Maximum spacing estimation of chronological boundaries in discrete and uncertainty-prone age sequences

Shi-Yong Yu

Key Laboratory of Regional Sustainable Development System Analysis and Simulation in Higher Education Institutions, School of Geography, Geomatics, and Planning, Jiangsu Normal University, Xuzhou, 221116, China

A B S T R A C T

Estimating the endpoints of age sequences is a crucial task in archaeological and geological sciences. Here we advance this process by addressing three major limitations in previous approaches. First, we introduce a maximum spacing estimation method to simplify the conventional maximum likelihood estimation approach. Second, we apply Monte Carlo simulations to account for uncertainties in laboratory-derived ages. Third, we utilize a range of probability distributions to manage sampling variability, improving the accuracy and reliability of chronological inferences. This method is versatile, applicable not only for estimating settlement time from smaller datasets but also for determining the timing of rise and fall of cultures from larger datasets. Rigorous testing on both simulated and real-world chronological data demonstrates the practical utility and robustness of this method in handling discrete and uncertainty-prone age sequences. Comparative analysis shows that different statistical models significantly impact the estimation of chronological boundaries. Uniform and exponential models provide more constrained estimates with higher confidence, while normal and log-normal models introduce greater uncertainty. These wider intervals may reflect underlying uncertainties, such as stratigraphic variability or mixed-age samples, which restrictive models may overlook. We also highlight the sensitivity of age sequence estimates to data size, with important implications for interpreting the temporal boundaries of archaeological and geological events. Therefore, researchers are encouraged to carefully assess the nature and frequency distribution of their chronological data before considering the narrower estimates from uniform and exponential models as well as the broader intervals from normal and log-normal models to build more reliable chronological frameworks.

1. Introduction

Estimating chronological boundaries from discrete and uncertainty-prone age sequences presents a clear challenge in archaeological and geological research due to the sporadic and imprecise nature of chronological data (Holland-Lulewicz and Ritchison, 2021). Discrete age sequences, often derived from radiocarbon dating, stratigraphy, and/or other dating methods, provide only limited and sometimes ambiguous information about the start, end, and span of archaeological and geological phenomena. Inferring the endpoints from these age sequences can be further affected by various uncertainties, including dating errors, calibration issues, and sampling biases (Saltré et al., 2015). For example, analogous to the Signor-Lipps effect in paleontology (Signor and Lipps, 1982), where the apparent last appearance of a species can be skewed by the incompleteness of the fossil record, the onset of an archaeological phenomenon will have occurred earlier than the earliest dated samples, while the termination will have happened more recently than any of the samples dated. This means that dated samples provide only a partial view of the actual timing of archaeological events. Therefore, resolving this temporal discrepancy is crucial in both archaeological and anthropological sciences, as it underpins our understanding of various phenomena ranging from the timing of site occupation and abandonment (Yu et al., 2021) to the rise and fall of regional cultures (Zheng et al., 2021) as well as the accurate interpretation of the across-continental dispersal of modern humans (Becerra-Valdivia and Higham, 2020; Saltré et al., 2024).

Many attempts have been made to address these challenges and sophisticated statistical and computational techniques have been developed to estimate chronological boundaries (Buck et al., 1992; Steier and Rom, 2000; Zeeden et al., 2018). Statistical techniques diverge into two primary approaches: Frequentist and Bayesian. Bayesian chronological modeling, such as OxCal (Bronk Ramsey, 1995, 2000), BCal (Buck et al.,

1999), and MatCalib (Yu, 2022), allow for the integration of discrete data points with their associated uncertainties, providing a probabilistic framework for inferring the timing of an interval within which the ages most likely fall, even when the data are incomplete or prone to error. In contrast to Bayesian statistics, which treats parameters as random variables with probability distributions, Frequentist statistics considers parameters to be fixed but unknown quantities. As such, Frequentist methods estimate these parameters based on sample data (Solow, 1993, 1997; Herrando-Pérez and Saltré, 2024), often using techniques such as maximum likelihood estimation (MLE), and focus on making inferences based on the long-run frequency properties of the estimators. For example, Solow (1997) proposed that laboratory-derived ages could be modeled using a normal distribution, with the mean representing the true age and the standard deviation reflecting the dating uncertainty. This approach leads to a normal likelihood function for the true ages given the laboratory ages. However, integrating this function over a parametric temporal interval yields an expression that is analytically complex. Additionally, even after applying a logarithmic transformation, this integrated likelihood function remains non-concave, complicating the assurance of finding an analytical solution. Unconstrained optimization using numerical methods often yields meaningless solutions, whereas constrained optimization consistently provides results close to the minimum and maximum of the age sequence. Therefore, this complexity in the MLE-based inference of chronological boundaries necessitates the use of alternative robust approaches.

Maximum spacing estimation (MSE) is a robust statistical technique that has been widely used to estimate distribution parameters by maximizing the spacing between observed data points (Ranneby, 1984; Anatolyev and Kosenok, 2005). It is particularly effective for continuous

We apply MSE, a powerful and versatile method that has been increasingly used for parameter estimation across various statistical contexts (Ekström and Inference, 2008). This method finds the parameter values that make the data points as evenly spaced as possible. This gives a simple, tidy solution. Unlike the MLE method, which looks for the parameters that make the observed data most probable, MSE focuses on the spacing pattern rather than the likelihood (Cheng and Amin, 1983). Without loss of generality, let $F(t; \theta)$ denote the cumulative distribution function of the true age t. Here t is a random variable defined over (or truncated to) the temporal interval $[\alpha, \beta]$, and θ represents the set of parameters of the distribution (θ is empty for a non-parametric distribution). Following Ranneby (1984), we define the one-step spacing, $d_i(\alpha, \beta; \theta)$, as the gap between the values of the cumulative distribution function at two consecutive age points

$$d_i(\alpha,\beta;\theta) = \begin{cases} \dfrac{F(x_i;\theta) - F(\alpha;\theta)}{F(\beta;\theta) - F(\alpha;\theta)} & , \quad i = 1 \\[2mm] \dfrac{F(x_i;\theta) - F(x_{i-1};\theta)}{F(\beta;\theta) - F(\alpha;\theta)} & , \quad i = 2, \cdots, n, \\[2mm] \dfrac{F(\beta;\theta) - F(x_{i-1};\theta)}{F(\beta;\theta) - F(\alpha;\theta)} & , \quad i = n+1 \end{cases} \tag{1}$$

