A New Robust Statistical Model for Radiocarbon Data

J. Andrés Christen    Sergio Pérez-Elizalde

Centro de Investigación en Matemáticas
Guanajuato, México.

Paleochronology Building Workshop, San Miguel Allende,
Motivation

Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a “radiocarbon date” and a standard error is the output of the dating process, \( y \pm \sigma \) (eg. 4500 \( \pm \) 30).

The general method currently used to analyze radiocarbon data (\( y \)) is conditional on the standard deviation (\( \sigma \)). Nevertheless, \( \sigma \) is assumed as known in the usual statistical model for radiocarbon data.

We want to propose a robust analysis in the presence of atypical data.

and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.
Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a “radiocarbon date” and a standard error is the output of the dating process, \( y \pm \sigma \) (eg. 4500 ± 30).

The general method currently used to analyze radiocarbon data \((y)\) is conditional on the standard deviation \((\sigma)\). Nevertheless, \(\sigma\) is assumed as known in the usual statistical model for radiocarbon data.

We want to propose a robust analysis in the presence of atypical data.

and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.
Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a “radiocarbon date” and a standard error is the output of the dating process, $y \pm \sigma$ (eg. 4500 ± 30).

The general method currently used to analyze radiocarbon data ($y$) is conditional on the standard deviation ($\sigma$). Nevertheless, $\sigma$ is assumed as known in the usual statistical model for radiocarbon data.

We want to propose a robust analysis in the presence of atypical data.

and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.
Radiocarbon dating is a method to approximate the age of organic samples and after a complex and costly process a “radiocarbon date” and a standard error is the output of the dating process, $y \pm \sigma$ (eg. 4500 ± 30).

The general method currently used to analyze radiocarbon data ($y$) is conditional on the standard deviation ($\sigma$). Nevertheless, $\sigma$ is assumed as known in the usual statistical model for radiocarbon data.

We want to propose a robust analysis in the presence of atypical data.

and understand and explain the scatter in radiocarbon data seen in interlaboratory studies.
Distribution of offsets relative to the dendro-dated samples
Traditional model

The traditional statistical model for a $^{14}$C determination $y_j$ is given by

$$y_j \sim N\left(\mu(\theta), \sigma^2_j\right), \quad j = 1, 2, \ldots, m \tag{1}$$

- $\mu(\cdot)$ is the calibration curve
- $\theta$ is the associated calendar year
- $\sigma_j$ is the reported standard deviation for $y_j$
- For a given $\theta$ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

$$y_j \sim N\left(\mu(\theta), \sigma^2_j + \sigma^2(\theta)\right), \tag{2}$$

where $\sigma_j$ is assumed as known; this is the basic statistical model currently used for the statistical analysis of $^{14}$C data
Traditional model

The traditional statistical model for a $^{14}$C determination $y_j$ is given by

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

(1)

- $\mu(\cdot)$ is the calibration curve
- $\theta$ is the associated calendar year
- $\sigma_j$ is the reported standard deviation for $y_j$
- For a given $\theta$ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

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

(2)

where $\sigma_j$ is assumed as known; this is the basic statistical model currently used for the statistical analysis of $^{14}$C data.
The traditional statistical model for a $^{14}\text{C}$ determination $y_j$ is given by

$$y_j \sim N \left( \mu(\theta), \sigma_j^2 \right), \quad j = 1, 2, \ldots, m$$  \hspace{1cm} (1)

- $\mu(\cdot)$ is the calibration curve
- $\theta$ is the associated calendar year
- $\sigma_j$ is the reported standard deviation for $y_j$
- For a given $\theta$ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

$$y_j \sim N \left( \mu(\theta), \sigma_j^2 + \sigma^2(\theta) \right),$$  \hspace{1cm} (2)

where $\sigma_j$ is assumed as known; this is the basic statistical model currently used for the statistical analysis of $^{14}\text{C}$ data
Traditional model

The traditional statistical model for a $^{14}$C determination $y_j$ is given by

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

(1)

- $\mu(\cdot)$ is the calibration curve
- $\theta$ is the associated calendar year
- $\sigma_j$ is the reported standard deviation for $y_j$
- For a given $\theta$ we use an estimate of both $\mu(\theta)$ and its standard deviation $\sigma(\theta)$ (for example INTCAL04). Model (1) becomes

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

(2)

where $\sigma_j$ is assumed as known; this is the basic statistical model currently used for the statistical analysis of $^{14}$C data
The traditional likelihood function

The likelihood function for $\theta$ given a random sample $y = (y_1, \ldots, y_m)$ of $m$ $^{14}$C determinations is

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

where $\omega_{j}^{2}(\theta) = \sigma^2(\theta) + \sigma_j^2$.

