NetC14: A Gaussian Mixture Neural Network for Calibrating and Summarizing Radiocarbon Dates

Kenneth B. Vernon[1,2,3], Brian F. Codding[4,5], Simon Brewer[1,2]

[1] *Scientific Computing and Imaging Institute, University of Utah*
[2] *School of Environment, Society, and Sustainability, University of Utah*
[3] *Department of Anthropology, University of Utah*
[4] *Environmental Studies Program, UC Santa Barbara*
[5] *Department of Geography, UC Santa Barbara*

Last updated: 2026-08-01

Abstract
Calibrating and summarizing radiocarbon dates are now central to archaeological inquiry, but standard approaches either require multiple steps that introduce statistical challenges or are computationally-intensive which limits general application. Here we propose Net14C, a quasi-parametric Gaussian mixture neural network for calibrating and summarizing radiocarbon dates using regularized Maximum Likelihood. Its principal methodological contribution is the application of the SparseMax activation to the mixture weights of a deliberately overspecified model, allowing it to prune components during training while preserving differentiability and satisfying the fixed-dimensionality requirement of neural architectures. The model is evaluated using simulated data and compared to two traditional approaches (Summed Probability Distribution and Composite Kernel Density Estimation) as well as two recently proposed Bayesian mixture models fit with MCMC (Bayesian Gaussian Mixture Model and Dirichlet Process Mixture Model), making it the first attempt to systematically evaluate these recent methods. All models are further applied to two published archaeological datasets from Iron Age Ireland and Paleoindian North America. Results indicate that NetC14 attains parity with existing mixture models on both age and density estimation while achieving substantially lower compute times and superior scaling, even absent GPU acceleration. NetC14 also inherits the extensibility of neural networks, allowing for integration of additional diagnostic data, debiasing techniques, and demographic constraints.

Keywords: Deep Learning, SparseMax, Regularized Maximum Likelihood, Density Estimation, Paleodemography, Bayesian Inference

Target Journal:

Correspondence: k.vernon@sci.utah.edu (Blake Vernon)

1. Introduction

Archaeological radiocarbon (^{14}C) dating has two primary objectives: (i) to estimate the true ages of a set of organic samples and (ii) to reconstruct the true calendar age distribution $p(t)$ from which those samples are drawn, with the latter serving as a proxy for population size over time. The first objective was established by the chemist Willard Libby in the late 1940s (Libby, 1946), the second by the archaeologist John Rick four decades later (Rick, 1987). Among archaeologists, a common strategy for achieving them is to first estimate the ages of samples individually and only then bring them together to reconstruct the distribution from which the samples are drawn, typically by summing the posterior probabilities of the individual age estimates, the Summed Probability Distribution (SPD) method (Blockley et al., 2000; Crema, 2022). This approach offers a decent approximation to the truth, but also suffers from a few well-known problems, like generating noisy estimates (Brown, 2015), but also perhaps most importantly, being statistically incoherent (Carleton & Groucutt, 2021; Crema, 2022; Heaton, 2022).

Recently, however, Heaton (2022) and Price et al. (2021) have proposed an alternative Bayesian strategy, pursuing both radiocarbon dating objectives simultaneously, estimating ages and the chronology they are drawn from at the same time using latent variable models. To the extent that their approaches differ, it is in the specification of $p(t)$. Price et al. (2021) recommend a parametric Bayesian Gaussian Mixture Model (BGMM), Heaton (2022) a non-parameteric Dirichlet Process Mixture Model (DPMM), with the latter learning the number of Gaussian mixture components during model training. While significant advances in understanding and integrating the two goals of archaeological radiocarbon dating, these approaches are computationally intensive and limited to a specific type of data, rendering general use challenging.

Here we propose a third mixture model, a Gaussian Mixture Neural Network that we call NetC14. NetC14 is quasi-parametric as it begins with an initially over-specified number of Gaussian components, then uses a SparseMax activation (Martins & Astudillo, 2016) on mixture weights to prune components out of the model. In this paper, we compare the ability of these three mixture models - BGMM, DPMM, and NetC14 - plus two traditional approaches - SPD and Composite Kernel Density Estimation (CKDE) - to calibrate and summarize radiocarbon dates using simulated data. We also apply them to real datasets used in Heaton (2022) and Bronk Ramsey (2017). Our results show that a minimal specification of NetC14 - severely limited in its complexity and implementation - is already competitive with the state of the art, but substantially more

computationally efficient and scalable to larger datasets, and that is before harnessing GPU acceleration and other optimizations. NetC14 also comes with all the advantages of neural networks, making them highly flexible and adaptable to a wide variety of archaeological data, which we discuss below.

2. Background

In this section, we review the Bayesian logic of the mixture models. For more background on archaeological radiocarbon dating, see (Bayliss, 2009; Bronk Ramsey, 2008; Hajdas et al., 2021; Taylor & Bar-Yosef, 2016; Wood, 2015).

2.1. Generative Model

Our data consist of fraction modern radiocarbon measurements, a dimensionless unit comparing the $^{14}C/^{12}C$ ratio in organic samples to a modern reference standard of atmospheric carbon levels in 1950 (McNichol et al., 2001). Adapting Heaton (2022) to this fraction modern context, we define the generative model for radiocarbon measurements F_i on $i = 1, ..., n$ organic samples as

$$t_1, ..., t_n \sim p(t)$$
$$F_i \sim N(\mu(t_i), z_i^2)$$

where $t_1, ..., t_n$ are the true but unknown ages of the n samples drawn from the age distribution $p(t)$, $\mu(t)$ is the fraction modern predicted by a radiocarbon calibration curve for a given age t, and z_i is uncertainty around the measured fraction modern for sample i. In plain English, we might think of this generative model as answering two archaeological questions. First, how likely is it that someone in the past did something at a particular time and place that leaves some organic material behind marked by the death of an organism, like butchering a deer or chopping down a tree? That is a random draw of a calendar age t_i from $p(t)$. Second, assuming that organic material preserves over the intervening years, how much ^{14}C should we expect to find in it relative to a modern reference standard when we actually get around to measuring it? That observation is a random draw of a fraction modern measurement F_i from $N(\mu(t_i), z_i^2)$, a Gaussian measurement model that incorporates uncertainty from both the measurements and the calibration curve.

We denote uncertainty in the calibration curve itself for a given age t as

$$\mu(t) \mid t \sim N(m(t), s(t)^2)$$

with $m(t)$ and $s(t)$ being the point-wise means and standard deviations of Gaussians defined by the calibration curve. Conventionally, these quantities are expressed as ^{14}C ages, but we assume that they have been converted to fraction modern using the mean-life function for radiocarbon. Given that F_i are also sampled from a Gaussian, the curve can be marginalized over, such that

$$F_i \mid t_i \sim N(m(t_i), s(t_i)^2 + z_i^2)$$

which shows that uncertainty propagates through the calibration curve and into the measurements.

2.2. Bayesian Mixture Models

The mixture models evaluated in this paper seek to reconstruct the calendar age distribution $p(t)$ by conditioning on the mixture parameters $\theta = \{w_k, \mu_k, \sigma_k\}$ (the weights, means, and standard deviations) of $k = 1, ..., K$ clusters or component Gaussians

$$p(t \mid \theta) = \sum_{k=1}^{K} w_k \, N(t \mid \mu_k, \sigma_k^2)$$

where the weights must sum to one, $\sum w_k = 1$. The use of mixture models is interesting here for a number of reasons. First, they are universal approximators, capable of reconstructing any continuous density function provided they have a sufficient number of component distributions, though naturally that comes with the risk of over-fitting (Li & Barron, 1999; McLachlan & Peel, 2000; McLachlan et al., 2019). They are relatively straightforward to grasp, too, especially if one already understands how SPD models are fit, as they both involve summing probability distributions. Finally, the latent clustering assignments are open to a great deal of archaeological interpretation - or over-interpretation as the case may be.

