Bayesian Modeling Practical

Rick Schulting

Using Bayesian Modeling

Baye's Theorem

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)}$$

P(A) = marginal or prior probability of A
P(B) = prior probability of B
P(A|B) = conditional or posterior probability of A given B
P(B|A) = posterior probability of B given A

But never mind this formula. What this all comes down to is that Bayesian modelling provides a means of modifying one's beliefs in the light of new information, i.e., beliefs about the likelihood of *A* are modified by having observed *B.*

In terms of ^{14}C chronology, we deal primarily with one of two situations (or, more often, a combination of the two):

1) In the first instance, we may have stratigraphic details (the 'prior probability') that represent additional information on the sequence of events. Bayesian modelling provides a way of formally using this information to improve the precision of the calibrated date ranges. It also quantifies the likelihood that a particular date is out of sequence, or that the sequence itself is improbable (e.g., there may have been significant disturbance in the deposits, etc.).

2) In the second case, we do not have reliable stratigraphic information. But (surprisingly to many) it is still possible, making some very basic assumptions, to use Bayesian modelling. The 'prior probability' in this case comes from information contained in the series of dates themselves, and in the properties of the calibration curve.

Case study 1. An imaginary court tomb somewhere in the west…

Just for illustrative purposes, say that we have a charred hazelnut sample (thereby avoiding any potential 'old wood' effect) directly underlying a Neolithic court tomb. It is assumed to relate to pre-construction feasting, and provides a determination of 4900 ± 30 BP (RJS-1).

Due to the bloody acidic soils that dominate Ireland, only a single human tooth survives from the chamber. It provides a determination of 4950 ± 30 BP (RJS-2).

First, calibrate these determinations, noting the 2-sigma (95% CI) ranges.

RJS-1	hazelnut shell	4950 ± 30 BP
RJS-2	human tooth	5000 ± 30 BP

But we can do better than this. The model, taking into account the stratigraphic relationship between the two samples, results in a substantial improvement. Note that this will not always be the case – much depends on the shape of the calibration curve over the relevant period.

Entering the model into OxCal 4.1

There are some short-cuts to entering data (dates and the model parameters), but in the first instance we will enter everything manually.

In OxCal 4.1, chose 'New' under the *File* menu. Double-click on the word 'Plot()' to name the file. A box entitled 'Edit' will appear to the left. Place the cursor between the brackets after 'Plot' in the Edit box and click. Type 'Court tomb model', and click on 'Ok'. The Edit box will close and the model box (on the right) will now read 'Plot(Court tomb model)'.

The default position for the cursor is the small blue circle (with interior black pulsing dot showing it is selected) under Plot, but if for any reason this is not selected, click on it.

The *Insert* menu on the left is used extensively, to set up the models and to enter 14C results. Click on it now – the default is an R_date, i.e., a radiocarbon date. We are not entering one of these at present, so click on the pull-down menu 'R_date' at the top left and select 'Sequence'.

Type 'Court tomb' in the 'Name' box and click on the '>>' symbol. This symbol is the equvialent of Return/Enter (which do not work in OxCal). (This can be re-named later by double-clicking on 'Sequence' in the model box, opening the 'Edit' box on the left, as for 'Plot'.)

Again, the default position for the cursor should be the blue box below 'Sequence'. Note that you should now have two nested boxes, with the 'Sequence' box inside the larger 'Plot' box.

Again under *Insert*, using the pull down menu at the top left (which now reads 'Sequence'), choose 'Boundary'. You will need to scroll up if you don't see it. Type 'start' in the 'Name' box and hit >> . Then type 'end' in the 'Name' box and hit >> again. Ignore the 'Expression' box.

We actually want to insert the dates between the 'start' and 'end' Boundaries, but it is often simpler to enter the two Boundaries at the same time. So we are ready for the dates. Single-click on the blue circle beside 'Boundary("end") in the model box on the right to position the cursor there (if you double-click you will open the 'Edit' box; just hit either 'Ok' or 'Cancel' if you do this by mistake).

Now use the *Insert* menu and select R_date. Type RJS-1 into the 'Name' box (this is usually for the 'lab code', but could also be a name of a site, or whatever, but don't make it too long), then click on the '14C Date' box to place the cursor there, and enter 4950, then 30 in the 'Uncertainty' box (this is the ± standard error term). Click on >> to enter the date, which should now appear in the model box between 'Boundary start' and 'end'.

Click on the 'Name' box back to the left, and modify this to RJS-2 for the second date. Enter 5000 and 30 and click on >>.