We derive the posterior distribution of $\theta$ by formal use of the Bayes’ rule; that is,

$$\pi(\theta \mid y) \propto L(\theta \mid y) \pi(\theta). \quad (4)$$
The traditional likelihood function

- The likelihood function for $\theta$ given a random sample $y = (y_1, \ldots, y_m)$ of $m$ $^{14}\text{C}$ determinations is

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

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

- We derive the posterior distribution of $\theta$ by formal use of the Bayes’ rule; that is,

$$\pi(\theta \mid y) \propto L(\theta \mid y)\pi(\theta). \quad (4)$$
The traditional posterior function

For the prior of $\theta$ we will use a uniform distribution on the interval $(B_2, B_1)$, $B_1 < B_2$.

Further prior information about $\theta$ may be included through any other prior distribution.

With the conventional normal model the posterior is proportional to the likelihood $L$, i.e.,

$$\pi(\theta | y) \propto L(\theta | y).$$
The traditional posterior function

- For the prior of $\theta$ we will use a uniform distribution on the interval $(B_2, B_1)$, $B_1 < B_2$.

- Further prior information about $\theta$ may be included through any other prior distribution.

- With the conventional normal model the posterior is proportional to the likelihood $L$, i.e.,

  $$\pi_N(\theta | y) \propto L_N(\theta | y).$$
The traditional posterior function

For the prior of $\theta$ we will use a uniform distribution on the interval $(B_2, B_1)$, $B_1 < B_2$.

Further prior information about $\theta$ may be included through any other prior distribution.

With the conventional normal model the posterior is proportional to the likelihood $L_N$, i.e.,

$\pi_N(\theta | y) \propto L_N(\theta | y)$. 

Cal BP
The traditional posterior function

- For the prior of $\theta$ we will use a uniform distribution on the interval $(B_2, B_1)$, $B_1 < B_2$.

- Further prior information about $\theta$ may be included through any other prior distribution.

- With the conventional normal model the posterior is proportional to the likelihood $L_N$; i.e.,

$$\pi_N(\theta | \mathbf{y}) \propto L_N(\theta | \mathbf{y}).$$
The traditional posterior function

- For the prior of $\theta$ we will use a uniform distribution on the interval $(B_2, B_1), \ B_1 < B_2$.

- Further prior information about $\theta$ may be included through any other prior distribution.

- With the conventional normal model the posterior is proportional to the likelihood $L_N$; i.e.,

$$\pi_N(\theta | y) \propto L_N(\theta | y).$$
The traditional posterior function

- For the prior of $\theta$ we will use a uniform distribution on the interval $(B_2, B_1)$, $B_1 < B_2$.

- Further prior information about $\theta$ may be included through any other prior distribution.

- With the conventional normal model the posterior is proportional to the likelihood $L_N$; i.e.,

$$\pi_N(\theta | y) \propto L_N(\theta | y).$$
Classic calibration (using software CALIB)

Calibration of a $^{14}$C determination $606 \pm 40$ using CALIB.
Traditional model: disadvantages

- This traditional model assumes that $\sigma_j$ is known exactly. However, $\sigma_j$ is calculated at each laboratory and strictly speaking is not known precisely.

- 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 common in practice (Blaauw et al., 2005).

- Even for the simplest of cases Christen (1994) approach to detect outliers requires the use of complex numerical techniques (eg. MCMC).

- International interlaboratory studies show “unexplained” scatter in $^{14}$C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal
This traditional model assumes that $\sigma_j$ is known exactly. However, $\sigma_j$ is calculated at each laboratory and strictly speaking is not known precisely.

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 common in practice (Blaauw et al., 2005).

Even for the simplest of cases Christen (1994) approach to detect outliers requires the use of complex numerical techniques (eg. MCMC).

International interlaboratory studies show “unexplained” scatter in $^{14}$C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal
Traditional model: disadvantages

- This traditional model assumes that $\sigma_j$ is known exactly. However, $\sigma_j$ is calculated at each laboratory and strictly speaking is not known precisely.

- 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 common in practice (Blaauw et al., 2005).

- Even for the simplest of cases Christen (1994) approach to detect outliers requires the use of complex numerical techniques (eg. MCMC).

- International interlaboratory studies show “unexplained” scatter in $^{14}$C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal
This traditional model assumes that $\sigma_j$ is known exactly. However, $\sigma_j$ is calculated at each laboratory and strictly speaking is not known precisely.

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 common in practice (Blaauw et al., 2005).

Even for the simplest of cases Christen (1994) approach to detect outliers requires the use of complex numerical techniques (eg. MCMC).

International interlaboratory studies show “unexplained” scatter in $^{14}$C data. An unexplored alternative would be to change the model to a heavier tailed distribution than the Normal
Christen (1994) approach to detect outliers considers that each radiocarbon determination might need a shift in the radiocarbon scale $\delta_j$, in order to be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim N \left( \mu(\theta) + \phi_j \delta_j, \sigma_j^2 \right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination $j$ does require or does not require a shift ($\delta_j$) to be properly explained (is or is not an outlier).

