# Analysis of the Abundance of Radiocarbon Samples as Count Data

Miguel de Navascués[1,2], Concetta Burgarella[2,3], and Mattias Jakobsson[2]

https://doi.org/10.5281/zenodo.13381596

## Abstract

The analysis of the abundance of radiocarbon samples through time has become a popular method to address questions of demography in archaeology. The history of this approach is marked by the use of the Sum of Probability Distributions (SPD), a key methodological development that first allowed researchers to visualize the abundance of radiocarbon samples on a calibrated temporal scale. However, the lack of a mathematical definition hinders the use of SPD in a proper statistical framework. Recent developments of model-based approaches have allowed a more rigorous statistical analysis of the abundance of radiocarbon data. Despite these advances, these methods inherit from the SPD an interpretation of the abundance of samples as a probability distribution. In this work we propose a change of perspective by treating radiocarbon data as count data. We present an approach that models the expected number of samples occurring at each year. We argue that this model provides more interpretable parameters and better accounts for the uncertainty in the number of samples. The performance of the proposed approach is evaluated through simulations and compared to an alternative state-of-the-art approach. Our new method is competitive with the state-of-the-art model. Furthermore, we demonstrate the computational burden of using the SPD as summary statistics under an approximate Bayesian computation analysis and propose more efficient summary statistics. Finally, we use a dataset of radiocarbon samples from Ireland and Britain to provide an application example. The results of these analyses are largely congruent with previous work on the same dataset except in revealing an earlier start of the Neolithic demographic expansion.

**Keywords:** radiocarbon dating, demography, likelihood-free inference, simulation

[1]CBGP, INRAE, CIRAD, IRD, Institute Agro, University of Montpellier, Montpellier, France, [2]Human Evolution Program, Department of Organismal Biology, Uppsala University, Uppsala, Sweden, [3]AGAP Institut, University of Montpellier, CIRAD, INRAE, Institut Agro, Montpellier, France

**Correspondence**
miguel.navascues@inrae.fr

## Introduction

The development of radiocarbon dating (Libby et al., 1949) has revolutionized the study of the past, finding applications in archaeology, geology, paleobiology, and paleoclimatology (Bronk Ramsey, 2008; Carleton and Groucutt, 2021; Taylor, 1995). As this technique became a standard in research, the accumulation of dated samples has led to the investigation of sample abundance over time, addressing various questions related to environmental processes (changes in sea level, forest fire frequency, or fluvial activity; Geyh, 1980; Pierce et al., 2004; Thorndycraft and Benito, 2006), as well as studying population size changes in humans and other species (*e.g.* Broughton and Weitzel, 2018; Rick, 1987) and ecological interactions (Marom and Wolkowski, 2024). There is a growing interest in the analysis of radiocarbon sample abundance, notably fueled by the recent availability of extensive $^{14}$C databases (e.g. Bird et al., 2022) and the development of new statistical methods (reviewed in Crema, 2022).

Until very recently, the analysis of radiocarbon data abundance relied predominantly on the Sum of Probability Distributions (SPD). The SPD is derived by aggregating the posterior distributions for the calibrated age of each sample in the dataset. However, the interpretation of the SPD encounters a main challenge because it lacks a precise definition of its underlying meaning. Despite speculation by some authors on the meaning of such sums of probabilities, a formal mathematical definition is notably absent ~~(*e.g.* Carleton and Groucutt, 2021; Crema, 2022, ; also see the supplementar~~ (*e.g.* Carleton and Groucutt, 2021; Crema, 2022, also see the supplementary text S.1). Despite this lack of a clear interpretation, the SPD is considered informative regarding changes in radiocarbon sample abundance over time. Nevertheless, the absence of a formal model hinders the full use of this intuition, as there is no established measure of the significance and uncertainty associated with variations in the SPD.

In recent years, significant progress has been made with the introduction of model-based methods, as extensively reviewed by Crema (2022). This advancement has opened up new avenues for analyzing the abundance of radiocarbon dates, enabling the testing of models, making model comparisons, and estimating model parameters. However, in these innovative approaches ~~still heavily rely on the SPD or a probabilistic interpretation thereof. In these models,~~ radiocarbon dates are conceptualized as independent samples drawn from a probability distribution~~, with the expectation that the distribution's shape mirrors that of the normalized SPD. Such models serve either to generate pseudo-observed data or as the foundation for inference (Carleton, 2021; Crema and Shoda, 2021; Porčić et al., 2020; Timpson et al., 2020)~~. This assumption implies an inmutable data-generating process in which each new radiocarbon sample is randomly drawn from the same distribution. However, ~~interpreting the SPD~~ treating the abundance of radiocarbon samples as a probability distribution overlooks the inherent nature of radiocarbon data, which is essentially count data. The number of samples (whether total or within a specific period) is an outcome of the whole data-generating process, not a fixed parameter ~~determined~~ set by the experiment or researcher. Consequently, models that assume a fixed number of samples fail to fully account for the inherent uncertainty associated with the sampling process.

~~In this study, we advocate for a departure from the use of the SPD or methods derived from it in statistical inference for the analysis of radiocarbon data abundance. Instead, we~~ Furthermore, this perspective imposes a static view of the abundance of radiocarbon samples, attributing it solely on factors acting at time the sample was formed (*e.g.* population size, intensitity of fire use or waste disposal practices). Under this framework, new samples are expected to come from

the same distribution and a statistical reanalysis of the new data merely refines the estimation of that distribution. However, the generation of radiocarbon-dated samples is influenced by factors specific to the each sample (*e.g.* research question or availability of alternative dating procedures) and factors that depend on the time of the sampling (*e.g.* damage or destruction of archaeological heritage, shift of research interests). The data-generating process is mutable and we argue that any new data should lead to a revised model that integrates all factors affecting the entire radiocarbon record.

In this study, we advocate for the use of model-based methods that more accurately describe the data-generating process. We propose a novel model that treats radiocarbon data as count data~~and characterizes the expected number of samples per year~~, allowing the total amount of samples to be determined by the model rather than imposed as if it were part of an experimental design. The parameters of this model ~~offer~~ provide a natural interpretation in the context of the studied process, characterizing the expected number of samples per year and can be interpreted as combining all the factors that affect the abundance of radiocarbon samples. Inference within this model is executed within the approximate Bayesian computation framework, and its application to pseudo-observed data allows for an exploration of differences with ~~approaches inspired by the SPD~~a state-of-the-art model-based approach.

To illustrate the application of our proposed model, we reanalyze a published dataset of archaeological radiocarbon dates from Britain and Ireland (Bevan et al., 2017). This case study serves as an example of the practical application of our approach, shedding light on its potential advantages over the SPD. Our results are in congruence with those by Bevan et al. (2017), but provide formal statistical support to the conclusions.

## Material and methods

### Models for the abundance of radiocarbon samples

*Model of counts.* In the newly proposed model, the radiocarbon-dated samples are represented as a vector $\mathbf{R}$ (refer to Table S1 for notation in the article). Each element $R_t$ of vector $\mathbf{R}$ denotes the number of samples at each year $t$ within a specified time range $[t_{\text{min}}, t_{\text{max}}]$. The abundance of radiocarbon samples is conceptualized as a Poisson distribution ($R_t \sim \text{Poisson}(\lambda_t)$). This model offers a straightforward formulation with interpretable parameters: assuming that at each year $t$ there were a potential number of ~~items~~ samples $n_t$ that could contribute to the dataset with a probability $p_t$, the rate parameter $\lambda_t = n_t p_t$ represents the expected number of samples at year $t$. In the context of archaeological data, for example, $n_t$ would encompass all organic objects associated with or connected to human activity, qualifying as anthropogenic samples. The probability $p_t$ encapsulates a multifaceted process, including sample deposition, preservation over time, discovery or sampling, and decision to conduct radiocarbon dating. Consequently, vector $\boldsymbol{\lambda}$ encapsulates this intricate data-generating process, with its values representing the expected values for vector $\mathbf{R}$ (see figure 1a and b for a visual representation).

However, $\mathbf{R}$ is not directly observable because the true age of each radiocarbon sample is unknown. Radiocarbon dating provides the Conventional Radiocarbon Ages (CRA), also referred to as uncalibrated dates. The CRA values would correspond to the true dates if the environment $^{14}$C proportion was constant through time and geography and equal to that of the atmosphere in 1950, if the true $^{14}$C half-life was Libby et al.'s (1949) estimate of 5730 years (Bronk Ramsey, 2008), and if the $^{14}$C proportions were measured without error. In reality, these assumptions

do not hold. Thus, radiocarbon data has the form of vector $\mathbf{R'}$, which contains the number of samples, $R'_u$, dated at uncalibrated date $u$ within a range $[u_{\text{min}}, u_{\text{max}}]$ (figure 1c). The relationship between $\mathbf{R}$ and $\mathbf{R'}$ is given by the calibration curve, as in well established Bayesian analysis of radiocarbon dates (Bronk Ramsey, 2008). Assuming that radiocarbon dating uncertainty can be modelled with a normal distribution, the radiocarbon age $u$ of a sample of age $t$ is modelled as $u \sim \mathcal{N}\left(u_{c,t}, \sqrt{e_{c,t}^2 + e_{\text{CRA}}^2}\right)$, where $u_{c,t}$ and $e_{c,t}$ are the values of the calibration curve for time $t$ and $e_{\text{CRA}}$ is the measurement error in the CRA for that sample. Using this normal distribution it is possible to model the observed number of uncalibrated dates ($\mathbf{R'}$) from the expected number of samples contributed by each 'calibrated' year ($\boldsymbol{\lambda}$).

*Changes through time of the abundance of radiocarbon samples.* The model for the abundance of radiocarbon samples as described above is determined by the set of parameters $\lambda_t$ in $\boldsymbol{\lambda}$. In most cases, periods of hundreds if not thousands of years will be analysed, which makes models with large number of parameters (one $\lambda$ per year). This is impractical because large amounts of data would be necessary to fit that many parameters and there would be a very likely risk of over-fitting the model. Instead, additional models can be used to determine the change of $\lambda$ through time, assuming that consecutive years will have similar $\lambda$ values. In this work, three models are explored. The first two are the exponential model ($\lambda_t = \lambda_0 e^{-rt}$, as in figure 1a) and the logistic model ($\lambda_t = \frac{k\lambda_0}{\lambda_0 + (k-\lambda_0)e^{-rt}}$). These are simple models often associated to demographic processes and used in the context of the analysis of abundance of radiocarbon samples (*e.g.* Bevan et al., 2017). For a demographic interpretation of the changes in abundance of radiocarbon samples, the parameters of these models represent the initial population size ($N_0 = C\lambda_0$), the carrying capacity ($K = Ck$) and the growth rate ($r$), with $C$ being an unknown constant of proportionality.