That's it! Our simple model is entered. The use of 'Sequence' tells OxCal that these dates are in chronological order ('earliest' is always first, by which we mean the stratigraphic order or other logical sequence, and NOT the actual date: here for example, the second date we entered is actually earlier than the first) and should be treated accordingly. The dates entered into a Sequence are always bounded at either end with the 'Boundary' function (it doesn't matter what we label these, but 'start' and 'end' are useful and descriptive).

We now need to save the file: select 'Save' from the *File* menu and enter a name for it by overwriting 'Untitled' at top left. Click on 'Save' at bottom right of this window.

We are ready to run the model, using 'Run' under *File*. So, select 'Run'. Depending the complexity of the model, OxCal will return a result almost immediately, or can take a considerable time. Our model is simple, so should just take a few seconds to return a set of results in the box to the right. Note that this now differs from the data/model entry page we used up to now, with a different menu to the left.

The two 'Unmodelled' columns shows the straightforward 2-sigma (95% CI) calibrated date range, without taking the model into account. The dates are negative (e.g., '-3787') when BC, and positive when AD. The 'Modelled' column shows the 2-sigma ranges taking into account the 'prior information' we entered into the model, i.e., the fact that we have good reason to expect that RJS-1 is earlier than RJS-2.

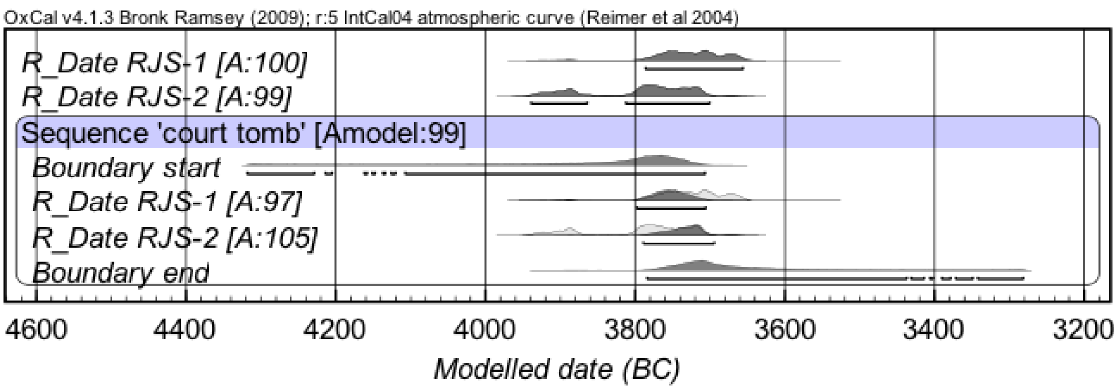
Take a moment to compare the two sets of results, and note how they differ.

To see the more detailed break-down of the calibrated ranges, select 'Log file' from the *View* menu. This opens a separate window with both 1-sigma and 2-sigma ranges in all their glorious detail.

Close this box, and select 'Plot single' from *View*. This opens a graph showing the probability distribution for the 'start Boundary'. Don't worry about the interpretation of this at this point. Click on the > button at the top right of the window to move to the next graph, which will show the calibrated result for RJS-1, against the appropriate portion of the calibration curve. The 'ghosted' distribution in light gray (or simple line if viewed in black and white) represents the standard, unmodelled calibrated date. The darker gray/black distribution represents the calibrated range taking the model into account. The 'agreement' value (97.1%) shown is a measure of how well this individual determination fits with the model's expectations. We'll come back to this shortly.

Click again on the > button to see the graph for the next determination. Take a moment to compare the unmodelled and modelled distributions, noting how and where they differ. Clicking again shows the Boundary 'end', which again we are not concerned with at this stage.

Finally, select 'Plot dates' under *View* to see all the results together in one graph. This is one of the most useful features in OxCal, and is the form in which multiple dates are usually presented. Various labels can be added, individual dates excluded, etc. Your graph should look something like the one below, though without the first two entries.



Bayesian models can also identify potential problems with our 'prior information', with what we thought we knew. Now, we will change RJS-2 to 5050 ± 30 BP to see what happens when the dates conform even less well to the proposed stratigraphy.

To do this, first click on the white arrow in blue circle at the top left, beside the OxCal version number (4.1). This brings us back to the data/model window. Now double-click on the text of 'R_date("RJS-2", 5000 30), which will open the Edit box to the left. Change the date to 5050 and click on 'Ok'.