The posterior probability $P[\phi_j = 1|\text{All data and prior info}]$ is calculated. This main be interpreted as our (posterior) probability that determination $j$ is an outlier.
Christen (1994) approach to detect outliers considers that each radiocarbon determination might need a shift in the radiocarbon scale \(\delta_j\), in order to be properly explained in terms of the rest of dates and the contextual information used.

That is:

\[
y_j \sim N\left(\mu(\theta) + \phi_j \delta_j, \sigma_j^2\right), \quad j = 1, 2, \ldots, m,
\]

where \(\phi_j = 1, 0\) depending on whether determination \(j\) does require or does not require a shift (\(\delta_j\)) to be properly explained (is or is not an outlier).

The posterior probability \(P[\phi_j = 1|\text{All data and prior info}]\) is calculated. This main be interpreted as our (posterior) probability that determination \(j\) is an outlier.
Christen (1994) approach to detect outliers considers that each radiocarbon determination might need a shift in the radiocarbon scale $\delta_j$, in order to be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim N \left( \mu(\theta) + \phi_j \delta_j, \sigma_j^2 \right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination $j$ does require or does not need a shift ($\delta_j$) to be properly explained (is or is not an outlier).

The posterior probability $P[\phi_j = 1|\text{All data and prior info}]$ is calculated. This may be interpreted as our (posterior) probability that determination $j$ is an outlier.
Christen (1994) approach to detect outliers considers that each radiocarbon determination might need a shift in the radiocarbon scale $\delta_j$, in order to be be properly explained in terms of the rest of dates and the contextual information used.

That is:

$$y_j \sim N\left(\mu(\theta) + \phi_j\delta_j, \sigma_j^2\right), \quad j = 1, 2, \ldots, m,$$

where $\phi_j = 1, 0$ depending on whether determination $j$ does require or does not require a shift ($\delta_j$) to be properly explained (is or is not an outlier).

The posterior probability $P[\phi_j = 1|\text{All data and prior info}]$ is calculated. This main be interpreted as our (posterior) probability that determination $j$ is an outlier.
The new model

Normal model with a variance multiplier

- We know that $\sigma_j$ varies jointly with $y_j$.
- The uncertainty about the variance of $y_j$ in model (1) may be introduced by considering the product $\alpha \sigma_j^2$, where $\alpha > 0$. The new model is
  
  $$y_j \sim N \left( \mu(\theta), \alpha \sigma_j^2 \right)$$

- $\alpha$ is an unknown “variance multiplier” to the laboratory reported variance $\sigma_j^2$.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,
  
  $$y_j \sim N \left( \mu(\theta), \alpha \left( \sigma^2(\theta) + \sigma_j^2 \right) \right)$$
The new model

Normal model with a variance multiplier

- We know that $\sigma_j$ varies jointly with $y_j$.
- The uncertainty about the variance of $y_j$ in model (1) may be introduced by considering the product $\alpha \sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim N \left( \mu(\theta), \alpha \sigma_j^2 \right)$$

- $\alpha$ is an unknown “variance multiplier” to the laboratory reported variance $\sigma_j^2$.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,

$$y_j \sim N \left( \mu(\theta), \alpha \left( \sigma^2(\theta) + \sigma_j^2 \right) \right)$$
Normal model with a variance multiplier

We know that $\sigma_j$ varies jointly with $y_j$.

The uncertainty about the variance of $y_j$ in model (1) may be introduced by considering the product $\alpha \sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim N \left( \mu(\theta), \alpha \sigma_j^2 \right)$$

$\alpha$ is an unknown “variance multiplier” to the laboratory reported variance $\sigma_j^2$.

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

$$y_j \sim N \left( \mu(\theta), \alpha \left( \sigma^2(\theta) + \sigma_j^2 \right) \right)$$
The new model

Normal model with a variance multiplier

- We know that $\sigma_j$ varies jointly with $y_j$.
- The uncertainty about the variance of $y_j$ in model (1) may be introduced by considering the product $\alpha \sigma_j^2$, where $\alpha > 0$. The new model is

$$y_j \sim N \left( \mu(\theta), \alpha \sigma_j^2 \right)$$

- $\alpha$ is an unknown “variance multiplier” to the laboratory reported variance $\sigma_j^2$.
- If we also consider a model which uses the variance $\sigma^2(\theta)$ in the calibration curve,

$$y_j \sim N \left( \mu(\theta), \alpha \left( \sigma^2(\theta) + \sigma_j^2 \right) \right)$$
The new model

- Calibration curve based on high quality data, so might need more optimistic $\alpha$ for $\sigma(\theta)$. But, then MCMC needed to infer model parameters.

