Anglo-Saxon Chronology I – the male graves

A commentary on Chapter 6 of *Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework.*

Mike Baxter[1], 16 Lady Bay Road, West Bridgford, Nottingham, NG2 5BJ, UK
(e-mail: michaelj.baxter@btconnect.com)

January 2014

ABSTRACT: The book *Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework* (2013) is a major contribution to, and revision of, Anglo-Saxon chronology, the importance of which has already attracted considerable attention. The central Chapters 6 and 7 on the chronology of male and female furnished inhumations depend heavily on statistical and scientific methodology and reasoning that not all Anglo-Saxon scholars are necessarily equipped to follow. This assessment is based on comment from such scholars, who have also commented that the book makes for very difficult reading. This paper, and its companion on the female graves, was conceived as a 'reader's guide' to the analysis that tries to separate out the more important aspects from those that the less statistically informed reader might wish to avoid.

The statistical methodology developed in the book is, in an archaeological context, innovative and more advanced than other comparable methodology I am familiar with. The value of this extends beyond Anglo-Saxon studies and should be of interest to a much wider audience. This and the companion paper have taken a critical approach concerning the 'readability' of some of the text, but the methodological importance should not be obscured. This paper concerns Chapter 6 on the chronology of the male graves. A separate paper on Chapter 7, the female graves, has been written because of the different, and interesting, statistical and interpretive challenges it posed.

[1]Emeritus Professor of Statistical Archaeology, Nottingham Trent University, UK

1 Introduction

These notes are based on the book *Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework* (Bayliss *et al.*, 2013), the subject of a large part of a one-day conference, *Re-dating Early England*, held in London in November 2013. It is a substantial and impressive tome that will clearly be of importance to Anglo-Saxon scholarship. This is not an area I am competent to comment on, but statistical aspects of the book (about which I know more), particularly in the central chapters 6 and 7, are impressive. I have a longstanding interest in applications of correspondence analysis (CA) in archaeology. This method is central to the analysis in the book and is one of the most ambitious and interesting applications I have seen. The manner in which CA is allied to Bayesian modelling is also, to the best of my knowledge, novel and merits recognition beyond the Anglo-Saxon community.

Were Chapters 6 and 7 to be rewritten in the form of papers suitable for submission to prestigious scientific archaeology journals such as the *Journal of Archaeological Science* or *Archaeometry*, and were I invited to referee such submissions, I would strongly recommend publication on the grounds of both substantive and scientific/statistical interest and importance. This and the companion paper are structured, in part, from the perspective of such a hypothetical referee of such hypothetical papers.

I begin in this way, in order to stress my admiration for what has been achieved in the book, because much of what is to follow will be read as critical. Massive as the book is, the writing of it represents only the tip of the iceberg of what was involved in its construction, particularly given the 'collegiate' nature of the production and time-scale involved. Were papers based on the book chapters to be structured in anything like the same way my strong recommendation for publication would be tempered with the caveat that substantial rewriting, including the omission of many figures, would be necessary to make the text readable and accessible to much of the intended audience. This opinion began to develop when I attempted to emulate analyses in the book – originally all I'd intended to do – and quickly became 'bogged down' in the detail provided.

I fully appreciate that the chapters were not drafted to be read as shorter scientific papers, nevertheless the issue of (un)readability matters. It seems obvious that the authors have been at some pains to record much of the 'iterative' and 'reflexive' analysis and thinking that led to the development of their final models. This desire for 'transparency', if it can be characterised as such, can again be admired, but as far as the readability of the text goes it has arguably resulted in 'opacity'.

I would hesitate to offer this opinion, but for the fact that I know of Anglo-Saxon scholars (and others) who have grappled with the text and have expressed similar views, sometimes more strongly. The substantive interest of the volume for many readers presumably lies in the chronology offered and its implications for the study of Anglo-Saxon archaeology. The development of the chronology rests, however, on statistical and scientific methodology (and their combination) quite complex in their application and presented in considerable detail in the book. Some understanding of this is desirable, so these notes were eventually conceived as a kind of 'reader's guide', clearing out some of the 'thicket' of detail in order to obtain a clearer view of the main ideas and results. I have therefore, and as noted above, adopted the position of a 'referee' asked to review the chapters with a view to journal publication. This paper deals with Chapter 6, the male graves. Discussion of the female graves, in Chapter 7, presents additional, more statistical, complications and forms the basis of a separate paper.

The editor of the hypothetical journal commissioning this would also engage an Anglo-Saxon specialist to review the more purely archaeological aspects and an expert on radiocarbon dating. I ask to be forgiven for archaeological and scientific mis-understandings. Section 2 provides a brief introduction to the main statistical ideas central to Chapters 6 and 7 of the book, discussed more fully in the context of Chapter 6 in Section 3. In Section 3 I'd originally intended to confine myself to aspects of the correspondence analysis and Bayesian models of Section 6.4.1 (Section 3.1 below) but was led inexorably into the later parts of Section 6.4 and Section 6.5 and areas of science related to radiocarbon dating about which I know considerably less than I do about the statistics. Section 6.5, however, is important, and the conclusion of Section 6.5.4 – an expressed preference for dates based on a purely terrestrial calibration curve – seems to me to be shaky on purely logical grounds. My thoughts on this are therefore offered rather tentatively in Section 3.3.2 below. For completeness a brief comment on Sections 6.3 and 6.6 of the book, concerning dating within artefact-classes, is provided in Section 3.4 below.

2 Methodological preliminaries

2.1 Correspondence analysis

Correspondence analysis (CA) is routinely used in archaeology and introductory accounts for an archaeological audience are available (e.g., Baxter 1994, 2003, Baxter and Cool 2010, Shennan 1997). As usually applied it takes a table of data and reduces it to a 'picture' - simply a two-dimensional plot - which is inspected for interpretable archaeological patterns in the data.

