Flexible Bayesian Methods for Archaeological Dating

Angela Jane Karlsberg

Thesis submitted to the University of Sheffield
for the degree of Doctor of Philosophy

Department of Probability and Statistics
School of Mathematics and Statistics
University of Sheffield
Sheffield, U.K.

July 2006

Acknowledgements

There are lots of people that I would like to thank for their support and encouragement and without whom completion of this thesis would not have been possible.

Firstly I would like to acknowledge the enthusiastic supervision of both Dr. Caitlin Buck and Prof. Paul Blackwell who provided advice, criticism and friendly encouragement whenever I needed it throughout the project.

Secondly, I would like to thank both EPSRC and English Heritage for providing resources and funding for my research. I owe particular gratitude to Alex Bayliss of English Heritage for her archaeological supervision and willingness for me to get involved in some practical archaeology.

I would also like to thank my office mate Lynsey McColl (for all those serious discussions and all those lunches) as well as all the other PhD students, the rest of the academic staff, research assistants and secretaries that made my time so enjoyable at Sheffield.

Finally I would like to thank my husband, Simon, and my parents, who have been great over the years and never raised an eyebrow when I claimed that my thesis would be "finished in the next month" for nearly a year...

Abstract

Statistical models for the calibration of both independent and related groups of radiocarbon determinations are now well established and there exists a number of software packages such as BCal, OxCal and CALIB that can perform the necessary calculations to implement them. When devising new statistical models it is important to understand the motivations and needs of the archaeologists. When researchers select samples for radiocarbon dating, they are often not interested in when a specific plant or animal died. Instead, they want to use the radiocarbon evidence to help them to learn about the dates of other events, which cannot be dated directly but which are of greater historical or archaeological significance (*e.g.* the founding of a site).

Our initial research focuses on formulating prior distributions that reliably represent *a priori* information relating to the rate of deposition of dateable material within an archaeological time period or phase. In archaeology, a phase is defined to be a collection of excavated material (context or layers) bounded early and late by events that are of archaeological importance. Current software for estimating boundary dates only allows for one possible type of *a priori* distribution, which assumes that material suitable for dating was deposited at a uniform rate between the start and end points of the phase. Although this model has been useful for many real problems, researchers have become increasingly aware of its limitations. We therefore propose a family of alternative prior models (with properties tailored to particular problems within archaeological research) which includes the uniform as a special case and allows for more realistic and robust modelling of the deposition process. We illustrate, via two case studies, the difference in archaeological conclusions drawn from the data when implementing both uniform and

non-uniform prior deposition models.

The second area of research, that we take the first steps towards tackling, is spatio-temporal modelling of archaeological calibration problems. This area of research is of particular interest to those studying the response of plants and animals, including humans, to climate change. In archaeological problems our temporal information typically arises from radiocarbon dating, which leads to estimated rather than exactly known calendar dates. Many of these problems have some form of spatial structure yet it is very rare that the spatial structure is formally accounted for. The combination of temporal uncertainty and spatial structure means that we cannot use standard models to tackle archaeological problems of this kind. Alongside this, our knowledge of past landscapes is generally very poor as they were often very different from modern ones; this limits the amount of spatial detail that can be included in the modelling.

In this thesis we aim to make reliable inferences in spatio-temporal problems by carefully devising a model that takes account of the temporal uncertainty as well as incorporating spatial structure, to provide probabilistic solutions to the questions posed. We illustrate the properties of both the conventional models and the spatio-temporal models using a case study relating to the radiocarbon evidence for the Late glacial reoccupation of NW Europe.

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The problem of interest throughout this thesis is to develop a more coherent and robust framework when modelling radiocarbon calibration problems. This is achieved by extending existing models to allow for a wider range of *a priori* information to be incorporated.

The content of this thesis falls broadly into two areas of research, these being

- modelling the deposition of datable material within an archaeological phase[1], and
- Spatio-temporal modelling of radiocarbon calibration problems.

The motivation for this thesis was driven by English Heritage, the industrial sponsor, which has commissioned radiocarbon dates on archaeological material from over 150 sites in the last eight years. The English Heritage Scientific Dating Coordinator, Alex Bayliss, routinely uses Bayesian methods to provide the core of the interpretative process. As one of the earliest routine users of the Bayesian chronology building framework, Bayliss has become increasingly aware of the limitations of the methodology that is currently implemented in software packages such as OxCal and BCal.

[1] A phase is defined to be a collection of excavated material (context or layers) bounded early and late by events that are of archaeological importance.

These packages perform the necessary calculations for the calibration of both independent and related groups of radiocarbon determinations. However, there are limitations to the type of *a priori* information that can be incorporated. These techniques only allow for one possible type of *a priori* distribution when modelling the relationship between observed radiocarbon determinations and successive start and end dates of phases of activity. The convention is to assume that the material suitable for radiocarbon dating was uniformly distributed between the start and end dates of the phase of activity, and that the start and end dates are themselves unknown. It was initially assumed that such models constituted suitably vague priors and they may be the easiest way to represent prior ignorance, when little is known *a priori* about the deposition rate.

Applied researchers, such as those at English Heritage, are increasingly concerned about how the assumption of uniform deposition rates will affect the inferences they make and are keen to explore alternatives allowing a more realistic and robust modelling of the deposition processes. It is felt that alternative models for the deposition of datable material could make an enormous impact on their day-to-day work.

In addition, there have been increasing numbers of case studies in the applied literature (*e.g.* Housley *et al.*, 1997 and Van Strydonck *et al.*, 2004) in which the authors believe that the rate of deposition/manufacture is not uniform over the proposed range. Suggestions have been made in Naylor and Smith (1988), Nicholls and Jones (2001) and Blackwell and Buck (2003) to look at alternatives to the conventional uniform model, but to date only one alternative has been implemented, see Section 2.6.

It is therefore our aim to seek alternatives to the conventional uniform model and thus build a more flexible range of *a priori* distributions that reliably represent information arising from archaeological research.

The second area of research covered in this thesis is that of spatio-temporal modelling. There are an increasing number of archaeological calibration problems that are concerned with studying the colonisation or recolonisation of past landscapes. Problems of this kind typically consist of multiple phases which are currently tackled by making use of the

existing temporal tools, *i.e.* assuming each phase is independent of the others. Although this allows us to calculate probabilistic answers to chronological questions of interest, *e.g.* to determine the order in which the phases were colonised, there are no formal methods for tackling such problems within a fully spatio-temporal framework.

We therefore propose a model which builds upon the existing models in Chapter 2, which takes account of both the spatial and temporal information that arises from archaeological excavations. This allows us to tackle problems within a fully spatio-temporal framework and hence combine data from related phases. There are also a number of other archaeological applications such as changes in culture, technologies *etc.* that all spread spatially and consequently could be tackled using the approach we suggested for incorporating spatial structure.

The aim of the material that follows in this chapter is to present both the motivation and the background needed to be able to fully understand the problems at hand, as well as presenting an outline of the chapters to come.

1.1 Background

This section provides a background to the project and, in particular, discuss the physical basis of radiocarbon dating and the cause of complications that arise in its use. The technical content of what follows has been obtained from two sources, Bowman (1990) and Aitken (1990).

1.1.1 Radiocarbon dating

The radiocarbon dating method, developed by a team of scientists led by Libby (who in 1960 was awarded the Nobel Prize in chemistry for his pioneering work in the development of the method), is now the most commonly used 'absolute dating

technique'[2].

The method has been a great contribution to the development of archaeology. Archaeologists grasped the importance of the technique as it provides a means to test the accuracy of the 'relative dating methods'[3]. By the mid 1950s a number of laboratories in Europe and the USA were producing radiocarbon measurements; today there are over 130 radiocarbon dating laboratories around the world. The radiocarbon dating technique has been, and continues to be, used in a number of different applications such as archaeology, geology, climatology and oceanography.

Basic principles

Radiocarbon dating is based on the carbon cycle and the radioactive properties of the isotope ^{14}C. Carbon occurs naturally in the form of three isotopes, carbon-12, carbon-13 and carbon-14 (denoted as ^{12}C, ^{13}C and ^{14}C respectively), which are all chemically identical; but differ as atoms of different isotopes have different numbers of neutrons in their nuclei.

Modern carbon consists of approximately 99% ^{12}C, 1% ^{13}C, but only about one part per million million of ^{14}C. ^{14}C is the only unstable and, therefore, radioactive carbon isotope. Radiocarbon (^{14}C) is continually being formed in the upper atmosphere due to the interaction of cosmic-ray neutrons with Nitrogen-14 (^{14}N). After formation, the ^{14}C atoms quickly combine with oxygen to form 'heavy' carbon dioxide (which is chemically indistinguishable from carbon dioxide containing either of the other carbon isotopes). This mixes with the ordinary carbon dioxide in the atmosphere and then via the photosynthesis process and the food chain, enters all plants and animal life. The carbon dioxide also enters the oceans as dissolved carbonate, so this too contains ^{14}C and consequently so do any shells and deposits formed from it. This collection of atmosphere,

[2]Definition from http://www.staff.ncl.ac.uk/kevin.greene/wintro/keyword.htm: Absolute dating: dates determined by methods whose accuracy is based on radioactive decay or regular natural phenomena such as tree rings.

[3]Definition from http://www.staff.ncl.ac.uk/kevin.greene/wintro/keyword.htm: Relative dating: relative ages cannot be used on their own but must be related to an absolute technique such as radiocarbon dating. Sequences of contexts established by the stratification of archaeological sites, or artefacts arranged into order by typology, are relative.

biosphere and oceans are commonly referred to as the carbon exchange reservoir.

Plants and animals, during their lifetime, constantly exchange carbon with the reservoir, so that the concentration ratio between ^{14}C and the non-radioactive isotopes is constantly maintained. Upon death, organisms cease to participate in carbon exchange with the atmosphere and there is loss of ^{14}C atoms by radioactive decay. The rate at which the atoms decay is determined by the law of radioactive decay, which is dependent upon the decay rate and the remaining proportion of ^{14}C in the sample. Each isotope has a specific decay rate; for ^{14}C this rate is 1% per 83 years, which is equivalent to a half-life[4] of 5730 years. The half-life of a radioactive isotope describes how long it takes for half of the atoms in a given mass to decay.

The law of radioactive decay is given by the following equation:

$$A = A_0\, e^{-\lambda t} \tag{1.1}$$

where A is the amount of remaining radioactive material (^{14}C) after time t, A_0 is the initial amount of radioactive material at time 0 and λ is the decay rate. The decay rate represents the amount of time it takes for the radioactive material to disintegrate and is related to the half-life, $t_{(1/2)}$, by

$$\lambda = \frac{\ln 2}{t_{(1/2)}}.$$

Another value of importance in the law of radioactive decay is the mean life, τ, which is simply the reciprocal of the decay rate.

Provided that we can estimate A relative to a standard A_0, the time elapsed (t) since the material died can be estimated by rearranging Equation 1.1 to give

$$t = \tau\, \ln\,(A/A_0). \tag{1.2}$$

[4]At the time that radiocarbon dating was been developed, Libby estimated the half-life of radiocarbon to be 5568 years, this value is known as the conventional or Libby's half-life. In later years the half-life of radiocarbon was revised by three independent laboratories and it was found that a more accurate value of the half-life was 5730 years. However, it is convention is to use the Libby's half-life in the calculation of radiocarbon results to avoid confusion.

So given, Equation 1.2, what types of material can be dated using radiocarbon? Basically materials which are composed of carbon and are, hence, organic. In the British Isles the most commonly preserved sample types are bone, shell and charcoal. However wood, peat, soil, pollen, textiles and fabrics are examples of the other types of materials that are commonly radiocarbon dated (See Bowman, 1990, pages 12-13 for further details).

1.1.2 Assumptions of the radiocarbon dating method

From above, it is seen that Equation 1.2 is the basis for the radiocarbon dating method and in particular, that the method is only useful if the two quantities A and A_0 are known or measurable. There are numerous assumptions necessary for the technique to work, which are summarized as follows.

- **Assumption 1**: the concentration of radiocarbon in each carbon reservoir (atmosphere, biosphere and oceans) has remained constant over time.
- **Assumption 2**: there has been rapid and complete mixing of ^{14}C within each carbon reservoir.
- **Assumption 3**: the half life of ^{14}C is accurately known.
- **Assumption 4**: after decay of an organism, the ^{14}C concentration in relation to ^{12}C and ^{13}C has not altered except by radioactive decay.

From the beginning of the development of radiocarbon dating these assumptions were believed to be correct given the techniques then available to check them. For example, James Arnold and William Libby published a 'curve of knowns' using the known age of samples ranging from approximately 900 to 4900 years ago. Using the best techniques then available, they were able to concluded that there was a good agreement between the theoretical and measured ^{14}C activities versus known age.

It was not until the late 1950s, when the technique had developed further, that discrepancies far from insignificant began to emerge. In some cases radiocarbon results

were found to be several centuries too young. Thus highlighting a problem with the technique, and the most likely explanation was the violation of one or all of the four assumptions above. The following sections discuss a number of issues that are now known to contribute to the violation of these assumptions.

Atmospheric ^{14}C variations

The most serious problem concerns Assumption 1, that the concentration of the radioactive isotope ^{14}C has remained constant over time. To help assess the problems concerning this assumption, timber samples already dated using dendrochronology[5] were radiocarbon dated. In the 1960s, a continuous tree-ring sequence stretching back approximately 8000 years was established and Hans Suess published the first calibration curve (Suess, 1970) using this data. This curve was referred to as Suess's curve and helped verify the discrepancies between the radiocarbon and calendar years and as a result confirmed that the ^{14}C equilibrium levels fluctuate slightly from year to year.

It became apparent from Suess's curve that there were two trends in atmospheric ^{14}C levels. The first trend is long-term and has been described as a sine wave with a period of approximately 9000 years. The second feature noticeable is superimposed onto the sine wave and takes the form of 'wiggles'. These wiggles, although only of a few decades on the calendar scale can have an amplitude of a several centuries on the radiocarbon axis.

A brief description of the geophysical causes of the fluctuations in the natural production rate of ^{14}C will be given (for further details see Bowman, 1990, pages 18-20), along with the effect of human activity on the atmospheric levels, such as the burning of fossil fuels.

The long term variation in atmospheric ^{14}C levels is seen to correlate well with fluctuations in the strength of the Earth's magnetic field. Cosmic rays are charged

[5]Dendrochronology is the dating of past events through the study of tree ring growth. Trees grow by the addition of an annual ring, yet the width of the rings vary from year to year depending upon climatic conditions. As a result patterns of greater and lesser growth from the same species of trees can be compared with the aim of creating tree ring chronologies spanning back several millennia.

particles which are deflected by the Earth's magnetic field. If the strength of the magnetic field becomes weaker, less cosmic rays will be deflected away from the Earth and production of ^{14}C will rise, and *vice versa*.

Short term variations in the atmospheric ^{14}C levels can be caused by 'sunspot activity'[6]. During periods of high sunspot activity the magnetic field increases, resulting in a higher number of cosmic rays being deflected and hence ^{14}C production decreasing. Records on sunspot activity over the past few centuries have shown that there tend to be cycles of two lengths. There is a cycle of a period of about 200 years which is superimposed by a cycle of 11 years. The effect of the 11 year cycle is unlikely to cause more than about 20 years variation in age. However the effect of the 200 year cycle is much more significant. The wiggles associated with this cycle represent changes in the radiocarbon age of a century or two, however, the corresponding calendar age changes by only a few decades. It is believed that it is these wiggles that cause the need for calibration of radiocarbon years to calendar years.

As well as the natural variations in the ^{14}C equilibrium levels in the atmosphere, humans have also had an effect on the global level of ^{14}C through the burning of fossil fuels and the effect of atomic bombs. The burning of fossil fuels[7] (such as coal, oil and natural gases) started in the last century and when burnt in large quantities the 'old' carbon released dilutes the ^{14}C concentration relative to ^{12}C and ^{13}C in the atmosphere. This results in a change in both size and isotopic composition of the atmospheric carbon reservoir, which in turn results in a lower than expected ^{14}C content in relatively modern samples. A more dramatic effect in the variation of the ^{14}C equilibrium level, arises from nuclear-weapon testing. It was seen, as a result of the testing carried out in the 1950s and 1960s, that the ^{14}C content measured in the atmosphere had approximately doubled in 1965 in comparison to the theoretical 1950 level. Atmospheric testing was quickly banned, and due to the gradual mixing of ^{14}C through the carbon exchange, by the 1990s the ^{14}C

[6]A sunspot is a region on the Sun's surface that is marked by a lower temperature than its surroundings, and intense magnetic activity.

[7]Naturally occurring fuels formed over millions of years from organic material (hence their ^{14}C has long since decayed).

levels had decreased to about 20% higher than the theoretical 1950 level.

Alteration and contamination effects

One assumption, Assumption 4, of radiocarbon dating method is that the ratio of ^{14}C to ^{12}C and ^{13}C has only altered through radioactive decay. However, there are two processes which may alter the ^{14}C content in the organism, these are referred to as alteration and contamination.

Alteration results in the ^{14}C content of a sample being different to that in the atmosphere or to the value expected from radioactive decay alone. There are a number of processes that cause this effect, these involve isotopic fractionation, recrystallization of shell carbonate and *in situ* production. In all cases the ^{14}C content is modified without the addition of extra sources of carbon.

The most important of these processes is isotopic fractionation. This involves a change in the ratios of the different isotopes of carbon in the samples, through processes such as photosynthesis. In any organisms, there is a tendency for lighter isotopes to be taken-up in preference to heavier isotopes. Therefore, growing plants and animals may have a lower ^{14}C level than that of the atmosphere in which they metabolize. There may well be small variations from species to species and therefore it is now common practise to evaluate the effect of alteration in each sample to be dated.

Contamination occurs when the ^{14}C content of a sample is altered through the addition of material containing carbon that has a different ^{14}C content. There are a number of ways in which this occurs, for example, calcium carbonate (*e.g.* limestone) dissolved in the ground water can be transferred into buried samples, thus greatly increasing their apparent age. Contamination can also occur from humic acids (partially decomposed organic material), these can either increase or decrease the apparent age of samples depending upon their origin. Due to the possibility of contamination it is important that all samples selected for radiocarbon dating are firstly pretreated. Such pretreatment

removes any additional sources of carbon which may have contaminated the sample so that the ^{14}C content will only reflect that of the original sample material and no other source (for further details see Bowman, 1990, Chapter 2).

Mixing rates of ^{14}C

Northern-to-Southern hemisphere effect

The ^{14}C mixing rate in most terrestrial carbon reservoirs is thought to be sufficiently rapid for dating purposes (Assumption 2). However, while there is good mixing within the Southern and Northern hemispheres, mixing between them is poor. This results from them having separate atmospheric circulation systems, hence their prevailing winds blow in opposite directions along the equator. As a result, the ^{14}C concentration in the Southern hemisphere is, on average, below that of the Northern hemisphere. Consequently samples from the Southern hemisphere are approximately 55-58 calendar years older than those in the Northern hemisphere, with uncertainties increasing from ± 7.9 at 1000 cal BP to ± 25 at 11000 cal BP (McCormac *et al.*, 2004). It is believed that the cause of this difference is due to the fact that the Southern hemisphere has a greater ocean surface area.

Marine mixing effects

Mixing rates in the deep oceans are known to be slow in comparison to those in the atmosphere and biosphere. The mixing of ^{14}C in the oceans is known to be complicated by phenomena such as upwelling (the upward movement of the deep waters; which is latitude dependent). In areas where upwelling occurs, material from the surface water measures on average about 400 radiocarbon years too 'old' compared to those of terrestrial samples and we refer to this as the marine effect. There are generalized measurements for the marine effect in broad oceanographic regions, however local effects can vary over relatively short distances and these can outweigh the regional effects. As a result, the marine reservoir effect is an additional source of uncertainty when dating

samples of marine origin (for further details see Bowman, 1990, pages 24–25).

1.1.3 Measurement of radiocarbon

In the previous section the basic principles of radiocarbon dating were discussed and the need for calibration was established. This section moves on to discuss briefly the two laboratory methods used to detect ^{14}C.

There are currently two methods that are routinely used to estimate radiocarbon ages of organic samples, these are known as the conventional radiocarbon dating method (or indirect method) and the AMS (accelerator mass spectrometry) method. A brief overview of the two methods will be given but more details of the techniques can be found in Chapter 3 of Bowman, 1990.

The conventional method relies on the detection of beta-particles[8] emitted (these are fairly easily detected as they are electrically charged) when ^{14}C decays to ^{14}N. The amount of beta-particles detected will reflect the amount of remaining ^{14}C in the sample and therefore the amount of decay that has occurred. Since the development of radiocarbon dating the counting of beta-particles has formed the basis of the technique. The two main methods used to count the beta-particles are gas proportional counting and liquid scintillation counting, for details see Bowman, 1990, pages 31-32.

AMS is a much more recent technique and became commercial in the 1980s. The basic principle of AMS is to separate the specific elements by their atomic weights through mass spectrometry. This enables a direct measurement of the proportion of ^{14}C atoms relative to ^{12}C and ^{13}C in the sample. One advantage of AMS is that the size of the samples required for dating are much smaller than those required by the conventional dating technique. This enables not only the dating of samples such as individual seeds but also the dating of valuable artefacts with minimum destruction. Although the conventional method and AMS are based on different principles it is assumed that the radiocarbon

[8]A beta particle is the name given to an electron resulting from the radioactive decay of a nucleus (Bowman, 1990, page 31).

results can be interpreted in the same way (Bowman, 1990, page 31).

When selecting samples for radiocarbon dating archaeologists take considerable care and thought in maximizing the information that may be returned from a sample. It is crucial to demonstrate that there is a meaningful relationship between the sample dated and the archaeological event of interest, as the radiocarbon technique is expensive and (in the case of the conventional method) is a lengthy procedure.

The resulting product from the radiocarbon dating technique is what we refer to as a 'radiocarbon determination'. This consists of an estimated 'radiocarbon age' x and a standard error σ, reflecting the uncertainty in the dating process. Throughout the thesis a radiocarbon determination will be written in the form $x \pm \sigma$. All ages will be expressed, as is conventional in the radiocarbon community, in 'years before present' (BP). For the purpose of radiocarbon dating, present is taken as 1950AD.

In any experimental process there is always inherent experimental error. Usually experimental error is evaluated through the replication of the measurement process. However, in radiocarbon dating, this is not feasible for a number of reasons, mainly the time, cost and (in the case of the conventional method) the size of the sample needed to be able to produce a radiocarbon determination. Therefore the convention is to estimate the error term, σ, and then treat it as if it were known.

Currently there is no convention regarding how a laboratory should evaluate their total error. However, all laboratories do include a contribution to the error term from counting the number of decaying atoms in a period of time. Although there are other errors that occur in the radiocarbon dating process (see Section 2.5.3) the laboratories feel that this is the only error that they can accurately measure.

One obvious concern to the buyers of radiocarbon determinations is: if they sent the same sample to two different radiocarbon laboratories, would there be variability between the two laboratories and how great would this variability be? Since the 1980s Marion Scott has been involved in the design and analysis of Inter-lab comparisons, with the

primary goal of investigating the comparability of results produced under quite different laboratory protocols. The details of the latest of these studies, the Fourth International Radiocarbon Intercomparison (FIRI), can be found in the specialized issue of *Radiocarbon* 45(2).

1.1.4 Modelling the relationship between calendar and radiocarbon years

Following the early work of Hans Suess, who published the first calibration curve (Suess, 1970) to help verify the discrepancies between radiocarbon and calendar years, the need to calibrate was recognized worldwide by the radiocarbon community. This resulted in internationally agreed high-precision calibration data sets with the first being published in 1986 (Stuiver and Pearson 1986, Pearson and Stuiver 1986, Pearson *et al.* 1986). These were derived by radiocarbon dating timber samples that had already been dated on the calendar time-scale using dendrochronology.

As calibration became routine in the radiocarbon community, improving the calibration process became an important issue. As a result, the journal *Radiocarbon* has published 4 special issues on calibration. IntCal98 (Stuiver *et al.*, 1998) is the name given to the internationally agreed curve published in the third of these special issues. It updates and extends two previous estimates of the curve (28(2B), 1986 and 35(1), 1993). For the first two years of my PhD, IntCal98 was the most up to date version of the calibration curve. However, in my final year IntCal98 was again updated and IntCal04 (Reimer *et al.*, 2004) was ratified and published. As a result the following material will discuss both IntCal98 and IntCal04, as well as the main differences between the two.

The primary aim of the IntCal working group when constructing IntCal98 was to collect high quality data. However, when it came to the curve construction, relatively simple data averaging methods were used. The calibration curve (generated at intervals of 10yr for the range 0-15585 cal BP and 1000yr for 16000-24000 cal BP) was constructed by taking a weighted average of all the data within a 10yr window and assigning the midpoint

of the decade as the calendar age. In some cases the available ^{14}C measurements were on blocks of timber whose rings covered twenty rather than ten years. In these cases the data were handled as if been two separate decadal measurements. As a result of the methods used to construct the curve it was realized that important uncertainties had been ignored, for example the uncertainty in the calendar age of the varved sediments[9] and U/Th-dated corals.

IntCal04 was constructed and ratified at the 18th International Radiocarbon Conference held in New Zealand and extends the possible time period of calibration by an additional 2000 years (0 to 26000 cal BP). Although it does not greatly extend the time period of calibration the curve is estimated with a much higher resolution beyond 11400 cal BP than IntCal98. Dendrochronologically-dated tree-ring samples now cover the period 0-12400 cal BP and marine data, with site specific marine reservoir corrections, cover the period 12400-26000 cal BP. Where as, in the past, one of the primary aims of the IntCal working group was to collect high quality data, it has now been acknowledged that the methods used to construct the curve from the raw data are of equal importance. As a result, IntCal04 has been constructed using a coherent statistical method (Buck and Blackwell, 2004) which takes into account the uncertainties in both the calendar age and the ^{14}C age.

The following will outline the underlying model for the construction of IntCal04, which takes the form of a random walk. However, only the simplest case (a single radiocarbon determination) will be discussed here, for further details see Buck and Blackwell (2004). Given a single ^{14}C determination, X, with the known calendar date θ it is usually assumed that X is given by the true value of the calibration curve at date θ, written as $\mu(\theta)$, plus an error term, ϵ. Thus, X can be represented as $X = \mu(\theta) + \epsilon$, where

$$\epsilon \sim N(0, \sigma^2). \tag{1.3}$$

[9]Definition taken from http://www.thefreedictionary.com: A layer or series of sediment deposited in a body of still water in one year.

If interest lies in learning about the curve at a particular point, θ^*, then it is believed that for any θ near θ^*, knowing $\mu(\theta)$ would help learn about $\mu(\theta^*)$. Prior beliefs about the relationships between different points of the curve can be expressed in terms of a random walk. Hence, prior beliefs about the changes in the calibration curve from one year to the next may be represented by a normal distribution with a mean of β and a variance (per year) of r^2, which gives the following:

$$\mu(\theta+1) = \mu(\theta) + Z_{\theta+1}, \tag{1.4}$$

where

$$Z_{\theta+1} \sim N(\beta, r^2) \tag{1.5}$$

or

$$\mu(\theta+1) \sim N(\mu(\theta) + \beta, r^2). \tag{1.6}$$

Buck and Blackwell (2004), assumed that a natural choice for the value of β would be 1, as it seems reasonable to assume that the calibration curve would change by approximately 1 radiocarbon year per calendar year. However, a sensible value for the parameter r, *a priori*, was not so clear. Sensitivity tests were carried out, initially using single year tree-ring data for the model and then using decadal measurements from the tree ring data set. The tests revealed that the value of the parameter r under both cases was essentially the same and on this basis a value of 8 was assigned.

In practice, there are usually many observation relevant to the estimated ^{14}C age for a given calendar year, not just a single observation. Buck and Blackwell (2004) choose to carry out the calculations by treating each point on the curve separately, as this enables them to limit the observations that have to be considered in any one calculation. For each point on the curve they choose a suitable window, of at least 100 observations, of data points to use, so that the effect of excluding the remaining points is negligible.

The IntCal04 radiocarbon calibration curve is generated by the model outlined above at intervals of 5yr for for the range 0-12400 cal BP, 10yr for 12400-15000 cal BP and

20yr for 15000-26000 cal BP. Figures 1.1 (a) and (b) illustrate the difference between the two curves, IntCal98 and IntCal04. Figure (a) shows the period 0 to 500 cal BP, in which only a slight difference between the two curves is apparent. However, Figure (b) shows the two curves for the period 12500 to 14500 cal BP, in which there is a distinct difference between the two. IntCal04, which is estimated with a much higher resolution, has a much smaller 1-standard deviation envelope and is a much smoother curve. Clearly this could make a huge difference to interpretations when working in the latter part of the calibration period, this is illustrated in Section 4.3.3.

(a) (b)

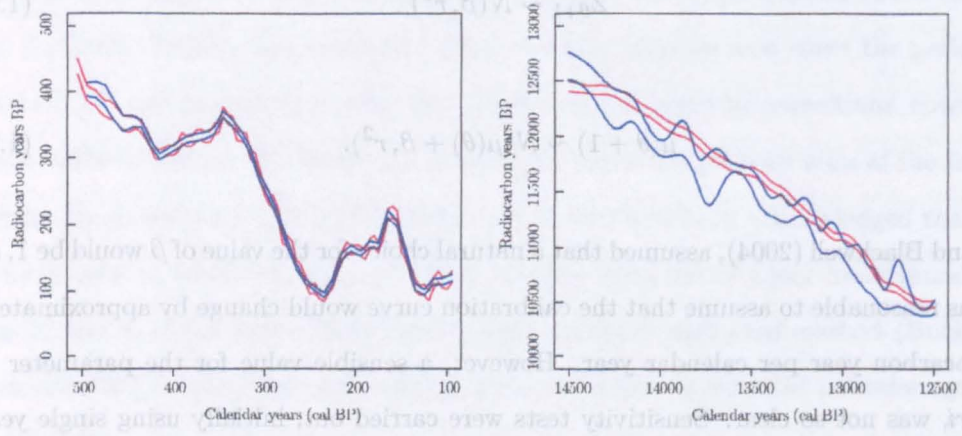


Figure 1.1: (a) IntCal98 (blue lines) and IntCal04 (red lines) terrestrial calibration curves both with a 1-standard deviation envelope for 0-500 cal BP (b) IntCal98 (blue lines) and IntCal04 (red lines) terrestrial calibration curves both with a 1-standard deviation envelope for 12500-14500 cal BP.

As well as IntCal04 two other calibration curves, SHCal04 and Marine04, were also constructed and ratified at the 18th International Radiocarbon Conference. A brief description of both will be given, but more details can be found in McCormac *et al.* (2004) and Hughen *et al.* (2004), respectively.

SHCal04: Southern hemisphere calibration, 0-11000 cal BP

As discussed on page 10, there is an offset between the ^{14}C concentration in the Northern and Southern hemisphere, resulting in samples from the Southern hemisphere being

'older' than those in the Northern hemisphere. As this is the case, there is clear need for two calibration curves, each constructed with calibration data obtained from the corresponding hemisphere. IntCal04 is the internationally agreed calibration curve for the calibration of terrestrial Northern hemisphere samples.

SHCal04 is the internationally agreed calibration curve for the terrestrial samples that originated form the Southern hemisphere. The data available for the construction of the Southern hemisphere calibration curve (McCormac *et al.*, 2004) are limited and only cover the period 50 to 990 cal BP. In order to extend the calibration curve back beyond 990 cal BP, the offset between SHCal04 and IntCal04 needs to be understood. By considering IntCal04 and SHCal04 separately, back to 990 cal BP, it is seen that the offset between them varies gradually over time but the direction and magnitude of the offset is fairly consistent. A random effects component is added to the random walk model (Buck and Blackwell, 2004) to allow for the offset to vary slowly over time. The offset is based on the variability of the offset found between 50-990 cal BP, which was 55-58 years, with an uncertainty that increases from ± 7.9 at 1000 cal BP to ± 25 at 11000 cal BP. Further details of the construction of SHCal04 can be found in Buck and Blackwell (2004).

Marine04: Marine radiocarbon age calibration, 0-26000 cal BP

So far only the calibration of terrestrial samples has been discussed. As seen on page 10 not only is there an offset between the concentration of ^{14}C in the Northern and Southern hemisphere over time but there also exists an offset between the concentration of ^{14}C in the oceans and on land. This results in the need to allow for additional sources of uncertainty when dating samples of marine origin. As a result a separate calibration curve, Marine04, has been constructed for the calibration of marine samples.

The Marine04 calibration curve is constructed in two parts using a combination of the tree-ring data and marine data sets. The first section of Marine04, from 0-10500 cal BP, is constructed using the dendrochronology based curve of IntCal04. The curve is converted using an ocean atmospheric box diffusion model (Hughen *et al.* 2004) to

provide 'global' ocean mixed layer ^{14}C ages. To construct the curve beyond 10500 cal BP, ^{14}C measurements are available from foraminifera[10] in varved sediment and U-series dated corals. The individual marine data sets were corrected, by subtracting ΔR (the difference between the regional surface ocean ^{14}C ages and the 'global' mixed layer ^{14}C ages). The output from both the ocean atmospheric box diffusion model and the corrected foraminifera and coral ^{14}C data are then combined using the random walk model (as detailed on page 15) to estimate the underlying Marine04 calibration curve. The Marine04 calibration curve is generated at intervals of 5yr for the range 0-10000 cal BP, 10yr for 10000-15000 cal BP and 50yr for 15000-26000 cal BP. Full details of the Marine04 calibration curve can be found in Hughen *et al.* (2004).

NotCal04: Comparison/calibration ^{14}C records 26000-50000 cal BP

IntCal04 has extended the time period of calibration back to 26000 cal BP. However, there are a number of case studies and situations in which we may want to calibrate radiocarbon determinations beyond the scope of IntCal04. Currently there is no internationally agreed curve that extends further, although there exist various potentially suitable data sets from individual research projects extending back as far as 50000 cal BP (van der Plicht *et al.*, 2004). One reason why these individual data sets cannot be recommended for construction of a calibration curve, in this period of time, is that they deviate too much from one another. Nevertheless, the data contains important information with regard to the natural ^{14}C variations prior to 26000 cal BP. As a result the IntCal working team has spent time trying to understand the underlying properties of the calibration curve for this period. They believe that each data set can be used to build a 'comparison curve', which would have its own offset from the true underlying calibration curve. To construct the curve, for the period 26000-50000 cal BP, each individual data set is used in a random effects extension to the random walk model (used for construction of IntCal04). This model allows for the possibility of offsets and for the possibility that

[10]Definition taken from http://www.bartleby.com/11/104.html: A class of animals of very low organisation, and generally of small size, having a jelly-like body, from the surface of which delicate filaments can be given off and retracted for the prehension of external objects, and having a calcareous or sandy shell.

the offsets might vary over time, between comparison curves and the true calibration curve. Further details of the NotCal04 can be found in van der Plicht *et al.* (2004).

1.2 Outline of the thesis

This section offers a summary of the contents in the following chapters of this thesis.

Chapter 2, titled Statistics in radiocarbon dating, begins by introducing the concepts of Bayesian inference (yet methods for its implementation are discussed later in Chapter 3). This is followed by a review of the most important publications relating to the interpretation of radiocarbon determinations. The early research carried out in this area, such as Ottaway (1973) and Ward and Wilson (1978), focused their interpretations using a purely classical approach, while researchers during the late 1980s and 1990s utilized a Bayesian framework so they could take account of the various uncertainties involved, when interpreting a set of radiocarbon determinations.

The first half of Chapter 3 discusses methods for Bayesian implementation. A range of methods are considered, but the main attention focuses on the use and practical considerations of MCMC. The second half of this chapter, introduces the first of the new ideas, in particular, the introduction of a trapezium or sigmoidal prior for modelling the rate at which datable material was deposited/manufactured between the start and end of an archaeological phase.

Chapter 4 presents two case studies, in which the authors believe that the rate of deposition/manufacture was not uniform over the proposed range. The aim of the chapter is illustrate the difference in the archaeological interpretations, drawn from the data, when assuming both uniform and non-uniform rates of deposition/manufacture.

Chapter 5 initially recaps the types of *a priori* information that may arise during an archaeological calibration problem, as well as discussing how it is integrated into the existing models. This chapter then focuses on incorporating *a priori* information about the relations between phase boundary dates, in the form of joint prior distributions. When working with multiple phases there are two forms in which data may occur; temporal and spatio-temporal. This chapter concentrates on temporal data, typically arising from multiple phases within the same archaeological site.

The final area of research that we aim to tackle is spatio-temporal modelling. Throughout the thesis, all research so far has been based on the interpretation of temporal data alone. Chapter 6 presents the first steps for incorporating any spatial structure as well as temporal information, that arises from excavations, in order to combine data from related sites and to be able to make more coherent and satisfactory interpretations of the data. In Chapter 7 we revisit the human reoccupation of NW Europe case study this time to illustrate the difference in archaeological interpretations, drawn from the data, when implementing both non-spatio-temporal and spatio-temporal models.

The final chapter, Chapter 8, gives conclusions from the analysis carried out within this thesis and discusses further improvements and developments which could be undertaken.

Chapter 2

Statistics in radiocarbon dating

2.1 Introduction

The interpretation of radiocarbon data has received increasing interest from the statistical community in the last few decades. Such work includes Naylor and Smith (1988) who were among the first to develop tools for chronology building within a Bayesian framework. Initial attention focused upon the calibration and interpretation of radiocarbon data, quickly moving to incorporate chronological information from a range of different sources, including stratigraphic sequences, historic evidence, *etc.* (Litton and Lesse 1991, Buck *et al.* 1991, 1992).

During the late 1980s and early 1990s researchers adopted an inference scheme, based on Markov chain Monte Carlo (MCMC) simulation. Christen (1994b) has given an outlier analysis, within the same Bayesian framework and Christen *et al.* (1995) and Christen and Litton (1995) suggested and implemented a Bayesian approach to include *a priori* information about the rate at which samples within a sequence were deposited. Software packages like OxCal, described in Ramsey (2005), and BCal, described in Buck *et al.* (1999), implement some of the methods presented. Buck *et al.* (1996) and Litton and Buck (1996) review the field in more detail. More recent work has been carried out by Nicholls and Jones (2001), who propose an alternative formulation for non-informative

priors and also suggest using Bayes factors to help select between competing models for chronology building when based on radiocarbon data.

What follows is intended to give an overview of some of the research carried out in this area and all of the above will be reviewed in more detail. However, before doing so the probability notation used throughout the thesis will be set up as well as briefly introducing the concept of Bayesian inference and the use of prior probability distributions. The reason for this is that a large majority of methods discussed in this chapter utilizes the Bayesian framework.

2.2 Bayesian inference

Probability notation

Firstly, the notation used throughout the thesis is defined. $p(.|.)$ denotes a conditional probability distribution and similarly $p(.)$ denotes a marginal distribution. The same notation is used for continuous density functions and discrete probability mass functions. Capital letters are used to denote random variables, such as X, and lowercase letters are used to represent realized values of the random variables, such as x. Also the use of boldface is to distinguish vectors such as $\boldsymbol{x} = \{x_1, \ldots, x_n\}$ from a scalar variable x.

If X and Y are two random variables, defined on the same sample space then $p(x, y)$ defines the joint probability density function of X and Y. Similarly $p(x)$ denotes the marginal distribution of X and $p(y)$ denotes the marginal distribution of Y.

Bayesian inference is a form of statistical inference in which parameters are considered as random variables having a probability distribution reflecting the current state of knowledge. The Bayesian approach takes a subjective view of probability which can be used to express uncertainty about an event; as a consequence it is possible to make probability statements about parameters. Hence, prior to observing the outcome of an event, the experimenters can express their uncertainty about the parameter (or parameters) $\boldsymbol{\phi}$ in terms of a probability distribution. This distribution is called the

prior distribution of ϕ, written as $p(\phi)$. A probability model for the data, x, given the parameter, which describes their relation, can be summarized as a likelihood and denoted by $p(x|\phi)$.

Inference for ϕ is then made by combining $p(\phi)$ and $p(x|\phi)$ using Bayes' Theorem, which states that

$$p(\phi|x) = \frac{p(\phi)p(x|\phi)}{p(x)} \tag{2.1}$$

where $p(\phi|x)$ is the conditional distribution of ϕ given x, known as the posterior distribution and $p(x) = \int p(\phi)p(x|\phi)d\phi$ and is referred to as the normalizing constant, which is denoted by k for the remainder of this thesis. Note that the integral is over the whole range of ϕ and would be written as the summation in a case of ϕ being discrete.

An equivalent form of Equation 2.1 omits the factor $p(x)$, which does not depend upon ϕ, and can be considered as a constant, resulting in the unnormalised posterior density,

$$p(\phi|x) \propto p(\phi)p(x|\phi). \tag{2.2}$$

This simple expression captures the core of Bayesian inference.

2.2.1 Prior probability distributions

There are a number of different types of *a priori* distributions used within the Bayesian framework, the following material will discuss some of the most common types of priors that arise throughout the thesis.

Informative prior: An informative prior expresses specific and definite information regarding a parameter of interest. In most cases informative priors arise from an expert's opinion or from a previous study of a similar nature.

Non-informative prior: A non-informative prior, also referred to as a vague prior, occurs when there is relatively little information concerning a parameter. A uniform prior (over a range of parameter values) is often used to represent situations where little or

no *a priori* information is available. When using a non-informative prior the posterior distribution is wholly determined by the information contained within the data.

When asked to express a prior, different priors may be expressed depending upon the person. It may well be necessary to have a conventional prior that is accepted by all parties that is non-informative and we refer to this as a *reference prior.* There are various theoretical approaches to defining reference priors; non of them is universally accepted, except in very simple problems.

Improper prior: An improper prior is when a probability mass function does not sum to 1 for a discrete distribution and the probability density function does not integrate to 1 for a continuous distribution. For example, an unbounded uniform prior would be classed as an improper prior and some statisticians use improper priors as non-informative priors.

The main reason for discussing the different types of priors that arise, so early on in the thesis, is that when interpreting radiocarbon data, archaeologists are becoming increasingly aware of making use of their *a priori* knowledge. Their prior knowledge might arise from either past excavations or from experts within the subject field. In the applied literature reviewed (see Sections 2.3, 2.4 and 2.5) there are a number of cases where the archaeologists have quite specific prior knowledge, for example

- the rate at which material is deposited within an archaeological phase
- the ordering of θ's or phases from stratigraphic information
- the likely time elapsed between the deposits of each sample in a sequence of radiocarbon determinations.

2.3 Interpreting radiocarbon data

The material in this section firstly looks at how a single radiocarbon determination can be calibrated to transform the radiocarbon age into calendar years, before moving on to look at how to interpret a group of related radiocarbon determinations which belong to an archaeological phase.

2.3.1 Basic model and calibration

Suppose that we are interested in dating some event, and that we have a suitable sample of organic material that ceased metabolizing at the moment of that event. Before the sample is dated, the calendar age θ (measured in years cal BP, where the prefix cal denotes the result of radiocarbon calibration) in which it ceased metabolizing is unknown. As well as the unknown calendar age, this sample will also have a unique radiocarbon age, which relates to the amount of ^{14}C currently contained within the sample. The radiocarbon age is conventionally denoted by $\mu(\theta)$. Due to the nature of the samples available for radiocarbon dating and the experimental error associated with the dating process, the radiocarbon laboratories do not provide $\mu(\theta)$ accurately. What they do provide is an estimate of $\mu(\theta)$, referred to as x, which is a realization of the random variable X (*i.e.* if the sample was dated a number of times, the values given for x would vary, each time). Thus, X can be represented as $X{=}\mu(\theta) + \epsilon$, where

$$\epsilon \sim N(0, \sigma^2). \tag{2.3}$$

Since it is assumed that σ is known (and provided by the radiocarbon laboratory), X is only conditioned upon the unknown parameter θ. X is therefore modelled using a Normal distribution with mean $\mu(\theta)$ and variance σ^2,

$$X \sim N(\mu(\theta), \sigma^2), \tag{2.4}$$

where $\mu(\theta)$ represents the calibration curve and is usually expressed in a piece-wise linear form

$$\mu(\theta) = \begin{cases} a_0 + b_0\theta & (\theta \leq t_0) \\ a_l + b_l\theta & (t_{l-1} \leq \theta \leq t_l,\ l = 1, 2, \ldots, L) \\ a_L + b_L\theta & (\theta \leq t_L) \end{cases} \tag{2.5}$$

where t_l are referred to as the knots of the calibration curve, $L + 1$ is the number of knots and a_l and b_l are assumed to be known constants which ensure continuity at the

knots.

This assumption of normality is the most common and widely accepted assumption in the statistical analysis of radiocarbon data. The assumption is made by the vast majority of researchers in the field (see Ward and Wilson, 1978) and is derived from the fact that conventional radiocarbon determinations arise from counting the number of ^{14}C atoms decaying in a period of time. These counts have a Poisson distribution which can be approximated by the Normal distribution.

The assumption of normality is a difficult assumption to test, this has not been pursued here and it does not appear to be widely discussed in the literature, however there could be arguments for relaxing the assumption or even assuming specifically a distributional model with heavier tails.

Consider a radiocarbon determination from a single organic sample. We would like to use this, to help learn about its calendar age. Using the Bayesian framework, we can learn about the calendar age, by formalizing a likelihood, which relates θ to x, σ and $\mu(\theta)$. The appropriate likelihood, based on Equation 2.4, can be written as:

$$p(x|\theta) \propto \exp\left\{-\frac{(x-\mu(\theta))^2}{2\sigma^2}\right\}. \tag{2.6}$$

In absence of informative prior information, the convention is to assume that the prior value for θ is equally likely to lie anywhere over the range of the calibration data. This is usually represented using a uniform, vague prior for θ, *i.e.* $p(\theta) \propto 1$, for $0 < \theta$. This implies that the posterior density of θ, $p(\theta|x)$, is essentially equivalent to the likelihood. Due to the wiggly nature of the calibration curve (see Figure 1.1), the posterior density can often be non-symmetric and multimodal, which can make interpretations difficult, as illustrated in Figure 2.1.