- We assume multiplier $\alpha$ also affects $\sigma(\theta)$, ensures mathematical tractability and analytically feasible representation of $\theta$ posterior distribution.

$$y_j \sim N \left( \mu(\theta), \alpha_1 \sigma^2(\theta) + \alpha_2 \sigma_j^2 \right).$$  \hspace{1cm} (5)

- Typically $\sigma(\theta)$ small compared to $\sigma$, model well behaved approximation to the more realistic model.
The new model

- Calibration curve based on high quality data, so might need more optimistic $\alpha$ for $\sigma(\theta)$. But, then MCMC needed to infer model parameters.

- We assume multiplier $\alpha$ also affects $\sigma(\theta)$, ensures mathematical tractability and analytically feasible representation of $\theta$ posterior distribution.

$$y_j \sim \mathcal{N} \left( \mu(\theta), \alpha_1 \sigma^2(\theta) + \alpha_2 \sigma^2_j \right).$$

- Typically $\sigma(\theta)$ small compared to $\sigma$, model well behaved approximation to the more realistic model.
The new model

- Calibration curve based on high quality data, so might need more optimistic $\alpha$ for $\sigma(\theta)$. But, then MCMC needed to infer model parameters.

- We assume multiplier $\alpha$ also affects $\sigma(\theta)$, ensures mathematical tractability and analytically feasible representation of $\theta$ posterior distribution.

- Typically $\sigma(\theta)$ small compared to $\sigma$, model well behaved approximation to the more realistic model.

$$y_j \sim N \left( \mu(\theta), \alpha_1 \sigma^2(\theta) + \alpha_2 \sigma_j^2 \right).$$

(5)
Double multiplier model vs. single multiplier model

Probability plot for the double multiplier normal model (vertical axis, blue region), single multiplier normal model (vertical axis, red line) vs. the traditional normal model (horizontal axis). $\theta = 500$, $\sigma = 50$. 
Double multiplier model vs. single multiplier model

Probability plot for the double multiplier normal model (vertical axis, blue region), single multiplier normal model (vertical axis, red line) vs. the traditional normal model (horizontal axis). $\theta = 500$, $\sigma = 50$. 
Double multiplier model vs. single multiplier model

Probability plot for the double multiplier normal model (vertical axis, blue region), single multiplier normal model (vertical axis, red line) vs. the traditional normal model (horizontal axis). $\theta = 500$, $\sigma = 50$. 
The prior for $\alpha$

We assume that the prior distribution for $\alpha$ is an inverted gamma with parameters $a$ and $b$

$$\pi(\alpha) = \text{InvGa}(\alpha | a, b).$$  \hfill (6)

Then, given $\theta$, the prior distribution of $\alpha \omega_j^2$ is the inverted gamma

$$\alpha \omega_j^2 | \theta \sim \text{InvGa}\left(a, b \left(\sigma_j^2 + \sigma^2(\theta)\right)\right),$$  \hfill (7)

such that

$$E\left(\alpha \omega_j^2 | \theta\right) = \frac{b}{a-1} \left(\sigma_j^2 + \sigma^2(\theta)\right)$$

is the prior expected variance of $y_j$.
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.
A proposed prior for $\alpha$

Prior density for the variance multiplier $\alpha$ with expected value $E(\alpha) = b/(a - 1) = 2$, mode $Mo(\alpha) = b/(a + 1) = 1$ and median $Me(\alpha) \approx 1.5$, $Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3$, $b = 4$.

$InvGa(\alpha | a = 3, b = 4)$ represents $Pr(\sqrt{\alpha} > 2) \approx 0.08$, $Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.

The most likely scenario is that the error term was correctly reported.

The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.
A proposed prior for $\alpha$

Prior density for the variance multiplier $\alpha$ with expected value $E(\alpha) = b/(a-1) = 2$, mode $Mo(\alpha) = b/(a+1) = 1$ and median $Me(\alpha) \approx 1.5$, $Pr(\alpha \leq 1) \approx 0.24$, $Pr(\alpha \geq 4) \approx 0.08$, $a = 3$, $b = 4$. 

InvGa $\alpha | a = 3, b = 4$ represents $Pr(\sqrt{\alpha} > 2) \approx 0.08$, $Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$. 

The most likely scenario is that the error term was correctly reported. The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.
A proposed prior for $\alpha$

Prior density for the variance multiplier $\alpha$ with expected value $E(\alpha) = b/(a - 1) = 2$, mode $Mo(\alpha) = b/(a + 1) = 1$ and median $Me(\alpha) \approx 1.5$, $Pr(\alpha \leq 1) \approx 0.24$, $Pr(\alpha \geq 4) \approx 0.08$, $a = 3$, $b = 4$.