The raw data, available in the archive deposited with the *Archaeology Data Service* can, with some manipulation, be extracted in the form of a table of grave-assemblages by artefact types, entries in the table being counts of an artefact type within a grave. This is what is called an *abundance matrix.* Mostly the counts are 0 or 1 and, for the purposes of analysis, counts greater than 1 are converted to 1, so that table entries simply record the presence (1) or absence (0) of an artefact-type in a grave. This is an *incidence matrix.* It is necessary for each row and column of the incidence matrix to have at least two non-zero entries.

For applications of the kind used in the book the pattern sought – and often expected – is that of a parabola or 'horseshoe'. In 'successful' applications, and the ideal case, an unambiguous ordering of the rows of the table (grave-assemblages) can be read around the horseshoe. This ordering is called a *seriation* of the data; with luck, and often what is desired, the seriation can be interpreted as a chronological ordering, external evidence being needed both to impose this interpretation and determine the early and late ends of the sequence.

A similar plot can be obtained for the columns of the data table (artefact-types). The two plots can be compared (imagine that they are overlaid) to see, for example, which artefact-types characterise the late and early ends of the sequence.

The other kind of pattern often looked for – and it can co-exist with the horseshoe – is some kind of clustering in the data. This is exploited in the book, and it is fundamental to the way CA is combined with Bayesian modelling, to define phases to which individual grave-assemblages can be assigned. This kind of clustering, associated with phasing, is referred to here and in the book, as based on the two-dimensional CA plot. It is possible to obtain a one-dimensional ordering by 'collapsing' the data points onto the first axis according the their co-ordinate values for that axis. This ordering underpins the derivation of 'leading-type' phasing as an alternative to direct interpretation of the two-dimensional plot. Further comment on this is provided in Section 3.

Once data are extracted in a suitable form it is easy enough to reproduce the CAs illustrated in Chapters 6 and 7 of the book. Baxter and Cool (2010) provide a guide, written with archaeologists in mind, about using `R` to do CA[2]. The paper has been used, and found to be useful, by archaeologists with little quantitative training. That is, this is the sort of analysis that you can easily do at home. The apparent 'complexity' of the analyses in Chapters 6 and 7 stems mainly from the iterative way CA is applied in pursuit of the horseshoe, and this is the subject of Section 3.

2.2 Bayesian modelling of radiocarbon dates

The archaeological literature is replete with applications of Bayesian modelling applied to radiocarbon dates. Despite this the suspicion exists that they are not always as well-understood as one would wish.

An individual radiocarbon date is determined with some uncertainty and can be represented in the form of a probabilty distribution or curve. To convert to a range of calendrical dates

[2] `R` is open-source and very powerful state-of-the-art statistical software I used to replicate analyses in the book; software written for `Excel` exists but I have not used it.

calibration is needed. This produces a calibration curve showing how probabilities are distributed over the date-range. The curve can be rather 'strung out' and 'lumpy' (multi-modal) making simple summaries in the form of a 'preferred' date problematic and sometimes meaningless.

What is typically done is to summarise the calibrated distribution in terms of 68% or 95% probabilities (highest posterior density intervals), which can be 'disjoint'. For example, from Table 6.1 the 95% probability for UB-6474 (Dover Buckland, grave 264) is summarised as Cal AD 425-495 (51%) and Cal AD 535-585 (44%). What this implies is that mid-5th and mid-6th century dates are plausible and more-or-less equally likely, but that dates around the turn of the century and early part of the sixth are not[3].

Bayesian models use prior information to 'sharpen' inferences about date. For instance, very simply and hypothetically, in the example above, if it was known with certainty that the cemetery was 6th century the earlier dates could be discounted. More subtly, if phasing information is available (and correct - i.e. the relative dating of some pairs of graves is known/assumed) this allows constraints to be placed on the individual calibrated dates. This often has the effect, after Bayesian modelling, of 'truncating' the extremes of the calibrated dates and removing some of the more unsightly lumps. For example, given three phases A, B, C from earliest to latest, the bulk of the probability for a grave in phase B, *after* modelling the relative dates implied by the phasing, is 'forced' to be spread over ranges later than those for phase A and earlier than phase C[4].

In principle, if the CA produces a perfect horseshoe, an exact relative ordering of date could be used as prior information to be used as input into the Bayesian model. This ideal is not attained and relative ordering at the level of phasing is sought from the CA with some emphasis on making this as fine as is realistically possible[5]. Phasing was accomplished in two ways. Most simply the CA is inspected for natural gaps in, or thinning of, the horseshoe. This can also be thought of (loosely) as looking for 'clustering' in the *two-dimensional* scatter of points provided by the CA. The second form of phasing is based on a *one-dimensional* ordering of the grave-assemblages determined by their position on the first axis of the CA. The phase-boundaries in this one-dimensional ordering are determined 'on the basis of the introduction, predominance or demise' of what are called 'leading types' (p.254)[6]. The CA plots labelled by phase in Chapters 6 and 7 use the phasing determined directly from the CA, rather than that based on leading types, though frequent reference is made to the latter in the text (e.g., Section 6.5.4).

Fitting Bayesian models can be computionally intensive; it is noted on p.85 that several weeks, and in some cases months, were needed to fit models described in the book. Unlike CA it is not something to attempt at home.

3 The male graves

3.1 Book Section 6.4.1

The focus of this section in on Section 6.4.1 of the book (pp.251–289) where CA is applied to the male grave assemblages. The bottom line is that many readers may find it sufficient to read pages 251–257 to get a good idea of the analytical strategy; accept that it eventually leads to the definition of five phases (on the basis of the CA and of leading types); and jump to pages 285–289

[3]A problem to be found in the literature is that such results are sometimes interpreted in the light of whatever archaeological 'theory' an investigator favours. For example, and hypothetically, an archaeologist whose theory favours the early date will cite the calibrated data as 'consistent' with the theory, ignoring the almost equally likely probability that it is wrong. Citing 68% intervals that conform to preconceptions, when the 95% interval doesn't, is another form of biased interpretation.

[4]This does not preclude overlap between date ranges in different phases, and things can go wrong if the phasing is incorrect or there are problems with the radiocarbon date. This kind of thing can be investigated after modelling by inspecting the model (as graphically displayed) and using measures of 'agreement' between individual grave-assemblages and the model, as is done in the book (see Section 2.4.5 and p.85 in particular).