Clearly, $\sum_{i=1}^{n+1} d_i(\alpha, \beta; \theta) = 1$. To estimate the endpoints α and β, we define product spacings, $D(\alpha, \beta; \theta)$, as the geometric mean of the gaps

$$D(\alpha,\beta;\theta) = \left[\prod_{i=1}^{n+1} d_i(\alpha,\beta;\theta) \right]^{\frac{1}{n+1}}. \tag{2}$$

Taking the natural logarithm on both sides, this expression becomes

$$\ln D(\alpha,\beta;\theta) = \frac{1}{n+1} \sum_{i=1}^{n+1} \ln[d_i(\alpha,\beta;\theta)] = \frac{1}{n+1} \left[\begin{array}{l} \ln(F(x_1;\theta) - F(\alpha;\theta)) + \sum_{i=2}^{n} \ln(F(x_i;\theta) - F(x_{i-1};\theta)) + \\ \ln(F(\beta;\theta) - F(x_n;\theta)) - (n+1)\ln(F(\beta;\theta) - F(\alpha;\theta)) \end{array} \right]. \tag{3}$$

distributions and offers advantages in handling small sample sizes due to its reduced sensitivity to outliers, providing analytically elegant parameter estimates that are often more reliable and less biased compared to other methods. In this study, we derive point estimates of chronological boundaries for discrete and uncertainty-prone age sequences with an MSE-based approach. The issue of sampling biases is comprehensively explored by considering a range of underlying distributions to assess their impact on parameter estimation. To address dating uncertainty, we construct credible intervals through Monte Carlo simulations, which enhance the robustness and reliability of our inference on chronological boundaries by incorporating and quantifying the inherent uncertainty in the data. The utility and effectiveness of this method are rigorously tested using both simulated datasets and real-world archaeological data, demonstrating its practical applicability and robustness in handling highly discrete and uncertain age sequences.

2. Description of the maximum spacing estimation method

Let $\mathbf{x} = \{x_1 < x_2, \cdots, < x_n\}$ denote a collection of ordered numerical ages in calendar years, obtained through various chronometric techniques such as radiocarbon dating, optically stimulated luminescence (OSL), and U-Th methods. The corresponding dating uncertainties are represented by $\sigma = \{\sigma_1, \sigma_2, \cdots, \sigma_n\}$. These ages are assumed to be independent and identically distributed (iid) within an interval $[\alpha, \beta]$, representing a limited number of observations of the uncertainty-averse true ages denoted as $\mathbf{t} = \{t_1 < t_2, \cdots, < t_n\}$. In this analysis, our goal is to estimate the endpoints of this interval (i.e., α and β) within which the true ages corresponding to $\mathbf{x}$ are likely to fall.

Our objective is to obtain robust point estimates for α and β by maximizing $\ln D(\alpha, \beta; \theta)$. Differentiating Eqn. (3) with respect to α and β, respectively, yields

$$\begin{cases} \dfrac{\partial}{\partial \alpha} \ln D(\alpha,\beta;\theta) = \dfrac{1}{n+1} \left[\dfrac{-f(\alpha;\theta)}{F(x_1;\theta) - F(\alpha;\theta)} \right] + \dfrac{f(\alpha;\theta)}{F(\beta;\theta) - F(\alpha;\theta)} \\[3mm] \dfrac{\partial}{\partial \beta} \ln D(\alpha,\beta;\theta) = \dfrac{1}{n+1} \left[\dfrac{f(\beta;\theta)}{F(\beta;\theta) - F(x_n;\theta)} \right] - \dfrac{f(\beta;\theta)}{F(\beta;\theta) - F(\alpha;\theta)} \end{cases}. \tag{4}$$

Equating $\frac{\partial}{\partial \alpha} \ln D(\alpha,\beta;\theta)$ and $\frac{\partial}{\partial \beta} \ln D(\alpha,\beta;\theta)$ with 0 in Eqn. (4) and solving these equations simultaneously gives

$$\begin{cases} F(\alpha;\theta) = \dfrac{1}{n-1}(nF(x_1;\theta) - F(x_n;\theta)) \\[3mm] F(\beta;\theta) = \dfrac{1}{n-1}(nF(x_n;\theta) - F(x_1;\theta)) \end{cases}. \tag{5}$$

Therefore, the optimal values for α and β that maximize $\ln D(\alpha, \beta; \theta)$ are given by

$$\begin{cases} \alpha = F^{-1}\left(\dfrac{1}{n-1}(nF(x_1;\theta) - F(x_n;\theta)); \theta \right) \\[3mm] \beta = F^{-1}\left(\dfrac{1}{n-1}(nF(x_n;\theta) - F(x_1;\theta)); \theta \right) \end{cases}, \tag{6}$$

where $F^{-1}(\cdot)$ is the inverse cumulative distribution function of the true age t.

The above formulation provides a general framework for estimating

chronological boundaries from age sequences, applicable to both parametric and non-parametric distributions. Although a closed-form expression of the cumulative distribution is not available for a non-parametric model (θ is empty in this case), point estimates of endpoints α and β can still be obtained from Eqn. (6) by linearly extrapolating the empirical inverse cumulative distribution. In the following section, we will explore endpoint estimations for the true ages across various parametric distributions. Each distribution has its own characteristics and assumptions, influencing the estimation process and robustness of the results. By deriving analytical estimates for the lower and upper bounds (i.e., α and β) of the age sequences, we aim to ensure that the estimation process is resilient to variations in the probabilistic model and data, thereby maintaining reliability across different distributional forms (Appendix).