In a Bayesian context, we aim to estimate the mixture parameters θ by sampling from their posterior probability given the observed fraction modern measurements F, that is

$$p(\theta \mid F) \propto p(F \mid \theta) \, p(\theta)$$

with $p(F \mid \theta)$ being the likelihood of the measurements given the mixture parameters, what we refer to as the *mixture likelihood*, and $p(\theta)$ the prior probability of the parameters. To sample from this posterior, we need to (i) define the mixture likelihood $p(F \mid \theta)$ and (ii) specify the priors $p(\theta)$ for the set of weights, means, and errors of the mixture model. We will elaborate on the priors in Heaton (2022) and Price et al. (2021) in the methods section. For now, the likelihood function warrants some attention.

To define the likelihood, we first define the joint probability of an observed fraction modern measurement F_i and its calendar age t_i as

$$p(F_i, t_i \mid \theta) = p(F_i \mid t_i)\, p(t_i \mid \theta)$$

where $p(F_i \mid t_i) = N(F_i \mid m(t_i), s(t_i)^2 + z_i^2)$ is the likelihood of a measurement conditioned on its calendar age, what we refer to as the *measurement likelihood*, and $p(t_i \mid \theta)$ is the global density estimated by the mixture model at t_i. This encapsulates the generative model, weighting the probability of obtaining measurements F_i with ages t_i by the prior probability of those ages. Note that we drop θ from the measurement likelihood because the generative model tells us that the observed measurements depend on the mixture density only through their sampled calendar ages. Provided we know those ages, in other words, knowing the mixture density offers us no additional information about what measurements we should expect. We also assume the measurements are conditionally independent given their ages, ignoring correlation induced by marginalizing over the shared calibration curve.

Because we do not know the true ages, however, we consider all possible ages instead, thus evaluating the mixture likelihood for sample i as

$$p(F_i \mid \theta) = \int p(F_i \mid t)\, p(t \mid \theta)\, dt$$

While the measurement likelihood $p(F_i \mid t)$ is strictly continuous, by convention, it is interpolated over an annually resolved grid. For n radiocarbon samples evaluated over m years, the result is an $n \times m$ measurement matrix $\mathbf{M}$. Each row of this matrix is a likelihood for $i = 1, ..., n$ measurements F_i over the $j = 1, ..., m$ calendar ages t_j. Normalizing the rows of $\mathbf{M}$ with $\widetilde{\mathbf{M}} = \mathrm{diag}(\mathbf{M}\mathbf{1})^{-1}\mathbf{M}$ (or `M/rowSums(M)` in R) transforms the likelihoods into posterior probabilities over

ages assuming a uniform prior, referred to by the authors of the R package `rcarbon` as a `calMatrix`, which is the main output of `rcarbon::calibrate()` (Crema & Bevan, 2021).

Given the conventional discretization of the measurement model, it is convenient to also discretize the global calendar age density $p(t \mid \theta)$ over the same grid of years. The result is a length m vector $\mathbf{v}$. Taking the matrix product of $\mathbf{M}$ and $\mathbf{v}$ gives the length n vector of mixture likelihoods

$$\mathbf{h} = \mathbf{M}\mathbf{v}$$

This is equation 14 in Price et al. (2021). The individual elements of $\mathbf{h}$ represent Riemann sum approximations of the integral of the mixture likelihood for each sample, $\mathrm{h}_i \approx p(F_i \mid \theta)$ with $\Delta t = 1$. Thus, the product of the h_i approximates the total likelihood

$$p(F \mid \theta) \approx \prod_{i=1}^{n} \mathrm{h}_i(\theta)$$

for the mixture model. With the likelihood now defined, it is just a matter of setting priors on the mixture parameters θ to arrive at a complete specification of the posterior.

2.3. Posterior Age Estimation

The posterior age distribution $[t \mid F, \theta]$ of each sample is obtained by evaluating the joint distribution at the observed fraction modern measurements $p(t \mid F, \theta)$, which according to Bayes' Theorem is equal to

$$p(t \mid F, \theta) = \frac{p(F \mid t)\, p(t \mid \theta)}{p(F \mid \theta)}$$

In matrix form, that is

$$\mathbf{q} = \mathrm{diag}(\mathbf{h})^{-1}\mathbf{M}\,\mathrm{diag}(\mathbf{v})$$

and in R (at least one very fast way to do this with recycling) is

```
q <- t(t(M) * v)
q <- q/rowSums(q)
```

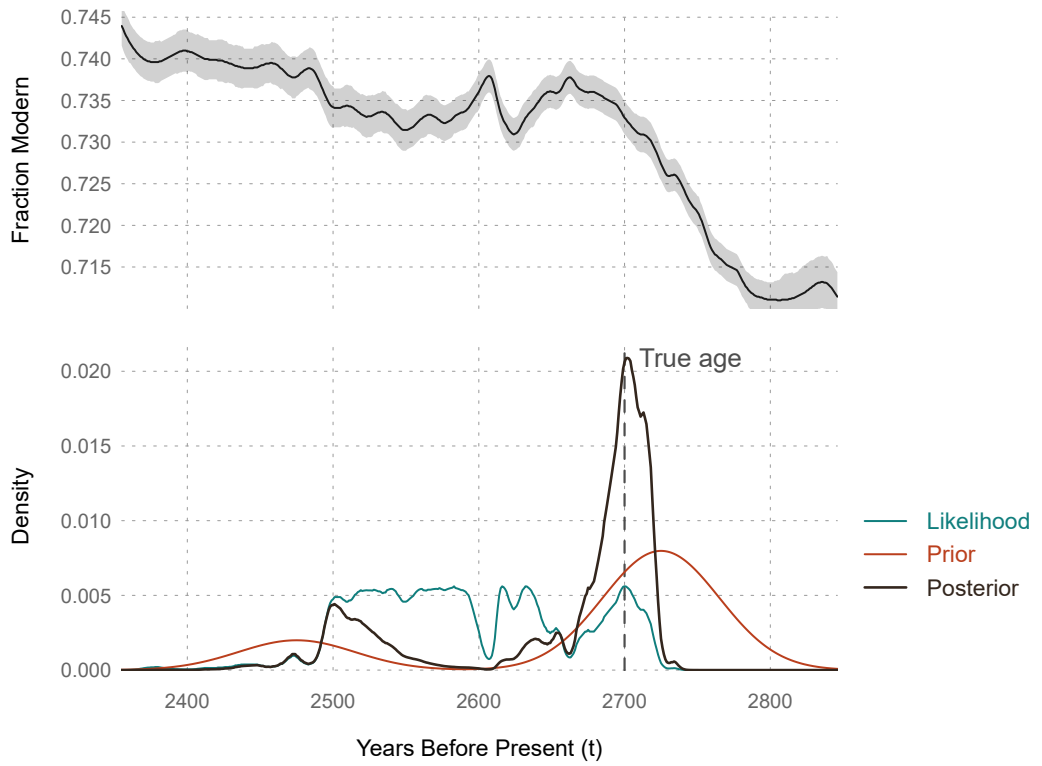

Figure 1: Radiocarbon calibration with measurement likelihood $p(F \mid t)$ weighted by the global density $p(t \mid \theta)$, which acts as an informative prior concentrating posterior probability mass near modes of the global density and also, as a consequence, around the true age of the sample. To avoid over-burdening the y-axis with multiple scales, the calibration curve is plotted above with a gray ribbon for the 95% confidence interval.

In words, we multiply each row of the matrix $\mathbf{M}$ element-wise by the global density $\mathbf{v} \approx p(t \mid \theta)$, and then normalize that quantity by dividing each element by the row sum $\mathrm{h}_i(\theta)$, the mixture likelihood of that sample - in effect, concentrating the posterior probability mass of each sample around the modes of the mixture model, as shown in Figure 1. The resulting matrix $\mathbf{q}$ has the same dimensions as $\mathbf{M}$. Getting a scalar estimate of age for each radiocarbon sample just requires drawing an age from each row of $\mathbf{q}$ using a method like inverse transform sampling (called *quantiles* in R), which for row i is roughly equivalent to `sample(t, size = 1, prob = q[i,])` in R.