- $InvGa(\alpha \mid a = 3, b = 4)$ represents $Pr(\sqrt{\alpha} > 2) \approx 0.08$, $Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.
A proposed prior for $\alpha$

Prior density for the variance multiplier $\alpha$ with expected value $E(\alpha) = b/(a - 1) = 2$, mode $Mo(\alpha) = b/(a + 1) = 1$ and median $Me(\alpha) \approx 1.5$, $Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3$, $b = 4$.

- $\text{InvGa}(\alpha | a = 3, b = 4)$ represents $Pr(\sqrt{\alpha} > 2) \approx 0.08$, $Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.
- The most likely scenario is that the error term was correctly reported.
- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.
A proposed prior for $\alpha$

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

$\Pr(\alpha \leq 1) \approx 0.24$,

$P(\alpha \geq 4) \approx 0.08$, $a = 3$, $b = 4$.

- $\text{InvGa}(\alpha \mid a = 3, b = 4)$ represents
  $\Pr(\sqrt{\alpha} > 2) \approx 0.08$, $\Pr(\sqrt{\alpha} < 1) \approx 0.248$ and
  $\Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.

- The most likely scenario is that the error term was correctly reported.

- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.
A proposed prior for $\alpha$

Prior density for the variance multiplier $\alpha$ with expected value $E(\alpha) = b/(a - 1) = 2$, mode $Mo(\alpha) = b/(a + 1) = 1$ and median $Me(\alpha) \approx 1.5$, $Pr(\alpha \leq 1) \approx 0.24$, $P(\alpha \geq 4) \approx 0.08$, $a = 3$, $b = 4$.

- $\text{InvGa}(\alpha \mid a = 3, b = 4)$ represents
  $Pr(\sqrt{\alpha} > 2) \approx 0.08$, $Pr(\sqrt{\alpha} < 1) \approx 0.248$ and $Pr(1 < \sqrt{\alpha} < 2) \approx 0.672$.

- The most likely scenario is that the error term was correctly reported.

- The choice of $a = 3$ and $b = 4$ should be regarded as a practical guideline only.
Examples for the prior for $\alpha$
The new (integrated) likelihood

- The parameter of interest is the true calendar age $\theta$, being $\alpha$ a nuisance parameter.

- In a Bayesian setting nuisance parameters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function. Here we derive the posterior distribution for $\theta$ using the integrated likelihood:

$$L_{a,b}(\theta \mid y) = \int_0^{\infty} \prod_{j=1}^{m} p(y_j \mid \theta, \alpha) \pi_{a,b}(\alpha) \, d\alpha.$$  

- Note that we are assuming prior independence of the parameters $(\theta, \alpha)$. 

The new (integrated) likelihood

- The parameter of interest is the true calendar age $\theta$, being $\alpha$ a nuisance parameter.

- In a Bayesian setting nuisance parameters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function. Here we derive the posterior distribution for $\theta$ using the integrated likelihood:

$$L_{a,b} (\theta \mid y) = \int_0^\infty \prod_{j=1}^m p(y_j \mid \theta, \alpha) \pi_{a,b}(\alpha) \, d\alpha.$$  \hspace{0.5cm} (8)

- Note that we are assuming prior independence of the parameters $(\theta, \alpha)$. 
The new (integrated) likelihood

- The parameter of interest is the true calendar age $\theta$, being $\alpha$ a nuisance parameter.

- In a Bayesian setting nuisance parameters are naturally eliminated by integrating out them from either the posterior distribution or the likelihood function. Here we derive the posterior distribution for $\theta$ using the integrated likelihood:

\[
L_{a,b}(\theta | \mathbf{y}) = \int_0^\infty \prod_{j=1}^m p(y_j | \theta, \alpha) \pi_{a,b}(\alpha) \, d\alpha.
\]  

Note that we are assuming prior independence of the parameters $(\theta, \alpha)$.
Therefore, under the prior distribution (6) the integrated likelihood, given $y = (y_1, \ldots, y_m)$, is

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

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

$$\propto t\left(y | \mu(\theta)1_m, \Sigma(\theta)b/a, 2a\right), \quad (9)$$

where $1_m = (1, \ldots, 1)^t$ and $\Sigma(\theta) = \text{diag}(\omega_1(\theta), \ldots, \omega_m(\theta))$.

The integrated likelihood for $\theta$ given $y$ is proportional to a $t$ distribution with location parameter $\mu(\theta)1_m$, covariance matrix $\Sigma(\theta)b/(a - 1)$ and $2a$ d.f.
The new (integrated) likelihood

Therefore, under the prior distribution (6) the integrated likelihood, given \( y = (y_1, \ldots, y_m) \), is

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

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

\[
\propto t\left( y \bigg| \mu(\theta)1_m, \Sigma(\theta)b/a, 2a \right), \quad (9)
\]

where \( 1_m = (1, \ldots, 1)^t \) and \( \Sigma(\theta) = \text{diag}(\omega_1(\theta), \ldots, \omega_m(\theta)) \).