[5]What is done, pp.294–295, is to determine an ordering based on the first CA axis only and use this as input to a Bayesian model to determine how well the chronologies implied by the CA and the subseuqent modelling match.

[6]The reader is referred back to Section 3 where, as far as I can see, *leading types* (emphasis in the original) receives a single *en passant* mention on p.63 with a further reference back to Section 1.3 where the term occurs on p.31.

for the end results. That is, it is important to arrive at the right station and platform to board the train, but once it is running stations on the way (Figures 6.23–6.48 and much of the text) need only be glanced at by the traveller primarily interested in what awaits at the terminus.

The discussion of the chronology of the male graves can be read as a linear narrative, doubtless imposed on what was in practice a 'non-linear' sequence of 'experiments'. The discussion is admirably thorough to the extent that one could, if one wished, replicate each stage of the CA analysis used to develop the preferred seriarion in Figure 6.49, but most readers may not want to do this. The downside, consequent on this thoroughness, is that the chapter makes for 'dense' reading.

The narrative constructed below, in the form of 'stages', condenses that of the book. I have borne in mind a hypothetical editorial injunction to identify 'unnecessary' figures that could be omitted with little damage done to understanding.

Stage 1 is fundamental to what follows in Section 6.4.1; the story begins with an analysis based on 38 artefact-types (shield bosses and spearheads)[7]. These types were selected because they are common and a strong chronological sequence was expected for the shield bosses, based on earlier analyses (p.252).

The initial CA seriation is less than perfect; this is remedied by omitting 13 grave-assemblages to get horseshoes that are almost as nice as you could wish for (Figure 6.21). The archaeological rationale for the omission is that the graves contain artefacts judged to be of some antiquity at the time of their burial; you get pretty much the same result if you simply eliminate points that don't sit nicely on the horseshoe.

Figure 6.18a is used to illustrate how the CA can be used to suggest two phases, which provides the prior information that informs the Bayesian model of Figure 6.19. This serves, also, to validate the interpretation of the seriation as a chronological one. Figure 6.21a is more relevant since it proposes three phases which, with minor variation, informs the next few illustrations of Bayesian models, starting with Figure 6.22. Figure 6.20 simply illustrates the effect of removing one 'anomalous' grave before removing all 13 and could be dispensed with in a more concise presentation.

Stage 1 is fundamental in the sense that it 'fixes' the horseshoe to which all subsequent additions to the data are expected to conform. *Stages 2–4* (pp.257–264) simply add different artefact types (3 seax and seax-fittings; 7 buckle and belt-fittings; and 7 sword and sword-fittings) to the data. This increases the size of the table of data being analyzed and includes more of the dated graves. This results in Figure 6.30 where it can be seen that the horseshoe remains pretty intact. At this stage of the proceedings the 'left-arm' of the horseshoe, representing the latest phase, is still rather thinly populated[8].

For a concise reading much of what occurs between Figures 6.22 and 6.30 might be omitted. A moderately detailed discussion of what are only intermediate stages in the development of the preferred seriation is presented. Nothing much untoward occurs to affect interpretation except that the three-phase model suggested by Figure 6.21 is tweaked slightly to incorporate some graves near the phase boundaries which don't otherwise fit well into the original phasing.

The point to emphasise is that, once it has been used to help validate a chronological interpretation of the seriation, the Bayesian models do not – at least as presented – feed much into the development of the seriation[9]. Having established methodology and provided a 'template' for the seriation, one could jump from Figure 6.21 to Figure 6.49 (the preferred seriation) before worrying too much about the Bayesian modelling.

---

[7] It says 37 on p.252 but there are 38 in Figure 6.18b. In replicating this I picked up slightly more graves using these 38 types than the 216 given in the text but this seems of little consequence. Some of the shield boss types given in the database, and discussed in Chapter 5, needed to be amalgamated to get the types used in the CA.

[8] For those unfamiliar with CA the orientation of the graphs is arbitrary, since the signs of the plotting coordinates can change from one analysis to the next (e.g., compare Figure 6.20 and 6.30). This is nothing to worry about. Purists insist on equal scaling of the axes in CA so that the graphs can be read as maps showing the 'distances' between graves. This interpretation is sacrificed in the figures because axes are not equally scaled, but it makes the graphs easier to read.

[9] I will discuss the relation between the CA and Bayesian modelling in a little more detail in the conclusion to the paper on female graves.

Stage 5, from p.263 on, bridges the gap between the two figures; essentially it involves adding as many other artefact-types and grave assemblages as possible to the data set without disturbing the pattern established at Stage 1 too much. This seems to be driven by a desire to include as many of the dated graves as possible in the seriation, something explicitly acknowledged on p.273 and p.277. Apart from increasing the size of the data set analysed, with the inclusion of more dated graves, the general effects of this are to (a) populate the thinner and later end of the seriation with more graves, including dated ones; (b) 'coarsen' the horseshoe a bit; (c) permit finer phasing; and (d) produce more 'precise' estimate of date-ranges.

To achieve this, extra complexity is introduced into the selection of graves/artefacts for inclusion in the seriation. Up to the end of Stage 4 grave assemblages are treated 'as is' and included or excluded from the analysis on the basis of conformity with the horseshoe established at Stage 1. In Stage 5 some effort is made to include in the seriation graves that initially do not fit by excluding individual artefacts from the grave assemblages that are judged responsible for a 'lack of fit' to the seriation. Call this 'data modification' – not to be construed pejoratively. Examples of this and other forms of 'modification', in order of occurrence, are

p.264 Type BU7, considered and rejected at Stage 3, is reintroduced improving 'coherency in the thinner, later arm' of the horseshoe, but requiring the removal of three grave-assemblages justified on the grounds of 'anomalously late survival of an otherwise relatively early type of spearhead'.

p.271 'Typological modification' is introduced at this point with the three smallest objects classified as SX1-a removed from analysis. This seems to be done because it leads to a more satisfactory 'sequential phase model'.

p.273 Here the process of adding additional grave-assemblages, 'if necessary' excluding 'incidences of what appeared to be anomalous artefact-types within them', is described. Discussion follows with a summary of graves-assemblages subjected to modification provided in Table 6.2 (p.274). At Stage 1 the first grave-assemblage to be dropped from the initial seriation was SutH01. In what is described as a 'major step' effort is put into re-incorporating this into the seriation, judged to be satisfactory after 'three adjustments'.

p.276 Seax type SX1-c is excluded from grave assemblage Fo2 as it 'may have been an old object refurbished with more up-to-date seax-fitting' allowing 'SX-1c to fall into the earlier part of the correspondence analysis parabola'.

p.276 Grave-assemblage WG01, judged to be unsatisfactory because of an 'anachronist' incidence of SB1-b, is dropped because after removal of this it only contains one artefact.