However, assuming that a single mathematical function governs the changes in $\lambda_t$ over large periods of time might not be appropriate. Piecewise models can be used to set a different relationship between $\lambda$ and $t$ at different periods. The whole range of time considered $[t_{\text{min}}, t_{\text{max}}]$ is divided in $m$ periods defined by $m+1$ times $t_0, t_1, \dots, t_m$ (with $t_0 = t_{\text{min}}$ and $t_m = t_{\text{max}}$). Here, we consider a piecewise exponential model defined by $m+1$ parameters $\lambda_{t_0}, \lambda_{t_1}, \dots, \lambda_{t_m}$. Within each period $x \in [1, m]$, $\lambda$ changes exponentially with rate $r_x = \frac{\log(\lambda_{t_x}) - \log\left(\lambda_{t_{x-1}}\right)}{t_x - t_{x-1}}$. For simplicity, we consider the specific case in which all time intervals are of the same length.

*Comparisons between two sets of radiocarbon data.* Some research questions require the comparison of two sets of samples of radiocarbon data (*e.g.* comparison of two geographical regions, or different food sources on the same region). For two sets a and b, the total data is $\mathbf{R} = \mathbf{R}^{\text{a}} + \mathbf{R}^{\text{b}}$. Each set of radiocarbon data can be modelled with a Poisson distribution: $R_t^{\text{a}} \sim \text{Pois}(\lambda_t^{\text{a}})$, $R_t^{\text{b}} \sim \text{Pois}(\lambda_t^{\text{b}})$, and $R_t \sim \text{Pois}(\lambda_t) = \text{Pois}(\lambda_t^{\text{a}} + \lambda_t^{\text{b}})$. We define $q_t = \frac{\lambda_t^{\text{a}}}{\lambda_t}$, which describes the proportion of category a contributing to the total amount of samples. Our interest here is to understand whether the relationship of the changes of $\lambda_t^{\text{a}}$ and $\lambda_t^{\text{b}}$ with time is determined by some common factors or if their histories are independent. We consider three scenarios for the relationship between two sets of samples. In the first scenario, that we name "independent", $\lambda_t^{\text{a}}$ and $\lambda_t^{\text{b}}$ values are independent. In the second scenario, that we name "interdependent", parameters $\lambda_t$ and $q_t$ determine $\lambda_t^{\text{a}} = q_t\lambda_t$ and $\lambda_t^{\text{b}} = (1-q_t)\lambda_t$. The third scenario, that we name "parallel", is a special case of the interdependent scenario in which $q_t$ is constant through time. The dependency among parameters in these scenarios is obtained through conditional prior probability distributions among $\boldsymbol{\lambda}^{\text{a}}$, $\boldsymbol{\lambda}^{\text{b}}$ and $\mathbf{q}$ (see below).

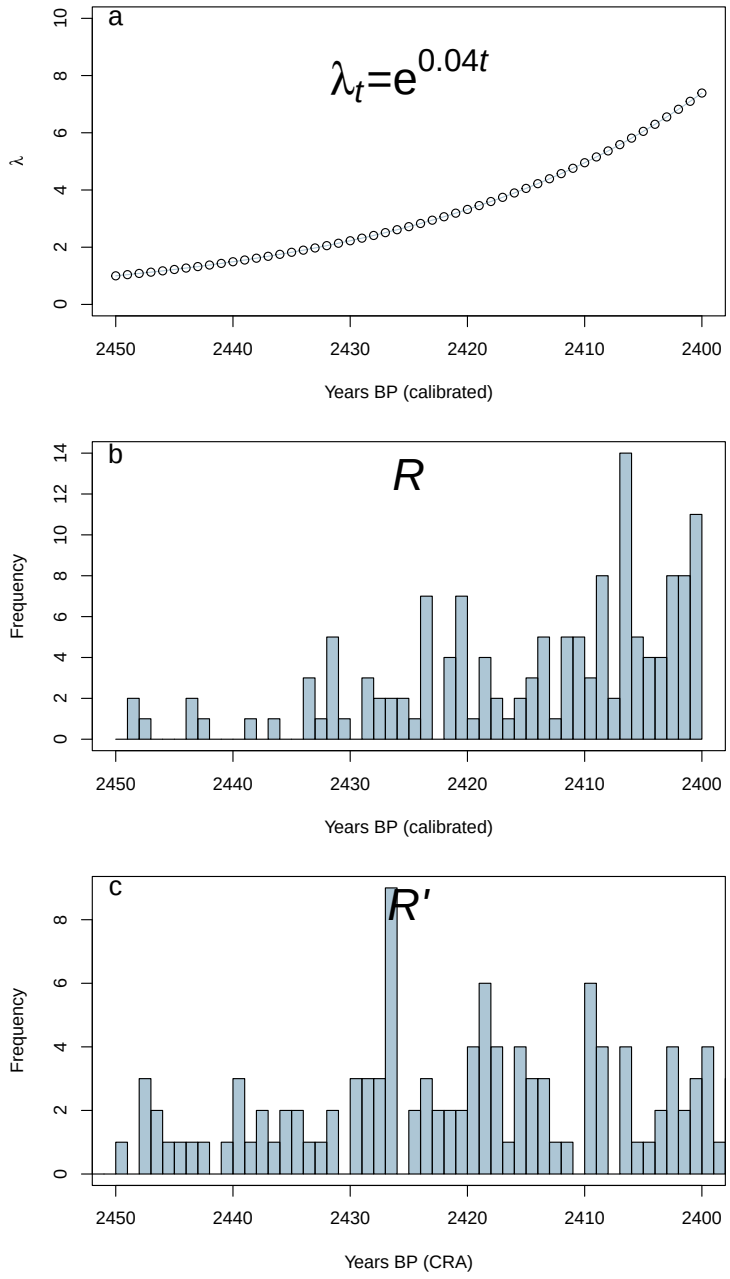


**Figure 1 – Model for the abundance of radiocarbon samples (example).** (a) A mathematical law determines the relationship between the expected number of samples per year ($\lambda$, the rate parameter of a Poisson distribution) and time ($t$): in this example an exponential law with initial value $\lambda_0 = 1$ at time $t_0 = 2450$YBP (*i.e.* $t = t_0 - t_{\text{YBP}}$) and growth rate $r = 0.04$. (b) Number of samples per year (true age) in the data set (**R**, not observable) of one random realization of model in (a); that is, random draws from Poisson distributions with parameters in $\boldsymbol{\lambda}$. (c) Number of samples per year (conventional radiocarbon age, CRA) in the data set (**R′**); that is, random draws from Normal distributions with parameters determined by the calibration curve and ages in (b).

*Model of probabilities.* Previous works have considered similar models (*e.g.* exponential change) to describe a probability distribution of the age of a sample. A fixed number of draws from this distribution is assumed to constitute the data set of radiocarbon dates (Crema and Shoda, 2021; Porčić et al., 2020; Timpson et al., 2020). These "probability distribution" models, describe the change of probability $\pi$ through time instead of the change of $\lambda$ through time. In this model, $\pi_t$

is the probability that a radiocarbon sample is from year $t$. For any of the models of counts, an equivalent model of probabilities can be obtained by setting $\pi_t = \frac{\lambda_t}{\sum_{t_{min}}^{t_{max}} \lambda_t}$, so the total probability of the model equals one for the period considered. It is important to note that by doing this normalization the model of probabilities has one degree of freedom less than the model of counts. For instance, the exponential count model has parameters $\lambda_0$ and $r$, while the exponential probability model is determined solely by $r$ (there is a single possible value of $\pi_0$ for each value of $r$). Also, the probability model is restricted to the studied period (formally, the probability outside the range is zero), while the count model can be extrapolated beyond that period of time.

**Inference using approximate Bayesian computation**

Approximate Bayesian Computation (ABC) is a statistical approach to make model-based inference without the calculation of likelihoods (see Sunnåker et al., 2013, for a review). ABC is often used for inference under models with analytically intractable likelihoods, which is not necessarily the case for the models of abundance of radiocarbon samples (*e.g.* Crema and Shoda, 2021). However, it has other advantages such as the fast implementation under different models and priors, which is one of the main reasons for its use in this work (see below for a discussion of other reasons). In ABC, the calculation of the likelihood of a model is substituted by the simulation of data under the model. The similarity between the real and simulated data reflects the likelihood of the model.

*Summary statistics.* The similarity between the real and simulated data is typically evaluated by comparing several summary statistics of the data. Previous applications of ABC to the analysis of the abundance of radiocarbon samples used the values of the SPD at each year as summary statistics (DiNapoli et al., 2021; Porčić et al., 2020). In this work we also explore the alternative of using summary statistics based directly on CRAs (*i.e.* $\mathbf{R}'$). Specifically we use: $T$, the total number of uncalibrated dates; $H_{u_i}$, the number of uncalibrated dates at interval $[u_i, u_i - \delta)$ (with values covering the whole period of analysis and $u_{i+1} = u_i + \delta$) and using several values of $\delta$ $(10, 50, 100, 500)$; and $\Delta H_{u_i}$, the difference between consecutive $H_{u_i}$ and $H_{u_{i+1}}$ values. A visual example of $H_{u_i}$ statistics is the histogram of CRA from Britain and Ireland in figure 2.

In the case of real archaeological data, dates belonging to the same site are given a lower weight for the calculation of all these statistics. This is done to compensate biases due to large variance in sample size among sites that could reflect, for instance, differences in the resources or research questions of the teams working on them rather than the abundance of materials. These weights are calculated by using the binning procedure proposed by Shennan et al. (2013). The weight for each uncalibrated date is the inverse of the number of dates within the bin. For instance, all the dates within a bin count as a single sample for computing $T$. Here we have used a binning range of 100 years.

