

NEW RADIOCARBON CALIBRATION SOFTWARE

Martin Jones

Centre for Archaeological Research, Auckland University, Private Bag 92019, Auckland, New Zealand.
Email: martin@analytic.co.nz.

Geoff Nicholls

Department of Mathematics, Auckland University, Private Bag 92019, Auckland, New Zealand.
Email: nicholls@math.auckland.ac.nz.

ABSTRACT. We have developed a software utility, “DateLab”, for conventional radiocarbon age (CRA) calibration and Bayesian analysis of CRAs. The current version has a smaller range of applicability than other similar utilities such as Bcal, Oxcal, and Mexcal. However, it enables analysis of some common types of CRA datesets. The main advantages of DateLab are its high quality sampling algorithm, the possibility of carrying out model comparison and hypothesis testing in a straightforward way, and the unbiased character of the summary statistics on which the analysis depends.

INTRODUCTION

A number of statistical packages for radiocarbon date calibration are freely available (see <http://www.radiocarbon.org/>). Those in most widespread use include Calib, described in Stuiver and Reimer (1993), which computes the likelihood¹ for each calibrated date, and OxCal, described in Ramsey (1995), which implements sample based Bayesian inference. We present another package, DateLab, implementing sample-based Bayesian inference. DateLab has, in some respects, a smaller range of applicability than existing packages. However, where it can be used DateLab offers high-quality, unbiased statistics and implements model comparison. This allows basic questions such as “Do these dates observe superposition?” to be quantified and statistically evaluated. As it is almost always the case that alternative models exist for Bayesian calibration this facility is a particularly useful feature in DateLab.

Calibration of independent CRAs is now a well-established process and a number of software packages (in particular Calib; Stuiver and Reimer 1993) exist to perform this calculation. However, in some cases it is useful to be able to calibrate CRAs that are related, in a manner that incorporates prior chronometric information into the calibration process. This is more complex than calibration by itself, and is typically performed via Bayesian inference (see Buck et al. 1996 and Christen 1994a for an overview of this type of application). Available packages include OxCal (Ramsey 1995), Bcal (Buck et al. 1999), and MexCal (Buck et al. 1999), all of which are designed to perform Bayesian calibration of CRAs. These tools are more than adequate for the job at hand, if used correctly. However, it is distressingly easy for users who are not experienced statisticians to inadvertently bias their calibration analyses. By restricting the class of models that can be applied and using more reliable algorithms to analyze those models, we hope to steer the user away from some uninteresting technical statistical problems that can otherwise invalidate their analysis. Here, we review the basic ideas of Bayesian inference, and then explain what those “uninteresting technical statistical problems” might be!

Bayesian Calibration

Bayesian methods for ¹⁴C date calibration are now in widespread use (e.g. Naylor and Smith 1988; Buck et al. 1991, 1992, 1994, 1996; Christen 1994a, 1994b; Christen and Buck 1998; Christen and Litton 1995; Christen et al. 1995; Litton and Leese 1991; Nicholls and Jones 1998, 2001; Zeidler et al. 1998). These statistical methods are attractive, as they allow associated chronometric information

¹The likelihood is equivalent to the calibrated distribution of a CRA.

to be taken into account, in an explicit way, in the chronometric analysis. For example, where artifacts are found in primary context, stratigraphic constraints can be applied to the calibrated dates. Information of this kind is made explicit in the analysis via a probability distribution, called the *prior*, which weights the calibrated dates toward values in line with our prior expectations. On the other hand, the data act through a distribution called the *likelihood*. A calibrated value that makes the observed CRA a likely outcome of the ^{14}C observation process has a high likelihood. The prior and likelihood distributions together determine a new probability distribution known as the *posterior*. Sets of calibrated dates agreeing with the data, and at the same time plausible in the light of prior information, yield a large posterior probability. In Bayesian calibration this posterior distribution is our analysis result.

Formally speaking, the un-normalized posterior distribution is given by:

$$\text{posterior} = \text{likelihood} \times \text{prior} \quad (1)$$

The likelihood is determined by the CRA data and the standard ^{14}C observation model (e.g. Buck et al. 1991). How do we summarize prior beliefs? An ideal approach is to try to specify a general “neutral”, or non-informative prior probability distribution, that is, a type of prior that assigns equal probability to any set of calibrated dates. However, a single neutral set of prior beliefs cannot exist in archaeology or any other domain. Prior beliefs that are non-informative with respect to any one hypothesis must be informative with respect to some other hypothesis. We must accept a prior that is in some respects informative and check that the information is representative of knowledge available in the problem at hand. So, for example, Nicholls and Jones (1998, 2001) write a simple prior model and check that it is non-informative for many (but not all!) situations of practical interest in ^{14}C dating. Finally, when conflicting models are proposed, we simply ask, “Which prior model does the data support?”. DateLab implements statistical tools that answer such questions.

