Radiocarbon, Vol 56, Nr 2, 2014, p 871–876 DOI: 10.2458/56.17468

SIMULATIONS AND OUTPUTS

Seren Griffiths

7 School Place, Oxford, United Kingdom. Email: griffithssg@cf.ac.uk.

ABSTRACT. Bayesian analysis is now routinely applied for the construction of site-specific stratigraphic chronological models. Other approaches have analyzed the chronology of phases of archaeological activity across regions. The available radiocarbon results—the nature of the samples and their associations—provide the basis for what chronological questions it is possible to address for any site or region. In dealing with regional analyses, due consideration must be made of data selection. While data selection might be a relatively self-evident consideration in the analysis of a site chronology, working with data from a larger region poses a number of specific data selection issues. Robust association between dated samples and a particular type of diagnostic material culture or site may provide one means of producing regional chronologies. However, if the activity under investigation includes a number of different cultural traits, which are related but with each having a slightly different chronological currency, defining the temporal end of data selection becomes more problematic. This article presents one approach, using a case study from the British Mesolithic-Neolithic transition, with 880 simulation OxCal models used to investigate the effect of variously defining the end of a regional archaeological phase. The results emphasize that for a regional case study, sensitivity analysis may provide a useful tool to ensure representative models; the study also highlights the importance of comparing multiple model posteriors.

INTRODUCTION

Bayesian modeling can provide estimates for the start or end of archaeological "phases" of activity (Buck et al. 1992, 1996). Two broad classes of archaeological phase can be analyzed (Bronk Ramsey 2008:265). The first is the chronology of single sites, where stratigraphic information about dated samples is used to relate samples to each other (e.g. Bayliss et al. 2007). The second type of phase is where radiocarbon dates are analyzed as representative of types of activity that took place across regions. The treatment of data on which these regional analyses are undertaken are "based on an interpretation, or a range of possible interpretations, of the regional chronology, and frequently make assumptions about synchronous changes that take place across a region" (Bronk Ramsey 2008:265). Regional studies may assume, for example, that phases of activity associated with a particular type of material culture—such as pottery or metal typologies (e.g. Needham et al. 1998; Garrow et al. 2010)—represent *related* activity, and the associated ^{14}C results are not therefore independent.

In complex Bayesian modeling that includes highly informative stratigraphic prior information (Bayliss et al. 2007; Bayliss 2009), it is essential that the prior information about the relationships between samples are accurate—or at least not significantly misleading. It is important to ensuring an appropriate distribution is applied to data, as it is possible to construct models that bias data in ways that are not intended (Steier and Rom 2000). In the use of less detailed prior information—such as for ^{14}C dates associated with types of material culture—results may be assumed to represent a sample of a uniform phase of activity over time (Buck et al. 1992; Bronk Ramsey 1995).

Application of an "uninformative" uniform model for the analysis of a group of results associated with an archaeological phase of activity processes associated statistical scatter, and is regarded as a relatively "neutral" assumption (Bronk Ramsey 2009:357), although it is possible to provide alternative estimates for the start and end of activity using more complex models (Karlsberg 2006; Bronk Ramsey 2009), which can be used to reflect archaeological interpretations of the available information. However, even using the relatively "neutral" assumption of a uniform phase of regional activity, consideration must be given as to which data are included in a model; different selections of data have the potential to produce different model outputs.

Proceedings of the Radiocarbon and Archaeology 7th International Symposium
Ghent, Belgium, April 2013 | Edited by Mark Van Strydonck, Philippe Crombé, and Guy De Mulder

The inclusion of data may be relatively simple if the definition of the phase of activity is based on the discrete presence of a type of material culture, for example, pottery type A. In this case, a model can estimate the start and end of the deposition of the pottery type in contexts that have been excavated archaeologically (there are of course a number of considerations in terms of how representative the sample of ^{14}C dates is of the duration of pottery use, and the association of the ^{14}C samples with the deposits from which pottery type A was recovered; Waterbolk 1971).