The CA plots between pages 264 and 284 simply illustrate the effect of some of these modifications, and the incremental addition of artefact-type.

While all this is going on the phasing is gradually being refined and quite a lot of Bayesian models are illustrated. A 'more active interpretation and exploration ... with a view to finding a finer chronological phasing' is announced on p.270 (at this point a three-phase model is being used with two of the dated graves being allowed to 'float' between two of the phases because otherwise aspects of the model are not satisfactory). The effect of the various changes to the seriation is to remove the need to allow two graves to 'float' and to sub-divide the central phase into 'three further sequential phases' (p.271). The phases, at this point named x, y1a1, y1a2, y1a3, y2, are renamed p, q, r, s, t on p.275, and five phases labelled in this way are retained in the final preferred correspondence analysis on p.286.

Bayesian models using phasing based on leading types are intermittently interleaved with those derived from the two-dimensional CA graphs and there are some differences, but the broad story is the same. Initially a three-phase model (a, b, c) based on 'chronology' associated with the introduction of shield boss types is used (p.254). This is eventually expanded into a five-phase model, (a, b, c, d, e), defined additionally by the introduction of sword types (p.285). The two phasing schemes are judged to be 'very similar' providing 'robust' date estimates for phase boundaries (p.289 and compare Figures 6.50 and 6.51 for the two partitions). These statements about robustness and similarity concern estimated *phase boundaries*. However, phases (a–e) and (p–t) are not congruent in the sense that phase a contains the same graves as phase p, and so on[10]. This is taken up in the next section (both here and in the book).

3.2 Book Section 6.4.2

The Bayesian models in Figures 6.50 and 6.51 are refined by adding dated graves not included in the final CA (pp.289-293). This is possible because graves not used to inform the seriation can nevertheless be assigned to subsets of the phases developed, providing additional prior information. Eleven graves, additional to the 28 used in Figures 6.50 and 6.51, are listed in Table 6.3 and the models represented in Figures 6.53 and 6.52 for the two-dimensional CA and leading types phasing become the preferred models. Typically the 95% intervals for the phase boundaries are shortened by 5 years, with a more noticeable shortening at the start of the sequences.

This dealt with, the issue of differences between the two forms of phasing is addressed (pp.293-296). Phase-boundaries for the two schemes (as opposed to their estimated range) as determined by ordering the data matrix on the first CA axis, do not coincide. For example, of the dated graves SPTip250 and EH012 occur in the same phase r, using the two-dimensional CA map phasing, but different phases, c and d, using the leading type phasing. These are described as 'liminal areas of the sorted data-matrix' on p.294 where it is suggested that the two models are approximately 95% similar since 36/38 of the dated graves used in the preferred models do not show potentially substantive difference in phasing. An alternative method of assessing 'chronological accuracy' is also described (pp.294–295) leading to an assessment of 89% 'accuracy'. The methodology is described in the Appendix, along with a simpler approach. This is not of real immediate concern since, using any of the measures, results are satisfactory, but is of more interest when it comes to the female graves discussed in the companion paper.

Possibly the most interesting part of Section 6.4.2 is the discussion on pp.295-296 of how the models might be used to phase further examples of graves not in the study. In principle the CA itself can be used for this since new data can be 'plugged-in' to see where it fits. This demands however, that the protocols followed are identical to those used in the study so this may not be a practical proposition. Therefore, with mandatory caveats, it is suggested that 'the definition of chronological phases by leading types has the advantage of being easier and quicker to use' (p.295). That is, phasing is to be based on artefact-typology, chronologically interpreted. The selection of artefact types to use is informed by the CA but, once established, the CA can be viewed as 'scaffolding', used to erect the typological scheme and which can then be stripped from it. The first paragraph of Section 6.5 provides a very concise summary of all this.

3.3 Book Section 6.5

3.3.1 Sections 6.5.1–6.5.3

Section 6.5 is concerned with aspects of the representativeness of the data and reliability of analysis under the heading 'Sensitivity Analysis'. It is concluded (pp.296–298) that there is some geographic imbalance concerning the geographical distribution of the dated graves relative to the seriated sample[11], but on the whole things are deemed to be OK. Sections 6.5.1 to 6.5.4 variously look at the effect of using an alternative calibration curve; varying an underlying modelling assumption about burial rates[12]; parallels with dated continental graves; and dietary effects.

The first two ways of varying the analysis produce nothing to worry about (see the first paragraph of Section 6.5.3, p.301). The inclusion of continental parallels equally does not throw

[10] You need to look at e-Fig. 6.6, referenced in the text, and the sheet named 'Sorted matrix' to see this.

[11] This is at slight variance with the statement in the concluding Section 6.7 (p.336) that the dated graves prove to be 'spatially ... representative' but is not, I think, a major issue.

[12] This involves comparing a model assuming a uniform-rate of burial across the period with rates that can differ between phases. I shall discuss this in some detail in the paper on female graves where the issue is of much greater importance.

up any problems and the authors prefer to stick with the models illustrated by Figures 6.52 and 6.53 to 'keep the English chronological sequences independent of those on the Continent' (p.305). Although they don't quite express it this way there is, interestingly, a 'get out clause' to the effect that it is dangerous to assume that similar artefact types found in widely separated geographical locations 'have strictly contemporary date-ranges'. This leads into a brief excursion about the possibility of 'regional distinctiveness' in the male graves but this is not considered an issue (p.309).