The posterior distribution defined in Equation 1 is the joint probability of a number of event date parameters. We summarize this multidimensional distribution by considering the distribution of some meaningful statistic of direct interest (the span of calibrated ages might be one such salient statistic). We take the original joint posterior distribution of all parameters and integrate out the uninteresting parameters, in order to compute the “marginal distribution” for the remaining salient statistic. It is most often necessary to carry out the integration numerically, using a computer, as the integrals involved cannot usually be done by hand.

Several software packages designed to perform Bayesian calibration already exist. DateLab differs from these in terms of both functionality and in some aspects of the underlying mathematical framework. In particular, the default prior probability distribution for calibrated dates used in DateLab follows that given in Nicholls and Jones (1998, 2001). This prior avoids an undesirable bias toward more widely spread calibrated dates, which is imposed by the priors in common use². A further difference is the manner in which the marginal distributions are integrated out of the posterior. These are typically calculated via some form of sampling procedure, which is usually based around Markov chain Monte Carlo (MCMC) techniques. MCMC samples require careful checking to ensure they really do represent the posterior distribution. It cannot be assumed that the typical user of Bayesian calibration software is versed in this highly technical subject. DateLab offers an

²This prior is now also applied in OxCal. However, it should be noted that the prior outlined by Nicholls and Jones (2001) has been explicitly developed for the limited phase model described in this paper. Application of their prior to more complex phase models can give rise to biased posterior distributions.

alternative to the MCMC sampling algorithm called the rejection sampling algorithm. Samples generated by the rejection algorithm can be relied on to represent the posterior distribution.

In the following, we outline the functionality of DateLab and detail the mathematical framework underlying the analysis process.

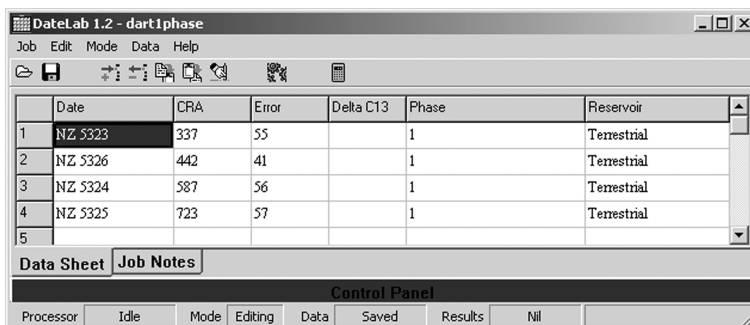


Figure 1 DateLab data entry screen

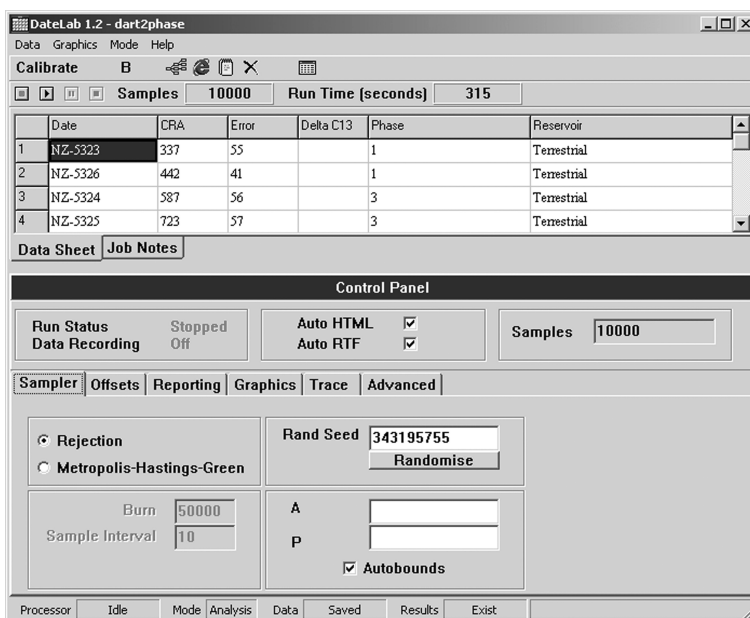


Figure 2 DateLab analysis screen

DateLab Functionality

The basic operation of DateLab is via a simple graphical user interface (Figures 1 and 2 above) and is detailed in the user manual (<http://www.datelab.org>). DateLab is user-friendly and produces high-quality graphical output (for examples see Figures 3 and 4) and detailed reports. Examples of standard reports can be seen at <http://www.datelab.org>.