The integrated likelihood for \( \theta \) given \( y \) is proportional to a \( t \) distribution with location parameter \( \mu(\theta)1_m \), covariance matrix \( \Sigma(\theta)b/(a - 1) \) and \( 2a \) d.f..
The new (integrated) likelihood

Note that with the sequence of parameters values $a = i + 1$, $b = i$, $i = 1, 2, \ldots$, we obtain a sequence of prior distributions which 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.

As expected, our new model has as limiting case the standard normal model when \textit{a priori} $\Pr(\alpha = 1) = 1$; that is, $\sigma_j$ is known exactly.
The new (integrated) likelihood

Note that with the sequence of parameters values $a = i + 1, b = i, i = 1, 2, \ldots$, we obtain a sequence of prior distributions which 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.

As expected, our new model has as limiting case the standard normal model when $a \text{ priori } \Pr(\alpha = 1) = 1$; that is, $\sigma_j$ is known exactly.
Now we derive the posterior distribution of $\theta$ by formal use of the Bayes’ rule; that is,

$$\pi(\theta \mid y) \propto L(\theta \mid y)\pi(\theta).$$  \hspace{1cm} (10)

As a prior of $\theta$ 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 $\theta$ they may properly include it through any other prior distribution, exactly the same as in the traditional normal case.

The posterior distribution for $\theta$ is

$$\pi_{a,b}(\theta \mid y) \propto L_{a,b}(\theta \mid y)\pi(\theta) \propto L_{a,b}(\theta \mid y),$$ \hspace{1cm} (11)
Now we derive the posterior distribution of $\theta$ by formal use of the Bayes’ rule; that is,

$$\pi(\theta \mid y) \propto L(\theta \mid y)\pi(\theta).$$

(10)

As a prior of $\theta$ 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 $\theta$ they may properly include it through any other prior distribution, exactly the same as in the traditional normal case.

The posterior distribution for $\theta$ is

$$\pi_{a,b}(\theta \mid y) \propto L_{a,b}(\theta \mid y)\pi(\theta) \propto L_{a,b}(\theta \mid y),$$

(11)
Now we derive the posterior distribution of $\theta$ by formal use of the Bayes’ rule; that is,

$$\pi(\theta \mid y) \propto L(\theta \mid y)\pi(\theta).$$ \hfill (10)

As a prior of $\theta$ 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 $\theta$ they may properly include it through any other prior distribution, exactly the same as in the traditional normal case.

The posterior distribution for $\theta$ is

$$\pi_{a,b}(\theta \mid y) \propto L_{a,b}(\theta \mid y)\pi(\theta) \propto L_{a,b}(\theta \mid y),$$ \hfill (11)
Single $^{14}$C calibration
Single $^{14}$C calibration

![Graph showing calibration process with specific parameters and data points]
Single $^{14}$C calibration

Calibration Curve $\mu(\theta)$

$\mu(\theta) \pm \sigma(\theta)$

$y = 606$

$\sigma = 37$

$a = 50$, $b = 51$
Simulated Example

We analyze a set of simulated of \( m = 5 \) radiocarbon observations. The parameter values are \( \theta = 650, \sigma(\theta) = 12, \) and \( m = 5. \)

<table>
<thead>
<tr>
<th>id</th>
<th>Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>S1</td>
<td>649 ± 25</td>
</tr>
<tr>
<td>S2</td>
<td>598 ± 25</td>
</tr>
<tr>
<td>S3</td>
<td>748 ± 69</td>
</tr>
<tr>
<td>S4</td>
<td>606 ± 37</td>
</tr>
<tr>
<td>S5</td>
<td>368 ± 37</td>
</tr>
</tbody>
</table>
Simulated Example

The figure exhibits the simulated radiocarbon determinations plotted over the calibration curve. Note that there is an atypical observation (S5).
Simulated Example

Posterior densities for $\theta$, and their corresponding %95 HPD credible sets, for (a) $\pi_{3,4}$, (b) $\pi_N$, (c) $\pi_R$ and (d) $\pi^*_N$ (normal model not including observation S5).
Simulated Example

- Note that $\pi_N$ looks rougher, reproducing the wiggles in the calibration curve, while $\pi_{3,4}$ is smoother, and concentrated over the most likely region given the data.
- The effect of the outlying observation S5 causes the normal likelihood to shrink and shift to the right, leaving the true value for $\theta$ out of the 95% HPD region for $\pi_N$.
Simulated Example

- Note that $\pi_N$ looks rougher, reproducing the wiggles in the calibration curve, while $\pi_{3,4}$ is smoother, and concentrated over the most likely region given the data.

- The effect of the outlying observation S5 causes the normal likelihood to shrink and shift to the right, leaving the true value for $\theta$ out of the 95% HPD region for $\pi_N$. 