In cases where multiple types of material culture all contribute to the definition of an archaeological phase (for example, the "Neolithic package"), it may be necessary to impose an arbitrary chronological cut-off point in the data selection used for analysis. This was relevant in a recent research project that examined the chronology of the start of the British Neolithic (Griffiths 2011); in this case, aspects of Neolithic material culture (for example, plant and animal domesticates) continued after the period of interest.

If data selection is undertaken using the age of results and pre-existing chronological frameworks, it is possible that dating exercises will simply perpetuate existing or preconceived chronological frameworks or periodizations. In regional chronological analyses, the timing of phases of activity should be the object of archaeological investigation, not a basis for data selection. Employing such existing chronological frameworks as a basis for sample selection is in fact to adopt an inappropriate approach to data collection. When then does one cut off data selection, without using pre-existing chronological period divisions, and without selecting an arbitrary point that might bias model output?

As part of a recent research project on the chronology of the Mesolithic-Neolithic transition in Britain (Griffiths 2011), 880 simulation models were produced to investigate the effects of data selection on model output, and whether the shape of the calibration curve at different points in the analysis would result in issues with interpretation (Bronk Ramsey 1995:425).

SIMULATION EXPERIMENTS

Some 880 simulated models were produced in OxCal v 4.1 (Bronk Ramsey 2009) with different "arbitrary" end points for the study period of the Mesolithic-Neolithic transition in Britain (4300–3600 BC). Each model comprised a simple OxCal `Bounded` defined `Phase`, which estimates the start and end of the simulated phase of activity. This study was designed to investigate the effect of the duration of a phase of activity, the shape of the calibration curve on model output, and to address how often a unique model might produce inaccurate ranges from this part of the calibration curve.

Each model consisted of an assemblage of simulated ^{14}C measurements (`R_Simulate`), which were constrained within a uniform distribution `Phase` and defined by `Boundary` parameters. Ten models were produced for each decade in the 37th to 44th centuries BC. For example, for the start date 3710 BC, 10 models were simulated to end at "arbitrary" cut-off points in 3150, 3200, 3250, 3300, 3350, 3400, 3450, 3500, 3550, and 3600 BC. Each model contained 31 `R_Simulate` parameters with identical error terms (±30).

SIMULATION RESULTS

The differing highest posterior density (HPD) intervals for the "start" `Boundary` and the effect of shifting the end of the `Phase`, and the different effects of the calibration curve can be compared (Figures 1 and 2). Figures 1–2 show the ranges of the start `Boundary` parameters at 68% and 95% for the 10 simulation models for each decade in the 37th to 44th centuries BC. Additional analyses were undertaken to explore the effects of shifting the real start date of phase, but keeping the end date the same (Figure 3).

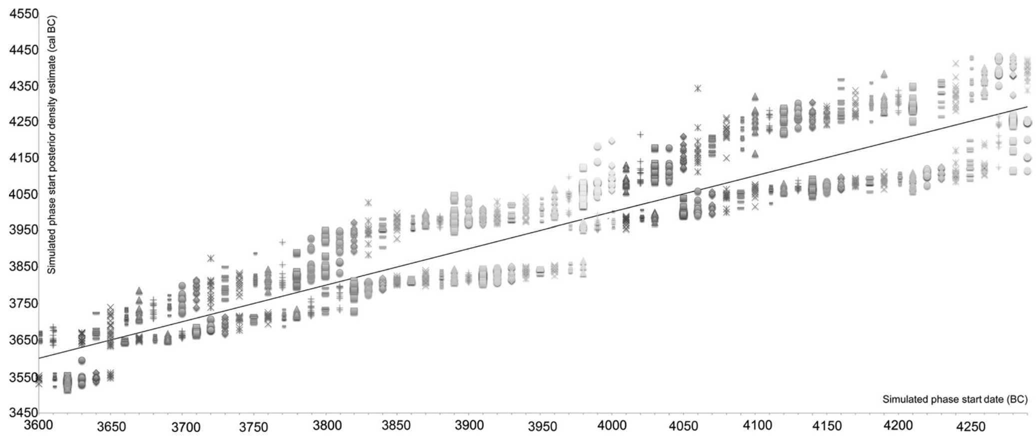

Figure 1 `Boundary` parameter ranges from simulations described in the text (95% probability). For each simulated start date, the end points of the ranges of each of the 10 analyses are indicated. The symbols have been arbitrarily attributed. The precision of the ranges derives in part from the shape of the calibration curve.