In the case of the analysis of two sets of radiocarbon data, we define additional summary statistics. These additional statistics capture the relationship between the two sets in their abundance of samples or its change. This is captured by the calculation of the correlation and covariance between $\mathbf{H}^{\mathrm{a}}$ and $\mathbf{H}^{\mathrm{b}}$, and between $\mathbf{\Delta H}^{\mathrm{a}}$ and $\mathbf{\Delta H}^{\mathrm{b}}$ (for sets a and b).

*Approximate Bayesian computation* via *random forests.* Previous applications of ABC to the analysis of abundance of radiocarbon samples have used the ABC rejection algorithm (DiNapoli et

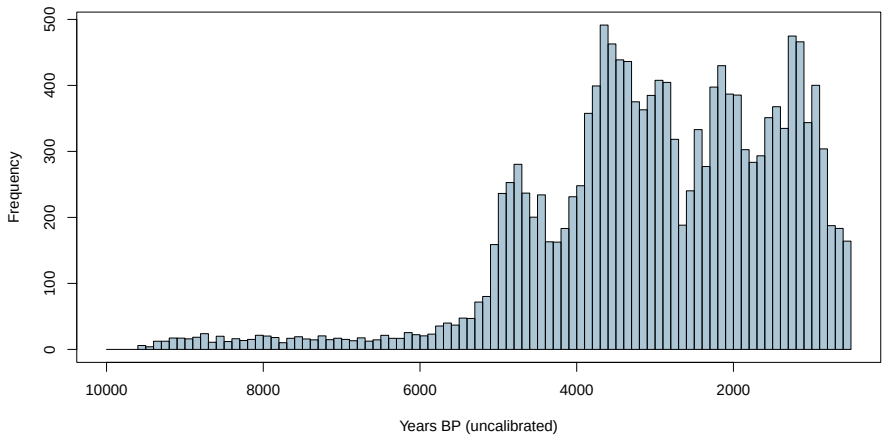


**Figure 2 – Conventional radiocarbon ages composing the dated archaeological record from Britain and Ireland (Bevan et al., 2017).** Histogram with number of radiocarbon samples in bins of 100 (uncalibrated) years (samples from the same site are weighted similarly to the procedure proposed by Shennan et al., 2013). This is a visual representation of summary statistics $H_{u_i}$ with $u_i$ taking values from 9900 to 500 YBP and $\delta = 100$.

al., 2021; Porčić et al., 2020). This algorithm represents the most basic way of performing ABC and presents several limitations respect to other algorithms proposed for the comparison of observed and simulated summary statistics. Here, we use ABC *via* Random Forests (ABCRF; Pudlo et al., 2016; Raynal et al., 2019), which uses the eponymous machine learning algorithm to learn the relationship between summary statistics similarity and posterior probability of the model or the parameters. In the learning step, random forests are grown from a training set constituted by a large number of simulations known as the reference table. One random forest is grown for each parameter or for each model comparison and they can be used to make predictions about the real data. An important advantage of this algorithm is that a lower number of simulations are required for inference (reducing the computational cost) and there is no need to set an arbitrary tolerance level.

*Simulation.* The simulation of radiocarbon data requires to set a specific model for the relationship between $\lambda_t$ with time (*e.g.* the logistic model) and the values of its parameters (*e.g.* $\lambda_0^*$, $r^*$ and $K^*$, for the logistic model; where $^*$ denotes simulation values). These will determine all values in $\boldsymbol{\lambda}^*$, which are then used to simulate $\mathbf{R}^*$ by sampling from Poisson distributions. The uncalibrated date $u^*$ for each sample of known date $t^*$ in $\mathbf{R}^*$ will be simulated by sampling from a Normal distribution with mean and standard deviation taken from the CRA and error associated to $t^*$ in the appropriate calibration curve (Shennan et al., 2013). This will result in $\mathbf{R}'^*$. An example of this procedure is presented in figure 1. This simulation process is rather similar to the procedure proposed by Shennan et al. (2013) and widely used in other works. The main difference is that the total number of samples in the simulated data set depends on the model, allowing to account for this additional source of stochasticity. The IntCal20 calibration curve (Reimer et al., 2020) is used throughout this paper.

## Evaluation of the proposed model and approach

The performance of the method can be evaluated on simulated data for which the generating model and parameter values are known. This is done by exploiting the properties of the random forest algorithm. Random forests are a collection of decision trees that are grown from random

subsets of the training data (the reference table in the case of ABC), in a processed called bootstrap aggregating or "bagging". Because of this, for each simulation in the reference table there is a subsets of trees in the random forest that have been grown without the information of that simulation. That subset of trees can be used to make inferences for that simulation, which are called the "out-of-bag" (OOB) predictions. True values and OOB predictions can be compared to estimate the error of the method, without the need of an additional testing set. In the context of model choice, OOB error is used to provide confusion matrices. In the context of parameter inference, OOB predictions are used to calculate mean squared error and the correlation coefficient between true values and their corresponding OOB prediction.

*Choice of summary statistics.* The reduction of the data to a set of summary statistics can produce loss of information for the ABC. Therefore, it is recommended to use a set of summary statistics that are informative about the models and parameters to be inferred. Using the SPD (as in DiNapoli et al., 2021; Porčić et al., 2020) is a logical choice since SPD is considered to be highly informative about the changes in abundance of radiocarbon samples. However, the calculation of SPD is computationally costly. Also, strictly speaking, the SPD is not a summary of the data but a combination of the data with the calibration curve. Here we propose an alternative set of summary statistics based on the CRA data as described above.

It is important to determine if these summary statistics are as informative as the SPD for the inference and if there is a gain in computational time by using them. First, the computational time for the calculation of the two sets of summary statistics was measured in 300 simulated datasets of 1343 CRA dates. This simulated CRA datasets were generated by sampling 1343 calibrated dates uniformly between 7000 and 5000 YBP and using a CRA error of 30 years for simulating their corresponding CRA. The bench-marking procedure compares the SPD calculation as implemented in `R` package `rcarbon` (Crema and Bevan, 2021), and an implementation of the new set of statistics in `R` (de Navascués, 2024). Then, the performance of the two sets of summary statistics for ABC inference was also evaluated. A reference table of 20000 simulations was produced using the model of probabilities with probability changing exponentially between 7000 and 5000 YBP (*i.e.* the same model used in the ABC example in Crema, 2022). The growth rate parameter, $r$, was sampled from a uniform prior distribution between $-0.01$ and $0.01$. Two random forest models with 5000 trees were trained from this reference table, one using the SPD as predictors and another one using the new set of summary statistics ($T$, $H_{u_i}$ and $\Delta H_{u_i}$).

*Model of probabilities* versus *model of counts.* The effect of using a model of counts instead of using a model of probabilities is studied by generating one reference table from each of the two models under exponential change on the period from 7000 YBP to 5000 YBP. Parameter values are sampled from the following prior probability distributions: uniform between $-0.005$ and $0.005$ for $r$, and log-uniform between $0.005$ and $5$ for $\lambda_0$. A condition of $\sum_{t=0}^{2000} \lambda_0 e^{-rt} < 5000$ is imposed to avoid simulations with an unrealistic high value of samples. For each parameter value combination, $r^*$ and $\lambda_0^*$, two simulations are run, one for each of the two separate reference tables. The first simulation uses $r^*$ and $\lambda_0^*$ to simulate under the model of counts and the second uses only $r^*$ to simulate 1343 CRA dates from a model of probabilities. Summary statistics based on the CRA data ($T$, $H_{u_i}$ and $\Delta H_{u_i}$) are calculated and random forests are trained for $r$ and $\lambda_0$ for the count model, and $r$ and $\pi_0$ for the probability model (note that the estimation of $\pi_0$ is done for comparison with $\lambda_0$ but is unnecessary in practice if parameter $r$ has already been estimated).

In addition to the evaluation through the OOB predictions, a separate set of independent simulations (pseudo-observed data-sets, PODs) was produced to study the properties of the posterior distributions obtained under the two different models. These PODs were simulated in a model of counts with $\lambda_0 = 0.01$, exponential rate change of $r = 0.003$, with starting time 7000 YBP and final time 5000 YBP. Simulations were run until 300 PODs were obtained that contained exactly 1343 samples (the expected number of samples under that model). Conditioning the simulation to a specific number of samples was done in order to analyse those PODs with the reference tables from both models (counts and probabilities) since the probabilities model assumes a fixed number of samples. The first 300 PODs were used to estimate posterior probability distributions for $r$ under the model of counts and the PODs with 1343 samples were used to estimate posterior probability distributions for $r$ under the model of probabilities.

**Case study: archaeological radiocarbon dates from Britain and Ireland**

In order to illustrate the approach presented in this work, we reanalyse data of archaeological radiocarbon dates from Britain and Ireland (Bevan, 2017). This data base comprises 30516 radiocarbon dates from 200 to 9580 uncalibrated YBP from Ireland (7797 entries), Scotland (6401 entries), North-West England and Wales (5333 entries), and South-East England (10985 entries). In more than three quarters of the entries, the taxonomic origin of the material is identified. The taxonomic level of this identification is heterogeneous across the data: sometimes identification is at species level but often it is only at genus or higher levels. Among the taxon identified, there are several food sources, such as as wheat (*Triticum*, 678 entries) and barley (*Hordeum*, 1102 entries).

The original article by Bevan et al. (2017) studies the change of human population size and usage of food resources based on those data. Our work is not intended as a thorough reanalysis of this dataset but as an illustration of the model and method proposed. Therefore, we only focus on two questions: the global pattern of change in abundance of radiocarbon samples (interpreted as a population size proxy in the original article) and the relationship between the abundance of samples of barley and wheat through time.

*Estimation of population size changes in Britain and Ireland.* For the analysis of the population size change in Britain and Ireland, we consider the three models described above: exponential change, logistic change and piecewise exponential change. The time period explored is restricted between 10000 and 500 YBP. For the exponential and the logistic models, the parameters $\lambda_0$ and $\lambda_f$ (value of $\lambda$ at 500 YBP) were taken from a log-uniform prior distribution in the range $[0.001, 12]$, conditional to histories of increasing $\lambda$ ($\lambda_0 < \lambda_f$). For the logistic model, parameter $k$ value was sampled from a log-uniform distribution in the range $[\lambda_f + 0.001, \lambda_f + 12]$. Rate of change $r$ is obtained from the values of those parameters.