3. Methods

For this analysis, we rely on the IntCal20 calibration curve, a Bayesian splines model similar to a Gaussian Process, hence defining a number of correlated Gaussians with means and standard

deviations - $m(t)$ and $s(t)$ - over calendar ages (Heaton et al., 2020; Reimer et al., 2020). The analysis is conducted in the R programming language and environment, v4.5.0 (R Core Team, 2025), with the actual analysis requiring just three packages: `torch` (Falbel & Luraschi, 2026), `carbondate` (Heaton et al., 2024; Heaton, 2022), and `parallel`, the last one shipping with the R installation. Five additional packages support literate programming with Quarto. Code and data can be found at the Zenodo archive and at the GitHub repository *kbvernon/netc14*.

3.1. Simulation

To compare models, we rely on both simulated and empirical data. Each simulation requires that we specify three variables: the number of clusters K, the number of samples n, and the number of years m over which to simulate a discretized calendar age distribution $\mathbf{v}$. Their possible values $K = \{1, 3, 9\}$, $m = \{500, 2500, 12500\}$, and $n = \{50, 250, 1250\}$ were chosen for spread and to achieve a range well beyond what was attempted in previous studies, making tests considerably more challenging. Every combination of these variables is evaluated resulting in 27 unique starting conditions or *scenarios*. For each scenario, a start date t_{min} is randomly sampled from a discrete uniform on the interval from 0 to $(55000 - m)$ years before present (ybp) so that the end date $t_{max} = t_{min} + m - 1$ never exceeds the upper limit on radiocarbon dating. All 27 scenarios are repeated 28 times, each with a unique start date, resulting in 756 total simulations spread out randomly across the viable range of radiocarbon dating.

A single simulation provided with one of those scenarios and starting years consists of two stages. In the first, we randomly generate a calendar age distribution $p(t)$. In the second, we implement the generative model. To simulate $p(t)$, we randomly sample the parameters $\theta = \{w_k, \mu_k, \sigma_k\}$ for a mixture of K Gaussians. The mixture weights w_k are drawn from a Gamma with shape $\alpha = 1$ and rate $\beta = 1$, then normalized to ensure they sum to one. The mixture means μ_k are sampled by partitioning a discrete, annually-resolved uniform on the interval $[t_{min}, t_{max}]$ into K equally-spaced bins and drawing a single mean from each as a way to encourage separation between components and multi-modality in the full mixture distribution. Finally, standard deviations σ_k are drawn from a Gamma with $\alpha = 10$ and $\beta = (10 - 1)/(m/20)$. This ensures that mixture component uncertainty scales with the time scale of the simulation. For example, a simulation covering 500 years would generate distributions with an expected uncertainty of $\alpha/\beta = 27.8$ years. For 2500 and 12500 year intervals, that increases to 138.9 and 694.4 years, respectively. The resulting mixture is then evaluated over a discrete, annually-resolved grid in the interval

$[t_{min}, t_{max}]$ and truncated to remove any probability mass outside of that interval, giving us the length m vector $\mathbf{v}$.

The only remaining quantity to simulate is the measurement uncertainty z_i^2, which we sample first in ^{14}C years from a Gamma with $\alpha = 2$ and $\beta = 1/3$, right-shift by 20 to obtain a mean measurement error of 26-years, then convert to fraction modern using the mean-life equation for radiocarbon. The generative model is then implemented to obtain measurements F_i, which are calibrated in turn with IntCal20 within the m length interval $[t_{min}, t_{max}]$, thus forming the $n \times m$ measurement matrix $\mathbf{M}$, the core data structure used by all models except the DPMM, which takes the uncalibrated fraction modern measurements F and errors z directly.

3.2. Models

3.2.1. Summed Probability Distribution (SPD)

The SPD model works directly on the measurement matrix $\mathbf{M}$, normalizing each row to obtain individual posteriors, then taking the column means. In matrix notation, that is

$$\hat{\mathbf{v}} = \frac{1}{n}\widetilde{\mathbf{M}}^\top \mathbf{1}$$

or

```
v <- colMeans(M/rowSums(M))
```

in R. The result is a noisy estimate $\hat{\mathbf{v}}$ of $\mathbf{v}$. Drawing scalar age estimates $\hat{t}_1, \ldots, \hat{t}_n$ with SPD involves sampling from the posteriors in the rows of $\widetilde{\mathbf{M}}$.

3.2.2. Composite Kernel Density Estimation (CKDE)

There are many versions of KDE for radiocarbon dating (Bronk Ramsey, 2017; McLaughlin, 2019). The version tested here is similar to the one implemented in `rcarbon` and described in Brown (2017). It is a composite KDE in the sense that it averages multiple KDE fit to different sets of samples drawn from the rows of $\widetilde{\mathbf{M}}$. KDE is then applied to those samples and recorded as a length m vector. This is iterated n_{iter} times to construct an $n_{iter} \times m$ matrix, with the column means of that matrix giving the estimate $\hat{\mathbf{v}}$. Quantiles can be estimated as well to get confidence intervals. For age estimation, we use the Bayesian posterior age sampling approach described above,

multiplying $\hat{\mathbf{v}}$ as a prior through the rows of $\mathbf{M}$ to obtain posterior age estimates, then inverse transform sampling from those posteriors to get $\hat{t}_1, ..., \hat{t}_n$, so this approach does some kind of information sharing across samples, though it is *ad hoc* at best. Note that we use R's defaults for KDE, meaning the density is estimated using a Gaussian kernel with a bandwidth determined by Silverman's Rule.

3.2.3. Bayesian Gaussian Mixture Model (BGMM)

We implement a posterior for the parametric BGMM in a torch module that is similar but not identical to the model described by Price et al. (2021). It includes a few customizations to make the model more flexible and performant. For instance, we choose the number of clusters K by fitting multiple k-means models with different K and choosing the one that minimizes the BIC. The weights and errors are then initialized as $1/K$ and m/K, respectively.

In our experiments, the k-means strategy tended to exaggerate the number of components, so we set the concentration parameter for the Dirichlet prior on the mixture weights to 0.6, which encourages sparsity. We also set the Gamma distribution as a prior on the mixture errors as in Price et al. (2021), but rather than code the shape and rate directly, they are determined by constraining the mean of the Gamma to be equal to the ratio of years to clusters, or $\alpha/\beta = m/K$. The coefficient of variation is then used to determine how much spread around the mean error there can be. The default in this case is 0.6, making it a weak prior. As in Price et al. (2021), the prior on the mixture means is flat, and an ordering constraint is imposed to prevent label switching.

Sampling from the posterior is conducted using an adaptive dual averaging Hamiltonian Monte Carlo (Hoffman & Gelman, 2014) implemented as a torch module. The number of iterations is 1000 by default, resulting in 1000 samples of each of the mixture parameters, with half of the samples removed as burn-in. The estimate $\hat{\mathbf{v}}$ is obtained by constructing mixture models for each Hamiltonian sample s of mixture parameters $\theta^{(s)}$, evaluating over the year grid in $[t_{min}, t_{max}]$, and then taking the mean of those density estimates, $\hat{\mathbf{v}} = 1/n_s \sum \hat{\mathbf{v}}^{(s)}$. Confidence intervals can be constructed in the same way. Age estimates $\hat{t}_1, ..., \hat{t}_n$ are obtained using the posterior age estimation approach described above.

3.2.4. Dirichlet Process Mixture Model (DPMM)

The non-parametric DPMM proposed by Heaton (2022) learns the number of mixtures K using a Gibbs sampler that allows for sampling sequentially from conditional distributions. The core of this

approach is a latent clustering variable c_i that holds the cluster assignment of each measurement $i = 1, ..., n$. The c_i are drawn using Polya Urn sampling (Neal, 2000)

$$c_i \mid t_i, \xi_k, n_k \sim \text{Categorical}(r_{i1}, ..., r_{iK+1}), \qquad r_{ik} \propto \begin{cases} n_k \, N(t_i \mid \mu_k, \tau_k^{-1}) & \text{existing cluster } k \\ \alpha_d \, T(t_i \mid \mu_\phi, \tau_\phi, 2\alpha) & \text{new cluster } K+1 \end{cases}$$

with $\xi_k = \{\mu_k, \tau_k\}$ being the mean and precision ($\tau = 1/\sigma^2$) of cluster k, n_k the size of k excluding i, r_{ik} the probability of sample i being assigned to k, α_d the concentration parameter of the Dirichlet Process, T the location-scale student's t-distribution with 2α degrees of freedom, μ_ϕ a hyper-parameter specifying the base mean for clusters, and τ_ϕ the base precision derived from the Gamma. So, the Polya Urn assigns a ^{14}C sample to an existing cluster k with probability equal to its size n_k weighted by the likelihood of its age given that *specific* cluster and to a new (currently empty) cluster with probability equal to the concentration α_d weighted by the likelihood of its age given an *average* cluster. Because assignment is redone in each MCMC iteration, there is also the potential for zeroing out clusters, so the number of clusters can shrink or grow.