3.3.2 Section 6.5.4 – Dietary effects

Older readers of the *Guardian* newspaper may remember that resident cartoonist Steve Bell's penguin had a world view that could be summarised as 'It's all about fish'. Sections 6.5.4 is all about fish. It is very 'densely' written, even opaque at times, and hard to follow. I don't pretend to understand the science much but the logic of the argument, which I think is flawed, can be followed on the basis of the authors' own analysis.

In summary, the analyses of Section 6.4.1 use a terrestial calibration curve which – if I understand it – assumes Anglo-Saxons don't eat fish. Eating fish will affect the radiocarbon date; may need 'correcting' for; and, if the correction is needed and matters, will result in later dates than the terrestial curve. Two methods of correction are evaluated, of which only one – the formal isotopic mixed balance model, ISOSOURCE – is discussed here[13]. It is judged in Section 6.5.4 to be the main contender. The authors eventually 'judge that the models based on fully terrestial calibration provides the most realistic date-estimates for the corpus of Anglo-Saxon graves included in this study at present'. I am not convinced by the logic leading to this conclusion and the rest of this section is an attempt to explain my unease. As well as the two 'competing' methods of 'correction' there are two 'competing' phasing models, the leading-type phasing and that based on the two-dimensional CA plot[14].

On p.58 it is stated that the 'stable isotopic data indicate that the protein sources in the diets of the dated individuals derived overwhelmingly from terrestrial sources'. The term 'overwhelmingly terrestrial' is repeated, with reference to the 'entire Anglo-Saxon data-set', in Section 4.3.1 (p.117) on 'Dietary analysis'. If 'overwhelmingly' is to be understood in the sense that most people use the term then the need to investigate alternatives to the terrestrial model (and hence Section 6.5.4) would seem superfluous.

In fact, what is stated in Section 6.5.4 (p.310) is that 'ISOSOURCE suggests that, overall, Anglo-Saxon people had *largely* terrestrial plant and protein-based diets, with freshwater fish *averaging* about 26% and marine fish from 1.4% to 12.7%' (my emphases). 'Largely' is not the same as 'overwhelming'; a 26% average for freshwater fish implies that quite a few graves must have more than this average so that, to my layman's eye, this seems to imply that fish consumption is hardly negligible.

The problem of dealing with *freshwater* fish is not ignored, but the way it is dealt with has that effect. The relevant discussion is on p.310 and concludes that there 'is no information on freshwater fish in [the region studied]' so no attempt is made to model the effect of consumption of freshwater protein resources. It is stated, however, that 'the possibility of such offsets must be acknowledged so that the radiocarbon dates used are, if anything, biased towards older dates'.

[13]See p.310. Both methods are based on estimating past diet from stable isotope values. The method discussed in detail in the book but not here involves 'linear interpolation from the δ^{13}C values of the dated individuals and the δ^{13}C values of their available food sources'. Discussion of the results from this method occupies much of pages 310–320, including Figures 6.63–6.67 and Tables 6.9 and 6.10. The eventual conclusion is that the terrestrial method is to be preferred and this is based partly on its greater compatibility with archaeological/historical evidence (p.320). A reader prepared to take this on trust might thus omit reading many of these pages, although the discussion of historical date ranges for some graves used as 'controls' to assess the models might be of interest and feeds into evaluation of the ISOSOUCE analysis. This latter method *is* judged to be as compatible with the archaeological/historical evidence as the terrestrial analysis.

[14]Results from the four combinations possible with these are each described in detail in the text and some of the discussion is repetitive. Since the ISOSOURCE method is clearly preferred and the phasing models give broadly similar results, only the results for the ISOSOURCE method and leading-type phasing are commented on below.

Focusing on the ISOSOURCE analysis with leading-type phasing (pp.325–327), the bones of the story are that, compared to the terrestrial analysis, estimated phase-boundary dates are 15-25 years later (p.325 and Table 6.11); that this is not much affected if dated Continental graves are included in the model as a 'control'; and that the model is broadly compatible with historical dates that *might* be assigned to three of the English graves. As the authors conclude, there is 'little reason to state a preference' between the models 'on archaeological and historical grounds' (p. 327).

So, to the unbiased observer, it might seem that the two contenders – ISOSOURCE and terrestrial models – are about equal on points but provide substantively different results as far as phase-boundaries are concerned. Most of p.327 is devoted to explaining why the terrestrial model wins on points (though not a knock-out, even a technical one) and is where I find the logic unconvincing. The eventually stated preference for the terrestrial model arises from concerns about technical aspects of the mixed-source models – again in the authors' words 'we are reluctant to prefer the mixed-source models based on the ISOSOURCE calculation in the absence of more convincing evidence of their validity' (p.327). This is followed by a list of technological 'complexities' or 'uncertainties' that afflict the ISOSOURCE calculations which therefore, it is judged, lack any great precision (expressed as being 'reflected in large errors quoted on the dietary estimates' on p.327).

The 'logic' implicit and explicit in Section 6.5.4 is as follows:

1. Terrestrial calculations are biased towards older dates if Anglo-Saxons eat fish (see p.57 for a concise summary of this issue).
2. Some Anglo-Saxons do eat fish.
3. Therefore the calculations are biased (my gloss on this).
4. Information to correct for freshwater fish consumption isn't available. The hope is that any effect is small.
5. Correcting for marine consumption gives substantively different results from a terrestrial model.
6. There is no external evidence for preferring one model over the other.
7. The terrestrial model is preferred because the ISOSOURCE dates can't be determined with 'precision'.

This is a simplified summary, but I hope not a caricature. Unless you think the potential 'fish effect' can be entirely discounted – and I do not think this is claimed – a model known to be biased is preferred to one that attempts to 'correct' for the bias because of 'uncertainty' associated with the precision of bias correction.

This is not, as it happens, a simple issue and I have some sympathy with the authors. Their logic, nevertheless, seems to me to be faulty and arises because of a desire to fix on a 'preferred' model when the evidence does not really support the expression of a single preference.