In the following sections, a number of the relevant publications related to interpretation of radiocarbon determinations are discussed. In doing so a wide variety of problems will be considered which use a range of statistical techniques. The review is presented in

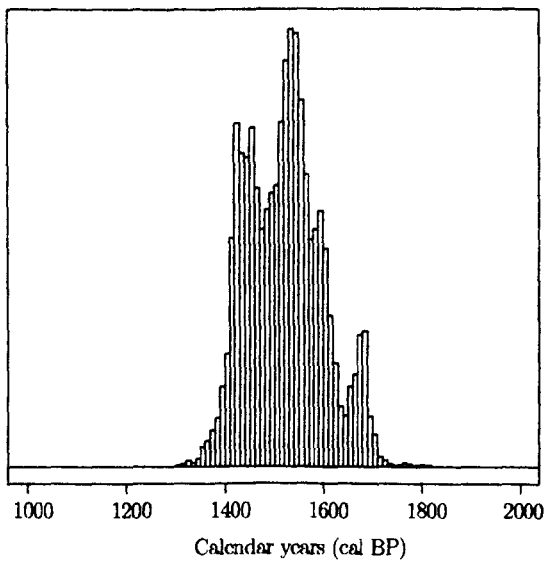


Figure 2.1: Calibrated date, illustrating non-symmetry and multimodality, for the radiocarbon determination 1630±60, when using the internationally agreed radiocarbon calibration curve, IntCal04.

approximate chronological order, as techniques have developed over several decades.

2.3.2 Interpreting groups of related radiocarbon determinations

Probably one of the most widely quoted works on the statistical analysis of sets of radiocarbon determinations is that of Ward and Wilson (1978). The paper is concerned with techniques for comparing and combining a set of radiocarbon determinations. In doing so the authors consider two separate cases,

- **Case I:** When two or more radiocarbon determinations are made on the same object.
- **Case II:** When one radiocarbon determination is made on two or more samples that are not known to be from the same object.

The notation used, extends that of Section 2.3.1, where n represents the number of radiocarbon determinations of the form $x_1 \pm \sigma_1, \ldots, x_n \pm \sigma_n$ and each x_i is a realization

of the random variable X_i. [Note, Ward and Wilson do not make use of a calibration data set and therefore no calibration procedure is used. This is no surprise since at the time the paper was written calibration was not routine.]

Case I, is used when all the radiocarbon determinations under consideration are known to be replicated from the same object, hence all have the same true (unknown) mean, μ. It is assumed that any differences found between the radiocarbon determinations will have resulted from errors in the dating process, ϵ_i, as given in Equation 2.3. Therefore a radiocarbon determination is modelled as

$$X_i \sim N(\mu, \sigma_i^2). \tag{2.7}$$

The authors are interested in testing the null hypothesis,

$$H_0 : X_i = \mu \quad \text{for } i = 1, \ldots, n.$$

That is to say, that a set of radiocarbon determination are consistent (*i.e.* all have the same true radiocarbon age) by using the following test statistic,

$$T = \sum_{i=1}^{n} \frac{(x_i - x_p)}{\sigma_i^2}, \tag{2.8}$$

where x_p is the pooled mean of the radiocarbon determinations and is given by

$$x_p = \left(\sum_{i=1}^{n} \frac{x_i}{\sigma_i^2} \right) / \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right). \tag{2.9}$$

The test statistic, T, has a Chi-square distribution with $n - 1$ degrees of freedom. If the null hypothesis is not rejected, hence the radiocarbon determinations are judged not to be significantly different then they can be combined to give a pooled age, x_p, and a corresponding variance,

$$V(x_p) = \left(\sum_{i=1}^{n} \frac{1}{\sigma_i^2} \right)^{-1}. \tag{2.10}$$

If the null hypothesis is rejected, hence the radiocarbon determinations are found to be significantly different, then they should not be combined.

Case II, is used when one does not know whether the set of radiocarbon determinations are estimating the same calendar age, or effectively indistinguishably different ages. For this reason, it cannot be assumed that each x_i has the same true mean, *i.e.* each x_i has its own mean μ_i for $i = 1, \ldots, n$.

Unlike Case I, where the only source of error considered is from the dating process, additional error terms are introduced. The authors feel there is a need to account for the "error factor in the calibration curve", f_i, for each radiocarbon determination and assumes that f_i is independent of $f_j (i \neq j)$. The authors also include an additional error term, g_i, to allow for the effect of 'sunspot activity' (see page 8). Both error terms, f_i and g_i, are assumed to be normally distributed with a mean of zero and a standard deviation σ_f and σ_g, respectively.

In Case II, taking account of the additional error terms, a radiocarbon determination can now modelled as

$$X_i \sim N(\mu_i, s_i^2) \tag{2.11}$$

where $s_i^2 = \sigma_i^2 + \sigma_f^2 + \sigma_g^2$.

The authors are now interested in testing the null hypothesis,

$$H_0 : \mu_1 = \mu_2, \ldots, \mu_n.$$

The test statistic, T, as given in Equation 2.8 is used replacing σ_i^2 with s_i^2. If the null hypothesis is not rejected, and from an archaeological consideration it seems appropriate, the radiocarbon determinations can be combined. The pooled radiocarbon age is given as in Equation 2.9 and the variance of the pooled age as in Equation 2.10 (in both cases replacing σ_i^2 with s_i^2).

The techniques suggested in Ward and Wilson (1978) are still widely used and also implemented in calibration software such as OxCal (Ramsey, 2005). Given the availability of the high precision calibration data, which gives rise to a non-monotonic calibration curve, it is clear that there is not a one-to-one relationship between radiocarbon and calendar years. This means that calendar age estimates are typically multimodal. Using Figure 1.1a, for example, consider what will happen if we obtain a radiocarbon determination with a mean radiocarbon age of 150 years BP. Even if we obtain this value with zero uncertainty, it could relate to any one of three calendar years. This suggests that a statistical model based on calibration should be used, and that any consistency checking should be undertaken on the calendar rather than the radiocarbon timescale.

In a series of papers, Ottaway and her colleagues discuss the desire to summaries sets of radiocarbon determinations. The first paper, Ottaway (1973), proposes a technique for summarizing sets of radiocarbon determinations diagrammatically, referred to as the 'inter-quartile range' or 'dispersion diagrams'. Ottaway constructed the dispersion diagrams by ordering the x_i's along the radiocarbon timescale and then calculated the lower quartile, median and upper quartile, see Figure 2.2. Ottaway then defines the period of time between the lower and upper quartile as the 'flourit'[1] of a culture.

Ottaway later proposes a method for "correcting" dispersion diagrams to the calendar time scale by using Suess's calibration curve, Suess (1970). As a result of the calibration curve being non-linear and non-monotonic, some of the x_i's correspond to more than one possible date on the calendar time scale. To overcome this problem Ottaway gave fractional weights, 1/(no. of possible corrected dates), to each of the ambiguous dates. The dispersion diagrams were then produced, as before, but now on the calendar time scale and with the fractional weights taken into account when calculating the quartiles.

There are several concerns with the approach presented by Ottaway. Firstly, only the x_i's are taken into consideration the standard deviations, σ_i's, are ignored. The second

[1]The flourit of a culture can be thought of as the most prolific period.

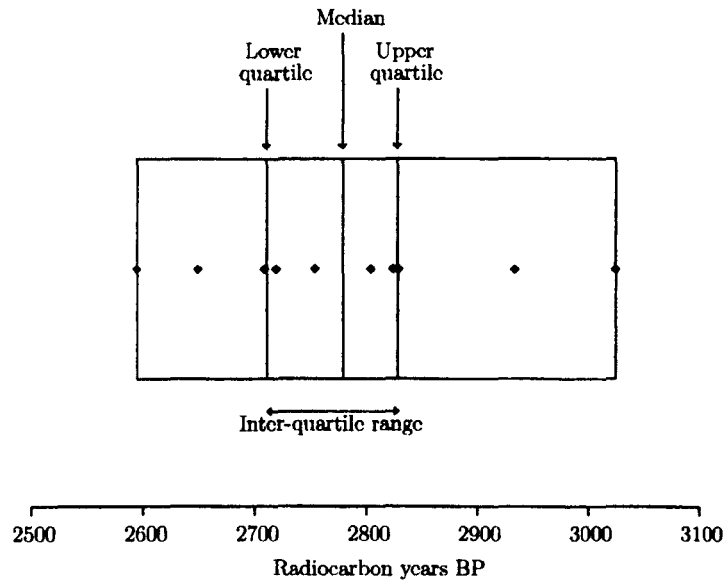


Figure 2.2: Construction of a dispersion diagram for a collection of 10 radiocarbon determinations, where ♦ represents the radiocarbon ages, x_i's.

concern, relates to the definition of the the flourit as being the period of time between the lower and upper quartiles. Ottaway suggests that the inter-quartile range may be "too cautious an estimate for the 'flourit' of a culture and that an expansion to include 2/3 or 3/4 of the data may give a more meaningful result".

The next paper in the series is Aitchison *et al.* (1990). The authors are interested in estimating the duration of an archaeological phenomenon, such as the occupation of a settlement. To do so the authors consider the suggestion proposed in Ottaway (1973). However, the authors point out that the lower and upper quartiles as calculated in Ottaway (1973) are just point estimates of the population quartiles, and that there is no measure of uncertainty accounted for in these estimates.

As a result, Aitchison *et al.* (1990) suggest a simple *ad hoc* extension to the method proposed in Ottaway (1973), in order to take account of uncertainty when estimating the flourit of a culture. The method they propose is referred to as the 'Extended quartile interval', which is calculated by constructing two series. The first series is calculated by

$x_i - 0.67\sigma_i$ and the second series by $x_i + 0.67\sigma_i$. Using the two series and evaluating the lower quartile from the first series and the upper quartile from the second series, the extended quartile interval is defined to be the difference between the two. Clearly, this method will extend the length of the flourit of a culture relative to that obtained using the conventional inter-quartile range.

The final issue addressed in Aitchison *et al.* (1990), is whether the individual radiocarbon determinations should first be calibrated before the flourit is calculated or whether it is sufficient to calibrate the lower and upper quartiles and use the resulting calibrated dates to calculate the flourit. To investigate, the authors chose not to work with the extended quartile interval, but to take the simplified definition of a flourit of culture, as defined in Ottaway (1973). As discussed in the review of Ottaway (1973) some of the x_i's correspond to more than one possible date on the calendar timescale, as a result of the wiggles in the radiocarbon calibration curve. Hence, there are multiple intercepts of x_i with $\mu(\theta)$. This results in a series of $\theta_{i,j}$'s where $i = 1, \ldots, n$ (n is the number of radiocarbon determinations) and $j = 1, \ldots, k_i$ where k_i represents the number of intercepts of x_i with $\mu(\theta)$.

To overcome this problem the authors considered two possible approaches. The first approach takes the average of the $\theta_{i,j}$'s, to arrive at a single date on the calendar time scale, for each x_i. The second approach applies the fractional weighting scheme as described in the review of Ottaway (1973).

The authors calculate the flourit, before and after calibration, using the above two approaches for a number of case studies, with varying numbers of radiocarbon determinations. The authors conclude that when using the first approach the flourit, whether calculated before or after calibration, is virtually the same. However, when using the second approach, differences did occur depending upon whether the calculations were carried out before or after calibration. The magnitude of the difference depends largely on where on the calibration curve a given data set falls.

However, it is felt that calibrating the individual radiocarbon determinations before calculating the flourit seems a much more sensible suggestion than calibrating the lower and upper quartiles and using these dates to calculate the flourit. As clearly, depending upon the part of the calibration curve under consideration, multiple dates on the calendar time scale might arise for the lower or upper quartile which could cause complications when calculating the flourit.

In the extension of the work of Ottaway (1973) and Aitchison *et al.* (1990) the next paper in the series is Aitchison *et al.* (1991). In this paper the authors develop a technique for summarizing sets of radiocarbon determinations on the calendar time scale with the inclusion of σ's. The authors aim to define and provide a sound statistical solution to rectify the problems of dispersion diagrams as defined in both Ottaway (1973) and Aitchison *et al.* (1990).

Their method is based on two assumptions

- "There exists a frequency distribution, $p(\theta)$, (with respect to the calendar time scale) of all possible artefacts or material from the phenomenon which might be sampled".
- "The actual artefacts or material sampled by the archaeologists are, as far as is possible, a reasonable representative sample from this frequency distribution".

The authors are interested in estimating the frequency distribution, $p(\theta)$, and then estimating the lower and upper quartiles in order to give an estimate of the flourit. Given a set of n radiocarbon determinations, of the form $x_1 \pm \sigma_1, \ldots, x_n \pm \sigma_n$, they proceed as follows

1. Solve the equation

$$x_i = \mu(\theta).$$

   Clearly, (as a result of the calibration curve, Pearson *et al.* (1986), being non-linear and non-monotonic) some of the x_i's may correspond to multiple dates, θ_i's, on

the calendar time scale. Hence, this results in a series of $\theta_{i,j}$'s, where $i = 1, \ldots, n$ and $j = 1, \ldots, k_i$ where k_i represents the number of intercepts of x_i with $\mu(\theta)$.

2. The second step is to provide an approximation of the standard error, $\hat{se}(\theta_{i,j})$, for each of the $\theta_{i,j}$'s by taking account of three factors

   (i) the σ_i's

   (ii) the errors, $\sigma_c(\theta_{i,j})$, from the calibration curve at the point $\theta_{i,j}$

   (iii) the slope or steepness, $\left|\frac{d\mu}{d\theta_{i,j}}\right|$, of the curve at point $\theta_{i,j}$.

   These three factors are combined to give an estimate

$$\hat{se}(\theta_{i,j}) = \frac{\sqrt{\sigma_i^2 + \sigma_c^2(\theta_{i,j})}}{\left|\frac{d\mu}{d\theta_{i,j}}\right|}.$$

3. The third step is to provide an estimate of the frequency distribution, $p(\boldsymbol{\theta})$, by combining the data $\theta_{i,j}$ as found in Step 1 with their estimated standard errors found in Step 2 using a non-parametric density estimation technique.

4. The final step is to provide an estimate of the flourit, by obtaining the cumulative distribution function, $F(\boldsymbol{\theta})$, and then estimating its lower and upper quartiles.

One main concern with the method used in Aitchison *et al.* (1991) is the violation of one of the assumptions relating to non-parametric density estimation. That is, the sample values, here the $\theta_{i,j}$'s, are assumed to be independent. Clearly, in the method outlined above, given $\theta_{1,1}$ then x_1 can be calculated and if x_1 is known then the rest of the $\theta_{1,j}$'s are known, hence the $\theta_{1,j}$'s are not independent.

The authors also use the definition of a flourit as "the period of time when the middle 50% of artefacts from the culture were produced, *i.e.* the lower and upper quartiles of the distribution", when it is clear from Ottaway (1973) that this definition may well be "too cautious an estimate".

All three of the papers discussed, Ottaway (1973), Aitchison *et al.* (1990) and Aitchison *et al.* (1991), realize the importance of estimating the duration of an archaeological phenomenon. Although not directly discussed, the following section sets up a novel approach to this problem. Bayesian methods are introduced to estimate the start and end dates of an archaeological phenomenon. Clearly, by knowing these two dates it would be possible to calculate the difference between the two, *i.e.* the duration of an archaeological phenomenon.

2.3.3 The first use of Bayesian statistics

Naylor and Smith (1988) offered a major contribution to the interpretation of radiocarbon determinations, by developing a model which takes account of the various uncertainties involved in relating observed radiocarbon determinations of artefacts to successive chronological start and end dates for significant phases or periods of activity. Just to recap, a phase is defined as collection of dateable material bounded early and late by events that are of archaeological importance.

The archaeological problem studied in Naylor and Smith (1988) relates to the Iron Age hillfort at Danebury. From this archaeological site it became apparent that there were four phases of pottery production. There was also a total of 65 radiocarbon determinations, each associated with a pottery fragment. On the basis of stylistic considerations, each pottery fragment was assigned to one of four ceramic phases by an expert in the subject field. Before moving on to discuss the detail of the Naylor and Smith (1988) paper, Figure 2.3 is intended to set up the basic notation used.

Consider a single vertical series of J abutting phases, with $J+1$ phase boundaries, in which the (unknown) calendar dates of the phase boundaries are represented by $\alpha_1, \alpha_2, \ldots, \alpha_{J+1}$. Phase 1, with boundary dates α_1 and α_2 represents the deepest phase (containing the oldest material) and Phase J, with boundary dates α_J and α_{J+1} represents the most recent phase (containing the youngest material). Within each phase

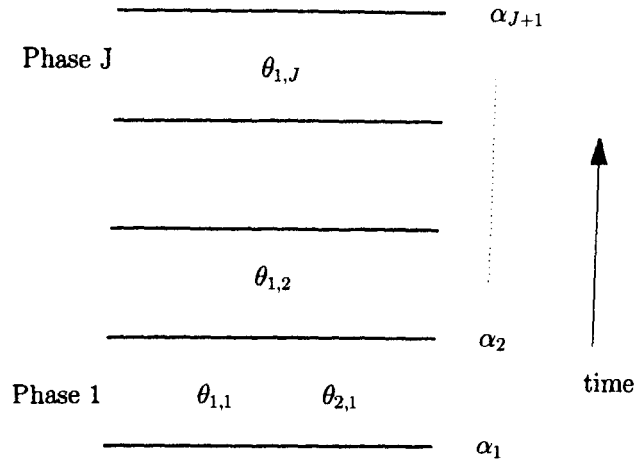


Figure 2.3: Schematic representation of abutting phases.

there are n_j organic samples suitable for radiocarbon dating and $x_{i,j}$ is used to denote the ith radiocarbon age in the jth phase with associated standard error $\sigma_{i,j}$. Each $x_{i,j}$ is associated with a true unknown calendar date $\theta_{i,j}$.

The authors are interested in making inferences, on the calendar time scale, about the dates of the four phases of pottery production. In particular, they would like to learn about the 5 unknown chronological dates, $\alpha_1, \alpha_2, \ldots, \alpha_5$, which represent the beginning of Phase 1, the end of Phase 1, the beginning of Phase 2 and so on. As the phases are regarded as abutting, this implies that the end of Phase 1, is the same event as the beginning of Phase 2. They use $\boldsymbol{\alpha} = (\alpha_1, \alpha_2, \alpha_3, \alpha_4, \alpha_5)$ to represent the start and end dates for each of the four phases with the assumption that $\alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 > \alpha_5$.

Having set up the problem in this hierarchical framework, the authors discuss their *a priori* beliefs about the rate of pottery production within the ceramic phases. They felt that representing the production rate within a ceramic phase as a Uniform distribution was an appropriate assumption to make, giving

$$p(\theta_{i,j}|\alpha_j, \alpha_{j+1}) = \begin{cases} (\alpha_j - \alpha_{j+1})^{-1} & \alpha_j > \theta_{i,j} \geq \alpha_{j+1}, \\ 0 & \text{otherwise.} \end{cases}$$

Naylor and Smith (1988) clearly state that "this uniform assumption could be replaced by say, a beta distribution over each interval, reflecting a gradual increase in production followed by a tailing off towards the end of the phase". However, it was not the intention of the paper to explore a range of possible models but to present a methodology for representing and analyzing a particular model.

The data consist of the 65 radiocarbon determinations in the form $x_{i,j} \pm \sigma_{i,j}$ and from Equation 2.4 we know that

$$p(x_{i,j}|\theta_{i,j}) \sim N(\mu(\theta), \sigma_{i,j}^2)$$

where $\mu(\theta)$ represents the piece-wise linear calibration curve discussed in Section 2.3.1.

All that remains now is to model the $p(\boldsymbol{\alpha})$. Naylor and Smith (1988) used a particularly simple form of prior, to represent a minimal state of prior knowledge, which can be expressed as follows

$$p(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{for } \alpha_1 > \alpha_2 > \alpha_3 > \alpha_4 > \alpha_5 > 0 \\ 0 & \text{otherwise.} \end{cases}$$

To calculate the posterior distributions of $\boldsymbol{\alpha}$, Naylor and Smith (1988) used numerical integration techniques (see Section 3.2.1), implemented using their own computer software. The authors also discuss briefly some specific posterior predictive functions that answer a variety of possible questions that may be of interest to archaeologists. For example, if we have a radiocarbon determination with an associated standard deviation what is the posterior predictive probability that the radiocarbon determination is from ceramic Phase j?

Although Naylor and Smith (1988) was a major contribution to the interpretation of radiocarbon determinations it contains two technical errors: The first being that they consider the year 0 BP as 1983 AD (as this was the year the data were obtained), where the current convention in the radiocarbon and archaeological community is to take 0 BP

as 1950 AD. The second error they made was to do with the use of the calibration curve. They used an old calibration data set, instead of the high-precision calibration data set which was published in 1986.

2.4 Some case studies and simple extensions

Litton and Lesse (1991) follows on from Naylor and Smith (1988) (written for a statistical audience) with the intention to review the basic modelling ideas of Naylor and Smith (1988) in an archaeological framework. In the hope that archaeologists would appreciate its significance. The authors see the work of Naylor and Smith (1988) as composed of five different stages, which they believe to be common to many archaeological calibration problems.

1. Defining the archaeological problem;
2. expressing the statistical model in terms of the question posed;
3. specifying the *a priori* information;
4. using statistical inference procedures;
5. interpreting the results.

They discuss each of the five aspects in a langauge and style better suited to the archaeological community. They also note that some archaeologists may be unhappy with some of the modelling assumptions made by Naylor and Smith (1988), such as: why are the phases non-overlapping?, why should the pottery fragments be uniformly distributed over a phase? but explain that the model can readily be adapted to allow for other complexities; while the overall Bayesian methodology will remain the same.

The combination of Naylor and Smith (1988) and Litton and Lesse (1991) leads to a new approach to the statistical analysis and interpretation of sets of radiocarbon

determinations that is based on the Bayesian framework. This new approach is followed up by Buck *et al.* (1991) and Buck *et al.* (1992).

Buck *et al.* (1991) seek further to bridge the gap between statistics and archaeology. They aim to explain Bayesian statistics to the archaeological community and to illustrate the approach taken by Bayesian statisticians when interpreting data.

A new case study is used to illustrate the methodology: a two-phase Neolithic village at Skara Brae, Orkney which preserves 3.5 meters of stratigraphy. The main part of the site has two occupation phases, Village 1 and Village 2, which are both preceded by a thin basal layer, separated from Village 1 by a thin sand layer. Both phases of occupation are associated with midden deposits[2]. Midden deposits of approximately 1m can be associated with the occupation of Village 1 and approximately 2m can be associated with Village 2. There is also a clear horizon (sand layer) between the two villages.

Since many factors affect midden accumulation rates, it is not possible to estimate the length of occupation from stratigraphy alone. Nor is it clear how to estimate the length of time elapsed between the end of Village 1 and the beginning of Village 2 (as a whole site can be immersed in sand in a single sand-storm or over an extended period of time). This suggests that a non-stratigraphic dating method is required to estimate the calendar dates of the start and end of the two villages and also the length of time elapsed between the two.

This is an unusual case study, in the sense that there are well marked phase horizons and a reasonable number of radiocarbon determinations associated with each. Rather than adopting the method used in Naylor and Smith (1988), at Skara Brae, the start and end dates of the two villages can be directly dated using the radiocarbon determinations available from the midden deposits. These four events are represented by the calendar years $\theta_1, \theta_2, \theta_3$ and θ_4 and (from the stratigraphic information) it is known that $\theta_1 > \theta_2 \geq \theta_3 > \theta_4$.

[2]Midden deposits are deposits of waste material and are commonly composed of domestic and food waste.

There are 14 radiocarbon determinations available, of the form $x_i \pm \sigma_i$, in which 1–4 provide information about the calendar date θ_1, 5–7 provide information about θ_2, 8–10 provide information about θ_3 and 11–14 about θ_4. This information is then explicitly introduced into the analysis and the marginal posterior distributions $p(\theta_i|\boldsymbol{x})$ are calculated. Full details of the calculations can be found in the Appendix of Buck *et al.* (1991).

2.5 The MCMC revolution

Buck *et al.* (1992) is slightly more technical than those immediately proceeding it and is aimed mainly at statisticians and archaeological scientists. This paper builds on the modelling ideas of Naylor and Smith (1988) and outlines the principles of Bayesian statistics, but also explains the technical difficulties that arise in the calculation of marginal densities for events, such as the start of a phase, on the basis of large number of radiocarbon determinations. Instead of adopting the numerical approximation techniques previously used, Buck *et al.* (1992) introduce the method of Gibbs sampling (see Section 3.2.2) to evaluate posterior densities. The methodology is illustrated through two case studies one of which will be discussed here, *i.e.* a reanalysis of the Danebury data used in the Naylor and Smith (1988) paper. There were two main reasons to reanalyze the data, firstly the two technical errors made by Naylor and Smith (1988) and secondly the recent innovations in statistical methods based on the Gibbs sampler. Initially the same basic model was adopted for the interpretation of the Danebury data.

Buck *et al.* (1992) felt that there was one major archaeological criticism of the model and this was the assumption of abutting phases. Therefore the problem was remodelled without this assumption so that the authors could test whether or not the phases are likely to be abutting.

The model for the Danebury data was reformalized as follows. Again, primary interest lies in estimating the start and end dates for each of the phases. So, let α_j and β_j

represent the start and end dates (cal BP) of phase j (for j=1,2,3,4). Due to the absence of any *a priori* information, it is conventionally assumed that α_j and β_j lie anywhere in the range of the calibration curve, subject to the constraint $\alpha_j > \beta_j$, as the dates are measured in years cal BP. No assumption on the ordering of the phases was made.

Now let n_j be the number of samples assigned to the jth phase and, using the previously notation, $\theta_{i,j}$ represents the calendar date of the ith radiocarbon determination in the jth phase. In the absence of any other *a priori* information about $\theta_{i,j}$, the rate of pottery production within any phase is still assumed to be uniform. It was also assumed that the phases are independent of each other and therefore,

$$p(\alpha_1, \beta_1, \alpha_2, \beta_2, \alpha_3, \beta_3, \alpha_4, \beta_4) = \prod_{j=1}^{4} p(\alpha_j, \beta_j)$$

where

$$p(\alpha_j, \beta_j) = \begin{cases} c_j & \text{for } \alpha_j > \beta_j \\ 0 & \text{otherwise} \end{cases}$$

and c_j is a constant.

So, the clear difference between the two models is the assumption of the relations between the phases. After the reanalysis of the data using the second model, the authors made the following assessment of the relations between phases. They do not believe that the ceramic phases are abutting in time. In fact, the suggestion is that although the phases show a clear progression through time there is considerable overlap in pottery production from the different ceramic phases.

With the basic models devised, illustrated and implemented using both numerical integration and MCMC, a basic chronology building framework was in place and (with enthusiasm building among the user community) extensions to the basic models soon began to be developed. The following section reviews a selection of the papers that reported on the most important of these extensions.

2.5.1 Prior information about time elapsed between deposits

Archaeologist sometimes have information about the likely time elapsed between the deposits of each sample in a sequence of radiocarbon determinations. This is particulary common when several radiocarbon measurements have been made on a piece of wood so that the time elapsed between the rings or layers can be estimated. In such situations tree-ring dating prior to radiocarbon dating may be used as *a priori* information about the time elapsed between successive radiocarbon determinations. Wiggle matching is

(a) (b)

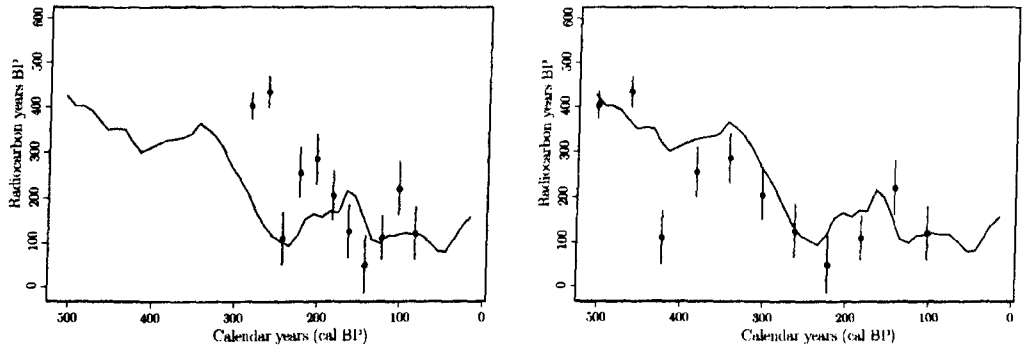


Figure 2.4: An example of a floating chronology with a gap of (a) 20 years (b) 40 years, between radiocarbon determinations, example reproduced from Christen *et al.* (1995).

the name given to the technique in which a sequence of related samples are dated using the high-precision radiocarbon calibration curve. The original technique used was to order the radiocarbon determinations according to the archaeological chronology *i.e.* the object dated by determination $x_i \pm \sigma_i$ is known to be earlier than the object dated by determination $x_{i+1} \pm \sigma_{i+1}$. Then the values $x_i - \sigma_i$, x_i, $x_i + \sigma_i$ were plotted along evenly spaced vertical lines where the gap between the lines is n calendar years. This can be referred to as a 'floating curve' which is subsequently compared with the high-precision calibration curve. The above step is repeated with different gaps of size n calendar years until a 'satisfactory' match is found. This match then provides an estimate for the calendar age for each of the dated objects, see Figure 2.4. It is clear that there is no

common approach to carrying out the necessary comparisons. Some work (Weninger, 1986) use highly subjective visual matching of graphs, which results in problems, such as what measure should we use to compare one wiggle match to another. While others use statistical methods based on least squares (Pearson, 1986), however this technique can only be used when the time elapsed gaps between related samples is known. This led Christen to propose a more general approach to archaeological wiggle matching which utilized the Bayesian framework, a detailed description of his work can be found in the following subsection.

Bayesian approach to wiggle matching

Given a set of radiocarbon determinations, with corresponding unknown calendar dates $\theta_1, \ldots, \theta_n$, when there is *a priori* information about the relative dates $\theta_i - \theta_{i-1}$ (for $i = 2, 3, \ldots, n$) it is referred to as a 'floating chronology'. There are two common examples that arise in archaeology. Firstly, it occurs when constructing a 'tree-ring chronology', when we radiocarbon date tree rings and the number of rings between samples is known. In this case it is assumed that $\theta_i - \theta_{i-1} = \gamma_i > 0$ (γ_i is the time interval between successive events) for $i = 2, 3, \ldots, n$. Secondly, it occurs when samples are known to have 'stratigraphic ordering' and we also have knowledge about the relative dates $\theta_i - \theta_{i-1}$. (Further *a priori* information about the relative dates $\theta_i - \theta_{i-1}$ might also be available, such as maximum and minimum time spans.) In both cases it is assumed that $\theta_n > \theta_{n-1} > \ldots > \theta_2 > \theta_1$, clearly this is an extension of the type of problem illustrated in Section 2.4.

Christen (1994a) and Christen and Litton (1995) suggested and implemented a general Bayesian approach to wiggle matching and implemented the method for a case study in which the γ_i's are unknown.

The key point in this problem is that there is *a priori* information about the relative dates $\theta_i - \theta_{i-1}$, which is included in the analysis through $p(\theta_i | \theta_{i-1}, \theta_{i+1})$ and it is assumed, *a priori*, that the relative dates $\theta_i - \theta_{i-1}$ are independent. Given this, the prior information

about $\theta_i - \theta_{i-1}$ is then defined by the density function g_i, giving

$$p(\theta_i|\theta_{i-1}, \theta_{i+1}) \propto g_i(\theta_i - \theta_{i-1})\, g_{i+1}(\theta_{i+1} - \theta_i) \quad \text{for} \quad i = 2, 3, \ldots, n-1.$$

Then using Equation 2.4 and assuming independence

$$p(\boldsymbol{x}|\boldsymbol{\theta}) = \prod_{i=1}^{n} p(x_i|\theta_i).$$

The full conditionals can then be written as,

$$p(\theta_i|x_i, \theta_{i-1}, \theta_{i+1}) \propto \exp\left\{-\frac{(x_i - \mu(\theta_i))^2}{2\sigma_i^2}\right\} g_i(\theta_i - \theta_{i-1})\, g_{i+1}(\theta_{i+1} - \theta_i).$$

For different cases of floating chronologies, the prior information about $\theta_i - \theta_{i-1}$ may be modelled using different functions g_i. For example, in Christen (1994a) an example of a 'tree-ring chronology' is given in which the γ_i's are known ($\theta_i - \theta_{i-1} = \gamma_i$). As a result, the function g_i is simply defined as

$$g_i(x) = \begin{cases} 1 & \text{if } x = \gamma_i, \\ 0 & \text{if } x \neq \gamma_i. \end{cases}$$

In this particular case it is only necessary to calculate the distribution of θ_1 as the rest of the θ_i's can be calculated from it via

$$\theta_i = \theta_1 + \sum_{j=2}^{i} \gamma_j, \tag{2.12}$$

and the posterior distribution of θ_1 can be given by

$$p(\theta_1|\boldsymbol{x}) \propto \prod_{i=1}^{n} \exp\left\{-\frac{(x_i - \mu(\theta_i))^2}{2\sigma_i^2}\right\}$$

where the form of θ_i is given as in Equation 2.12. In Christen (1994a), numerical integration procedures were used to calculate the posterior distribution of θ_1 for a fairly

simple case study, but these kinds of models can also be implemented more generally using MCMC and are now available in software packages such as OxCal (Ramsey, 2005) and Bwigg (Christen, 2003).

2.5.2 Prior information about the rate of deposition

Christen and colleagues have also developed methods that allow the inclusion of *a priori* knowledge about the rate of deposition. In particular Christen *et al.* (1995) and Christen (1994a), discuss work in which they include *a priori* information about the rate of deposition of dry mass in lower levels (catotelm) of peat bogs.

Mathematical models of the growth of the catotelm have been proposed which relate the cumulative mass of peat above a particular depth to the calendar age of peat at that depth. Christen demonstrates how radiocarbon dating and the use of Bayesian statistics can be used to make inferences about the relationship between calendar age and cumulative mass, and to estimate the accumulation and decay rates.

Christen suggests that the problem should be modelled as follows. Consider an arbitrary fixed datum at depth d_0 (below which all other samples will lie). Then θ_0 represents the unknown age (in cal BP) of the peat at depth d_0. Now consider peat at depth d ($d > d_0$), let its calendar age be θ ($\theta > \theta_0$) cal BP. Let M represent the cumulative dry mass (in g cm^{-2}) of material deposited between d_0 and d. Let p be the rate at which dry mass is added to the peat, and let a be proportional to the rate of decay of peat after deposition (a is assumed constant over the entire depth of peat).

Adapting earlier work carried out by one of the co-authors (Clymo), Christen *et al.* (1995) makes several suggestions for modelling peat deposition, one of which is

$$M = \frac{p}{a}\Big(1 - \exp^{-a(\theta-\theta_0)}\Big). \tag{2.13}$$

There are two components to such a model: the first relates the cumulative mass to

the calendar age and the second relates the calendar age to the radiocarbon age. The second has been discussed in detail in Section 2.3.1. The first will be discussed here. Suppose there is a series of n radiocarbon determinations $x_1 \pm \sigma_1, x_2 \pm \sigma_2, \ldots, x_n \pm \sigma_n$ from peat samples taken at successive depths $d_1 < d_2 < \ldots < d_n$ corresponding to calendar years $\theta_1, \theta_2, \ldots, \theta_n$. Since peat age increases with depth, it implies that the θ's must be ordered, hence $\theta_1 < \theta_2 < \ldots < \theta_n$.

The authors then defines m_i, the cumulative dry mass at depth d_i, in accordance with Equation 2.13,

$$m_i = \frac{p}{a}\Big(1 - \exp^{-a(\theta_i - \theta_0)}\Big).$$

Rearranging the above equation in terms of θ_i gives

$$\theta_i = \theta_0 - a^{-1} \ln (1 - p^{-1} a m_i). \tag{2.14}$$

This results in the calendar age θ_i been expressed in terms of the unknown parameters $\varphi = (\theta_0, p, a)$ which we wish to learn about. Thus, the likelihood can then be written as

$$p(\varphi|x) \propto \prod_{i=1}^{n} \exp\left\{-\frac{(x_i - \mu(\theta_i))^2}{2\sigma_i^2}\right\} \tag{2.15}$$

where θ_i is given by Equation 2.14.

The authors then use the likelihood as defined in Equation 2.15 and assume informative priors for the two parameters, p and a. In applying the above method to data arising from specific peat formations, the authors demonstrate that incorporating *a priori* information and other known sources of error can elegantly be accounted for.

2.5.3 Outlier detection

Outliers in radiocarbon dating are thought to be relatively common, since there are a number of factors that affect the quality of radiocarbon dating and which could lead to the production of an outlying age estimate. Such factors include the following

1. Contamination with older or younger material (see Bowman 1990, page 27).

2. The quality of sample handling and preparation in the laboratory to ensure samples undergo the appropriate pretreatment (see Bowman 1990, pages 28-30).

3. The quality of care taken to ensure that samples can realistically provide calendar date estimates for the events we wish to learn about (see Bowman 1990, Chapter 5).

Radiocarbon laboratories only have control over the second factor and no control over the other two factors. This implies that radiocarbon laboratories can be producing top quality radiocarbon determinations, which may still contain outliers in relation to the event of interest. It is therefore important that any statistical methods used with a set of radiocarbon determinations should be robust to outliers. Christen (1994a, b) implemented a Bayesian approach for modelling and identification of outliers in groups of related radiocarbon determinations.

The approach taken is as follows. The problem is simplified by taking only a single phase and using the same notation as in Buck *et al.* (1991). Let α represent the start of the phase and β represent the end of the phase. It is assumed that there are n radiocarbon determinations within the phase and it is assumed that they are uniformly distributed over the interval α to β.

Christen suggested that if x_i needs to be shifted by δ_i (radiocarbon years) in order for it to be consistent with the rest of the samples in the same phase then it is said to be an outlier. This is formulated as

$$x_i|\theta_i, \phi_i, \delta_i \sim N(\mu(\theta_i) + \phi_i\delta_i, \sigma_i^2) \tag{2.16}$$

where

$$\phi_i = \begin{cases} 1 & \text{if } x_i \text{ needs a shift} \\ 0 & \text{otherwise.} \end{cases}$$

So essentially the phrase 'needs a shift' can be interpreted as 'is an outlier'. The prior uncertainty concerning whether or not a shift is needed is measured by $P(\phi_i = 1)$ and $P(\phi_i = 0)$, respectively. It is assumed in Christen (1994a, b) that whether or not the ith radiocarbon determination needs a shift is independent of other dates and also of α and β. As a result of this, the likelihood derived can be expressed as

$$p(\boldsymbol{x}|\boldsymbol{\theta}, \boldsymbol{\delta}, \boldsymbol{\phi}, \alpha, \beta) = \prod_{i=1}^{n} p(x_i|\theta_i, \delta_i, \phi_i)\, p(\theta_i|\alpha, \beta).$$

As seen previously $p(\theta_i|\alpha, \beta) \sim U(\beta, \alpha)$ for $\beta < \theta_i < \alpha$ and as a result of this

$$p(x_i|\theta_i, \delta_i, \phi_i) \propto \exp\left\{-\frac{(x_i - \delta_i\phi_i - \mu(\theta_i))^2}{2\sigma_i^2}\right\}.$$

As in previous models, *a priori* information about the boundary dates are typically vague, $p(\alpha, \beta) \propto 1$ for $\beta < \alpha$. Since there is no *a priori* knowledge about the size of any shifts, it is also reasonable to assume a vague prior for δ_i. Typically, however, the prior belief about the probability that a sample is an outlier is more informative. The prior probability that the ith sample is an outlier is represented by q_i. That is $P(\phi_i = 1) = q_i$ and the prior probability that the ith sample is not an outlier is $P(\phi_i = 0) = 1 - q_i$. Christen (1994a, b) suggested that unless there is case specific expert knowledge then taking $q_i = 0.1$ represents a sensible vague prior.

The method devised in Christen (1994a,b) for detecting outliers has been implemented using MCMC and is readily available in the software BCal (Buck *et al.*, 1999).

2.5.4 Remodelling the calibration curve

Christen (1994a) suggests remodelling the radiocarbon calibration curve, within the Bayesian framework, in order to take account of the uncertainties in the calibration data. This suggestion arose as a consequence of the radiocarbon dating technique improving and hence producing high-precision radiocarbon determinations.

The suggestion made in Christen (1994a) is that X should be modelled as being normally distributed with a mean $\mu(\theta)$ and a variance given by $\omega^2(\theta)$,

$$X|\theta \sim N(\mu(\theta), \omega^2(\theta)), \tag{2.17}$$

where $\omega^2(\theta) = \sigma^2 + \sigma^2(\theta)$.

The first term of $\omega^2(\theta)$ represents the reported standard deviation from the radiocarbon laboratory and the second term reflects the uncertainty in the calibration data.

Christen shows that a reasonable estimate of $\sigma^2(\theta)$ can be given as

$$\sigma^2(\theta) = \sigma_k^2 \left(\frac{\theta - t_{k-1}}{t_k - t_{k-1}} \right) + \sigma_{k-1}^2 \left(\frac{t_k - \theta}{t_k - t_{k-1}} \right)^2 + \lambda^2 \frac{(\theta - t_{k-1})(t_k - \theta)}{t_k - t_{k-1}} \tag{2.18}$$

where t_k is the calendar date of the kth knot and σ_k is the standard deviation of the calibration curve at the kth knot. The term λ in Equation 2.18 is used to account for the short term variability in the atmospheric ^{14}C levels, estimated as 20, based on a sample from published experimental data.

Thus the likelihood, $p(x|\theta)$, as seen in Equation 2.6 is now corrected with the addition of the variance $\sigma^2(\theta)$. Hence, assuming a vague prior for θ, the posterior density is essentially equivalent to the likelihood and is given by

$$p(\theta|x) \propto \frac{1}{\omega(\theta)} \exp\left\{ -\frac{(x - \mu(\theta))^2}{2\omega^2(\theta)} \right\}. \tag{2.19}$$

2.6 Alternative prior specification

Nicholls and Jones (2001) discuss the prior, $p(\alpha)$, currently used in the literature (*i.e.* Buck *et al.*, 1992) which was intended to be reasonably non-informative. The authors are particulary concerned with modelling groups of related radiocarbon determinations and the inferences on the calendar dates of phase boundary parameters that are made when

the conventional prior is used. The authors believe that when using the conventional prior and when the duration of an archaeological phenomenon (such as the occupation of settlement) is over a relatively small timescale in comparison to the precision of the radiocarbon determinations, that the non-informative priors generate a bias towards wider date ranges of the archaeological phenomenon, which does not reflect substantial prior knowledge.

This prompted the authors to propose an alternative formulation for a non-informative prior, in which the distribution of the difference between the earliest and latest dates has a uniform distribution. The following material discusses both the properties and motivation behind their alternative prior, as well as demonstrating how it is derived from a simple physical model of deposition.

Building on the notation used by Naylor and Smith (1988), there are J abutting phases and $J + 1$ phase boundaries (*i.e.* Phase 1, with boundaries α_1 and α_2, represents the deepest phase containing the oldest material). Within each phase there are n_j radiocarbon determinations. The only difference in notation is that Nicholls and Jones (2001) define a sequence of data to be modelled as lying in a finite interval (P, A) of length R, where the P stands for *terminus post quem* - "date after which" and A stands for *terminus ante quem* - "date before which". (In Nicholls and Jones (2001) all dates are given in years AD, where the convention throughout this thesis is to give dates in calendar years BP. Thus, this review will differ slightly from the notation given in Nicholls and Jones, 2001).

The main interest lies in the distribution of $p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ which summarizes the *a priori* knowledge before the radiocarbon determinations are available. As in Naylor and Smith (1988) and Buck *et al.* (1992) it seems natural to model θ conditionally on the layer boundary dates, $\boldsymbol{\alpha}$, so $p(\boldsymbol{\theta}, \boldsymbol{\alpha})$ is split into two components

$$p(\boldsymbol{\theta}, \boldsymbol{\alpha}) = p(\boldsymbol{\theta}|\boldsymbol{\alpha})p(\boldsymbol{\alpha}).$$

In the absence of radiocarbon data, the parameter $\theta_{i,j}$ might take any value between α_j and α_{j+1} with equal probability,

$$p(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \prod_{j=1}^{J} (\alpha_j - \alpha_{j+1})^{-n_j}, \tag{2.20}$$

this is the conventional prior used by previous researchers *e.g.* Naylor and Smith (1988) and Buck *et al.* (1992).

The main question of interest raised in Nicholls and Jones (2001), is what prior density should be used for the set of phase boundary dates, $p(\boldsymbol{\alpha})$. One natural choice of prior is to assume that any legal set of dates is equally likely, hence a uniform prior density (Nicholls and Jones refer to this as the *constant prior density*). This is the prior found in the earlier work by Buck *et al.* (1992) which Nicholls and Jones believes weights the prior in favour of more widely spread sets of dates. They suggest an alternative prior density,

$$\tilde{p}(\boldsymbol{\alpha}) = \frac{s(\alpha)^{1-J}}{R - s(\alpha)} \tag{2.21}$$

which they believe is a more suitable non-informative prior for $\boldsymbol{\alpha}$, since it gives a uniform marginal prior for the span. One drawback to their non-informative prior is that the marginal prior densities for the α_j's do not have an intuitive archaeological interpretation.

Modelling the deposition process

This section looks at how the authors motivate their alternative prior density, $\tilde{p}(\boldsymbol{\alpha})$, which has the property that the marginal prior density of the span is uniform and is derived from a simple physical model of the deposition process, based on properties of Poisson processes.

- The parameters α_1 and α_{J+1} represent the start (α_1) and the end (α_{J+1}) of the phases of activity.
- The dateable material is assumed to be generated according to a Poisson process P_λ with a piece-wise constant rate $\lambda(t)$, for times t in the interval $[\alpha_{J+1}, \alpha_1]$.

- The parameters $\alpha_2, \ldots, \alpha_J$ mark the change-points in $\lambda(t)$. These are themselves a realization of Poisson process P_Λ of constant intensity in the interval $[\alpha_{J+1}, \alpha_1]$.

- The datable material (generated by P_λ) are randomly thinned (in archaeology this occurs naturally, by samples decaying or not being found in the excavation process) and a Poisson process thinned in this way remains a Poisson process[3].