The Normal-Gamma serves as the joint prior for the means and precisions of the mixture distribution

$$\xi_k \mid \mu_\phi \sim \text{NormalGamma}(\mu_\phi, \lambda, \alpha, \beta)$$

with λ a weight on the component errors. The shape and rate parameters of the Gamma are updated each iteration based on the previous iteration's calendar age estimates $\hat{t}_i$ and the current assignments c_i. As the number of radiocarbon samples assigned to a cluster goes up, the scale parameter increases, and with it the expected precision. Conversely, when the spread of the calendar ages increases, the rate parameter also goes up, thereby decreasing the expected precision. The precision is also penalized when the mean of the cluster drifts from μ_ϕ.

Given the hard cluster assignments c_i, the conditional for the calendar ages can use the Gaussians of the clusters as priors for age estimation, rather than the full mixture density

$$t_i \mid c_i = k, F_i, \xi_k \sim g_{ik}, \qquad g_{ik}(t) \propto p(F_i \mid t) \, N(t \mid \mu_k, \tau_k^{-1})$$

Draws from this distribution are obtained using slice sampling (Neal, 2003). So, the Gibbs sampler assigns a scalar age estimate to each ^{14}C during each MCMC iteration. Because cluster assignment

happens during each MCMC iteration, too, radiocarbon samples will wander across clusters in proportion to their likelihood. The individual age estimates from each iteration are thus equivalent to draws from each row of $\mathbf{q}$.

Gibbs sampling from these conditionals is implemented in performant C++ code in the R package `carbondate`. The default number of MCMC iterations is 100,000, but that proved impractical for the number of simulations given the resources of a typical desktop computer, so we limit sampling to 10,000 iterations. The final density estimate is obtained by taking the mean of a subset s of the individual Gibbs samples, $\hat{\mathbf{v}} = 1/n_s \sum \hat{\mathbf{v}}^{(s)}$. Age estimates are retained as draws from the posterior age distribution $[\hat{t} \mid F, \theta]$ or $\mathbf{q}$ for each radiocarbon date, the per iteration sampled ages.

3.2.5. Gaussian Mixture Neural Network (NetC14)

NetC14 estimates the mixture parameters θ that maximize the mixture likelihood $p(F \mid \theta)$, so one can think of it as an instance of regularized Maximum Likelihood. The main innovation in NetC14 is that it learns the number of mixture components by applying the SparseMax activation to the logits of the mixture weights w_k of an artificially inflated number of components, which pushes many weights to zero without compromising back-propagation. The activation works by centering the logits around a threshold and then applying ReLU, with the threshold determined so that positive centered logits sum to one and the rest are set to zero (Martins & Astudillo, 2016). Pruning components in this way rather than adding them satisfies the requirement imposed by neural networks that the dimensions of the data be known in advance.

The model is encapsulated in a single shallow torch module initialized with $K_0 = \min(\text{round}(\sqrt{m}), K_{max})$ clusters, where K_{max} is the maximum number of allowed clusters, set to 35 by default. Three $1 \times K_0$ vectors hold the log-transformed mixture parameters, with initial values set in the same manner as the BGMM. As a crude approximation to Bayesian uncertainty, a small amount of Gaussian noise is added during model training to the logits before constraining them to parameter space. That includes the same ordering constraint on means as in the BGMM. The logits of the errors are exponentiated, and the SparseMax is applied to the logits of the weights with a temperature π controlling how aggressively the activation zeroes them out. The forward pass then constructs the mixture distribution and evaluates the mixture over the annually-resolved grid interval $[t_{min}, t_{max}]$, returning an estimate of the density $\hat{\mathbf{v}}$.

A custom loss function that computes $\log p(F \mid \theta)$ is implemented in a separate torch module. The loss embeds the measurement matrix $\mathbf{M}$ and takes $\hat{\mathbf{v}}$ as input, allowing us to compute $\sum \log \mathrm{h}_i$ efficiently. To prevent component singularities, the loss includes a penalty that increases with the precision according to the function $\gamma \cdot nm \sum \sigma_k^{-2}$ for n radiocarbon samples and m years, with γ a weighting parameter that controls the size of the penalty. Tuning was done through a simple grid search over a range of values of π and γ using the evaluation method described in the next section, resulting in default values $\pi = 1.75$ and $\gamma = 2.5$. The model is trained over 512 epochs using the Adam optimizer with a learning rate of 0.005 (Kingma & Ba, 2017). Density estimates $\hat{\mathbf{v}}$ are obtained by setting the model to training mode and running multiple forward passes, generating unique estimates $\hat{\mathbf{v}}^{(s)}$ with Gaussian noise that are then averaged, $\hat{\mathbf{v}} = 1/n_s \sum \hat{\mathbf{v}}^{(s)}$, with the potential to calculate confidence intervals. Posterior age estimation is conducted in the manner described above.

3.3. Evaluation

We evaluate the performance of the models at calibrating and summarizing radiocarbon dates, the two goals of radiocarbon dating. Following Heaton (2022), we measure age estimation performance using a Monte Carlo estimate of the mean absolute error

$$\mathrm{MAE} = \frac{1}{n} \sum_{i=1}^{n} \mathbb{E}\big|[\hat{t}_i \mid F_i] - t_i\big|$$

with $\hat{t}_i \mid F_i \sim p(t_i \mid F_i, \theta)$ (suppressing θ) being the posterior age distribution of the ith sample (q_i in matrix notation). Technically, this involves drawing 1,000 scalar samples from $\hat{t}_i \mid F_i$ for each sample, comparing each to the true age t_i, and then taking the mean of those $1000 \cdot n$ comparisons. For interpretability, we define an oracle model that knows the true age distribution $p(t)$ perfectly, and have it do posterior age estimation with the measurement matrix $\mathbf{M}$. The idea here is to define a theoretical floor MAE_o on attainable error in age estimation in the face of measurement and calibration curve uncertainty while controlling for all other sources of uncertainty. The exact comparison is

$$\mathrm{Recovery} = \frac{\mathrm{MAE} - \mathrm{MAE}_o}{\mathrm{MAE}_o}$$

showing the model's proportional excess error over the oracle, with zero indicating parity.

To measure density estimation performance, we calculate total variation distance for the discrete evaluation of the density $\mathbf{v} \approx p(t)$ and its estimate $\hat{\mathbf{v}}$ over the age grid, which is just one-half the sum of the absolute differences at each year $j = 1, ..., m$ for the normalized distributions, or

$$\text{TVD} = 0.5 \sum_{j=1}^{m} |v_j - \hat{v}_j|$$

Both metrics were recorded for all models over all 756 simulations. Some crude benchmarks are conducted as well using the same simulation architecture to evaluate compute times across different sample sizes n ranging from 500 to 5000, holding the other simulation parameters constant at $m = 3000$ years and $K = 6$.

