

A NEW ROBUST STATISTICAL MODEL FOR RADIOCARBON DATA

J Andrés Christen • Sergio Pérez E

Centro de Investigación en Matemáticas, A. C. (CIMAT), A.P. 402, Guanajuato, Guanajuato 36000, Mexico.

Corresponding author. Email: sergiop@cimat.mx.

ABSTRACT. The general method currently used to analyze radiocarbon data (y) is conditional on the standard deviation (σ), reported by ^{14}C laboratories, which reflects the uncertainty in the dating process. This uncertainty is measured through a series of empirical as well as theoretical considerations about the dating process, chemical preprocessing, etc. Nevertheless, σ is assumed as known in the statistical model for ^{14}C data used since the dawn of the discipline. This paper proposes a method for the analysis of ^{14}C data where the associated variance is taken as the product of an unknown constant α with the sum of the variance reported by the laboratory σ^2 and the variance of the calibration curve $\sigma^2(\theta)$ (that is, an unknown error multiplier). Using this approach, assuming that the ^{14}C determination y arises from a Normal population and that, *a priori*, α has an inverse gamma distribution $\text{InvGa}(a, b)$, the resulting dating model is a t distribution with $2a$ degrees of freedom. The introduction of parameters a and b allows a robust analysis in the presence of atypical data and at the same time incorporates the uncertainty associated with the intra- and interlaboratory error assessment processes. Comparisons with the common Normal model show that the proposed t model produces smoother posterior distributions and seem to be far more robust to atypical data, presenting a simpler alternative to the standard ^{14}C outlier analysis. Moreover, this new model might be a step forward in understanding and explaining the otherwise elusive scatter in ^{14}C data seen in interlaboratory studies.

INTRODUCTION

Traditional Model

The established statistical model for a radiocarbon determination y_j is given by

$$y_j \sim N(\mu(\theta), \sigma_j^2), j = 1, 2, \dots, m \quad (1)$$

where $\mu(\cdot)$ is the calibration curve, θ is the associated calendar year (the year in which the organic material in the dated sample ceased metabolizing), and σ_j is the reported standard deviation for the y_j ^{14}C determination (see e.g. Clark 1975; Ward and Wilson 1978; Naylor and Smith 1988; Christen 1994; Blaauw et al. 2007). Indeed, the calibration is not known exactly, and for a given θ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (Buck and Blackwell 2004; Buck et al. 2006). Assuming that the error of $\mu(\theta)$ follows a Normal distribution, model (1) becomes

$$y_j \sim N(\mu(\theta), \sigma_j^2 + \sigma^2(\theta)) \quad (2)$$

where σ_j is assumed as known. This is the basic statistical model currently used for the statistical analysis of ^{14}C data (it is used in software like BCal, OxCal, etc. and used by the IntCal04 study group, Reimer et al. 2004). The likelihood function for given a random sample $\mathbf{y} = (y_1, \dots, y_m)$ of m ^{14}C determinations is

$$L_N(\theta|\mathbf{y}) \propto \prod_{j=1}^m \frac{1}{\omega_j(\theta)} \exp\left\{-\frac{1}{2\omega_j^2(\theta)}(y_j - \mu(\theta))^2\right\} \quad (3)$$

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

As mentioned above, this traditional model assumes that σ_j is known exactly. It is not clear, however, how σ_j is calculated at each laboratory, with each laboratory following different practices, including error multipliers and other considerations, besides “counting statistics,” which themselves have quite different interpretations depending on the counting method used (AMS versus beta-

counting dating). Also, the presence of outliers is a constant factor in the analysis of ^{14}C data, which may influence notably the inference results given the small sample sizes commonly in practice (Blaauw et al. 2005). Christen's (1994) approach to *detect* outliers basically consists in finding, for each ^{14}C determination, the posterior probability of it being an outlier (based on a shifted model for each determination). Observations with high posterior probabilities may be dropped from the data set and the analysis proceeds as usual. In general, this approach works in most situations, is in current use, and is implemented in popular software for the analysis of ^{14}C data (Bpeat, Blaauw and Christen 2005; BCal, Buck et al. 1999; and a beta-test version of OxCal, Bronk Ramsey, personal communication). Even for the simplest of cases, the Christen (1994) approach requires the use of the Markov chain Monte Carlo (MCMC) numerical method in order to simulate from the joint posterior distribution of all parameters and adds 2 parameters per each determination in the sample.

On the other hand, after years of quality control and interlaboratory studies, still there is "unexplained" scatter in ^{14}C data, while considering the above-mentioned Normal model. Regarding the most recently published international radiocarbon intercomparison (FIRI), it was concluded that "The chances of outliers occurring, assuming they occur by chance, is roughly 1 in 20" (5%) and that from "duplicate results, we see that, on average, the difference in duplicates is zero ... but the magnitude of difference is frequently large relative to the quoted errors (and larger than expected given the [Normal distribution] interpretation of the quoted error)" (Scott 2003: section 10). Given these conclusions, an outlier identification process is in order to explain spread-out results, but an as yet unexplored alternative would be to change the model to a heavier-tailed distribution than the Normal. By not considering the standard errors (σ_j) as known exactly, but simply as an additional piece of information possibly needing a correcting error multiplier, the net result is to consider a t distribution as an alternative for the Normal model, as we show in the next section. This new model will attempt to *accommodate* the scatter, hopefully leading to a better understanding of ^{14}C data.