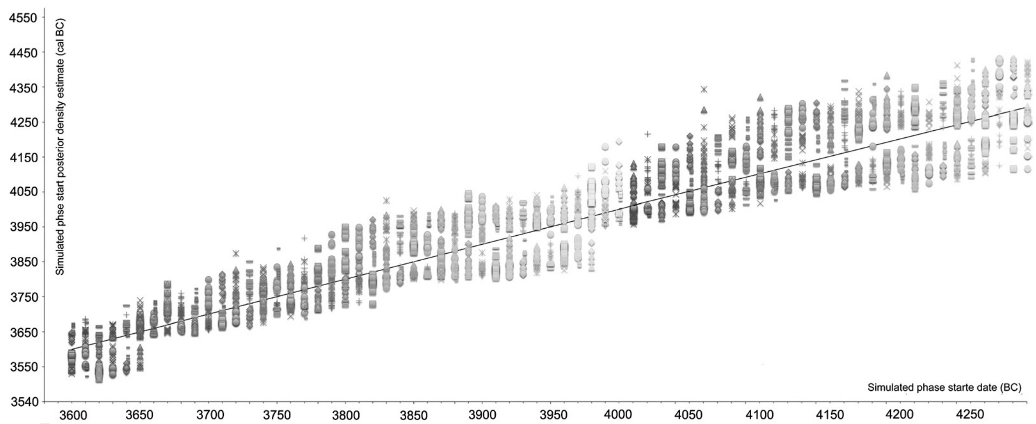

Figure 2 `Boundary` parameter ranges from simulations described in the text (68% probability). For each simulated start date, the end points of the ranges of each of the 10 analyses are indicated. The symbols have been arbitrarily attributed. The precision of the ranges derives in part from the shape of the calibration curve.

In terms of interpreting model output, the effects of shifting the data sample, and the effects of the relevant part of the calibration curve, a number of considerations should be made. Firstly, it is apparent that the choice of end date influences the precision of the posterior density estimate start ranges. There is no simple relationship between the duration of the phase and the precision of the posterior density estimates (Figure 3). The posterior distributions of the parameters appear rather to be informed by the specific part of the calibration curve on which they fall.

Secondly, the choice of end date influences the shape of the posterior distributions. Analysis of multiple simulations for specific periods can suggest possible issues for the interpretation of model output and for data selection. Some posterior density estimates for start `Boundary` parameters are well reproduced regardless of when the phase of Neolithic archaeological activity was curtailed. Some groups include individual distributions that include a limited spike of higher probability (against the background of other output), indicating the start of the simulated phase. All the posteriors for some start dates show relatively accurate spikes of higher probability independent of when the phase ended. Some groups of posterior density estimates show variability in their most elevated probabilities of when the phases began, depending on when the phase ended.