3.4. Archaeological Application

We apply all models to two real world datasets used in Heaton (2022) and Bronk Ramsey (2017), including dates from Iron Age Ireland (Armit et al., 2014) and Paleoindian North America (Buchanan et al., 2008). Armit et al. (2014) use a substantial radiocarbon dataset ($n = 2023$) from Ireland to reconstruct the population history of the island during the transition from the Bronze Age to the early Iron Age, from roughly 3200 to 2500 ybp. Comparing their population reconstruction to climate reconstructions over the same interval allows them to evaluate the hypothesis that a climate downturn led to population decline leading into the Iron Age. Buchanan et al. (2008) use a smaller sample ($n = 628$) to reconstruct the population history of Paleoindians in North America at the onset of a prolonged drying period known as the Younger Dryas, from 15000 to 9000 ybp. The population reconstruction allows them to evaluate the hypothesis that large low-density extraterrestrial objects impacted North America around 12900 ybp, causing the Younger Dryas and subsequent population decline among North American megafauna and Paleoindians (Firestone et al., 2007). Both papers apply SPD to the calibrated dates matrix $\mathbf{M}$ to estimate $p(t)$, using it as a proxy for population size over time. For comparative purposes, we apply all models to these data and evaluate differences graphically.

4. Results

General results are reported in Table 1. Figure 2 and Figure 3 show error metrics for models fit to simulated fraction modern data across both number of samples and number of clusters in each simulation. The three mixture models are roughly equivalent, approaching the origin as sample

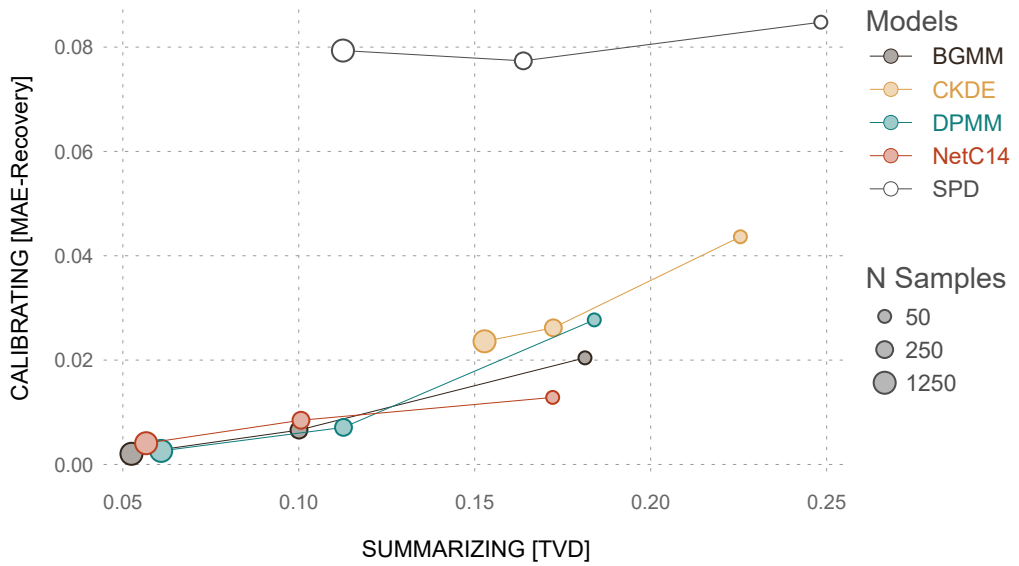

Figure 2: Evaluation results by model and sample size. Coordinates are median values across both dimensions.

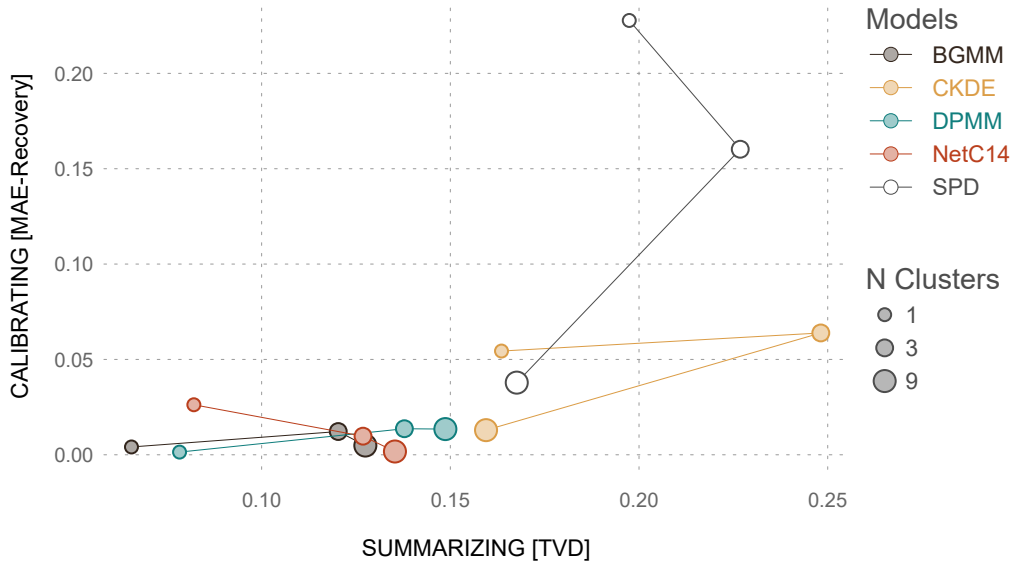

Figure 3: Evaluation results by model and number of clusters. Coordinates are median values across both dimensions.

size increases, though there is some notable spread in recovery for small samples. Figure 3 shows that as the number of clusters in the simulation increases, the ability of the mixtures to reconstruct the true density diminishes, but performance in age estimation remains constant. Note that the models are largely stable across age estimation and that most of the improvement happens along the axis of density estimation. The one exception is SPD for age estimation across number of

Table 1: Evaluation results across all models, sample sizes, and number of clusters. Parentheses show 95% confidence intervals.

Model	MAE-Recovery	MAE	TVD	Elapsed (s)
BGMM	0.007 (-0.067, 0.540)	86.457 (25.738, 222.719)	0.105 (0.013, 0.508)	49.313 (27.601, 91.853)
CKDE	0.029 (-0.024, 1.514)	90.909 (33.150, 242.602)	0.185 (0.038, 0.537)	1.451 (1.133, 1.932)
DPMM	0.009 (-0.078, 0.800)	87.528 (23.985, 223.569)	0.122 (0.010, 0.506)	71.463 (11.639, 141.552)
NetC14	0.007 (-0.042, 0.636)	86.348 (29.871, 215.878)	0.111 (0.019, 0.411)	8.496 (6.858, 12.721)
SPD	0.080 (-0.004, 2.290)	97.217 (39.701, 260.636)	0.190 (0.053, 0.617)	0.040 (0.007, 0.089)

clusters, which gets better at age estimation as the number of clusters grow. Nevertheless, SPD performs considerably worse than the other methods on both dimensions.

While the mixture models are more or less equally competitive in age and density estimation, compute times vary substantially, as shown in Figure 4. The SPD is clearly fastest, followed closely by CKDE, with median compute times of 0.04 and 1.45 respectively for $m = 3000$ and $K = 6$, with n ranging between 500 and 5000. The next fastest is NetC14 at a median compute time of 8.496 s, followed by the torch implemented BGMM at 49.313 s and DPMM at 71.463 s. BGMM and NetC14 also appear to scale much better than the DPMM over different sample sizes. Finally, applications to real world datasets show general agreement across models regarding the shape of the age distribution that the samples are drawn from, as shown in Figure 5 and Figure 6. The main exceptions are, one, that the Bayesian mixture models place probability spikes at several places and, two, that the CKDE tends to smooth out the distribution.