Now the density $\tilde{p}(\boldsymbol{\alpha}, \boldsymbol{\theta})$ can be expressed as conditional components of the values generated by the above processes

$$\tilde{p}(\boldsymbol{\theta}, \boldsymbol{\alpha}) = \tilde{p}(\boldsymbol{\theta}|\boldsymbol{\alpha})\tilde{p}(\alpha_2, \ldots, \alpha_J|\alpha_1, \alpha_{J+1})\tilde{p}(\alpha_1, \alpha_{J+1}).$$

The authors then condition on n_j and as a result it can easily be shown that the density is uniform over the interval in which they are generated[4], since the events are Poisson.

$$\tilde{p}(\boldsymbol{\theta}|\boldsymbol{\alpha}) = \prod_{j=1}^{J} (\alpha_j - \alpha_{j+1})^{-n_j}.$$

It is then assumed that all change points generated by P_Λ are recorded. As a result of this assumption, and conditioning on the number of events generated by P_Λ, the $\alpha_2, \ldots, \alpha_J$ are again uniform,

$$\tilde{p}(\alpha_2, \ldots, \alpha_J|\alpha_1, \alpha_{J+1}) \propto (\alpha_1 - \alpha_{J+1})^{1-J}.$$

So far, the above process model has been used to determine the prior density for all unknowns except α_1 and α_{J+1}. Nicholls and Jones (2001) do not attempt to model the process which determines the density $\tilde{p}(\alpha_1, \alpha_{J+1})$. Instead they impose a weak bias

[3]Given a Poisson process with rate λ, each occurrence has a constant probability p of being recorded and the recording of an occurrence is independent of that of each other occurrence. Then if $N'(t)$ is the number of occurrences recorded in an interval of length t then $N'(t)$ has a poisson process of rate λp. This particular type of thinning is referred to as geometric thinning.

[4]In the interval $(0, t)$ given that the number of occurrences is $N(t) = n$, then the times of these n occurrences are independent and uniformly distributed in the interval.

towards a shorter interval with

$$\tilde{p}(\alpha_1, \alpha_{J+1}) = 1/R - s(\alpha). \tag{2.22}$$

Another alternative non-informative choice is $\tilde{p}(\alpha_1, \alpha_{J+1}) = 1$, which would lead to a marginal prior span density being proportional to $R - S$. However, the authors slightly favour Equation 2.22, as it gives a uniform prior span, which they believe is a convenient property, hence it is non-informative in respect to the span.

Non-abutting phases

The authors also briefly discuss multiple phase models (Buck *et al.*, 1992), where the phases are not abutting, but may in fact overlap. So from Buck *et al.* (1992), α_j and β_j represent the beginning and ending of the phases j, where $(j = 1, \ldots, J)$.

The main difference with phases that may well overlap are the constraints of the phase boundaries. For example, depending upon the prior constraints, α_1 and β_J may not represent the latest and earliest dates. Hence, the authors suggest calculating the span as

$$s(\alpha, \beta) = \max(\boldsymbol{\alpha}) - \min(\boldsymbol{\beta}).$$

Summary

This subsection offers both a summary and a critical evaluation of Nicholls and Jones (2001). The motivation behind this paper arose as the authors believed that the conventional uniform prior distribution, first discussed in Naylor And Smith (1988), is more informative than first though with respect to intervals or spans of time (*i.e.* $\alpha_j - \alpha_{j+1}$). In particular, it is biased towards longer time spans which is undesirable for at least some real applications. This led Nicholls and Jones to propose an alternative prior distribution, derived from a simple physical model of the deposition process, which they believe leads to a more suitable non-informative prior for $\boldsymbol{\alpha}$, as it gives the property of a uniform marginal prior for the span. However there is a drawback to their choice

of prior. It leads to very informative marginal priors for α_1 and α_{J+1} that do not have intuitive archaeological interpretations.

2.7 Model comparison

In the Bayesian context, the most widespread model choice criterion is the Bayes factor, defined as the ratio of the marginal likelihoods for a pair of models, which represents the evidence provided by the data in favour of a certain model.

To explain the idea behind Bayes factors it is assumed that there are only two models of interest, M_1 and M_2, and interest lies in the relative probabilities of the two models given the data $\boldsymbol{x}$.

The data are assumed to have arisen from Model 1 with a probability density $p(\boldsymbol{x}|M_1)$ or from Model 2 with a probability density $p(\boldsymbol{x}|M_2)$. Given prior probabilities $p(M_1)$ and $p(M_2)$, the data produce posterior probabilities $p(M_1|\boldsymbol{x})$ and $p(M_2|\boldsymbol{x})$. Then using Bayes theorem it can be seen that,

$$p(M_i|\boldsymbol{x}) = \frac{p(\boldsymbol{x}|M_i)p(M_i)}{p(\boldsymbol{x}|M_1)p(M_1) + p(\boldsymbol{x}|M_2)p(M_2)} \tag{2.23}$$

where $i = 1, 2$. The posterior odds in favour of model M_1 over the alternative model M_2 can then be rewritten as,

$$\frac{p(M_1|\boldsymbol{x})}{p(M_2|\boldsymbol{x})} = \frac{p(\boldsymbol{x}|M_1)}{p(\boldsymbol{x}|M_2)} \times \frac{p(M_1)}{p(M_2)}.$$

Posterior odds = Bayes factors $\times$ Prior odds.

However, under the non-informative choice that $p(M_1) = p(M_2)$, hence both models are equally likely, the ratio of the prior odds equal 1, implying that the posterior odds equal the Bayes factor

$$B(1v2) = \frac{p(\boldsymbol{x}|M_1)}{p(\boldsymbol{x}|M_2)}.$$

Essentially the Bayes factor is the ratio of the marginal likelihoods. The densities $p(\boldsymbol{x}|M_i)$ for $i = 1, 2$ are obtained by integrating over the parameter space, so that

$$p(\boldsymbol{x}|M_i) = \int p(\boldsymbol{x}|\boldsymbol{\phi}_{M_i}, M_i)p(\boldsymbol{\phi}_{M_i}|M_i)d\boldsymbol{\phi}_{M_i}.$$

So the Bayes factor is a summary of the evidence provided by the data in favour of one model over another.

Agreement indices

The calibration software OxCal (Ramsey, 2005) offers a tool that it calls the agreement index, A, to allow users to test for outliers or unreliable chronological models, *i.e.* the user community use it as a tool for model comparison. Checks can be made on both individual dated items and on the model as a whole to ensure a reasonable level of consistency between dating evidence and other information. For each dated item the agreement index is calculated by, in Ramsey's notation,

$$A = \frac{\int p(t)p'(t)dt}{\int p(t)p(t)dt}. \tag{2.24}$$

Here $p(t)$ represents the probability distribution before a chronological model has been taken into account. As seen in Section 2.3.1 we refer to this case as the basic model and as a result $p(t)$ is essentially equivalent to the likelihood *i.e.* $p(t) = p(\theta|x, M_0) \approx p(x|\theta)$. The probability distribution $p'(t)$ represents the posterior distribution given a chronological model, therefore we could rewrite $p'(t)$ as $p(\theta|x, M_1)$. By substituting $p(x|\theta)$ for $p(t)$ and $p(\theta|x, M_1)$ for $p'(t)$ in Equation 2.24, we get

$$A = \frac{\int p(x|\theta)p(\theta|x, M_1)d\theta}{\int p(x|\theta)p(x|\theta)d\theta}. \tag{2.25}$$

An overall agreement index, $A_{overall}$, is also calculated for the model as a whole which

is given by

$$A_{overall} = \left[\prod_{i=1}^{n} A_i\right]^{1/\sqrt{n}}. \tag{2.26}$$

Ramsey (2005) refers to the overall agreement index as a pseudo-Bayes-factor, with the exception of the power term, and refers the readers to Gilks *et al.* (1996), Chapter 9 for details. However, after careful manipulation of A, and the use of Gilks *et al.* (1996), it appears that $A_{overall}$ is in fact the ratio of marginal posterior predictive densities (with exception of the power term) rather than a pseudo-Bayes factor; the proof of this is given below.

Using the definition of a marginal posterior predictive density from Gilks *et al.* (1996), page 151, Equation 9.4, which states that

$$p(y_r|Y_{obs}) = \int f(y_r|\theta) f(\theta|Y_{obs}) d\theta. \tag{2.27}$$

We can rewrite Equation 2.27 in terms of the above two models of interest M_0 (the basic model) and M_1 (which represents any realistic chronological model).

For M_0 the marginal posterior predictive densities can be written as follows

$$p(x_i|\boldsymbol{x}, M_0) = \int p(x_i|\theta_i) p(\theta_i|x_i) d\theta_i \tag{2.28}$$

$$= \int p(x_i|\theta_i) p(x_i|\theta_i) p(\theta_i) d\theta_i \tag{2.29}$$

$$= \int p(x_i|\theta_i) p(x_i|\theta_i) d\theta_i. \tag{2.30}$$

Note the form of Equation 2.28 arises as we are only concerned with basic model, *i.e.* each x_i is independent. Also, from Section 2.3.1, we assume a vague prior for θ *i.e* $p(\theta_i) \propto 1$ for $0 < \theta_i$. For this reason Equation 2.29 simplifies to Equation 2.30.

For M_1 we can write the marginal posterior predictive density as

$$p(x_i|\boldsymbol{x}, M_1) = \int p(x_i|\theta_i) p(\theta_i|\boldsymbol{x}, M_1) d\theta_i. \tag{2.31}$$

If we now write A_i in terms of the marginal posterior predictive density, given in Equation 2.30 and Equation 2.31, we get

$$A_i = \frac{\int p(x_i|\theta_i)p(\theta_i|\boldsymbol{x}, M_1)d\theta_i}{\int p(x_i|\theta_i)p(x_i|\theta_i)d\theta_i}. \tag{2.32}$$

Clearly, Equation 2.32 takes the same form as Equation 2.25, hence $A_{overall}$ is in fact the ratio of the posterior predictive densities.

When interested in learning about chronological models through the use of radiocarbon data, model comparison becomes an important issue. OxCal is currently the only calibration software which offers the users some form of model comparison. Although model comparison is not an area of research that we choose to tackle within the thesis, a more detailed discussion on ideas for alternative methods can be found in Section 8.2.1.

2.8 Implementation of methods discussed within the chapter

Many of the techniques for interpreting radiocarbon determinations (described in this chapter) are not readily available to the archaeological community because they have not been implemented in suitable software. However, there are exceptions, such as BCal (http://bcal.shef.ac.uk), OxCal (http://www.rlaha.ox.ac.uk/orau/oxcal.html) and CALIB (http://radiocarbon.pa.qub.ac.uk), which were written for this purpose.

There are other software packages such as WinBugs, for constructing Bayesian statistical models using Markov chain Monte Carlo methods (see Section 3.2.2), which can be used for implementing many of the Bayesian methods for interpreting radiocarbon data. However, WinBugs is only really suitable for use by those who have some understanding of the mathematics behind the modelling and some knowledge about MCMC, to check that the output is reliable.

Andrew Millard of Durham University has written WinBugs code for the implementation

of archaeological problems, many of which are taken from Buck *et al.* (1996), where he aims at providing a step forward on the kind of problems that can be tackled. The WinBugs code for all examples implemented (which include simple radiocarbon calibration, incorporating stratigraphic information, archaeological phase models assuming a uniform deposition rate, and many more) can be found on his web page: http://www.dur.ac.uk/a.r.millard/.

Chapter 3

Modelling the deposition process

3.1 Introduction

This chapter will initially discuss a number of methods for Bayesian implementation, and in particular MCMC, before moving on to look at a range of alternative prior distributions for modelling the deposition rate of the dateable material within a phase of archaeological activity, all of which have a meaningful archaeological interpretation. However, two of these alternative priors are believed to have a much wider range of uses and therefore will be discussed in greater detail. This will include their motivation, their parametrization and the methods used for their implementation.

3.2 Methods for general Bayesian inference

This section of the chapter is intended to give a background to some of the statistical methods used in Sections 2.3 – 2.6 and is also intended to discuss the methods adopted throughout the rest of the thesis.

The aim in Bayesian statistics is to devise a suitable statistical model $p(\boldsymbol{x}|\boldsymbol{\phi})$, formulate the prior knowledge $p(\boldsymbol{\phi})$ and perform the necessary calculations to summarize the

posterior distribution, $p(\boldsymbol{\phi}|\boldsymbol{x})$, as given in Equation 2.1. In Bayesian statistics we are often faced with the problem that we cannot derive the posterior distribution analytically, as the evaluation of the normalizing constant, k, is often difficult.

There are situations in which the posterior can be derived analytically, *e.g.* when using a conjugate prior[1]. However in practice the prior distribution must reflect accurately the available prior information, which may lead to complex modelling, in which case the use of conjugate priors is not applicable. Other methods, such as numerical integration, can be used to approximate the integral k. For completeness, the following sub-section discusses the concept of numerical integration, however, those interested only in the techniques used within the thesis may skip to Section 3.2.2.

3.2.1 Numerical integration

In Bayesian inference numerical integration, also referred to as quadrature, can be used to evaluate the normalizing constant, k, when analytical solutions fail. Consider a general one-dimensional problem, in which we want to approximate the integral

$$I = \int_a^b f(x)dx. \tag{3.1}$$

The integral I is approximated by evaluating f at a number of points $x_1, x_2, \ldots, x_n$. The simplest solution is given by the weighted average

$$\hat{I} = \sum_{i=1}^{n} w_i f(x_i)$$

where w_i (for $i = 1, \ldots, n$) are known as the weights.

Different quadrature methods are characterized by using different points of evaluation $x_1, x_2, \ldots, x_n \in [a, b]$ and/or different corresponding weights, $w_1, w_2, \ldots, w_n$.

[1]Choose a prior with a suitable form so the posterior belongs to the same functional family as the prior. The choice of the functional family depend upon the likelihood and choosing a prior in this way is said to be conjugate. For example, given a normal likelihood and choosing a normal prior, the posterior is still normal.

Classical Newton-Cotes formulae

There are two distinct approaches to numerical integration. The first method, referred to as the Newton-Cotes formula, is when $f(x)$ is evaluated at regularly spaced points. The simplest of all quadrature rules is the *midpoint rule.* This rule divides the interval [a,b] into n subintervals, the function $f(x)$ is then evaluated at the midpoint of each subinterval and equal weights are then applied. In this case

$$\hat{I}_{MP} = h\sum_{i=1}^{n} f(a + (2i-1)h/2))$$

where $h = (b-a)/n$. This method basically approximates the integral I by summing the areas of rectangles (the rectangles have an equal base $(b-a)/n$).

There are slight variations within the Newton-Cotes formulae, such as the *trapezoidal rule*, this method uses unit weights except at the extremes of the interval. The trapezoidal rule gives the approximation

$$\hat{I}_T = h\left[\frac{f(a)}{2} + \sum_{i=1}^{n} f(a + (2i-1)h/2)) + \frac{f(b)}{2}\right].$$

Another variation is given by the *Simpson's rule.* In this case weights alternating between 4/3 and 2/3 are used except in the extremes of the interval where a value of 1/3 is used. In this case the integral is approximated by

$$\hat{I}_S = \frac{h}{3}\left[f(a) + 4\sum_{i=1}^{n/2} f(a + (4i+1)h/2)) + 2\sum_{i=1}^{n/2} f(a + (4i+3)h/2)) + f(b)\right].$$

Gaussian quadrature

The second approach to numerical integration is the idea of Gaussian quadrature, where the evaluation points are no longer restricted to be equally spaced, and that they can be chosen to give higher accuracy. Gaussian quadrature is constructed to yield exact results for polynomials of order $2n-1$ (or less), by a suitable choice of evaluation points and weights.

There are a number of Gaussian quadrature methods, which evaluate $f(x)$ over a finite or infinite range. The problem is to evaluate

$$I = \int_a^b w(x)f(x)dx. \tag{3.2}$$

Depending upon the choice of a, b and w, will result in different Gaussian quadrature rules, some of the most common rules are given below in Table 3.1.

Interval	$w(x)$	Integration rules
$[-1, 1]$	1	Gauss-Legendre quadrature
$[-1, 1]$	$1/(\sqrt{1-x^2})$	Gauss-Chebshev quadrature
$[0, \infty)$	e^{-x}	Gauss-Laguerre quadrature
$(-\infty, \infty)$	e^{-x^2}	Gauss-Hermite quadrature

Table 3.1: Rules of Gaussian quadrature

Naylor and Smith (1988), one of the first to model radiocarbon determinations in a Bayesian framework, used numerical integration techniques to summarize the posterior distributions of interest. The particular method they used was Gauss-Hermite quadrature, for further details of this methods see O'Hagan (1994).

The one-dimensional quadrature rules discussed above can be directly generalized to higher dimensions, see O'Hagan (1994). However, one drawback to methods of numerical integration is that, for problems involving a large numbers of dimensions (*e.g.* a Bayesian inference problem with a large number of parameters), using such approaches becomes computationally intense.

3.2.2 Simulation methods

As seen in Equation 2.2 we can express the posterior distribution up to a constant of proportionality. As a result, we can generate random samples from the distribution of interest, which in this case is the posterior distribution. In general the dimensionality of ϕ will be too high to use methods such as rejection sampling. However, simulation methods based on Markov chains are available, these methods are better known as Markov chain

Monte Carlo (MCMC) methods.

Markov chains

Markov chains are sequences of random variables $X_1, X_2, \ldots$ such that, for $t \geq 0$, $P(X_{t+1}|X_1, \ldots, X_t) = P(X_{t+1}|X_t)$, hence X_{t+1} only depends upon the previous state X_t. $P(.|.)$ can be referred to as the *transition kernel* and describes how we move from X_t to X_{t+1}.

Let $P^t(X_t|X_0)$ denote the distribution of X_t, where X_0 represents the starting state of the Markov chain. It can be shown that as $t \to \infty$, $P^t(X_t|X_0)$ will converge to the *stationary distribution*, $\psi(X)$, which does not depend upon t or X_0, given that the Markov chain is irreducible, aperiodic and positive recurrent. If the Markov chain has reached equilibrium by time T, then we can say that $X_{T+1}, \ldots, X_{T+n}$ is a sample from the density function $\psi(X)$.

Markov chain Monte Carlo methods

Markov chain Monte Carlo (MCMC) methods essentially construct a Markov chain for the parameters $\boldsymbol{\phi}$ whose stationary distribution, $\psi(\boldsymbol{\phi})$, is equal to the posterior distribution $p(\boldsymbol{\phi}|\boldsymbol{x})$. MCMC is now one of the most popular approaches when dealing with complicated models in which it is rare that samples from the posterior distribution can be obtained directly. There are two main methods used. The first is the Gibbs sampler which is used when it is possible to sample from each of the 1-dimensional full conditional distributions, $p(\phi_i|\boldsymbol{x}, \phi_1, \ldots, \phi_{i-1}, \phi_{i+1}, \ldots, \phi_n)$. The second method is the Metropolis-Hastings algorithm, this is used when it is not possible to sample from the conditional distributions of interest.

Gibbs sampler

The Gibbs sampler is a Markov chain algorithm that is particularly useful in high-dimensional problems, when it is possible to sample from each of the one dimensional conditional distributions. Suppose that there are k parameters $\phi_1, \ldots, \phi_k$ of interest,

denoted by $\boldsymbol{\phi}$, and we wish to make inferences about their joint posterior distribution, $p(\boldsymbol{\phi}|\boldsymbol{x})$, as well as their marginal posterior distributions $p(\phi_i|\boldsymbol{x})$. Then the Gibbs sampler can be used to sample from the conditional distributions in the following way.

1. Choose arbitrary starting values $\boldsymbol{\phi^{(0)}} = (\phi_1^{(0)}, \phi_2^{(0)}, \ldots, \phi_n^{(0)})$

2. Generate a series of random values $\boldsymbol{\phi^{(1)}}$, $\boldsymbol{\phi^{(2)}} \ldots \boldsymbol{\phi^{(t)}}$ in the following way:

   - draw $\phi_1^{(1)}$ from $p(\phi_1|\boldsymbol{x}, \phi_2^{(0)}, \ldots, \phi_k^{(0)})$,
   - draw $\phi_2^{(1)}$ from $p(\phi_2|\boldsymbol{x}, \phi_1^{(1)}, \phi_3^{(0)}, \ldots, \phi_k^{(0)})$,
   - draw $\phi_3^{(1)}$ from $p(\phi_3|\boldsymbol{x}, \phi_1^{(1)}, \phi_2^{(1)}, \phi_4^{(0)}, \ldots, \phi_k^{(0)})$,
   -
   -
   - draw $\phi_k^{(1)}$ from $p(\phi_3|\boldsymbol{x}, \phi_1^{(1)}, \phi_2^{(1)}, \ldots, \phi_{k-1}^{(1)})$.

   This completes one iteration of the algorithm.

3. Repeat step 2 for t iterations.

The idea behind the Gibbs sampler is to draw samples from the posterior distributions $p(\phi_i|\boldsymbol{x})$ using Markov chains which have the stationary distribution, $\psi(\phi_i) = p(\phi_i|\boldsymbol{x})$. In particular, as $t \rightarrow \infty$, $\phi_i^{(t)}$ tends to a random quantity whose density is $p(\phi_i|\boldsymbol{x})$. Thus for large t, the values $(\phi_1^{(t)}, \ldots, \phi_k^{(t)})$ are approximately a random sample from $p(\boldsymbol{\phi}|\boldsymbol{x})$. There are two different methods of the Gibbs sampler, the 'deterministic-scan Gibbs sampler', which updates ϕ_i, in order. The second type is the 'random-scan Gibbs sampler', where the ϕ_i to be updated is chosen randomly.

The Gibbs sampler was first introduced into archaeological problems by Buck *et al.* (1992), see Section 2.5, and is used in both OxCal and BCal to implement chronological models.

Metropolis–Hastings algorithm

The Metropolis–Hastings algorithm is a general term for a family of Markov chain simulation methods, used to draw samples from the posterior distribution, when it is not possible to sample from the full conditional distributions of the parameters. The Metropolis–Hastings algorithm was described by Hastings (1970), generalizing the algorithm of Metropolis *et al.* (1953). The Gibbs sampler, as discussed above, can be viewed as a special case of Metropolis Hastings. This section will present the Metropolis–Hastings algorithm and discuss several implementation issues.

Suppose that $\psi(\boldsymbol{\phi})$ is the density of interest, hence the stationary distribution of the Markov chain. Suppose further that we have some (arbitrary) proposal distribution $q(\boldsymbol{\phi}'|\boldsymbol{\phi}^t)$ which is easy to simulate from, but does not necessarily define a Markov chain having $\psi(\boldsymbol{\phi})$ as its stationary distribution. Consider the following algorithm:

1. Generate *proposed* values $\boldsymbol{\phi}'$ using the proposal distribution $q(\boldsymbol{\phi}'|\boldsymbol{\phi}^t)$
2. Evaluate the acceptance probability $p(\boldsymbol{\phi}', \boldsymbol{\phi}^t)$ of the proposed move, where

$$p(\boldsymbol{\phi}', \boldsymbol{\phi}^t) = \min\left[1, \frac{p(\boldsymbol{\phi}'|\boldsymbol{x})q(\boldsymbol{\phi}^t|\boldsymbol{\phi}')}{p(\boldsymbol{\phi}^t|\boldsymbol{x})q(\boldsymbol{\phi}'|\boldsymbol{\phi}^t)}\right] \tag{3.3}$$

3. Put $\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}'$ with probability $p(\boldsymbol{\phi}', \boldsymbol{\phi}^t)$, and put $\boldsymbol{\phi}^{t+1} = \boldsymbol{\phi}^t$ otherwise.

In other words, at each stage, a new value is generated from the proposal distribution. This is either accepted, in which case the chain moves, or rejected, in which case the chain stays where it is. Whether the move is accepted or rejected depends on an acceptance probability which itself depends on the relationship between the density of interest and the proposal distribution.

3.2.3 Practical considerations in MCMC

There are several issues which arise when implementing MCMC methods. These include the choice of the proposal distribution and its corresponding standard deviation, choosing

suitable starting values for parameters ϕ and also how to judge when the Markov chain has reached equilibrium. Each of these issues will be discussed in turn below.

Choice of proposal distribution

The first consideration to take account of is the choice of proposal distribution, $q(\phi'|\phi^t)$.

- The Metropolis Algorithm considers only symmetric proposal distributions, having the form $q(\phi^t|\phi') = q(\phi'|\phi^t)\ \forall\ \phi^t$ and ϕ'. Here the acceptance probability simplifies to

$$p(\phi', \phi^t) = \min\left[1, \frac{p(\phi'|\boldsymbol{x})}{p(\phi^t|\boldsymbol{x})}\right] \tag{3.4}$$

  A special case of the Metropolis algorithm is the random-walk Metropolis, in this case the proposed value ϕ' at each stage is the parameter value from the previous iteration adjusted by adding a displacement, from some symmetric distribution. For example $q(\phi'|\phi^t) \sim U(\phi^t - 1, \phi^t + 1)$.

- The independence sampler is the Metropolis-Hastings algorithm where $q(\phi', \phi^t) = q(\phi')$, does not depend on ϕ^t. The proposal distribution needs to be a good approximation of (and heavier tailed than) the stationary distribution, for this method to work well.

- Instead of updating the whole of ϕ=$(\phi_1, \ldots, \phi_n)$, it is often more convenient and computationally efficient to update components one by one and this method is known as the single component Metropolis-Hastings algorithm. Gibbs Sampling is a special case of the single component Metropolis -Hastings algorithm in which the proposal distribution for each component is its full conditional distribution. In this case the acceptance probability is always 1, hence a proposed value is always accepted.

Care is also need when choosing values for the corresponding standard deviations of the proposal distributions. One method is to monitor the acceptance probability and then it is possible to adjust the standard deviations accordingly. A cautious proposal

distribution, generating small steps, is the consequence of a proposal distribution with a too low standard deviation and will generally have a high acceptance rate, but will nevertheless move slowly around the parameter space. A bold proposal distribution generating large steps will often propose moves from the body to the tail of the posterior distribution, as a consequence of a too large standard deviation. Such a chain will frequently not move, giving a low acceptance rate and resulting in slow exploration of the posterior distribution. Ideally the overall proportion of accepted moves should be around 25% and it is therefore possible to experiment with values of the standard deviation to get an overall acceptance rate around this level.

Choice of starting values

In theory, since we are only interested in values once the Markov chain has reached equilibrium, the choice of starting values should be irrelevant. In practise, it is important to choose starting values for the parameters, ϕ, carefully as a poor choice of starting value may result in the Markov chain taking longer to converge. In addition, the choice of starting values can help check the behaviour of the algorithm is correct. One way to choose starting values is to experiment with a number of different starting values and see if they converge to similar distributions, as they should.

Convergence of the Markov chain

One of the most difficult assessments to make regarding MCMC output, is how to identify when a Markov chain has reached equilibrium. MCMC methods can vary considerably, sometimes they can be quite slow to converge, requiring long runs, while other times runs of a much shorter length are adequate. So in order to check the stationarity of an MCMC chain, convergence diagnostics are commonly used.

Bayesian Output Analysis (BOA), available from www.public-health.uiowa.edu/boa (Smith, 2005), offers four of the most commonly used methods to check convergence of MCMC output, these being Brooks, Gelman & Rubin, Geweke, Heidelberg & Welch and Rafferty & Lewis, as well a visual methods such as time series plot of parameter

values against iteration numbers, and autocorrelation plots. When checking convergence of MCMC output it is suggested that no one method should be thought of as superior, but a combination of diagnostics are used rather than any one single diagnostic. During the course of this project the two diagnostics Geweke and Heidelberg & Welch were commonly used as well as a number of visual methods. The section below will outline these two diagnostics in more detail.

Geweke

This diagnostic was devised by Geweke in 1992 and requires a single MCMC chain. The method is based on a standard time series method and for each parameter the chain is divided into two windows, one containing the first $x\%$ and the other containing $y\%$ of the iterations. In both windows the sample mean and the asymptotic variance are calculated. A Z-statistic is then produced, by calculating the difference between the two means divided by the asymptotic standard error of their difference. As the number of iterations increases the distribution of the Z-statistic approaches the standard normal. Therefore, large values of Z *i.e.* which fall in the extreme tails of $N(0,1)$ suggest that the chain has not fully converged and a longer run is needed.

Heidelberger & Welch

This convergence diagnostic is a two-stage test for a single chain (devised in 1983), the first being the 'stationarity test' and the second called the 'interval half-width test'. The first test is based on the Cramer-von-Mises statistic, to test the null hypothesis that the sampled values come from a stationary distribution. Initially the test is applied to the whole chain, if the null hypothesis is rejected, then the test is repeated with the first 10% of the iterations discarded. This is repeated until either more than 50% of the chain has been discarded or the test has passed. If the test fails then a longer MCMC run is needed. However, if the stationarity test is passed then the portion of iterations that passed the test are subject to the half-width test. This test calculates the standard error of the mean for the portion of iterations that passed the stationarity test and the half-width of the associated 95% confidence interval for the mean (*i.e.* 1.96×standard

error). If the half-width is less than ϵ times the sample mean of the retained iterations then the half-width test is passed. If the half-width test fails this suggests that there is evidence against convergence and a longer chain is needed.

While the above two convergence diagnostics are routinely used, visual methods are also very useful. For example a calibrated date, as seen in Figure 2.1, is typically multi-modal and non-symmetric. Depending upon the part of the calibration curve under consideration, the calibrated date might be clearly bimodal but the chain may become stuck in one of the modes. In this case, although convergence diagnostics may indicate that the chain has converged, the chain is not fully representing the posterior distribution of interest. When this occurs, one solution may be to run multiple chains each with diverse starting values and then compare within and between chain properties.

3.3 Devising alternative deposition models

This section of the chapter is concerned with the modelling of the deposition rate of the datable material within a phase of archaeological activity. As previously discussed, the current convention is to assume that the material suitable for dating was deposited at a uniform rate (see Section 2.3.3) between the start and end dates of unknown calendar age. Initially it was assumed that such models constituted vague priors and that they were suitable for use in a wide range of applications. This has proven to be the case for a great number of real problems and may also be the simplest way to represent prior ignorance, when little is known *a priori* about the rate of deposition.

However, there have been an increasing number of cases in the applied literature (Housley *et al.*, 1997 and Van Strydonck *et al.*, 2004) in which the authors discuss the rate at which material is deposited or the rate at which material is manufactured within a period of time. In these cases the authors believe that the rate of deposition or manufacture is not uniform over the proposed range. However, the majority of archeologists have very little mathematical background and they rely wholly on the software available to them

to interpret their data, often resulting in them contradicting their own prior beliefs. So the aim here is to seek alternatives to the conventional uniform deposition model and thus build a more flexible range of prior distributions, that reliably represent the archaeologist's *a priori* information.

It might be assumed that the first obvious step to progress from the uniform prior for the rate of deposition, would be a triangular prior. Clearly a triangular prior could be defined in a number of different ways, see Figures 3.1(b) & (c). These priors represent situations where there has been no period of established activity. For example, Figure 3.1(b) could be useful in situations where we know a process slowly started but suddenly came to an end; this might represent a settlement building phase which was ended abruptly by fire or during a battle. However, triangular priors are not perceived to be particularly useful in many other situations.

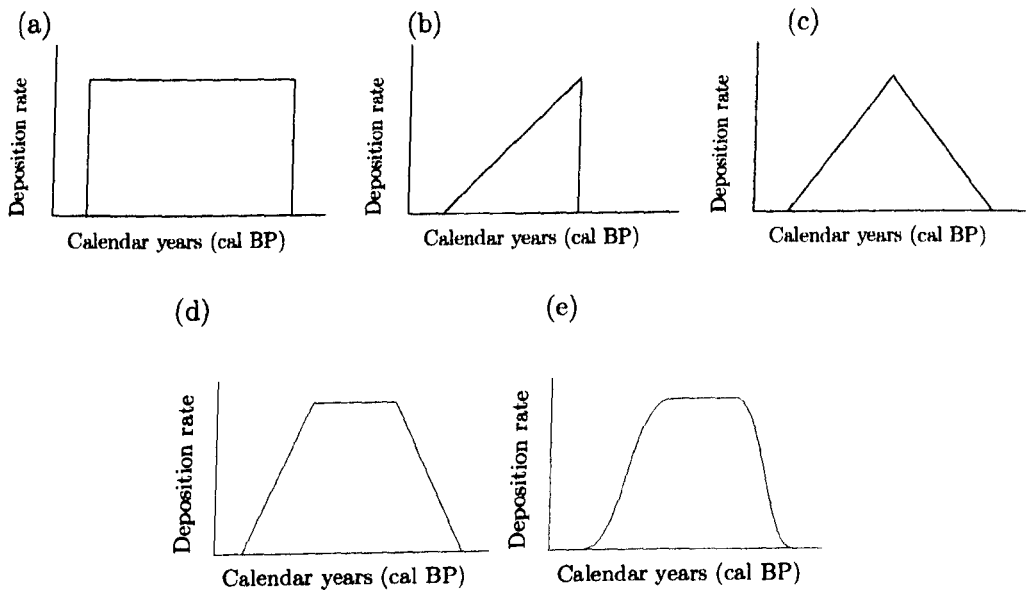


Figure 3.1: Schematic representations of prior deposition models (a) conventional uniform (b) right-angled triangle (c) general triangle (d) trapezium and (e) sigmoidal.

Figures 3.1(d) & (e) however, which we refer to as the trapezium and sigmoidal priors, their names reflecting their shapes. In both cases it is believed that they can be seen as formalizations of a well accepted archaeological model, for the use of sites or whole

regions and also for the development of technologies and fashions. The reason is, that rather than there being a sudden increase from zero to the maximum rate of deposition, as in the uniform case (see Figure 3.1(a)), there would be a gradual increase, followed by a period of constant deposition, followed by a gradual decrease in the deposition rate. For this reason it seems likely that both prior distributions will have a range of general uses and reflect more accurately the uncertainties in such processes *i.e.* sites or landscapes being established over a finite period of time, rather than instantaneously.

Note that the shape of Figure 3.1(d) & (e) need not be symmetric. Also, Figures 3.1(a) – (c) can all be thought of as special cases of the trapezium prior (this will be discussed in more detail in the next section).

So far throughout this thesis we have talked about modelling the rate of deposition of datable material within an archaeological phase. As seen from Figure 3.1, a number of alternative deposition rates are plausible and we can think of these plots as possible intensity functions. However, to use such intensity functions as prior distributions it must initially be assumed that an individual sample is randomly chosen from the material suitable for dating. Then the intensity function needs to be normalized to enable us to use it as a probability distribution.

There are a number of case studies from the applied literature which motivate the use of non-uniform priors. Two in particular will be used in Chapter 4 to illustrate the difference in inferences when adopting the conventional uniform prior and alternative non-uniform priors, these being Van Strydonck *et al.* (2004) and Housley *et al.* (1997). In both papers the authors are interested in using radiocarbon dating as a means of answering chronological questions posed. Both authors adopt *ad hoc* statistical methods which do not really answer the questions posed and are not statistically sound (full details of their methodology can be found in Sections 4.2 & 4.3). However, the authors do state their *a priori* beliefs about the rate at which the material dated was deposited/manufactured within the phase of interest. For example, in Van Strydonck *et al.* (2004) the authors state "... the manufacturing dates of related textiles are not uniformly distributed over the

proposed range, but there exists an introduction phase, a blooming period and a period of decline". From this statement it is clear that there is a need to develop the existing models further to allow for a more robust and realistic modelling of the deposition rate.

This subsection is concerned with modelling the deposition rate of datable material and in particular devising a range of alternative non-uniform a priori distributions. In the majority of archaeological calibration problems there will typically be some form of *a priori* knowledge. Whether this relates to the time period of interest or the rate at which the material is deposited/manufactured within an archaeological phase.

We did not elicit *a priori* information regarding the rate of deposition/manufacture within a period of time from experts directly (e.g. archaeologists) as many have very little mathematical/statistical background which may make it difficult for them to translate their beliefs into a suitable/realistic statistical distribution. Although, this is not always the case as Alex Bayliss of English Heritage was very keen to be involved in the development of the alternative distributions and on many occasions we discussed typical manufacture/deposition rates for a range of archaeological calibration problems.

The main approach we adopted was to devise alternative distributions based on a priori knowledge stated in cases studies found in the applied literature. In many cases the authors would state their a priori beliefs about the rate of deposition/manufacture within a period of time as being non-uniform, see Sections 4.2.3 and 4.3.2.

3.4 Set-up of the trapezium and sigmoidal priors for the deposition rate of datable material within an archaeological phase

In the sections to follow, the modelling of both the trapezium and sigmoidal priors will be discussed. Initially, we describe the parametrization of the two prior distributions before going on to talk about the set-up of the models for a single phase of activity as

well as multiple phases of activity.

3.4.1 Trapezium prior

The first alternative prior distribution considered is the trapezium prior distribution as seen in Figure 3.2. The same parameters, α and β, are used to represent the start and end dates (cal BP) of the phase of activity, to be consistent with the conventional parameterization of the uniform prior. Now we also have two extra parameters in the case of the trapezium prior. Firstly, γ which represents the beginning of the period of constant deposition and, secondly, δ which represents the end of the period of constant deposition (see Figure 3.2). The trapezium can be explicitly defined as

$$X \sim \text{Trap}(\alpha, \gamma, \delta, \beta) \iff p(x|\alpha, \gamma, \delta, \beta) = \begin{cases} 0 & \text{for } x < \beta \\ h(x-\beta)/(\delta-\beta) & \text{for } \beta < x < \delta \\ h & \text{for } \delta < x < \gamma \\ h(\alpha-x)/(\alpha-\gamma) & \text{for } \gamma < x < \alpha \\ 0 & \text{for } x > \alpha \end{cases} \quad (3.5)$$

where h is a constant, whose value is determined by the values of α, γ, δ and β.

Notice that the trapezium prior is a generalization of uniform prior, in the sense that the uniform is a special case of the trapezium when $\alpha = \gamma$ and $\beta = \delta$. It is also seen that if $\gamma = \delta$ then this represents Figure 3.1(c) and if $\gamma = \delta = \beta$ this represents Figure 3.1(b).

3.4.2 Sigmoidal prior

The second alternative prior implemented is the sigmoidal prior suggested by Blackwell and Buck (2003), illustrated in Figure 3.3. The sigmoidal prior again has four parameters. Again, to be consistent, α and β are used to represent the start and end dates (cal BP) of the phase of activity and γ and δ represent the internal parameters.

This prior is similar to the trapezium prior, but the main difference lies in the tails of the

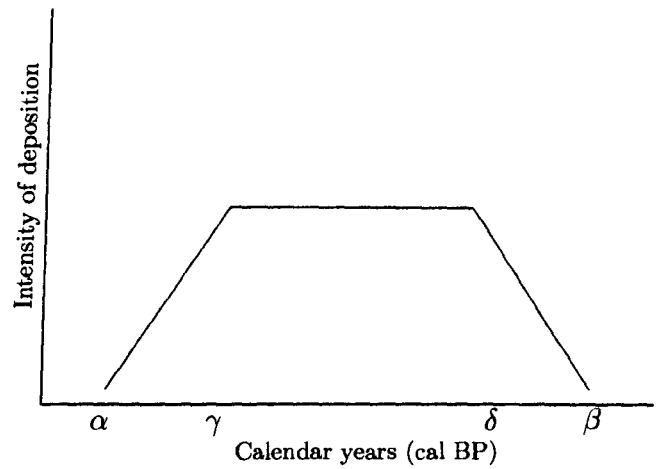


Figure 3.2: Schematic representation of trapezium prior for the deposition rate.

distribution. Hence the constant part on (δ, γ) is the same, but the linear interpolation in (γ, α) and (β, δ) in the trapezium prior is replaced by something much smoother, *e.g.* some sort of sigmoid, such that

$$X \sim \text{Sig}(\alpha, \gamma, \delta, \beta) \iff p(x|\alpha, \gamma, \delta, \beta) = \begin{cases} 0 & \text{for } x < \beta \\ hg((x-\beta)/(\delta-\beta)) & \text{for } \beta < x < \delta \\ h & \text{for } \delta < x < \gamma \\ hg((\alpha-x)/(\alpha-\gamma)) & \text{for } \gamma < x < \alpha \\ 0 & \text{for } x > \alpha \end{cases} \quad (3.6)$$

where $g(.) : [0,1] \to [0,1]$ is some known monotonic function. The monotonic function $g(.)$ used in the implementation here is defined as

$$g(x) = \frac{x^2}{x^2 + (1-x)^2}.$$

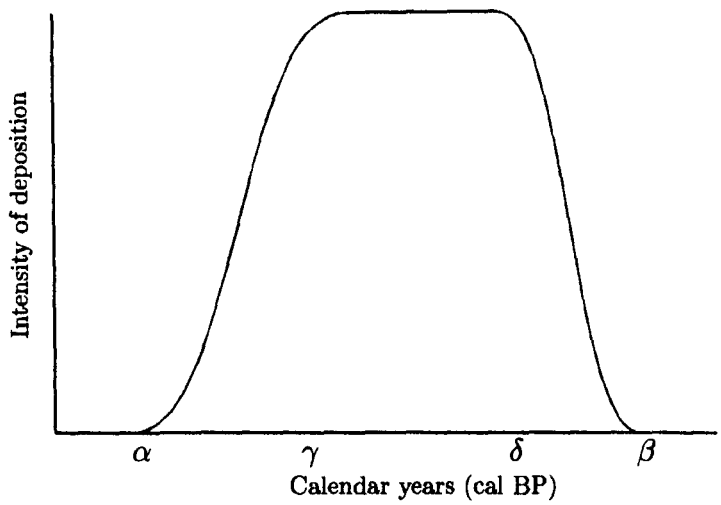


Figure 3.3: Schematic representation of the sigmoidal prior for the deposition rate as suggested in Blackwell and Buck (2003).

3.4.3 Single phase of activity

When modelling a single phase of activity, with respect to the trapezium or sigmoidal prior, interest lies in estimating the calendar dates of the phase boundaries, α, γ, δ and β. Within the phase, there are n samples suitable for radiocarbon dating. For $i = 1, \ldots, n$ $x_i \pm \sigma_i$ represents the ith radiocarbon determination each associated with a true unknown calendar date θ_i. Using Equation 2.17, the likelihood can be written as

$$p(x_i|\theta_i) \propto \frac{1}{w_i(\theta_i)} \exp\left\{-\frac{(x_i - \mu(\theta_i))^2}{2w_i^2(\theta_i)}\right\} \tag{3.7}$$

where $w_i^2(\theta_i) = \sigma_i^2 + \sigma_i^2(\theta_i)$. This follows the suggestion in Christen (1994a) *i.e.* to take account of the uncertainties in the calibration data.

In the absence of any prior knowledge, it is assumed that α, γ, δ and β lie anywhere in the finite interval (P, A) of length R, where the P stands for *terminus post quem* ("date after which") and A stands for *terminus ante quem* ("date before which"). Note that, as the dates are measured in years BP (before present) they are subject to the constraint that $A > \alpha \geq \gamma \geq \delta \geq \beta > P$. A particularly simple form of joint prior, $p(\alpha, \gamma, \delta, \beta)$,

is used to represent a minimal state of prior knowledge, which can be represented as follows, as

$$p(\alpha, \gamma, \delta, \beta) \propto \mathrm{I}_C(\alpha, \gamma, \delta, \beta)$$

where

$$\mathrm{I}_C(\alpha, \gamma, \delta, \beta) = \begin{cases} 1 & \text{if } (\alpha, \gamma, \delta, \beta) \epsilon\, C \\ 0 & \text{otherwise} \end{cases}$$

and C is the set of α, γ, δ and β that satisfy the above constraints *i.e.* $A > \alpha \geq \gamma \geq \delta \geq \beta > P$.

Our main interest lies in the form of $p(\theta_i|\alpha, \gamma, \delta, \beta)$. This represents the prior knowledge about the rate at which datable material is deposited within an archaeological phase. The conventional method, as discussed in Section 2.3.3, is to assume that

$$\theta_i|\alpha, \beta \sim U(\alpha, \beta) \quad \text{for } i = 1, \ldots, n.$$

However, $p(\theta_i|\alpha, \gamma, \delta, \beta)$ can now take two alternative forms, these are

$$p_{\text{Trap}}(\theta_i|\alpha, \gamma, \delta, \beta) \sim \text{Trap}\ (\alpha, \gamma, \delta, \beta) \tag{3.8}$$

$$p_{\text{Sig}}(\theta_i|\alpha, \gamma, \delta, \beta) \sim \text{Sig}\ (\alpha, \gamma, \delta, \beta) \tag{3.9}$$

as defined in Equations 3.5 and 3.6. As a result, $p(\boldsymbol{\theta}|\alpha, \gamma, \delta, \beta)$ can either be written

$$p(\boldsymbol{\theta}|\alpha, \gamma, \delta, \beta) \propto I_D(\boldsymbol{\theta}) \prod_{i=1}^{n} p_{\text{Trap}}(\theta_i|\alpha, \gamma, \delta, \beta)$$

when implementing the trapezium prior or

$$p(\boldsymbol{\theta}|\alpha, \gamma, \delta, \beta) \propto I_D(\boldsymbol{\theta}) \prod_{i=1}^{n} p_{\text{Sig}}(\theta_i|\alpha, \gamma, \delta, \beta)$$

when implementing the sigmoidal prior where

$$I_D(\boldsymbol{\theta}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta} \in D \\ 0 & \text{otherwise} \end{cases}$$

and D is the set of values of $\boldsymbol{\theta}$ that satisfy some constraints. For example there might be stratigraphic ordering between some of the θ_i's.

Consequently, the joint posterior density of $\boldsymbol{\theta}, \alpha, \gamma, \delta$ and β is given by

$$p(\alpha, \gamma, \delta, \beta, \boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta})p(\boldsymbol{\theta}|\alpha, \gamma, \delta, \beta)p(\alpha, \gamma, \delta, \beta).$$

It is not possible to write down explicitly the conditional distributions of interest, *e.g.* $p(\alpha|\boldsymbol{\theta}, \boldsymbol{x}, \beta)$ due to the form of $p(\boldsymbol{\theta}|\alpha, \gamma, \delta, \beta)$. We are therefore unable to use the Gibbs sampler, consequently we use a Metropolis-Hastings algorithm to evaluate $p(\alpha, \gamma, \delta, \beta, \boldsymbol{\theta})$; details of the algorithm used can be found in Section 3.5.

3.4.4 Multiple phases of activity

This section sets up both the trapezium and sigmoidal prior deposition models within a multiple phase framework. The ideas are very closely linked to those in the previous section.