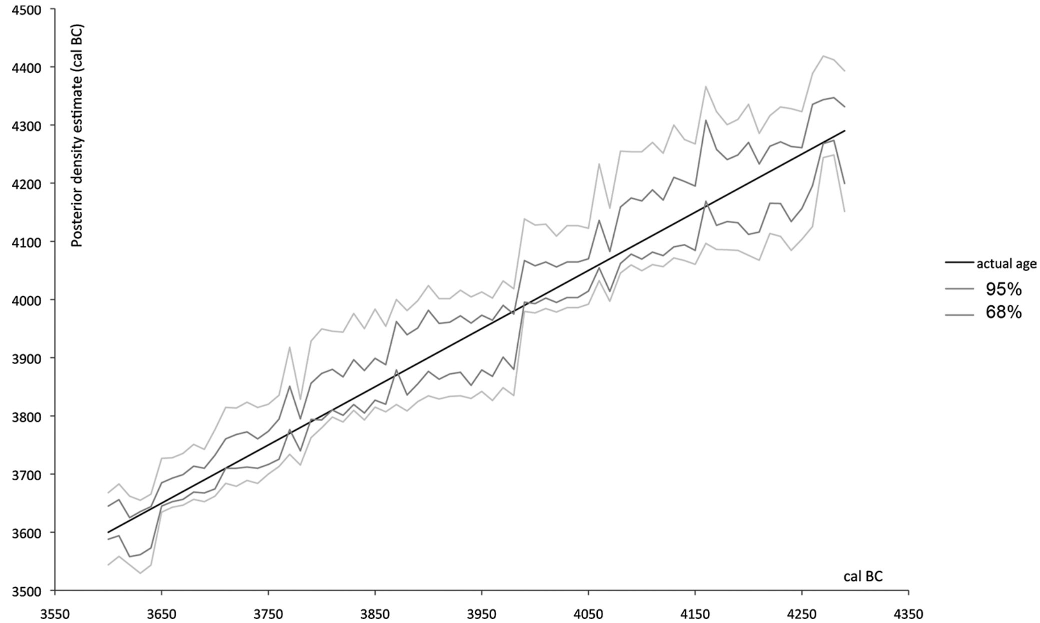

Figure 3 The various precisions produced for start `Boundary` estimates for a `Phase` ending in 3200 BC, which are achieved by changing the start date of the `Phase`. The plots in light gray represent the changes in precision on the 95% probability range. The plots in dark gray represent the changes in precision on the 68% probability range. The dark line represents the actual age of the simulated start date.

At 95% probability, 3.5% of the ranges produced did not include the simulated start date; at 68% probability, only 7.9% of the ranges did not include the simulated start date. In these analyses, the output HPD intervals are accurate more often than we might expect on statistical grounds. It should be emphasized that analyses of these types will depend on the sampled area of the calibration curve, the number of data, and so on. From these analyses, there do not appear to be specific start dates that consistently produced inaccurate outputs. While one-off models may therefore produce inaccurate ranges, multiple simulation models have the potential to highlight and identify these cases. This observation emphasizes the importance of undertaking repeated simulation modeling as part of the quality assessment of any real-life project output.

Occasionally, there is a tendency for bimodality that does not always reflect the simulated start point of the phase (Figure 4). There are occasions where individual ranges include elevated probability away from the simulated start of the phase (Figure 5).

These examples demonstrate the potential importance of sensitivity analysis—especially when modeling assemblages of ^{14}C data associated with regional chronologies or other currency chronologies beyond the level of site stratigraphic models. In archaeological case studies, it may be appropriate to produce multiple models, including models with simulated "dummy" data that artificially change the span of the chronological data sample, to investigate whether output is sensitive to the shape of the calibration curve or the data sample selection (e.g. Griffiths 2011: Chapter 9). If such simulations produce results that are all similar, the model output can be demonstrated to be insensitive to the data selection and treatment (Bronk Ramsey 2000:201).