I find it helpful to think in terms of analogies with more formal statistical thinking and hope I can be indulged in an excursion into the ideas involved. There is a fundamental difference between 'accuracy' and 'precision'. For the purposes of exposition introduce the idea of a 'correct' uncalibrated radiocarbon date whose 'true' value is Ω. Imagine (this is a 'thought experiment') that replicate determinations are made of the date using each of two different methods (terrestrial and ISOSOURCE). (In practice only one measurement is made, call it A using the first method and B using the second method but pretend we can replicate.) If, on average, measurements A recovered the true value Ω it is an unbiased (accurate) estimate, otherwise it is *biased* (inaccurate).

In terms of the logic outlined above, terrestrial measurements, which we associate with A, are biased; furthermore we know the direction of bias – they give dates that are too old – but don't know the magnitude of this. Were such measurments repeatable, different values of A would be obtained and this could be measured, providing an estimate of the precision of A (i.e. the spread of values obtained). In practice the 'error' attached to a single determination A is an attempt to measure its precision. It is hypothetically possible for replicated measurements to return the same value of A which, in the circumstances envisaged, means we have a biased (inaccurate) measurement determined with perfect precision.

Conversely, let method B – ISOSOURCE – return a value that is unbiased so that on average, if we could replicate, we recover the true value Ω. This is 'accurate' but if the variation in hypothetical replication is large (in practice the reported 'error') then precision will be poor. In statistical theory this is a situation, where one has to choose between A (inaccurate, precise) and B (accurate, imprecise), that has engaged a lot of attention. It is possible to combine measures of accuracy and precision into a single measure that is used to guide the choice between A and B. What it amounts to is that the possibly conflicting desires for accuracy and precision may need to be balanced and there are circumstances when an inaccurate measure will be preferred when the bias (inaccuracy) is *small* if the precision is much greater than that of an unbiased (accurate) estimate.

The preference expressed for the (biased) terrestrial estimate seems to me to be underpinned by this sort of thinking – whether consciously or not – but begs some unanswered and possibly unanswerable questions[15]. One issue is that we don't know if the bias is small enough to, in effect, ignore. If method B is unbiased, even if imprecise which seems to be a main objection to it, then on average it will improve on method A. In this circumstance it is not obvious that method A should be preferred[16].

The other possibility is that method B is itself biased. It 'corrects' in the right direction regardless. If it is biased and 'under-corrects' then it is nevertheless correcting in the right direction and is to be preferred. If it is 'over-correcting', and producing dates that are too late then the issue would be whether the bias in favour of later dates is greater than that in favour of earlier dates inherent in the terrestrial model.

My reading of the second column of p.327 – which is where the reasons for preferring the terrestrial model are outlined – is that problems concerning the 'validity' of ISOSOURCE calculations that have been expressed can't, collectively, be related easily to concerns about bias. If there is an assurance that any bias in terrestrial calculations is small then, OK, it might be preferred. I see no obvious assertion of this, and if it could be made easily Section 6.5.4 would be unnecessary. Otherwise the main reason for preferring the terrestrial model would be if the ISOSOURCE model was seriously biased in favour of later dates and more so than the terrestrial model is in favour of earlier dates. I don't think this argument is made.

The conclusion to be drawn from this is that it is probably a mistake to try and identify a preferred model. Unless the effect of fish can be discounted *a priori* it must be accepted that the terrestrial model gives dates that are too old; we have no real idea by how much; but as far as phase-boundaries go it could be by up to 15–25 years.

In fact the authors eventually say as much. That terrestrial calibration is preferred is reiterated, in identical terms, in the discussion of the female graves and several other places. However, in Chapter 8 in discussion of the date of the end of furnished burial (Section 8.2.2, p.468), there is an almost 'throw-away' comment that dietary effects of the order of 10–20 years can't be discounted, followed by an even-handed discussion of the consequences of this compared to those to be derived from the terrestrial model.

This was obviously written later than Chapters 6 and 7; the thought occurs that in terms of 'reader comprehension' it should have invited a rewrite of the sections on dietary effects in the earlier chapters. Rather than subjecting readers to the detail to which they are exposed, an exposition in these earlier chapters that made clear that two alternatives to terrestrial calculations

[15] There are difficulties with the analogy I am trying to draw. In practice, and the whole point about radiocarbon dating, we are dealing with date ranges and calibration. A shift in the uncalibrated date between methods A and B does not simply shift the distribution of the calibrated curve but changes its shape as well, and the shape of the calibrated curves also depends on A and B. This introduces complications but I think the point I'm trying to make here stands.

[16] It may be worth noting at this point that, whatever the precison associated with individual graves is, the *range* of the dates (as opposed to actual dates) for the terrestrial and ISOSOURCE phase-boundaries are broadly similar (Table 6.11).

were investigated; that one was found wanting (without all the detail) but the other not; that the competing ISOSOURCE models were subject to the same tests as the terrestrial calculations; and that the terrestrial and ISOSOURCE model were 'competing' would have sufficed. There is no necessity to state a preference, and the exposition could have been much shorter and comprehensible without any sacrifice of integrity. I was, regrettably, reminded of student theses imbued – understandably but not always appropriately – with a desire to impress on the reader the thoroughness of investigation, without due selectivity in highlighting what the reader really wants and needs to know and what is important.

3.4 Section 6.3 and Section 6.6

Other than its independent interest I wasn't initially and entirely clear about the role of Section 6.3, 'Male Artefact Types', in the context of what follows. Where sufficient examples exist in dated graves a variety of Bayesian models are fitted to individual artefact-types providing chronological information about their 'currency'. There is an excursion into phyletic seriation in Section 6.3.2 (p.248) – the idea that within an artefact-class single traits 'evolve' over time 'revealing chronological variation in Anglo-Saxon artefact-types'. That is, I think, within some artefact-classes it may be possible to impose a chronological sequence of development on artefact-types within the class. This appears to feed into the leading-type phasing used later.