THE NEW MODEL

Normal Model with a Variance Multiplier

Since measured activity precision depends on ^{14}C age, we know that σ_j varies jointly with y_j and, therefore, both quantities may be seen as data arising from a bivariate distribution. The uncertainty about the variance of y_j in Equation 1 may be introduced by considering the product $\alpha\sigma_j^2$, where α is a strictly positive unknown constant. Here, the meaning of α is that of an unknown "variance multiplier" to the laboratory-reported variance σ_j^2 . In fact, this formulation is equivalent to a Normal model with unknown variance, but also considering the reported error σ_j as part of the data. In this fashion, we may formally introduce the uncertainty related to the variance of y_j . Then, our modified statistical model for the ^{14}C determination y_j is

$$y_j \sim N(\mu(\theta), \alpha\sigma_j^2) \quad (4)$$

Here, we also consider a model that uses the variance $\sigma^2(\theta)$ in the calibration curve, accordingly

$$y_j \sim N(\mu(\theta), \alpha(\sigma^2(\theta) + \sigma_j^2)) \quad (5)$$

Then, given a random sample of ^{14}C determinations $\mathbf{y} = (y_1, \dots, y_m)$ the likelihood function for (θ, α) is

$$L(\theta, \alpha | \mathbf{y}) \propto \alpha^{-m/2} \prod_{j=1}^m \frac{1}{\omega_j(\theta)} \exp \left\{ -\frac{1}{2\alpha\omega_j^2(\theta)} (y_j - \mu(\theta))^2 \right\} \quad (6)$$

The high-precision calibration curves are based on arguably high-quality screened data and it could well be that in some cases a more optimistic multiplier α_1 should be used for $\sigma^2(\theta)$. That is,

$$y_j \sim N(\mu(\theta), \alpha_1 \sigma^2(\theta) + \alpha_2 \sigma_j^2) \quad (7)$$

Unfortunately, with this more realistic modeling more complex numerical methods are needed in order to make inferences about other model parameters. Our formulation, which assumes that the multiplier α also affects $\sigma^2(\theta)$, will ensure mathematical tractability and an analytically feasible representation of the marginal posterior distribution of θ . Moreover, since typically $\sigma(\theta)$ is small compared with σ , the proposed model is a well-behaved approximation to the 2-multiplier model with 1 multiplier for the ^{14}C determination and another one for the calibration curve variance. Figure 1 shows the probability plot (Gibbons and Chakraborti 2003:144) of the double-multiplier Normal model (Equation 7) and the single-multiplier Normal model (Equation 5) against the traditional Normal model (Equation 2) represented by the dashed line. We see that there is very little difference between the family of models indexed by the multipliers $\alpha_1 \in [0, 4]$, $\alpha_2 = 2$ (gray colored region), and our proposed model (Equation 5) given by the continuous black line ($\alpha_1 = \alpha_2 = \alpha = 2$). That is, our proposed model with 1 multiplier is a parsimonious and mathematically tractable alternative to the more complex double-multiplier model. We therefore proceed with the former model in what follows.

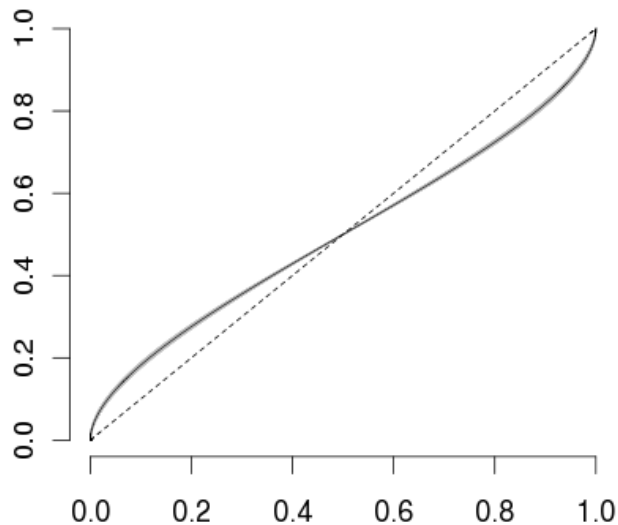


Figure 1 Probability plot for the double-multiplier Normal model (vertical axis, gray region), single-multiplier Normal model (vertical axis, black line) versus the traditional Normal model (horizontal axis): $\theta = 500$, $\sigma = 50$.

The New (Integrated) Likelihood

Usually, Bayesian statisticians use gamma distributions to model uncertainty about scale parameters, which has as support the positive real line and is flexible enough to model most prior uncertainties about scale parameters. In particular, it is common practice to assign a gamma distribution to the precision (reciprocal of the variance) of a Normal model, implying an inverse gamma distribution for the Normal variance. Accordingly, we assume that the prior distribution for α is an inverse gamma with shape parameter a and scale parameter b :

$$\pi(\alpha) = \text{InvGa}(\alpha|a, b) = \frac{b^a}{\Gamma(a)} \alpha^{-(a+1)} e^{-\frac{b}{\alpha}} \quad (8)$$

Then, given θ , the prior distribution of $\alpha\omega_j^2$ is the inverse gamma

$$\alpha\omega_j^2|\theta \sim \text{InvGa}(a, b(\sigma_j^2 + \sigma^2(\theta))) \quad (9)$$

such that $E(\alpha\omega_j^2|\theta) = \frac{b}{a-1}(\sigma_j^2 + \sigma^2(\theta))$ is the prior expected variance of y_j . However, the prior density for α does not depend on θ and may be established in a general way. It is clear that for particular applications, $\pi(\alpha)$ should be set according to *a priori* considerations about possible error multipliers for the sample at hand. However, it is possible to consider general guidelines applicable to modern ^{14}C data sets in most cases.