5. Discussion

In simulations, NetC14 achieves parity with existing mixture models, but interesting differences emerge when models are applied to real world datasets. Note that both Bayesian mixture models place a large probability spike at 1,550 ybp when applied to the Iron Age Ireland dataset in Figure 5,

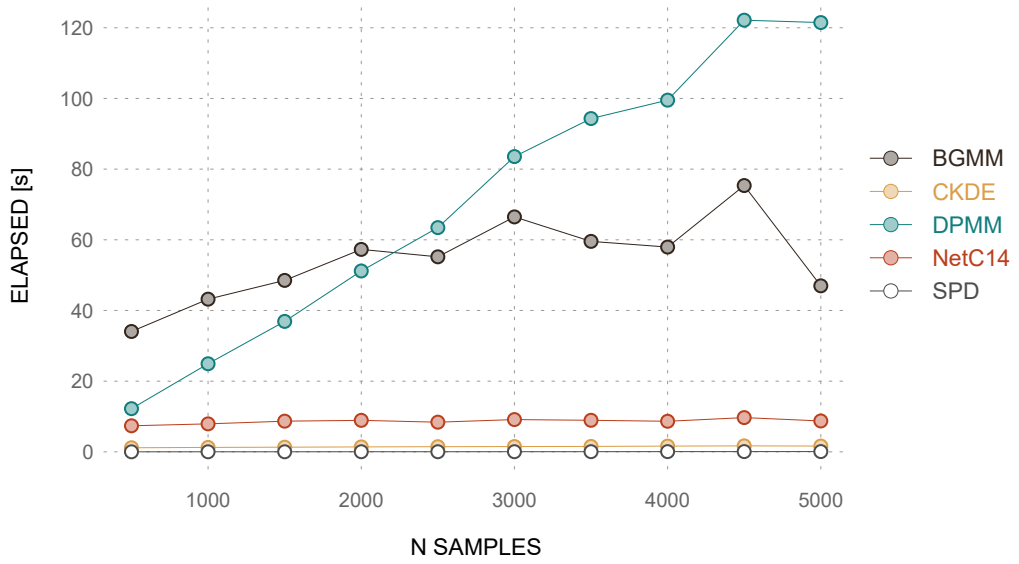

Figure 4: Median elapsed time for model training across different sample sizes.

but the other models do not. The DPMM also places spikes in the Paleoindian North America dataset at around 9,100 and 9,500 ybp in Figure 6. NetC14, on the other hand, appears to smooth the SPD, as one might expect given that it is a ML method that effectively tries to minimize distance from SPD estimates, while CKDE exhibits a familiar spread, smearing the probability over a wider range of years (Bronk Ramsey, 2017). The probability spikes are almost certainly owing to the hyperparameter settings for the Gamma prior on the uncertainty or precision. Particularly for the DPMM, both the number of samples assigned to a cluster and their concentration around the mean of the cluster - their likelihood - can affect the precision.

While all three mixture models perform more or less equally well at age and density estimation, they differ in compute efficiency, with NetC14 being much faster. Owing to its use of ML, it only needs to compute the log posterior once per iteration or epoch. The BGMM has to calculate the gradient each jump of the leapfrog integrator (12 jumps by default) during each MCMC iteration, and the DPMM relies on a heavy slice sampling loop (Neal, 2003) over the posterior ages that continues indefinitely until a calendar age proposal is accepted for each sample and MCMC iteration. It is worth noting, too, that NetC14 was severely handicapped in the benchmarks and simulations, not taking advantage of torch's parallelism or its GPU acceleration. Still, we should not put too much weight on these simple benchmarks. They are only meant to be suggestive, not definitive. More thorough testing would be required for that.

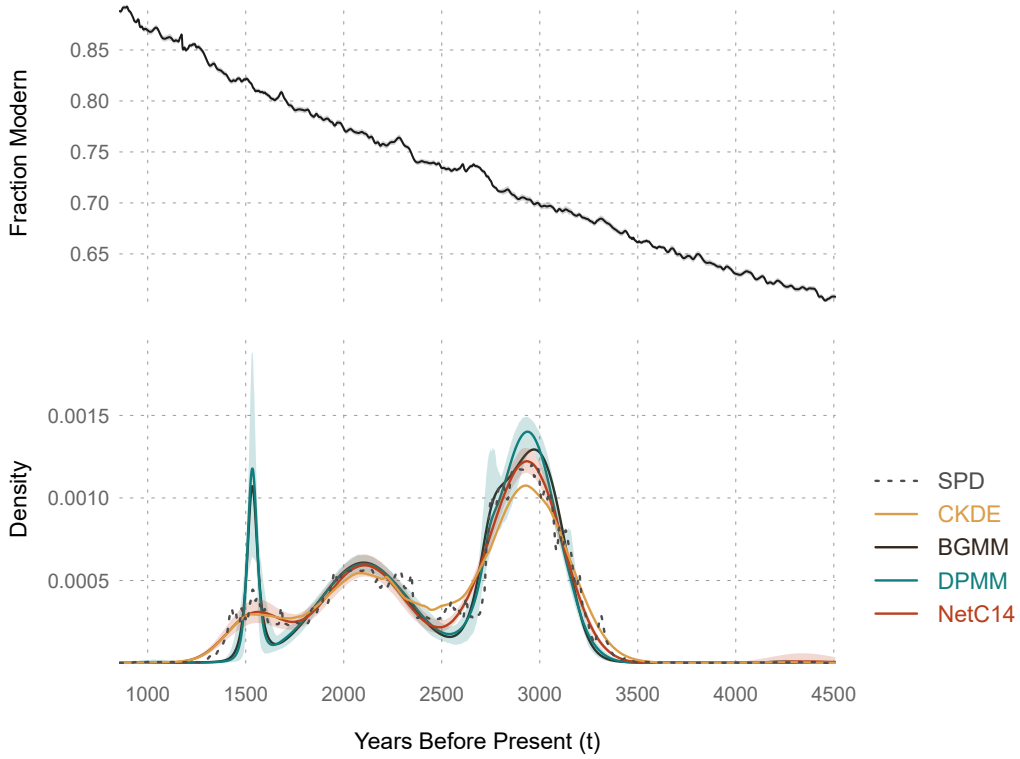

Figure 5: Iron Age Ireland. Methods applied to radiocarbon dates from Armit et al. (2014).

A more compelling difference between NetC14 and the Bayesian mixture models lies in its currently untapped potential as a neural network. Of all the models, it has the greatest capacity to condition the mixture parameters on additional diagnostic data $p(\theta \mid F, x)$, with the mixture parameters becoming functions of that data $\theta = \{w_k(x), \mu_k(x), \sigma_k(x)\}$ such that mixture parameters can vary across the radiocarbon samples $i = 1, ..., n$. This is known as a Mixture Density Network (MDN) (Bishop, 1994). As a neural network, the MDN can ingest all manner of data, including familiar matrices of ceramic or projectile point assemblages, but also higher dimensional arrays like imagery, even Lidar over radiocarbon sample contexts. Letting mixture parameters vary over radiocarbon samples in this way would allow the model to handle multi-modality in the calibration curve directly. Indeed, that was the motivation for the development of MDN in the first place (Bishop, 1994). The model would then have the ability to output artifact production profiles that have long been of interest to archaeologists (Ford, 1938; Lipo et al., 2015). Incorporating such data would also allow for indirect dating, meaning the model could make predictions about occupation sequences for sites and regions that lack direct dates but have diagnostics available (Crema, 2025; Ortman, 2016). Finally, incorporating non-diagnostic but