3. Distributional framework

3.1. Uniform distribution

The uniform distribution assumes that all outcomes of the true age t equally likely fall in a parametric interval $[\alpha, \beta]$. For the true ages distributed uniformly over this unknown interval, the estimation process is relatively straightforward. Elaborative algebraic work (Appendix) yields point estimates for α and β

$$
\begin{cases}
\alpha = \dfrac{1}{n-1}(nx_1 - x_n) \\
\beta = \dfrac{1}{n-1}(nx_n - x_1)
\end{cases}.
\tag{7}
$$

3.2. Exponential distribution

The exponential distribution is used for modeling the time between events in a Poisson process. It is defined by a single rate parameter $\lambda = \frac{1}{n}\sum_{i=1}^{n}x_i$. This distribution is often applied in contexts where events are assumed to occur continuously and independently over the temporal interval $[\alpha, \beta]$. Endpoint estimation under this distribution requires careful handling of the rate parameter to ensure accurate interval estimates. Elaborative algebraic work (Appendix) yields point estimates for α and β

$$
\begin{cases}
\alpha = -\lambda \ln\left(\dfrac{1}{n-1}\left(n\exp\left(-\dfrac{x_1}{\lambda}\right) - \exp\left(-\dfrac{x_n}{\lambda}\right)\right)\right) \\
\beta = -\lambda \ln\left(\dfrac{1}{n-1}\left(n\exp\left(-\dfrac{x_n}{\lambda}\right) - \exp\left(-\dfrac{x_1}{\lambda}\right)\right)\right)
\end{cases}.
\tag{8}
$$

3.3. Normal distribution

The normal distribution is characterized by its mean and standard deviation given as $\mu = \frac{1}{n}\sum_{i=1}^{n}x_i$ and $\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(x_i - \mu)^2}$. It is a common choice due to its simplicity and well-understood properties. For the true ages following a normal distribution truncated to a parametric interval $[\alpha, \beta]$, the estimation involves assumptions about the mean and variance known from the laboratory-derived ages. Therefore, the resultant estimates of the interval are sensitive to these parameters. Elaborative algebraic work (Appendix) yields point estimates for α and β

$$
\begin{cases}
\alpha = \mu + \sigma\Phi^{-1}\left(\dfrac{1}{n-1}\left(n\Phi\left(\dfrac{x_1-\mu}{\sigma}\right) - \Phi\left(\dfrac{x_n-\mu}{\sigma}\right)\right)\right) \\
\beta = \mu + \sigma\Phi^{-1}\left(\dfrac{1}{n-1}\left(n\Phi\left(\dfrac{x_n-\mu}{\sigma}\right) - \Phi\left(\dfrac{x_1-\mu}{\sigma}\right)\right)\right)
\end{cases},
\tag{9}
$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ is the cumulative distribution function and inverse cumulative distribution function of the normal distribution, respectively.

3.4. Log-normal distribution

When the natural logarithm of the true age t follows a normal distribution, the mean and standard deviation are defined as $\mu = \frac{1}{n}\sum_{i=1}^{n}\ln x_i$ and $\sigma = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}(\ln x_i - \mu)^2}$, respectively. This distribution is useful for modeling skewed data within a parametric interval $[\alpha, \beta]$. After transforming the data into the logarithmic space (Appendix), applying the MSE method gives the optimal values for α and β

$$
\begin{cases}
\alpha = \exp\left[\mu + \sigma\Phi^{-1}\left(\dfrac{1}{n-1}\left(n\Phi\left(\dfrac{\ln x_1-\mu}{\sigma}\right) - \Phi\left(\dfrac{\ln x_n-\mu}{\sigma}\right)\right)\right)\right] \\
\beta = \exp\left[\mu + \sigma\Phi^{-1}\left(\dfrac{1}{n-1}\left(n\Phi\left(\dfrac{\ln x_n-\mu}{\sigma}\right) - \Phi\left(\dfrac{\ln x_1-\mu}{\sigma}\right)\right)\right)\right]
\end{cases},
\tag{10}
$$

where $\Phi(\cdot)$ and $\Phi^{-1}(\cdot)$ is the cumulative distribution function and inverse cumulative distribution function of the normal distribution, respectively.

4. Implementation and usage instructions

We have developed *MatEndpt*, a MATLAB package for maximum spacing estimation of chronological boundaries of an age sequence. The package features the following functionality: (1) Data handling. *MatEndpt* includes a tab-delimited text file, which users can customize with their own data. Users can specify the age type (e.g., ^{14}C, OSL, U-series), the reservoir age (R) or reservoir age offset (ΔR), the calibration curve for ^{14}C ages, and the year that samples are dated. For ^{14}C ages, *MatEndpt* supports calibration for both marine and terrestrial samples from different hemispheres. Note that calibrations for post-bomb (post-AD, 1950) ^{14}C ages are not supported. For marine ^{14}C ages, if R is provided, a terrestrial calibration curve should be specified; otherwise, a marine curve should be used. (2) Data preprocessing. After importing the data into the working space, *MatEndpt* processes them by correcting for any reservoir effect, converting conventional ^{14}C ages to the F^{14}C (fraction modern) space (Reimer et al., 2004), and transforming non-^{14}C ages to the BP (before present, with present referring to AD, 1950) scale. (3) Monte Carlo simulations. Numerical ages are derived from various chronometric techniques, with uncertainties typically reported as 1σ standard deviations by laboratories. The Frequentist statistical method (Solow, 1997) assumes that errors in each age follow a normal distribution, which is generally assumed for non-^{14}C ages. However, ^{14}C age calibrations produce a probability density function (PDF) rather than a point estimate, often resulting in asymmetric and multimodal distributions (Guilderson et al., 2005). Therefore, dating errors may not follow a normal distribution in this case (Andrés Christen and Pérez E, 2009; Yu, 2022). Monte Carlo simulations are used to tackle the nonnormality of the calibrated ^{14}C ages. For datasets with ^{14}C ages, the calibration curves are first loaded and then the ^{14}C ages and errors corresponding to the calendar ages in the curve are converted to the F^{14}C space. A unique probability distribution of calendar ages is generated for each ^{14}C age. In total, 20,000 age sequences are produced by resampling from these PDFs for ^{14}C ages and from a theoretically normal distribution for non-^{14}C ages. (4) Calculating endpoints of age sequences. Endpoints for each simulated age sequence are calculated using Eqn. (6)−10 for a non-parametric, uniform, exponential, normal, and log-normal distribution, respectively. Since ages closer to the beginning and end of the sequence contribute more significantly to the estimate of the chronological boundaries, temporal gaps are down-weighted as described by Herrando-Pérez and Saltré (2024). Empirical PDFs of the estimated endpoints are constructed through numerically differentiating their empirical cumulative distribution (Yu, 2024). This allows for the generation of point estimates such as mean age, age at median probability, and modal age (Telford et al., 2004), as well as 1σ and 2σ credible intervals (i.e., the 68.2 % and 95.4 % highest posterior density regions) for