Figure 2 shows the prior density of α for $a = 3$ and $b = 4$. A decision needs to be made on the values of a and b . The former round numbers represent a distribution with $E(\alpha) = b/(a - 1) = 2$, mode $Mo(\alpha) = b/(a + 1) = 1$, median $Me(\alpha) \approx 1.5$, $P(\alpha \leq 1) \approx 0.24$, and $P(\alpha \geq 4) \approx 0.08$. For a series of mathematical tractability issues to be discussed below, we use α as a *variance* multiplier ($\alpha\sigma_j^2$). In more common terms, the multiplier has been applied to the standard error term. Considering the distribution of $\sqrt{\alpha}$, the *a priori* $\text{InvGa}(\alpha|a = 3, b = 4)$ represents a low probability (8%) that the standard error multiplier is above 2, some probability that it is overestimated (24%), and a strong prior probability ($\approx 70\%$) that the reported standard error was underestimated with a factor above 1 and below 2. Moreover, this distribution represents the *a priori* belief that the most likely scenario is that the error term was correctly reported ($\pi(\alpha)$ has its mode at 1), but we expect a 40% underestimation (since $E(\alpha) = 2$ and $\sqrt{2} \approx 1.4$). We believe these figures represent more or less the current state of knowledge for modern (quality-controlled) ^{14}C determinations, perhaps even tending to the optimistic side. We must stress the fact that the choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only and not by any means taken as the final word on establishing a prior distribution on α . Indeed, our approach does allow for other choices of these parameters that could be explored for particular applications.

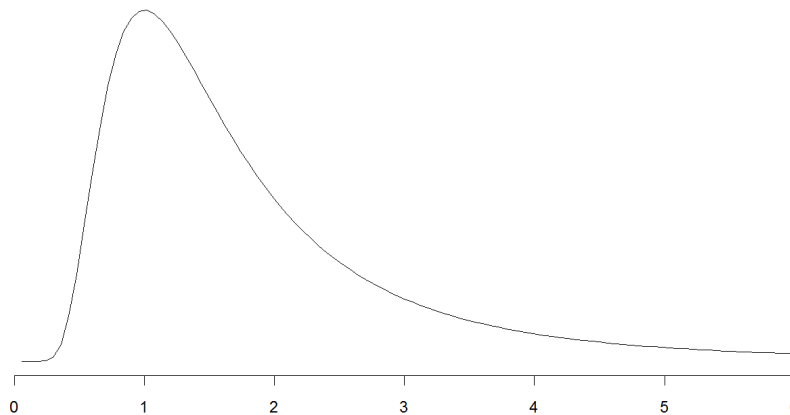


Figure 2 Prior density for the variance multiplier with expected value $E(\alpha) = b/(a - 1) = 2$, mode $Mo(\alpha) = b/(a + 1) = 1$, and median $Me(\alpha) \approx 1.5$, $a = 3$, $b = 4$.

Certainly, the parameter of interest is the true calendar age θ , while α is not a model parameter of interest for the researcher; that is, it is a nuisance parameter. In a Bayesian setting, nuisance param-

eters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function, using their prior distribution. Here, we derive the posterior distribution for θ using the integrated likelihood

$$L_{a,b}(\theta|\mathbf{y}) \propto \int_0^\infty \prod_{j=1}^m p(y_j|\theta, \alpha) \pi_{a,b}(\alpha) d\alpha \quad (10)$$

Note that we are assuming prior independence of the parameters (θ, α) . Therefore, under the prior distribution (Equation 8) the integrated likelihood is

$$L_{a,b}(\theta|\mathbf{y}) = \int_0^\infty \prod_{j=1}^m N(y_j|\mu(\theta), \alpha\omega_j(\theta)) \text{InvGa}(\alpha|a, b) d\alpha \quad (11)$$

$$\propto \left[b + \sum_{j=1}^m \frac{(y_j - \mu(\theta))^2}{2\omega_j^2(\theta)} \right]^{-\frac{2a+m}{2}} \quad (12)$$

$$\propto \left[1 + a^{-1} \sum_{j=1}^m \frac{(y_j - \mu(\theta))^2}{\omega_j^2(\theta)b/a} \right]^{-\frac{2a+m}{2}} \\ \propto t(\mathbf{y}|\mu(\theta)\mathbf{1}_m, \Sigma(\theta)b/a, 2a)$$

where $\mathbf{y} = (y_1, \dots, y_m)$, $\mathbf{1}_m = (1, \dots, 1)^t$, and $\Sigma(\theta) = \text{diag}(\omega_1^2(\theta), \dots, \omega_m^2(\theta))$ is the covariance matrix under the Normal model. Then, the integrated likelihood for θ given \mathbf{y} is proportional to a t distribution with location parameter $\mu(\theta)\mathbf{1}_m$, covariance matrix $\Sigma(\theta)b/(a-1)$ and $2a$ degrees of freedom. Note that with the sequence of parameters values $a = i+1$, $b = i$, $i = 1, 2, \dots$, we obtain a sequence of prior distributions that converges to the degenerate Dirac distribution at $\alpha = 1$, leading to a sequence of integrated heavy-tail models, with covariance matrix $\Sigma(\theta)$, which converges to the traditional Normal model (a standard result stating that the t distribution converges to a Normal distribution as the degrees of freedom of the t distribution increase). As expected, our new model has as a limiting case the standard Normal model when *a priori* $Pr(\alpha = 1) = 1$; that is, when σ_j is known exactly.