There are now m phases of activity that have been identified by archaeologists. In this case we let $\alpha_j, \gamma_j, \delta_j$ and β_j represent the four parameters, of the trapezium or sigmoidal prior, for phase j ($j = 1, \ldots, m$). That is to say, α_1 represents the beginning of Phase 1 and β_1 represents the end of Phase 1. As seen in Section 3.4.3, as we are working in calendar years BP, the parameters are subject to the constraint $A > \alpha_j \geq \gamma_j \geq \delta_j \geq \beta_j > P$.

As we are now working with multiple phases, *a priori* information with regard to the phase boundary parameters may arise *e.g.* $\alpha_1 > \alpha_2$, $\beta_1 = \alpha_2$, $\alpha_2 > \beta_1$. Again a particularly simple form of joint prior, $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta})$, is used, which assumes that all

values of $\alpha_j, \gamma_j, \delta_j$ and β_j are equally likely. That is

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \propto \mathrm{I}_C(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta})$$

where

$$\mathrm{I}_C(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \in C \\ 0 & \text{otherwise} \end{cases}$$

where C is the set of values of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ which satisfy some given constraints, such as the ordering of parameters.

Let n_j represent the number of samples assigned to the jth phase. The ith radiocarbon determination in the jth phase is represented by $x_{i,j} \pm \sigma_{i,j}$, associated with $\theta_{i,j}$, the corresponding calendar date (cal BP). By adapting Equation 3.7, the likelihood $p(x_{i,j}|\theta_{i,j})$ can be written as

$$p(x_{i,j}|\theta_{i,j}) \propto \frac{1}{w_{i,j}(\theta_{i,j})} \exp\left\{ -\frac{(x_{i,j} - \mu(\theta_{i,j}))^2}{2w_{i,j}^2(\theta_{i,j})} \right\} \tag{3.10}$$

where $w_{i,j}^2(\theta_{i,j}) = \sigma_{i,j}^2 + \sigma_{i,j}^2(\theta_{i,j})$.

Again our main interest lies in the form of $p(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j)$. This represents the prior knowledge about the rate at which datable material is deposited within the jth phase. The conventional method, as discussed in Section 2.3.3, is to assume that

$$\theta_{i,j}|\alpha_j, \beta_j \sim U(\alpha_j, \beta_j) \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, 8.$$

However, $p(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j)$ can now take two alternative forms, these are

$$p_{\text{Trap}}(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j) \sim \text{Trap}\,(\alpha_j, \gamma_j, \delta_j, \beta_j) \tag{3.11}$$

$$p_{\text{Sig}}(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j) \sim \text{Sig}(\alpha_j, \gamma_j, \delta_j, \beta_j) \tag{3.12}$$

as defined in Equations 3.5 and 3.6. As a result, $p(\boldsymbol{\theta_j}|\alpha, \gamma, \delta, \beta)$ [note $\boldsymbol{\theta_j}$ represents the

set of θ's belonging to the jth phase] can either be written

$$p(\boldsymbol{\theta_j}|\alpha_j, \gamma_j, \delta_j, \beta_j) \propto I_{D_j}(\boldsymbol{\theta_j}) \prod_{i=1}^{n_j} p_{\text{Trap}}(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j)$$

when implementing the trapezium prior or

$$p(\boldsymbol{\theta_j}|\alpha_j, \gamma_j, \delta_j, \beta_j) \propto I_{D_j}(\boldsymbol{\theta_j}) \prod_{i=1}^{n_j} p_{\text{Sig}}(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j)$$

when implementing the sigmoidal prior where

$$\mathrm{I}_{D_j}(\boldsymbol{\theta_j}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta_j} \in D_j \\ 0 & \text{otherwise} \end{cases}$$

but now D_j is used to represent the set of values that $\boldsymbol{\theta_j}$ can take within phase j.

So if we wish to estimate the calendar dates of the phase boundaries, $\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}$ and $\boldsymbol{\theta}$, *e.g.* $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, the joint posterior distribution can be written as

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}, \boldsymbol{\theta}|\boldsymbol{x}) \propto p(\boldsymbol{x}|\boldsymbol{\theta}) p(\boldsymbol{\theta}|\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}).$$

The remainder of this chapter will concentrate on the methods used to implement both the trapezium and the sigmoidal prior deposition models, as well as discussing the problems encountered along the way.

3.5 Implementing the trapezium and sigmoidal prior distributions using a Metropolis–Hastings algorithm

As discussed in Section 3.4.3 it is not possible to sample from the conditional distributions of the parameters, $\boldsymbol{\theta}, \alpha, \gamma, \delta$ and β. As a result a Metropolis-Hastings algorithm, as described in Section 3.2.2, has been implemented in the programming language C, in order to sample values from the posterior distribution $p(\boldsymbol{\theta}, \alpha, \gamma, \delta, \beta|x)$. Only details of

the algorithm used for a single phases of activity are given here, but using the extensions in the previous section the algorithm can easily be generalized to multiple phases. The particular type of algorithm used is a single-component Metropolis–Hastings algorithm.

Let $\boldsymbol{\phi} = \{\alpha, \gamma, \delta, \beta, \theta_1, \ldots, \theta_n\}$. Instead of updating the whole of $\boldsymbol{\phi}$ at once it is more convenient to update the parameters $\boldsymbol{\theta}, \alpha, \gamma, \delta$ and β separately. As a result, an iteration of the single component Metropolis-Hastings comprises of h updating steps (where $h = n + 4$ and n represents the number of radiocarbon determinations). Although it is not necessary, a fixed updating order for the parameters is assumed. The parameter β is updated first, followed by δ, γ, α and then $\theta_1, \ldots, \theta_n$. [Note, as a result of using a single-component algorithm, the ith update of $\boldsymbol{\phi}$ at iteration t may depend on the current values of any of the components of $\boldsymbol{\phi}$ *i.e.* $\boldsymbol{\phi}_{(i)}^t = \{\phi_1^t, \ldots \phi_{i-1}^t, \phi_{i+1}^{t-1}, \ldots, \phi_h^{t-1}\}$, where $\boldsymbol{\phi}_{(i)}$ denotes all $\boldsymbol{\phi}$ except the ith element].

Updating β, δ, γ and α

Although a single-component algorithm is used which means that each of the four parameters β, δ, γ and α are updated separately, each is updated in the same way. For this reason only the steps for updating β will be outlined below.

At each iteration t, the next state β^{t+1} is chosen by sampling a candidate point β' from a proposal distribution $q(\beta'|\beta^t)$. A convenient choice of proposal distribution is a truncated Normal distribution, as β is constrained to lie between P and δ, see Section 3.4.3.

Truncated Normal distribution

A truncated Normal distribution is a Normal distribution that is restricted to lie within a finite (or semi infinite) range, by truncating the tails of the distribution. The truncated Normal distribution is expressed in terms of the Normal distribution as follows

$$\phi_{L,U}(x|\mu,\sigma) = \begin{cases} \frac{1}{\sigma}\left[\Phi\left(\frac{U-\mu}{\sigma}\right) - \Phi\left(\frac{L-\mu}{\sigma}\right)\right]^{-1} \phi\left[\frac{x-\mu}{\sigma}\right] & L \leq x \leq U \\ 0 & \text{otherwise,} \end{cases}$$

where $\phi_{L,U}(x|\mu,\sigma)$ denotes the density of a normal random variable truncated at $[L, U]$

and ϕ and Φ are the probability density and the cumulative distribution function respectively for the standard Normal distribution.

As a result the proposal distribution, $q(\beta'|\beta^t)$, can be expressed as

$$q(\beta'|\beta^t) = \phi_{P,\delta}(\beta'|\beta^t, \sigma_\beta)$$

with mean β^t and corresponding standard deviations σ_β and with the left tail truncated at P and the right tail truncated at the minimum of δ and $\min(\boldsymbol{\theta})$.

The candidate point β' is then accepted with probability

$$p(\beta^t, \beta') = \min\left(1, \frac{p(\alpha,\gamma,\delta,\beta')p(\boldsymbol{\theta}|\alpha,\gamma,\delta,\beta')q(\beta^t|\beta')}{p(\alpha,\gamma,\delta,\beta^t)p(\boldsymbol{\theta}|\alpha,\gamma,\delta,\beta^t)q(\beta'|\beta^t)}\right).$$

As the algorithm used is a single-component Metropolis-Hastings algorithm, the term $p(\boldsymbol{x}|\boldsymbol{\theta})$ does not contribute to the updating of β and thus can be treated as a constant. Clearly, $p(\boldsymbol{\theta}|\alpha,\gamma,\delta,\beta)$ will differ depending on whether a trapezium or sigmoidal prior is being implemented.

Updating $\boldsymbol{\theta}$

Each θ_i is updated separately yet the method for updating the individual θ's is the same. Hence, a general methodology for updating θ_i is given below.

A proposal distribution, $q(\theta_i'|\theta_i^t)$, is needed to generate the next value, θ_i^{t+1}, in the Markov chain given the current value of θ_i^t. Again, a convenient choice is the truncated Normal distribution as $\boldsymbol{\theta}$ is constrained to lie between β and α, see Section 3.4.3. This is represented as

$$q(\theta_i'|\theta_i) = \phi_{\beta,\alpha}(\theta_i'|\theta_i, \sigma_{\theta_i}).$$

The candidate point θ_i' is then accepted with probability

$$p(\theta_i^t, \theta_i') = \min\left(1, \frac{p(x_i|\theta_i')p(\theta_i'|\alpha,\gamma,\delta,\beta)q(\theta_i^t|\theta_i')}{p(x_i|\theta_i^t)p(\theta_i^t|\alpha,\gamma,\delta,\beta)q(\theta_i'|\theta_i^t)}\right).$$

As a result of using a single-component Metropolis-Hastings algorithm the term $p(\alpha, \gamma, \delta, \beta)$ does not contribute when updating θ_i and thus can be treated as a constant. Also the term $p(\boldsymbol{\theta}|\alpha, \gamma, \delta, \beta)$ can be simplified to $p(\theta_i|\alpha, \gamma, \delta, \beta)$ as only θ_i is being updated and therefore $\boldsymbol{\theta}_i$ stays the same. Again, the form of $p(\theta_i|\alpha, \gamma, \delta, \beta)$ will differ depending on whether a trapezium or sigmoidal prior is being implemented.

This section was intended briefly to explain the algorithm used to estimate the calendar dates of $\boldsymbol{\theta}, \alpha, \gamma, \delta$ and β. However, full details of the algorithms can be found in the C code in Appendix A.

3.6 Coding the MCMC: problems encountered

As discussed, in Section 2.8, WinBugs code for implementing archaeological phase models (assuming a uniform deposition rate) is available via Andrew Millard's web page. As a result, it was initially decided to work in WinBugs and extend the existing code to incorporate a range of alternative a priori deposition rates.

Clearly, the trapezium distribution is not a standard distribution, however, WinBugs offers an option to use sampling distributions that are not included in their list of standard distributions by using, what they call the 'zeros trick' or alternatively the 'ones trick' (see the WinBugs manual found at http://www.mrc-bsu.cam.ac.uk/bugs for details). While trying to adopt these tricks, a number of problems were encountered as well as peculiarities in the output. The main peculiarity was that the overall span (the difference between α and β) when using the trapezium prior was smaller than the overall span when using the uniform prior. This caused concern, as it was felt that the trapezium prior would allow for more uncertainty in the tails of α and β consequently resulting in the difference between the two being greater under the trapezium prior. After correspondence with Andrew Thomas (WinBugs technical queries), it became apparent that it was not possible to define a new sampling distribution using one of their 'tricks' when the parameters of the distribution are ordered *i.e.* in our case $\alpha \geq \gamma \geq \delta \geq \beta$.

As using WinBugs for extending the existing models was clearly not an option, it was felt that the best alternative was to write our own MCMC algorithms in R. Although R has many advantages, e.g. we have more control over the algorithms used, there is one main disadvantage of using R. R is very intense in terms of computer memory and each iteration is expensive in terms of computer time. Nonetheless, for a while, because of advantages associated with speed of coding we used R programs, running them as batch jobs, carrying out a smaller number of iterations each time to avoid potential memory problems, which decreased the total computer time needed.

However, as my research progressed and the case studies became more complicated, (in the sense of having multiple phases and larger numbers of radiocarbon determinations) memory problems occurred and the total amount of computer time needed increased greatly. For example, the reoccupation case study (see Section 4.3) consists of eight regions with a total of 133 radiocarbon determinations. The Upper Rhine only contains a small number of radiocarbon determinations (eight) and consequently results in problems with convergence. As a result the case study was run for approximately 1 million iterations to ensure all parameters in each region had reached convergence, this took around six days to run in R as a batch job. It was felt that this case study still had a relatively small number of radiocarbon determinations and that, if we were to tackle more complex problems in the future an alternative to R was needed. Thus it seemed sensible to invest time in learning the programming language C.

Due to my lack of experience in programming, this was not a straight forward task and it took several months to be able to write, implement and check code for both the conventional uniform and trapezium prior for multiple phases. Careful checking of the output from the C code against that from my own R code and BCal ensured that estimates being produced by my new C Code were robust. Although this has taken a great deal of time it has been beneficial as case studies such as the reoccupation case study can now be solved in a few hours rather than many days.

Consequently, all case studies for this thesis have been implemented in C and the

algorithms used can be found in Appendix A.

3.6.1 Computing aspects of the case studies within Chapter 4

This subsection briefly discusses the computational aspects of the case studies in Chapter 4. In both cases studies we monitor the main parameters (α, γ, δ and β) which are exported into R enabling us to use BOA to determine the length of burn-in and assess convergence. To assess the length of burn-in two alternative methods were used visual inspection of traces of output for each parameter in the MCMC chain and a more formal method developed Hiedelberger & Welch (1983) and implemented within BOA.

For the Coptic Textiles case study (Section 4.2) a burn-in length of approximately 1000 iterations was required. While a much longer burn-in length was required for the reoccupation case study (Section 4.3), all regions except the Upper Rhine required a burn-in length of approximately 10000 iteration, where as the Upper Rhine required a longer burn-in of approximately 20000 iterations.

Formal convergence diagnostics were carried out on all parameter chains once the burn-in lengths had been determined. The two diagnostics Geweke (1992) and Hielderger & Welch (1983) were used to assess convergence of each chain. In Section 4.2.5 we draw inferences for the Coptic textile case study, which are based on an MCMC sample of 200000 iterations. In Section 4.3.5 we draw inference for the reoccupation case study. The data available for the reoccupation case study stretch over approximately 4000 years on the radiocarbon time scale, and some individual regions (*e.g.* the Upper Rhine) have radiocarbon determinations spanning some 2500 years. As a result of this and as a result of the algorithms adopted for implementing this case study, it was found that the MCMC chains needed to run for a total of 1 million iterations, in order for all parameter chains to reach convergence.

Chapter 4

Case studies for the uniform, trapezium and sigmoidal models

4.1 Introduction

In this chapter, two case studies are presented to illustrate use of the alternative non-uniform prior distributions developed in the previous chapter. In both case studies, the authors believe that the rate of deposition or manufacture was not uniform over the proposed range. Therefore, the aim of this chapter is to illustrate the difference in the archaeological conclusions drawn from the data when implementing both uniform and non-uniform prior deposition models.

The first case study arises from Van Strydonck *et al.* (2004) where the authors' interest lies in comparing radiocarbon dating to art historic dating of both Roman and Coptic textiles. There are a number of groups of stylistically related textiles of interest, each treated as having its own phase of manufacture. However, only one of the groups of textiles will be used for illustration. This results in a simple, single phase, case study which contains a small number of radiocarbon determinations and no relative chronological *a priori* information.

The second case study arises from Housley *et al.* (1997). This case study is concerned with the human reoccupation of NW Europe after the last ice age. This is a much larger case study in which radiocarbon determinations are available from eight different regions within NW Europe. When moving to interpreting radiocarbon determinations which relate to several phases, questions arise about the relationships between the phases. For example: "do the phases overlap?", "do they abut?", "when did Phase A stop and when did phase B begin?", and so on. In this chapter we take the first step in modelling this problem by treating the calendar dates of phase boundaries as independent of one another. This allows us to calculate probabilistic answers to the above kind of questions. The next step, outlined in Chapter 5, will be to move on to incorporate *a priori* information about the relations between phase boundary dates, in the form of joint prior distributions.

Note from this point forward we will refer to a 'uniform model' when a uniform prior for the rate of deposition of datable material has been implemented. Similarly, for a trapezium and sigmoidal model.

4.2 Case study: Radiocarbon dating and art historic dating of Roman and Coptic textiles

4.2.1 Introduction

The first case study arises from Van Strydonck *et al.* (2004), briefly discussed in Chapter 3 as an example from the applied literature which motivates my research. The case study is concerned with dating Roman and Coptic textiles from Egypt and, in particular, comparing the radiocarbon results with chronologies proposed by art historians. The majority of art historians base their chronologies on a comparison of stylistic features which may arise from a variety of media such as paintings, sculptures, mosaics and archaeological features.

The authors are essentially interested in learning about the chronologies of a variety of Roman and Coptic textiles through the use of radiocarbon dating. Particular interest lies in the length of time over which each type of textile was manufactured and estimates for the last date of manufacture. For the purpose of this case study only one of the stylistic groups of textiles, which consists of twelve woollen tunics, will be discussed; see Table 4.1.

Sample id.	Det.	Sample id.	Det.
UtC-9431	1630±60	UtC-9049	1615±40
UtC-9051	1590±40	KIA-10569	1585±30
UtC-2612	1540±60	UtC-2619	1530±70
UtC-9050	1485±40	KIA-10570	1470±35
UtC-7253	1450±50	UtC-7240	1420±60
UtC-9052	1380±40	UtC-2620	1350±70

Table 4.1: Radiocarbon determinations associated with each of the 12 woolen tunics from Van Strydonck *et al.* (2004).

The method adopted by the authors, to summarize the radiocarbon determinations, was rather *ad hoc* as it neither appeared to answer the questions posed nor to be based on sound statistical arguments. This method is referred to as the 'summed probability distribution' method which has become increasingly popular over the last decade within the archaeological community, it is believed that the reason behind its selection is due to the following statement from the OxCal manual (Ramsey, 2005).

"Combining probability distributions by summing is usually difficult to justify statistically but will generate a probability distribution which is a best estimate for the chronological distribution of the items dated."

The summed probability distribution is calculated by calibrating the radiocarbon determination from each artefact separately to produce a posterior estimate of its calendar date. Then the posteriors for the individual calendar dates, which relate to the same stylistic group of textiles, are combined by summing. The result is then normalized to give a probability distribution. It is this resulting probability distribution that is referred to as the 'best estimate of the chronological distribution'. However, averaging

the posteriors for the individual calendar dates (which do not date the same event) does not estimate the form of the underlying distribution that is of interest. In fact, it is not clear what interpretation can legitimately be made from the distribution produced by this method, or what quantity has the distribution constructed in this way. However, the authors calculate an inter-quartile range and a 95% probability interval from this probability distribution, which they believe represents a proxy for the chronology of the woollen tunics.

Note that the distributions being combined are posterior estimates of the calendar dates of the individual textiles. As a result, the combined distribution gives simply the posterior distribution of the date of an unknown sample selected at random from that set of textiles. It does not directly say anything about the process from which those samples might come.

The summed probability distribution for the group of twelve woollen tunics has been reproduced (using IntCal04) for illustrative purposes, see Figure 4.1. While Figure 4.1 may look appealing it is not at all clear that it relates to the chronology of the woollen tunics. The 95% HPD interval for the summed probability distribution is calculated as 350-710 AD, suggesting that the tunics are no older than the 8^{th} century AD. As a result, Van Strydonck *et al.* (2004) concludes that it is very unlikely that the woollen tunics belong to the 11^{th} or 12^{th} century AD as suggested by some art historians.

The problem tackled in Van Strydonck *et al.* (2004) is similar to that discussed in Buck *et al.* (1992), in that it can be thought of as a group of related textiles from a single phase of manufacture. The authors could therefore adopt a model-based Bayesian framework as detailed in Buck *et al.* (1992), summarized in Section 2.5, to estimate the start and end date of manufacture and then use these two dates to calculate the length of time over which the textiles were manufactured.

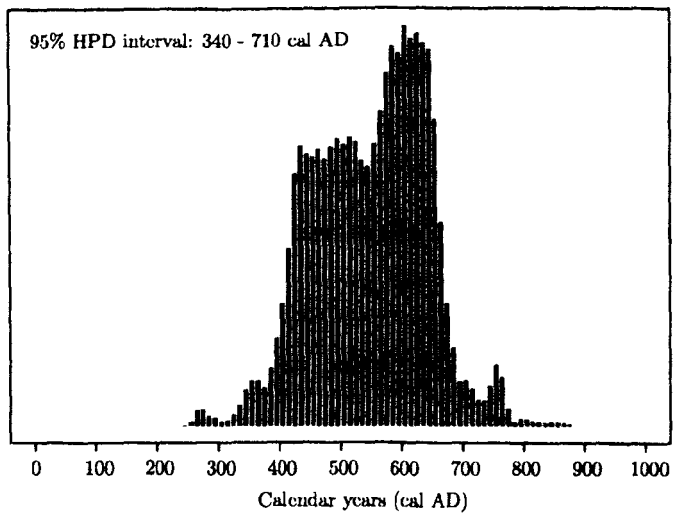


Figure 4.1: Summed probability distribution of the 12 radiocarbon dates (in Table 4.1) each associated with one of the woollen tunics.

4.2.2 A simple temporal model

As it is extremely unlikely that we will have access to samples that directly date the start and end dates of the manufacturing phase, the radiocarbon determinations, $x_1 \pm \sigma_1, \ldots, x_{12} \pm \sigma_{12}$ each associated with a true unknown calendar dates θ_i, are used to learn indirectly about these dates of interest. Adopting the notation used in Buck *et al.* (1992) the earliest date for the manufacturing phase is labelled α cal AD[1] and the latest date for the manufacturing phase is labelled β cal AD.

In this case study we can use *a priori* information arising from the art historians concerning the period of time in which the manufacturing phase occurred. This results in the following prior distributions for $\alpha \sim U(0\text{AD}, \beta)$ and $\beta \sim U(\alpha, 1600\text{AD})$. Note that as we are now working in calendar years AD, $\beta > \alpha$. By adopting the methods devised in Buck *et al.* (1992) we assume that the material suitable for dating were

[1]To be consistent with the original authors, all calibrated ages are given in terms of calendar years AD. This is to enable a direct comparison with the results we obtain, to those given in Van Strydonck *et al.* (2004).

manufactured uniformly between the start and end date of the manufacturing phase, *i.e.* $\theta_i|\alpha, \beta \sim U(\alpha, \beta)$ for $i = 1, \ldots, 12$.

4.2.3 Prior beliefs

Using the conventional uniform model would be one way to tackle the problem in Van Strydonck *et al.* (2004). However, my particular interest in that paper lies in the following statement "... the manufacturing dates of related textiles are not uniformly distributed over the proposed range, but there exists an introduction phase, a blooming period and a period of decline". Clearly, if the methods devised in Buck *et al.* (1992) were implemented, which assume a uniform rate of manufacture throughout the whole phase, the authors would be contradicting their prior beliefs about the rate at which the textiles were manufactured.

Ideally, we would like to be able to incorporate the authors' prior beliefs into the analysis to allow for a more realistic and robust modelling of the manufacturing process of the textiles. As seen in Section 3.4, both the trapezium and sigmoidal models would better reflect the reality of the manufacturing phase, over the conventional uniform model. The reason is that, rather than there being a sudden increase from zero to the maximum rate of manufacture, as in the uniform model (see Figure 3.1(a)), there would be a gradual increase in the rate of manufacture, followed by a period in which there is a constant rate of manufacture, followed by a gradual decrease in the manufacturing rate. After correspondence with Van Strydonck, it was agreed that a suitable prior distribution, to reliably represent his problem, would be a trapezium distribution, as discussed in Section 3.4.1. The three stages of the manufacturing phase are illustrated in Figure 4.2, in terms of the *a priori* beliefs stated in Van Strydonck *et al.* (2004).

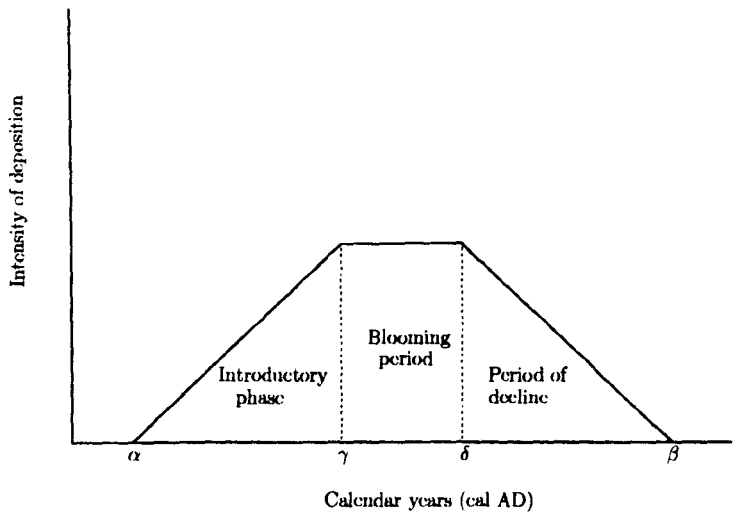


Figure 4.2: Schematic representation of the trapezium prior for the manufacturing phase of the stylistically related textiles in terms of the prior beliefs stated in Van Strydonck *et al.* (2004).

4.2.4 The trapezium model

When modelling the manufacturing phase in terms of the trapezium model the same parameters, α and β, are used to represent the start and end dates (cal AD) of the manufacturing phase. However, there are now two extra parameters, firstly, γ which represents the beginning of the 'blooming period' and secondly, δ which represents the end of the 'blooming period' (see Figure 4.2). Again, the chronologies proposed by the art historians are regarded as *a priori* information about the period of time over which the woollen tunics were manufactured. Therefore, prior distributions for α, γ, δ and β defined by the following relationships are used

$$\alpha \sim U(0\text{AD}, \gamma), \quad \gamma \sim U(\alpha, \delta), \quad \delta \sim U(\gamma, \beta) \quad \text{and} \quad \beta \sim U(\delta, 1600\text{AD})$$

and as we are working in calendar years AD $\beta \geq \delta \geq \gamma \geq \alpha$. When implementing the trapezium model, samples suitable for dating are assumed to be manufactured with a

trapezium rate of manufacture through the period α to β,

$$\theta_i|\alpha,\gamma,\delta,\beta \sim \text{Trap}(\alpha,\gamma,\delta,\beta) \text{ for } i = 1,\ldots,12.$$

4.2.5 Dating Coptic textiles

This next section offers a reanalysis of the twelve stylistically related woollen tunics to illustrate the difference in inferences incurred when the conventional uniform model and the alternative trapezium prior model are implemented. To evaluate the posterior date estimates, under both models, the Metropolis-Hastings algorithm detailed in Section 3.5 has been implemented. In this case study the relevant internationally agreed calibration curve is IntCal04 (Reimer *et al.*, 2004).

Last date of manufacture

Van Strydonck *et al.* (2004) are primarily interested in comparing their results to those obtained by the art historians, as there are many discrepancies between art historians as to when the woollen tunics were last manufactured. Initial interest focuses on the estimates of the last date of manufacture, β, under the two alternative prior models. Using the results provided under both the uniform and trapezium models should help clarify whether the textiles were likely to be manufactured as late as some art historians believe.

The estimates of the last date of manufacture, β, (based on an MCMC sample of 200000 iterations) were found to be unimodal but not all symmetric. The date estimates obtained are summarized by their modal values and their 95% highest posterior density (HPD) regions (in Table 4.2). HPD regions define the shortest interval within which 95% of the posterior probability occurs.

Using Table 4.2, we can make the following interpretations. Under the uniform model,

Model	95% HPD interval for the start and end dates of manufacture (cal AD)	
	α	β
Uniform	350 - 520	610 - 765
Trapezium	190 - 500	620 - 900

Table 4.2: The 95% HPD intervals for the start and end date of the manufacturing phase under the conventional uniform and trapezium models.

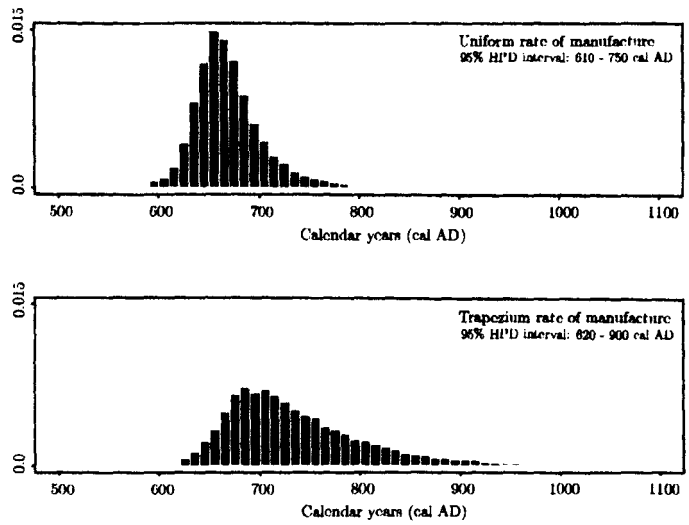


Figure 4.3: Marginal posterior distributions for the end date of the manufacturing phase under the two alternative models.

the HPD interval for the estimate of the last date manufacture is 610-750 cal AD, while under the trapezium model, the interval is 620-905 cal AD. Clearly, under the trapezium model the HPD interval is much wider, resulting from the posterior distribution being right-skewed, as shown in Figure 4.3. This suggests that the trapezium model allows for more uncertainty in the estimate of this date. However, the modal values under the two alternative models are relatively similar. Under the uniform model the modal value is 664 cal AD and under the trapezium prior model the modal value is 678 cal AD.

Duration of manufacture

Additional interest lies in the total length of time over which the textiles were manufactured. We define the duration of manufacture as the distribution of the difference

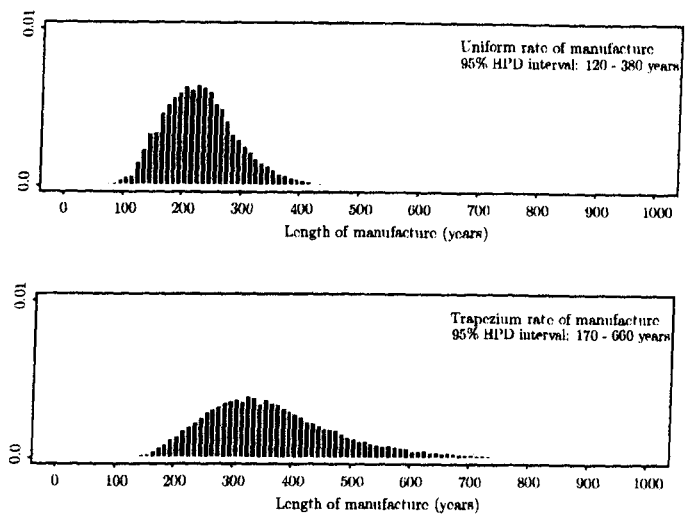


Figure 4.4: The marginal posterior distributions for the duration of manufacture under the uniform and trapezium models.

between α and β and summarize it by its modal value and 95% highest posterior density (HPD) regions, see Table 4.3. Using these values, along with Figure 4.4 the following interpretations can be made. Under the uniform model the 95% HPD interval is 120-380 years while under the trapezium model the 95% HPD interval is 170-660 years. Again, under the trapezium model the HPD interval is much wider, resulting from the posterior distribution being right-skewed. However, the modal values under the two models differ. Under the uniform model the modal value for the duration of manufacture is 225 years while under the trapezium model the corresponding value is 330 years.

Model	95% HPD interval for the length of manufacture
Uniform	120 - 380 years
Trapezium	170 - 660 years

Table 4.3: The 95% HPD regions for the duration of manufacture under the uniform and trapezium models.

Conculsions

The aim of the paper Van Strydonck *et al.* (2004) was to compare the radiocarbon results of Roman and Coptic textiles with the chronologies proposed by the art historians. It

is hard to make any direct comparisons with the results obtained from the summed probability distribution method, used by Van Strydonck *et al.* (2004), to those presented by the art historians, due to our concerns stated in Section 4.2.1. However, as illustrated, by adopting a model-based Bayesian approach we can obtain estimates for the last date that a particular type of textile was manufactured. This allows us to make a direct comparison to the dates estimated by the art historians.

The chronology proposed by art historians, for the woolen tunics, varies somewhat. Some historians date the tunics as late as the 11^{th} or 12^{th} century AD, while others date them no older than the 6^{th} to 9^{th} century AD. When implementing the conventional uniform model, there is no evidence to support the suggestion that the woollen tunics are older than the 8^{th} century AD, while under the trapezium model the woollen tunics can be thought to be as late as the 9^{th} to 10^{th} century AD. So the evidence from the radiocarbon dating appears to rule out the later dates proposed on the basis of art-historical evidence.

The conventional uniform model offers a considerably different date estimate, from the trapezium model, for the last date of manufacture. However, we feel that the trapezium model better reflects the prior beliefs stated in Van Strydonck *et al.* (2004) and has considerably more intuitive archaeological interpretation than the uniform model in this particular case.

4.3 Case study: Human reoccupation of NW Europe after the last Ice Age

4.3.1 Introduction

This case study is motivated by three papers, Housley *et al.* (1997), Blockley *et al.* (2000) and Blackwell and Buck (2003), all of which aim to interpret a large collection of radiocarbon determinations which relate to the human reoccupation of NW Europe as the climate improved at the end of the last Ice Age. Radiocarbon determinations are

available from eight different regions within NW Europe and it is of interest to know the earliest date for which there is evidence for reoccupation and the order in which the regions were reoccupied.

The first attempt to address these issues was Housley *et al.* (1997). In this paper the authors treated the radiocarbon determinations in one region as being independent from the radiocarbon determinations in other regions. Then within each region, in order to answer questions concerning the timing of the reoccupation process, the authors compute a 'moving sum' of the 1σ ranges for all radiocarbon determinations associated with each region. This method produces a series of histograms, one for each of the eight regions, where the 'bin widths' of the histograms were chosen to be roughly the same as the average 1σ ranges (which essentially allows the radiocarbon determinations to be treated as point estimates in the reoccupation process). These histograms were then interpreted as "supporting a model of population movement", in which the earliest non-zero bin in each histogram was taken as identifying the start of reoccupation ('pioneer phase') and the modal bin in each histogram was interpreted as a 'residential phase' *i.e.* when the population was fully established. Figure 4.5 is given to illustrate the 'moving sum' methodology adopted in Housley *et al.* (1997) for the set of radiocarbon determinations available from the Upper Rhine region.

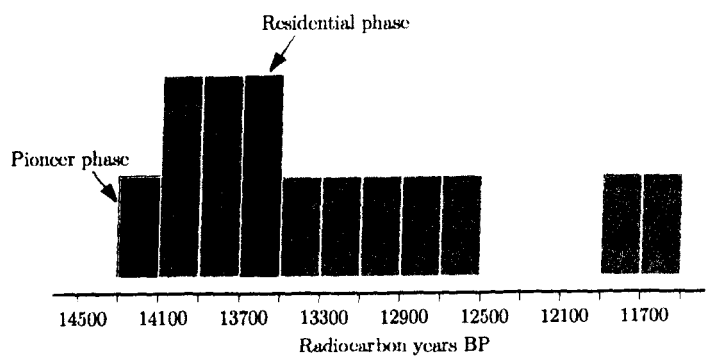


Figure 4.5: The moving sum distribution of the radiocarbon determinations available from the Upper Rhine region.

The second attempt to address these issues was Blockley *et al.* (2000). Initially, they

review Housley *et al.* (1997) and make two main criticisms: firstly that the authors only account for 1σ on the uncalibrated dates and not 2σ and, secondly, that they should have calibrated the radiocarbon determinations so that interpretations can be made on the calendar time scale. Blockley *et al.* (2000) then undertake a reanalysis of the same data, seeking to address these weaknesses. They start by calibrating each radiocarbon determination within the region onto the calendar time scale and then adopt the same "summed probability distribution" method as Van Strydonck *et al.* (2004). Blockley *et al.* (2000) believe that this method gives 'a best estimate of the chronological distribution of events'. However, they conclude that there is no clear evidence for a pioneer or residential phase from the summed probability plots for each region.

These two papers motivated the third, Blackwell and Buck (2003), in which the authors suggest a model-based Bayesian approach. Blackwell and Buck implement the conventional uniform model as devised in Buck *et al.* (1992) to estimate the first date of reoccupation in each of the eight regions. Using these estimates, for the first date of reoccupation, and treating the eight regions as being independent of one another, Blackwell and Buck (2003) gave a probabilistic answer to the question "in which order were the regions reoccupied?". Full details of their approach is given in Section 4.3.5.

4.3.2 Prior beliefs

In the first paper (Housley *et al.*, 1997) the authors discuss their prior beliefs about the mechanics of the reoccupation process and describe it as

> ... a two stage process. There was an initial pioneer phase when only a few small hunting parties moved to explore and exploit the previously unpopulated area. It was followed by the establishment of larger, but possibly not permanent, occupation in each of the eight regions, termed the residential camp phase.

This led Blackwell and Buck (2003) to note that it might be worth exploring " ... models that reflect the likely sparseness of material for dating from the early stages of reoccupation in each region." This is thus another example in the applied literature in which there is a need to seek alternatives to the conventional uniform model. The sigmoidal prior, as described in Section 3.4.2, was the example illustrated in Blackwell and Buck (2003), although, it can argued that the trapezium prior would adequately represent the properties of the non-uniform prior that they describe.

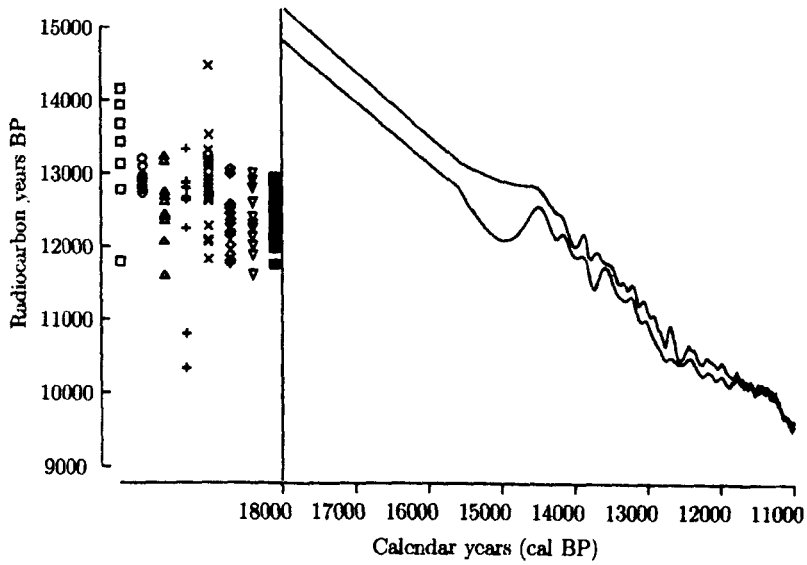
The following sub-sections provide a reanalysis of the collection of radiocarbon determinations from Housley *et al.* (1997). As discussed in Section 1.1.4, IntCal98, the internationally agreed calibration curve at the time Blackwell and Buck (2003) was written, has since been updated to IntCal04. For this reason, the first step taken in this reanalysis will be to compare the relevant sections of IntCal98 and IntCal04 to see whether it is likely that the interpretations made from the data using IntCal04 will differ greatly to those when IntCal98 was used.

Then (in Section 4.3.5) we will look at the difference in interpretations when the uniform, trapezium and sigmoidal models are implemented.

4.3.3 IntCal98 versus IntCal04: Will it make a difference to the archaeological interpretations?

Figure 4.6 shows the radiocarbon ages ($x_{i,j}$'s) from Housley *et al.* (1997), for the eight different regions, against the relevant section of (a) IntCal98 and (b) IntCal04 [Table B.1, Appendix B, gives the full set of radiocarbon determinations]. There is a clear difference between the two sections of curves, with the main differences lying between 10000-13000 (radiocarbon years BP). IntCal04 is much smoother, for the section under consideration, and it is therefore believed that the archaeological interpretations inferred from the data will differ considerably, depending upon which curve is used.

(a): IntCal98



(b): IntCal04

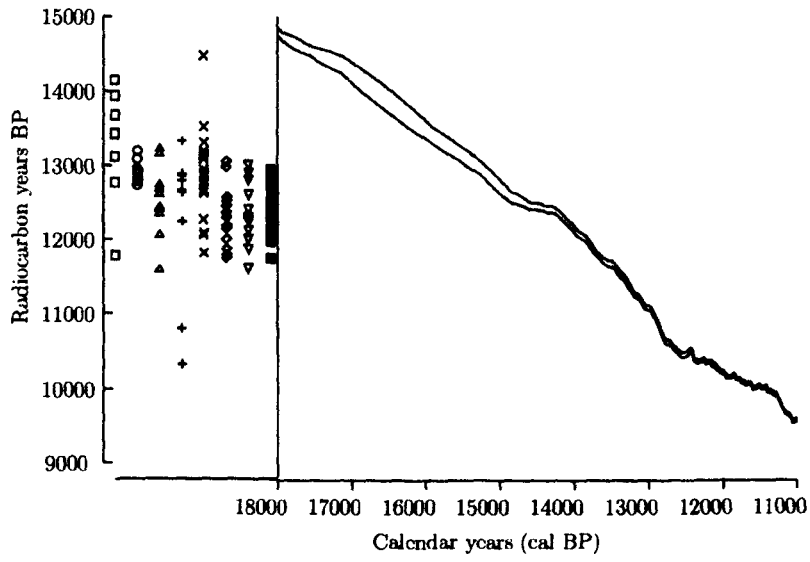


Figure 4.6: The radiocarbon ages ($x_{i,j}$'s) from Housley *et al.* (1997) shown alongside the relevant sections of the calibration curves, (a) IntCal98 and (b) IntCal04, with the corresponding number of radiocarbon determinations in each region. □=Upper Rhine (7), ○=Middle Rhine (9), △=Southern Germany (10), +=Belgium (13), ×=Thuringian Basin (23), ◇=Northern Germany (16), ▽=Paris (14) and ⊠=British Isles (41).

4.3.4 Setting up the multiple phase problem

Conventional uniform model

Each region is assumed to have its own phase of reoccupation, hence the dates of events in a given region are independent of the dates of events in other regions. The primary aim is to estimate the first date of human reoccupation in each region. The earliest boundary for the phase in each region, which marks the beginning of the 'Pioneer sub-phase', is labelled α_j cal BP (where $j = 1, 2, \ldots, 8$ denoting the number of regions). The latest boundary for the phase in each region marks the end of the reoccupation phase and is labelled β_j cal BP. While it is α_j, the early boundary, that is of interest the reoccupation phase is seen as bounded in time, for practical and theoretical reasons.

There are no samples that directly relate to the α_j or β_j within each region, therefore the radiocarbon determinations for events occurring in region j are used to learn indirectly about α_j and β_j. Since there is no *a priori* knowledge about the period of time in which the reoccupation phase occurred, it is assumed that α_j and β_j lie somewhere in the range of the calibration curve (with the constraint that α_j will always be greater than β_j as we are working on the cal BP time scale). It is also assumed that the samples suitable for dating in region j were deposited uniformly between the start and end date of the reoccupation phase.

Trapezium and Sigmoidal models

When implementing both the trapezium and sigmoidal models the same notation α_j and β_j cal BP (for $j = 1, \ldots, 8$) is used to represent the first and last date of the reoccupation phase, respectively. There are also two extra parameters to be considered, firstly γ_j which in this case represents the beginning of the 'residential camp sub-phase' and secondly δ_j which represents the end of the 'residential camp sub-phase'.

As there is no *a priori* information relating to the period of time in which the reoccupation phase occurred, it is assumed that $\alpha_j, \gamma_j, \delta_j$ and β_j lie somewhere in the range of the calibration curve (with the constraint $A > \alpha_j \leq \gamma_j \leq \delta_j \leq \beta_j > P$ as we

are working on the cal BP timescale).

When implementing the trapezium model the samples suitable for dating are assumed to be deposited with a trapezium rate of deposition through the period α_j to β_j, that is to say $\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j \sim \text{Trap}(\alpha_j, \gamma_j, \delta_j, \beta_j)$, where $\theta_{i,j}$ denotes the ith radiocarbon determination in the jth phase (for $j = 1, \ldots, 8$).

Similarly, when implementing the sigmoidal model it is assumed that the material suitable for dating was deposited between α_j and β_j at a sigmoidal rate

$$\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j \sim \text{Sig}(\alpha_j, \gamma_j, \delta_j, \beta_j).$$

Full details of both these models, for a multiple phase problem, can be found in Section 3.4.4, along with the algorithms used to evaluate the posterior distributions of interest in Section 3.5. In this case study the latest version of the calibration curve, IntCal04, will be used.

4.3.5 The human reoccupation of NW Europe

First date of reoccupation

The primary objective is to estimate the earliest date of reoccupation of each region under study. The first dates of reoccupation (α_j) under the uniform, trapezium and sigmoidal model are directly comparable. The estimates of these dates (based on a sample of a million iterations) are summarized by their 95% HPD regions (in Table 4.4) and visually by the marginal posterior distributions (in Figure 4.7).

Using Table 4.4 it can be seen that the first dates of reoccupation for each region under the trapezium and sigmoidal models are similar to one another, but differ considerably from that obtained with the uniform model.

Figure 4.7 shows the estimates of the marginal posterior distributions for the first date of reoccupation in each region under the conventional uniform (red lines), the trapezium

(blue lines) and the sigmoidal models (green lines). This confirms that the first dates of reoccupation, for each region, under the trapezium and sigmoidal models are similar to one another. In particular it is seen that the two models result in their estimates being earlier in time than those derived using the conventional uniform model. There are slight differences between the posterior estimates obtained when using the trapezium and sigmoidal models, as illustrated in Figure 4.7. The most noticeable differences occur in the estimates of the first dates of reoccupation of the British Isles and the Thuringian Basin.

Region	95% HPD interval for the date of first reoccupation (cal BP)		
	Uniform model	Trapezium model	Sigmoidal model
Upper Rhine	16500 - 19580	16740 - 23470	16770 - 23720
Thuringian Basin	16350 - 17650	16620 - 18600	16740 - 19150
Southern Germany	15300 - 16750	15440 - 18530	15470 - 18850
Middle Rhine	15150 - 15770	15260 - 16580	15250 - 16680
Belgium	15320 - 16900	15440 - 18690	15520 - 19480
Paris Basin	15030 - 16010	15140 - 17010	15170 - 17330
Northern Germany	14310 - 15440	14530 - 16320	14510 - 16530
British Isles	14670 - 15180	14750 - 15530	14780 - 15750

Table 4.4: The 95% HPD intervals for the first date of reoccupation of the eight regions under the conventional uniform, trapezium and sigmoidal models.