Simulated Example

- If we drop from the data S5 the resulting posterior arising from $\pi^*_N$ is not more informative than $\pi_N$.
- $\pi_{3,4}$ is based on all the data and the heavy tails of the underlying model ensure that we are properly including the information provided by possible extreme values.
- Our new approach is more cautious and results in wider smoother distributions.
- Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion.
Simulated Example

- If we drop from the data S5 the resulting posterior arising from $\pi^*_N$ is not more informative than $\pi_N$.
- $\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.
- Our new approach is more cautious and results in wider smoother distributions.
- Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion.
If we drop from the data S5 the resulting posterior arising from $\pi^*_N$ is not more informative than $\pi_N$.

$\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.

Our new approach is more cautious and results in wider smoother distributions.

Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion.
If we drop from the data S5 the resulting posterior arising from $\pi^*_N$ is not more informative than $\pi_N$.

$\pi_{3,4}$ is based on all the data and the heavy tails of the underlaying model ensure that we are properly including the information provided by possible extreme values.

Our new approach is more cautious and results in wider smoother distributions.

Shorter intervals may be obtained by dropping outlier determinations, but the gain in precision, given the amount of atypical information, is an illusion.
Table: Radiocarbon determinations for the ‘Shroud of Turin’.

<table>
<thead>
<tr>
<th>Laboratory</th>
<th>id</th>
<th>Determination</th>
</tr>
</thead>
<tbody>
<tr>
<td>Arizona</td>
<td>A1</td>
<td>591 ± 30</td>
</tr>
<tr>
<td></td>
<td>A2</td>
<td>690 ± 35</td>
</tr>
<tr>
<td></td>
<td>A3</td>
<td>606 ± 41</td>
</tr>
<tr>
<td></td>
<td>A4</td>
<td>701 ± 33</td>
</tr>
<tr>
<td>Oxford</td>
<td>O1</td>
<td>795 ± 65</td>
</tr>
<tr>
<td></td>
<td>O2</td>
<td>730 ± 45</td>
</tr>
<tr>
<td></td>
<td>O3</td>
<td>745 ± 55</td>
</tr>
<tr>
<td>Zurich</td>
<td>Z1</td>
<td>733 ± 61</td>
</tr>
<tr>
<td></td>
<td>Z2</td>
<td>722 ± 56</td>
</tr>
<tr>
<td></td>
<td>Z3</td>
<td>635 ± 57</td>
</tr>
<tr>
<td></td>
<td>Z4</td>
<td>639 ± 45</td>
</tr>
<tr>
<td></td>
<td>Z5</td>
<td>679 ± 51</td>
</tr>
</tbody>
</table>
Differences in the determination process suggest the use a different $\alpha$ for each laboratory. The likelihood is

$$L(\theta, \alpha | y) = \prod_{i=1}^{n} \prod_{j=1}^{m_i} (2\pi \alpha_i \omega_{ij}^2(\theta))^{-m_i/2} \exp \left\{ -\frac{1}{2\alpha_i \omega_{ij}^2(\theta)} (y_{ij} - \mu(\theta))^2 \right\} \quad (12)$$

Where $n = 3$, $m_1 = 4$, $m_2 = 3$ and $m_3 = 5$. Integrating the likelihood function w.r.t. an InvGa prior for each $\alpha_i$, the integrated likelihood is

$$L_{a,b}(\theta | y) \propto \prod_{i=1}^{3} t \left( y_i \bigg| \mu(\theta)1_{m_i}, \Sigma_i(\theta) b/a, 2a \right), \quad (13)$$

where $\Sigma_i(\theta) = \text{diag} (\omega_{i1}(\theta), \ldots, \omega_{m_1}(\theta))$. Thus, $L_{a,b}$ is the product of three multivariate $t$ densities.
Figure: Posterior densities and 95% HPD regions (under shaded area) of $\theta$ for the Shroud of Turin data. (a) $\pi_N$, (b) $\pi_N^*$, and (c) $\pi_{3,4}$. 
Operating Characteristics

In order to analyze the performance of our proposed model we estimate with Monte Carlo simulation the “coverage probability” of 95% HPD sets.

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

<table>
<thead>
<tr>
<th>Posterior Distribution</th>
<th>$p$ (outlier probability)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>0.0</td>
</tr>
<tr>
<td>$\pi_{a=3,b=4}$</td>
<td>0.9806</td>
</tr>
<tr>
<td>$\pi_{a=15,b=16}$</td>
<td>0.9650</td>
</tr>
<tr>
<td>$\pi_N$</td>
<td>0.9556</td>
</tr>
<tr>
<td>$\pi_R$</td>
<td>0.9558</td>
</tr>
</tbody>
</table>
Operating Characteristics

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.