the endpoints (Hyndman, 1996).

Upon completion of the Monte Carlo simulations, *MatEndpt* outputs the following results: probability density functions, point estimates, and credible intervals for both the estimated endpoints and the simulated calendar ages. The output is organized as a list containing three objects: (1) point estimates and credible intervals for the lower bound of the age sequence, (2) point estimates and credible intervals for the upper bound, and (3) point estimates and credible intervals for the simulated calendar ages. Users can visualize the estimated endpoints by plotting their PDFs alongside the 68.2 % and 95.4 % highest posterior density regions. These visualizations can be displayed either individually or in conjunction with the PDFs of the simulated calendar ages. The duration of settlement at a site can be computed by subtracting the lower endpoint from the upper endpoint using the random samples generated during the Monte Carlo step. This process facilitates the construction of an empirical PDF and the calculation of point estimates and credible intervals. *MatEndpt* supports three age scales: (1) BP (before present, with present referring to AD, 1950), (2) B2K (before AD, 2000), and (3) BC/AD. Conversion between these age scales is straightforward, as detailed by Yu (2022).

5. Comparative analyses

5.1. Sensitivity analysis using synthesized data

5.1.1. Sensitivity to underlying distributions

We evaluate the sensitivity of endpoint estimates to variations in the parameters of underlying distributions and sample size. This analysis is crucial for understanding the robustness of the MSE method across different conditions. To achieve this, we first synthesize an age sequence comprising 50 calendar ages, drawn randomly from a temporal interval over 3000–5000 years BP along with associated errors sampled from a gamma distribution. The gamma distribution was chosen because dating uncertainties are inherently positive and typically right-skewed, with most values being small but occasionally much larger; the gamma distribution reproduces this pattern while avoiding nonsensical negative values. From this population, we randomly select 10 ages and convert them to radiocarbon ages using the IntCal20 calibration curve (Reimer et al., 2020). We then apply the MSE method to estimate the endpoints

Table 1
Estimated endpoints of a synthesized age sequence spanning 3000−5000 years BP under different probability distributions.

Probability distribution	Lower bound (years BP)			Upper bound (years BP)		
	Mean	Mode	2σ credible interval	Mean	Mode	2σ credible interval
Uniform	3026	3075	[2845, 3180]	4961	4945	[4815, 5115]
Exponential	3025	3040	[2860, 3170]	4959	4940	[4815, 5115]
Normal	3015	3075	[2815, 3180]	4974	4970	[4805, 5155]
Log-normal	3018	3030	[2825, 3185]	4964	4920	[4810, 5145]

of the synthesized age sequence based on the above parametric distributions. Finally, we assess the accuracy and reliability of these estimates, exploring how variations in distributional choice and sample size influence the robustness of the MSE method.

The latest possible ages in the chronological sequence, represented by the lower bounds on the BP scale, range from about 2600 to 3200 years BP (Fig. 1A; Table 1). The different distributions offer varying interpretations of the likely lower bound. The uniform distribution shows a relatively flat, equal-probability interval with slightly higher probabilities concentrated around the true age (i.e., 3000 years BP). The exponential distribution introduces a stronger peak in the probability density around the same time, suggesting a more pronounced likelihood for the lower bound to be closer to the true age but still with some uncertainty extending toward earlier dates. The normal and log-normal distributions exhibit wider credible intervals, indicating a greater degree of uncertainty in the lower bound estimate. These models suggest that, while the lower bound is most likely to fall around 3000 years BP, there is a broader spread of potential estimates, extending to both earlier and later dates, reflecting the higher flexibility of these models in estimating the lower bound.