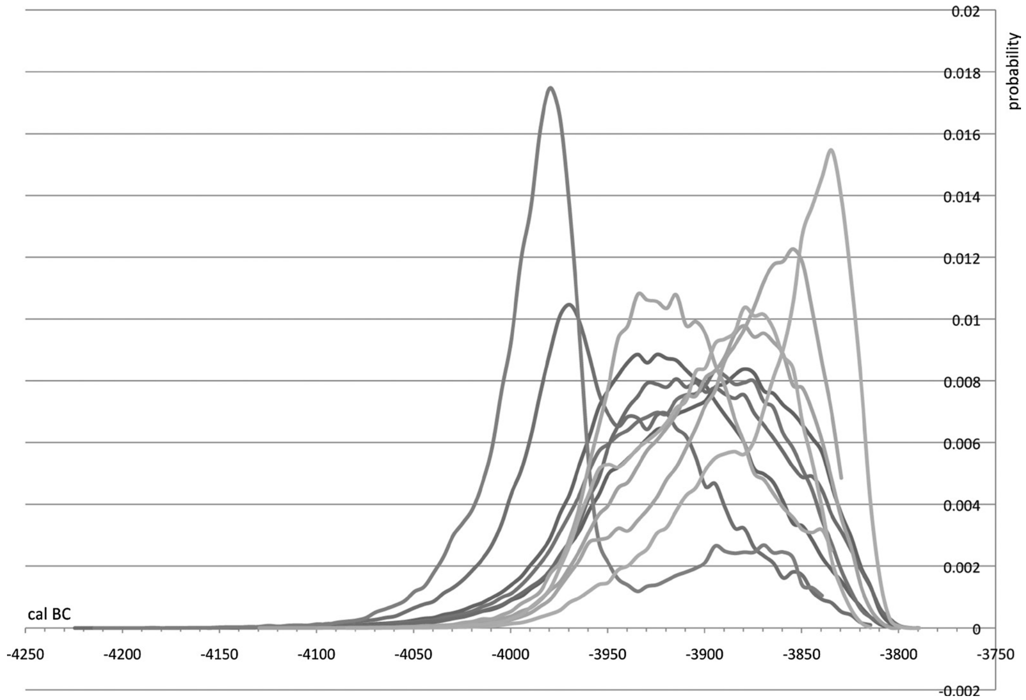

Figure 4 Example of the types of output distributions from the simulation analyses undertaken, in this example for `Phase` models beginning in 3890 BC. This example shows bimodality in simulation output regardless of when a `Phase` ends.