I will limit further comment to shield bosses because of the role they play in developing 'leading-type phasing' in conjunction with CA. The initial venture into such phasing is noted on p.254 where three phases are defined roughly on the basis of the introduction of artefact-types within the SB2, SB3 (phase a), SB4 (phase b) and SB5 (phase c) classes[17]. This selection is justified by reference back to the phyletic seriation in Figures 6.15–6.17 (pp.248–250)[18]. My reading of this is that the classification was constructed with an eye to chronology and in the expectation that a chronologically interpretable typology existed. If so, the role shield bosses play in later phasing is not a surprise. The CA 'validates' what you suspect is already 'known', but in conjunction with the Bayesian models allows more precise dating for (the introduction of) the types, as well as refining phasing with the introduction of leading types from other artefact classes.

Section 6.6 revisits the dating for individual artefact classes *after* the CA. I shall comment only briefly though, given the dating provided for types/classes, I assume this section will be of interest to typological specialists.

Graves are sorted by relative chronology on the basis of scores on the first CA axis. From this a 'model' of the chronological relationship between the currencies of different artefact-types is developed. These can be quite complicated; that for shield bosses is illustrated in Figure 6.78. This chronological model of currencies then provides additional prior information that is fed into Bayesian models for those graves containing the artefact-types that define the model of currencies. Figure 6.79 (p.335), for example, shows the Bayesian model for seven shield boss types.

The English chronology suggested by the Bayesian model is compared briefly with that of Continental chronologies where it is concluded that the two do not align well so that the English sequence must 'stand alone' (p.334). The extent to which English and Continental chronologies appear to align for other artefact-classes is summarised on p.338.

4 Postscript

My original intention was simply to see how easy it was to reproduce the CA analysis in Section 6.4.1 and whether the preferred seriation might have been established via a quicker route (which,

[17] SB4-a1-2 is associated with phase c.

[18] I felt this might have been spelled out a little more explicitly. My initial thought, before I read the text more carefully, was that the selection of leading types represented some sort of 'typological truism' that did not need spelling out for Anglo-Saxon scholars. In fact the shield boss classification appears to have been developed 'from scratch' (p.148) for the project. The typological scheme is outlined in Table 5.2 on p.152. Some of the thinking behind the construction of the scheme is outlined on p.149. This is where you are told that some artefact-types were paired for the purposes of seriation.

in the event, I have yet to pursue). In order to understand the implications of what was done I found myself reading the rest of the chapter, as well as other parts of the book necessary to follow some of the argument. In writing these notes I found myself using words like 'dense' and 'opaque' to describe the style of presentation. To be frank, reading the book makes for pretty heavy going; I know I am not alone in thinking this, and it was that which prompted me to put my thoughts in writing.

At the time of writing I'm half-way through reading a biography of Charlie Chaplin. He was in the habit of shooting enormous amounts of footage and then rejecting most of it in the editing process. In *The Gold Rush*, for example, 231,505 feet of film were edited down to 8,555 feet. Extant footage exists of scenes rejected by Chaplin from his films that are hilarious and other directors would have died for. They were rejected, however entertaining, because he did not wish to distract from the more central elements of a film.

Chaplin cut his material ruthlessly, for artistic purposes. Erich von Stroheim, one of the other most noted directors of the 1920s, was, by contrast, a reluctant cutter for precisely the same reasons. Accounts vary, but for his most notorious film, *Greed*, one story is that he produced a version that ran for over 8 hours; was prevailed on to edit it but was only prepared to do so down to about 4 hours; and was then removed from the film by the studio who edited it down to a bit over 2 hours. One view is that the studio butchered a masterpiece; another is that nobody would have watched it had it not been reduced to a length that the viewing public could cope with.

My own thought about the book is that it looks like an unmolested von Stroheim that would have benefited from a much more ruthless Chaplineqsue editing approach. As I understand it the book is regarded as important and will therefore be read, but one suspects that large chunks of it will go unread by those who feel obliged to tackle it. Some parts of the book would be more readily subject to précis than others but some of the more statistical/technical sections might have been produced as separate working papers, with only the essentials summarised in the main text. Section 6.4.1, which is all I'd originally intended to look at, occupies some 38 pages and includes 33 figures of which I'd judge six to be essential. My academic training is supposed to equip me to read this sort of thing, but I did not find it light reading; I do not know what Anglo-Saxon scholars without a quantitative/scientific training will make of it. In my capacity as a hypothetical reviewer of Section 6.4.1 submitted for journal publication (and it could form the basis of such a paper) I would accept it on the basis of its potential interest and importance, but would ask for a major revision involving considerable reduction, including omission of figures, more selectivity in the detail presented and so on.

A much shorter and much more readable and accessible book is yearning to be born from the body of the present text.

References

Baxter, M.J. 1994: *Exploratory Multivariate Analysis in Archaeology.* Edinburgh: Edinburgh University Press.

Baxter, M.J. 2003: *Statistics in Archaeology.* Arnold: London.

Baxter, M.J. and Cool, H.E.M. 2010b: Correspondence Analysis in R for archaeologists: an educational account. *Archeologia e Calcolatori* **21**, 211-28.

Bayliss, A., Hines, J., Høilund Nielsen, K.H., McCormac, G. and Scull, C. 2013 *Anglo-Saxon Graves and Grave Goods of the 6th and 7th Centuries AD: A Chronological Framework.* London: Society for Medieval Archaeology.

Shennan, S. 1997: *Quantifying Archaeology: Second Edition.* Edinburgh: Edinburgh University Press.

Appendix

On the chronological accuracy of the seriation

A strict relative chronological ordering for the seriated graves can be based on coordinates on the first CA axis. A Bayesian model is fitted with this ordering providing prior information, and graves with poor agreement are sequentially deleted from the model until those remaining are judged to have satisfactory agreement. Three of 28 graves are so deleted leading to a measure of 'chronological accuracy' of 25/28, or 89% (p.295).