The earliest possible ages in the chronological sequence, represented by the upper bounds on the BP scale, span approximately 4700–5300 years BP (Fig. 1B). As with the lower bound estimates, the different statistical models yield distinct patterns of probability. The uniform
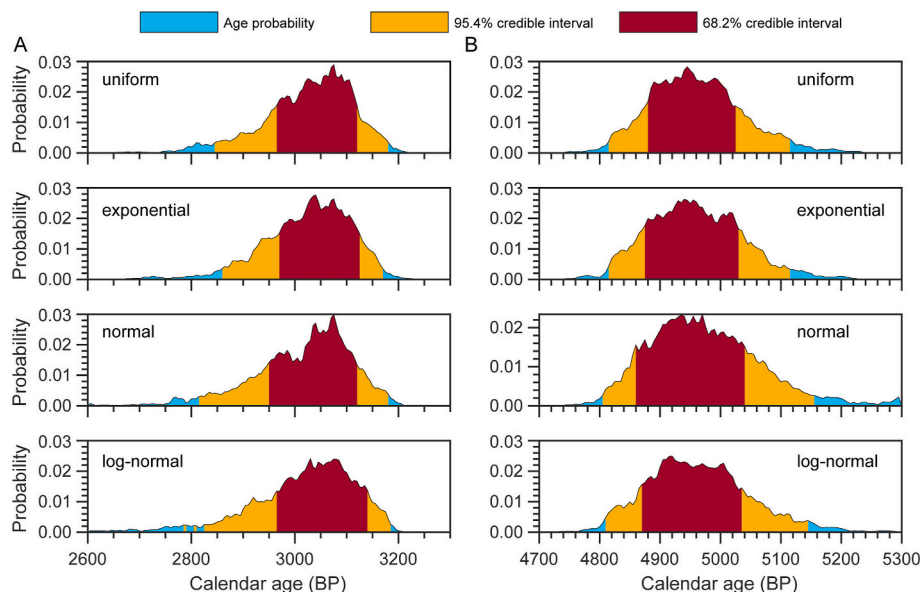
Fig. 1. Probability density distributions for estimated chronological endpoints under four statistical models based on synthesized chronological data spanning 3000–5000 years before present. Panel A presents results for the upper bound (latest possible age), while Panel B shows results for the lower bound (earliest possible age) of the sequence. Uniform and exponential models produce more constrained intervals, whereas normal and log-normal models yield broader intervals, reflecting greater uncertainty in endpoint estimation. These differences illustrate how distributional assumptions influence the inferred temporal boundaries.

distribution presents a wide, relatively flat probability distribution, though there is a noticeable peak around the true age (i.e., 5000 years BP), indicating a slightly higher likelihood for the upper bound in that range. The exponential distribution produces a sharper peak, concentrating the highest probability around 5000 years BP. The normal and log-normal distributions again show wider credible intervals for the upper bound, suggesting greater uncertainty in the age estimates. These models have a broader spread, particularly the log-normal distribution, which displays a pronounced skewness, with a longer tail extending toward later dates. This suggests that under the log-normal model, there is a higher probability that the upper bound may be later than 5000 years BP, whereas the normal distribution provides a more symmetric probability spread centered around the same time.

Our results highlight key differences in how the statistical distributions influence the estimation of the lower and upper bounds of the age sequence. The normal and log-normal models, which assume symmetric and skewed probability distributions, respectively, consistently yield wider credible intervals for both the lower and upper bounds, capturing greater uncertainty in these estimates. This broader range reflects the flexibility of these models in accommodating variability in the data. In contrast, the uniform and exponential models provide narrower credible intervals, suggesting more constrained age estimates. The exponential distribution, in particular, exhibits sharper peaks in both panels, indicating more concentrated likelihoods for both the estimated lower and upper bounds to fall within known age ranges (around 3000 years BP for the lower bound and 5000 years BP for the upper bound).

5.1.2. Sensitivity to sample size

We illustrate the impact of sample size on estimating the lower and upper bounds of the same synthesized age sequence under four different probability distributions (Fig. 2). The sensitivity is assessed by calculating relative error in percentage. For both estimates, the four parametric distributions display similar trends, though there are notable differences in how they behave for small sample sizes. For the lower bound estimates (Fig. 2A), all models exhibit high relative error when

the sample size is small (below 50), with the error peaking around 6 % for the uniform distribution and slightly lower for the other models. As sample size increases, the relative error rapidly declines, stabilizing at below 1 % for all distributions once the sample size exceeds 100. This indicates that smaller sample sizes lead to higher uncertainty in the lower bound estimates, while larger sample sizes reduce error across all distribution models. Similarly, for the upper bound estimates (Fig. 2B), the relative error starts high for small sample sizes (below 50), but the magnitude of the relative error is lower than that observed for the lower bound, peaking at around 2 % for the uniform and exponential models and slightly less for the normal and log-normal distributions. The error decreases as sample size increases, leveling off at values close to 0 once the sample size exceeds 100.

Our result highlights the sensitivity of lower and upper bound estimates to sample size, demonstrating that smaller sample sizes introduce greater relative error. The results suggest that the normal and log-normal models tend to perform slightly better in terms of minimizing relative error compared to the uniform and exponential models. However, as the sample size grows, all distributions converge to similar, minimal error levels, suggesting that sample size is a critical factor in reducing estimation uncertainty, regardless of the chosen distribution. This emphasizes the importance of using sufficiently large sample sizes to achieve reliable and accurate chronological estimates in archaeological and geological studies.