In the piecewise exponential model there are $m+1$ parameters $\lambda$ ($\lambda_{t_0}, \lambda_{t_1}, \ldots, \lambda_{t_m}$). The value of $m$ is set to divide the analysed total range of ages in periods of approximately 400 years. Thus, for the range 10000 to 500 YBP, $m = 24$. The value of $\lambda_{t_0}$ is taken from a log-uniform prior distribution in the range $[\lambda_{\min}, \lambda_{\max}]$ and consecutive values $\lambda_{t_x} = \max(\min(\phi\lambda_{t_{x-1}}, \lambda_{\max}), \lambda_{\min})$ with $\phi$ taken from a log-uniform distribution in the range $[0.1, 10]$ (as in Boitard et al., 2016). This way of sampling the evolution of $\lambda$ through time reflects the prior belief that large jumps over a short period of time are unrealistic (this prior prevents changes larger that one order of

magnitude for consecutive $\lambda_{t_x}$ values). The minimum and maximum $\lambda$ values for the whole model are $\lambda_{\text{min}} = 0.001$ and $\lambda_{\text{max}} = 12$.

For each model, a reference table of 30 000 simulation was ~~build~~ built taking parameter values from the prior distributions described above. Model choice and posterior probability for the observed data were obtained through ABCRF, using 2000 trees for the training of the random forest and 2000 trees for the calculation of the posterior probability. The pertinence of the approach was evaluated in two ways. First, OOB prediction were used to calculate the confusion matrix to evaluate the general performance of the approach. Second, a visual evaluation of the goodness-of-fit of the model to the observed data is also provided: the variability of patterns produced by the different models is represented using Principal Component Analysis (PCA) of the summary statistics in the reference table, then the observed data set is projected into the PC space.

A larger reference table of 100 000 simulation was used to estimate the parameters of each model. Random forest of 2000 trees were trained on $\log(\lambda_0)$, $\log(\lambda_f)$, $\log(k)$ for the exponential and logistic model. For the piecewise exponential model random forest of 2000 trees were trained for $\log(\lambda_{t_x})$ at each the 25 time points defining the periods and $r_x$ for each of the 24 periods. Refer to the full description of the model and parameters above for more details.

*Testing the relationship between abundances of wheat and barley in Britain and Ireland.* For the study of the abundances of wheat and barley, we considered a piecewise exponential model and explored the time range from 6000 to 500 YBP divided in $m = 14$ periods of approximately 400 years. The model describes the abundance of radiocarbon samples of two categories: wheat (w) and barley (b). The samples were ascribed to these two categories following the same criteria as in Bevan et al. (2017). The relationship between the changes of abundance through time of these two categories was modelled according to the above mentioned independent, interdependent and parallel scenarios. All three scenarios are produced with the same model, which have 15 parameters $\lambda^{\text{w}}_{t_x}$ and 15 parameters $\lambda^{\text{b}}_{t_x}$; the differences among scenarios reside in the conditional prior probability distributions.

For the independent scenario, parameters $\lambda^{\text{w}}_{t_x}$ and $\lambda^{\text{b}}_{t_x}$ are sampled independently using the same procedure as described above. That is, $\lambda^{\text{w}}_{t_0}$ is sampled from a log-uniform distribution in the range $[\lambda_{\text{min}}, \lambda_{\text{max}}]$ and consecutive values $\lambda_{t_x} = \max(\min(\phi\lambda_{t_{x-1}}, \lambda_{\text{max}}), \lambda_{\text{min}})$ with $\phi$ taken from a log-uniform distribution in the range $[0.1, 10]$. For the interdependent scenario, parameters $\lambda_{t_x}$ (*i.e.* $\lambda$ for the sum of both categories), are sampled with the same procedure; then $q_{t_x}$ are sampled from a uniform distribution in the range $[0, 1]$ which determines the proportion of categories $\lambda^{\text{w}}_{t_x}$ and $\lambda^{\text{b}}_{t_x}$ at each time $t_0, t_1, \dots, t_m$. Finally, the parallel scenario is a special case of the interdependent scenario, in which a single $q$ value is taken from a uniform distribution in the range $[0, 1]$ and the proportion of the two categories does not change through time. The minimum and maximum $\lambda$ values for the whole model are $\lambda_{\text{min}} = 0.001$ and $\lambda_{\text{max}} = 2$.

## Implementation

All the calculations presented in this work were done in R (R Core Team, 2021) with scripts (available in de Navascués, 2024) that use: package extraDistr (Wolodzko, 2020) to sample from prior distributions; package rcarbon (Crema and Bevan, 2021) to simulate CRA; packages Hmisc

(Harrell Jr, 2022), moments (Komsta and Novomestky, 2022) and weights (Pasek, 2021) to calculate summary statistics; and package abcrf (Marin et al., 2022) to perform ABC analyses. Simulations are run in parallel using doParallel (Microsoft Corporation and Weston, 2022b), doSNOW (Microsoft Corporation and Weston, 2022a) and doRNG (Gaujoux, 2023).

## Results

### The choice of model and summary statistics for ABC inference

We evaluated the performance of two distinct sets of summary statistics. One set comprises the values of the SPD for each year, while the second set is calculated from counts of uncalibrated dates as detailed in the Methods section. Summary statistics based on counts of uncalibrated dates offer a significant computational advantage, being approximately 250 times faster to calculate. Despite this difference, both sets of summary statistics demonstrate very high accuracy in inference with no discernible difference in statistical results (Figure 3).

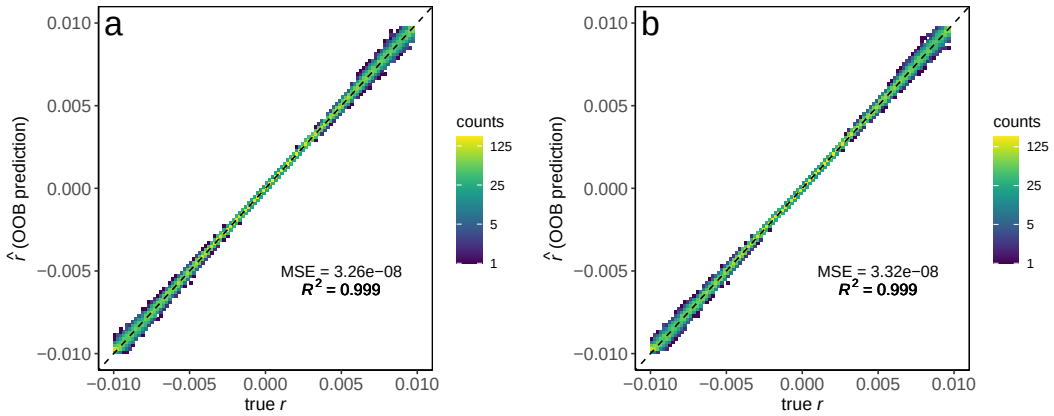


**Figure 3 – Influence of choice of summary statistics on the estimation of parameters.** Out-of-bag (OOB) estimates of the exponential growth rate, $r$, compared to the true value from simulations in the reference table. The ABC was performed either using: (a) the SPD as summary statistics, or (b) a set of summary statistics calculated from the count of CRA. The performance is very similar despite the much higher computational cost of using the SPD.

We also evaluated the use of two different models, referred to as the model of probabilities and the model of counts, each offering a distinct perspective on the process generating radiocarbon data. Notably, parameter inference under the model of probabilities demonstrated higher accuracy compared to the model of counts (see Figure 4). This discrepancy in accuracy primarily stems from larger errors observed in simulations with low values of $r$ or $\lambda_0$ (refer to Figure 4a and b), which consequently results in less data for the model of counts.

For PODs generated under the model of counts ($\lambda_0 = 0.01$, $r = 0.003$, with a range of 2000 years), analysis under either the model of probabilities or the model of counts yielded comparable levels of error (mean squared error of $9.57 \times 10^{-9}$ and $9.31 \times 10^{-9}$ respectively for parameter $r$). However, it's worth noting that the 95% credibility intervals were wider for the model of counts (refer to Figure 5b). Furthermore, nominal coverage was more accurate for the model of probabilities (see Figure 5a).

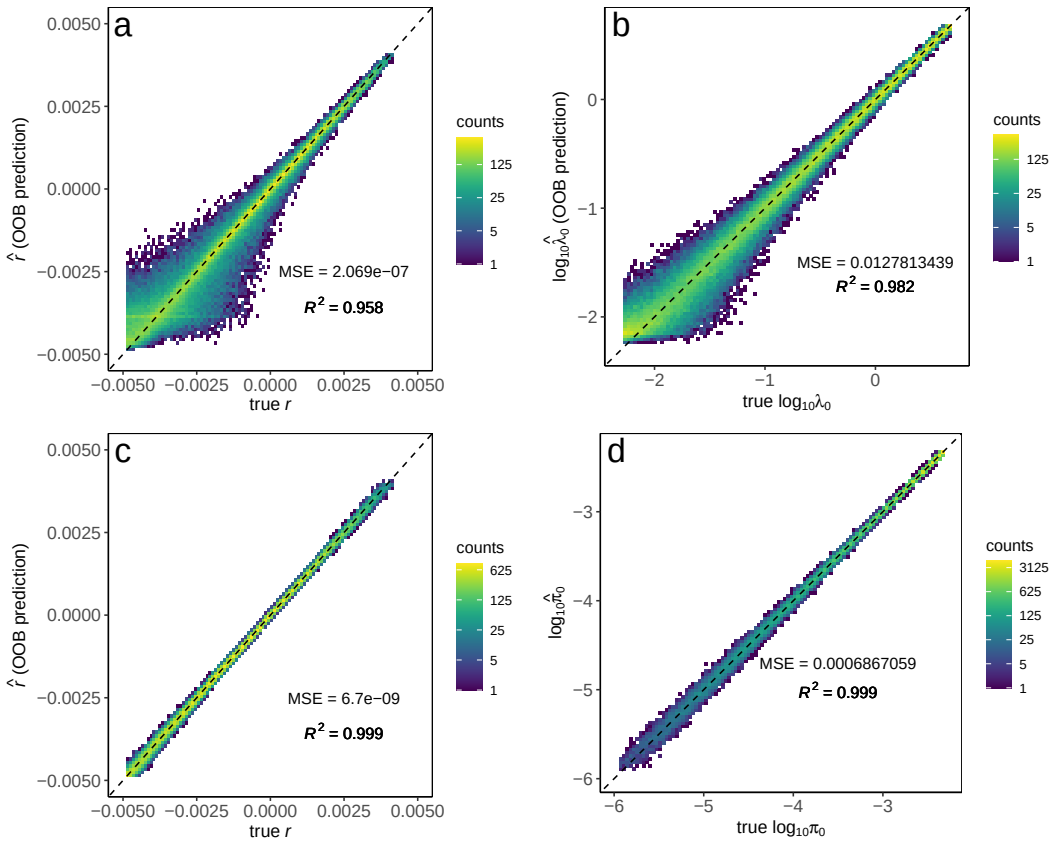


**Figure 4 – Parameter inference differences between modeling counts and modeling probabilities.** Out-of-bag (OOB) estimates compared to the true value from simulations in the reference table. (a) and (b) model of counts. (c) and (d) model of probabilities. (a) and (c) growth rate, $r$. (b) and (d) initial value ($\lambda_0$ or $\pi_0$).