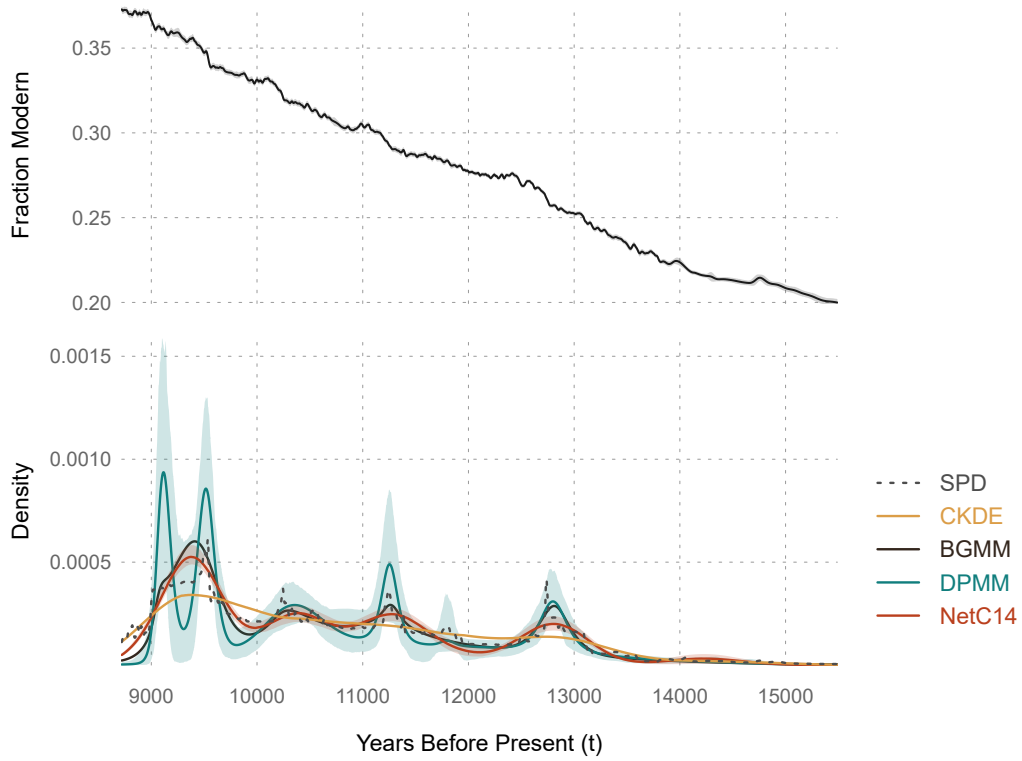

Figure 6: Paleoindian North America. Methods applied to radiocarbon dates from Buchanan et al. (2008).

presumably causal variables would allow for testing archaeological hypotheses regarding population change over time in a single model.

All of the methods evaluated here assume that radiocarbon dates are representative of the true age distribution, and in our simulations they are, but in reality, all sorts of intervening forces may bias the sample (Attenbrow & Hiscock, 2015; Contreras & Codding, 2024; Contreras & Meadows, 2014; Crema, 2022; Rick, 1987). From the moment an organic sample is deposited, it begins to break down, but the rate at which it does so and its potential for surviving into the future varies based on features of the organic sample itself and its environment (Contreras & Codding, 2024; Surovell et al., 2009). And even if the organic sample did manage to preserve to the present, its chances of being sampled are not random because dating is costly and thus biased toward the scientific interests of well-funded researchers. A region's ability to attract or sustain more archaeologists would also presumably tend to generate larger samples (Codding et al., 2024). Some techniques have already been developed to address some of these issues, like the Surovell correction for taphonomic decay (Surovell et al., 2009) and stratified sampling for research intensity (like

`rcarbon::thinDates`), both of which could easily be integrated into NetC14, for instance with stratified mini-batch sampling. Research into debiasing techniques for neural networks is also quite vast (Wang & Liu, 2023) and offers new potentially more powerful means of handling this persistent challenge.

Archaeologists have also long recognized that population reconstructions based on radiocarbon dates often imply demographic rates that are implausible (Brown, 2015; Contreras & Meadows, 2014). Because the estimated density is free to take any shape the data suggest, reconstructions can exhibit growth or decline over short intervals far exceeding what human populations are capable of sustaining, particularly where the calibration curve introduces artificial peaks and troughs. Bayesian methods exist to constrain these reconstructions with explicit growth models (Crema & Shoda, 2021; Wilson et al., 2026), but they inherit all the limitations of the MCMC machinery they depend on, chief among them the computational cost that makes repeated model fitting and sensitivity analysis impractical. Neural networks offer a more tractable route to the same end, as demographic expectations can be imposed directly through the loss function rather than through the prior. Models that incorporate theoretical constraints in this manner are known as physics-informed neural networks (Raissi et al., 2019). They represent an exceptionally promising avenue of research for radiocarbon dating, one that NetC14 and other models like it can easily embrace.

6. Conclusion

Here we systematically compare the major extant methods and introduce NetC14, a novel application of neural networks using regularized Maximum Likelihood to simultaneously estimate the age of an archaeological event and the distribution of many dated events. The results show that NetC14 performs as well as recent advances that rely on more cumbersome and computationally-intensive Bayesian methods. The computational advantages of NetC14, plus the promise to expand the application to include additional covariates allowing to address archaeological questions, should improve the uptake of these approaches while also expanding the potential to explain past human behavior.

Bibliography

Armit, I., Swindles, G. T., Becker, K., Plunkett, G., & Blaauw, M. (2014). Rapid climate change did not cause population collapse at the end of the European Bronze Age. *Proceedings of the National Academy of Sciences*, *111*(48), 17045–17049. https://doi.org/10.1073/pnas.1408028111

Attenbrow, V., & Hiscock, P. (2015). Dates and demography: Are radiometric dates a robust proxy for long-term prehistoric demographic change?. *Archaeology in Oceania*, *50*(S1), 30–36. https://doi.org/10.1002/arco.5052

Bayliss, A. (2009). Rolling out revolution: Using radiocarbon dating in archaeology. *Radiocarbon*, *51*(1), 123–147. https://doi.org/10.1017/S0033822200033750

Bishop, C. M. (1994). *Mixture Density Networks* (Technical Report NCRG/94/004; pp. 1–25). Aston University Neural Computing Research Group. https://publications.aston.ac.uk/id/eprint/373/1/NCRG_94_004.pdf

Blockley, S. P. E., Donahue, R. E., & Pollard, A. M. (2000). Radiocarbon calibration and Late Glacial occupation in northwest Europe. *Antiquity*, *74*(283), 112–119. https://doi.org/10.1017/S0003598X00066199

Bronk Ramsey, C. (2008). Radiocarbon dating: Revolutions in understanding. *Archaeometry*, *50*(2), 249–275. https://doi.org/10.1111/j.1475-4754.2008.00394.x

Bronk Ramsey, C. (2017). Methods for Summarizing Radiocarbon Datasets. *Radiocarbon*, *59*(6), 1809–1833. https://doi.org/10.1017/RDC.2017.108

Brown, W. A. (2015). Through a filter, darkly: population size estimation, systematic error, and random error in radiocarbon-supported demographic temporal frequency analysis. *Journal of Archaeological Science*, *53*, 133–147. https://doi.org/10.1016/j.jas.2014.10.013

Brown, W. A. (2017). The past and future of growth rate estimation in demographic temporal frequency analysis: Biodemographic interpretability and the ascendance of dynamic growth models. *Journal of Archaeological Science*, *80*, 96–108. https://doi.org/10.1016/j.jas.2017.02.003

Buchanan, B., Collard, M., & Edinborough, K. (2008). Paleoindian demography and the extraterrestrial impact hypothesis. *Proceedings of the National Academy of Sciences*, *105*(33), 11651–11654. https://doi.org/10.1073/pnas.0803762105

Carleton, W. C., & Groucutt, H. S. (2021). Sum things are not what they seem: Problems with point-wise interpretations and quantitative analyses of proxies based on aggregated radiocarbon dates. *The Holocene*, *31*(4), 630–643.

Codding, B. F., Meyer, J., Brewer, S. C., Kelly, R. L., & Jones, T. L. (2024). Do counts of radiocarbon-dated archaeological sites reflect human population density? A preliminary empirical validation examining spatial variation across late Holocene California. *Radiocarbon*, *66*(4), 610–623. https://doi.org/10.1017/RDC.2024.81

Contreras, D. A., & Codding, B. F. (2024). Landscape taphonomy predictably complicates demographic reconstruction. *Journal of Archaeological Method and Theory*, *31*(3), 1102–1128. https://doi.org/10.1007/s10816-023-09634-5

Contreras, D. A., & Meadows, J. (2014). Summed radiocarbon calibrations as a population proxy: A critical evaluation using a realistic simulation approach. *Journal of Archaeological Science*, *52*, 591–608. https://doi.org/10.1016/j.jas.2014.05.030

Crema, E. R. (2022). Statistical inference of prehistoric demography from frequency distributions of radiocarbon dates: A review and a guide for the perplexed. *Journal of Archaeological Method and Theory*, *29*(4), 1387–1418. https://doi.org/10.1007/s10816-022-09559-5

Crema, E. R., & Bevan, A. (2021). Inference from large sets of radiocarbon dates: Software and methods. *Radiocarbon*, *63*(1), 23–39. https://doi.org/10.1017/RDC.2020.95

Crema, E. R. (2025). A Bayesian alternative for aoristic analyses in archaeology. *Archaeometry*, *67*(S1), 7–30. https://doi.org/10.1111/arcm.12984