5.2. Real-world data application

5.2.1. Arrival time of anatomically modern humans in southern China

The dispersal of modern humans into China represents one of the key migrations of *Homo sapiens* out of Africa. Fossil and archaeological evidence suggest that modern humans reached China as early as between 80,000 and 120,000 years ago. Key sites such as those in Zhoukoudian and Tianyuan Cave, have yielded fossils and artifacts indicating a relatively early presence of modern humans in the region. Genetic studies support this timeline, revealing that modern Chinese populations share
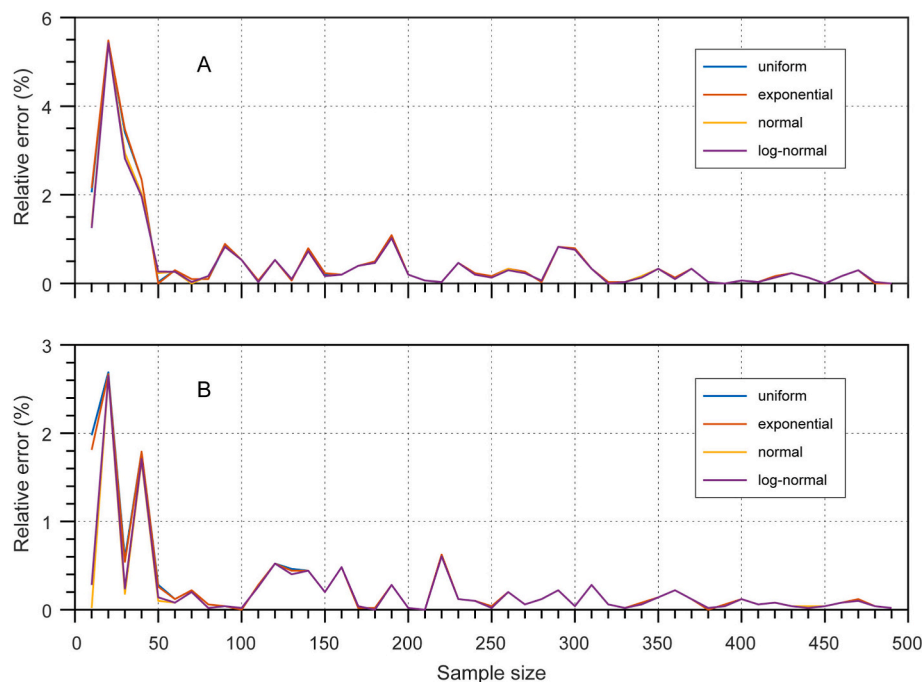
Fig. 2. Sensitivity of estimated chronological endpoints to sample size for a synthesized age sequence spanning 3000–5000 years before present, evaluated under four statistical models. Relative error (%) is plotted against sample size for the lower bound (Panel A) and the upper bound (Panel B) of the sequence, respectively. Across all models, relative errors are highest at small sample sizes (<50) and decline rapidly with increasing sample size, stabilizing at values generally below 1 % for samples >100. The similarity of trends among models indicates that sample size exerts a stronger influence on estimation accuracy than the choice of statistical distribution, though minor differences are apparent at very small sample sizes.

ancestry with ancient human groups that arrived during this period. The arrival of *Homo sapiens* in China marks an important phase of interaction with local archaic human populations, such as Denisovans and possibly late Neanderthals, contributing to the genetic diversity observed in contemporary East Asian populations. This migration not only shaped the demographic and genetic landscape of East Asia, but also facilitated the spread of new technologies, cultural practices, and adaptations to different environmental challenges. Here, we aim to refine the estimated arrival time of anatomically modern humans in southern China by integrating radiocarbon (^{14}C), optically stimulated luminescence (OSL), and uranium-series (U-series) dating results within a probabilistic modeling framework.

The Liujiang hominin, found in the Tongtianyan cave (24°10′59″N, 109°25′56″E), Guangxi Province, provides strong evidence for a late-Pleistocene presence of modern humans in East Asia. Geological and stratigraphic observations indicate that the sedimentary sequence in the cave can be subdivided into three irregularly bedded units. Unit II is characterized by yellowish-brown clay sediments, with abundant mammal remains and a nearly complete human cranium. We estimate the arrival of modern humans in South China from a collection of 23 ages obtained through indirect dating of fossil-bearing sediments of Unit II using the ^{14}C and OSL method as well as direct dating of human tooth remains using the U-series method (Ge et al., 2024). Fig. 3 depicts the estimated arrival time of modern humans to South China, with two different probability models used to constrain the age distribution. Both distributions demonstrate a peak age probability centered around 30,500 years BP, with slightly more spread in the exponential model than the uniform model. The narrower credible intervals in the uniform model suggest a more concentrated age estimate, while the exponential model introduces a slight right skew, indicating a greater probability of later dates.

Our comparative study reveals that, while both models suggest a similar timeframe for the arrival of modern humans to South China, the choice of the distribution more or less influences the uncertainty around the estimate. The exponential model allows for a broader range of possible ages, particularly beyond 31,000 years BP, whereas the uniform model presents a more confined age range. This suggests that, depending on the underlying assumptions about the nature of the event (e.g., gradual versus instantaneous arrival), different models can yield slightly different interpretations of the timing of modern human dispersal into South China. The overall convergence of the estimates around 30,500 years BP, however, supports the idea of a relatively stable and reliable period for the arrival of modern humans in this region. Our age estimates for the Liujiang skeletal remains are consistent with those at the Tianyuan (Shang, 2010), Bailian (Zhou et al., 2019), and Laoya caves (Xing et al., 2017), implying that these early populations likely migrated

along coast and inland waterways, moving through South Asia and Southeast Asia and reaching East Asia synchronously. Note that these conclusions are constrained by the limited number of directly dated human remains in southern China, and by the assumption that the selected probability models adequately reflect the true distribution of arrival times. Additional well-dated sites are needed to further refine the chronology of modern human dispersal into this region.

5.2.2. Duration of human occupation at a mound site in the lower Yellow River area

The Yellow River floodplain offers a unique opportunity to examine the interplay between human activity and fluvial system during the Holocene (Yu et al., 2023), being recognized as the cradle of Chinese
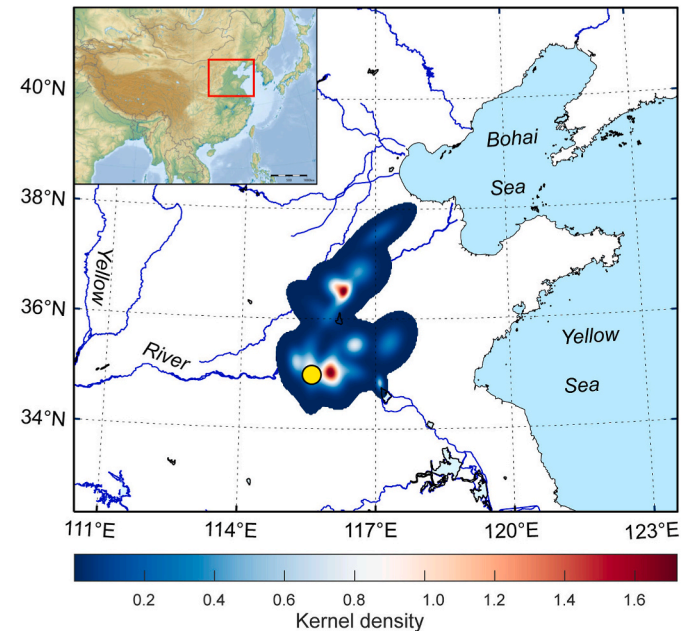