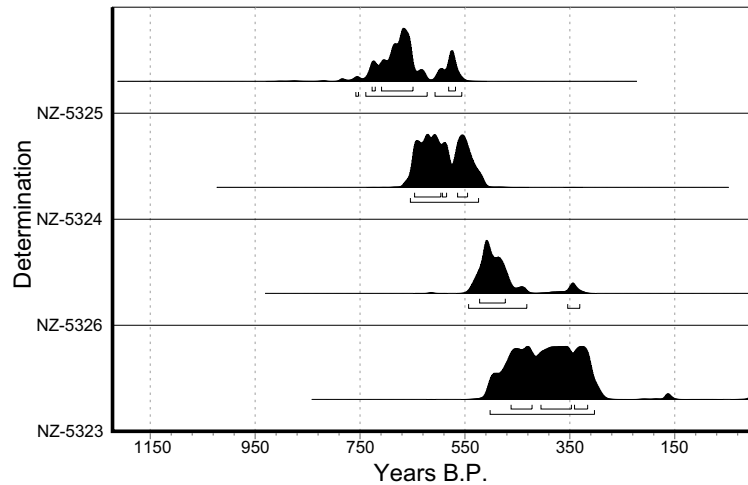


Figure 3 Combined histogram output from DateLab for the likelihoods (standard calibrated distributions) of the CRAs given in Table 2

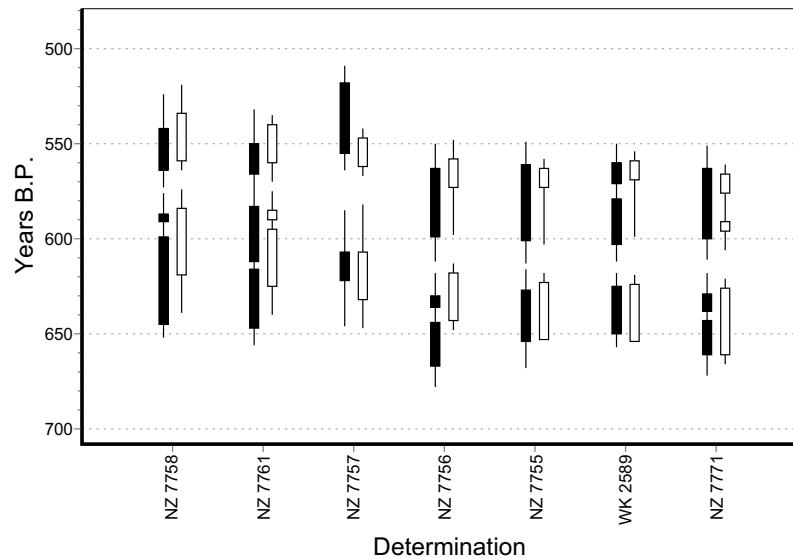


Figure 4 Posterior bar graph for the Shag Mouth data under S_1 : The solid bars represent the likelihoods (standard calibrated distributions) and the white bars represent the posterior distribution.

DateLab is currently only available for Win-32 systems. Future versions of the full DateLab system will be platform independent. The basic DateLab functionality is as follows:

1. Basic calibration of independent CRAs in the same fashion as found in packages such as Calib or Oxcal.
2. Calibration of dates using correlated reservoir offsets as discussed in Jones and Nicholls (2001) and Nicholls and Jones (1998, 2001).
3. Bayesian calibration of serial phases of dates following the methodology outlined by Nicholls and Jones (1998, 2001). Nicholls and Jones (1998, 2001) define a new class of prior models for

Bayesian calibration that are in general more appropriate than those used in existing software packages.

4. Sampling of the posterior via either rejection or MCMC analysis. As already discussed, rejection is an optimal sampling approach. However, rejection will be too slow for some problems and in this case a Metropolis-Hastings MCMC analysis following that given in Nicholls and Jones (1998, 2001) is implemented.
5. Model comparison via Bayes factors. In applying Bayesian analysis, it is often the case that a number of plausible models could be used. DateLab offers a sound method to objectively compare the relative likelihood of different models.
6. Results of analysis are summarized as intelligible documents in either HTML, RTF, or LaTeX formats with multiple graphics formats supported.

In the following, we detail the mathematical framework underlying the analyses described above.

Models

The Bayesian Calibration model used in DateLab is a direct representation of the relative chronometric information imposed by stratigraphic constraints in an archaeological excavation (see Nicholls and Jones 1998, 2001 for a detailed description and motivation for this model). All dates are treated as coming from one of a number of phases that occur as a single series. Within phases there are no prior constraints on the relative age of any of the dates. However, we know *a priori* the relative ordering of the phases and add a further constraint that there is no overlap of the phases. While this can be extended to analyze multi-phase models where there is complete independence between the phases, more general phase models, of the kind applied in Zeidler et al. (1998), cannot currently be analyzed using DateLab.