Crema, E. R., & Shoda, S. (2021). A Bayesian approach for fitting and comparing demographic growth models of radiocarbon dates: A case study on the Jomon-Yayoi transition in Kyushu (Japan). *PLOS ONE*, *16*(5), e0251695. https://doi.org/10.1371/journal.pone.0251695

Falbel, D., & Luraschi, J. (2026). *torch: Tensors and neural networks with 'GPU' acceleration* [Technical report]. https://torch.mlverse.org/docs

Firestone, R. B., West, A., Kennett, J. P., Becker, L., Bunch, T. E., Revay, Z. S., Schultz, P. H., Belgya, T., Kennett, D. J., Erlandson, J. M., Dickenson, O. J., Goodyear, A. C., Harris, R. S., Howard, G. A., Kloosterman, J. B., Lechler, P., Mayewski, P. A., Montgomery, J., Poreda, R., … Wolbach, W. S. (2007). Evidence for an extraterrestrial impact 12,900 years ago that contributed to the megafaunal extinctions and the Younger Dryas cooling. *Proceedings of the National Academy of Sciences*, *104*(41), 16016–16021. https://doi.org/10.1073/pnas.0706977104

Ford, J. A. (1938). A chronological method applicable to the Southeast. *American Antiquity*, *3*(3), 260–264. https://doi.org/10.2307/275264

Hajdas, I., Ascough, P., Garnett, M. H., Fallon, S. J., Pearson, C. L., Quarta, G., Spalding, K. L., Yamaguchi, H., & Yoneda, M. (2021). Radiocarbon dating. *Nature Reviews Methods Primers*, *1*(1), 62. https://doi.org/10.1038/s43586-021-00058-7

Heaton, T. J., Bard, E., Bayliss, A., Blaauw, M., Bronk Ramsey, C., Reimer, P. J., Turney, C. S. M., & Usoskin, I. (2024). Extreme solar storms and the quest for exact dating with radiocarbon. *Nature*, *633*(8029), 306–317. https://doi.org/10.1038/s41586-024-07679-4

Heaton, T. J., Blaauw, M., Blackwell, P. G., Bronk Ramsey, C., Reimer, P. J., & Scott, E. M. (2020). The IntCal20 approach to radiocarbon calibration curve construction: A new methodology using Bayesian splines and errors-in-variables. *Radiocarbon*, *62*(4), 821–863. https://doi.org/10.1017/RDC.2020.46

Heaton, T. J. (2022). Non-parametric calibration of multiple related radiocarbon determinations and their calendar age summarisation. *Journal of the Royal Statistical Society Series C: Applied Statistics*, *71*(5), 1918–1956. https://doi.org/10.1111/rssc.12599

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research*, *15*(1), 1593–1623.

Kingma, D. P., & Ba, J. (2017). *Adam: A method for stochastic optimization*. https://arxiv.org/abs/1412.6980

Li, J., & Barron, A. (1999). *Mixture density estimation* (S. Solla, T. Leen, & K. Müller, Eds.; Vol. 12). MIT Press. https://proceedings.neurips.cc/paper_files/paper/1999/file/a0f3601dc682036423013a5d965db9aa-Paper.pdf

Libby, W. F. (1946). Atmospheric helium three and radiocarbon from cosmic radiation. *Physical Review*, *69*(11–12), 671–672. https://doi.org/10.1103/PhysRev.69.671.2

Lipo, C. P., Madsen, M. E., & Dunnell, R. C. (2015). A theoretically-sufficient and computationally-practical technique for deterministic frequency seriation. *PLOS ONE*, *10*(4), e0124942. https://doi.org/10.1371/journal.pone.0124942

Martins, A., & Astudillo, R. (2016). From Softmax to Sparsemax: A Sparse Model of Attention and Multi-Label Classification. In M. F. Balcan & K. Q. Weinberger (Eds.), *Proceedings of The 33rd International Conference on Machine Learning: Vol. 48. Proceedings of The 33rd International Conference on Machine Learning*.

McLachlan, G. J., & Peel, D. (2000). *Finite Mixture Models*. Wiley.

McLachlan, G. J., Lee, S. X., & Rathnayake, S. I. (2019). Finite mixture models. *Annual Review of Statistics and Its Application*, *6*(Volume 6, 2019), 355–378. https://doi.org/https://doi.org/10.1146/annurev-statistics-031017-100325

McLaughlin, T. R. (2019). On applications of space–time modelling with open-source 14C age calibration. *Journal of Archaeological Method and Theory*, *26*(2), 479–501. https://doi.org/10.1007/s10816-018-9381-3

McNichol, A. P., Jull, A. J. T., & Burr, G. S. (2001). Converting AMS data to radiocarbon values: Considerations and conventions. *Radiocarbon*, *43*(2A), 313–320. https://doi.org/10.1017/S0033822200038169

Neal, R. M. (2000). Markov chain sampling methods for Dirichlet process mixture models. *Journal of Computational and Graphical Statistics*, *9*(2), 249–265. https://doi.org/10.1080/10618600.2000.10474879

Neal, R. M. (2003). Slice sampling. *The Annals of Statistics*, *31*(3), 705–767. https://doi.org/10.1214/aos/1056562461

Ortman, S. G. (2016). Uniform Probability Density Analysis and Population History in the Northern Rio Grande. *Journal of Archaeological Method and Theory*, *23*(1), 95–126. https://doi.org/10.1007/s10816-014-9227-6

Price, M. H., Capriles, J. M., Hoggarth, J. A., Bocinsky, R. K., Ebert, C. E., & Jones, J. H. (2021). End-to-end Bayesian analysis for summarizing sets of radiocarbon dates. *Journal of Archaeological Science*, *135*, 105473. https://doi.org/10.1016/j.jas.2021.105473

R Core Team. (2025). *R: A Language and Environment for Statistical Computing*. https://www.r-project.org/

Raissi, M., Perdikaris, P., & Karniadakis, G. (2019). Physics-informed neural networks: A deep learning framework for solving forward and inverse problems involving nonlinear partial differential equations. *Journal of Computational Physics*, *378*, 686–707. https://doi.org/10.1016/j.jcp.2018.10.045

Reimer, P. J., Austin, W. E. N., Bard, E., Bayliss, A., Blackwell, P. G., Bronk Ramsey, C., Butzin, M., Cheng, H., Edwards, R. L., Friedrich, M., Grootes, P. M., Guilderson, T. P., Hajdas, I., Heaton, T. J., Hogg, A. G., Hughen, K. A., Kromer, B., Manning, S. W., Muscheler, R., … Talamo, S. (2020). The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon*, *62*(4), 725–757. https://doi.org/10.1017/RDC.2020.41

Rick, J. W. (1987). Dates as data: An examination of the Peruvian preceramic radiocarbon record. *American Antiquity*, *52*(1), 55–73. https://doi.org/10.2307/281060

Surovell, T. A., Byrd Finley, J., Smith, G. M., Brantingham, P. J., & Kelly, R. (2009). Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science*, *36*(8), 1715–1724. https://doi.org/10.1016/j.jas.2009.03.029

Taylor, R. E., & Bar-Yosef, O. (2016). *Radiocarbon dating: An archaeological perspective* (2nd ed.). Routledge.

Wang, Y., & Liu, H. (2023). *2023 International Conference on Machine Learning and Cybernetics (ICMLC)*. 458–463. https://doi.org/10.1109/ICMLC58545.2023.10327985

Wilson, K. M., Codding, B. F., McCool, W. C., Medina, I., Vernon, K. B., & Brewer, S. C. (2026). The impact of subsistence change on demography: Theory-informed Approximate Bayesian Computation reveals 4-fold increase in carrying capacity following economic intensification. *Journal of Archaeological Science*, *189*, 106523. https://doi.org/10.1016/j.jas.2026.106523

Wood, R. (2015). From revolution to convention: The past, present and future of radiocarbon dating. *Scoping the Future of Archaeological Science: Papers in Honour of Richard Klein*, *56*, 61–72. https://doi.org/10.1016/j.jas.2015.02.019