The order in which reoccupation took place

Blackwell and Buck (2003) used two summary methods to discuss the order in which reoccupation took place. The first gives probabilities that each region is temporally ranked one (earliest) through to eight (latest), the second calculates the probabilities of particular orderings, in which reoccupation occurred.

The first summary method to be discussed is the probability that each region is temporally ranked one (earliest) through to eight (latest). Table 4.5(a) gives the corresponding probabilities when the conventional uniform model was been used, Table 4.5(b) when the trapezium model was been used and Table 4.5(c) when the sigmoidal model was been used. Under all three models the Upper Rhine was the first

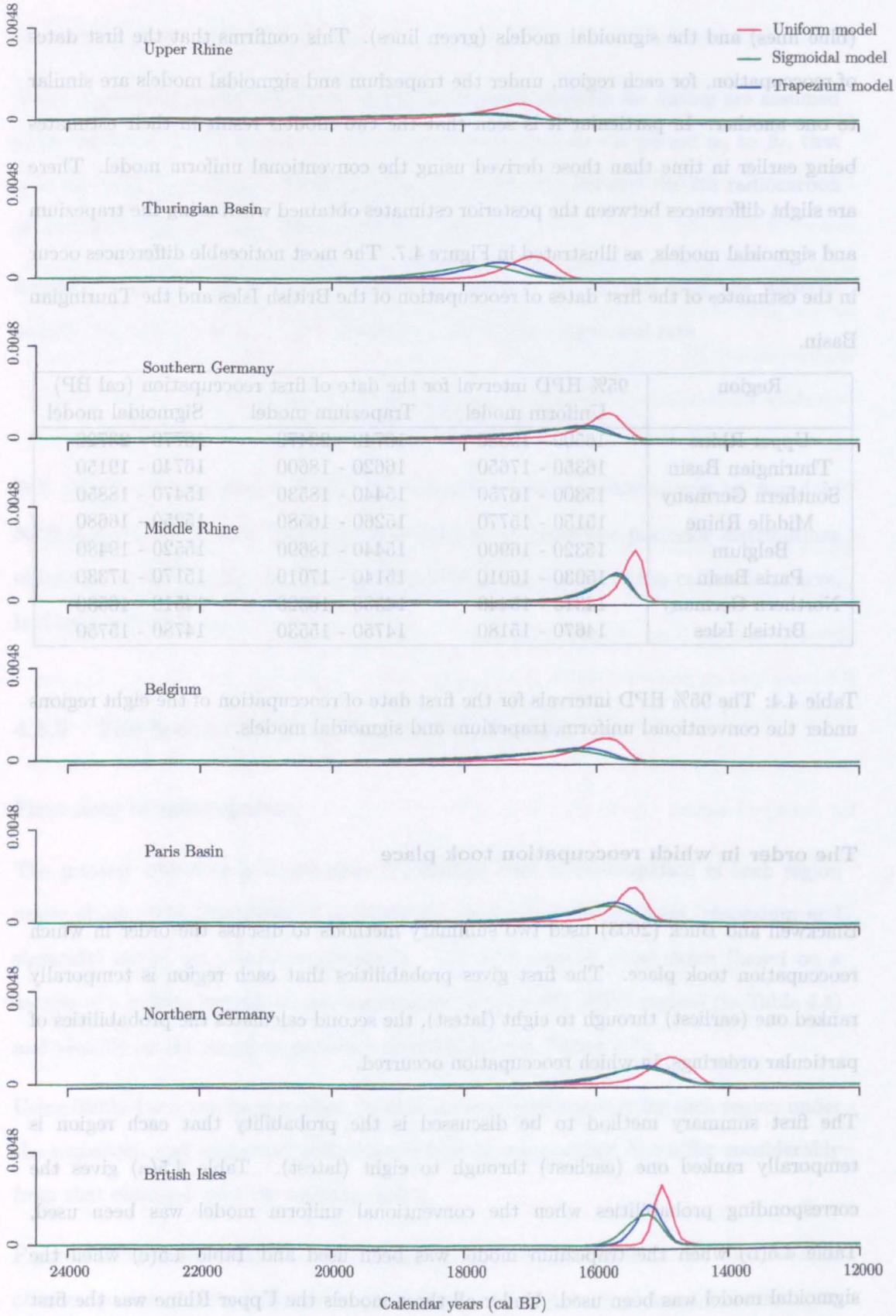


Figure 4.7: Marginal posterior distributions for the first date of reoccupation in each region under conventional uniform (red), trapezium (blue) and sigmoidal (green) models.

region to be reoccupied (with probability of 0.75 when the conventional uniform model was used, 0.75 when the trapezium model was used and 0.69 when the sigmoidal model was used). The British Isles was the last region to be reoccupied, with a probability of 0.61 under the trapezium model and 0.58 under the sigmoidal model, but under the conventional uniform model the corresponding probability was only 0.39. Again it can be seen that the trapezium and sigmoidal models give similar results to one another, but that these differ from the results obtained when using the uniform model.

The second summaries reported are the posterior probabilities of particular order in which reoccupation might have occurred. The ten most likely of those are given in Tables 4.6(a), (b) and (c) for the conventional uniform model, the trapezium model and the sigmoidal model, respectively. As there are eight regions of interest, there are a possible (8!) 40430 different orderings in which the regions could have been reoccupied.

The most likely ordering differs between the three models. Under the uniform model the most likely order is the Upper Rhine, Thuringian Basin, Belgium, Southern Germany, Paris, Middle Rhine, British Isles and Northern Germany with a posterior probability of 0.082. Under the trapezium model, the most likely order is the Upper Rhine, Thuringian Basin, Belgium, Southern Germany, Paris Basin, Middle Rhine, Northern Germany and the British Isles with a posterior probability of 0.0260. Under the sigmoidal model the most likely ordering is the Upper Rhine, Thuringian Basin, Belgium, Southern Germany, Paris Basin, Middle Rhine, Northern Germany and the British Isles with a posterior probability of 0.0237. The most likely order under the trapezium model and sigmoidal model is the same, under the uniform model the sequence differs in that the ordering of the the British Isles and Northern Germany is reversed.

Under all three models it is clear that the ten most likely orderings account for only a small amount of the total posterior probability. Thus when trying to make inferences concerning the order of reoccupation, it is clear that there is a large amount of uncertainty to be taken into consideration. From Tables 4.6(a), (b) and (c), it is clear that there is more uncertainty in the order of reoccupation under the trapezium and sigmoidal models

a): Uniform model

Region	1	2	3	4	5	6	7	8
Upper Rhine	0.75	0.23	0.02	0.00	0.00	0.00	0.00	0.00
Thuringian Basin	0.22	0.71	0.07	0.00	0.00	0.00	0.00	0.00
Southern Germany	0.01	0.03	0.38	0.42	0.13	0.03	0.00	0.00
Middle Rhine	0.00	0.00	0.02	0.08	0.37	0.49	0.04	0.00
Belgium	0.02	0.04	0.46	0.35	0.10	0.03	0.00	0.00
Paris Basin	0.00	0.00	0.06	0.13	0.37	0.39	0.05	0.00
Northern Germany	0.00	0.00	0.00	0.00	0.02	0.05	0.32	0.61
British Isles	0.00	0.00	0.00	0.00	0.00	0.01	0.60	0.39

b): Trapezium model

Region	1	2	3	4	5	6	7	8
Upper Rhine	0.75	0.19	0.06	0.00	0.00	0.00	0.00	0.00
Thuringian Basin	0.14	0.58	0.24	0.04	0.00	0.00	0.00	0.00
Southern Germany	0.05	0.10	0.28	0.32	0.17	0.07	0.01	0.00
Middle Rhine	0.00	0.01	0.04	0.11	0.29	0.40	0.14	0.01
Belgium	0.06	0.12	0.29	0.30	0.16	0.06	0.01	0.00
Paris Basin	0.00	0.01	0.08	0.17	0.30	0.30	0.12	0.02
Northern Germany	0.00	0.00	0.02	0.04	0.08	0.14	0.35	0.37
British Isles	0.00	0.00	0.00	0.00	0.00	0.04	0.35	0.61

c): Sigmoidal model

Region	1	2	3	4	5	6	7	8
Upper Rhine	0.69	0.21	0.08	0.02	0.00	0.00	0.00	0.00
Thuringian Basin	0.17	0.53	0.25	0.05	0.00	0.00	0.00	0.00
Southern Germany	0.05	0.10	0.25	0.33	0.18	0.07	0.02	0.00
Middle Rhine	0.00	0.00	0.03	0.10	0.25	0.41	0.19	0.02
Belgium	0.09	0.14	0.29	0.27	0.14	0.06	0.01	0.00
Paris Basin	0.00	0.02	0.08	0.19	0.31	0.25	0.13	0.02
Northern Germany	0.00	0.00	0.02	0.05	0.10	0.15	0.32	0.37
British Isles	0.00	0.00	0.00	0.00	0.01	0.06	0.35	0.58

Table 4.5: The probability that each region is temporally ranked one (earliest) through to eight (latest) under a) the conventional uniform b) the trapezium and c) the sigmoidal models.

a): Uniform model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	1	1	1	2
Thuringian Basin	2	2	2	2	2	2	2	2	2	1
Southern Germany	4	4	3	3	4	4	3	3	5	4
Middle Rhine	6	5	6	5	6	5	6	5	6	6
Belgium	3	3	4	4	3	3	4	4	3	3
Paris Basin	5	6	5	6	5	6	5	6	4	5
Northern Germany	8	8	8	8	7	7	7	7	8	8
British Isles	7	7	7	7	8	8	8	8	7	7
Probability	0.082	0.078	0.069	0.065	0.044	0.040	0.037	0.033	0.026	0.025

b): Trapezium model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	1	1	1	1
Thuringian Basin	2	2	2	2	2	2	2	2	2	2
Southern Germany	4	4	3	3	4	4	3	3	5	5
Middle Rhine	6	6	6	6	5	5	5	5	6	6
Belgium	3	3	4	4	3	3	4	4	3	3
Paris Basin	5	5	5	5	6	6	6	6	4	4
Northern Germany	7	8	7	8	8	7	8	7	7	8
British Isles	8	7	8	7	7	8	7	8	8	7
Probability	0.026	0.025	0.025	0.025	0.023	0.023	0.022	0.021	0.014	0.013

c): Sigmoidal model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	1	1	1	1
Thuringian Basin	2	2	2	2	2	2	2	2	2	2
Southern Germany	4	4	3	3	4	4	3	5	3	5
Middle Rhine	6	6	6	6	5	5	5	6	5	6
Belgium	3	3	4	4	3	3	4	3	3	4
Paris Basin	5	5	5	5	6	6	6	4	6	4
Northern Germany	7	8	8	7	8	7	8	8	7	7
British Isles	8	7	7	8	7	8	7	7	8	8
Probability	0.024	0.022	0.020	0.019	0.018	0.016	0.013	0.013	0.013	0.012

Table 4.6: The ten most likely orders of the reoccupation of the eight regions under study (1=earliest, 8=latest) when implementing a) the conventional uniform b) the trapezium and c) the sigmoidal models.

than under the uniform model.

4.3.6 Does the reoccupation process overlap in the eight regions?

This next section offers a slightly different perspective by looking in more detail at how the reoccupation process spread across NW Europe. The aim here is to look at how different regions became reoccupied in relation to each other. For example, it might be expected that two regions close together spatially may well be reoccupied at similar times compared to two regions which are further apart.

Two spatially close regions from the case study have been selected for illustrative purposes, these are the Thuringian Basin and Southern Germany. Interest lies in when Southern Germany was reoccupied in relation to the Thuringian Basin. It is questions of this kind that archaeologists are seeking solutions to.

For the purpose of this illustration the Thuringian Basin will be referred to as Phase A and Southern Germany as Phase B. The parameters $\alpha_A, \gamma_A, \delta_A$ and β_A belong to Phase A and $\alpha_B, \gamma_B, \delta_B$ and β_B belong to Phase B. Also, only inferences for the trapezium model will be discussed.

From Figure 4.7 it appears that the Thuringian Basin was reoccupied first and Southern Germany shortly afterwards and clearly there is overlap between the two phases. When implementing the trapezium model for example, there are several different types of overlap that may occur. In this particular case we are interested in the relation between the process of reoccupation in the two regions.

1. What is the probability that the reoccupation process in the Thuringian Basin began before the reoccupation process in Southern Germany? "$P(\alpha_A > \alpha_B)$"

2. What is the probability that the reoccupation process in Southern Germany stated within the Pioneer sub-phase of the Thuringian Basin? "$P(\alpha_A > \alpha_B > \gamma_A)$"

3. What is the probability that the reoccupation process in Southern Germany stated

within the Residential camp sub-phase of the Thuringian Basin?

"P($\gamma_A > \alpha_B > \delta_A$)"

These are relatively simple questions to calculate answers to, once the output from the MCMC chains is available. For example, to calculate the probability that the reoccupation process in the Thuringian Basin began before the reoccupation process in Southern Germany, the estimates of α_A and α_B are treated as being independent and ranked in ascending order. We can then count the number of times in which $\alpha_A > \alpha_B$. As a result, we can calculate that the posterior probability that the process of reoccupation of the Thuringian Basin started before the process of reoccupation of Southern Germany is 0.82. In the same way, we can calculate the posterior probabilities for questions 2 and 3.

Figures 4.8 and 4.9 provide graphical illustrations of the answers to question 2 and 3, above. The posterior probability that the process of reoccupation in Southern Germany started within the Pioneer sub-phase of the Thuringian Basin is 0.69, whereas the posterior probability that the process of reoccupation in Southern Germany started within the Residential camp sub-phase of the Thuringian Basin is 0.12. From this, it can be seen that the two phases appear to be reoccupied within a similar time period, but that the reoccupation process probably started earlier in the Thuringian Basin than it did in Southern Germany.

4.4 Summary

The aim of this chapter was to illustrate, via two case studies, the difference in archaeological interpretations of the data when we assuming different *a priori* information about the rate at which the material was deposited within a phase of archaeological activity.

In the second case study, which relates to the human reoccupation of NW Europe, the trapezium and sigmoidal models led to very similar estimates for the dates of interest. By

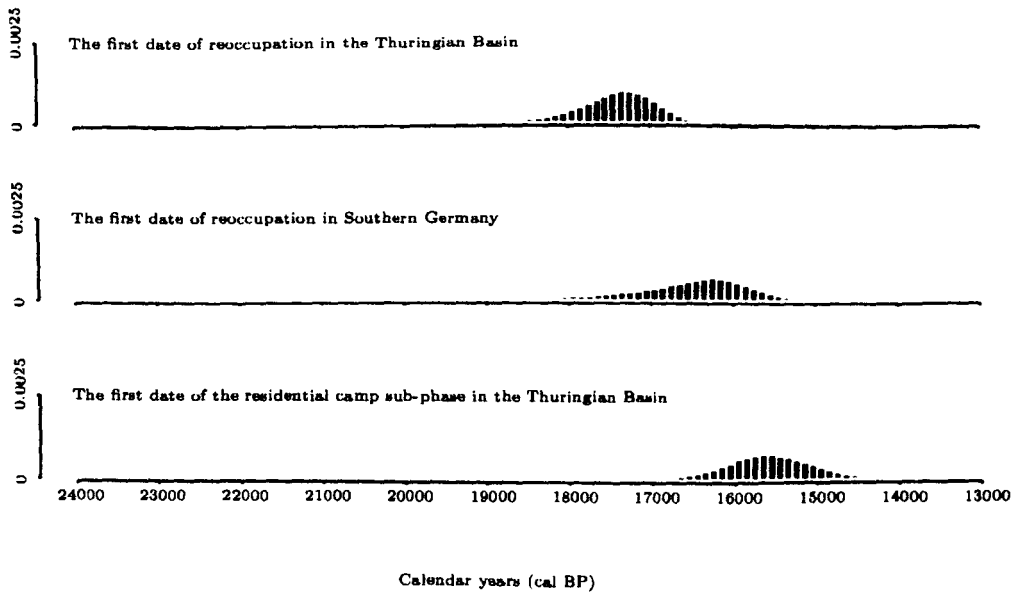


Figure 4.8: The probability that reoccupation in Southern Germany started within the Pioneer sub-phase of the Thuringian Basin

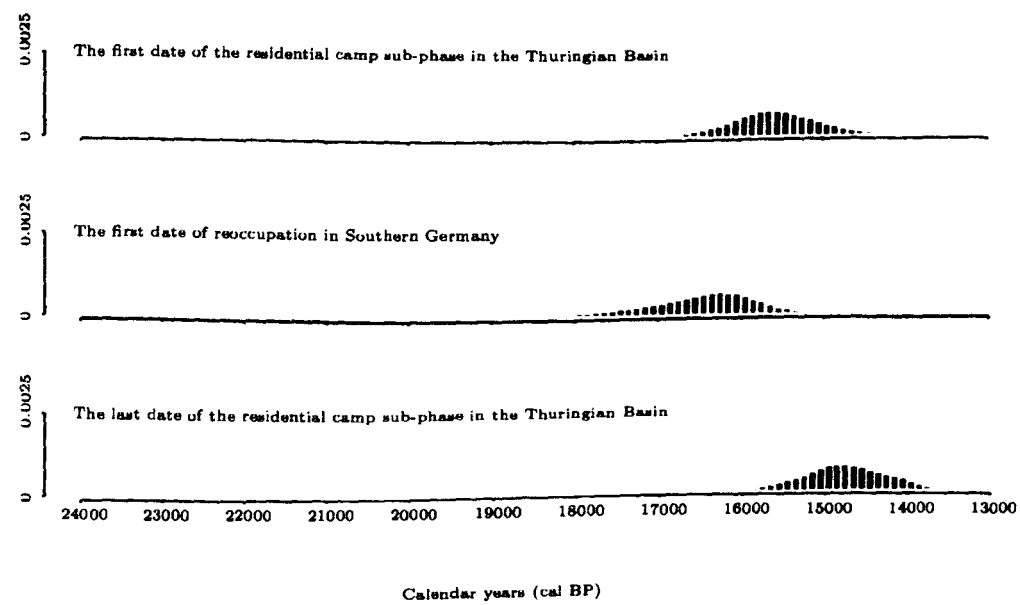


Figure 4.9: The probability that reoccupation in Southern Germany started within the Residential camp sub-phase of the Thuringian Basin

contrast, the uniform model offers considerably different date estimates. The uniform model seems much less realistic since, as noted by Blackwell and Buck "the uniform model is most appropriate for short lived phases at single spatial locations and may not be so sensible in the case of reoccupation/colonisation of landscapes."

As seen in Section 3.4.1 the uniform model is a special case of the trapezium model (when $\alpha = \gamma$ and $\beta = \delta$). For this reason, we feel that the trapezium model has more intuitive archaeological interpretation than the sigmoidal model and in the following chapters we focus on the use of the trapezium model.

In the second case study, although we have made inferences of both a temporal and spatial nature, the models we have used are purely temporal. This case study is a typical example of a spatio-temporal problem that is currently (at best) tackled as if it were only a temporal problem. Any spatial information has been ignored in order to learn about temporal aspects of the reoccupation of the individual regions of NW Europe, rather than the reoccupation of NW Europe as a whole. It seems a sensible assumption that regions spatially closer together were more likely to be reoccupied at similar times than those regions further apart. If this is the case then we could use a joint prior distribution to express *a priori* information about relationships between the regions.

Therefore the following chapters will discuss ideas to help us move towards tackling such problems in a fully spatiao-temporal framework.

Chapter 5

First steps towards fully spatio-temporal modelling

5.1 Introduction

Initially, a brief review of the thesis so far is given. Chapter 3 is concerned with the modelling of the deposition rate of datable material within a phase of archaeological activity. Four alternative non-uniform prior deposition models have been discussed but only two in great detail, as they are believed to have a much wider use to the archaeological community. These are referred to as the trapezium and sigmoidal models. Chapter 4 then presents two case studies to illustrate the difference in archaeological conclusions drawn from the data when implementing both uniform and non-uniform prior deposition models.

However, this is only the first step in developing a more robust statistical framework within which we can tackle a wider variety of archaeological problems. Now that a statistical framework is in place for modelling the deposition rate, within a phase of activity, in a more robust and coherent manner, the next step is to extend the existing models to allow for a wider range of *a priori* information to be incorporated.

As seen in the previous chapters, archaeological dating is a much wider problem than simply dating the individual objects. The reason for this is that only on rare occasions do the dates of individual objects allow us to answer directly the chronological question posed. It is our aim to develop a general framework within which we can tackle problems concerning chronology building, whether these relate to developments in fashions and technologies or to the understanding of how landscapes were colonised by humans, animals or plants.

The next section is intended to review the types of *a priori* information that may arise during an archaeological calibration problem, as well as discussing how they are integrated into the existing models.

5.2 Types of *a priori* information

When faced with archaeological calibration problems, there are a number of different types of *a priori* information that may arise.

1. We may have *a priori* information about the time period between successive events, see Section 2.5.1.
2. Very often, we also have *a priori* information relating to the time period, *e.g.* the Bronze Age period.
3. Most commonly, *a priori* information exists about the chronological orderings of events or phases.

Such *a priori* information can easily be integrated into the existing models. Christen *et al.* (1995) and Christen (1994a) discusses how prior information about the likely time elapsed between the deposits of each sample in a sequence of radiocarbon determinations may be incorporated, see Section 2.5.1. Prior information relating to the historical time period of interest can also be easily integrated, by choosing suitable values of P and A which correspond, for example, to the start and end dates of the Bronze age period.

Most commonly, we may want to incorporate *a priori* information relating to the chronological ordering of events or phases. As seen in Section 3.4.4, when working with problems which comprise of multiple phases of activity, *a priori* information may arise about the relations between phase boundary parameters. To incorporate such information we use a relatively simple form of joint prior,

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \propto \mathrm{I}_C(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta})$$

where

$$\mathrm{I}_C(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \in C \\ 0 & \text{otherwise} \end{cases}$$

where C is some set of values of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$,$\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$. One constraint on the parameters which must always be satisfied (as we are working in calendar years BP) is $A > \alpha_j \geq \gamma_j \geq \delta_j \geq \beta_j > P$. This form of joint prior allows us to incorporate a wide range of *a priori* information, whether it is in the form of sequence information (such as $\alpha_1 > \alpha_2$) or more specific information (such as $0 < \alpha_1 - \beta_1 < 100$). As well, as *a priori* information arising about the relations between phase boundary parameters, there is often stratigraphic information leading to prior information about the ordering of radiocarbon dates within a phase. As seen in Section 3.4.4 this is represented as

$$\mathrm{I}_{D_j}(\boldsymbol{\theta_j}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta_j} \in D_j \\ 0 & \text{otherwise} \end{cases}$$

where $D_j = (D_1, \ldots, D_m)$ represents the set of values that $\boldsymbol{\theta_j}$ can take within each of the phases [note $\boldsymbol{\theta_j}$ represents the set of θ's belonging to the jth phase].

If this type of *a priori* information is easily integrated into the analysis of radiocarbon data, what else is it that we want to achieve? We want to be able to go one step further and be able to incorporate more complicated relations between phase boundary parameters, by introducing joint prior distributions. A simple example of the type of *a*

priori information that we may want to incorporate is given below.

Example

Suppose there is an archaeological excavation in which there appears to be two phases of activity. Firstly, there is a settlement inside an enclosure, then secondly there is evidence of a settlement outside of the enclosure. In such a case, interest may lie in whether the settlement inside the enclosure was contemporary or successive to that outside. In both settlements there is evidence of pottery that dates to the Late Bronze Age[1] (LBA). We label activity associated with the settlement inside the enclosure as Phase A and that outside the enclosure as Phase B.

How could we incorporate this information into the analysis? The Late Bronze Age in the British Isles dates between approximately 1000BC to 600BC, a period of around 400 years. As a result of there being evidence in both phases of only LBA pottery we may assume that the start of both phases occurred during the LBA period. This implies that the start of the two phases must be no greater than 400 years apart. However, we know nothing about the ordering of the phases, as this is the question we are interested in learning about. Taking this into account we could express this knowledge of the relation between the start date of the two phases, α_A and α_B, as a joint prior distribution, $p(\alpha_A, \alpha_B)$.

Adopting the algorithms discussed in Section 3.5 the joint prior, $p(\alpha_1, \alpha_2)$, could be incorporated to enable us to arrive at more coherent and stratificatory estimates of the phase boundary parameters α_1 and α_2. Given these estimate, we can calculate a probabilistic answer to the question "Is Phase A contemporary or successive to Phase B?". However, before this is possible we need to think about the types of joint prior distributions that may be appropriate.

The next section will focus on ideas for defining joint (bivariate) distributions, which we feel have a wide range of uses when developing chronological models, rather than being

[1]The Bronze age is characterized by the first use of copper or bronze and its chronology is strictly local, hence it started at different times in different parts of the world.

problem specific. We start by thinking about two-phase problems, *i.e.* defining bivariate distributions.

5.3 Defining joint (bivariate) prior distributions

The idea behind defining a joint prior distribution is to enable us to incorporate *a priori* information about the relationships between parameters, when nothing is known about specific values of individual parameters. When defining a joint prior distribution to represent such information it is important to see what marginal distributions are induced by the joint distributions we choose. Ideally, we want robust joint prior distributions that lead to marginal distributions which have an intuitive archaeological interpretation.

In this section we discuss a range of plausible bivariate distributions, which are felt to have properties which would be of use in a number of archaeological problems. In the example given above, we are interested in defining a joint prior distribution, $p(\alpha_1, \alpha_2)$, *i.e.* with respect to the start dates of the phases. However, depending upon the problem at hand there may be a wide range of joint prior distributions, with respect to the parameters of interest, that we may want to define. For this reason, in the following section we discuss a general framework that provides flexibility in defining a joint prior distribution for the parameters of interest.

Notation

For the rest of this chapter, X and Y represent random variables that both take values in [0,1] and are defined on the same sample space. In practice, both will usually represent (linearly transformed) calendar dates.

If we think back to the example given on page 115, we know that $|\alpha_A - \alpha_B| \leq 400$ years. Figure 5.1 represents an example of a joint prior distribution that takes into account this information. In Figure 5.1, both the x-axis and y-axis range from 0 to 1. We define a parameter w to represent our prior knowledge about the maximum time between the two start dates *e.g.* $|\alpha_A - \alpha_B| \leq w$. The parameter w relates to the parameter u, seen

in Figure 5.1, by $w = u\sqrt{2}$.

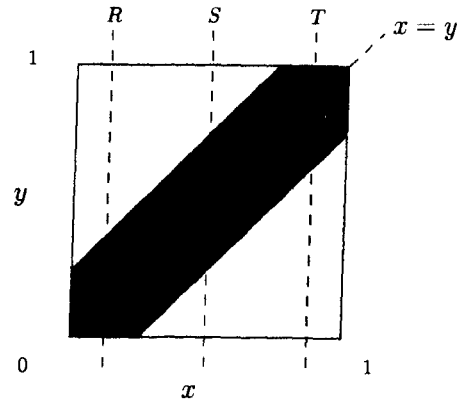


Figure 5.1: An example of a joint prior distribution $p(x, y)$

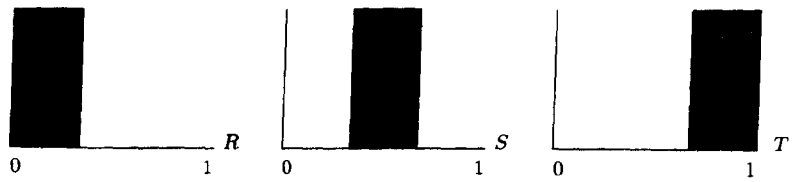


Figure 5.2: Cross sections of the joint prior distribution from Figure 5.1, along the lines R, S and T in the direction of early to late with respect to y.

The probability density is uniform over the dark grey shaded region and the light grey shaded areas represent a probability density of zero. Figure 5.2 (intended to help visualize the shape of the joint distribution) shows the cross sections of the joint prior distribution along the lines R, S and T in the direction 0 to 1, with respect to y.

The probability density function for the joint prior distribution, $p(x, y)$, as seen in Figure 5.1, can be written in a general form as

$$p(x, y) = \begin{cases} c & \text{if } |x - y| \leq u\sqrt{2} \\ 0 & \text{otherwise,} \end{cases} \tag{5.1}$$

where c is just a normalizing constant. However, $p(x, y)$ can be written more explicitly, in three parts corresponding to the cross-sections in Figure 5.2.

When $x < u\sqrt{2}$

$$p(x,y) = \begin{cases} c & \text{if } 0 \leq y \leq x + u\sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$

When $u\sqrt{2} \leq x \leq 1 - u\sqrt{2}$

$$p(x,y) = \begin{cases} 0 & \text{if } y < x - u\sqrt{2} \\ c & \text{if } x - u\sqrt{2} \leq y \leq x + u\sqrt{2} \\ 0 & \text{if } y > x + u\sqrt{2}. \end{cases}$$

When $x > 1 - u\sqrt{2}$

$$p(x,y) = \begin{cases} 0 & \text{if } y < x - u\sqrt{2} \\ c & \text{if } x - u\sqrt{2} \leq y \leq 1. \end{cases}$$

We are only interested in using joint prior distributions that lead to marginal distributions, $p(x)$ and $p(y)$, which have an intuitive archaeological interpretation. For this reason we are interested to see what marginal distributions are induced by Equation 5.1. As $p(x,y)$ is a symmetric distribution the marginal distributions, $p(x)$ and $p(y)$, will take the same form and can be calculated as follows,

$$p(x) = \int p(x,y)dy \quad \text{and} \quad p(y) = \int p(x,y)dx. \tag{5.2}$$

Using Equation 5.2, the normalized marginal probability density function, $p(x)$, is given as

$$p(x) = \begin{cases} (x + u\sqrt{2})/(2u\sqrt{2} + 2u^2) & \text{if } 0 \leq x < u\sqrt{2} \\ (2u\sqrt{2})/(2u\sqrt{2} + 2u^2) & \text{if } u\sqrt{2} \leq x \leq 1 - u\sqrt{2} \\ (1 - x + u\sqrt{2})/(2u\sqrt{2} + 2u^2) & \text{if } 1 - u\sqrt{2} < x \leq 1. \end{cases} \tag{5.3}$$

Clearly, Equation 5.3 will differ depending upon the value of the $u\sqrt{2}$. However, it is unlikely that we would wish to use a joint distribution where the value of $u\sqrt{2} \geq 0.5$.

Using Equation 5.3, we can draw from $p(x)$ to visualize the marginal prior distribution.

Figure 5.3 represents $p(x)$ for the case $u = 0.1$, in which we see the marginal prior $p(x)$ is essentially uniform over the region [0,1], except at the end of the region.

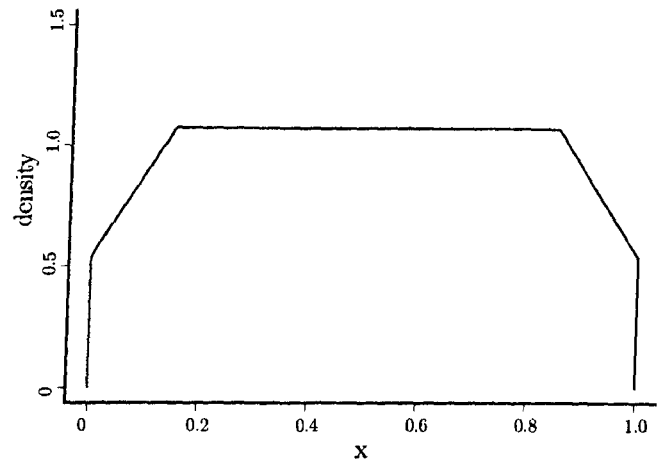


Figure 5.3: The marginal prior distribution $p(x)$ given in Equation 5.3, where $u = 0.1$

Implementing this form of joint prior would be relatively simple. One way to achieve this, would involve generating a new value for α_A using a truncated normal proposal distribution, as described in Section 3.5, where L takes the value $\alpha_B - u\sqrt{2}$ and U takes the value of $\alpha_B + u\sqrt{2}$, likewise for generating new values of α_B. These bounds would need to be modified slightly to take account of the edge effect, *i.e.* when $0 \leq x < u\sqrt{2}$ and $1 - u\sqrt{2} < x \leq 1$. However, when working with archaeological calibration problems $U - L$ may be relatively small in comparison to $A - P$, if this is the case then we would only in theory be interested in the region $u\sqrt{2} \leq x \leq 1 - u\sqrt{2}$. However, this is not always the case.

A range of distributions will be needed since, for specific case studies, our *a priori* information about w will vary. As well as the value of w varying, the actual shape of the joint prior distribution may also vary, depending upon the case study and the type of *a priori* information we want to incorporate. Figure 5.4 shows an alternative joint prior distribution, again with cross sections of the joint prior distribution along the lines R, S

and T, to help visualize the shape of the distribution.

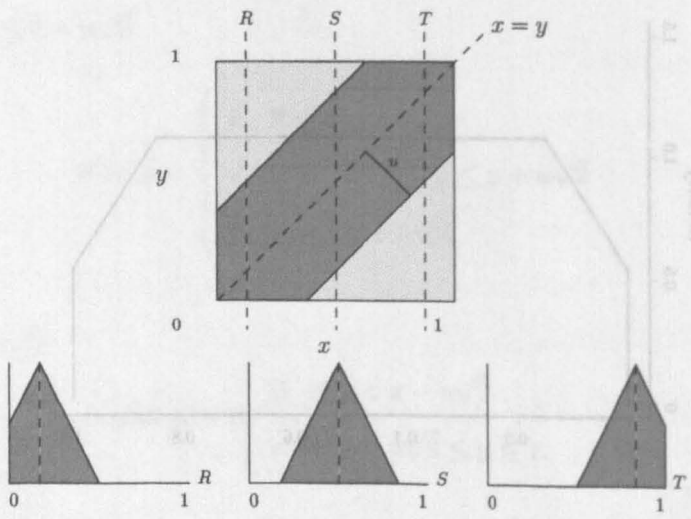


Figure 5.4: An example of a joint prior distribution with corresponding cross sections

The probability density function for the joint prior distribution, $p(x, y)$, as seen in Figure 5.4 can be written in a general form as

$$p(x,y) \propto \begin{cases} 1 - (u\sqrt{2})^{-1}|x-y| & \text{if } |x-y| \leq u\sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$

A typical example of a situation in which a joint prior distribution with the above properties might be used, is when the archaeologists believe they are dating the transition from one phase of activity to the next *i.e.* β_A and α_B are essentially dating the same event.

Figure 5.5 shows another alternative prior distribution, again with cross sections, this time along the lines R, S, T, V and W. The corresponding probability density function for the joint prior distribution, $p(x, y)$, as seen in Figure 5.5 can be written in a general

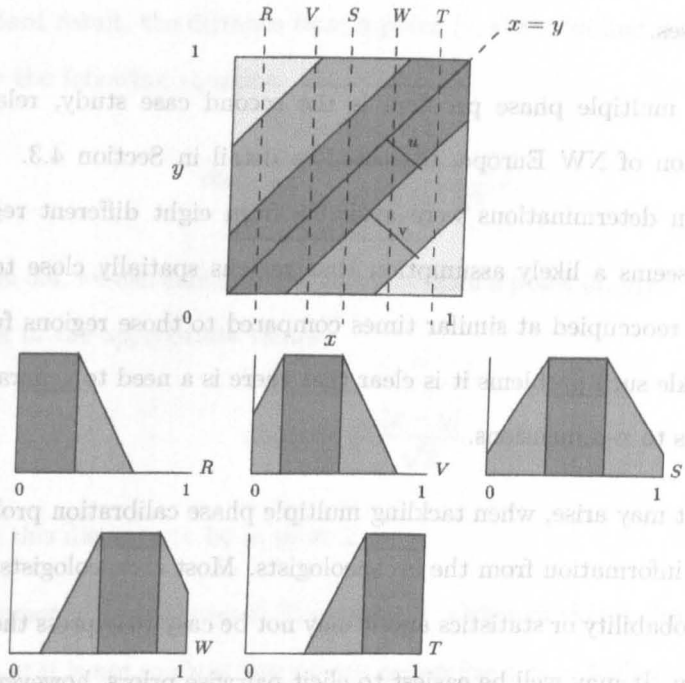


Figure 5.5: An example of an alternative joint prior distribution with corresponding cross sections

form as

$$p(x,y) \propto \begin{cases} 1 & \text{if } |x-y| \leq u\sqrt{2} \\ 1-(v\sqrt{2})^{-1}|x-y| & \text{if } u\sqrt{2} \leq |x-y| \leq (u+v)\sqrt{2} \\ 0 & \text{otherwise.} \end{cases}$$

This is slightly more complicated joint prior distribution, but still has a number of archaeological uses.

So far, this chapter has only discussed bivariate prior distributions *i.e. a priori* information arising from the relation between two parameters of interest, for example $p(\alpha_A, \alpha_B)$, $p(\beta_A, \alpha_B)$ or $p(\beta_A, \beta_B)$. However, a typical archaeological calibration problem may consists of complex systems of phases based on artefact typologies and stratigraphy. As a result, the next step is to extend the joint prior distributions discussed in this section to multiple dimensions to allow us to incorporate *a priori* information arising

from multiple phases.

An example of a multiple phase problem is the second case study, relating to the human reoccupation of NW Europe, discussed in detail in Section 4.3. In this case study, radiocarbon determinations were available from eight different regions within NW Europe. It seems a likely assumption that regions spatially close together were more likely to be reoccupied at similar times compared to those regions further apart. To be able to tackle such problems it is clear that there is a need to generalize the joint prior distributions to n-dimensions.

One problem that may arise, when tackling multiple phase calibration problems, is how to elicit *a priori* information from the archaeologists. Most archaeologists are certainly not experts in probability or statistics and it may not be easy to express their beliefs in a probabilistic form. It may well be easiest to elicit pairwise priors, however, a great deal of care will be needed to ensure that the pairwise priors do not contradict one another.

5.4 Generalizing to higher dimensions

The idea behind the following sections is to be able to generalize the joint prior distributions in the previous section to higher dimensions. When working in $\mathbb{R}^2$ and $\mathbb{R}^3$ we can visualize the form of the joint prior distribution that we would like to incorporate. However, beyond $\mathbb{R}^3$ this is no longer possible. We propose an idea in Section 5.4.2 and briefly explain the reasons for choosing the methods that we adopted.

5.4.1 Rewriting the two-dimensional case

If we think back to Figure 5.1 we are essentially defining a set of points, S, as

$$S = \{(x, y) \, \epsilon \, \mathbb{R}^2 : \text{distance from a point } (x, y) \text{ to the line } x = y \text{ is at most } u\}.$$

Using a standard result, the distance from a point (r, s) to the line $Ax + By + C = 0$ is calculated, by the following equation

$$\text{distance} = \frac{|Ar + Bs + C|}{\sqrt{A^2 + B^2}}. \tag{5.4}$$

Using Equation 5.4, we can calculate the distance from a point $(x, y) \epsilon S$ to the line $x = y$ by substituting in the appropriate values

$$\text{distance} = \frac{|x - y|}{\sqrt{2}} \tag{5.5}$$

and we require this distance to be at most u.

This gives the exact form as shown in Equation 5.1, although this is clearly what we are trying to arrive at it is not so clear how we can generalize this to higher dimensions. The following section offers a non-standard way of arriving at the same information as in this sub-section, but at the same time allows us to setup a framework which we believe can easily be generalized to higher dimensions.

5.4.2 Alternative formalization

Using Figure 5.6, we wish to calculate the nearest point, $Q(q, q)$, on the line $x = y$ to the point $P(x, y)$ *i.e.* the shortest distance from point P to the line. This distance will be measured along the perpendicular from P to the line *i.e.* the vector $QP(x - q, y - q)$. Before we can calculate the length of the vector QP we need to know q in terms of x and y. We know that we want QP to be orthogonal to $(1, 1)$ and we also know that two vectors are orthogonal if their dot product is equal to zero. As a result, we can calculate q as

$$\begin{aligned} 1 \times (x - q) + 1 \times (y - q) &= 0 \\ x - q + y - q &= 0 \\ x + y &= 2q \\ q &= (x + y)/2. \end{aligned} \tag{5.6}$$

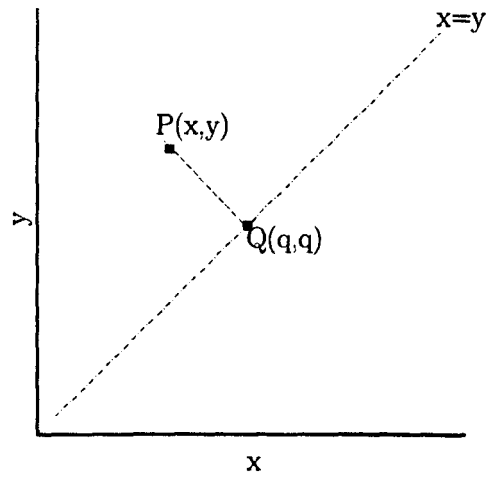


Figure 5.6: Illustration of the shortest distance from a point P to the line $x = y$.

Hence, the nearest point to P on the line $x = y$ is the point $Q\left(\frac{x+y}{2}, \frac{x+y}{2}\right)$. Next we need to calculate the length of the vector QP *i.e.* calculate the distance between the two points P and Q. We know that the length of a vector can be calculated (in $\mathbb{R}^2$) using Pythagoras' Theorem, as

$$\text{length} = \sqrt{x^2 + y^2}. \tag{5.7}$$

Manipulating Equation 5.7 allows us to write the length of QP (or distance from Q to P), in a non-standard way, such as

$$\left\{(x,y) - \left(\frac{x+y}{2}\right)\mathbf{1}\right\}^{\mathrm{T}} I \left\{(x,y) - \left(\frac{x+y}{2}\right)\mathbf{1}\right\} \tag{5.8}$$

where I represents the identity matrix. Equation 5.8 represents the length squared of the vector QP and simplifies to $\frac{1}{2}(x-y)^2$. The main reason for writing the distance in this way is so that we can introduce the identity matrix. As we generalize to higher dimensions (*e.g.* $\mathbb{R}^3$) we may have pairwise *a priori* information with regard to the time between parameters of interest such as $|x-y| \leq a$, $|x-z| \leq b$ and $|y-z| \leq c$ where

$a \neq b \neq c$. The idea, is that by changing the diagonal elements of I we can scale our measure of distance according to the *a priori* information that arises.

Figure 5.7 gives two visual representation of the prior defined by

$$\left\{(x,y) - \left(\frac{x+y}{2}\right)1\right\}^{\mathrm{T}} I \left\{(x,y) - \left(\frac{x+y}{2}\right)1\right\} \leq u^2 \tag{5.9}$$

with varying values of u. Both these figures are seen to represent the same distribution as that illustrated in Figure 5.1

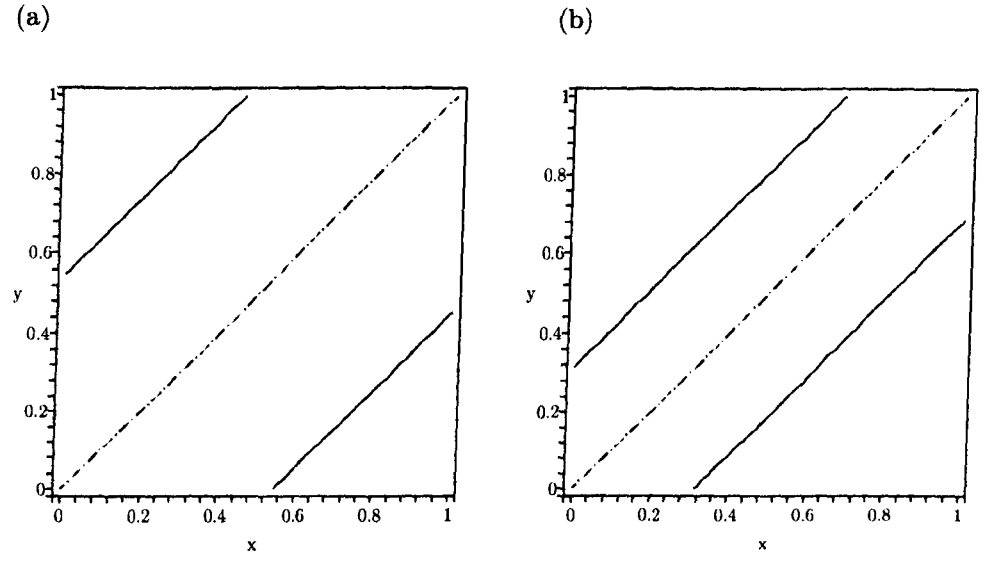


Figure 5.7: The joint prior distribution resulting from Equation 5.9, with values (a) $u^2 = 0.15$ and (b) $u^2 = 0.05$.

Given the ideas presented in this section, in the next section we discuss use of similar ideas to generalize to $\mathbb{R}^3$. In theory once the framework for a three-dimensional prior distribution is established it should be fairly easily to extend to $\mathbb{R}^n$.

5.4.3 Priors in $\mathbb{R}^3$

As discussed in Section 5.2 it is quite unlikely that we we will be able to elicit anything more than pairwise prior distributions for the parameters of interest, when working with

more than two archaeological phases of interest. The aim of this section is to look at extending the ideas proposed in the previous section while ensuring that the types of *a priori* information that may arise can easily be incorporated.

If we think about the joint prior distribution in general terms, as we did in Section 5.4.2, we are essentially interested in defining a set of points, S, such that

$$S = \{(x, y, z) \in \mathbb{R}^3 : \text{distance from a point } (x, y, z) \text{ to the line } x = y = z \text{ is at most } u\}.$$

If we consider the joint prior distribution $p(x, y, z)$ and fix one of the parameters, x for example (as we are interested in incorporating pairwise prior information), we can visualize ways in which we could represent the joint prior distribution $p(y, z)$. Figure 5.8 shows two ways in which we could represent the joint prior distribution $p(y, z)$. The centre represents the point (x, x), then given a fixed x we can think about how y and z could vary. The first way is represented by the dashed line, which defines a square about the centre, while the dotted lines defines a circle about the centre. We favour the idea of using a circle as we are not allowing z to be at its extreme while y is also at its extreme. By using the general equation of a circle, we are essentially interested in defining points of (y, z) such that $(y - x)^2 + (z - x)^2 \leq u^2$; similarly for $p(x, y)$ and $p(x, z)$.