**Table:** Average rounded count of the 95% HPD sets for different values of $p$.

<table>
<thead>
<tr>
<th>Posterior Distribution</th>
<th>$p$ (outlier probability)</th>
</tr>
</thead>
<tbody>
<tr>
<td>$\pi_{a=3,b=4}$</td>
<td>0.0</td>
</tr>
<tr>
<td></td>
<td>72</td>
</tr>
<tr>
<td>$\pi_{a=15,b=16}$</td>
<td>65</td>
</tr>
<tr>
<td>$\pi_N$</td>
<td>61</td>
</tr>
<tr>
<td>$\pi_R$</td>
<td>73</td>
</tr>
</tbody>
</table>
More complex dating problems

- We may consider the above formulation as the hierarchical model

\[ y \leftarrow \theta \leftarrow (a, b) \]

hyperparameters are introduced to model specific features of the data

- The general dating model is of the form

\[ y \leftarrow \theta \leftarrow \psi, \]

where \( y \) is a generic representation of data obtained under diverse sampling schemes, \( \theta \) is a vector of several calendar BP years and \( \psi \) contains \( (a, b) \) and quantities related to the phenomena being dated
We may consider the above formulation as the hierarchical model

\[ y \leftarrow \theta \leftarrow (a, b) \]

hyperparameters are introduced to model specific features of the data.

The general dating model is of the form

\[ y \leftarrow \theta \leftarrow \psi, \]

were \( y \) is a generic representation of data obtained under diverse sampling schemes, \( \theta \) is a vector of several calendar BP years and \( \psi \) contains \((a, b)\) and quantities related to the phenomena being dated.
More complex dating problems

- This representation is given by Christen (1994) and implemented in BCal, OxCal, Bpeat, etc.
- That is, for any dating problem we obtain a more robust analysis of radiocarbon data by substituting the normal model (2) by the $t$ model in (9).
- Setting $a = i + 1$, $b = i$ for large $i$ the normal model is recovered. Our working recommendation, both from conceptual and analytical perspectives, is $a = 3$ and $b = 4$. 
More complex dating problems

- This representation is given by Christen (1994) and implemented in BCal, OxCal, Bpeat, etc.
- That is, for any dating problem we obtain a more robust analysis of radiocarbon data by substituting the normal model (2) by the \( t \) model in (9).
- Setting \( a = i + 1, b = i \) for large \( i \) the normal model is recovered.
- Our working recommendation, both from conceptual and analytical perspectives, is \( a = 3 \) and \( b = 4 \).
More complex dating problems

- This representation is given by Christen (1994) and implemented in BCal, OxCal, Bpeat, etc.
- That is, for any dating problem we obtain a more robust analysis of radiocarbon data by substituting the normal model (2) by the $t$ model in (9).
- Setting $a = i + 1$, $b = i$ for large $i$ the normal model is recovered. Our working recommendation, both from conceptual and analytical perspectives, is $a = 3$ and $b = 4$. 
Using MCMC we estimate

$$\pi(\alpha \mid y) \propto \int_{\theta \in \Theta} p(y \mid \theta, \alpha) \pi(\theta, \alpha) \, d\theta.$$
Discussion

- The effect of outlier observations is reduced without additional parameters nor removing determinations.
- The posterior for $\theta$ has a smoother shape and the coverage of HPD regions is closer to the posterior $1 - \alpha$ probability.
- By plugging in the new model into the general statistical framework proposed by Christen (1994) and Buck et al. (2003) we obtain a method robust to outlier observations and other causes of overdispersed data, with far fewer parameters.
Discussion

- The effect of outlier observations is reduced without additional parameters nor removing determinations.
- The posterior for $\theta$ has a smoother shape and the coverage of HPD regions is closer to the posterior $1 - \alpha$ probability.
- By plugging in the new model into the general statistical framework proposed by Christen (1994) and Buck et al. (2003) we obtain a method robust to outlier observations and other causes of overdispersed data, with far fewer parameters.
The effect of outlier observations is reduced without additional parameters nor removing determinations.

The posterior for $\theta$ has a smoother shape and the coverage of HPD regions is closer to the posterior $1 - \alpha$ probability.

By plugging in the new model into the general statistical framework proposed by Christen (1994) and Buck et al. (2003) we obtain a method robust to outlier observations and other causes of overdispersed data, with far fewer parameters.
As we increase either $a$ or sample size, the proposed method will produce inferential results similar to those obtained under the normal model.

Further research is needed regarding the sensitivity of posterior summaries to the choice of the prior $\alpha$ and the general applicability of our choice ($a = 3$ and $b = 4$).
Discussion

- As we increase either $a$ or sample size, the proposed method will produce inferential results similar to those obtained under the normal model.
- Further research is needed regarding the sensitivity of posterior summaries to the choice of the prior $\alpha$ and the general applicability of our choice ($a = 3$ and $b = 4$).