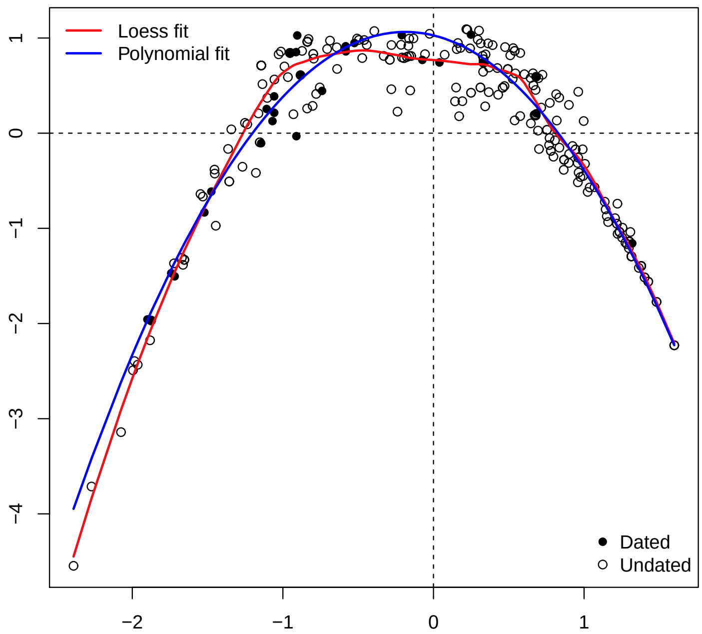

Figure 1: *The preferred correspondence analysis for the male graves labelled by whether or not a grave is dated, with fitted non-parametric and parametric regression curves.*

An alternative and possibly simpler method of measuring chronological accuracy is presented here, which will assume more importance in the companion paper on the female graves. Figure 1 above shows the preferred CA (Figure 6.49a, p.286, in the book) labelled by whether or not the grave is dated. It has a satisfactory parabolic shape and a second-order polynomial regression model has been fitted to it[19] (the blue curve) with 'goodness of fit' of 93.1% as measured by the coefficient of determination R^2. An alternative method of summarising the regression is shown using a non-parametric regression fit, sufficiently close to the polynomial fit to justify using the latter as a summary description[20].

The R^2 statistics can be viewed as measuring the quality of the seriation as a whole. The measure of chronological accuracy used in the book, which requires the sequential fitting of Bayesian

[19] I believe the latest version of Torsten Madsen's `CAPCA` program, an `Excel` add-in for CA, allows something like this. As described (http://www.archaeoinfo.dk/), second-order polynomial fits are used. I understand Madsen's software was used for the book, but the latest implementation will not have been available. I don't recall seeing published applications of the idea, though it would not surprise me if they exist.

[20] The loess method using the `loess` function in `R` was used with defaults except that a `span` of 0.25, which controls the level of smoothing, was used. A 'psuedo-R^2' analogous to R^2 can be constructed from the summary output and gives a value of 95.2%.

models, is directed to assessing the quality of seriation of the dated graves only. A simpler method of doing this, that avoids the need for sequential fitting would be to correlate the rank-ordering of the graves inferred from the preferred Bayesian models with their rank-ordering inferred from the seriation. This is illustrated in Figure 2 where the rank-order obtained from the first axis of the CA is plotted against the rank-order inferred from the ordering of grave-assemblages in the preferred Bayesian model of Figure 7.52.

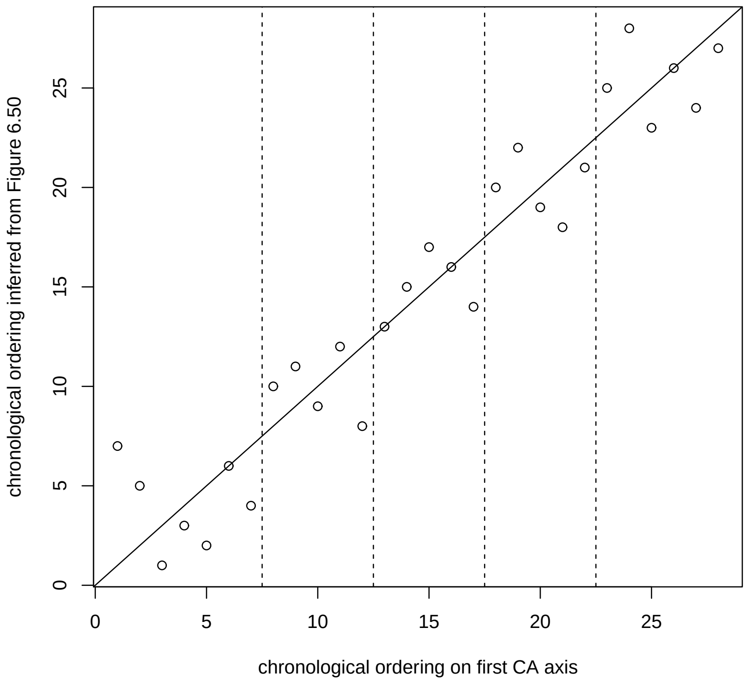

Figure 2: *The chronological rank-order of grave-assemblages inferred from Figure 7.52 plotted against the rank-order inferred from the first axis of the preferred correspondence analysis. Dashed lines indicate the phasing p–t reading from left to right.*

The solid line is that on which points should sit were the chronological accuracy perfect. The dashed lines indicate the phasing; that for the Bayesian model is determined by that for the seriation so the rank-orders are not independently determined[21]. Distances between phase boundaries cannot be interpreted in terms of calendrical dates – they simply reflect the number of dated graves in each phase.

That there is generally good agreement betwen the rankings is evident. Spearman's rank-correlation coefficient is 0.95 and the R^2 value from a linear regression fit is 91.1%. This is much the same as the 89% determined using the method in the book. They are different methods of assessing 'chronological accuracy', however, and need not be expected to be similar though they both tell the same story (this will be elaborated on in discussion of the female graves). The approach described here is simpler to implement since it avoids the need for sequential Bayesian modelling, but could be used as a starting point to investigate why some graves do not fit the pattern well, were that desired.

[21] That is, and for example, rank-orders in phase q are constrained in both orderings to be greater than those in phase p, etc.