## Analysis of archaeological radiocarbon dates from Britain and Ireland

*Estimation of $\lambda$ as a proxy of population size.* Three models (exponential, logistic and piecewise exponential) were explored to explain the change in abundance of radiocarbon samples from Britain and Ireland. According to the OOB estimates from the training set, the ABCRF approach is able to identify the piecewise exponential model with very little error and the logistic model with a somehow higher error. However, the exponential model is difficult to identify, being often wrongly classified as the logistic model (table S2). For the real data, the piecewise exponential model has a clear superior fit than the two alternative models, with the ABCRF analysis indicating a high posterior probability for that model (0.869). The PCA of the summary statistics diversity across simulations further reveals the lack of fit of the exponential and logistic models, which are unable to reproduce the patterns found in real data (figure S1b and c). Parameter estimates under the piecewise model reveal a history of fluctuations of $\lambda$ through time closely resembling the SPD curve (figure 6a). Five of the periods (6833–6438, 6042–5646, 4854–4458, 4458–4062 and 1688–1292 YBP) have estimates of the rate of change $r$ with credibility intervals excluding ~~the~~ zero, indicating a significant increase of $\lambda$ during those periods (figure 6b).

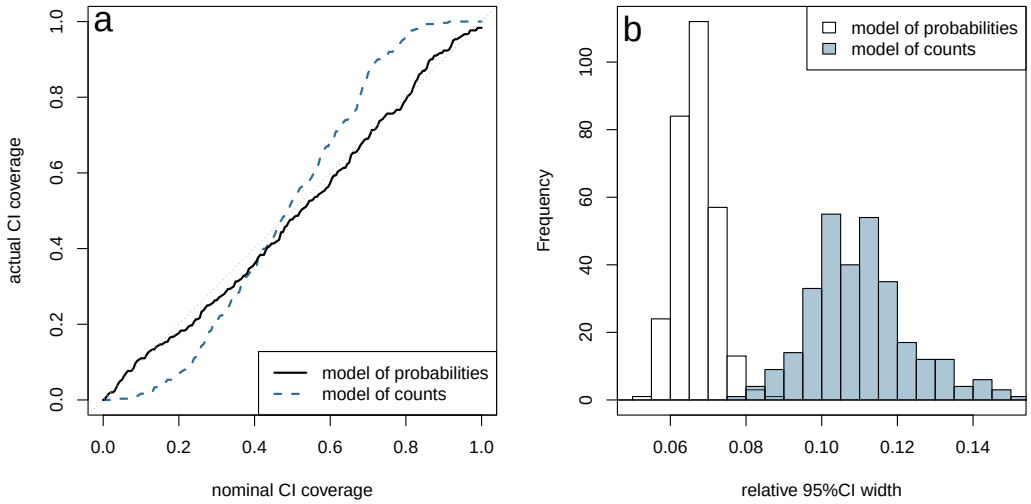


**Figure 5 – Influence of the model on the credibility interval width.** (a) Quantile-quantile plot of the actual and nominal coverage of 300 estimated posterior distributions estimated from pseudo-observed data-sets. (b) Relative width of the 95% credibility interval of the parameter $r$ estimated under a model of probabilities or under a model of counts. The histogram represent 300 values from pseudo-observed data-sets generated with a model of exponential change with $r = 0.003$ and $\lambda = 0.01$.

*Relationship between the abundances of radiocarbon samples from wheat and barley.* The change in abundance of radiocarbon samples for wheat and barley was modeled using a piecewise exponential model. Three different scenarios within this model were considered based on the degree of independence between the trajectories of the two cereals: the independent scenario, where both trajectories are completely independent; the parallel scenario, where both trajectories change in parallel; and the interdependent scenario, where both trajectories are correlated. These three scenarios can be distinguished with relatively low error using ABCRF, as indicated by the confusion matrix (table S2). For the empirical data, the chosen scenario is the interdependent scenario, with a posterior probability of 0.863. Visual evaluation of the goodness of fit through Principal Component Analysis (PCA) shows that the observed data falls within the expected diversity of summary statistics values for the interdependent scenario (figure S2). Parameter $\lambda$ estimates indicate that the abundance of both cereals sharply increased between 6000 and 5500 years before present (YBP), then decreased until 4500 YBP, and increased again until 3500 YBP, remaining stable with some minor fluctuations until 500 YBP (figure 7a). The relative abundance of the two cereals ($q$) also changed dramatically around 5500 YBP, with wheat being more abundant in the first period and barley becoming more abundant in the second period (figure 7b).

## Discussion

### Performance of the new approach

The present work proposes a novel approach to analyzing the abundance of radiocarbon samples. This new method moves away from using the ~~Sum of Probability Distributions (SPD), which we regard as useful only for visualization purposes. Insteadof trying to fit models to the SPD~~assumption that the abubdance of radiocarbon samples can be described with a probability distribution, which is the currently widespread view (*e.g.* Carleton, 2021; Crema and Shoda, 2021; Porčić et al., 2

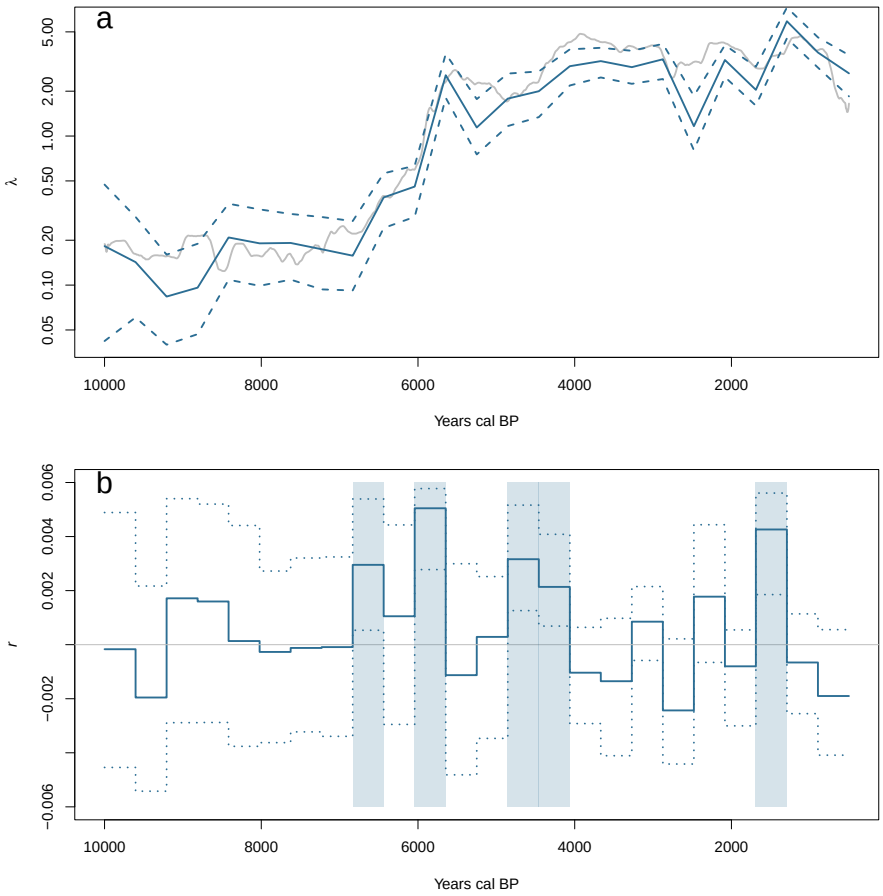


**Figure 6 – Parameter estimates under piecewise exponential model.** (a) Abundance of dated archaeological record through time measured as the expected number of dated archaeological samples per year ($\lambda$). Solid blue line indicate the point estimate ($\hat{\lambda}$) and dashed lines indicate 95% credibility interval. The Sum of Probability Distributions (grey line) of the data is plotted for reference. Note log scale for $\lambda$. (b) Rate of change in the abundance of dated archaeological record through time ($r$). Solid blue line indicate the point estimate ($\hat{r}$) and dotted lines indicate 95%CI. Periods in which the 95%CI for $r$ does not include zero (horizontal grey line) are marked with ~~an asterisk (*)~~a light blue vertical band.

. Instead, we propose using a model based on ~~the Poisson distribution~~ Poisson draws to represent the number of samples per year. We argue that this model offers a parametrization with a natural interpretation, where $\lambda_t$ is the expected number of radiocarbon samples at year $t$, and aligns better with the inherent nature of the data.