An alternative to the inverse gamma prior is a reference prior (Bernardo 2005), which may be used to obtain a reference posterior dependent only on the data and the assumed model for y . Reference priors use only model-based prior information and have been successfully used to eliminate nuisance parameters from the likelihood function (Berger et al. 1999). For completeness, and to compare with informative priors like the one given in Figure 2, we present a proposal for a reference prior for α . Assuming $\mu = \mu(\theta)$ and α as the interest and nuisance parameters, respectively, it can be verified that the conditional reference prior for α is

$$\pi(\alpha|\mu) = \pi(\alpha) \propto \alpha^{-1} \quad (13)$$

However, our parameter of interest is θ and the reference prior should be derived for the ordered parameters of interest (θ, α) . Since μ is not a monotone function, it will be very difficult to define the marginal reference prior for θ , as the derivation of the reference prior requires the asymptotic posterior distribution of θ , which due to the non-monotonicity of μ could be multimodal. In such a

case, the application of the standard algorithm to obtain reference priors becomes troublesome. Following Bernardo (1979), we know that the reference prior for (θ, α) is of the form $\pi(\theta)\pi(\alpha|\theta)$, where $\pi(\alpha|\theta) = \pi(\alpha) \propto \alpha^{-1}$, since α is a scale parameter in the Normal model (Equation 2). Thus, in order to integrate α out of the likelihood function (Equation 3) it is safe to use the conditional reference prior (Equation 13).

Integrating the likelihood with respect to the prior (Equation 13), we obtain the integrated likelihood:

$$L_R(\theta|\mathbf{y}) = \int_0^\infty \prod_{j=1}^m N(y_j|\mu(\theta), \alpha\omega_j(\theta))\alpha^{-1}d\alpha \quad (14)$$

$$\propto \left[\sum_{j=1}^m \frac{(y_j - \mu(\theta))^2}{\omega_j^2(\theta)} \right]^{-\frac{m}{2}}$$

Note that L_R is proportional to the integrated likelihood given in (Equation 12) when $a = 0$ and $b = 0$ and is a proper probability model if and only if $m \geq 2$. Also, it may be verified that for large sample size m : $L_N \approx L_{a,b} \approx L_R$.

Posterior Distribution

Now we derive the posterior distribution of θ by formal use of the Bayes' rule, i.e.:

$$\pi(\theta|\mathbf{y}) \propto L(\theta|\mathbf{y})\pi(\theta) \quad (15)$$

Here, the likelihood function L could be L_N , $L_{a,b}$, or L_R and for the prior of θ we will use a uniform distribution on the interval (B_2, B_1) , $B_1 < B_2$. Of course, if the researcher has further prior information about θ they may properly include it through any other prior distribution. We are principally interested in the posterior obtained with the integrated likelihood $L_{a,b}$, then the posterior distribution for θ is

$$\pi_{a,b}(\theta|\mathbf{y}) \propto L_{a,b}(\theta|\mathbf{y})\pi(\theta) \propto L_{a,b}(\theta|\mathbf{y}), B_1 < \theta < B_2 \quad (16)$$

Similarly, the posterior distribution of θ using the reference integrated likelihood L_R is

$$\pi_R(\theta|\mathbf{y}) \propto L_R(\theta|\mathbf{y}), B_1 < \theta < B_2 \quad (17)$$

With the conventional Normal model, the posterior is proportional to the likelihood L_N , i.e.:

$$\pi_N(\theta|\mathbf{y}) \propto L_N(\theta|\mathbf{y}), B_1 < \theta < B_2 \quad (18)$$

In the next section, we give numerical examples to explore these alternative models.

EXAMPLES

Simulated Data

We first analyze a set of simulated of $m = 5$ ^{14}C observations (Table 1). For this data, the true value of the calendar year is 650 cal BP and we assume a constant standard deviation $\sigma(\theta) = 12$. Each σ_j^2 was simulated from InvGa(2, 2500) such that $\sqrt{E(\sigma_j^2)} = 50$. Figure 3 exhibits the simulated ^{14}C determinations plotted over the calibration curve. Note that there is an atypical observation (S5).

Table 1 Simulated ^{14}C data. The parameter values are $\theta = 650$, $\sigma(\theta) = 12$, and $m = 5$.

ID	Determination
S1	649 ± 25
S2	598 ± 25
S3	748 ± 69
S4	606 ± 37
S5	368 ± 37

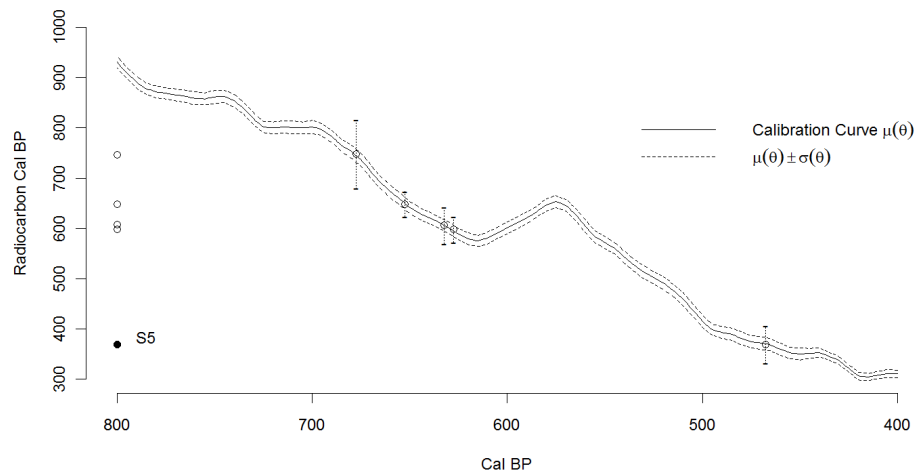


Figure 3 Simulated data displayed on the calibration curve

Figure 4 shows the posterior distributions π_N , π_R , and $\pi_{3,4}$ for the simulated data in Table 1 as explained above in the section “Posterior Distribution,” where $B_1 = 350$ and $B_2 = 950$. Also in Figure 4, we present the posterior density using the traditional Normal model but removing the outlying observation S5; this is density π_N^* . Note that π_N looks rougher, reproducing the wiggles in the calibration curve, while $\pi_{3,4}$ and π_R are smoother and concentrated over the most likely region given the data.