By manipulating Equation 5.8 we can write the distance squared, from a point $(x, y, z) \in S$ to the line $x = y = z$, as

$$\left\{(x, y, z) - \left(\frac{x + y + z}{3}\right)\mathbf{1}\right\}^T I \left\{(x, y, z) - \left(\frac{x + y + z}{3}\right)\mathbf{1}\right\} \leq u^2. \tag{5.10}$$

Expanding Equation 5.10 and multiplying through by $\frac{3}{2}$ gives

$$x^2 - xy - xz + y^2 - yz + z^2 \leq \frac{3u^2}{2}, \tag{5.11}$$

which essentially defines a shape resembling a cylinder running around the line $x = y = z$. Figure 5.9(a) gives an illustration of Equation 5.11 when $u = 0.4$ while Figure 5.9(b) represents the same figure rotated so that we are looking along the line $x = y = z$.

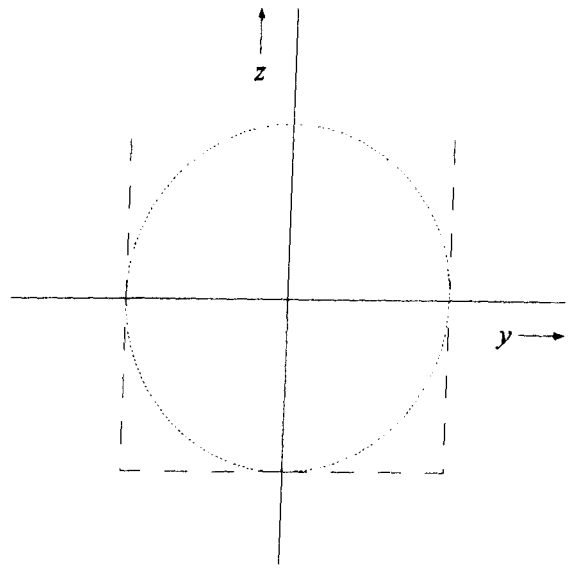


Figure 5.8: Cross-section through $\mathbb{R}^3$ with fixed x.

Despite the symmetry in x, y and z the joint pairwise prior distributions will differ in shape from those derived in Section 5.4.2. We can derive the joint pairwise prior distribution $p(y, z)$ by integrating with respect to x, which gives

$$\frac{2}{9} - \frac{1}{3}z - \frac{1}{3}y - \frac{2}{3}yz + \frac{2}{3}z^2 + \frac{2}{3}y^2 \leq u^2. \tag{5.12}$$

Figure 5.10 gives visual representations of the joint pairwise distribution $p(y, z)$ with varying values of u. As Equation 5.11 is symmetric (as a result of using I) in x, y and z all pairwise distributions, $p(x, y), p(x, z)$ and $p(y, z)$ will take the same form.

5.4.4 Asymmetric priors in $\mathbb{R}^3$

Clearly what we would like to be able to incorporate into the joint distribution, $p(x, y, z)$, is *a priori* information with regard to the pairwise prior distributions *e.g.* $|x-y| < |x-z|$. As briefly discussed in Section 5.4.2 the idea is to change the diagonal elements of I so

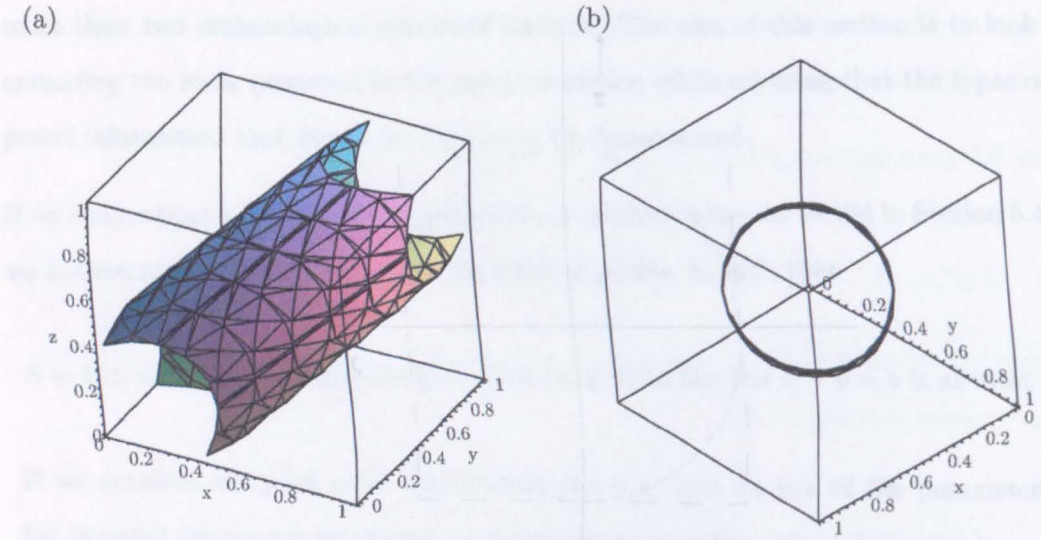


Figure 5.9: The joint prior distribution $p(x, y, z)$ resulting from Equation 5.11, when $u = 0.4$.

that we can scale u according to the *a priori* information that may arise *e.g.*

$$M = \begin{pmatrix} 2 & 0 & 0 \\ 0 & 1 & 0 \\ 0 & 0 & 1 \end{pmatrix}.$$

By replacing I with M in Equation 5.11 we can see what affect that this has on the form of the pairwise distributions. Equation 5.11 becomes

$$\frac{10}{9}x^2 - \frac{10}{9}xy - \frac{10}{9}xz + \frac{7}{9}y^2 - \frac{4}{9}yz + \frac{7}{9}z^2 \leq u^2. \tag{5.13}$$

Figure 5.11 is given to help visualize the three joint pairwise prior distributions, $p(x, y)$, $p(x, z)$ and $p(y, z)$, that arise from Equation 5.13 *i.e.* when I is replaced with M in Equation 5.11. There are two noticeable differences. Firstly, the shape of the pairwise prior distributions, $p(x, y)$ and p(x, z) are more elongated than $p(y, z)$. Secondly, that $p(x, y)$ is not symmetric about the line $x = y$; similarly for $p(x, z)$.

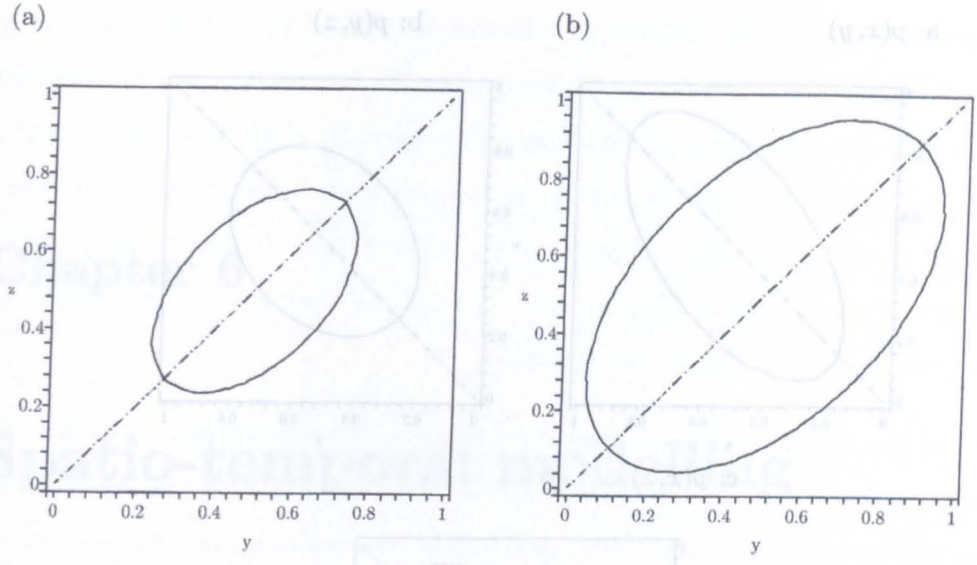


Figure 5.10: The joint pairwise prior distribution, $p(y, z)$ resulting from Equation 5.12, with varying values of u (a) $u = 0.3$ and (b) $u = 0.4$.

Clearly, changing more than one of the diagonal elements of I will result in all three pairwise prior distributions differing. However, before we can consider using these ideas, we need to be able to understand how the changes we make in I relate to the *a priori* information that we wish to incorporate. For example if we wish to incorporate $|x-y| \leq 0.1$, $|x-z| \leq 0.3$ and $|y-z| \leq 0.2$ what value of u do we use and what changes do we make to I?

This is where the difficulties lie. In Section 5.4.3 we could clearly see that choosing different values of u and changing the diagonal elements of I would allow for a variety of pairwise prior distributions to be incorporated. Ideally, this is what we want to be able to do in higher dimensions too. Unfortunately, in higher dimensions it is very unclear how we can relate the changes we make to the elements of I to the *a priori* information that we want to incorporate. As a result, it seems difficult to use this approach in practice, and so pointless to extend the ideas in Sections 5.4.2 and 5.4.3 further to $\mathbb{R}^n$.

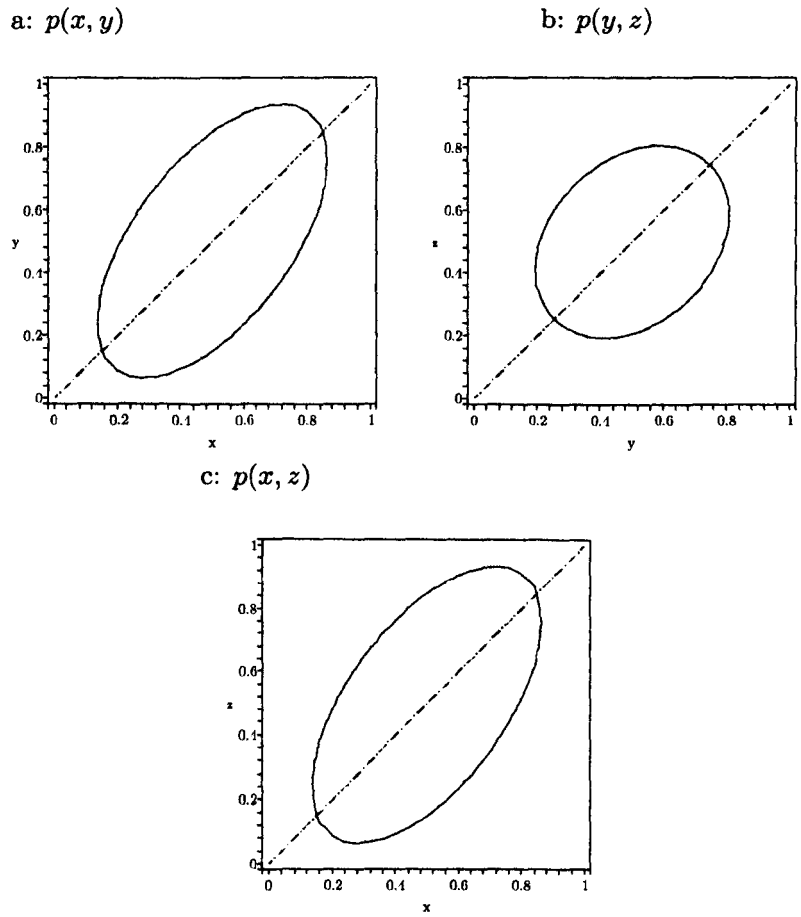


Figure 5.11: The joint pairwise prior distributions $p(x, y)$, $p(x, z)$ and $p(y, z)$ resulting from Equation 5.13, using a value of $u = 0.4$.

5.5 Summary

Although the ideas in this chapter seemed sensible, and may well be worth pursuing again in the future, for the purpose of this thesis none of the methods discussed in this chapter have been implemented. In Chapter 6, we tackle the same problem, but approach it from a different perspective.

Chapter 6

Spatio-temporal modelling

6.1 Introduction

The general problem is to understand when, how and why plants and animals spread into regions and landscapes that were not previously occupied. Such problems are of interest to those studying the response of plants and animals to climate change, for example the human reoccupation of NW Europe case study discussed in Chapter 4. In that particular case study archaeologists are interested to learn about how quickly NW Europe became reoccupied as the ice sheets retreated after the last ice age and what routes people took during the process.

When studying colonisation and recolonisation of past landscapes/environments, we often have a shortage of data points. For example in the reoccupation case study we have approximately 135 radiocarbon determinations, 41 of which are from the British Isles. As well as a lack of data points we also have a lack of archaeological information. However, on rare occasions we may find information preserved in lake beds, peat bogs or under the sea but as a result of towns being redeveloped, roads being built and sea levels changing we only have access to a sub-set of this information.

As there are currently no formal methods for tackling problems in a fully spatio-temporal

framework researchers either calibrate the radiocarbon determinations individually and then draw isochron maps or divide the landscape into regions and use the earliest available date in each region to determine the order in which the regions were re(colonised), as in Housley *et al.* (1997). In Chapter 4 we extended the latter idea by making use of the existing temporal tools, assuming that the eight regions were independent of one another, to allow us to calculate a probabilistic answer to the question "in which order were the regions reoccupied?".

In this chapter we aim to take the first steps towards fully spatio-temporal modelling by taking into account any spatial as well as the temporal information that arises from archaeological excavations, in order to combine data from related sites. We first revisit the human reoccupation of NW Europe case study, to discuss the type of *a priori* information that arises, before proposing a general spatio-temporal framework in which we can tackle such problems.

6.2 Example of an archaeological spatio-temporal problem

Initially we review the human reoccupation of NW Europe case, seen in Section 4.3. This represents a typical problem that is currently at best tackled as if it were a temporal problem but ideally we would like to tackle within a spatio-temporal framework. As seen in Section 4.3 this case study has been tackled a number of times in the published literature, each time aiming at interpreting the data in a more coherent manner than the previous, yet each time, ignoring any spatial information that we would ideally like to incorporate. The primary objective is to estimate the earliest date of reoccupation in each region under study and to calculate the probabilities of particular orderings in which reoccupation occurred.

Our first attempt to tackle this problem (detailed in Section 4.3.5) made simple use of the temporal tools from which we could make inferences on the ordering of the regions. However, as seen in Section 4.3.5, the most likely ordering accounted for only a very

small proportion (approximately 2%) of the total posterior probability under both the trapezium and sigmodial models. As a result, it is very difficult to make any inferences on the movement of humans across NW Europe after the last glacial period.

We believe that incorporating previously ignored spatial information into the analysis may make a huge difference to the archaeological interpretations that we can draw from the data. For example, "is there a pair of regions whose ordering is resolved by incorporating spatial information"?

As seen in Section 4.3, Housley *et al.* (1997) divides NW Europe into 8 regions (see Figure 6.1 [1].) It seems sensible to assume that those regions closer together (spatially) were more likely to be reoccupied at similar times compared to those regions further apart. Also, it seems sensible to assume that the first areas to be reoccupied were those furthest south. Unfortunately, for the human reoccupation of NW Europe case study, we do not have any quantitative spatial information, such as exact location of the sites, in terms of latitude and longitude readings or distances between sites. As a result, we use a similar approach to that adopted by Housley *et al.* (1997). Although, it is very crude, we use Figure 6.1 which indicates the eight different regions defined in Housley *et al.* (1997) to measure the approximate Euclidean distances between the pairs of regions (in kms, see Section 7.3). Given these distances we need to think about how we can incorporate the spatial information into the modelling framework.

In this section we have discussed an example that we would like to be able to tackle in a fully spatio-temporal framework. As a result, the following section presents our first thoughts for ways in which we could incorporate both the temporal and spatial information, that arises from the reoccupation case study.

[1]Note that the during the time period of interest the coastal line of Europe differed considerably from today's coastal line *e.g.* the British Isles was connected to the rest of Europe

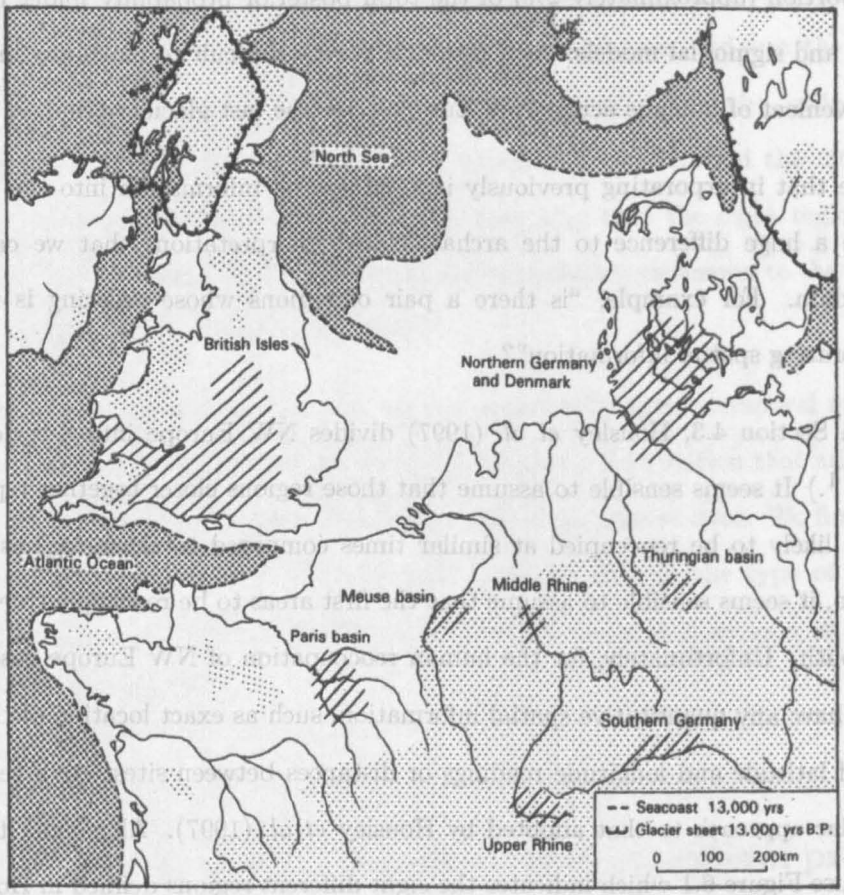


Figure 6.1: The eight regions of Late glacial NW Europe (Figure taken from Housley *et al.*, 1997), where the diagonal lines are taken to represent the individual regions.

6.3 Incorporating spatio-temporal information

In this section we propose a novel approach, which extends our basic model for chronology building (see Section 3.4.4), for tackling archaeological calibration problems in a fully spatio-temporal framework.

We aim at constructing a general framework in which we can incorporate a range of spatial as well as temporal information into the *a priori* information in a structured way. For example, in the human reoccupation of NW Europe case study, we are primarily interested in the first date of reoccupation (α's) for each region, from which we can

derive the order in which the regions were reoccupied. As it seems sensible to assume that regions closer together spatially were more likely to be reoccupied at similar times, we seek to define a joint prior distribution $p(\boldsymbol{\alpha})$ which incorporates the spatial structure between the α's. Note that the methodology proposed in the following section can be easily generalized *i.e* if we want to express the spatial structure between other parameters of interest such as β.

Our first thoughts for incorporating the spatial structure, in terms of the reoccupation case study, is to represent the prior for $\boldsymbol{\alpha} = \{\alpha_1, \ldots, \alpha_k\}$ as being uniform over S, where we define S to be the set

$$S = \bigcap_{j<k} \{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\} \tag{6.1}$$

and $c_{j,k}$ is defined to be a measure of difficulty of spread between regions j and k ($c_{j,k}$ may be ∞ in some cases and there may also be occasions in which we wish to constrain $\alpha_j > \alpha_k$.)

An alternative way of writing this down, is to say

$$p(\boldsymbol{\alpha}) \propto \prod_{j<k} I\{|\alpha_j - \alpha_k| \leq c_{j,k}\} \tag{6.2}$$

where $I(\cdot)$ represents an indicator function. If we were happy to assume a uniform ease of spread, then $c_{j,k}$ could be proportional to Euclidean distance ($d_{j,k}$) between regions j and k.

The two following subsections look at both the marginal and conditional prior distributions induced from using the joint prior distributions, $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$ and $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta})$, under both the conventional uniform and the trapezium models. We then move on to look at the corresponding distributions induced when using the joint prior distributions in which a spatial component is added into the model.

6.3.1 The prior distributions when no spatial dependence is incorporated

Conventional uniform model

When modelling multiple phases of activity with respect to the uniform model, the convention is to assume a particularly simple form of joint prior for $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$. That is, all values of α_j and β_j that satisfy a given set of constraints, C, are equally likely. The joint prior for $\boldsymbol{\alpha}$ and $\boldsymbol{\beta}$ is given by

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}) \propto I_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{6.3}$$

where

$$I_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in C \\ 0 & \text{otherwise} \end{cases}$$

where C is the set of values of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ which satisfy some given constraints *e.g.* the ordering of the parameters.

If there is no prior ordering between regions, that is to say each region is assumed to be independent of the others, then the conditional prior distributions[2] for α_j and β_j can be written as

$$\alpha_j | \beta_j \sim U(\beta_j, A) \quad \text{and} \quad \beta_j | \alpha_j \sim U(P, \alpha_j)$$

where A represents the maximum outer limit and P the lower outer limit *i.e.* $A < \alpha_j < \beta_j < P$.

Figure 6.2 is given to illustrate the properties of the marginal prior distributions $p(\alpha_j)$ and $p(\beta_j)$ under the conventional uniform model, when no spatial dependence has been incorporated into the model.

[2]We don't generally tend to write down the conditional priors explicitly, but as the conditional priors are of importance when incorporating the spatial structure, we introduce them here to enable us to make comparisons later in the chapter.

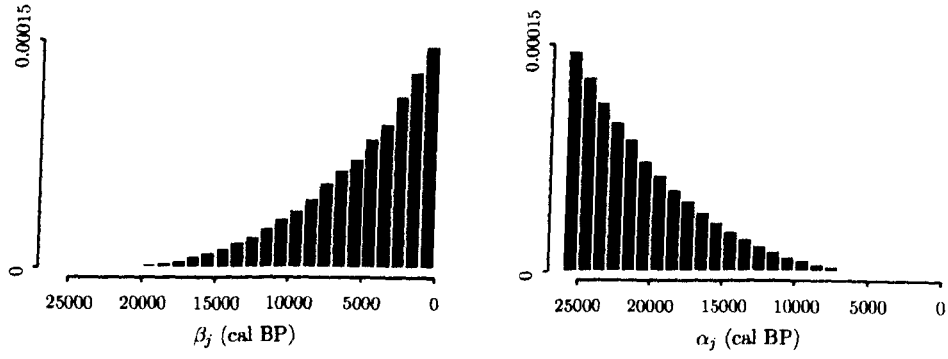


Figure 6.2: The marginal prior distributions, $p(\alpha_j)$ and $p(\beta_j)$, induced by Equation 6.3 over the period [0,26000] cal BP.

Trapezium model

When modelling multiple phases of activity with respect to the trapezium model, we have two additional parameters γ_j and δ_j. However, we use a similar form of joint prior (as used for the conventional uniform model), that assumes that all values of $\alpha_j, \gamma_j, \delta_j$ and β_j are equally likely, given they satisfy the set of constraints C. That is

$$p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \propto I_C(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \tag{6.4}$$

where

$$I_C(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \in C \\ 0 & \text{otherwise} \end{cases}$$

where C is the set of values of $\boldsymbol{\alpha} = (\alpha_1, \ldots, \alpha_m)$, $\boldsymbol{\gamma} = (\gamma_1, \ldots, \gamma_m)$, $\boldsymbol{\delta} = (\delta_1, \ldots, \delta_m)$ and $\boldsymbol{\beta} = (\beta_1, \ldots, \beta_m)$ which satisfy some given constraints *e.g.* the ordering of the parameters.

If we again assume that there is no prior ordering between regions *i.e.* the chronology of each region is assumed to be independent of the others. Then the conditional prior distributions, induced from Equation 6.4, can be written as

$$\alpha_j|\gamma_j, \delta_j, \beta_j \sim U(\gamma_j, A), \quad \gamma_j|\alpha_j, \delta_j, \beta_j \sim U(\delta_j, \alpha_j), \quad \delta_j|\alpha_j, \gamma_j, \beta_j \sim U(\beta_j, \gamma_j)$$

and $\beta_j|\alpha_j, \gamma_j, \delta_j \sim U(P, \delta_j)$.

Figure 6.3 represents the properties of the four marginal prior distributions $p(\alpha_j)$, $p(\gamma_j)$, $p(\delta_j)$ and $p(\beta_j)$ under the trapezium model when no spatial dependence has been incorporated into the model.

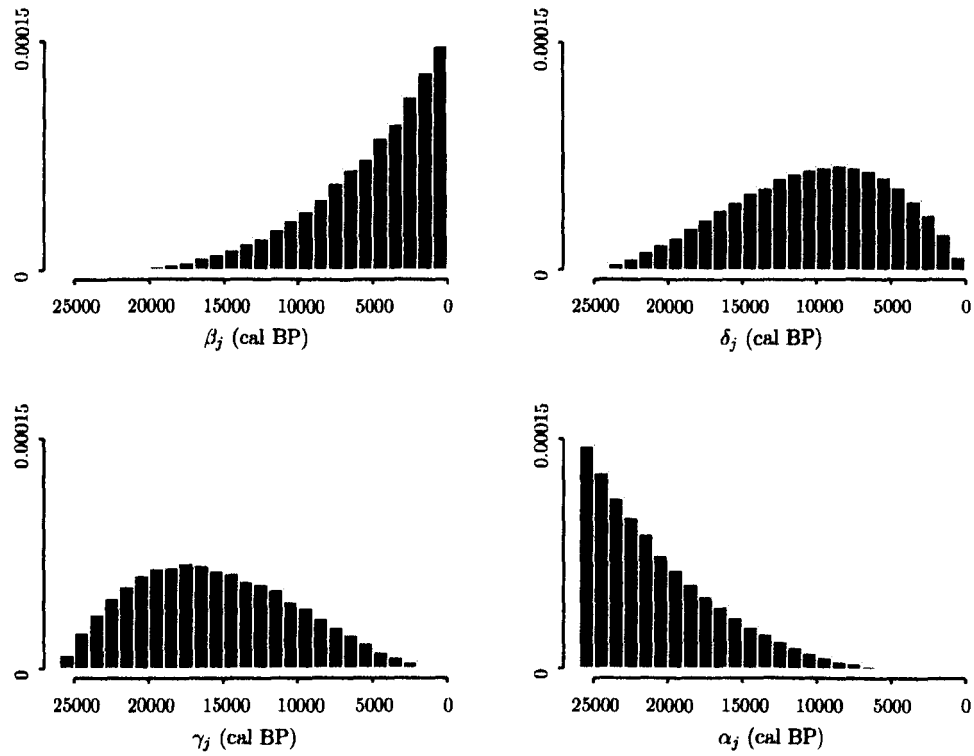


Figure 6.3: The marginal prior distributions, $p(\alpha_j)$, $p(\gamma_j)$, $p(\delta_j)$ and $p(\beta_j)$, under the trapezium model for the period [0,26000] cal BP.

6.3.2 The prior distributions when incorporating spatial dependence between regions

In this section we look at the differences between $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$, $p(\alpha_j|\beta_j)$ and $p(\beta_j|\alpha_j)$, under the uniform model when we incorporate a spatial structure between the α's as discussed in Section 6.3 and those in the previous section where no spatial dependence was

incorporated. Similar comparisons are also made for the trapezium model.

When incorporating the spatial dependence we can use a similar form of joint prior, $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$, which we can write as

$$p(\boldsymbol{\alpha}, \boldsymbol{\beta}) \propto I_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{6.5}$$

where

$$I_C(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in C \\ 0 & \text{otherwise} \end{cases}$$

where $C = \{\boldsymbol{\alpha}, \boldsymbol{\beta} : \boldsymbol{\alpha} \in S,\ \beta_j < \alpha_j\ \forall\ j\}$ and $S = \bigcap_{j<k}\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\}$.

As discussed on page 135, $c_{j,k}$ is defined to be a measure of difficulty of spread between regions j and k and if we assume a uniform ease of spread then $c_{j,k}$ could be proportional to the Euclidean distances between regions j and k.

Defining the joint prior distribution as in Equation 6.5, results in a change in the conditional prior distribution $p(\alpha_j|\beta_j)$ while the conditional distribution $p(\beta_j|\alpha_j)$ remains unchanged. That is

$$\alpha_j|\beta_j \sim U(\max(\beta_j, \max\Big(\alpha_k - \frac{d_{j,k}}{s}\Big)k \neq j), \min(A, \min\Big(\alpha_k + \frac{d_{j,k}}{s}\Big)k \neq j) \tag{6.6}$$

$$\text{and} \quad \beta_j|\alpha_j \sim U(P, \alpha_j)$$

where s represents our 'minimum speed' parameter. We introduce the parameter s (in kms per year) so that we can represent the maximum length of time (in years) in which it will take the hunter gathers to move from one region to another, assuming a uniform ease of spread. Note we could extend the spatio-temporal models further by treating s as an unknown parameter, however, for simplicity we chose to treat s as a constant. Further discussion on extending the spatio-temporal models can be found in Section 8.2.2.

One of the easiest ways to ensure that the conditional prior distributions have an intuitive archaeological meaning, with respect to the problem at hand, is visually. As a result we

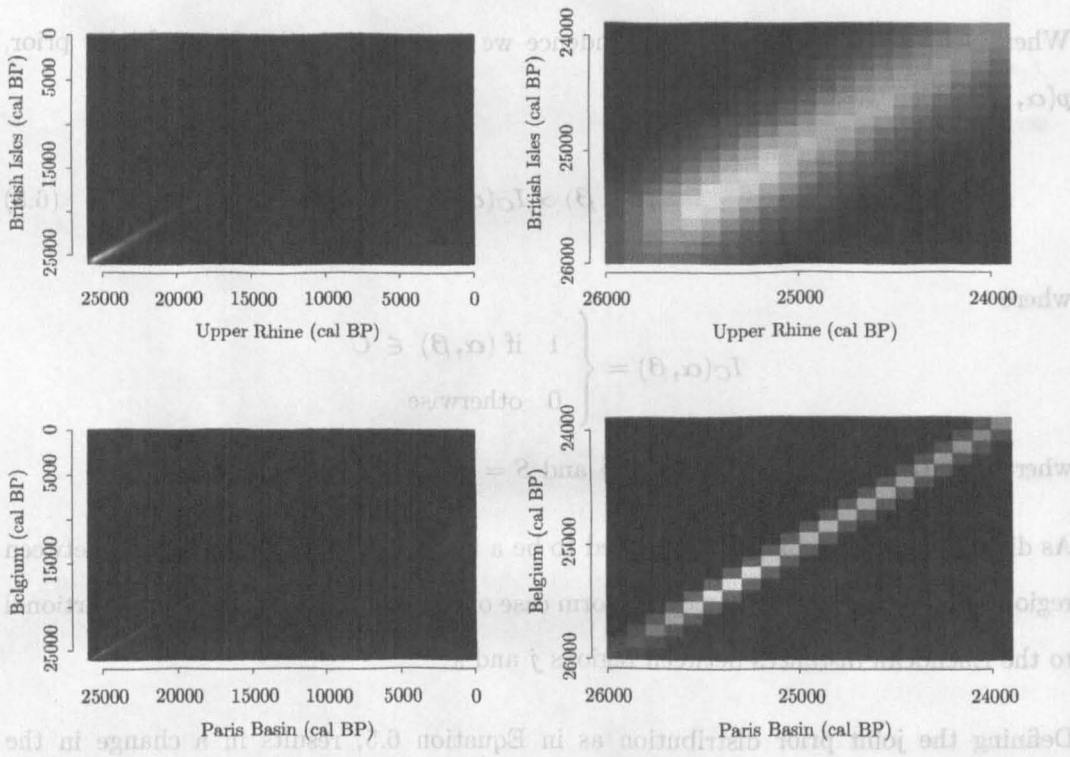


Figure 6.4: The joint prior distribution $p(\alpha_{UR}, \alpha_{BI})$ for The Upper Rhine and the British Isles, top two plots, and the joint prior distribution $p(\alpha_{PB}, \alpha_B)$ for the Paris Basin and Belgium, bottom two plots.

have simulated from the joint prior distribution $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in order to see whether the joint priors $p(\alpha_j, \alpha_k)$ and the conditional prior distributions have the desired properties.

Figure 6.4 represents two examples of $p(\alpha_j, \alpha_k)$, one in which the regions are known to be spatially close together, *e.g* the Paris Basin and Belgium, and one in which the regions are further apart, *e.g* the Upper Rhine and the British Isles. The lighter the shading, in Figure 6.4, the higher the probability density *i.e.* the black areas represents a probability density of zero. It is clear, from the top and bottom right plots which zoom in on these areas, that the spatial dependence between α_j and α_k behaves as we expected.

However, as seen in the top and bottom left plots the lighter shaded areas only cover the region approximately 18000-26000 cal BP. This suggests, as a result of incorporating

a spatial dependence between the regions, that the α_j's are being forced towards the upper limit A *i.e.* forcing the α_j's towards being very old.

Figure 6.5 illustrates the marginal prior distributions $p(\alpha_j)$ and $p(\beta_j)$. If we compare these plots to Figure 6.2, we can see more clearly the impact of the spatial dependence and how it has forced the α_j's towards A. The same property occurs for the trapezium model when we use a similar form of joint prior for $p(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta})$, see Figure 6.6.

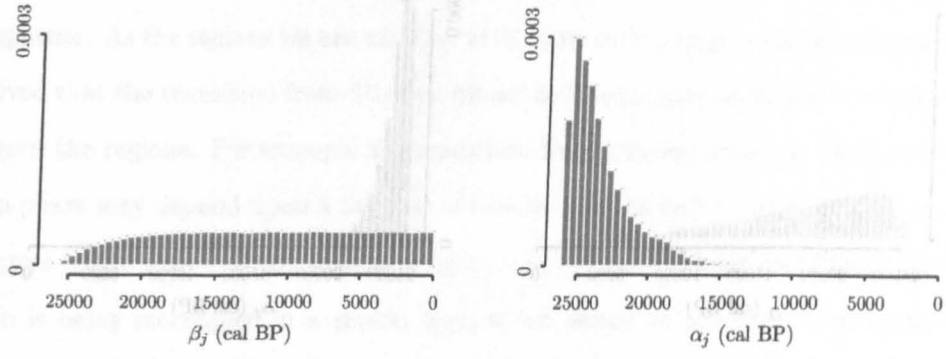


Figure 6.5: The marginal prior distributions, $p(\alpha_j)$ and $p(\beta_j)$, under the conventional uniform model when a spatial dependence between the α_j's is incorporated, in the period [0,26000] cal BP where $s = 1$.

Although the marginal distributions have an intuitive archaeological interpretation, we feel that they represent unrealistic dates for both the α_j's and β_j's under the uniform model and similarly for the trapezium model. As a result we feel that incorporating the spatial dependence between the regions, as defined in Equation 6.5, is not acceptable, and therefore an alternative approach is required.

The following section initially discusses why the marginal distributions induced by the joint prior, $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$, represent unrealistic dates for both the α_j's and β_j's, before moving on to look at alternative ways in which we can incorporate the spatial dependence between the α's.

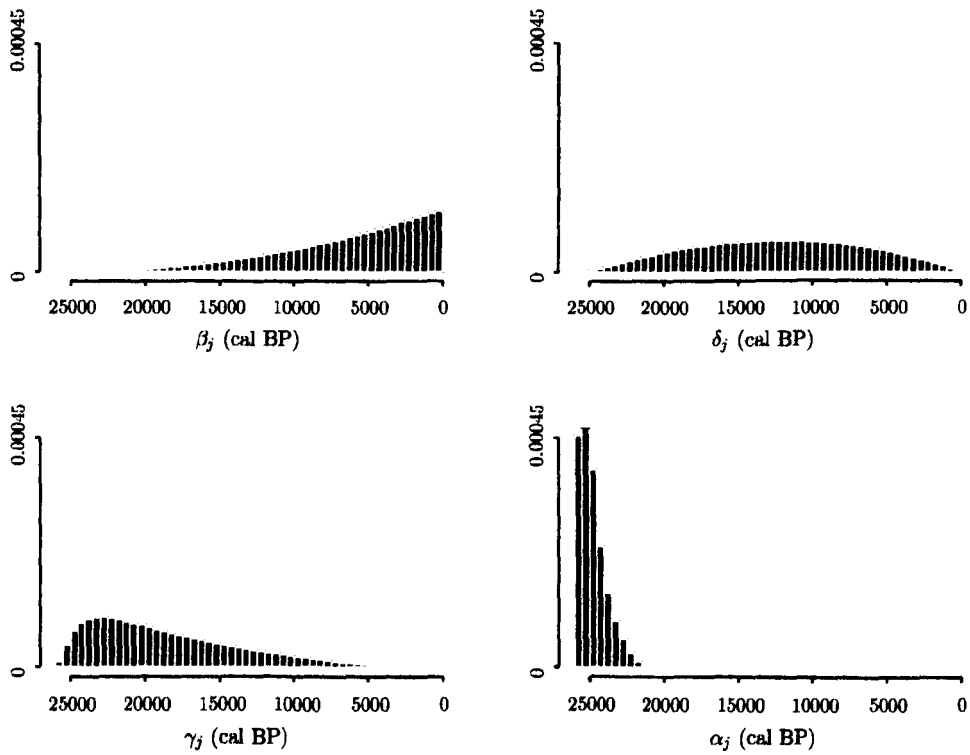


Figure 6.6: The marginal prior distributions, $p(\alpha_j)$, $p(\gamma_j)$, $p(\delta_j)$ and $p(\beta_j)$, under the trapezium model when a spatial dependence between the α_j's is incorporated, in the period [0,26000] cal BP where $s = 1$.

6.4 Alternative spatio-temporal priors

As discussed in the previous subsection the joint prior, $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$, as defined in Equation 6.5 induces unrealistic dates for the marginal prior distributions of interest, and in particular for $p(\alpha_j|\beta_j)$. If we think about the trapezium model and take the extreme case in which all the α_j's were known, we would effectively have eight different γ's, with one α forced to be older than all the γ's. Hence, on average that α would be very old.

So, how can we develop a spatio-temporal framework, in which we can incorporate the spatial dependence between the α's without changing the marginal prior distributions too much?

6.4.1 First idea - constrain $\alpha_j - \gamma_j \leq c$ years

Our next thoughts were to constrain $\alpha_j - \gamma_j \leq c$ years, that is to constrain the length of time it takes from a region first being reoccupied to becoming fully established. However, although this might appear to be sensible for a number of applications, after further thought it is probably not appropriate for the human reoccupation of NW Europe case study. In this case study α represents the first date of reoccupation, that is when a few small hunting parties have moved in to explore and exploit the previously unpopulated area. The γ represents the establishment of a larger, but possibly not permanent, occupation. As the regions we are working with cover such a large geographical area it is believed that the transition from 'Pioneer phase' to 'Residential camp phase' would vary between the regions. For example the transition from Pioneer phase to the Residential camp phase may depend upon a number of factors, such as ease to reach the region, the resources in the region *etc.*By constraining $\alpha_j - \gamma_j \leq c$ years we are assuming that each region is being reoccupied in a similar way, which seems an unrealistic assumption to make.

6.4.2 Second idea - define the prior differently

A second possibility is to define the joint prior differently. In the human reoccupation case study we are primarily interested in the first date of reoccupation of NW Europe, within each region. As this is the case it seems sensible to incorporate the spatial dependence between the α's. So our initial thoughts were to define a joint prior, $p(\boldsymbol{\alpha})$, which incorporates the spatial dependence, then define a prior for β_j conditionally on α_j, *i.e.* $p(\beta_j|\alpha_j)$. So $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$ is defined by $p(\boldsymbol{\alpha})$ and $p(\boldsymbol{\beta}|\boldsymbol{\alpha})$, rather than implicity by $p(\alpha|\beta)$ and $p(\beta|\alpha)$.

Conventional uniform model

We no longer define a joint prior, $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$, as in the previous sections. We initially define

a prior for $\boldsymbol{\alpha}$ which we can write as

$$p(\boldsymbol{\alpha}) \propto I_S(\boldsymbol{\alpha}) \tag{6.7}$$

where

$$I_S(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} \in S \\ 0 & \text{otherwise} \end{cases}$$

and $S = \bigcap_{j<k}\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\}$.

Then secondly we define a prior for each β_j conditionally on α_j such as

$$p(\boldsymbol{\beta}|\boldsymbol{\alpha}) \propto I_B(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{6.8}$$

where

$$I_B(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in B \\ 0 & \text{otherwise} \end{cases}$$

and $B = \{\boldsymbol{\alpha}, \boldsymbol{\beta} : \boldsymbol{\alpha} \in S, P < \beta_j < \alpha_j < A \; \forall j\}$.

Defining the prior in two parts, as seen above, results in the following marginal prior for α_j and the conditional prior $p(\beta_j|\alpha_j)$,

$$\alpha_j \sim U(\max(P, \max\Big(\alpha_k - \frac{d_{j,k}}{s}\Big)k \neq j), \min(A, \min\Big(\alpha_k + \frac{d_{j,k}}{s}\Big)k \neq j) \tag{6.9}$$

$$\text{and} \quad \beta_j|\alpha_j \sim U(P, \alpha_j).$$

Comparing Equation 6.9 with Equation 6.6 we can see that the conditional prior $p(\beta_j|\alpha_j)$ remains unchanged, but the marginal prior for $\boldsymbol{\alpha}$ differs as we no longer need to condition on β_j. If we compare Figure 6.7, which shows the marginal prior distributions for α_j and β_j, to Figure 5.3 from Chapter 5, we can see that they essentially represent the same distribution. That is the marginal prior distribution for α_j is essentially uniform over the region [0,26000] except at the end of the region. We believe the marginal priors induced from $p(\boldsymbol{\alpha})$ and the conditional priors $p(\beta_j|\alpha_j)$ represent much more realistic dates

for the parameters of interest and at the same time have a meaningful archaeological interpretation.

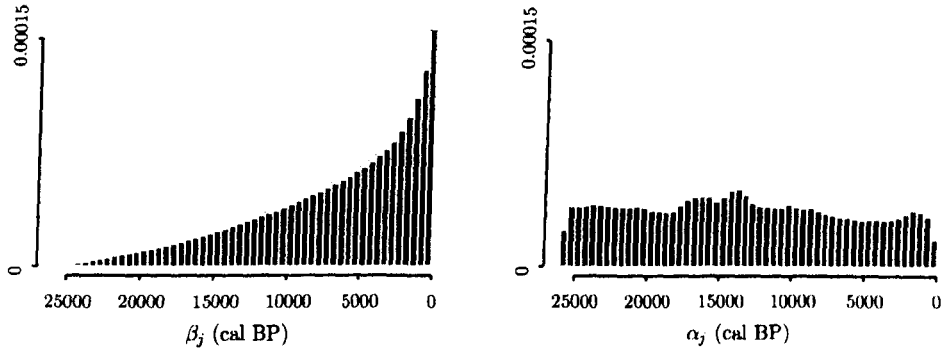


Figure 6.7: The marginal prior distributions for α_j and β_j, over the period [0,26000] cal BP (where $s = 1$), when assuming a uniform rate of deposition.

Trapezium model

Under the trapezium model we use the same form of prior as in previous section, that is we initially define the joint prior for $\boldsymbol{\alpha}$ as

$$p(\boldsymbol{\alpha}) \propto I_S(\boldsymbol{\alpha}) \tag{6.10}$$

where

$$I_S(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} \in S \\ 0 & \text{otherwise} \end{cases}$$

and $S = \bigcap_{j<k}\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\}$. We then condition on α_j and define a joint conditional prior distribution as

$$p(\boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}|\boldsymbol{\alpha}) \propto I_B(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \tag{6.11}$$

where

$$I_B(\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta}) \in B \\ 0 & \text{otherwise} \end{cases}$$

and $B = \{\boldsymbol{\alpha}, \boldsymbol{\gamma}, \boldsymbol{\delta}, \boldsymbol{\beta} : \boldsymbol{\alpha} \in S, P < \beta_j \leq \delta_j \leq \gamma_j \leq \alpha_j < A \; \forall j\}$.

We can see from Figure 6.8, which represents the marginal prior distributions for $\alpha_j, \gamma_j, \delta_j$ and β_j, that the distributions represent more realistic dates for the parameters than those shown in Figure 6.6.

As a result, in the following sections we will discuss, for both the uniform and trapezium models, the full details for modelling archaeological calibration problems as fully spatio-temporal problems, as well as discussing how we can implement them by extending the Metropolis-Hastings algorithm discussed in Section 3.5.

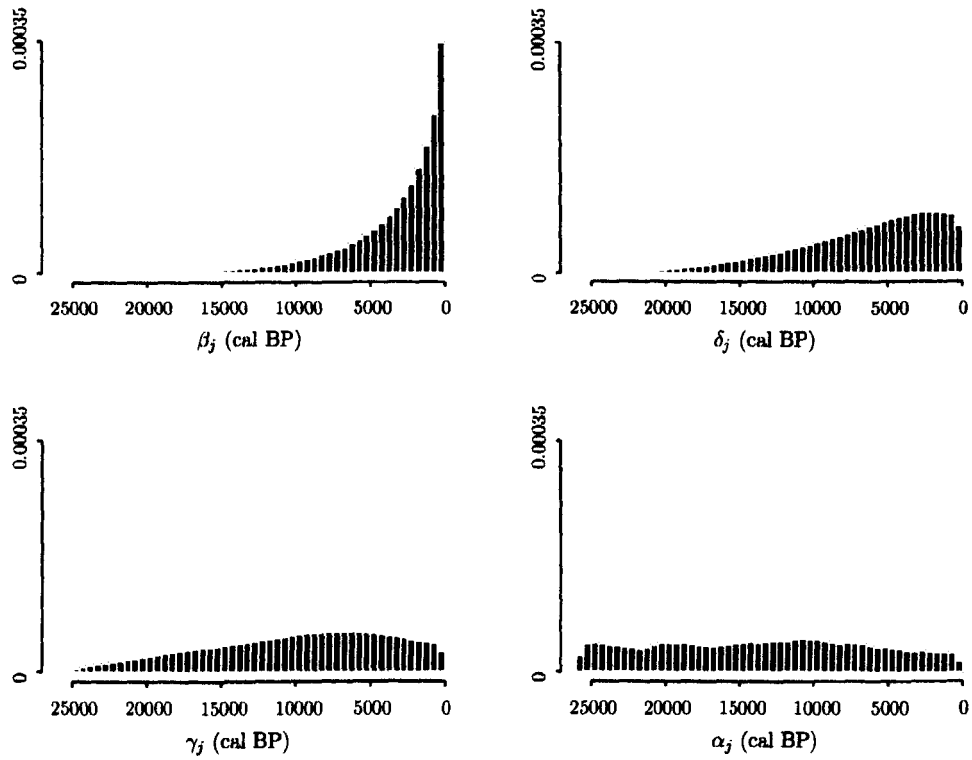


Figure 6.8: The marginal prior distributions for $\alpha_j, \gamma_j, \delta_j$ and β_j when assuming a trapezium rate of deposition and a spatial dependence between the α_j's is incorporated, for the period [0,26000] cal BP, where $s = 1$.