For further discussion of the DateLab model it is necessary to define the following notation. Dates are regarded as arising from a single series of M abutting phases. N_m ^{14}C age determinations are gathered from phase m , making $K \equiv \sum N_m$ dates in all. For $n \in \{1, 2, \dots, N_m\}$ let $y_{m,n}$ denote the value of the n 'th ^{14}C age measured in the m 'th phase, reported with associated standard error $\sigma_{m,n}$. For all quantities $X_{m,n}$ let X denote the corresponding vector in the natural ordering, so that $y \equiv (y_{1,1}, \dots, y_{M,NM})$, etc. Let $\theta_{m,n}$ be a calibrated date for specimen (m,n) , with units calendar years AD, and assumed to equal the context date associated with the (m,n) 'th specimen. For $m \in \{0, 1, \dots, M\}$ let ψ_m denote the boundary date at the lower boundary of phase m . We have a total $K+M+1$ unknown parameters: the $M+1$ layer boundary dates $\psi_0 \dots \psi_M$, and the K unknown object dates, $\theta_{1,1} \dots \theta_{M,NM}$. Let P and A , $P \leq A$ be given termini, setting lower and upper bounds on the dates. Possible parameter sets (ψ, θ) take some value in a parameter space Ω . This space is simply the set of all states (ψ, θ) satisfying the stratigraphic constraints:

$$\Omega \equiv \{(\psi, \theta); P \leq \psi_M \leq \theta_{M,1} \leq \psi_{M-1} \leq \dots \leq \psi_1 \leq \theta_{1,1} \leq \psi_0 \leq A\}$$

Following the standard Bayesian inferential framework, the posterior distribution of Ψ and θ conditional upon the observed dates y (with density $h(\Psi, \theta | y)$) is defined in terms of an unnormalized prior density $f(\Psi, \theta)$, and likelihood $L(y, \theta)$, as in Equation 1, as

$$h(\theta, \Psi | y) = L(y | \theta) \times f(\Psi | \theta) \tag{2}$$

We now outline the likelihood employed in DateLab and the family of prior models that DateLab is presently able to analyze.

Likelihood

The likelihood $L(y|\theta)$ used here follows the standard definition of the ^{14}C likelihood (e.g. Buck et al. 1991). The observation model for CRA $y_{m,n}$ is

$$y_{m,n} \sim \text{Normal}(\mu(\theta_{m,n}), \sigma(\theta_{m,n})^2)$$

where

$$\sigma(\theta_{m,n})^2 \equiv \hat{\sigma}(\theta_{m,n})^2 + \tilde{\sigma}_{m,n}^2$$

and $\mu(\theta_{m,n})$ and $\tilde{\sigma}(\theta_{m,n})$ are standard, empirically determined ^{14}C calibration functions (e.g. Stuiver et al. 1998). For DateLab we use the INTCAL98 calibration data available from <http://depts.washington.edu/qil/> in decadal tabulation. We spline these decadal values so that μ and σ^2 are functions piecewise linear by year. When a specimen (m,n) is of terrestrial origin, terrestrial calibration functions are used. Otherwise, the marine calibration functions are used. Material type dependence is implicit in our notation. Let $\ell(y_{m,n}|\theta_{m,n})$ denote the likelihood of parameter $\theta_{m,n}$,

$$\ell(y_{m,n}|\theta_{m,n}) = \frac{1}{\sigma(\theta_{m,n})\sqrt{2\pi}} \exp\left(-\frac{(y_{m,n} - \mu(\theta_{m,n}))^2}{2\sigma(\theta_{m,n})^2}\right)$$

Thus $\ell(y_{m,n}|\theta_{m,n})$ is a density distribution normalized over $y_{m,n}$ values, and unnormalized over $\theta_{m,n}$ values. Observations are assumed independent, so the joint likelihood, $L(y|\theta)$, is

$$L(y|\theta) = \prod_{m=1}^M \prod_{n=1}^{N_m} \ell(y_{m,n}|\theta_{m,n})$$

In the case that a correlated reservoir offset (Jones and Nicholls 2001) is used the likelihood needs to be modified. For details, see Nicholls and Jones (1998, 2001).

Prior

The prior density, $f(\psi, \theta)$, summarizes our state of knowledge before the ^{14}C determinations are available. It is natural to model the object dates θ conditional on the layer boundary dates ψ , so we break our prior up in two pieces,

$$f(\psi, \theta) = f_{\theta|\psi}(\theta|\psi) f_{\psi}(\psi)$$

In the absence of ^{14}C age determinations, the age parameters $\theta_{m,n}$ might take any value between $\psi_{m-1,n}$ and $\psi_{m,n}$ with equal probability, a state of knowledge represented by the choice

$$f_{\theta|\psi}(\theta|\psi) = \prod_{m=1}^M \frac{1}{(\psi_{m-1} - \psi_m)^{N_m}}$$

for (ψ, θ) restricted to Ω . What prior density $f_{\psi}(\psi)$ should we take for the set ψ of phase boundary event dates? A set of ψ values is “legal” if $P \leq \psi_M \leq \psi_{M-1} \dots \leq \psi_0 \leq A$. It seems natural to say “any legal set of dates ψ is *a priori* equally likely”, and we will call this choice, $f_{\psi}(\psi) = 1$, the uniform