Due to the non-monotonicity of the calibration curve, and the position of the observed values, there are 3 highest-density regions and $\pi_{3,4}$ may better reflect this fact. We see that the effect of the outliers is much more evident in π_N .

The effect of the outlying observation S5 causes the Normal likelihood to shrink and shift to the right, leaving the true value for θ out of the 95% HPD (highest posterior density) region for π_N . If we drop S5 from the data, we observe that the resulting posterior arising from the Normal likelihood (π_N^* in Figure 4) is not more informative than π_N ; the N shape of the posterior density implies 2 unconnected credible intervals for θ , one of them now containing the true value of θ . On the other hand, $\pi_{3,4}$ and π_R are based on the whole data, and the heavy tails of the underlying model ensure that we are properly including the information provided by possible extreme values. In this case, the 95% HPDs of both distributions include the true value for θ . This example depicts the behavior of our new model. While the Normal model is quite sensitive to outliers and its light tails and short spread lead to the very common wiggle distribution, our new approach is more cautious and results in wider, smoother distributions. This makes our approach more robust and less prone to “mistaken”

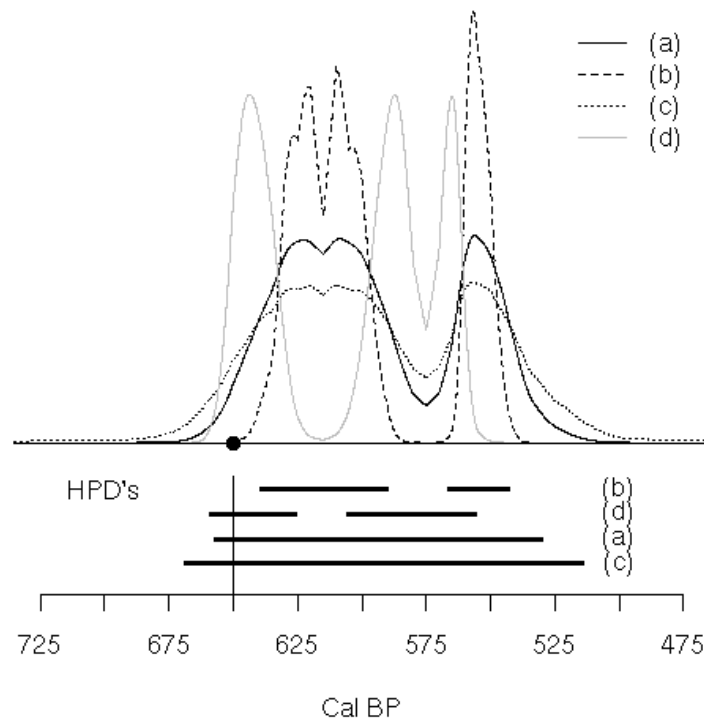


Figure 4 Posterior densities for θ over the interval ($B_1 = 350$, $B_2 = 950$), and their corresponding 95% HPD credible sets, for (a) $\pi_{3,4}$, (b) π_N , (c) π_R , and (d) π_N^* (Normal model not including observation S5). The true value of $\theta = 650$ BP is also plotted.

observations. Note that credible intervals under the new model will be in general wider than under the Normal. This is not a bad behavior of our approach; on the contrary, it means a more reliable measure of the posterior uncertainty in presence of outliers. Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion.

The Shroud of Turin

These data were taken from Damon et al. (1989) and were analyzed by Christen (1994) to show the performance of his proposed Bayesian method for detecting outliers in ^{14}C data. In this analysis, interest focuses on the associated calendar year θ in which the organic matter in the Shroud of Turin died. The data \mathbf{y} is a set of 12 ^{14}C determinations made in $n = 3$ AMS laboratories: Arizona, Oxford, and Zurich (Table 2). Following Christen (1994), as far as the Normal model is concerned, observations A1.1b and O1.1u have a high posterior probability of them being outliers and should be removed from the data in order to obtain a more reliable analysis, to obtain the reduced data \mathbf{y}^* .

We assume that each determination arises from a population with mean $\mu(\theta)$ as in (5). However, differences between laboratories in the determination process suggest to use a different multiplier α for each laboratory. Then, the likelihood function is of the form

$$L(\theta, \alpha | \mathbf{y}) = \prod_{i=1}^n \prod_{j=1}^{m_i} (2\pi\alpha_i\omega_{ij}^2(\theta))^{-m_i/2} \exp\left\{-\frac{1}{2\alpha_i\omega_{ij}^2(\theta)}(y_{ij} - \mu(\theta))^2\right\} \quad (19)$$

where $y = (y_1, y_2, \dots, y_n)$, $y_i = (y_{i1}, y_{i2}, \dots, y_{im_i})$, and $\alpha = (\alpha_1, \alpha_2, \dots, \alpha_n)$. In our example, $n = 3$, $m_1 = 4$, $m_2 = 3$, and $m_3 = 5$. Integrating the likelihood function as in (Equation 11) with respect to an inverse gamma prior density for each α_i , the integrated likelihood is

$$L_{a,b}(\theta|y) \propto \prod_{i=1}^3 t(\mathbf{y}_i | \mu(\theta) \mathbf{1}_{m_i}, \Sigma_i(\theta) b/a, 2a) \quad (20)$$

where $\Sigma_i(\theta) = \text{diag}(\omega_{i1}^2(\theta), \dots, \omega_{im_i}^2(\theta))$. Thus, $L_{a,b}$ is the product of 3 multivariate t densities. Similarly, the reference integrated likelihood is

$$L_R(\theta|y) \propto \prod_{i=1}^3 \left[\sum_{j=1}^{m_i} \frac{(y_{ij} - \mu(\theta))^2}{\omega_{ij}^2(\theta)} \right]^{-\frac{m_i}{2}} \quad (21)$$