Fig. 4. Kernel density map of mound sites in the lower Yellow River region, illustrating spatial concentrations of archaeological mounds. Warmer colors indicate higher site densities, with clustering patterns reflecting areas of intensive past human activity and settlement. Red box outlines the study area, while filled yellow circle marks the location of the Shilipu archaeological site, a key reference point for understanding regional occupation patterns. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)
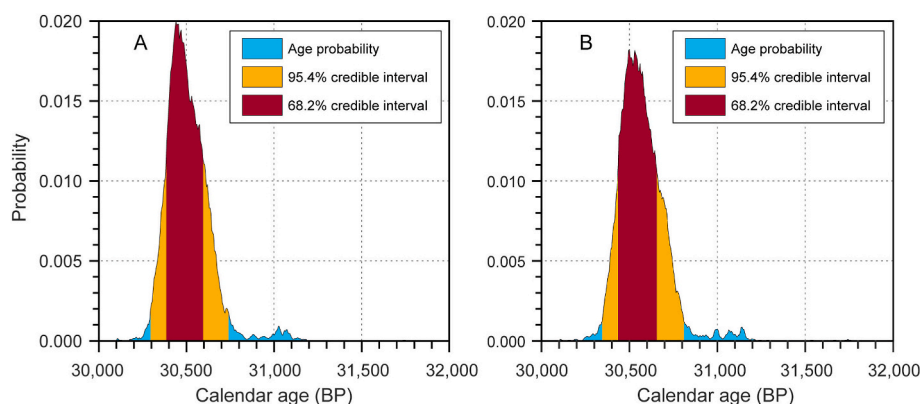

Fig. 3. Estimated arrival time of modern humans in southern China based on two alternative statistical models for the distribution of calendar ages. (A) uniform, and (B) exponential. Both models produce closely aligned central estimates (~30,500 years before present), but the exponential model yields slightly narrower credible intervals, indicating higher confidence in the estimated arrival date. These results highlight the influence of distributional assumptions on temporal inferences regarding early human dispersal.

civilization due to its fertile soil and water access. Archaeological excavations have revealed a continuous history of human habitation marked by increasing social complexity, evident in settlement hierarchies and urban development. Excavations in the lower Yellow River region have uncovered mound sites that date back to the Neolithic (Jing et al., 1997). These mounds, typically oblong or oval and several meters high, are composed of yellowish coarse sands and small pebbles, reflecting the hilly landscape formed by an incising river system in the late Pleistocene and early Holocene. Kernel density estimate of Neolithic and Bronze Age mound sites in the Lower Yellow River region reveals significant clusters in the Heze and Liaocheng areas of Shandong Province (Fig. 4). The spatial concentration of these sites, with the highest densities found in the central zones of these two regions, likely corresponding to key settlement or cultural hubs during these periods. The proximity of the mound sites to the Yellow River highlights the central role of the river in shaping early agricultural societies by providing essential resources for farming, transportation, and trade, which facilitated the growth and sustainability of these ancient communities. Situated in a floodplain, the mounds also provided ideal habitats and refuges during floods (Jing et al., 1995), and as sediment accumulation increased (Shi et al., 2010), inhabitants elevated the mounds artificially to adapt to rising water levels during the middle and late Holocene.

This case study aims to reconstruct the duration of human occupation at a representative mound site in the lower Yellow River floodplain, in order to better understand how early societies balanced the agricultural benefits of a fertile but flood-prone environment with the risks posed by hydrological instability. We estimate the settlement time of this site using a small dataset of radiocarbon ages from the Shilipu mound (35°06′54″N, 115°31′52″ E), which may deepen our understanding about how early societies adapted to the challenges and balanced its agricultural benefits with the risks of flooding of the Yellow River (Yu et al., 2020). The age distribution reveals a distinct period of human activity or mound aggradation (Fig. 5; Table 2). The lower bound of the age sequence shows a calendar age cluster primarily around 2500 BC, indicating the commencement of human occupation at this site closely related to the rise and expansion of the Longshan culture into this area (Sun, 2013). The unpredictable and frequent flooding of the Yangtze River during the late Holocene may have significantly impacted settlement patterns. Communities tend to settle in elevated areas and

constructed mound sites as protective measures against flooding, as well as for ceremonial or defensive purposes. The upper bound is shown by a prominent peak around 200 BC, suggesting a more recent abandonment of this site. As shown in Fig. 6, the most likely duration of settlement at this site is around 2190−2480 years (68.2 % credible interval) or 1960−2570 years (95.4 % credible interval).

The termination of human occupation at this site may signify an important shift in environmental dynamics of the middle Yellow River. The introduction of iron tools and the resulting population expansion during the Warring States period (475–221 BC) led to profound alterations in the landscape, particularly on the Chinese Loess Plateau. Iron tools made large-scale agriculture and deforestation easier, resulting in widespread soil erosion across the plateau. The loose, fine-grained loess soil, when disturbed, became highly vulnerable to erosion by water. Consequently, the lower Yellow River became a sediment-laden system (Shi et al., 2010), prone to flooding and changing its course frequently since the beginning of the second millennium BC (Li et al., 2021; Du et al., 2024). This shift in environmental dynamics could have led to a rapid aggradation of the floodplain and the inundation of the site by Yellow River floods, thus prompting the end of occupation at the site. The settlement time of this site represents the resilience and ingenuity of Neolithic and Bronze Age cultures in responding to the volatile hydrological regime of the Yellow River, reflecting the critical interaction between human settlement and environmental condition in the lower Yellow River basin. Overall, the settlement history of the Shilipu mound illustrates the resilience and ingenuity of Neolithic and Bronze Age cultures in responding to the volatile hydrological regime of the Yellow River. However, the analysis is limited by the small number of radiocarbon ages available, which constrains the precision of the estimated occupation duration and may overlook short-term episodes of abandonment or reoccupation.

