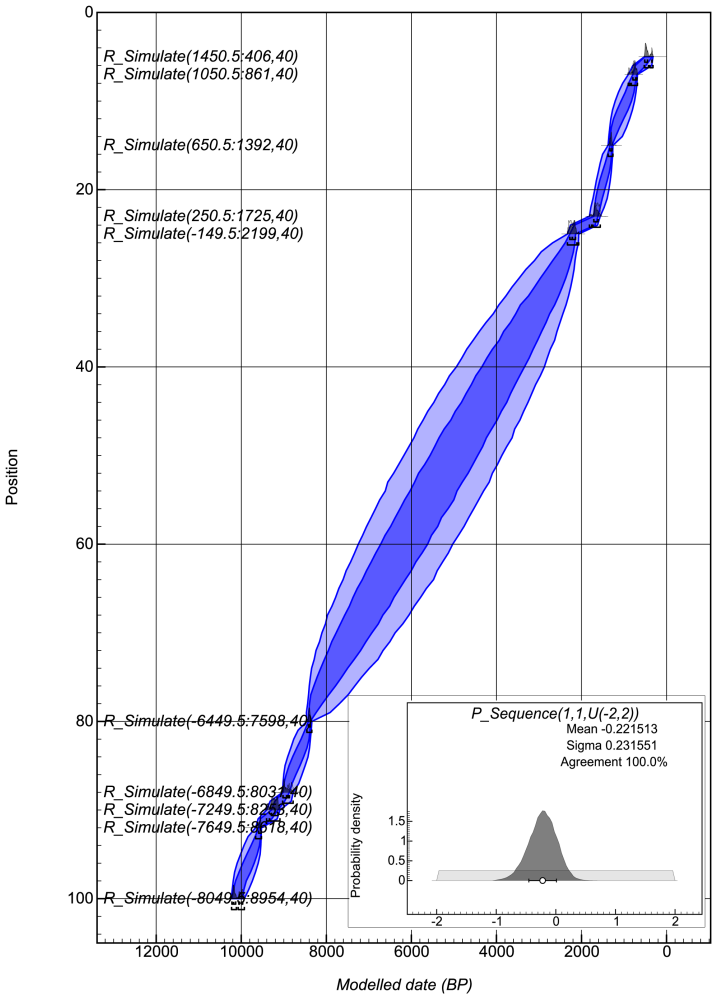


Figure 3 Compared to Figures 1 and 2, the deposition rate here is even more variable and the posterior for $\log_{10}(k/k_0)$ falls below zero, suggesting that the most appropriate k values lie below k_0.

In general, this approach to modeling best falls into what might be termed a "model averaging" approach, in that the outputs from the model are essentially a realistic average over different possible scenarios where the sedimentation rate is more or less uniform over time. Practical examples of the use of this more flexible prior are given in Lee et al. (2013a, these proceedings).

Covariance Information

The final significant addition to the modeling component in OxCal is the ability to pull out covariance information from the MCMC modeling stage. Two commands have been added to facilitate this. The syntax is the same as that for other summary queries (`First`, `Last`, `Sum`, `Order`) in that the query can be set to either refer to specific parameters or, if entered within a group, can pull out covariance information from all elements of, for example, a `P_Sequence`, `Phase`, or `Sequence`. The 2 new commands are `Covar_Matrix`, which produces a covariance matrix for the selected parameters, and `Correl_Matrix`, which produces instead a matrix of Pearson prod-

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

uct-moment correlation coefficients. This can be very useful in seeing the degree to which parameters with the model are correlated with one another. The program retains the `Correlation` command that produces a visual plot of the correlation between 2 parameters (but does not calculate a correlation coefficient).

The following CQL command snippet shows how these queries are to be used in generating the relevant matrices for all parameters within a `P_Sequence`:

```
P_Sequence("P",1,3,U(-2,2))
{
 Covar_Matrix();
 Correl_Matrix();
 ...
};
```

Note that for large models these matrices can become unmanageably large.

DATA PRESENTATION

The third area of development is in the tools for data presentation. All graphical data presentation is in the form of scalable vector graphics (SVG), with online tools available for conversion to portable document (PDF) and raster (PNG) formats.

Now that there is a very wide range of statistical tools available within the analysis part of the program, there is demand for a wider range of methods to present the output easily. There are 2 ways to do this: one is to interface to other packages and the other is to provide further graphical tools within the OxCal toolkit itself. Both approaches have their benefits and development in both of these areas is envisaged over the next few years. It makes most sense to develop tools within OxCal that are likely to find widespread use throughout the research community so that the time saved by researchers (not having to manually manipulate large amounts of data) is significant.