Table 2 ^{14}C determinations for the Shroud of Turin; nomenclature is the same as in the original publication, see Damon et al. (1989).

Laboratory & sample code	Sample (rep.)	Determination
Arizona AA-3367	A1.1b	591 ± 30
	A1.2b	690 ± 35
	A1.3a	606 ± 41
	A1.4a	701 ± 33
Oxford 2575	O1.1u	795 ± 65
	O1.2b	730 ± 45
	O1.1b	745 ± 55
Zurich ETH-3883	Z1.1u	733 ± 61
	Z1.1w	722 ± 56
	Z1.1s	635 ± 57
	Z1.2w	639 ± 45
	Z1.2s	679 ± 51

Figure 5 shows the posterior densities $\pi_{3,4}$, π_R , and π_N under a uniform prior for θ . Here, the value of B_1 in the uniform prior for θ is 600, the earliest recorded time in historical records for the appearance of the Shroud of Turin. For this data, $\pi_{3,4}$ and π_R have similar shapes and are located to the left of the π_N . It appears that posterior $\pi_{3,4}$ is a compromise between π_R and π_N . Accordingly, with the 95% HPD probability interval for $\pi_{3,4}$ the calendar year in which the organic matter in the Shroud of Turin died lies between 669 and 648 cal BP (cal AD 1281–1302). The posterior π_R has the largest 95% HPD interval of 670–647 cal BP. In fact, given the sample size, in this case the Normal model is not so sensitive to the outlier information. By removing the outliers, the resulting posterior π_N^* does not change noticeably; however, we are losing the information provided by 2 probable outliers (Christen 1994 establishes that determinations A1.1b and O1.1u are outliers, both with posterior probabilities lower than 0.5).

COVERAGE PROBABILITY OF THE HPD SETS

In order to analyze the performance of our proposed model, we estimate with Monte Carlo simulation the “coverage probability” of 95% HPD sets (the proportion of cases in which the 95% HPD sets of repeated simulations of samples include the true value of θ) for models π_N , $\pi_{(3,4)}$, and π_R . The simulation was made as follows: a sample $\mathbf{y} = (y_1, \dots, y_n)$ of size $n = 5$ was simulated from the proba-

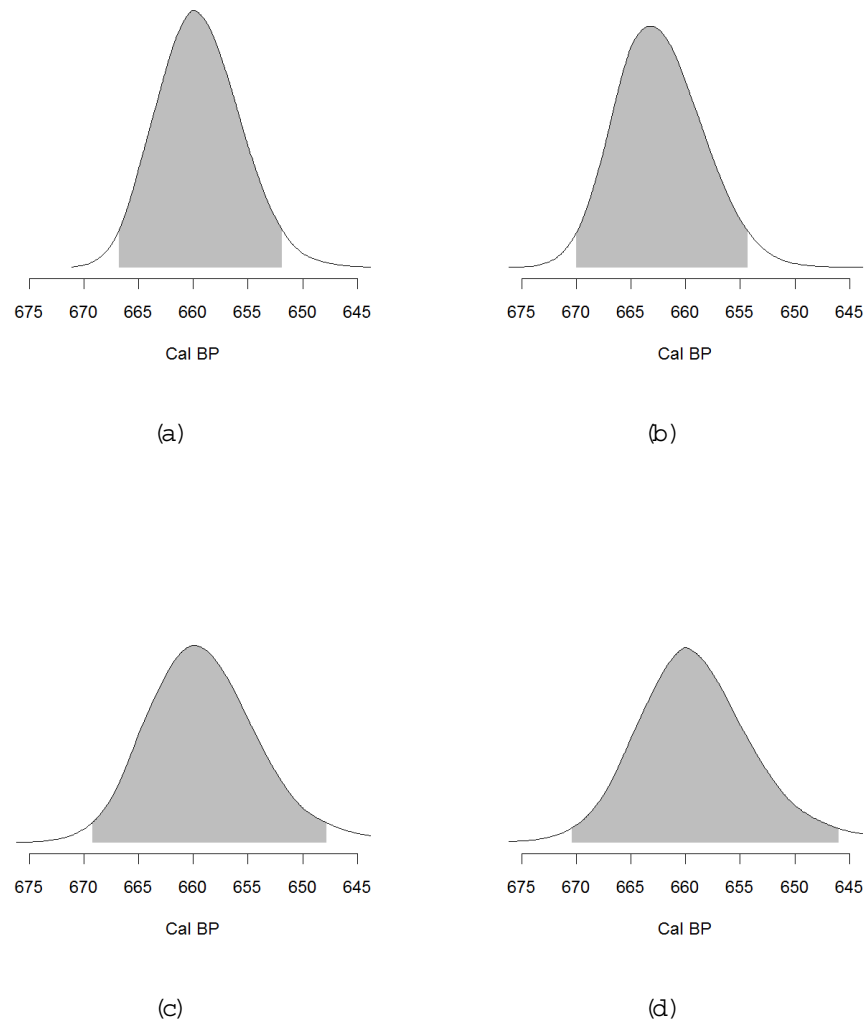


Figure 5 Posterior densities and 95% HPD regions (under shaded area) of θ for the Shroud of Turin data: (a) π_N , (b) π_N^* , (c) $\pi_{3,4}$, and (d) π_R .