prior density. Surprisingly enough, the uniform density for ψ weights the prior in favor of more widely spread sets of dates, and this can bias the whole analysis, though the effect is often slight. The date span, $\delta(\psi) \equiv \psi_0 - \psi_M$, which measures the number of years spanned by dated strata, exhibits the bias: it can be shown that, under the uniform prior density, a span of 2δ is favored over a span of δ by a factor of approximately 2^{M-1} (the approximation is good when $\delta \ll R$, where $R \equiv A - P$). Nicholls and Jones (1998, 2001) give several reasons for favoring the choice

$$f(\psi, \theta) = \frac{1}{(R - \psi_0 - \psi_M)} \frac{1}{(\psi_0 - \psi_M)^{M-1}} \tag{3}$$

the most important being, that this is simply more representative of the typical state of knowledge prior to the arrival of the data. It is non-informative for $(\psi_0 - \psi_M)$, the span of events. The prior in Equation 3 is akin to the prior belief that “any span value is equally likely, and then any legal set of dates ψ is equally likely given the span”. Equation 3 can be motivated in a number of ways. It may be derived by taking the simplest plausible physical model of the specimen deposition process, assuming constant deposition rates in layers, and allowing random thinning of deposited specimens. The suitability of the model choice expressed in Equation 3 over the constant prior may be tested, for any particular data set, using the model comparison tools described below.

Note that the prior density is not normalizable when no *terminus post quem* can be established, since $P = -\infty$ in that case. A prior with no finite normalization is said to be *improper*. Sample-based Bayesian inference with an improper prior is meaningful if the posterior density is normalizable, which is the case whenever $\sigma(\theta_{m,n}) < \infty$, under mild conditions on μ , the calibration function. Rejection sampling is infeasible when a very conservative P is asserted.

SAMPLING PROCEDURES

Sample-based inference is a numerical mode of analysis that allows us to form summarizing statements from the posterior density, i.e. integrate out marginal posterior distributions of interest. The inference is quite straightforward. The probability assertion X is the case, given the data and prior knowledge represented in the posterior, is estimated by sampling parameter sets (ψ, θ) from the posterior distribution, and then calculating the proportion of samples in which the event X occurs.

For example, a scientist may propose that the number of yr spanned by the modeled phases, $\delta(\psi) \equiv \psi_0 - \psi_M$, is less than 100 yr. In that case, S is the set $S = \{(\psi, \theta): \delta(\psi) \leq 100\}$ and we are interested in $\Pr\{(\Psi, \Theta) \in S | y\}$, which stands for the probability the scientist is correct, given the data, and any other substantial knowledge quantified in the prior distribution. Let

$$Z^h = \int_{\Omega} h(\psi, \theta | y) d^{(M+1)}\psi d^{(k)}\theta$$

denote the normalizing constant for the posterior density $h(\psi, \theta | y)$. Then, in terms of $h(\psi, \theta | y)$, the probability we have to estimate is

$$\Pr\{(\Psi, \Theta) \in S | y\} = \frac{1}{Z^h} \int_{S \cap \Omega} h(\psi, \theta | y) d^{(M+1)}\psi d^{(k)}\theta \tag{4}$$

Instead of calculating the integral given in Equation 4 in closed form, we estimate its value by Monte Carlo integration. The posterior probability for “the span is less than 100 yr” is estimated by the proportion of samples from the posterior distribution in which $\psi_0 - \psi_M \leq 100$. Similarly, histograms of sampled parameter sets may be used to summarize marginal posterior probability distributions.

The problem then is to generate samples from the posterior. In general, this is done using Markov chain Monte Carlo (MCMC) algorithms, which employ Gibbs sampling or some more general Metropolis-Hastings algorithm. There is a problem here. The output of MCMC algorithms is not in general guaranteed to have the desired (posterior) distribution. A sufficient condition for MCMC convergence to equilibrium, which may readily be checked, does not exist. In many applications, it can be established with reasonable confidence that the output is correctly distributed. Unfortunately, in the case of ^{14}C calibration, with a posterior density of the kind we have defined, it is particularly difficult to get Markov Chain Monte Carlo algorithms to converge reliably.

Thus, close attention needs to be paid to the output from MCMC analyses to ensure that the results are correct. However, for reasonably small data sets, the rejection algorithm given by Nicholls and Jones (1998, 2001) is suitable. This sampling algorithm has the advantage that, when it returns a sample, that sample is returned with a probability density that coincides with $h(\psi, \theta | y)$ up to machine precision. Moreover, samples are independent, so estimation of standard errors is straightforward. The main disadvantage of rejection sampling, compared with MCMC in general, is that it is very slow for certain types of calibration problems: roughly speaking, those in which the number of dates is large. DateLab implements both the rejection and Metropolis-Hastings MCMC sampling algorithms of Nicholls and Jones (1998, 2001). While the Metropolis-Hastings MCMC sampling routine is much faster than rejection, rejection sampling is the preferred approach. Unfortunately, under some sampling problems the rejection routine is too slow to be practical, and we are obliged to fall back on less reliable MCMC methods.