Through simulations, we demonstrate that analyzing data under this model allows for accurate inference of the expected abundance of radiocarbon samples and the rate of change in these abundances (figure 4a and b). The model of probabilities exhibits lower errors in estimating equivalent parameters (figure 4c and d). While this may appear desirable, we view it as a failure to capture the full uncertainty of the data. The model of probabilities treats the total number of samples, $T$, as a sample size controlled by the researcher. This model implies an experimental process where the researcher decides to sample $T$ times from a probability distribution. In reality, the researcher has no control over the number of radiocarbon samples in the dataset, which usually derives from the accumulation of prior research in the geographical area of interest. By fixing the total number of samples, a strong dependence is generated on the modeling

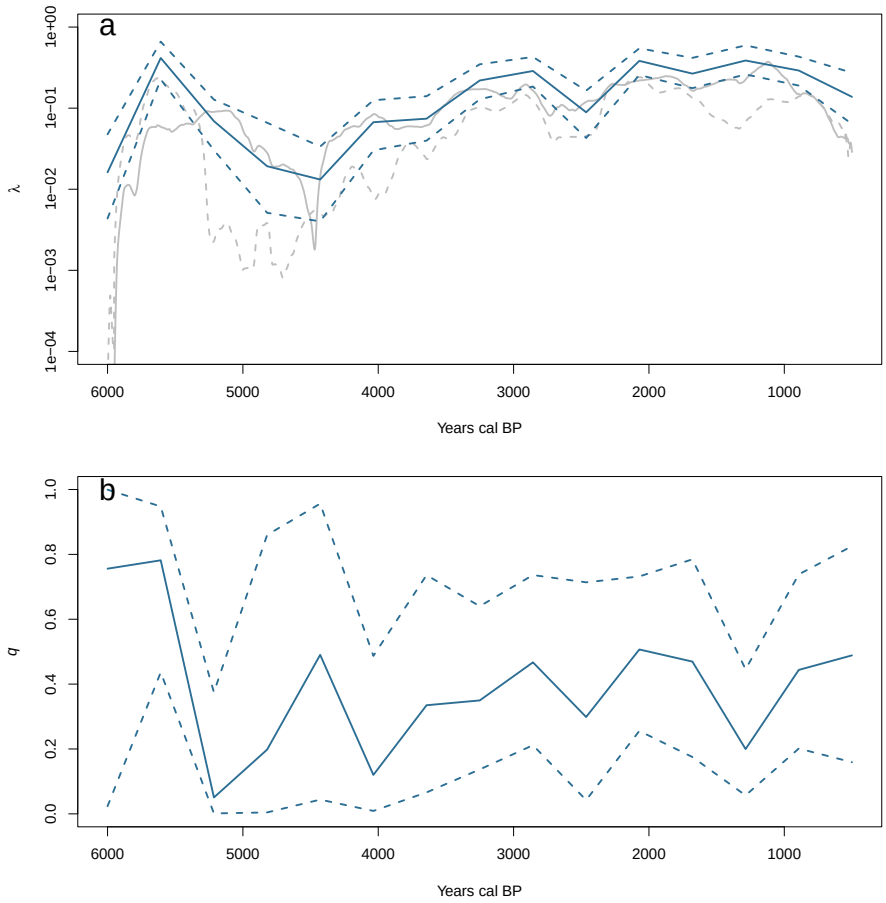


**Figure 7 – Parameter estimates under the interdependent scenario for cereals (*Hordeum*/*Triticum*).** (a) Estimate of the expected number of *Hordeum* and *Triticum* samples through time ($\lambda$). Solid blue line indicate the point estimate ($\hat{\lambda}$) and dashed lines indicate 95% credibility interval. The Sum of Probability Distributions (grey line) of *Hordeum* (solid line) and *Triticum* (dashed line) are plotted for reference. Note log scale for $\lambda$. (b) Proportion of *Triticum* ($q$) expected among the samples. Solid blue line indicate the point estimate ($\hat{q}$) and dotted lines indicate 95%CI.

of abundance of samples from different periods. Consider the initial abundance (expressed as $\lambda_0$ or $\pi_0$) and the final abundance ($\lambda_f$ or $\pi_f$) of samples. If we consider the extreme case in which all dates occur close to $t = 0$, the model of counts will provide a good estimate of parameter $\lambda_0$ but will have little information to estimate $\lambda_f$, this is why we observe larger error for estimates of $\lambda$ when the true value is small (figure 4b). However, the model of probabilities will provide an accurate estimate of $\pi_f$, despite having little information from samples of that period, since all the information is coming from the initial period and the constraints of the model.

The credibility interval coverage obtained under the model of probabilities is closer to the nominal value than that of the model of counts (figure 5). This difference likely arises from the limitation of ABCRF, which estimates the marginal posterior probability distribution of each parameter independently, rather than their joint posterior probability distribution. Since the model of probabilities has only one parameter in the example, its posterior probability distribution is

easier to estimate. For future applications, it would be interesting to consider other simulation-based inference approaches that estimate the joint distribution of parameters and provide confidence intervals with better coverage properties (Rousset et al., 2017, and unpublished results by F. Rousset).

An additional advantage of the model of counts is that simulations of the reference table can be reused to analyze different datasets. The generation of a reference table requires to assume a calibration curve and some prior distribution for the parameters; however, the number of samples is not fixed. Therefore, a reference table generated with non informative priors could be reused for dataset for which the same calibration curve are appropriate. This will save computational time and reduce the carbon footprint of the analyses (Lannelongue et al., 2021).

The use of the SPD values as summary statistics for simulation-based inference is a good idea only at first sight. The intuition that SPD is informative about the abundance of radiocarbon samples is confirmed independently of the type of model assumed (*e.g.* DiNapoli et al., 2021, and figure 6a). However, the computational cost of calculating the SPD offers no gain compared to simpler summaries of the CRA data. In addition, transforming the data into the SPD assumes a specific calibration curve for the data. This could be appropriate for many cases, but we can also imagine cases in which a proportion of samples might have some intake of marine carbon in an uncertain proportion. In a Bayesian framework, it would be straightforward to propose prior distributions for the proportion of marine carbon intake of those samples and incorporate their uncertainty on the calibration by doing so. In such a case, summarising the data with a different calibration curve could be prone to produce misunderstandings.

## Application of the new approach to data from Britain and Ireland

The main features of the demography of Britain and Ireland inferred in the original analysis by Bevan et al. (2017) are recovered in our analysis. Three periods (Early Neolithic, around 6000 YBP; Late Neolithic/Early Bronze Age, approximately between 5000 and 4000 YBP; and Early Medieval, around 1500 YBP) of demographic expansion show strong support, with distinct estimates (*i.e.* non-overlapping 95% credibility intervals) of $\lambda$ at the beginning and end of those periods and positive estimates of the growth rate $r$ with 95% credibility intervals excluding zero (figure 6). Some other features discussed by Bevan et al. (2017) have lower support, such as the decline after the peak at the Early Neolithic, the decline after the Bronze Age and its recovery: 95%CI of $r$ include zero, but they have non-overlapping 95%CI of $\lambda$. What it is noteworthy is the significant increase during the period 6833–6438 YBP which was not noted by Bevan et al. (2017). This observation of an early increase in $\lambda$ suggests a demographic expansion that starts several hundreds years before the peak around 6000 YBP. This last result highlights the importance of using the model-based statistical analyses developed in the last few years by several authors (reviewed by Crema, 2022) over visual evaluation of SPD curves.

The analysis of abundances of wheat and barley reveals that they are not independent of each other. Their total abundance (figure 7a) appears to follow the same fluctuations as the whole data, suggesting that demographic size could determine the amount of wheat and barley cultivated, potentially generating this correlated pattern of abundance between these two cereals. Bevan et al. (2017) already noticed the similarity between demography (*i.e.* abundance of all samples) and starchy food plants abundance, which is congruent with them being staple food. Regarding the relative abundance of these two cereals, the notably result is the transition of higher abundance of wheat in the Early Neolithic to the larger abundance of barley thousand years later. Relative

abundance might have fluctuated afterwards, but the largely overlapping confidence intervals prevent any meaningful discussion of these results.

**Combining radiocarbon data with other types of data**

The analysis of the abundance of radiocarbon data in archaeology has primarily focused on its use as a proxy for demography. The validity of this proxy has been thoroughly discussed in the literature (*e.g.* Rick, 1987; Williams, 2012). Some criticisms regarding the statistical uncertainty of the SPD (*e.g.* Carleton and Groucutt, 2021) are addressed in the recent works developing model-based approaches such as the one presented here. Biases arising from archaeological research (questions drive which sites or periods are studied and which samples are dated) might be attenuated by the weighting procedure proposed by Shennan et al. (2013). Taphonomy can also be modelled to correct for different preservation across samples (Contreras and Codding, 2023). However, some differences in the abundance of samples might be driven by changes in human practices (*e.g.* use of fire dependent on climate or cultural changes in the way to dispose objects) that might be more difficult to take into account. In that sense, it is important to highlight that the proposed model studies the abundance of radiocarbon samples. Its application to study demography (or any other process) requires understanding the limits of the data and model used and acknowledging those caveats. Nevertheless, inferring the past demographic dynamics is an important component for understanding the prehistoric populations and, despite its limits, the abundance of radiocarbon samples seems to be informative about it.

In order to produce more robust inferences of demography, the use of multiple sources of information has been suggested (Crema and Kobayashi, 2020; Hinz et al., 2022). The approach that we propose here could be developed for this purpose. First, the Poisson law proposed to model the number of artifacts that a given year contributes to the archaeological record can be extended to other types of dated archaeological remains. The key to this is to properly model the uncertainty about the age of those remains (*i.e.* the equivalent to the calibration curve for the radiocarbon samples). For dating methods based on the natural sciences (radiocarbon and optically stimulated luminescence dating) there is a wealth of information about how to model those uncertainties. For other methods of assigning dates (numismatic, aoristic approaches) proper statistical models can also be proposed (*e.g.* Crema, 2024).

Analyzing the data under an ABC approach may also facilitate the combination with other sources of information such as genetic diversity. ABC is widely used in population genetics to obtain demographic inferences, including in studies using ancient DNA from prehistoric sites. Simulating both archaeological and genetic data based on the same demographic trajectory can be envisioned and would allow the combination of two disparate sources of information. Nevertheless, it must be noted that population genetics "demographic" inference provides a measure of genetic drift (the so-called effective population size) rather than census population size. As with for the abundance of radiocarbon samples, there are good reasons to assume that the effective population size offers information about demography, but the interpretation of results should bear in mind the limits of the data and the models used. For instance, there is ancient DNA data from Britain and Ireland (*e.g.* Patterson et al., 2022) that could be used jointly with the radiocarbon data from Bevan et al. (2017) to infer demography. However, the genetic structure and admixture of those people would need to be taken into account to disentangle their effects on genetic diversity from that of population size. Studies trying to combine archaeological and genetic data will need to address the question of whether they are indeed inferring a common

process, two separate processes with irreconcilable differences, or, more likely, an intermediate situation.