For data linkage with other packages, there are a number of features of the overall structure that are designed to help. The online version of the program runs under PHP and from v 4.2 the PHP code will also be released in the downloaded package, excepting those parts that control logging in and out. This will mean that users can install a single-user OxCal server on their local machines—and if they have the IT expertise can also implement a multi-user version. OxCal in this form can be integrated with databases and other software packages. The file output routines allow data to be exported in spreadsheet (CSV) format, which can be imported into other applications. Since all of the output information is also coded into the JavaScript® (JS) output files, some users have also sorted out ways to extract information themselves directly from the output.

The OxCal analysis package is set up to allow for additional information included within the commands. Some information is used within the program itself (such as depth information given by the *z* attribute), but, if OxCal keywords are avoided, any other information can be added and will be passed to the output files. This is particularly useful for data presentation purposes.

As an example of this, v 4.2 includes a mapping package that allows the output from the program to be presented on maps. This can act as a rudimentary geographic information system (GIS). Items need to be tagged with position information, and optionally also the color and marker required on the map:

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

```
R_Date("A",4514,38)
{
 longitude=3.004;
 latitude=47.894;
 color="blue";
 marker="triangle";
};
```

There, once the analysis has been run, you can use the menu item [View > Show on map] to put your data on a map. Underlays can be chosen between a simple gray coastline or any of the Google® map formats. The map can be displayed either as an interactive single map or as a time-slice map sequence (see Figure 4). In the interactive map, you can play a sequence that will show how the probabilities of individual data points change through time, or you can select a particular time using a slider. The representation of probability depends on the marker type chosen. If no marker is chosen, circles are used, which change in size depending on the probability; if a marker is selected, then its degree of opacity is used to show the probability.

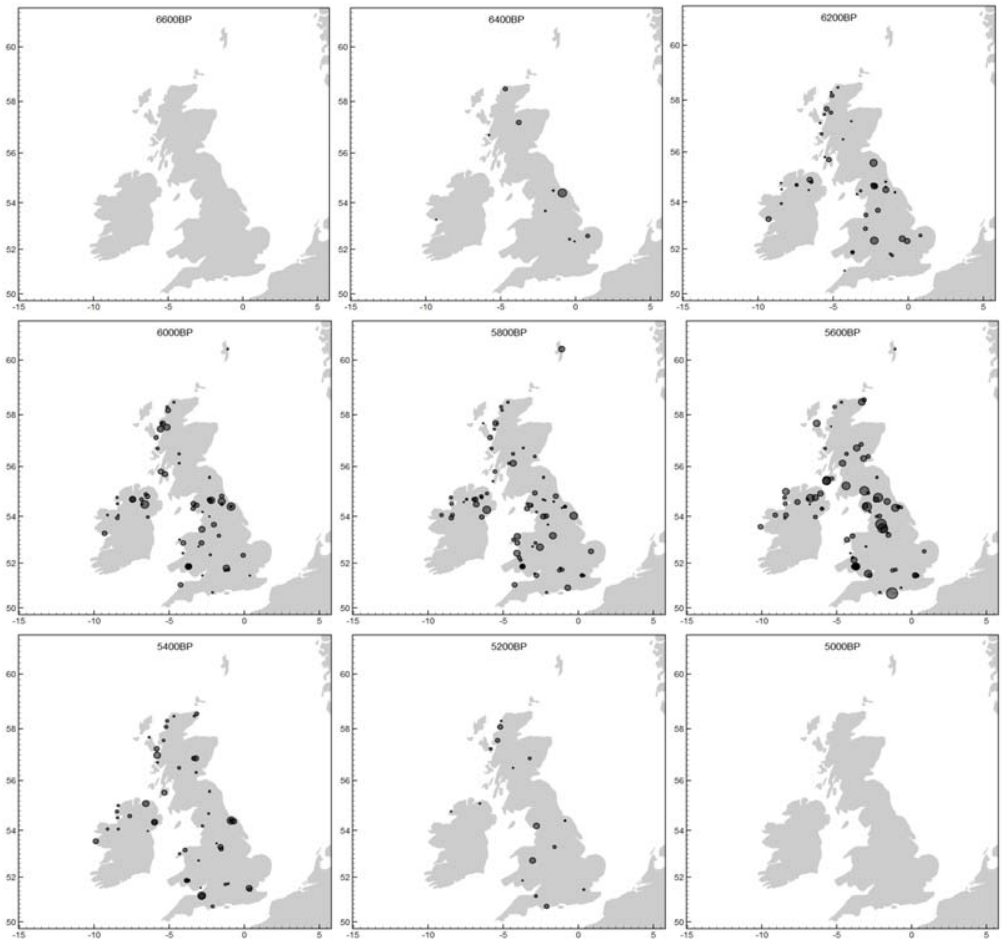


Figure 4 Maps showing the movement of elm decline across the British Isles at 200-yr intervals, modeled using a single trapezoidal phase (Lee 2012).