Model Comparison

When we carry out Bayesian calibration, it is usual that there will be several prior models that we may realistically apply to the data. In this case it is useful to be able to make a comparison between the competing models. For example, this allows us to address questions such as do these dates actually come from sequentially ordered strata (i.e. are the dates from a primary context) or has the material been mixed. In DateLab, model comparison is based around Bayes factors.

Methodology

Bayesian model comparison is based on a quantity called the Bayes factor. This quantity plays a similar role, in Bayesian inference, to the p-value of frequentist inference. Suppose \mathfrak{N}_1 and \mathfrak{N}_0 are 2 models, and the Bayes factor for Model \mathfrak{N}_1 over model \mathfrak{N}_0 is some number B . The analysis is telling us that model \mathfrak{N}_1 is B times more probable than model \mathfrak{N}_0 , in the light of the data. This statement may be interpreted quite literally. See the examples below. Table 1, taken from Raftery (1996), gives a standard interpretation of the Bayes factor (see Raftery 1996 for further discussion and references).

We now define the Bayes factor in more detail. For $i \in \{0, 1\}$, let $f(\psi, \theta | \mathfrak{N}_i)$ be the unnormalized prior density under model M_i , with normalizing constant Z_i^f . Let Ω_i denote the space of parameters in the i th model. The mean likelihood, $f_{Y|\mathfrak{N}_i}(y | \mathfrak{N}_i)$, for the data under model i is

$$f_{Y|\mathfrak{N}_i}(y | \mathfrak{N}_i) = \frac{1}{Z_i^f} \int_{\Omega} L(y | \theta) f(\psi, \theta | \mathfrak{N}_i) d^{(M+1)} \psi d^{(k)} \theta$$

DateLab uses the method outlined by Meng and Wong (1996) to calculate the mean likelihood. Details of the algorithm employed are given in Nicholls and Jones (1998, 2001).

The Bayes factor for comparison of models is

$$B(1v.0) \equiv \frac{P(y|\mathcal{N}_1)}{P(y|\mathcal{N}_0)}$$

On the basis of this, we can say Model 1 is $B(1v.0)$ more times likely than Model 0. The level of support this indicates for Model 1 is shown in Table 1.

Table 1 Interpretation of Bayes factor $B(1v.0)$, from Rafterty (1996)

$B(1v.0)$	Support for \mathcal{N}_1
<1	Supports \mathcal{N}_0
1 to 3	Barely worth mentioning
3 to 12	Positive
12 to 150	Strong
>150	Very strong

Examples

To demonstrate the use of Bayes factors we consider 2 archaeological date sets in Tables 2 and 3.

Table 2 Charcoal dates from Dart River Mouth, New Zealand (Anderson and Ritchie 1986; Simmons 1973). Column $y_{m,n}$ lists standard ^{14}C determinations, with no corrections. Column (m,n) lists m , the layer index and n , the specimen index within a layer. See text.

Date nr	(m,n)	$y_{m,n}$	$\sigma_{m,n}$
NZ 5323	(1,1)	337	55
NZ 5326	(1,2)	442	41
NZ 5324	(3,1)	587	56
NZ 5325	(3,2)	723	57

Table 3 Charcoal dates from Shag River Mouth, New Zealand (Anderson et al. 1996). The column $y_{m,n}$ lists standard ^{14}C determinations, with no corrections. Column (m,n) lists m , the layer index and n , the specimen index within a layer. See text.

Date nr	(m,n)	$y_{m,n}$	$\sigma_{m,n}$
NZ 7758	(1,1)	580	47
NZ 7761	(2,1)	600	50
NZ 7757	(3,1)	537	44
NZ 7756	(4,1)	670	47
NZ 7755	(5,1)	646	47
WK 2589	(5,2)	630	35
NZ 7771	(6,1)	660	46