6.5 Fully spatio-temporal modelling

This section of the chapter is split into two sub-sections. The first subsection sets up the spatio-temporal model (using the ideas presented in Section 6.4.2), when assuming a uniform prior distribution for the deposition rate of datable material, within a multiple phase framework. We also discuss the changes we need to make to the Metropolis-Hastings algorithm (detailed in Section 3.5) as a result of incorporating the spatial dependence between the regions. The second subsection follows the same format, yet discusses the case in which we assume a trapezium prior distribution for the deposition rate of datable material within an archaeological phase.

6.5.1 Uniform spatio-temporal model

The ideas in this section are very closely linked to those in Section 3.4.4.

Let n_j represent the number of sample assigned to the jth phase. The ith radiocarbon determination in the jth phase is represented by $x_{i,j} \pm \sigma_{i,j}$ each associated with $\theta_{i,j}$, the corresponding calendar date (cal BP). From Section 3.4.4, the likelihood $p(x_{i,j}|\theta_{i,j})$ can be written as

$$p(x_{i,j}|\theta_{i,j}) \propto \frac{1}{w_{i,j}(\theta_{i,j})} \exp\left\{-\frac{(x_{i,j}-\mu(\theta_{i,j}))^2}{2w_{i,j}^2(\theta_{i,j})}\right\} \tag{6.12}$$

where $w_{i,j}^2(\theta_{i,j}) = \sigma_{i,j}^2 + \sigma_{i,j}^2(\theta_{i,j})$.

We define m to represent the number of phases of activity identified by archaeologists. Consequently we can define α_j and β_j to represent the beginning and end of phase j for $j = 1, \ldots, m$. As seen in Section 3.4.3, as we are working in calendar years BP, the parameters are subject to the constraint $A > \alpha_j > \beta_j > P$.

We assume that the material suitable for dating within phase j was deposited uniformly between the start and the end of phase, that is to say

$$\theta_{i,j}|\alpha_j, \beta_j \sim U(\alpha_j, \beta_j) \quad \text{for } i = 1, \ldots, n \text{ and } j = 1, \ldots, m. \tag{6.13}$$

As a result, $p(\boldsymbol{\theta_j}|\alpha, \beta)$, where $\boldsymbol{\theta_j}$ represents the set of θ's belonging to the jth phase, can be written as

$$p(\boldsymbol{\theta_j}|\alpha_j, \beta_j) \propto I_{D_j}(\boldsymbol{\theta_j}) \prod_{i=1}^{n_j} p(\theta_{i,j}|\alpha_j, \beta_j)$$

where

$$\mathrm{I}_{D_j}(\boldsymbol{\theta_j}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta_j} \in D_j \\ 0 & \text{otherwise} \end{cases}$$

and D_j represents the set of values that $\boldsymbol{\theta_j}$ can take within phase j such as stratigraphic ordering between the θ's.

The spatial component is incorporated by defining a joint prior for $\boldsymbol{\alpha}$ and then defining the conditional prior $p(\beta_j|\alpha_j)$, independently for each j. That is

$$p(\boldsymbol{\alpha}) \propto I_S(\boldsymbol{\alpha}) \tag{6.14}$$

where

$$I_S(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} \in S \\ 0 & \text{otherwise} \end{cases}$$

and $S = \bigcap_{j<k}\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\}$. By assuming a uniform ease of spread ($c_{j,k} \propto d_{j,k}$ - the approximate Euclidean distances between regions) we can rewrite S as $S = \bigcap_{j<k}\left\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq \frac{d_{j,k}}{s}\right\}$. As a result $p(\boldsymbol{\alpha})$ represents the maximum length of time it would take the Late glacial hunters to travel between pairwise regions.

We then define β_j conditionally on α_j by

$$p(\boldsymbol{\beta}|\boldsymbol{\alpha}) \propto I_B(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{6.15}$$

where

$$I_B(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in B \\ 0 & \text{otherwise} \end{cases}$$

and $B = \{\boldsymbol{\alpha}, \boldsymbol{\beta} : \boldsymbol{\alpha} \in S, P < \beta_j < \alpha_j < A \ \forall j\}$.

Evaluating the posterior distribution $p(\alpha_j|\beta_j, \boldsymbol{\theta}_j)$

As discussed in Section 3.5 we are using a single component Metropolis-Hastings algorithm to evaluate the posterior distributions of interest. Consequently the updating of all parameters except $\boldsymbol{\alpha}$ remains as in Section 3.5. This section discusses the changes that need to be made to the algorithm for updating α_j in order to implement a fully spatio-temporal model.

If we think back to the conventional uniform model, when updating α_j we are essentially evaluating the posterior

$$p(\alpha_j|\beta_j, \boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\alpha_j, \beta_j)p(\alpha_j|\beta_j). \tag{6.16}$$

The first term is the likelihood for the θ's (as seen in Equation 6.13) and the second term is uniform on (β_j, A), hence it does not directly affect the updating of α_j and can therefore be treated as a constant.

As a result, we update α_j as

$$p(\alpha_j|\beta_j, \boldsymbol{\theta}) \propto (\alpha_j - \beta_j)^{-n_j} \quad \text{for} \quad \alpha_j > \max\{\boldsymbol{\theta}, \beta_j\}.$$

where n_j represents the number of radiocarbon determinations in the jth phase.

So how does the updating of α_j change when we incorporate the spatial dependence into $p(\boldsymbol{\alpha})$? Equation 6.16 becomes

$$p(\alpha_j|\beta_j, \boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\alpha_j, \beta_j)p(\alpha_j|\beta_j) \tag{6.17}$$

$$p(\alpha_j|\beta_j, \boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\alpha_j, \beta_j)p(\alpha_j)p(\beta_j|\alpha_j).$$

Again, the first term is the likelihood for the θ's (as seen in Equation 6.13), and the second term is uniform on (P, A), hence it does not directly affect the updating on α_j and thus can be treated as constant. However, the third term is the conditional prior for $\beta_j|\alpha_j$, which we do not normally write down explicitly.

Conditional on α_j, the parameter β_j is distributed uniformly on (P, α_j) and the density is therefore

$$p(\beta_j|\alpha_j) = \begin{cases} (\alpha_j - P)^{-1} & \text{for } P < \beta_j < \alpha_j \\ 0 & \text{otherwise.} \end{cases}$$

Therefore, when incorporating spatial dependence, we update α_j as

$$p(\alpha_j|\beta_j, \boldsymbol{\theta}) \propto (\alpha_j - \beta_j)^{-n_j}(\alpha_j - P)^{-1} \quad \text{for} \quad \alpha_j > \max\{\boldsymbol{\theta}, \beta_j\}.$$

Implementing the uniform spatio-temporal model using a Metropolis-Hastings algorithm

In Section 3.5 we discussed how we implemented the trapezium model using a single component Metropolis-Hastings algorithm. In this section we extend the ideas in Section 3.5 by discussing the changes that we need to make to the algorithm so that the spatial component can be incorporated. No changes to the algorithm are required for β_j and $\boldsymbol{\theta}$, so they remain as in Section 3.5. However, we do need to make changes to the updating algorithm for α_j.

At each iteration t, the next state α_j^{t+1} is chosen by sampling a candidate point α_j' from a proposal distribution $q(\alpha_j'|\alpha_j^t)$. Again, a convenient choice of proposal distribution is a truncated Normal distribution, with the left tail truncated at the lower outer limit P and the right tail truncated at the upper outer limit A. We can express the proposal distribution as

$$q(\alpha_j'|\alpha_j^t) = \phi_{L,U}(\alpha_j'|\alpha_j^t, \sigma_{\alpha_j}).$$

where $L = \max\Big(\max(\boldsymbol{\theta_j}), \max\Big(\alpha_k - \frac{d_{j,k}}{s}\Big) \text{ for } j \neq k\Big)$ and $U = \min\Big(A, \min\Big(\alpha_k + \frac{d_{j,k}}{s}\Big)$ for $j \neq k\Big)$ and with mean α_j^t and standard deviation σ_{α_j}.

The candidate point α_j' is then accepted with probability

$$p(\alpha_j^t, \alpha_j') = \min\left(1, \frac{p(\alpha_j')p(\beta_j|\alpha_j')p(\boldsymbol{\theta}|\alpha_j'\beta_j)p(\boldsymbol{x}|\boldsymbol{\theta})q(\alpha_j^t|\alpha_j')}{p(\alpha_j^t)p(\beta_j|\alpha_j^t)p(\boldsymbol{\theta}|\alpha_j^t, \beta_j)p(\boldsymbol{x}|\boldsymbol{\theta})q(\alpha_j'|\alpha_j^t)}\right). \tag{6.18}$$

As the algorithm used is a single component Metropolis-Hastings algorithm, the term $p(\boldsymbol{x}|\boldsymbol{\theta})$ does not contribute to the updating of α_j and thus can be treated as a constant.

6.5.2 Trapezium model

The ideas in this section are linked to those in the previous section, with the main difference being the choice of prior distribution used to represent the *a priori* information arising about the rate at which the material was deposited within an archaeological phase.

As in Section 6.5.1, n_j represents the number of samples assigned to the jth phase and $x_{i,j} \pm \sigma_{i,j}$ represents the ith radiocarbon determination in the jth phase, associated with the corresponding calendar date $\theta_{i,j}$ (cal BP). The form of the likelihood $p(x_{i,j}|\theta_{i,j})$ remains unchanged and is given in Equation 6.13.

There are m phases of activity identified by archaeologists. In this model we define $\alpha_j, \gamma_j, \delta_j$ and β_j to represent the four parameters of the trapezium prior (see Section 3.4.1) for phase j ($j = 1, \ldots, m$) subject to the constraint $A > \alpha_j \geq \gamma_j \geq \delta_j \geq \beta_j > P$, as we are working in calendar years BP.

We represent the prior knowledge about the rate at which datable material is deposited within the jth phase by

$$p_{\text{Trap}}(\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j) \sim \text{Trap}\,(\alpha_j, \gamma_j, \delta_j, \beta_j) \tag{6.19}$$

as defined in Equations 3.5. Consequently $p(\boldsymbol{\theta_j}|\alpha, \gamma, \delta, \beta)$ [note $\boldsymbol{\theta_j}$ represents the set of

θ's belonging to the jth phase] can be written as

$$p(\boldsymbol{\theta_j}|\alpha_j,\gamma_j,\delta_j,\beta_j) \propto I_{D_j}(\boldsymbol{\theta_j}) \prod_{i=1}^{n_j} p_{\text{Trap}}(\theta_{i,j}|\alpha_j,\gamma_j,\delta_j,\beta_j)$$

where

$$\mathrm{I}_{D_j}(\boldsymbol{\theta_j}) = \begin{cases} 1 & \text{if } \boldsymbol{\theta_j} \in D_j \\ 0 & \text{otherwise} \end{cases}$$

but now D_j is used to represent the set of values that $\boldsymbol{\theta_j}$ can take within phase j.

To incorporate the spatial component into the model we define a joint prior for $\boldsymbol{\alpha}$ and then a joint conditional prior $p(\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta}|\boldsymbol{\alpha})$.

$$p(\boldsymbol{\alpha}) \propto I_S(\boldsymbol{\alpha}) \tag{6.20}$$

where

$$I_S(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} \in S \\ 0 & \text{otherwise} \end{cases}$$

and $S = \bigcap_{j<k}\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\}$. Again, we assume a uniform ease of spread, so that we can assume that $c_{j,k}$ is proportional to $d_{j,k}$ (the Euclidean distances between regions).

We then define γ_j, δ_j and β_j conditionally on α_j as

$$p(\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta}|\boldsymbol{\alpha}) \propto I_B(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta}) \tag{6.21}$$

where

$$I_B(\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta}) \in B \\ 0 & \text{otherwise} \end{cases}$$

and $B = \{\boldsymbol{\alpha},\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta} : \boldsymbol{\alpha} \in S, P < \beta_j \leq \delta_j \leq \gamma_j \leq \alpha_j < A\ \forall j\}$.

Evaluating the posterior distribution for α_j

Again, as a result of incorporating a spatial component into the model we need to make changes to the Metropolis-Hastings algorithm (see Section 3.5) for the updating of α_j. If we think about the trapezium model, in which there is no spatial component, when updating α_j we are evaluating

$$p(\alpha_j|\gamma_j,\delta_j,\beta_j,\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\alpha_j,\gamma_j,\delta_j,\beta_j)p(\alpha_j|\gamma_j,\delta_j,\beta_j). \tag{6.22}$$

The first term is the likelihood for the θ's under the trapezium model (see Equation 6.19). The second term is uniform on (γ_j,α_j), hence it does not directly affect the updating on α_j and thus can be treated as constant.

So how does the updating of α_j under the trapezium spatio-temporal model differ to the above? As seen in Section 6.4.2 we initially define $p(\boldsymbol{\alpha})$ and then condition on α_j, to get the joint conditional distribution $p(\boldsymbol{\gamma},\boldsymbol{\delta},\boldsymbol{\beta}|\boldsymbol{\alpha})$. Consequently we can write the posterior distribution for α_j as

$$p(\alpha_j|\gamma_j,\delta_j,\beta_j,\boldsymbol{\theta}) \propto p(\boldsymbol{\theta}|\alpha_j,\gamma_j,\delta_j,\beta_j)p(\alpha_j)p(\gamma_j,\delta_j,\beta_j|\alpha_j). \tag{6.23}$$

The first term is again the likelihood for the θ's, the second term is uniform on (P,A) hence it does not affect the updating on α_j and thus can be treated as a constant. However, the third term is a bit more tricky, as we do not normally write the joint conditional prior for $\gamma_j,\delta_j,\beta_j|\alpha_j$ explicitly. Conditional on α_j, the parameters γ_j,δ_j and β_j could be generated directly by generating 3 independent uniforms on (P,A) and then ordering them and calling the smallest β_j, the middle one δ_j and the largest one γ_j. However, we must note as a result of generating the parameters by this method, that there are actually $6 = 3!$ possible orderings that get mapped to the same values γ_j,δ_j

and β_j. As a result the density is written as

$$p(\gamma_j, \delta_j, \beta_j|\alpha_j) = \begin{cases} 6\,(\alpha_j - P)^{-3} & \text{for } P < \beta_j < \gamma_j < \delta_j < \alpha_j \\ 0 & \text{otherwise.} \end{cases}$$

The 6 does not directly affect the updating, since it is constant, but the $(\alpha_j - P)^{-3}$ is important.

Implementing the trapezium spatio-temporal model using a Metropolis-Hastings algorithm

This section links closely to Section 6.5.1 by discussing the changes that we need to make to the Metropolis-Hastings algorithm in Section 3.5 to allow for the spatial component to be incorporated. Again, no changes to the algorithm are required for $\gamma_j, \delta_j, \beta_j$ and $\boldsymbol{\theta}$, they remain as in Section 3.5. Changes to the algorithm are required for the updating of α_j. Therefore, this section focuses on the details of the algorithm used for updating α_j under the trapezium spatio-temporal model.

The proposal distribution, $q(\alpha'_j|\alpha_j^t)$, is needed to generate the next value, α_j^{t+1}, in the Markov chain given the current value of α_j^t. As seen in Section 6.5.1 a convenient choice of proposal distribution is the truncated Normal distribution as α_j is constrained to lie between P and A. We can express the proposal distribution as

$$q(\alpha'_j|\alpha_j^t) = \phi_{L,U}(\alpha'_j|\alpha_j^t, \sigma_{\alpha_j})$$

where $L = \max\Big(\max(\boldsymbol{\theta_j}), \gamma_j, \max\Big(\alpha_k - \frac{d_{j,k}}{s}\Big)$ for $j \neq k\Big)$ and $U = \min\Big(A, \min\Big(\alpha_k + \frac{d_{j,k}}{s}\Big)$ for $j \neq k\Big)$ with mean α_j^t and standard deviation σ_{α_j}.

The candidate point α'_j is then accepted with probability

$$p(\alpha_j^t, \alpha'_j) = \min\left(1, \frac{p(\alpha'_j)p(\gamma_j, \delta_j\beta_j|\alpha'_j)p(\boldsymbol{\theta}|\alpha'_j, \gamma_j, \delta_j, \beta_j)p(\boldsymbol{x}|\boldsymbol{\theta})q(\alpha_j^t|\alpha'_j)}{p(\alpha_j^t)p(\gamma_j, \delta_j, \beta_j|\alpha_j^t)p(\boldsymbol{\theta}|\alpha_j^t, \gamma_j, \delta_j, \beta_j)p(\boldsymbol{x}|\boldsymbol{\theta})q(\alpha'_j|\alpha_j^t)}\right). \quad (6.24)$$

As the algorithm used is a single component Metropolis-Hastings algorithm, again the term $p(\boldsymbol{x}|\boldsymbol{\theta})$ does not contribute to the updating of α_j and thus can be treated as a constant. Full details of the algorithm can be found in the C code in Appendix A.

6.6 Summary

The aim of this chapter was to present the first steps for modelling archaeological calibration problems within a fully spatio-temporal framework. The model proposed builds on the existing models (as detailed in Chapters 2 & 3) for archaeological chronology building.

In this chapter the model proposed is based around the human reoccupation of NW Europe case study. However, the model can easily be generalized, for example if we want to incorporate the spatial dependence between the β's rather than the α's.

As we now have a framework in which we can incorporate a spatial structure between regions, our next aim is to implement both the uniform and trapezium spatio-temporal model and to illustrate, via the human reoccupation of NW Europe case study, the difference in archaeological interpretations drawn from the data when a spatial component is incorporated.

Chapter 7

Spatio-temporal case study: human reoccupation of NW Europe

7.1 Introduction

In this chapter the case study data relating to the human reoccupation of North Western Europe at the end of the last Ice Age are revisited in order to illustrate the use of the spatio-temporal models developed in the previous chapter. Although we believe that the conventional uniform model is not a suitable approach for tackling problems concerning colonisation/recolonisation of landscapes, there are a number of applied researchers who would disagree. Thus for completeness, we implement two spatio-temporal models with differing *a priori* distributions for the rate at which material is deposited within the phase: the conventional uniform prior and the trapezium prior.

From this point forward, we will refer to a 'uniform spatio-temporal model' when a uniform prior on the rate of deposition of datable material has been implemented. Equivalently we will refer to a 'trapezium spatio-temporal model' when using a trapezium

prior.

7.2 Recapping the inferences obtained from the human reoccupation case study in Chapter 4

In this section we recap the inferences obtained when we investigated the human reoccupation case study in Chapter 4, before moving on to discuss ways in which we can tackle this case study within a spatio-temporal framework in Sections 7.3 & 7.4.

Our first approach to this case study took into account the main concern raised in Blackwell and Buck (2003). This being, that although the conventional uniform model is useful in many situations, *i.e.* short lived phases at single spatial locations, it is not believed to be the most sensible approach for problems concerning colonisation/recolonisation of past landscapes *i.e.* multiple phases at different spatial locations. This led Blackwell and Buck (2003) to suggest an alternative to the conventional uniform model, which we refer to as the sigmoidal prior (see Section 3.4.2), that allows landscapes to be established over a finite period of time, rather than instantaneously.

This suggestion indicated to us that there was a need to seek alternatives to the conventional uniform model in order for archaeologists to reliably represent their *a priori* information. Subsequently we proposed a range of alternative non-uniform *a priori* distributions, two of which (the trapezium and sigmoidal) would adequately represent the *a priori* beliefs with regard to the mechanics of the reoccupation process as described in Housley *et al.* (1997); see Section 4.3.2.

It became apparent from the work carried out in Chapter 4 that the trapezium and sigmoidal models led to very similar date estimates for the first date of reoccupation within each region. However, these differed considerably from those derived under the uniform model. In particular the trapezium and sigmoidal models resulted in earlier date estimates than those obtained under the conventional uniform model. Another

conclusion arising from Chapter 4 is that, given the similarities in date estimates obtained from the trapezium and sigmoidal models, it seems unnecessary to implement both models. We slightly favour the trapezium model over the sigmoidal model, as we believe it has a more intuitive archaeological interpretation (see Section 3.4).

Although we made inferences in Chapter 4 of both a temporal nature (*e.g.* the first date of reoccupation) and spatial nature (*e.g.* the order in which the regions may have been reoccupied) by utilizing relatively simple temporal models, we ignored any spatial information that was available. As a result, our next aim is to use the spatio-temporal model developed in the previous chapter to incorporate a spatial structure between the regions and illustrate, via the human reoccupation of NW Europe case study, the difference in results when adopting fully spatio-temporal models as opposed to purely temporal models.

7.3 Spatial information arising from the reoccupation case study

Housley *et al.* (1997) defines NW Europe as southern Scandinavia, Germany, the Netherlands, Belgium, Northern France and the British Isles (see Figure 6.1). It is believed that NW Europe was mostly unoccupied by humans during the glacial period (approximately 18,000 years ago). The questions of greatest interest to many archaeologists are: 'how long did this period of abandonment last?' and 'when did it begin and end?'

An additional interest, in Housley *et al.* (1997), was to approximate the rate of human expansion across NW Europe (in km/yr) after the ice sheets had retreated. The authors approached this question by assuming that the Upper Rhine was the point of origin and then approximated the distances (in kms) from the Upper Rhine to the other regions, using Figure 6.1. The authors then used these distances, along with their estimates for the first date of reoccupation in each region (see Section 4.3), to calculate the rate of

human expansion between the Upper Rhine and the other seven regions.

As we have previously discussed, it seems likely (*a priori*) that regions spatially close together (*e.g.* Paris Basin and Meuse Basin) are more likely to be reoccupied at similar times compared to those regions further apart (*e.g.* The Upper Rhine and the British Isles). Also, it seems sensible to assume that the first areas to be reoccupied, after the ice sheets had retreated, were those furthest south (*e.g.* The Upper Rhine). Unfortunately, as discussed in Chapter 6, there is no quantitative spatial information available for this case study. This led us to adopt a similar approach to that used by Housley *et al.* (1997). Figure 6.1 illustrates the eight regions of NW Europe, where the diagonal lines represent the approximate regional areas. As a result we measure the approximate Euclidean distances between the centroids of pairs of regions (in kms). Given these distances (see Table 7.1) the next step is to contemplate ways in which we can incorporate this information into the existing modelling framework. However, before this is possible we also need some indication of the ease of spread between the regions of NW Europe.

Region	URhine	Thuringian	SGermany	MRhine	Belgium	Paris	NGermany	BIsles
URhine	–	460	270	330	360	400	750	800
Thuringian	–	–	270	270	420	620	380	850
SGermany	–	–	–	300	430	590	620	940
MRhine	–	–	–	–	140	350	460	890
Belgium	–	–	–	–	–	230	520	480
Paris	–	–	–	–	–	–	780	460
NGermany	–	–	–	–	–	–	–	720
BIsles	–	–	–	–	–	–	–	–

Table 7.1: Euclidean distances (to the nearest 10 kms) between the centroids of pairs of regions measured from Figure 6.1.

In order to determine the ease of spread from one region to another in a coherent manner, one would need to know about the topography of NW Europe during the Late glacial period. For example, knowing the locations of river networks, mountain ranges and dense forests would make a huge impact on possible routes of migration *i.e.* the closest region may not necessarily be the easiest or quickest region to reach. Unfortunately we do not (currently) have access to such information and consequently rely on eliciting,

from either archaeologists or the applied literature, as much detail as possible of past landscapes.

In Housley *et al.* (1997) the authors make reference to the paper by Ammerman and Cavelli-Sforza (1973) in which the authors discuss the movement of European farmers across northern Europe. In this paper the authors estimated a 1km per year advance through northern Europe, based on the assumption that individuals disperse randomly in all directions. Housley *et al.* (1997) assume that the rate of expansion of the Late glacial hunters into NW Europe, after the ice sheets had retreated, would be similar to the movement of the European farmers (assuming that the Upper Rhine was the point of origin). Given the assumption of a 1km per year advance in all directions, we are essentially saying that we can assume a uniform ease of spread between the regions of NW Europe. As a result of this assumption we can say that $c_{j,k}$ (the measure of difficulty of spread between regions) is proportional to the approximate Euclidean distance, $d_{j,k}$ between the regions. We can then incorporate this spatial information into the modelling framework as detailed in Section 6.4.2.

The aim of the remainder of the chapter is to illustrate the difference in archaeological conclusions drawn from the data when implementing spatio-temporal models compared to those previously obtained. Namely, we will look at the difference in interpretations between

1. the conventional uniform model *versus* the uniform spatio-temporal model, and

2. the trapezium model *versus* the trapezium spatio-temporal model.

7.4 Setting up the spatio-temporal models

In this section we initially review the material presented in Section 3.5 which discusses the set up the reoccupation case study (within a multiple phase framework) for both the conventional uniform and trapezium model. We also discuss how to incorporate the

spatial information arising from the case study (detailed in Section 7.3). Full details of the set up of both spatio-temporal models can be found in Section 6.5, along with a discussion of the MCMC algorithms used to evaluate the posterior distributions of interest. Here we present a shorter explanation of how to incorporate the spatial structure into the modelling framework.

The uniform model

Each region is assumed to have its own phase of human reoccupation; the dates of these phases were taken to be independent in Section 3.5, but we will modify this in the spatio-temporal case. The earliest date (which we refer to as the beginning of the 'pioneer sub-phase') for the reoccupation phase in region j is represented by α_j cal BP ($j = 1, \ldots, 8$ denoting the number of regions). Similarly, the last date for the reoccupation phase in each region is labelled β_j cal BP. Under the uniform spatio-temporal model it is assumed that the samples suitable for radiocarbon dating within region j were deposited uniformly between the start (α_j) and the end (β_j) date of the reoccupation phase. Individual events are taken to be independent within and between phases, given the dates of the phases; this assumption remains unchanged in the spatio-temporal case.

The trapezium model

When implementing the trapezium spatio-temporal model we use the same notation as in the uniform spatio-temporal model, α_j and β_j cal BP, to represent the first and last date of the reoccupation phase within each region, respectively. There are also two extra parameters, γ_j which represents the beginning of the 'residential camp sub-phase' and δ_j which represents the end of the 'residential camp sub-phase'. When implementing the trapezium spatio-temporal model it is assumed that the material suitable for dating was deposited between α_j and β_j at a trapezoidal rate, that is to say $\theta_{i,j}|\alpha_j, \gamma_j, \delta_j, \beta_j \sim \text{Trap}(\alpha_j, \gamma_j, \delta_j, \beta_j)$; see Section 3.4.1.

Incorporating the spatial aspect

As seen in both Sections 6.5.1 and 6.5.2 the general framework for modelling

archaeological calibration problems follows a similar format to Section 3.4.4, the main difference that occurs is that we no longer define a joint prior $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$. In order to incorporate the spatial structure arising from the case study we initially define a joint prior for $\boldsymbol{\alpha}$ as

$$p(\boldsymbol{\alpha}) \propto I_S(\boldsymbol{\alpha}) \tag{7.1}$$

where

$$I_S(\boldsymbol{\alpha}) = \begin{cases} 1 & \text{if } \boldsymbol{\alpha} \in S \\ 0 & \text{otherwise} \end{cases}$$

and $S = \bigcap_{j<k}\{\boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq c_{j,k}\}$. We then define a conditional prior $p(\beta_j|\alpha_j)$, independently for each j.

$$p(\boldsymbol{\beta}|\boldsymbol{\alpha}) \propto I_B(\boldsymbol{\alpha}, \boldsymbol{\beta}) \tag{7.2}$$

where

$$I_B(\boldsymbol{\alpha}, \boldsymbol{\beta}) = \begin{cases} 1 & \text{if } (\boldsymbol{\alpha}, \boldsymbol{\beta}) \in B \\ 0 & \text{otherwise} \end{cases}$$

and $B = \{\boldsymbol{\alpha}, \boldsymbol{\beta} : \boldsymbol{\alpha} \in S, P < \beta_j < \alpha_j < A \; \forall j\}$.

This clearly represents the uniform spatio-temporal model. When assuming a trapezium rate of deposition we firstly define a joint prior for $\boldsymbol{\alpha}$ as in Equation 7.1 and then define a joint conditional prior $p(\beta_j, \delta_j, \gamma_j|\alpha_j)$ independently for each j.

As discussed in the previous subsection Housley *et al.* (1997) assumes a 1km per year advance, based on the assumption that individuals disperse randomly in all directions (as well as assuming that the Upper Rhine was the point of origin). Consequently, on the basis of this assumption we assume a uniform ease of spread across NW Europe *i.e.* we can now think of $c_{j,k}$ (measure of difficulty of spread) as being proportional to the approximate Euclidean distances ($d_{j,k}$) between the regions (see Table 7.1). We define

the set S

$$S = \bigcap_{j<k} \left\{ \boldsymbol{\alpha} : |\alpha_j - \alpha_k| \leq \frac{d_{j,k}}{s} \right\}$$

in order to represent the maximum length in terms of time in which the hunter gathers would travel between regions. We define the parameter s to be 1 *i.e.* to represent the 1km per year advance as suggested in Housley *et al.* (1997). However as s represents the 'minimum speed' parameter by defining $s = 1$ it also covers all speeds greater than 1km per year.

Although Housley *et al.* (1997) assumes that the Upper Rhine was the point of origin, and we agree that this seems a sensible assumption to make, we do not necessarily know whether this is true. Therefore we decided not to implement this assumption into our modeling framework. However, given the MCMC output we can calculate the posterior probability of the Upper Rhine being the first region to be reoccupied, which allows us to test their assumption.

7.5 Results: Conventional uniform model *versus* the uniform spatio-temporal model

We use a similar format to that of Section 4.3.5 to discuss the reinterpretation of the reoccupation data. That is, we initially discuss the first date of reoccupation of each region under the non-spatial and spatio-temporal models, before discussing the order in which the regions may have been reoccupied. An additional question that we investigate (in Section 7.6.1) is, 'given the order in which the regions were reoccupied, what can we learn about the routes of migration through NW Europe?'.

First date of reoccupation

As seen in Section 4.3 the primary objective is to estimate the earliest date of reoccupation of each region under study. We summarise these estimates by their 95% HPD regions (in Table 7.2) and visually by their marginal posterior distributions (in

Figure 7.1). Our main interest lies in comparing the estimates obtained from the uniform spatio-temporal model to those obtained in Section 4.3.5 (*i.e.* using the conventional uniform model).

Region	95% HPD interval for the date of first reoccupation (cal BP)	
	no spatial info	1 km per year
Upper Rhine	16500 - 19580	16160 - 16520
Thuringian Basin	16350 - 17650	16020 - 16390
Southern Germany	15300 - 16750	15940 - 16420
Middle Rhine	15150 - 15770	15880 - 16260
Belgium	15320 - 16900	15720 - 16100
Paris Basin	15030 - 16010	15720 - 16110
Northern Germany	14310 - 15440	15580 - 16030
British Isles	14670 - 15180	15260 - 15680

Table 7.2: The 95% HPD intervals for the first date of reoccupation of the eight regions under the conventional uniform when incorporating no spatial dependence (conventional uniform model) and the uniform spatio-temporal model when assuming a minimum of a 1km per year rate of expansion.

From Table 7.2 we can see that the HPD regions are much narrower under the spatio-temporal model compared to those under the conventional uniform model. The most noticeable difference occurs in the estimates for the Upper Rhine. Under the conventional uniform model the 95% HPD region is 16500-19580 cal BP which compares with 16160-16520 cal BP under the uniform spatio-temporal model, *i.e.* the spatio-temporal model results in the Upper Rhine being reoccupied later in time. Similarly, the estimates for the first date of reoccupation of the British Isles differs between the two models. Again, the 95% HPD region under the spatio-temporal model is narrower than under the conventional uniform model, yet this time the first date is estimated to be earlier in time.

Figure 7.1 shows the estimates for the marginal posterior distributions for the first dates of reoccupation in each region under the conventional uniform (red lines) and uniform spatio-temporal models (blue lines). From Figure 7.1 we can see as a result of incorporating the spatial dependence between the α_j's that the estimates for the first date of reoccupation are being pulled closer together.

The order in which reoccupation took place

As the chronological ordering is not clearly defined by looking at the HPD regions and marginal posterior plots from Figure 7.1 we use the two summary methods as detailed in Section 4.3.5, to discuss the order in which the regions were reoccupied. Table 7.3 gives the probabilities that each region is temporally ranked 1 (earliest) through to eight (latest) for both the conventional uniform and uniform spatio-temporal models. The numbers highlighted (in light grey for the conventional uniform model and dark grey for the spatio-temporal model) represent the most probable order in which the regions were reoccupied. Under both the conventional uniform and spatio-temporal model the Upper Rhine was the first region to be reoccupied with corresponding posterior probabilities of 0.75 and 0.90. The British Isles was the last region to be reoccupied, with a probability of 1.00 under the spatio-temporal model but the corresponding posterior probability under the conventional uniform model was only 0.39. Consequently, there is less uncertainty associated with the order in which the regions were reoccupied when the spatial structure has been imposed.

The second summary we use reports the posterior probability of particular orderings in which reoccupation might have occurred. Table C.1, in Appendix C, represents the ten most likely orderings of the eight regions. We summarize these by giving the most likely order in which reoccupation may have occurred under the two differing models. Under the conventional uniform model the most likely order is the Upper Rhine, Thuringian Basin, Belgium, Southern Germany, Paris, Middle Rhine, British Isles and Northern Germany with a posterior probability of 0.082, while under the spatio-temporal model the most likely order is the Upper Rhine, Thuringian basin, Southern Germany, Middle Rhine, Paris, Belgium, Northern Germany and the British Isles with a posterior probability of 0.19. It is clear that by implementing a spatio-temporal model the order in which some regions were reoccupied (such as northern Germany and the British Isles)

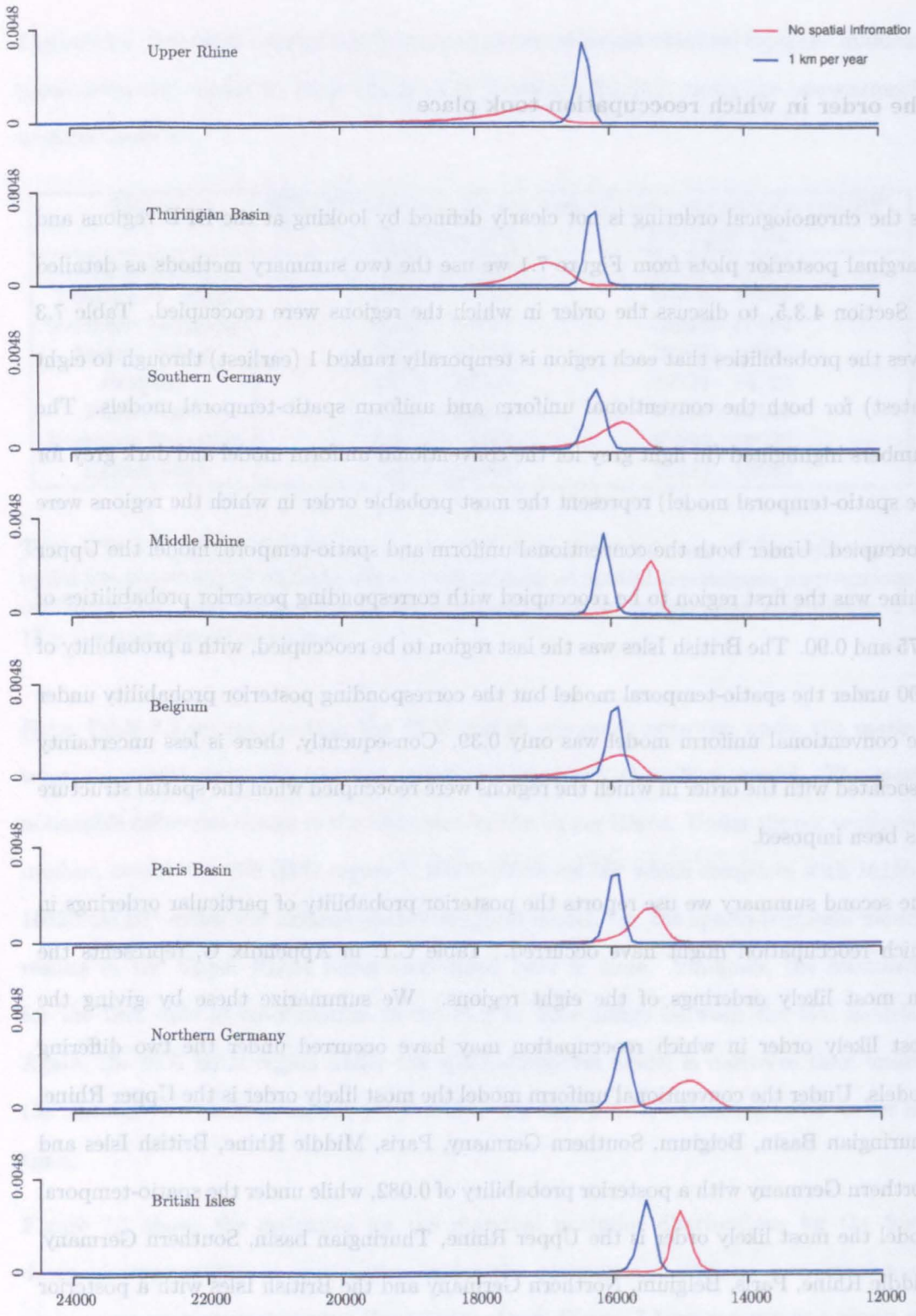


Figure 7.1: Marginal posterior distributions for the first date of reoccupation in each region under conventional uniform model (red) and the uniform spatio-temporal model (blue), assuming a minimum of a 1km per year expansion rate.

Region	1	2	3	4	5	6	7	8
Upper Rhine	0.75	0.23	0.02	0.00	0.00	0.00	0.00	0.00
	0.90	0.09	0.01	0.00	0.00	0.00	0.00	0.00
Thuringian Basin	0.22	0.71	0.07	0.00	0.00	0.00	0.00	0.00
	0.03	0.55	0.40	0.02	0.00	0.00	0.00	0.00
Southern Germany	0.01	0.03	0.38	0.42	0.13	0.03	0.00	0.00
	0.06	0.36	0.42	0.16	0.00	0.00	0.00	0.00
Middle Rhine	0.00	0.00	0.02	0.08	0.37	0.49	0.04	0.00
	0.00	0.01	0.17	0.76	0.04	0.02	0.00	0.00
Belgium	0.02	0.04	0.46	0.35	0.10	0.03	0.00	0.00
	0.00	0.00	0.00	0.02	0.40	0.51	0.07	0.00
Paris Basin	0.00	0.00	0.06	0.13	0.37	0.39	0.05	0.00
	0.00	0.00	0.00	0.04	0.49	0.36	0.11	0.00
Northern Germany	0.00	0.00	0.00	0.00	0.02	0.05	0.32	0.61
	0.00	0.00	0.00	0.00	0.08	0.10	0.82	0.00
British Isles	0.00	0.00	0.00	0.00	0.00	0.01	0.60	0.39
	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00

Table 7.3: The probability that each region is temporally ranked one (earliest) through to eight (latest) under the conventional uniform model (light grey) and the spatio-temporal model (dark grey), assuming a minimum of a 1km per year expansion rate

can be resolved.

Note that when using the conventional uniform model the ten most likely orderings accounted for approximately 50% of the total posterior probability. However, this rose to approximately 75% of the total posterior probability under the spatio-temporal model.

Conclusions

We conclude, as a result of incorporating a spatial structure between the regions that there appears to be a more clearly defined order in which the regions were reoccupied. We also believe that the Upper Rhine was most probably the point of origin, as assumed by Housley *et al.* (1997), as the posterior probability of the Upper Rhine being the first region to be reoccupied is 0.90 (see Table 7.3). In addition we also believe that there may have been two points of origin, as previously discussed one from the south (the Upper Rhine) and one from the east (the Thuringian Basin). We draw this conclusion on the basis of Table 7.3 and Figure 6.1. From Table 7.3 we see that the Thuringian Basin was most likely to be reoccupied second (with a posterior probability of 0.55). Using this conclusion and looking at Figure 6.1 it seems very unlikely that the hunter gathers would travel directly from the Upper Rhine to the Thuringian Basin without passing through Southern Germany. However, without the use of topography we cannot draw any firm conclusions. For further discussion on possible routes of migration through NW Europe after the last Glacial period see Section 7.6.1.

7.6 Results: Trapezium model *versus* the trapezium spatio-temporal model

In this section we make a comparative interpretation of the data, under the trapezium and trapezium spatio-temporal models. Again, discussing the first dates of reoccupation in each of regions, as well as the order in which the regions were reoccupied.

First date of reoccupation

The estimates for the first dates of reoccupation under the trapezium and trapezium spatio-temporal models are summarized in Table 7.4 by their 95% HPD regions and in Figure 7.2 by their marginal posterior distributions. We see the same affect on the length of the HPD regions as we saw in Section 7.5. That is, under the spatio-temporal model the HPD regions for the first date of reoccupation are much narrower *e.g.* under the trapezium model the 95% HPD region for the Upper Rhine is 16740-23470 cal BP compared to 16390-16940 cal BP under the trapezium spatio-temporal model. The same is true of the other seven regions.

Region	95% HPD interval for the date of first reoccupation (cal BP)	
	No spatial info	1 km per year
Upper Rhine	16740 - 23470	16390 - 16940
Thuringian Basin	16620 - 18600	16230 - 16810
Southern Germany	15440 - 18530	16210 - 16850
Middle Rhine	15260 - 16580	16210 - 16720
Belgium	15440 - 18690	16000 - 16610
Paris Basin	15140 - 17010	16000 - 16620
Northern Germany	14530 - 16320	15850 - 16550
British Isles	147500 - 15530	15570 - 16280

Table 7.4: The 95% HPD intervals for the first date of reoccupation of the eight regions under the trapezium model when incorporating no spatial dependence and when assuming a minimum expansion rate of 1km per year.

From Figure 7.2, which shows the estimates for the marginal posterior distributions of the first dates of reoccupation in each region under the trapezium model (red lines) and trapezium spatio-temporal (blue lines), we can see that the estimates under the trapezium model have more uncertainty associated with them than the estimates obtained from the spatio-temporal model, which makes a huge difference to the archaeological interpretations drawn from the data under the two models.

The order in which reoccupation took place

Using Figure 7.2 it appears that the Upper Rhine was the first region to be reoccupied and the British Isles the last, however, the chronological ordering of the regions in between

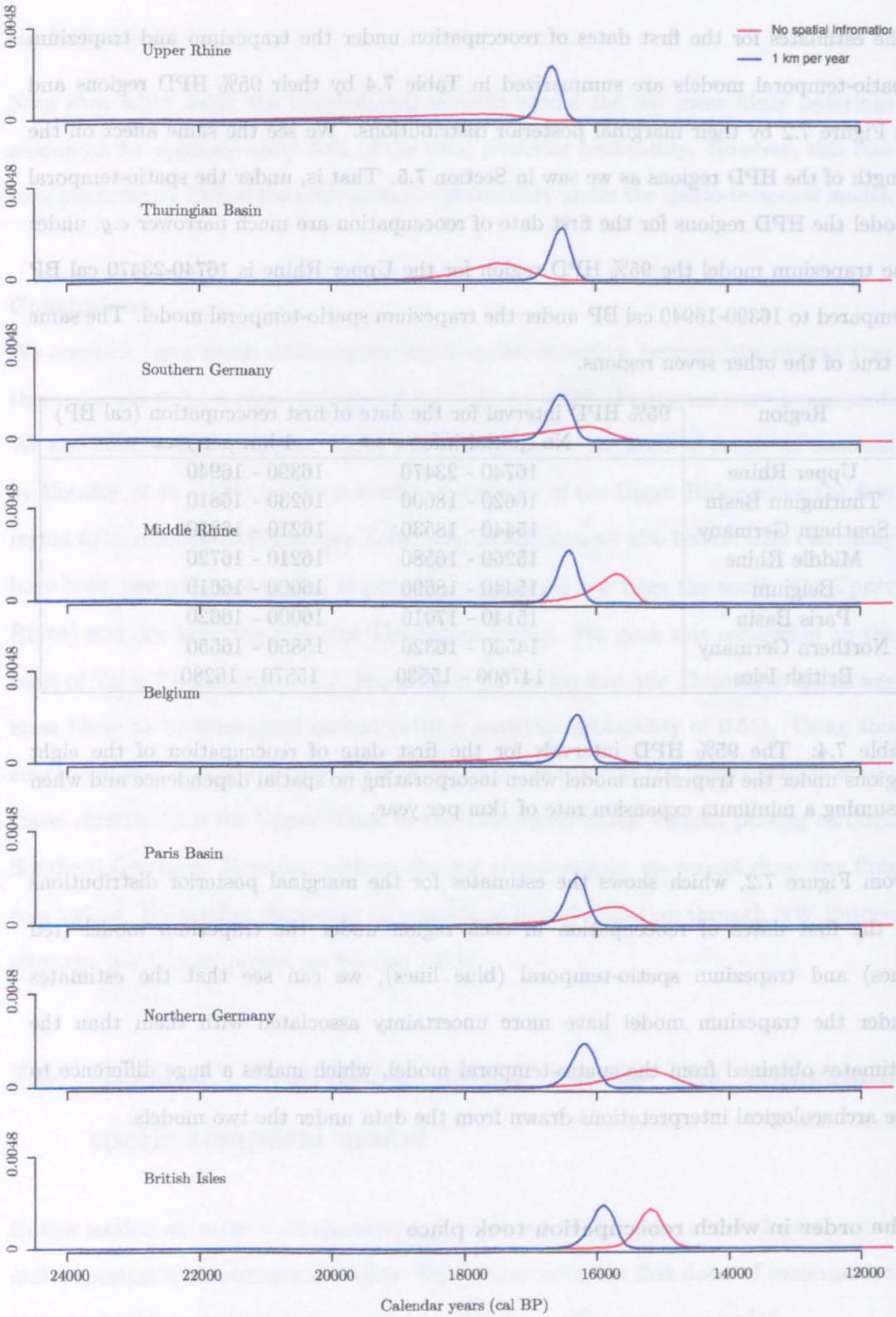


Figure 7.2: Marginal posterior distributions for the first date of reoccupation in each region under trapezium model (red) and the trapezium spatio-temporal model (blue), assuming a minimum expansion rate of 1km per year.