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

It should be stressed that this new mapping functionality is not a spatial model as such; it is simply a way to present the information from models in a spatial way. The modeling of individual sites needs to be done using existing tools within OxCal or by importing information from other packages. Within OxCal, it is likely to be most appropriate to apply the tags to phase boundaries or to `Sum` distributions from modeled phases rather than to individual ^{14}C dates. Although it is set up to look at information on maps, the plotting tool used is generic, and in principle could be used to look at spatial variation within a site or on a section.

CONCLUSIONS

It is a fundamental aspect of Bayesian analysis that the prior should reflect your state of expert knowledge before modeling. OxCal now has new ways to express these priors relating to different parameters in chronological models. Most significantly, the prior for local reservoir offset (or any other type of offset) no longer has to be expressed as a Normal distribution, and neither does that for mixtures of reservoirs. This should allow people to use more neutral priors and recover unbiased estimates for these important parameters.

No models are perfect, but sometimes the most widely used Bayesian models are actually too prescriptive to represent our true state of understanding. Models that allow for more gradual changes between phases should, in some instances, allow more realistic modeling of processes of non-instantaneous change. Likewise, the averaging over different depositional scenarios should give more robust results when our true state of knowledge about a sedimentary sequence is limited.

The spatial aspects of the OxCal package are a priority for development over the next few years. Another area, which is under active development, is the ability of the software to present data on different timescales, where there is some defined relationship. This should allow presentation of information, for example on the timescale of a particular record, where this has been related to the IntCal timescale itself in some way.

ACKNOWLEDGMENTS

The authors would like to acknowledge the important work of Angela Karlsberg, who, under the supervision of Caitlin Buck, first implemented the trapezoid model. The research and developments behind this paper were conducted in support of a NERC-funded project on Response of Humans to Abrupt Environmental Transitions (NE/E015670/1).

REFERENCES

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

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

Bronk Ramsey C. 2001. Development of the radiocarbon calibration program. *Radiocarbon* 43(2A):355–63.

Bronk Ramsey C. 2008. Deposition models for chronological records. *Quaternary Science Reviews* 27(1–2): 42–60.

Bronk Ramsey C. 2009a. Bayesian analysis of radiocarbon dates. *Radiocarbon* 51(1):337–60.

Bronk Ramsey C. 2009b. Dealing with outliers and offsets in radiocarbon dating. *Radiocarbon* 51(3):1023–45.

Bronk Ramsey C, Dee MW, Rowland JM, Higham TFG, Harris SA, Brock FA, Quiles A, Wild EM, Marcus ES, Shortland AJ. 2010. Radiocarbon-based chronology for Dynastic Egypt. *Science* 328(5985):1554–7.

Bronk Ramsey C, Staff RA, Bryant CL, Brock F, Kitagawa H, van der Plicht J, Schlolaut G, Marshall MH, Brauer A, Lamb HF, Payne RL, Tarasov PE, Haraguchi T, Gotanda K, Yonenobu H, Yokoyama Y, Tada R, Nakagawa T. 2012. A complete terrestrial radiocarbon record for 11.2 to 52.8 kyr B.P. *Science* 338(6105): 370–4.

Buck CE, Litton CD, Smith AFM. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19(5): 497–512.

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press

Higham T, Jacobi R, Basell L, Bronk Ramsey C, Chiotti L, Nespoulet R. 2011. Precision dating of the Palaeolithic: a new radiocarbon chronology for the Abri Pataud (France), a key Aurignacian sequence. *Journal of Human Evolution* 61(5):549–63.

Karlsberg AJ. 2006. Flexible Bayesian methods for archaeological dating [PhD thesis]. Sheffield: University of Sheffield.

Lee S. 2012. Bayesian methods for the construction of robust chronologies [PhD thesis]. Oxford: University of Oxford.

Lee S, Bronk Ramsey C. 2012. Development and application of the trapezoidal model for archaeological chronologies. *Radiocarbon* 54(1):107–22.

Lee S, Bronk Ramsey C, Hardiman M. 2013a. Modeling the age of the Cape Riva (Y-2) tephra. *Radiocarbon*, these proceedings, doi:10.2458/azu_js_rc.55.16214.

Lee S, Bronk Ramsey C, Mazar A. 2013b. Iron Age chronology in Israel—results from modeling with a trapezoidal Bayesian framework. *Radiocarbon*, these proceedings, doi:10.2458/azu_js_rc.55.16213.

https://doi.org/10.1017/S0033822200057878 Published online by Cambridge University Press