Save and 'Run'. The results table will include 'Warning! Poor agreement…' and give agreement values. Don't worry at this stage about the details of how these are calculated. The cut-off point OxCal uses for acceptance is 60%, so anything less than this will bring up the 'Warning' message.

'Poor agreement' can refer to either a single date (for example, one date out of sequence in an otherwise stratigraphically coherent group of 10 dates) or to the entire model, or, as in this case, to both. Obviously when there are only a few dates, the model as a whole is more likely to 'fail' when one or two are irreconcilable with the proposed order of events.

Note that OxCal will still plot the results – have a look at 'Plot dates' under *View*.

What happens next depends on the archaeologist. Is another sample available? Is there some reason the tooth might actually be earlier (from an earlier occupation, or 'ancestral' when introduced. Or, possibly a marine reservoir correction was needed, i.e., if the tooth had a high $\delta^{13}C$ value).

Case study 2: Mound of the Hostages, Tara (Brindley et al. 2005; O'Sullivan 2005)

In this case we do have some basic stratigraphic information that can be entered into the model. We have simplified this here to look at only two phases, charcoal from pre-cairn contexts, and unburnt human bone from the chamber, presumably relating to the passage tomb's primary use.

**Note that this example is intended for illustrative purposes only, and is not a formal application of a Bayesian model to the site. Such a model would need to take all the dates into account, and to explore different readings of the sequence, etc.

This example combines the two kinds of models. The two phases are placed into a sequence, which identifies the pre-cairn phase as necessarily earlier than the burial phase. But within each of these phases the dates are unordered (and so are comparable to the Gransha example, with the difference that the 'end' of the pre-cairn phase and the 'start' of the burial phase influence one another).

Table 2. A selection of ^{14}C determinations from Mound of the Hostages, Tara (Brindley et al. 2005)

Context	Material	Lab code	date BP	±
Pre-cairn phase				
pre-cairn fire	charcoal	GrA-17675	4900	50
pre-cairn surface	charcoal	GrA-26064	4840	80
pre-cairn	charcoal	GrA-17672	4680	40
pre-cairn	charcoal	GrA-17525	4555	45
pre-cairn	charcoal	GrA-26065	4550	90
pre-cairn	charcoal	GrA-17674	4525	40
pre-cairn	charcoal	GrA-17676	4485	40
Primary use-phase				
tomb fill	human bone	GrA-17679	4415	40
tomb fill	human bone	GrA-17678	4390	45
tomb fill	human bone	GrA-17682	4370	40
tomb fill	human bone	GrA-18352	4370	50
tomb fill	human bone	GrA-17668	4355	40
tomb fill	human bone	GrA-18354	4230	50

This has already been entered into OxCal as 'Tara_model'. Go to 'Open' under the File menu. A new window will open. Double-click on the symbol to the left of the file name. This will open the data/model window.

You will note that this kind of model is set up in OxCal as follows:

```
Sequence
Boundary 'start'
     Phase
     the 14C determinations for the pre-cairn phase
Boundary 'end'
Boundary 'start'
     Phase
     the 14C determinations for the burial phase
Boundary 'end'
```

Now hit ‘Run’. Take a moment to look at the results. This is probably best done using ‘Plot dates’ under *View*. Where are the greatest effects of the model?

In the graph showing all the dates, have a look at the two probability distributions for ‘Boundary end pre-cairn’ and ‘Boundary start primary use’. There seems to be a hint of a gap between the two. This can be more formally explored.

Return to the model window using the arrow to the top left.
We will add a new command, ‘Interval’, which estimates the time between two events, in this case the two phases at Tara. In other words, is there a gap, and, if so, of what duration?

Place the cursor on the blue circle to the left of ‘Boundary start primary use’, and using the submenu under *Insert* select ‘Interval’. Name this ‘phases 1 & 2’ and click on >> to enter. ‘Interval(“phases 1 & 2”) should now appear between the two Boundaries.

Save and ‘Run’.

You will see in the results Table that the interval is modelled as lasting between 0 and 267 (or so) years at 2 sigma. So there could be quite a large gap, or none at all! Models do not always provide the most precise answers! But they can at least quantify and qualify our impressions more exactly than our subjective ‘eyeballing’.

Have a look at ‘Plot dates’ under *View*. You’ll note that Interval does not appear. You need to look at the single graphs to see the its distribution. This is worth doing. Go to ‘Plot single’ under View and go through the graphs using the > key to the top right until you find the Interval plot.