bilistic model $N(y|\mu(\theta) + \phi\delta(-1)^j, \sigma^2 + \sigma^2(\theta))$, with $\theta = 650$ cal BP, $j \sim \text{Bernoulli}(1/2)$, $\phi \sim \text{Bernoulli}(p)$, $\delta = 250$, $\sigma^2 \sim \text{InvGa}(2, 2500)$ (expected reported error around 50), and $\sigma(\theta) = 12$. This is a simplified version of the common outlier model used by Christen (1994), but instead of δ having a Normal distribution we have $P(\delta = -250) = P(\delta = 250) = 1/2$; 250 is 5 times the expected standard error and thus we have a direct probability p for each determination to be an outlier. For each simulated \mathbf{y} , the posterior 95% HPD regions $C_{1-\alpha}^{\pi}(\mathbf{y})$ for π_N , $\pi_{3,4}$, and π_R were calculated using a Monte Carlo sample of size $s = 5000$ of θ , $1-\alpha = 0.95$ (see Chen and Shao 1999, for numerical details). If the true value of θ was contained in $C_{1-\alpha}^{\pi}(\mathbf{y})$, a success was recorded. This entire process was repeated $r = 5000$ times. For each posterior, the coverage probability was estimated with the success relative frequency. Also, since multimodal posteriors lead commonly to unconnected HPD regions, at each iteration the size of each HPD region (in cal years) was counted and the average used as an indicator of the precision. Note that for each \mathbf{y} there is a probability p for it to arise from the Normal

model with mean $\mu(\theta) \pm \delta$, where the location shift parameter δ means that y is an outlier. Thus, as the probability $p = \Pr(\phi = 1)$ decreases we expect a coverage probability close to the nominal $(1 - \alpha)$ under a Normal model with mean $\mu(\theta)$.

The simulation results are displayed in Tables 3 and 4. Note that for the considered cases and different choices of the inverse gamma prior $\pi(\theta)$ we obtain a coverage close to the nominal (0.95) and the average HPD length counts are not much greater than under the alternative Normal model. Note that with the posterior $\pi_{3,4}$ the average count is not so large, while the coverage probability is close to 0.95. The Normal model is the most sensitive to the presence of atypical data since the coverage probability decreases dramatically as the probability of outliers increases. It already goes down to 70% while the probability of outliers is only 10% (see Table 3). Our new t distribution approach remains robust in all cases, taking a more conservative approach resulting in distributions being slightly more spread out. With the reference integrated model, we obtain a coverage always greater than 0.95 and it appears to be robust to outliers; however the average length count is relatively large. Also, we included an analysis for the distribution $\pi_{a=15, b=16}$ in Tables 3 and 4. As explained in the section “The New (Integrated) Likelihood,” this distribution should approximate π_N . This is confirmed in the performing characteristics of both models seen in either table.

Table 3 Estimated coverage probability of the 95% HPD sets for different values of p .

Posterior distribution	p (outlier probability)			
	0.0	0.01	0.05	0.1
$\pi_{a=3, b=4}$	0.9806	0.9700	0.9564	0.9332
$\pi_{a=15, b=16}$	0.9650	0.9372	0.8714	0.7936
π_N	0.9556	0.9198	0.8194	0.7014
π_R	0.9558	0.9550	0.9582	0.9692

Table 4 Average rounded count of the 95% HPD sets for different values of p .

Posterior distribution	p (outlier probability)			
	0.0	0.01	0.05	0.1
$\pi_{a=3, b=4}$	72	74	82	92
$\pi_{a=15, b=16}$	65	65	66	68
π_N	61	61	58	56
π_R	73	80	102	130

MORE COMPLEX DATING PROBLEMS

In the previous sections, we have discussed examples where the parameter of interest was indeed the true calendar BP year θ . Moreover, in “The Shroud of Turin” section we analyzed a set of observations arising from 3 laboratories and we assumed that the parent population from which data arose is governed by the common parameters a and b . In fact, we may consider the above formulation as the hierarchical model $\mathbf{y} \leftarrow \theta \leftarrow (a, b)$, where the second-level parameters (hyperparameters) are introduced to model-specific features of the data; namely, the clustering induced by the intralaboratory determination variance and the presence of atypical data. In general, the dating model is of the form $\mathbf{y} \leftarrow \theta \leftarrow \psi$, where \mathbf{y} is a generic representation of data obtained under diverse sampling schemes, θ is a vector of several calendar BP yr, and ψ contains (a, b) and quantities related to the phenomena being dated (e.g. the early and late boundaries of a specific archaeological phase). Except for the hyperparameters a and b included in ψ , this hierarchical representation is the same as

that given by Christen (1994) and in common usage for the Bayesian interpretation of ^{14}C dates, including the corresponding implementation of BCal, OxCal, Bpeat, etc. That is, for any dating problem within this general statistical framework for ^{14}C analysis we obtain a more robust analysis of ^{14}C data by substituting the Normal likelihood (Equation 3) by the t model in (Equation 12). Note that this is a general form of the former since, as explained earlier, using $a = i + 1$, $b = i$ for large i , the Normal model is recovered. Certainly, our working recommendation, both from conceptual and analytical perspectives, is $a = 3$ and $b = 4$.