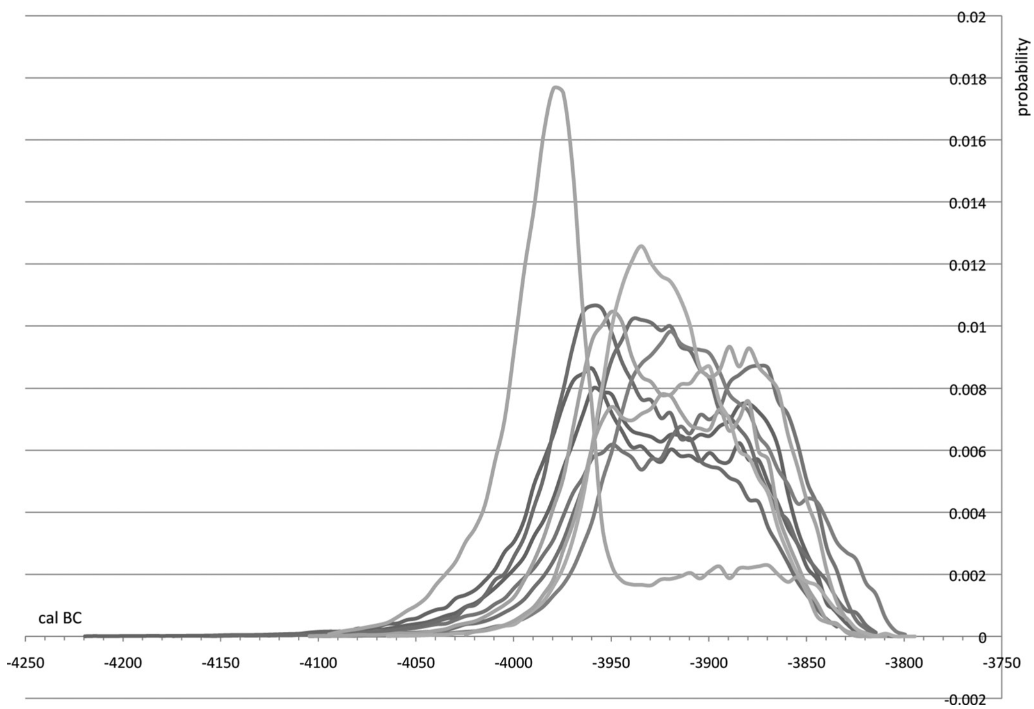

Figure 5 Example of the types of output distributions from the simulation analyses undertaken, in this example for `Phase` models beginning in 3940 BC. In this case, variability in simulation output is evident depending on the `Phase`. Simulation exercises may have value as sensitivity exercises for archaeological data.

CONCLUSIONS

Regional chronological analyses should not rely primarily on existing chronological periodizations, as such chronological frameworks may bias data sample selection, producing model output that simply reifies existing interpretations. In regional analyses, the timing of phases of activity should be the objects of archaeological investigation, not a basis for data selection. In cases where "uninformative" prior information is applied to groups of ^{14}C dates, and it is necessary to include an "arbitrary" cut-off point in data selection, it may be helpful to undertake sensitivity analysis to explore the impact of the calibration curve and the data sample selection.

Simulation sensitivity analyses may be useful in ensuring a robust and representative solution. The sensitivity analysis presented here emphasizes the importance of presenting and engaging with their HPD intervals, as well as quoting posterior ranges, when considering analysis outputs. In archaeological interpretations of the start or end of regional phases of activity, the effects of data selection or data cleansing and the shape of the calibration curve should always be considered in further analyses.

Selecting an arbitrary end point for simulated phases sometimes gives an inaccurate range for the simulated start age of the phase. However, the number of inaccurate HPD intervals were fewer than might be expected statistically. Employing different arbitrary end points for a phase of activity (as part of a project quality assurance) may provide a means of identifying specific model choices or periods that are problematic.

ACKNOWLEDGMENTS

Andrew Millard and Alex Bayliss provided invaluable comments on an earlier version of this paper. Two anonymous reviews made constructive comments. The analysis presented here was undertaken as part of an Arts and Humanities Research Council and English Heritage–funded PhD project. Alex Bayliss and Alasdair Whittle supervised the project with kindness and patience.

REFERENCES

Bayliss A. 2009. Rolling out the revolution: using radiocarbon dating in archaeology. *Radiocarbon* 51(1):123–47.

Bayliss A, Bronk Ramsey C, van der Plicht J, Whittle A. 2007. Bradshaw and Bayes: towards a timetable for the Neolithic. *Cambridge Archaeological Journal* 17(S1):1–28.

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

Bronk Ramsey C. 2000. Comment on 'The use of Bayesian statistics for ^{14}C dates of chronologically ordered samples: a critical analysis.' *Radiocarbon* 42(2):199–202.

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

Bronk Ramsey C. 2008. Radiocarbon dating: revolutions in understanding. *Archaeometry* 50(2):249–75.

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

Buck C, Cavanagh W, Litton C. 1996. *Bayesian Approach to Interpreting Archaeological Data*. Chichester: John Wiley & Sons.

Buck C, Litton CD, Smith AFM. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19(5):487–512.

Garrow D, Gosden C, Hill J, Bronk Ramsey C. 2010. Dating Celtic art: a major radiocarbon dating programme of Iron Age and Early Roman metalwork in Britain. *Archaeological Journal* 166:79–123.

Griffiths S. 2011. Chronological modelling of the Mesolithic-Neolithic transition in the North and Midlands of England and Wales [PhD thesis]. Cardiff University.

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

Needham S, Bronk Ramsey C, Coombs D, Cartwright C, Pettitt P. 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. 2000. The use of Bayesian statistics for ^{14}C dates of chronological ordered samples: a critical analysis. *Radiocarbon* 42(2):183–98.

Waterbolk H. 1971. Working with radiocarbon dates. *Proceedings of the Prehistoric Society* 37:15–33.