In Table 2, we list a set of $K=4$ charcoal dates from the Dart River Mouth site (Anderson and Ritchie 1986; Simmons 1973), southern New Zealand. Anderson and Ritchie (1986) suggest that 2 discrete occupation phases are represented in the Dart River archaeological record, with 2 of the dates presented in Table 2 deriving from each of these proposed phases. A real question is, do the dates support this model? A graph of the calibrated date distributions (Figure 3) does not make it obvious that 2 phases of occupation would necessarily be favored over a single longer-term phase of activity. We will compare the model where the phase structure suggested by Anderson and Ritchie (1986) is imposed (\mathcal{N}_1) with the model in which the dates are regarded as arising from a single phase of occu-

pation (\aleph_0). The form of the prior density is the non-informative density $f(\psi, \theta)$ specified above. Under model $\aleph = \aleph_1$, $M=3$, and the parameter space is

$$\Omega_1 \equiv \{(\psi, \theta); P \leq \psi_3 \leq \theta_{3..} \leq \psi_2 \leq \psi_1 \leq \theta_{1..} \leq \psi_0 \leq A\}$$

while under model $\aleph = \aleph_0$, we have $M = 1$, and

$$\Omega_0 \equiv \{(\psi, \theta); P \leq \psi_1 \leq \theta_{M..}, \theta_{M-1}, \mathbf{K} \theta_{1..} \leq \psi_0 \leq A\}$$

Computing the mean likelihood under each model with $A = 100$ BP and $P = 1000$ BP we report

$$P(y|\aleph_0) = 1.1(1) \times 10^{-12} \quad P(y|\aleph_1) = 5.2(2) \times 10^{-12}$$

Throughout this report, the quantity quoted in brackets is a standard error for the last reported digit. Here, the mean likelihoods have been estimated sufficiently accurately to determine their relative magnitude. $B(1v.0) = 4.7(2)$, which means that the occupation phase sequence described by Anderson and Ritchie is around 5 times more likely in light of the available dates than the suggestion that there is simply a single phase of occupation. From Table 1, we can state that the Bayes factor analysis provides positive support for the two occupation phase model described by Anderson and Ritchie (1986) in contrast to a single occupation phase model.

In the 2nd example, we consider a set of 7 dates from a single series of strata (Table 3, Figure 4). Do the dates support the assertion that the strata are primary context for the dated artifacts? Many stratigraphic models might be constructed. Each model constrains the θ in different ways. We will compare the model with the full set of stratigraphic constraints \aleph_1 with the model in which all stratigraphic constraints have been removed \aleph_0 : a result significantly in favor of \aleph_0 would be sufficient grounds to reject \aleph_1 , and thereby reject the strata as primary context. The form of the prior density is the non-informative density $f(\psi, \theta)$ specified above. Under model $\aleph = \aleph_1$, $M=6$, the N_m $m=1, 2, \dots, 6$ are given by the data in Table 3, and the parameter space is

$$\Omega_1 \equiv \{(\psi, \theta); P \leq \psi_M \leq \theta_{M..} \leq \psi_{M-1} \leq \theta_{M-1} \leq \mathbf{K} \psi_1 \leq \theta_{1..} \leq \psi_0 \leq A\}$$

while under model $\aleph = \aleph_0$, we have $M=1$, $N_1=K=7$ and

$$\Omega_0 \equiv \{(\psi, \theta); P \leq \psi_1 \leq \theta_{M..}, \theta_{M-1}, \mathbf{K} \theta_{1..} \leq \psi_0 \leq A\}$$

Computing the mean likelihood under each model with $A = 100$ BP and $P = 1000$ BP we report

$$P(y|\aleph_0) = 1.3(1) \times 10^{-18} \quad P(y|\aleph_1) = 1.7(2) \times 10^{-18}$$

Since $B(1v.0) = 1.3(2)$, there is evidence in favor of \aleph_1 . Thus, the data provides some support for the proposed stratigraphic sequence, however it is inconclusive. Certainly, there is no reason to reject the notion that the observed stratigraphic sequence observes temporal superposition.

DATELAB OUTPUT

DateLab produces a standard output report of each analysis run. This is produced as an HTML document by default, however LaTeX and RTF versions can also be produced. The reports are self-

explanatory, and consist of a number of pages of summary analysis and output depending upon the type of analysis that has been performed. All reports contain:

- Details of the Analysis Models that have been applied.
- Summary reporting of analysis statistics (such as the mean Likelihood, HPDs etc.).
- Graphical output of analysis results in a range of formats.

Readers are encouraged to view the examples on the DateLab website (www.datelab.org).

Graphics

DateLab produces 2 basic forms of graphics, histograms (e.g. Figure 3) and bar graphs (e.g. Figure 4), which can be combined or presented individually. Histograms are a standard graphic for presentation of calibrated distributions and are well understood. However, histograms are not entirely suitable for presenting large numbers of calibrated distributions in a combined graph. While combined histograms such as those given in Figure 3 are useful, this type of graph rapidly becomes unmanageable for a large number of dates. Combined bar graphs, such as Figure 4, are more satisfactory for this type of data. In these plots the boxes represent the 68% highest posterior density (HPD) for the distribution (roughly analogous to the 1-sigma interval) and the lines represent the 95% HPD. Other types of graph are produced where a different format is more appropriate.

All images can be output to GIF, EMF, EPS, TIFF, or BMP versions. Full details are given in the manual.