DISCUSSION

In this paper, we propose a new model for ^{14}C determinations that provides a more robust approach for the analysis of ^{14}C data. Using this new model, the effect of outlier observations is reduced in the relevant posterior distributions, without the need to include additional parameters nor removing determinations from the data set. The resulting posterior distribution for θ has a smoother shape in comparison to the common Normal model, and the coverage probability of HPD regions is closer to the posterior $1 - \alpha$ probability. By plugging in the new model into the general statistical framework for ^{14}C data proposed by Christen (1994) and Buck et al. (2003), we obtain a method of ^{14}C dating robust to outlier observations and other causes of overdispersed data, with far fewer parameters (the common outlier analysis model presented by Christen 1994 introduces 2 parameters per ^{14}C determination). Certainly, using standard results related to the Student's t distribution, it may be seen that as we increase the sample size, the proposed method will produce inferential results similar to those obtained under the conventional Normal model. Further research is needed regarding the sensitivity of posterior summaries, in a wide range of dating scenarios, to the choice of the prior for the variance multiplier and the general applicability of our choice ($a = 3$ and $b = 4$).

ACKNOWLEDGMENTS

This research was partially founded by the Consejo de Ciencia y Tecnología del Estado de Guanajuato (CONCyTEG), Mexico, project number 08-02-K662-075. Sergio Pérez Elizalde was partially funded by a CONACyT postdoctoral fellowship.

REFERENCES

- Berger JO, Liseo B, Wolpert RL. 1999. Integrated likelihood methods for eliminating nuisance parameters (with discussion). *Statistical Science* 14(1):1–28.
- Bernardo JM. 1979. Reference posterior distributions for Bayesian inference. *The Journal of the Royal Statistical Society B* 41(2):113–47.
- Bernardo JM. 2005. Reference analysis. In: Dey DK, Rao CR, editors. *Handbook of Statistics* 25. Amsterdam: Elsevier. p 17–90.
- Blaauw M, Christen JA. 2005. Radiocarbon peat chronologies and environmental change. *Applied Statistics* 54(5):805–16.
- Blaauw M, Christen JA, Guilderson TP, Reimer PJ, Brown TA. 2005. The problems of radiocarbon dating. *Science* 308(5728):1551–3.
- Blaauw M, Bakker R, Christen JA, Hall VA, van der Plicht J. 2007. A Bayesian framework for age modeling of radiocarbon-dated peat deposits: case studies from the Netherlands. *Radiocarbon* 49(2):357–67.
- Buck CE, Blackwell PG. 2004. Formal statistical models for estimating radiocarbon calibration curves. *Radiocarbon* 46(3):1093–102.
- Buck CE, Christen JA, James GN. 1999. BCal: an on-line Bayesian radiocarbon calibration tool. *Internet Archaeology* 7. Available at http://intarch.ac.uk/journal/issue7/buck_index.html.
- Buck CE, Higham TFG, Lowe DJ. 2003. Bayesian tools for tephrochronology. *The Holocene* 13(5):639–47.
- Buck CE, Gómez Portugal D, Litton CD, O'Hagan A. 2006. Bayesian non-parametric estimation of the cali-

- bration curve for radiocarbon dating. *Bayesian Analysis* 1(2):265–88.
- Chen MH, Shao QM. 1999. Monte Carlo estimation of Bayesian credible and HPD intervals. *Journal of Computational and Graphical Statistics* 8(1):69–92.
- Christen JA. 1994. Summarizing a set of radiocarbon determinations: a robust approach. *Applied Statistics-Journal of the Royal Statistical Society Series C* 43(3): 489–503.
- Clark R. 1975. A calibration curve for radiocarbon dates. *Antiquity* 49(196):251–66.
- Damon PE, Donahue DJ, Gore BH, Hatheway AL, Jull AJT, Linick TW, Sercel PJ, Toolin LJ, Bronk CR, Hall ET, Hedges REM, Housley R, Law IA, Perry C, Bonani G, Trumbore S, Ambers JC, Bowman SGE, Leese MN, Tite MS. 1989. Radiocarbon dating of the Shroud of Turin. *Nature* 337(6208):611–5.
- Gibbons JD, Chakraborti S. 2003. *Nonparametric Statistical Inference*. Boca Raton: CRC Press. 4th edition. 306 p.
- Naylor JC, Smith AFM. 1988. An archaeological inference problem. *Journal of the American Statistical Association* 83(403):588–95.
- Reimer PJ, Baillie MGL, Bard E, Bayliss A, Beck JW, Bertrand CJH, Blackwell PG, Buck CE, Burr GS, Cutler KB, Damon PE, Edwards RL, Fairbanks RG, Friedrich M, Guilderson TP, Hogg AG, Hughen KA, Kromer B, McCormac G, Manning S, Bronk Ramsey C, Reimer RW, Remmele S, Southon JR, Stuiver M, Talamo S, Taylor FW, van der Plicht J, Weyhenmeyer CE. 2004. IntCal04 terrestrial radiocarbon age calibration, 0–26 cal kyr BP. *Radiocarbon* 46(3):1029–58.
- Scott EM. 2003. Section 10: Summary and Conclusions. In: The Third International Radiocarbon Intercomparison (TIRI) and the Fourth International Radiocarbon (FIRI), 1990–2002. Results, Analysis and Conclusions. *Radiocarbon* 45(2):285–91.
- Ward GK, Wilson SR. 1978. Procedures for comparing and combining radiocarbon age determinations: a critique. *Archaeometry* 20(1):19–31.