Region	1	2	3	4	5	6	7	8
Upper Rhine	0.75	0.19	0.06	0.00	0.00	0.00	0.00	0.00
	0.80	0.15	0.04	0.01	0.00	0.00	0.00	0.00
Thuringian Basin	0.14	0.58	0.24	0.04	0.00	0.00	0.00	0.00
	0.08	0.37	0.41	0.09	0.03	0.02	0.00	0.00
Southern Germany	0.05	0.10	0.28	0.32	0.17	0.07	0.01	0.00
	0.12	0.42	0.27	0.14	0.03	0.02	0.00	0.00
Middle Rhine	0.00	0.01	0.04	0.11	0.29	0.40	0.14	0.01
	0.00	0.04	0.21	0.56	0.09	0.08	0.02	0.00
Belgium	0.06	0.12	0.29	0.30	0.16	0.06	0.01	0.00
	0.08	0.10	0.42	0.37	0.08	0.03	0.00	0.00
Paris Basin	0.00	0.01	0.08	0.17	0.30	0.30	0.12	0.02
	0.00	0.01	0.03	0.09	0.37	0.35	0.15	0.00
Northern Germany	0.00	0.00	0.02	0.04	0.08	0.14	0.35	0.37
	0.00	0.01	0.01	0.03	0.11	0.11	0.69	0.04
British Isles	0.00	0.00	0.00	0.00	0.00	0.04	0.35	0.61
	0.00	0.00	0.00	0.00	0.00	0.01	0.04	0.95

Table 7.5: The probability that each region is temporally ranked one (earliest) through to eight (latest) under the conventional uniform model (light grey) and the spatio-temporal model (dark grey), assuming a minimum expansion rate of 1km per year.

is not clear. Again, we summarize the order in which reoccupation occurred by giving the probabilities that each region is temporally ranked 1 (earliest) through to 8 (latest) for both the trapezium and spatio-temporal models, see Table 7.5. Again, we highlight the most probable order of a region being reoccupied in light grey for the trapezium model and dark grey for the spatio-temporal model. Under both models the Upper Rhine was the most probable region to be reoccupied first, with a posterior probability of 0.75 under the trapezium model and 0.80 under the spatio-temporal model. If we are interested in the most probable rank of Northern Germany (under the trapezium model) we see that the posterior probability of it being reoccupied seventh is 0.35 and being recolonized last as 0.37. As there is very little difference between the two probabilities it is

difficult to make any inferences on the order in which Northern Germany was reoccupied. However, under the spatio-temporal model the most probable order of Northern Germany being reoccupied is seventh with a posterior probability of 0.69. Hence as a result of incorporating a spatial structure we obtain more definite (higher posterior probabilities, less uncertain) orderings of the eight regions.

In Appendix C, Table C.2, reports the ten most likely orderings under both the trapezium and spatio-temporal models, respectively. Here we just report the most likely order under the two models. Under the trapezium model the most likely order is the Upper Rhine, Thuringian Basin, Belgium, Southern Germany, Paris Basin, Middle Rhine, Northern Germany and the British Isles with a posterior probability of 0.03. While under the spatio-temporal model the most likely order is the Upper Rhine, Southern Germany, Thuringian Basin, Middle Rhine, Paris Basin, Belgium, Northern Germany and the British Isles with a posterior probability of 0.10. Although this probability is still relatively small under the spatio-temporal model it has tripled in comparison to the corresponding probability under the trapezium model. It is clear that there is still a large amount of uncertainty to be taken into consideration when trying to make inferences concerning the order of reoccupation. We see from Table C.2 that under the trapezium model the ten most likely orderings account for approximately 22% of the total posterior probability, while under the spatio-temporal model the ten most likely orderings account for 45% of the total posterior probability.

7.6.1 Possible routes of migration through NW Europe

The primary objective of our work on the reoccupation case study was to estimate the earliest date of reoccupation of each region under study, from which we can derive the most likely order in which the regions may have been reoccupied. An additional interest to applied researchers, which follows on naturally from the most likely order, is the question 'what routes did people take through NW Europe during the reoccupation process?'.

Thus, the final topic that we discuss (with regard to this case study) is possible routes of human migration through NW Europe at the end of the Late glacial period. We cannot reach any definite conclusions about such routes with the tools developed in this thesis, but we can begin to discuss methods that might lead us to them.

Our starting point in discussing possible routes of migration are the inferences obtained in Section 7.5, in particular, those relating to the most likely ordering of reoccupation (see Table C.1). Alongside this, we need to refer to Figure 7.3 which defines the eight regions of NW Europe and indicated the most likely order of each region in bold typeface. Note that, in what follows here, only inferences obtained from the uniform spatio-temporal model will be discussed.

In Housley *et al.* (1997) the authors assumed a single point of origin for humans returning to NW Europe as the ice sheets retreated; namely the Upper Rhine. However on page 168 we discussed the idea of there being two points of origin one from the South (the Upper Rhine) and one from the East (the Thuringian Basin). We based these assumptions on the posterior probabilities reported in Table 7.3 and Figure 6.1. From Table 7.3 we see that the Upper Rhine was the most likely region to be reoccupied first with a posterior probability of 0.90 and that the Thuringian Basin was the second most likely region to be reoccupied with a corresponding probability of 0.55. Using this information along with Figure 7.3 it seems very unlikely that the hunter gathers would travel directly from the Upper Rhine to the Thuringian Basin without travelling through Southern Germany.

Although we can calculate the probability of a particular order in which the regions may have been reoccupied using our MCMC samples, it is not possible to use them to draw any firm conclusions concerning possible routes of migration. This is particularly true if we allow for the possibility of two points of origin. For example, we found that the Middle Rhine is most likely to have been the fourth region to be reoccupied. Given which areas are likely to have been reoccupied before this, the Middle Rhine may have been colonised by people travelling from the Thuringian Basin, Southern Germany or even possibly the Upper Rhine. Given the most likely posterior ordering and the possibility

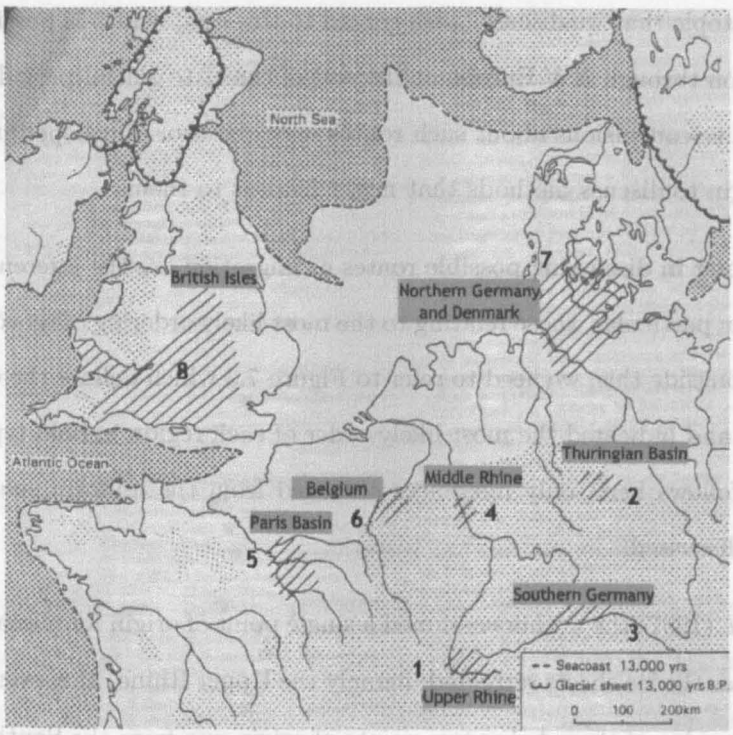


Figure 7.3: NW Europe after the last Glacial period assuming that regions were reoccupied in the order, the Upper Rhine, Thuringian basin, Southern Germany, Middle Rhine, Paris Basin, Belgium, Northern Germany and the British Isles.

of two points of initial origin, it is also possible that the Middle Rhine may have been reoccupied by people travelling in from two different directions simultaneously *e.g.* north from the Upper Rhine and west from the Thuringian Basin.

Thus, the results of this thesis do not lead to clear-cut conclusions about likely routes of migration. This may, in part, be because we have not accounted for topography in our models. It may be possible to eliminate some routes of migration if we take account of the topography of NW Europe during the Late glacial period *e.g.* the locations of geographic features such as river networks, mountain ranges and dense forests. To do this, extensions to the fully spatio-temporal models that we developed in Chapter 6 will be needed. Information about current and past landscapes can be now readily managed and plotted using geographical information systems (GIS). Incorporating information derived from such tools into our modelling framework (perhaps in the form of cost

surfaces) may be a sensible next step to take if we really want to begin to tackle issues relating to routes of migration/reoccupation, see Section 8.2.2 for further details.

7.7 Summary

In this chapter we implemented the spatio-temporal models developed in Chapter 6, with the aim of looking at any differences in the archaeological conclusions drawn from the human reoccupation case study when implementing non-spatial-temporal and spatio-temporal models.

As discussed in Section 7.3 the spatial information available, and thus incorporated in this case study was rather crude. However, it is clear that incorporating such information has made a huge difference to the archaeological conclusions drawn from the data, in particular, to the order of the regions in which reoccupation may have occurred. Hence the most likely order of reoccupation under both spatio-temporal models accounts for a higher proportion of the total posterior probability than under the non-spatio-temporal models. As well as this, a difference in the estimates for the first dates of reoccupation within each region is seen. Under the spatio-temporal model the HPD regions tend to be narrower than those obtained under the non-spatial models, which results in the conclusion that some regions were reoccupied earlier in time and others later in time *i.e.* hence the estimates under the spatio-temporal models are pulled closer to one another.

Chapter 8

Conclusions and further work

The aim of this thesis was to develop a more flexible statistical framework for the formal modelling of radiocarbon calibration problems by allowing a wider range of *a priori* information to be incorporated. Including this information allows us to obtain a more coherent and satisfactory interpretation of the data. As described in Chapter 1, the contents of this thesis may be broadly divided into two areas of research, namely

1. modelling the deposition of datable material within an archaeological phase
2. spatio-temporal modelling.

Below we offer detailed conclusions on both areas of research as well as some final thoughts and suggestions for future work on the material covered within this thesis.

8.1 Conclusions

The first of the new material is presented in Chapter 3. In this chapter we investigate alternative priors to the conventional uniform model currently used, and although these ideas have been suggested by a number of authors in the applied literature, we are among the first to study these suggestions in depth. Two case studies were briefly discussed

which motivate the use of non-uniform *a priori* distributions. As a result, we proposed a range of non-uniform distributions (see Section 3.4), all of which have an intuitive archaeological interpretation. Two of these priors, the trapezium and sigmoidal priors, are thought to have a wide range of uses (such as to describe the occupation of sites or whole regions as well as for the development of technologies and fashions) and formalise well-accepted heuristic representations of archaeological models.

The aim of Chapter 4 was to illustrate any differences in inferences drawn from the data. Two case studies were implemented using different *a priori* distributions for the rate at which material was deposited/manufactured within a phase of activity. The reason for choosing these two particular case studies is that both authors of the papers in which they originally appeared state (indirectly) their *a priori* beliefs about the rate at which the material dated was deposited/manufactured within the phase of interest. In both cases they are believed to be non-uniform.

In the original work on the first case study (Van Strydonck *et al.*, 2004) the authors were keen to explore formal Bayesian chronology-building tools, such as those available in OxCal, to compare their radiocarbon results with chronologies proposed by art historians. However, due to their lack of statistical knowledge they chose to implement a method commonly referred to as the 'summed probability distribution' method. In Section 4.2.1 we discussed our concerns with regard to implementing this method and our doubts that any meaningful interpretations can be drawn from the distributions produced.

Instead we suggest an alternative more logical way in which the authors could tackle this problem by implementing a simple temporal model as detailed in Buck *et al.* (1992). However, this approach did not account for the *a priori* beliefs that the authors state in their paper with regard to the rate at which the textiles are believed to have been manufactured over the proposed range. As a result we implemented both the conventional uniform model and the trapezium model for a single phase of textiles (12 stylistically related woollen tunics). Our main interest focused on the archaeological interpretations drawn from the data relating to the last date of manufacture of the

textile when assuming different *a priori* distributions.

Considerably different date estimates were derived for the last date of manufacture under the two models, both of which ruled out the later dates proposed by some art historians for these particular tunics. Under the uniform model there was no evidence to suggest that the tunics were older than the 8th century AD, while the trapezium model suggests that these tunics could be as old as the 9-10th century AD. Due to the *a priori* beliefs stated in Van Strydonck *et al.* (2004), we are inclined to believe the estimates produced under the trapezium model.

The second case study used in Chapter 4 is the human reoccupation of NW Europe case study, which has been utilized a number of times throughout the thesis. This case study has been analysed several times within the applied literature and has great importance in understanding the Late glacial period. This case study was chosen due to the *a priori* information by the original authors, relating to the mechanics of the reoccupation process.

Again, we use this case study as an illustrative example for implementing a range of models with differing *a priori* distributions for the rate at which the material available was deposited (within each of the regions). Blackwell and Buck (2003) suggested an alternative to the conventional uniform model; this being the sigmoidal prior. As a result we implemented three different models: the conventional uniform model, the trapezium model and the sigmoidal model.

Our primary interest was to estimate the first date of reoccupation within each region. The trapezium and sigmoidal models gave very similar date estimates, which differed considerably from those inferences obtained under the conventional uniform prior model. The trapezium and sigmoidal models allowed more uncertainty in the date estimates for the first dates of reoccupation, and in particular resulted in them being earlier in time than those derived under the conventional uniform model. In addition, the trapezium and sigmoidal models produced the same sequence for the most likely order in which the regions were reoccupied, again differing from the sequence derived under the conventional

uniform model.

From Chapter 4 it became clear that there are huge differences in the archaeological conclusions drawn from the data depending upon the *a priori* distribution used to model the rate at which material was deposited/manufactured within the phase of activity. The difference between the conclusions drawn under the trapezium and sigmoidal models are less dramatic than those between either of these and the uniform prior. For this reason we feel that it is not always necessary to implement both the trapezium and sigmoidal models. We slightly favour the use of a trapezium *a priori* distribution over the sigmoidal for a number of reasons. In particular the conventional uniform prior is a special case of the trapezium prior (when $\alpha = \gamma$ and $\delta = \beta$). For this reason, we feel that the trapezium prior has a more intuitive archaeological interpretation than that of the sigmoidal prior. We also feel that for those with a non-statistical background the properties of the trapezium prior have an easier and more meaningful interpretation.

Although we were satisfied that we had made advances in modelling the reoccupation process within each individual region more coherently, there were still a number of aspects that we felt had been ignored. Consequently, this lead us to the final area of research, attempting to tackle this problem using a spatio-temporal model.

Our first suggestions for spatio-temporal models were outlined in Chapter 5. The idea of this Chapter was to use joint *a priori* distributions to capture information between pairwise parameters of interest. As explained these ideas proved unsatisfactory when we failed to make a connection with the *a priori* distributions proposed and ways in which to adequately represent the *a priori* information arising from archaeological research. Subsequently, we decided to take a different approach to the same problem, which led to the development detailed in Chapter 6.

Chapter 6 offers what we believe to be the first advance towards tackling archaeological calibration problems within a fully spatio-temporal framework. It is by no means complete, but is aimed to give the basic groundwork for those wishing to pursue this line of research in the future.

The idea behind Chapter 6 was to find ways in which to incorporate spatial as well as temporal information into the modelling framework. It is believed that many archaeological problems such as colonisation/recolonisation of past landscapes are not purely temporal and thus including spatial information in the modelling framework is necessary and we would expect it to affect the interpretations drawn from the data greatly.

As our first approach in Chapter 5 failed we decided to tackle this problem from a purely practical perspective, *i.e.* with the human reoccupation case study in mind. The spatial information available from this case study consisted of (crudely measured) Euclidean distances between pairs of regions. We found no good way to use this kind of information under the framework outlined in Chapter 5. However, this form of spatial information is available for many case studies relating to the spread of plants and animals into regions/landscapes that were not previously occupied and so it seems sensible to approach the problem with this form of spatial information initially in mind.

Our approach in Chapter 6 was to extend the existing chronology building models, detailed in Chapters 2 & 3, with the aim of constructing a general framework in which both temporal and spatial information can be incorporated in a structured way. As our primary interest, with regard to the reoccupation case study, is to estimate the first date of reoccupation within each region it seemed sensible to seek a joint *a priori* distribution which incorporates the spatial structure between the α's.

In Housley *et al.* (1997) the authors assumed that the movement of the Late glacial hunters across NW Europe would follow a similar behaviour to those of European farmers. Subsequently, the authors assumed a 1km per year advance through NW Europe given the assumption that individuals disperse randomly in all directions. As a result we assumed a uniform ease of spread between the regions of NW Europe.

As seen in Chapters 2-4, when modelling archaeological calibration problems the convention is to assume a particularly simple form of joint prior for $p(\boldsymbol{\alpha}, \boldsymbol{\beta})$ in which all values of α_j and β_j that satisfy a given set of constraints, C, are equally likely. When

incorporating the spatial dependence we define a similar form of joint prior distribution, which incorporates the Euclidean distances and the ease of spread between pairwise α's. This allows us to represent the maximum length of time we believe it would take the Late glacial hunters to move between the regions.

When incorporating the spatial structure in this manner it quickly became apparent that this joint prior induces unrealistic dates for the marginal prior distributions of interest. In particular, the α's were forced to be much older. To combat this problem we developed an alternative definition of the joint prior.

Our next thoughts were to incorporate the spatial structure between the α's in the form of a joint prior, $p(\boldsymbol{\alpha})$, and then define a prior for β_j conditionally on α_j, *i.e.* $p(\beta_j|\alpha_j)$ in the case of the uniform model. This idea led to much more realistic dates for the marginal priors induced from $p(\boldsymbol{\alpha})$ and the conditional priors $p(\beta_j|\alpha_j)$ as well the parameters having meaningful archaeological interpretations. As a result, we incorporated the spatial structure for the trapezium model using the same approach.

The aim of Chapter 7 was to implement the fully spatio-temporal models and illustrate any differences in archaeological conclusions drawn from the data for the human reoccupation case study when implementing fully spatio-temporal models as opposed to purely temporal models. We implemented two spatio-temporal models with different *a priori* distributions for the rate at which material is deposited within the phase, to enable us to make direct comparisons with the inferences drawn in Chapter 4.

From Chapter 7 it became clear that incorporating spatial structure into the modelling framework, regardless of the *a priori* distributions used to model the deposition rate, made a huge difference to the archaeological conclusions drawn from the data. One of the most noticeable differences relates to the most likely order in which the regions were reoccupied. For example (when assuming a uniform prior on the deposition rate) the most likely order of reoccupation under the spatio-temporal model accounted for 1/5 of the total posterior probability, which was double the corresponding probability under the non-spatial model. Hence, by incorporating a spatial structure between the regions, we

are able more clearly to define the order in which the regions were reoccupied. Another difference that arose was our inferences about the first date of reoccupation. The HPD regions for these dates under the spatio-temporal models are much narrower than those under the non-spatial models. In particular, the first date is estimated to be earlier in time for some regions (*e.g.* The British Isles) and later in time for others (*e.g.* The Upper Rhine) thus moving the date estimates closer to one another.

This section has summarized the main findings within this thesis. The following section will discuss some final thoughts on the material covered in the thesis as well as some ideas for extending the work further.

8.2 Further work

We feel that we have developed a more flexible and coherent statistical framework in which to incorporate a wider range of *a priori* information, arising from either experts within the field or archaeological research. Using our framework we have been successful in tackling a number of archaeological calibration problems. Our greatest advance is being able to tackle problems within a spatio-temporal framework which were in the past tackled using a range of *ad hoc* methods. However there are still several further issues that need to be addressed, some of which are outlined below

- model choice
- extending spatio-temporal models
- outlier detection.

8.2.1 Model choice

One area of research that we did not tackle but is of great importance to the models developed within the thesis is that of model choice. Within the thesis we have proposed a range of models with differing *a priori* distributions for the rate at which material was

deposited/manufactured within a phase. So how do applied researchers choose between competing models? There are situations in which we can make model choices on the basis of the archaeological information (such as the case study arising from Van Strydonck *et al.* (2004), however there are equally as many cases in which this is not the true. Although we did not attempt to tackle this problem, we give our thoughts on ways in which to proceed.

As there is uncertainty regarding the model, interest lies in comparing models (in our case) which have different sets of parameters, with varying dimensions. One way to tackle this problem would be to use a reversible jump MCMC algorithm. This allows us to construct a Markov chain whose state can be of different dimensions, yet has the correct stationary distribution.

The reversible jump MCMC algorithm extends from the Gibbs sampler and Metropolis-Hastings algorithm by allowing for moves between models with varying dimensions. The algorithm updates the parameters, given the model, using standard MCMC algorithms then updates the model using the reversible jump procedure. In summary, the algorithm constructs a Markov chain whose stationary distribution is the joint posterior distribution of the models and parameters.

By using the reversible jump algorithm to move between different models, the algorithm can be used to estimate the proportion of time that the Markov chain spent in each model, referred to as posterior model probabilities. Thus enabling us to say which model is preferred. Further details on the reversible jump algorithm can be found in Givens and Hoeting (2005).

An alternative approach to using the reversible jump MCMC is to calculate Bayes factors for the different models, this being the most widespread model choice criteria. We discussed the ideas behind Bayes factors in Section 8.2.1.

One of our main interests would be to choose between competing models such as the conventional uniform model and the trapezium model. Consequently we want to compare

nested models[1], for details on how to calculate Bayes Factors see O'Hagan (1994).

8.2.2 Extending Spatio-temporal models

In Chapter 6 we took the first steps towards modelling colonisation/recolonisation problems in a spatio-temporal framework, yet there are a number of aspects that could be extended further.

As seen in Chapters 6 & Chapter 7 we assumed, given the *a priori* beliefs stated in Housley *et al.* (1997), that there was a uniform ease of spread between the regions of NW Europe. We defined the measure of difficulty of spread between regions ($c_{j,k}$) to be proportional to the Euclidean distance ($d_{j,k}$) between regions. Next we introduced s to represent the minimum speed parameter (kms per year) which allowed us to represent the maximum length of time in which the Late glacial hunters would move between regions. We chose s to be a constant for simplicity, however we could alternatively treat s as an unknown parameter. Another difficulty that may arise is that we are still assuming that the ease of spread between each region is similar. Clearly, this assumption is unrealistic, as is the assumption of a uniform ease of spread between regions.

Before the models can be developed further it is important to understand what information, in terms of the topography of the landscapes, is available. By incorporating such information (*e.g* river networks, mountain ranges, dense forests) may make a huge difference to the archaeological conclusions drawn from the data. Information of current landscapes is available through the use of GIS (geographical information systems) which can provides locational data (grid reference or latitude and longitude) along with information about the topography. Although the landscapes of interest will have changed over the thousands of years that have elapsed the main features will have remained similar (*e.g.* mountain ranges). We feel that incorporating the GIS information could help advance the spatio-temporal modelling framework and help us to arrive at more

[1]Nested models arise when one of our proposed models is a spacial case of the other and, in many cases, may be expressed using fewer parameters

coherent interpretations of the data. In addition it may add to the level of sophistication of the interpretations that can be drawn from the data so that, not only would we be able to provide sequences for the ordering in which reoccupation occurred, we would also be able to make more coherent statements about possible routes of migration.

As discussed in Chapter 6 there was no quantitative spatial information (available to us at the time we carried out the research) for the human reoccupation case study, as a result we used crudely measured Euclidean distances. There are, however, a number of case studies arising (concerning similar recolonisation problems) such as Gamble *et al.* (2004) who make use of the S2Age database (consisting of over 2000 radiocarbon determinations across Western Europe, each associated with Latitude and Longitudinal readings). It would therefore be sensible, within each phase, to incorporate a spatial dependence between archaeological sites (each with their own radiocarbon determinations). From Figure 6.1 it is clear that the regions in NW Europe each cover a large area and even though we can estimate when they first became reoccupied we are unclear as to the exact location of reoccupation and the directional movement through the region. By incorporating a spatial dependence between sites (within a region) it might be possible to shed light on a number of archaeological queries that are currently not tackled using formal statistical models.

8.2.3 Outlier detection

One final area of research that we would like to discuss briefly is outlier detection. As we saw in Section 2.5.3 outliers in radiocarbon dating are relatively common and Christen (1994a,b) proposed a Bayesian approach for modelling and identifying outliers in groups of related radiocarbon determinations. We did not extend the ideas in Christen (1994a,b) to account for outlier detection when implementing non-uniform *a priori* distributions, however we feel that this would not require a huge amount of additional work, consequently incorporating outlier detection into the analysis would be a useful next step.

As the trapezium *a priori* distribution accounts for uncertainty in the tails of the distribution, it is quite likely that a date which is referred to as an outlier under the uniform model will not necessary be an outlier under the trapezium model. This suggests that caution will be needed in helping the archeologists interpret data and understand why different models may results in different outliers.

Appendix A

C code

The attached CD Rom is divided into a three subdirectories. Each subdirectory is of the same structure. The first subdirectory contains the programs written in C to implement the single phase Coptic textiles case study (see Section 4.2) assuming a trapezium prior distribution for the rate at which the textiles were manufactured. The second subdirectory contains the programs needed to implement the human reoccupation case (multiple phases) of NW Europe when assuming a sigmoidal prior for the deposition rate. The final subdirectory contains the programs needed to implement the reoccupation case study in a fully spatio-temporal framework.

Each subdirectory contains one folder and 7 files. The folder called 'input' contains the archaeological information for each case study (*i.e.* radiocarbon determinations, the starting values to be used, the data need to use IntCal04 *etc.*). The three files random.h, random.c and drand48.c are common to all subdirectories, these enable us to generate from standard probability distributions (files obtained from Marc Kennedy).

Each subdirectory also contains a DEV C++ project file which links the two header and four source files to one another. Note that header files (.h) contain small parts of program code which contains the variable definitions along with the function prototypes. The source files (.c) are the files containing the text (the main code) and commonly

referred to as the program.

The two files functions.h and functions.c are also common to all subdirectories, containing written mathematical functions which are not standard in C. Each of these file varies between the subdirectories as a result of which model is being implemented.

1. **Trapezium model - single phase**
    - DEV C++ project file: 'Coptic_trapezium'
    - Main source file: 'Coptic_trap' (code annotated)
    - Secondary source files: random.c, drand48.c and functions.c
    - Header files: random.h and functions.h
2. **Sigmoidal model - multiple phase**
    - DEV C++ project file: 'recol_sigmoidal'
    - Main source file: 'recol_sig' (code annotated)
    - Secondary source files: random.c, drand48.c and functions.c
    - Header files: random.h and functions.h
3. **Uniform spatio-temporal model**
    - DEV C++ project file: 'uniform_spatiotemporal'
    - Main source file: 'uniform_spatio' (code annotated)
    - Secondary source files: random.c, drand48.c and functions.c
    - Header files: random.h and functions.h

Appendix B

Archaeological data

This Appendix contains a simplified version of Table 1 from Housley *et al.* (1997) which contains the data available for the Human reoccupation of NW Europe case study. Each radiocarbon determination is given with its corresponding lab code as well as indicating which of the eight regions it belongs to.

Table B.1: Radiocarbon determinations associated with each of the 8 regions from Housley *et al.* (1997).

Laboratory identifier	Determination BP	Phase
OxA5750	13670±100	Upper Rhine
OxA5749	14150±100	Upper Rhine
OxA5745	13940±100	Upper Rhine
OxA5747	13430±100	Upper Rhine
OxA5746	13120±90	Upper Rhine
OxA5748	12770±90	Upper Rhine
OxA5744	11780±90	Upper Rhine
OxA1126	12890±140	Middle Rhine
OxA1128	13200±140	Middle Rhine
OxA1129	13090±130	Middle Rhine
OxA1130	12950±140	Middle Rhine
OxA1125	12930±180	Middle Rhine
OxA1127	12820±130	Middle Rhine
OxA1130	12790±120	Middle Rhine
OxA1129	12910±130	Middle Rhine
OxA1128	12730±130	Middle Rhine
OxA4854	13230±130	Southern Germany
OxA5756	11590±90	Southern Germany
OxA5755	12060±90	Southern Germany
OxA5751	12610±90	Southern Germany
OxA5754	12680±100	Southern Germany
OxA5753	12740±90	Southern Germany
OxA5752	12410±90	Southern Germany
OxA5720	12440±140	Southern Germany
OxA5719	12350±130	Southern Germany
OxA5718	13160±130	Southern Germany
OxA3635	12870±95	Belgium
OxA4191	10800±110	Belgium
OxA4190	10330±110	Belgium
OxA3634	10320±80	Belgium
OxA4200	13330±160	Belgium
OxA4014	12870±110	Belgium
OxA4197	12800±130	Belgium
OxA3633	12880±100	Belgium
OxA4192	12860±140	Belgium
OxA3632	12790±100	Belgium
OxA4198	12660±140	Belgium
OxA4195	12630±140	Belgium
OxA4199	12240±130	Belgium
OxA177	12300±220	Paris Basin
OxA467	12250±160	Paris Basin
OxA176	12000±220	Paris Basin
OxA391	11870±130	Paris Basin
OxA175	12900±220	Paris Basin
OxA173	12800±220	Paris Basin

Laboratory identifier	Determination BP	Phase
OxA149	12400±200	Paris Basin
OxA148	12600±200	Paris Basin
OxA138	12900±300	Paris Basin
OxA139	13000±130	Paris Basin
OxA740	12120±200	Paris Basin
OxA178	11600±200	Paris Basin
OxA731	12240±160	Paris Basin
OxA730	12300±160	Paris Basin
H38121A	12300±300	Northern Germany
AAR1036	12140±110	Northern Germany
H38121B	12300±200	Northern Germany
W281	11870±200	Northern Germany
W264	11790±200	Northern Germany
W271	11750±200	Northern Germany
W261	12450±200	Northern Germany
K4261	12190±125	Northern Germany
AAR906	12520±190	Northern Germany
H136116	12980±370	Northern Germany
H3167	13050±270	Northern Germany
K4577	12440±115	Northern Germany
K4332	12570±115	Northern Germany
K4331	12440±115	Northern Germany
K4329	12360±110	Northern Germany
K4328	12180±130	Northern Germany
OxA5726	12640±130	Thuringian Basin
OxA5725	12990±130	Thuringian Basin
OxA5724	12940±140	Thuringian Basin
OxA5723	13080±140	Thuringian Basin
OxA5722	12860±130	Thuringian Basin
OxA5717	12670±110	Thuringian Basin
OxA5716	12790±110	Thuringian Basin
OxA5715	11810±110	Thuringian Basin
OxA5714	12620±120	Thuringian Basin
OxA5713	12740±120	Thuringian Basin
OxA5712	12270±110	Thuringian Basin
OxA5711	12050±110	Thuringian Basin
OxA5710	12080±110	Thuringian Basin
OxA5709	12270±120	Thuringian Basin
OxA4853	13090±130	Thuringian Basin
OxA4852	13520±130	Thuringian Basin
OxA4851	14470±140	Thuringian Basin
OxA4850	13160±140	Thuringian Basin
OxA4849	13130±120	Thuringian Basin
OxA4848	13150±130	Thuringian Basin
OxA4846	13190±130	Thuringian Basin
OxA4845	13120±130	Thuringian Basin
OxA4832	13310±110	Thuringian Basin

Laboratory identifier	Determination BP	Phase
OxA3413	12940±140	British Isles
OxA4106	12670±120	British Isles
OxA3411	12650±120	British Isles
OxA3416	12580±110	British Isles
OxA4107	12550±130	British Isles
OxA4102	12540±140	British Isles
OxA3404	12540±110	British Isles
OxA3412	12490±120	British Isles
OxA3452	12400±110	British Isles
OxA4109	12370±120	British Isles
OxA3415	12340±120	British Isles
OxA3400	12340±110	British Isles
OxA3398	12280±110	British Isles
OxA735	12240±150	British Isles
OxA4110	12110±120	British Isles
OxA4108	12110±120	British Isles
OxA1493	11970±120	British Isles
OxA1950	11740±150	British Isles
OxA150	12400±300	British Isles
OxA1467	12350±120	British Isles
OxA1616	12600±170	British Isles
OxA1618	12480±170	British Isles
OxA1619	12450±150	British Isles
OxA1617	12420±200	British Isles
OxA1670	12290±120	British Isles
OxA3718	12250±90	British Isles
OxA3717	12020±100	British Isles
OxA1494	12000±120	British Isles
OxA1500	12350±160	British Isles
OxA1789	12320±130	British Isles
OxA466	12800±170	British Isles
OxA3414	12570±120	British Isles
OxA464	12470±160	British Isles
OxA590	12370±150	British Isles
OxA465	12360±170	British Isles
OxA589	12340±150	British Isles
OxA1071	12300±180	British Isles
OxA1890	12170±130	British Isles
OxA587	12530±150	British Isles
OxA1121	12380±130	British Isles
OxA535	12210±160	British Isles

Appendix C

Spatio-temporal modelling: results

This Appendix contains two sets of tables, arising from the human reoccupation of NW Europe case study in Chapter 7. The first set of tables makes a comparison between the most likely orders of the reoccupation (for the eight regions under study) for the conventional uniform and uniform spatio-temporal models. While the second set of tables makes the same comparisons but between the trapezium model and trapezium spatio-temporal models.

Table C.1: The ten most likely orders of the reoccupation of the eight regions under study (1=earliest, 8=latest) when implementing a) the conventional uniform model and b) the uniform spatio-temporal model.

a): Uniform model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	1	1	1	2
Thuringian Basin	2	2	2	2	2	2	2	2	2	1
Southern Germany	4	4	3	3	4	4	3	3	5	4
Middle Rhine	6	5	6	5	6	5	6	5	6	6
Belgium	3	3	4	4	3	3	4	4	3	3
Paris Basin	5	6	5	6	5	6	5	6	4	5
Northern Germany	8	8	8	8	7	7	7	7	8	8
British Isles	7	7	7	7	8	8	8	8	7	7
Probability	0.082	0.078	0.069	0.065	0.044	0.040	0.037	0.033	0.026	0.025

b): Uniform spatio-temporal model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	2	2	1	1
Thuringian Basin	2	3	2	3	2	2	3	3	2	2
Southern Germany	3	2	3	2	4	4	1	1	3	3
Middle Rhine	4	4	4	4	3	3	4	4	4	4
Belgium	6	6	5	5	6	5	6	5	5	7
Paris Basin	5	5	6	6	5	6	5	6	7	5
Northern Germany	7	7	7	7	7	7	7	7	6	6
British Isles	8	8	8	8	8	8	8	8	8	8
Probability	0.19	0.15	0.13	0.10	0.054	0.049	0.022	0.020	0.019	0.018

Table C.2: The ten most likely orders of the reoccupation of the eight regions under study (1=earliest, 8=latest) when implementing a) the trapezium model and b) the trapezium spatio-temporal model.

a): Trapezium model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	1	1	1	1
Thuringian Basin	2	2	2	2	2	2	2	2	2	2
Southern Germany	4	4	3	3	4	4	3	3	5	5
Middle Rhine	6	6	6	6	5	5	5	5	6	6
Belgium	3	3	4	4	3	3	4	4	3	3
Paris Basin	5	5	5	5	6	6	6	6	4	4
Northern Germany	7	8	7	8	8	7	8	7	7	8
British Isles	8	7	8	7	7	8	7	8	8	7
Probability	0.026	0.025	0.025	0.025	0.023	0.023	0.022	0.021	0.014	0.013

b): Trapezium spatio-temporal model

Region	Position in ordering									
Upper Rhine	1	1	1	1	1	1	2	2	1	1
Thuringian Basin	3	3	2	2	2	2	3	3	3	4
Southern Germany	2	2	3	3	4	4	1	1	2	2
Middle Rhine	4	4	4	4	3	3	4	4	5	3
Belgium	6	5	6	5	5	6	6	5	6	6
Paris Basin	5	6	5	6	6	5	5	6	4	5
Northern Germany	7	7	7	7	7	7	7	7	7	7
British Isles	8	8	8	8	8	8	8	8	8	8
Probability	0.097	0.078	0.074	0.061	0.030	0.029	0.024	0.023	0.020	0.017

Bibliography

[1] Aitchison, T. C., Ottaway, B. S., and Al-Ruzaiza, A. S. (1991). Summarizing a group of ^{14}C dates on the historical time scale: with a woked example from the Late Neolithic of Bavaria. *Antiquity*, 65:108–116.

[2] Aitchison, T. C., Ottaway, B. S., and Scott, E. M. (1990). Statistical treatment of groups of related radiocarbon dates. In *Second International Symposium ^{14}C and Archaeology*. Mook, W. G. and Waterbolk, H. T. (ed.), PACT, Groningen, 95–104.

[3] Aitken, J. M. (1990). *Science-based Dating in Archaeology*. Longman, London.

[4] Ammerman, A. J. and Cavalli-Sforza, L. L. (1973). A population model for the diffusion of early farming in Europe. In Renfrew, A. C., editor, *The explanation of Culture Changes*, pages 343–358. London: Duckworth.

[5] Blackwell, P. G. and Buck, C. E. (2003). The Late Glacial human reoccupation of the north-wetern Europe: new approaches to space-time modelling. *Antiquity*, 77:232–240.

[6] Blockley, S. P. E., Donahue, R. E., and Pollard, A. M. (2000). Radiocarbon calibration and the Late Glacial occuptaion of northwest Europe. *Antiquity*, 74:112–119.

[7] Bowman, S. (1990). *Interpreting the Past: Radiocarbon Dating*. British Museum Publications, London.

[8] Buck, C. E. and Blackwell, P. G. (2004). Formal statistical models for estimating radiocarbon calibration curves. *Radiocarbon*, 46(3):1039–1102.

[9] Buck, C. E., Cavanagh, W. G., and Litton, C. D. (1996). *Bayesian Approach to Interpreting Archaeological Data.* Wiley and Sons.

[10] Buck, C. E., Christen, J. A., and James, G. N. (1999). BCal: an on-line Bayesian radiocrabon calibration tool. *Internet Archaeology*, 7. URL http://intarch.ac.uk/journal/issue7/buck/.

[11] 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.

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

[13] Christen, J. A. (1994a). *Bayesian interpretation of ^{14}C results.* PhD thesis, University of Nottingham, Nottingham,UK.

[14] Christen, J. A. (1994b). Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics*, 43:489–503.

[15] Christen, J. A. (2003). Bwigg: An Internet facility for Bayesian radiocarbon wiggle-matching. http://intarch.ac.uk/journal/issue13/christen_index.html.

[16] Christen, J. A., Clymo, R. S., and Litton, C. D. (1995). A Bayesian approach to the use of ^{14}C dates in the estimation of the age of peat. *Radiocarbon*, 37:431–442.

[17] Christen, J. A. and Litton, C. (1995). A Bayesian aproach to wiggle-matching. *Journal of Archaeological Science*, 22:719–725.

[18] Gamble, C. S., Davis, W., Pettitt, P. B., and Richards, M. B. (2004). Climate change and evolving human diversity in Europe during the last glacial. *Philosophical Transactions of the Royal Society: Biological Sciences*, 359(12):243–254.

[19] Gilks, W., Richardson, S., and Speigelhalter, D. (1996). *Markov chain Monte Carlo in practice.* Chapman and Hall, London.

[20] Givens, G. H. and Hoeting, J. A. (2005). *Computational Statistics.* Wiley and Sons.

[21] Hastings, W. K. (1970). Monte Carlo sampling methods using Markov chains and their applications. *Biometrika*, 57:97–109.

[22] Housley, R. A., Gamble, C. S., Street, M., and Pettitt, P. (1997). Radiocarbon evidence for the lateglacial human recolonisation of Northern Europe. *Proceedings of Prehistorica Society*, 63:25–54.

[23] Hughen, K. A., Baillie, M. G. L., Bard, E., Warren Beck, J., Bertrand, C. J. H., Blackwell, P. G., Buck, C. E., Burr, G. S., Cutler, K. B., Damon, P. E., Edwards, R. L., Fairbanks, R. G., Friedrich, M., Guilderson, T. P., Kromer, B., McCormac, F. G., Manning, S., Bronk Ramsey, C., Reimer, P. J., Reimer, R. W., Remmele, S., Southon, J. R., Stuiver, M., Talamo, S., Taylor, F. W., van der Plicht, J., and Weyhenmeyer, C. E. (2004). Marine04 Marine radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon*, 46(3):1059–1086.

[24] Litton, C. and Buck, C. E. (1996). An arachaeological example: radiocarbon dating. In Gilks, S. R. W. and Spiegelhalter, D., editors, *Markov chain Monte Calro in practice*, pages 465–480. Chapman and Hall, London.

[25] Litton, C. D. and Lesse, M. N. (1991). Some statstical problems arising in radiocarbon calibration. In *Computer Applications and Quantitative Methods in Archaeology*, pages 101–109. Lockyear, K. and Rahtz, S. P. Q. (eds.).

[26] McCormac, F. G., Hogg, A. G., Blackwell, P. G., Buck, C. E., Higham, T. F. G., and Reimer, P. J. (2204). SCHAL04 Southern hemisphere calibration, 0–11.0 cal kyr BP. *Radiocarbon*, 46(3):1087–1092.

[27] Metropolis, N., Rosenbluth, A. W., Rosenbluth, M. N., Teller, A. H., and Teller, E. (1953). Equations of state calculations by fast computing machine. *Journal of Chemical Physics*, 21:1087–1092.

[28] Naylor, J. C. and Smith, A. F. M. (1988). An archaeological inference problem. *Journal of the American Statistical Association*, 83:588–595.

[29] Nicholls, G. and Jones, M. (2001). Radiocarbon dating with temporal order constraints. *Applied Statistics*, 50(4):503–521.

[30] O'Hagan, A. (1994). *Kendall's Advanced Theory for Statisticians: Bayesian Inference*, volume 2B. Edward Arnold, London.

[31] Ottaway, B. S. (1973). Dispersion diagrams: a new approach to the display of ^{14}C dates. *Archaeometry*, 15(1):5–12.

[32] Pearson, G. W., Pilcher, J. R., Ballie, M. G. L., Corbett, D. M., and Qua, F. (1986). High-precision ^{14}C measurments of Irish oaks to show the natural ^{14}C variations from AD 1840–5210 BC. *Radiocarbon*, 28(2B):911–34.

[33] Pearson, G. W. and Stuiver, M. (1986). High-precision calibration of radiocarbon time scale, 500–2500 BC. *Radiocarbon*, 28(2B):839–862.

[34] Ramsey, C. B. (2005). OxCal program v3.10 manual. http://www.rlaha.ox.ac.uk/oxcal/arch_cmb.htm#sum.

[35] Reimer, P. J., Baillie, M. G. L., Bard, E., Bayliss, A., Beck, W. J., Bertrand, C. J. H., Blackwell, P. G., Buck, C. E., Burr, G. S., Cutler, K. B., Damon, P. E., Edwards, R. L., Fairbanks, R. G., Friedrich, M., Guilderson, T. P., Hogg, A. G., Hughen, K. A., Kromer, B., McCormac, F. G., Manning, S., Bronk Ramsey, C., Reimer, R. W., Remmele, S., Southon, J. R., Stuiver, M., Talamo, S., Taylor, F. W., van der Plicht, J., and Weyhenmeyer, C. E. (2004). IntCal04 Terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon*, 46(3):1029–1058.

[36] Scott, M. (2000). Bayesian methods: what can we gain and at what cost? *Radiocarbon*, 42(2):181.

[37] Smith, B. J. (2005). Bayesian Output Analysis program (BOA), version 1.1.5. http://www.public-health.uiowa.edu/boa.

[38] Stuiver, M. and Pearson, G. W. (1986). High-precision calibration of the radiocarbon time scale, AD 1950–500 BC. *Radiocarbon*, 28(2B):805–838.

[39] Stuiver, M., Reimer, P. J., Bard, E., Becks, J. W., Burr, G. S., Hughen, K. A., Kromer, B., McCormac, F. G., Plicht, J. V. D., and Spurk, M. (1998). IntCal98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon*, 40:1041–1083.

[40] Suess, H. E. (1970). Bristlecone-pine calibration of the radiocarbon time scale 5200 b.c to the present. In *Proceedings of the twelfth Nobel Symposium*. Olsson, I. U. (ed.), Wiley and Sons, New York, 303–311.

[41] van der Plicht, J., Beck, W. J., Bard, E., Baillie, M. G. L., Blackwell, P. G., Buck, C. E., Friedrich, M., Guilderson, T. P., Hughen, K. A., Kromer, B., McCormac, F. G., , Bronk Ramsey, C., Reimer, P. J., Reimer, R. W., Remmele, S., Richards, D. A., Southon, J. R., Stuiver, M., and Weyhenmeyer, C. E. (2004). NOTCAL04 comparison/calibration ^{14}C records 26–50 cal kyr BP. *Radiocarbon*, 46(3):1225–1238.

[42] Van Strydonck, M., Moor, A. D., and Benazeth, D. (2004). ^{14}C dating compared to art historical dating of Roman and Coptic textiles from Egypt. *Radiocarbon*, 46(1):231–244.

[43] Ward, G. K. and Wilson, S. R. (1978). Procedures for comparing and combing radiocarbon age determinations: A critique. *Archaeometry*, 20(1):19–31.

[44] Weninger, B. (1986). High-precision calibration of archaeological radiocarbon dates. *Acta Interdiciplinaria Archaeologica*, 4:11–53.