CONCLUSION

DateLab is a simple software package that offers a limited range of Bayesian calibration models. However, DateLab can analyze many common CRA datesets and has been designed to produce robust, high-quality results and output. DateLab will be a useful tool for researchers wishing to perform simple calibration of CRA data or Bayesian calibration of serially ordered date sets, especially when the problem allows rejection sampling to be performed at a reasonable speed.

The DateLab software may be obtained via the internet from <http://www.datelab.org>.

ACKNOWLEDGMENTS

This work was funded through the Royal Society of New Zealand under Marsden Fund grant UOA-713 and the Foundation of Research, Science and Technology in New Zealand under grant UOA-807/UOAX0013.

REFERENCES

- Anderson A, Ritchie N. 1986. Pavements, Pounamu and Ti: the Dart Bridge site in western Otago, New Zealand. *New Zealand Journal of Archaeology* 8:115–41.
- Anderson A, Smith I, Higham T. 1996. Radiocarbon chronology. In: Anderson A, Smith I, Allingham B, editors. *Shag river mouth: the archaeology of an early Southern Maori village*. The Australian National University, Canberra: ANH Publications, RSPAS. p 61–9.
- Buck CE, Cavanagh WG, Litton CD. 1996. *The Bayesian approach to interpreting archaeological data*. Chichester: Wiley.
- Buck CE, Christen JA, James GN. 1999. BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeology* 7. <<http://intarch.ac.uk/journal/issue7/buck/>>
- Buck CE, Christen JA, Kenworthy JB, Litton CD. 1994. Estimating the duration of archaeological activity using ¹⁴C determinations. *Oxford Journal of Archaeology* 13:229–40.
- Buck CE, Kenworthy JB, Litton CD, Smith AFM. 1991. Combining archaeological and radiocarbon information: a Bayesian approach to calibration. *Antiquity* 65: 808–21.

- Buck CE, Litton CD, Smith AFM. 1992. Calibration of radiocarbon results pertaining to related archaeological events. *Journal of Archaeological Science* 19:497–512.
- Christen JA. 1994a. *Bayesian interpretation of ¹⁴C results*. PhD thesis. University of Nottingham, Nottingham, England.
- Christen JA. 1994b. Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics* 43:489–503.
- Christen JA, Buck CE. 1998. Sample selection in radiocarbon dating. *Applied Statistics* 47:543–57.
- Christen JA, Clymo RS, Litton CD. 1995. A Bayesian approach to the use of ¹⁴C dates in the estimation of the age of peat. *Radiocarbon* 37(2):431–42.
- Christen JA, Litton CD. 1995. A Bayesian approach to wiggle-matching. *Journal of Archaeological Science* 22:719–25.
- Jones MD, Nicholls GK. 2001. Reservoir offset models for radiocarbon calibration. *Radiocarbon* 43(1):119–24.
- Litton CD, Leese MN. 1991. Some statistical problems arising in radiocarbon calibration. In: Lockyear K, Rahtz SPQ, editors. *Computer applications and quantitative methods in archaeology 1990*. Oxford: Tempus Reparatum. p 101–9.
- Meng X-L, Wong WH. 1996. Simulating ratios of normalising constants via a simple identity: a theoretical exploration. *Statistica Sinica* 6:831–60.
- Naylor JC, Smith AFM. 1988. An archaeological inference problem. *Journal of the American Statistical Association* 83:588–95.
- Nicholls GK, Jones MD. 1998. Radiocarbon dating with temporal order constraints. Technical report. Mathematics Department, Auckland University, New Zealand. No. 407 <<http://www.math.auckland.ac.nz/~nicholls>>
- Nicholls GK, Jones MD. 2001. Radiocarbon dating with temporal order constraints. *Journal of the Royal Statistical Society*, series C, 50(4):503–21.
- Raftery AE. 1996. *Markov chain Monte Carlo in practice*. Chapman and Hall.
- Ramsey CB. 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. *Radiocarbon* 37(2):425–30.
- Simmons D. 1973. Radiocarbon dates from the Dart Valley region. *New Zealand Archaeological Association Newsletter* 16:175.
- Stuiver M, Reimer PJ. 1993. Extended ¹⁴C database and revised CALIB radiocarbon calibration program. *Radiocarbon* 35(1):215–30.
- Stuiver M, Reimer PJ, Bard E, Beck JW, Burr GS, Hughen KA, Kromer B, McCormac FG, van der Plicht J, Spurk M. 1998. INTCAL98 radiocarbon age calibration, 24,000–0 cal BP. *Radiocarbon* 40(3):1041–83.
- Zeidler JA, Buck CE, Litton CD. 1998. The integration of archaeological phase information and radiocarbon results from the Jama River Valley, Ecuador: a Bayesian approach. *Latin American Antiquity* 9:135–59.