## Conclusion

The analysis of the change in frequency of radiocarbon samples though time is an attractive and useful approach to address diverse questions in archaeology and other sciences of the past. A key development for that analysis was the SPD, which allows to visualize the abundance of radiocarbon samples in the natural (calibrated) time scale, yet it lacks a formal mathematical definition. We argue that the ~~SPD has constrained later developments on the statistical inference of~~ use of the SPD has lead recent model-based approaches to conceptualize the abundance of radiocarbon samples ~~, misleading them to suboptimal concepts~~ as a probability distribution, which we consider a suboptimal model for the underlying data~~and models~~. We propose a new model for the conceptualization of abundances of radiocarbon samples, allowing powerful statistical inference of parameters that have a natural interpretation, as the number of expected samples contributed by each year to the total.

## Acknowledgements

This work has greatly benefited from insightful discussions with several people, and we would like to express our gratitude to Jan Apel, Kristian Brink, Magdalena Fraser, Anders Högberg, Helena Malmström, and Rita Peyroteo Stjerna for their valuable feedback on various aspects of radiocarbon data analysis. We would like to thank the input given by Jonathan Hanna, Thomas Huet and one anonymous reviewer on the first submission of our manuscript, as well as the feedback on the preprint given by Michael Holton Price. Additionally, we acknowledge the collaborative efforts facilitated by the memorandum of understanding between Uppsala University (~~IOB and EUG departments~~department of Organismal Biology and department of Ecology and Genetics) and the INRAE (UMR CBGP). This agreement has played a pivotal role in strengthening the collaboration between these institutions and has significantly contributed to the advancement of this research.

## Funding

This project has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 791695 (TimeAdapt).

## Conflict of interest disclosure

The authors of this article declare that they have no financial conflict of interest with the content of this article. Miguel de Navascués and Concetta Burgarella are recommenders for PCI Evolutionary Biology.

## Data, script, code, and supplementary information availability

Scripts to reproduce the analyses presented here are available at Zenodo (de Navascués, 2024). Data used in this work are from previous publications and were made available by Bevan (2017).

## Author contributions

MdN (Conceptualization, Formal analysis, Funding acquisition, Investigation, Methodology, Project administration, Software, Validation, Writing — original draft), CB (Conceptualization, Writing — review & editing), MJ (Conceptualization, Funding acquisition, Resources, Writing — review & editing).

## References


Bevan A (2017). *Radiocarbon Dataset and Analysis from Bevan, A., Colledge, S., Fuller, D., Fyfe, R., Shennan, S. and C. Stevens 2017. Holocene fluctuations in human population demonstrate repeated links to food production and climate*. UCL Discovery. https://doi.org/10.14324/000.ds.10025178.

Bevan A, Colledge S, Fuller D, Fyfe R, Shennan S, Stevens C (2017). *Holocene fluctuations in human population demonstrate repeated links to food production and climate*. *Proceedings of the National Academy of Sciences* **114**, E10524–E10531. https://doi.org/10.1073/pnas.1709190114.

Bird D, Miranda L, Vander Linden M, Robinson E, Bocinsky RK, Nicholson C, Capriles JM, Finley JB, Gayo EM, Gil A, Guedes J, Hoggarth JA, Kay A, Loftus E, Lombardo U, Mackie M, Palmisano A, Solheim S, Kelly RL, Freeman J (2022). *p3k14c, a synthetic global database of archaeological radiocarbon dates*. *Scientific Data* **9**, 27. https://doi.org/10.1038/s41597-022-01118-7.

Boitard S, Rodríguez W, Jay F, Mona S, Austerlitz F (2016). *Inferring population size history from large samples of genome-wide molecular data - An approximate Bayesian computation approach*. *PLOS Genetics* **12**. Ed. by Mark A Beaumont, e1005877. https://doi.org/10.1371/journal.pgen.1005877.

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

Broughton JM, Weitzel EM (2018). *Population reconstructions for humans and megafauna suggest mixed causes for North American Pleistocene extinctions*. *Nature Communications* **9**, 5441. https://doi.org/10.1038/s41467-018-07897-1.

Carleton WC (2021). *Evaluating Bayesian Radiocarbon-dated Event Count (REC) models for the study of long-term human and environmental processes*. *Journal of Quaternary Science* **36**, 110–123. https://doi.org/10.1002/jqs.3256.

Carleton WC, Groucutt HS (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**, 630–643. https://doi.org/10.1177/0959683620981700.

Contreras DA, Codding BF (2023). *Landscape Taphonomy Predictably Complicates Demographic Reconstruction*. *Journal of Archaeological Method and Theory*. https://doi.org/10.1007/s10816-023-09634-5.

Crema ER (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**, 1387–1418. https://doi.org/10.1007/s10816-022-09559-5.

Crema ER (2024). *A Bayesian alternative for aoristic analyses in archaeology*. *Archaeometry*. https://doi.org/10.1111/arcm.12984.

Crema ER, Bevan A (2021). *Inferences from large sets of radiocarbon dates: software and methods*. *Radiocarbon* **63**, 23–39. https://doi.org/10.1017/RDC.2020.95.

Crema ER, Kobayashi K (2020). *A multi-proxy inference of Jōmon population dynamics using bayesian phase models, residential data, and summed probability distribution of $^{14}$C dates*. *Journal of Archaeological Science* **117**, 105136. https://doi.org/10.1016/j.jas.2020.105136.

Crema ER, 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**, e0251695. https://doi.org/10.1371/journal.pone.0251695.

de Navascués M (2024). *DARth ABC: Analysis of the Dated Archaeological Record through Approximate Bayesian Computation (v1.0.0)*. *Zenodo*. https://doi.org/10.5281/zenodo.7560876.

DiNapoli RJ, Crema ER, Lipo CP, Rieth TM, Hunt TL (2021). *Approximate Bayesian Computation of radiocarbon and paleoenvironmental record shows population resilience on Rapa Nui (Easter Island)*. *Nature Communications* **12**, 3939. https://doi.org/10.1038/s41467-021-24252-z.

Gaujoux R (2023). *doRNG: Generic reproducible parallel backend for 'foreach' loops*. R package version 1.8.6. URL: https://CRAN.R-project.org/package=doRNG.

Geyh MA (1980). *Holocene Sea-Level History: Case Study of the Statistical Evaluation of $^{14}$C Dates*. *Radiocarbon* **22**, 695–704. https://doi.org/10.1017/S0033822200010067.

Harrell Jr FE (2022). *Hmisc: Harrell Miscellaneous*. R package version 4.7-2. URL: https://CRAN.R-project.org/package=Hmisc.

Hinz M, Roe J, Laabs J, Heitz C, Kolář J (2022). *Bayesian inference of prehistoric population dynamics from multiple proxies: a case study from the North of the Swiss Alps*. https://doi.org/10.31235/osf.io/dbcag.

Komsta L, Novomestky F (2022). *moments: Moments, Cumulants, Skewness, Kurtosis and Related Tests*. R package version 0.14.1. URL: https://CRAN.R-project.org/package=moments.

Lannelongue L, Grealey J, Bateman A, Inouye M (2021). *Ten simple rules to make your computing more environmentally sustainable*. *PLOS Computational Biology* **17**, e1009324. https://doi.org/10.1371/journal.pcbi.1009324.

Libby WF, Anderson EC, Arnold JR (1949). *Age Determination by Radiocarbon Content: World-Wide Assay of Natural Radiocarbon*. *Science* **110**, 678–680. https://doi.org/10.1126/science.109.2827.227. (Visited on 03/14/2022).

Marin JM, Raynal L, Pudlo P, Robert CP, Estoup A (2022). *abcrf: Approximate Bayesian Computation via Random Forests*. R package version 1.9. URL: https://CRAN.R-project.org/package=abcrf.

Marom N, Wolkowski U (2024). *A note on predator-prey dynamics in radiocarbon datasets*. *Peer Community Journal* **4**. https://doi.org/10.24072/pcjournal.395.

Microsoft Corporation, Weston S (2022a). *doSNOW: Foreach Parallel Adaptor for the 'snow' Package*. R package version 1.0.20. URL: https://CRAN.R-project.org/package=doSNOW.

Microsoft Corporation, Weston S (2022b). *doParallel: Foreach Parallel Adaptor for the 'parallel' Package*. R package version 1.0.17. URL: https://CRAN.R-project.org/package=doParallel.

Pasek J (2021). *weights: Weighting and Weighted Statistics*. R package version 1.0.4. URL: https://CRAN.R-project.org/package=weights.

Patterson N, Isakov M, Booth T, Büster L, Fischer CE, Olalde I, Ringbauer H, Akbari A, Cheronet O, Bleasdale M, Adamski N, Altena E, Bernardos R, Brace S, Broomandkhoshbacht N, Callan K,

Candilio F, Culleton B, Curtis E, Demetz L, et al. (2022). *Large-scale migration into Britain during the Middle to Late Bronze Age*. *Nature* **601**, 588–594. https://doi.org/10.1038/s41586-021-04287-4.

Pierce JL, Meyer GA, Timothy Jull AJ (2004). *Fire-induced erosion and millennial-scale climate change in northern ponderosa pine forests*. *Nature* **432**, 87–90. https://doi.org/10.1038/nature03058.

Porčić M, Blagojević T, Pendić J, Stefanović S (2020). *The Neolithic Demographic Transition in the Central Balkans: population dynamics reconstruction based on new radiocarbon evidence*. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20190712. https://doi.org/10.1098/rstb.2019.0712.

Pudlo P, Marin JM, Estoup A, Cornuet JM, Gautier M, Robert CP (2016). *Reliable ABC model choice via random forests*. *Bioinformatics* **32**, 859–866. https://doi.org/10.1093/bioinformatics/btv684.

R Core Team (2021). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing. Vienna, Austria. URL: https://www.R-project.org/.

Raynal L, Marin JM, Pudlo P, Ribatet M, Robert CP, Estoup A (2019). *ABC random forests for Bayesian parameter inference*. *Bioinformatics* **35**, 1720–1728. https://doi.org/10.1093/bioinformatics/bty867.

Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Ramsey CB, Butzin M, Cheng H, Edwards RL, Friedrich M, Grootes PM, Guilderson TP, Hajdas I, Heaton TJ, Hogg AG, Hughen KA, Kromer B, Manning SW, Muscheler R, Palmer JG, et al. (2020). *The IntCal20 Northern Hemisphere Radiocarbon Age Calibration Curve (0–55 cal kBP)*. *Radiocarbon* **62**, 725–757. https://doi.org/10.1017/RDC.2020.41.

Rick JW (1987). *Dates as Data: An Examination of the Peruvian Preceramic Radiocarbon Record*. *American Antiquity* **52**, 55–73. https://doi.org/10.2307/281060.

Rousset F, Gouy A, Martinez-Almoyna C, Courtiol A (2017). *The summary-likelihood method and its implementation in the Infusion package*. *Molecular Ecology Resources* **17**, 110–119. https://doi.org/10.1111/1755-0998.12627.

Shennan S, Downey SS, Timpson A, Edinborough K, Colledge S, Kerig T, Manning K, Thomas MG (2013). *Regional population collapse followed initial agriculture booms in mid-Holocene Europe*. *Nature Communications* **4**, 2486. https://doi.org/10.1038/ncomms3486.

Sunnåker M, Busetto AG, Numminen E, Corander J, Foll M, Dessimoz C (2013). *Approximate Bayesian Computation*. *PLOS Computational Biology* **9**, e1002803. https://doi.org/10.1371/journal.pcbi.1002803.

Taylor RE (1995). *Radiocarbon dating: The continuing revolution*. *Evolutionary Anthropology: Issues, News, and Reviews* **4**, 169–181. https://doi.org/10.1002/evan.1360040507.

Thorndycraft VR, Benito G (2006). *The Holocene fluvial chronology of Spain: evidence from a newly compiled radiocarbon database*. *Quaternary Science Reviews* **25**, 223–234. https://doi.org/10.1016/j.quascirev.2005.07.003.

Timpson A, Barberena R, Thomas MG, Méndez C, Manning K (2020). *Directly modelling population dynamics in the South American Arid Diagonal using $^{14}C$ dates*. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**, 20190723. https://doi.org/10.1098/rstb.2019.0723.

Williams AN (2012). *The use of summed radiocarbon probability distributions in archaeology: a review of methods*. *Journal of Archaeological Science* **39**, 578–589. https://doi.org/10.1016/j.jas.2011.07.014.

Wolodzko T (2020). *extraDistr: Additional Univariate and Multivariate Distributions*. R package version 1.9.1. URL: https://CRAN.R-project.org/package=extraDistr.

# Supplementary Materials

## S.1. An interpretation of Sum of Probability Distributions (SPD) as the expected number of samples per year

As far as we know, a formal mathematical interpretation of the SPD is lacking, as the algorithm defining it lacks a rigorous mathematical justification for aggregating independent probability distributions. Nevertheless, we find it more intuitive to interpret the SPD as the expected number of samples from each year rather than as a probability distribution. Consider the meaning of the SPD value at a specific year. Each sample in the radiocarbon data set has a probability of being 'sampled' in that year, given by the posterior probability distribution from the calibration process. Consequently, each sample actually originating from that year can be viewed as a 'success' and samples from other years as 'failures' akin to independent binomial trials. The sum of these successes, *i.e.*, the number of samples from that year, follows a Poisson-binomial distribution. The expected value of a Poisson-binomial distribution (*i.e.* the expected number of samples at that year) is the sum of the probabilities of each trial, representing the SPD value for that year. This rationale provides the insight that the SPD somehow quantifies the expected number of samples for each year, suggesting that the analysis of radiocarbon abundance data should focus on modeling the number of samples at each year. However, SPD values from different years are not independent, rendering the Poisson-binomial model inapplicable to the entire SPD. Remarkably, our proposed model, utilizing Poisson distributions to model the number of samples at each year, yields inferences closely aligning with SPD values (figure 6).

**Table S1 – Notation[a]**

| | meaning |
|---|---|
| CRA | conventional radiocarbon age |
| $H_u$ | number of samples in an interval starting at uncalibrated year $u$ and ending at uncalibrated year $u+\delta$ |
| $k$ | upper bound value of $\lambda$ under a model of logistic change |
| $n_t$ | number of objects in year $t$ that can potentially become a radiocarbon sample in the data set |
| $\mathcal{N}$ | normal or Gaussian distribution |
| $p_t$ | probability of an object to become a sample in the radiocarbon data at year $t$ |
| $q$ | ratio between $\lambda$ of subset a ($\lambda^a$) and $\lambda$ for the total dataset |
| $r$ | growth rate of $\lambda$ under a model of exponential or logistic change |
| $\mathbf{R}$ | vector of number of radiocarbon samples for each year between $t_{\text{min}}$ and $t_{\text{max}}$ |
| $\mathbf{R}^a$ | vector of number of radiocarbon samples for each year between $t_{\text{min}}$ and $t_{\text{max}}$ for the subset of samples $a$ |
| $R_t$ | number of radiocarbon samples at year $t$, an element of vector $\mathbf{R}$ |
| $\mathbf{R'}$ | vector of number of radiocarbon samples with CRAs between $u_{\text{min}}$ and $u_{\text{max}}$ |
| $R'_u$ | number of radiocarbon samples with CRA $= u$, an element of vector $\mathbf{R'}$ |
| $t$ | time in years (calibrated) |
| $T$ | total number of samples in the radiocarbon data |
| $u$ | uncalibrated radiocarbon year (measurement unit for CRA) |
| $\delta$ | size of the interval of uncalibrated years used to calculate summary statistics of the data |
| $\Delta H_u$ | Difference between values $H_u$ and $H_{u+\delta}$ |
| $\boldsymbol{\lambda}$ | vector of expected number of radiocarbon samples for each year between $t_{\text{min}}$ and $t_{\text{max}}$ |
| $\lambda_t$ | expected number of radiocarbon samples at year $t$, an element of vector $\boldsymbol{\lambda}$ |
| $\pi_t$ | probability that the age of a sample is $t$ |

[a] We follow the convention of marking vectors with bold font.

**Table S2** – ~~Confusion matrix~~Confusion matrices

| true model | prediction | | | error rate |
|---|---|---|---|---|
| | exponential | logistic | piecewise | |
| exponential | 15067 | 14873 | 60 | 0.498 |
| logistic | 4631 | 25277 | 92 | 0.157 |
| piecewise | 132 | 324 | 29544 | 0.015 |
| | independent | interdependent | parallel | |
| independent | 26348 | 2480 | 588 | 0.104 |
| interdependent | 980 | 25290 | 3146 | 0.140 |
| parallel | 619 | 2479 | 26318 | 0.105 |

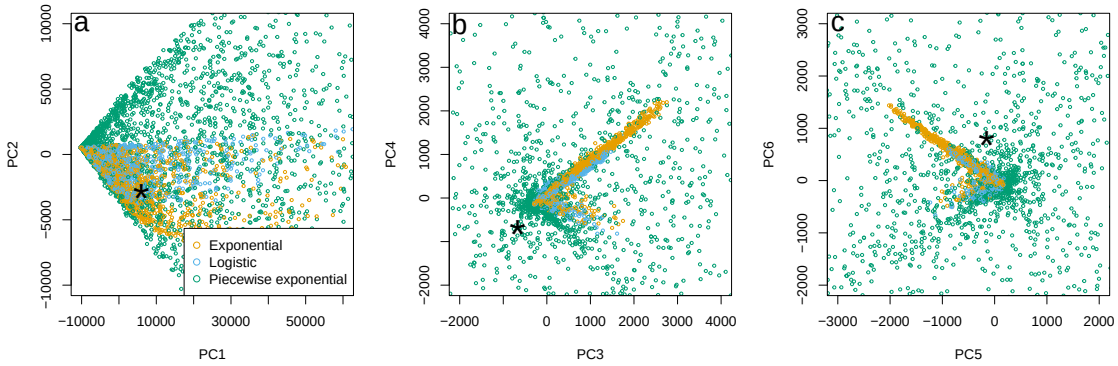


**Figure S1** – **PCA for goodness-of-fit evaluation.** PC values from 3000 randomly selected simulations are plotted for each model for the first six axes. The projection of the observed summary statistics is represented by an asterisk (*). The first six principal components capture 97.06% of the variance in the data and are presented by consecutive pairs in panels (a), (b) and (c).

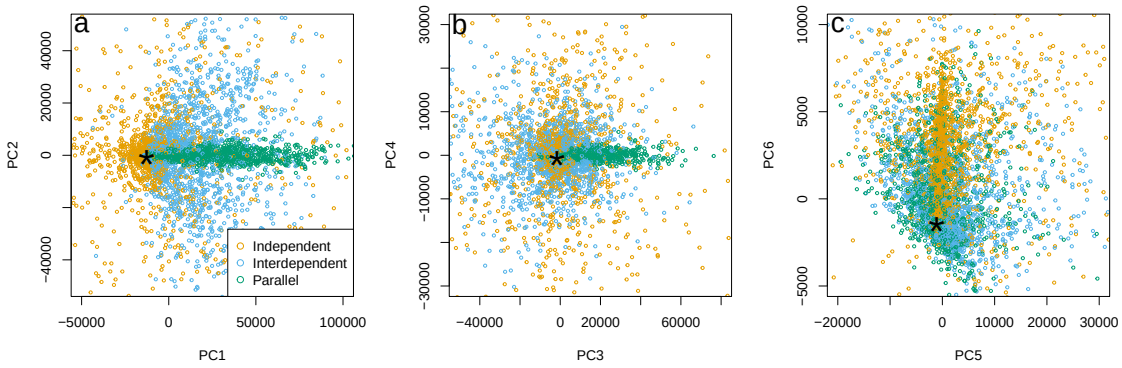


**Figure S2** – **PCA for goodness-of-fit evaluation, cereals.** PC values from 3000 randomly selected simulations are plotted for each model. The projection of the observed summary statistics is represented by an asterisk (*). The first six principal components capture 99.60% of the variance in the data and are presented by consecutive pairs in panels (a), (b) and (c).