6. Concluding remarks

Estimating chronological boundaries from age sequences is crucial in archaeology, where dating accuracy and reliability are paramount. In a Frequentist framework, we arrive at an analytical solution to this problem, addressing notable gaps in existing methodologies. Key points in the implementation of this method include: (1) Simplifying the maximum likelihood estimation method through maximum spacing estimation. By deriving an analytical expression through maximizing the spacing in the data, we streamline the process while preserving the robustness of the endpoint estimation. This makes the method more accessible for practical use, reducing computational complexity while maintaining numerical accuracy. (2) Applying Monte Carlo simulations. The incorporation of a Monte Carlo approach acknowledges the inherent uncertainties in radiocarbon or other laboratory-derived numerical ages, which are rarely treated with such rigor. This step enhances the precision of the model by simulating a range of potential outcomes, adding depth to the estimates. (3) Accounting for multiple probability distributions. By comparing four parametric models, we acknowledge the complex nature of chronological datasets. A comparative study using synthesized data shows that different statistical models offer different levels of confidence, with more restrictive models (uniform and exponential) yielding tighter estimates, while normal and log-normal distributions provide a broader range. This dual approach allows for flexibility and highlights the need to carefully select the appropriate distribution based on the characteristics of the chronological dataset.

The applicability of this method to both small and large datasets broadens its use, whether estimating the settlement duration of a single site or the broader rise and decline of a regional culture. This flexibility is valuable particularly in the archaeological field, where data availability and quality can vary greatly across sites and cultural periods. It is noteworthy that the broader credible intervals in the normal and log-normal models may indeed reflect real-world uncertainties, such as stratigraphic mixing or dating inconsistencies. By acknowledging this
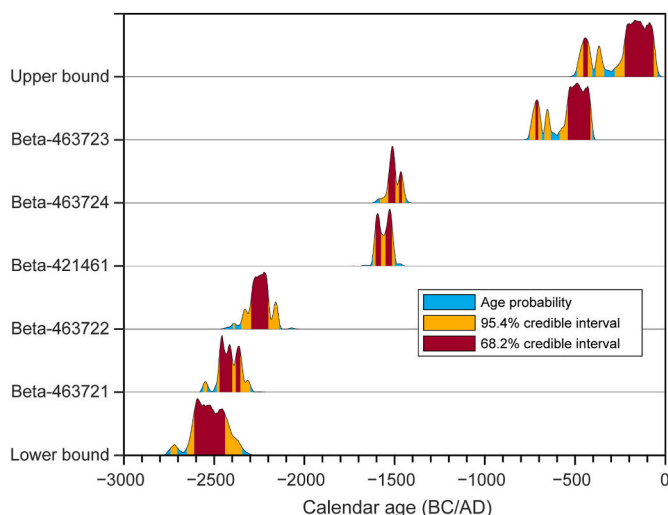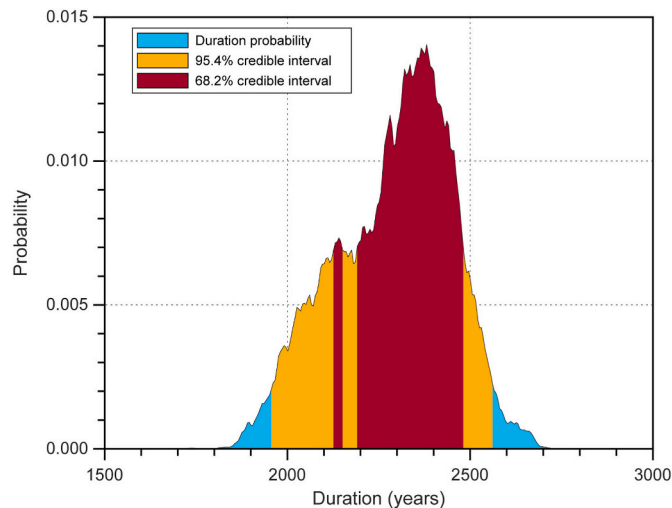

Fig. 5. Posterior probability distributions for the calibrated radiocarbon dates from the Shilipu mound in the lower Yellow River area, modeled under the assumption of a uniform distribution of calendar ages. Each horizontal row represents an individual ^{14}C sample, with the bottom and top rows showing the inferred lower and upper bounds of human occupation, respectively. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 2
Estimated human settlement time on the Shilipu mound in the lower Yellow River.

Sample number	Mean (BC/AD)	Median (BC/AD)	Mode (BC/AD)	1σ credible interval (BC/AD)	2σ credible interval (BC/AD)
Lower bound	−2523	−2525	−2595	[−2440, −2610] (1.0000)	[−2340, −2650] (0.9671) [−2705, −2740] (0.0329)
Beta-463721	−2415	−2420	−2455	[−2400, −2470] (0.7301) [−2355, −2380] (0.2699)	[−2300, −2490] (0.9691) [−2540, −2555] (0.0309)
Beta-463722	−2249	−2250	−2220	[−2200, −2295] (0.9712) [−2160, −2160] (0.0288)	[−2140, −2350] (0.9870) [−2385, −2395] (0.0130)
Beta-421461	−1559	−1560	−1530	[−1515, −1550] (0.5296) [−1575, −1605] (0.4704)	[−1500, −1620] (1.0000)
Beta-463724	−1503	−1510	−1510	[−1495, −1535] (0.7682) [−1460, −1475] (0.2318)	[−1440, −1570] (0.9938) [−1585, −1585] (0.0062)
Beta-463723	−546	−515	−500	[−415, −540] (0.8681) [−705, −725] (0.1319)	[−400, −580] (0.7220) [−635, −670] (0.0872) [−685, −750] (0.1908)
Upper bound	−229	−190	−160	[−60, −225] (0.8939) [−430, −450] (0.1061)	[−45, −275] (0.7236) [−345, −390] (0.0864) [−405, −490] (0.1900)

Note: Numbers in parentheses represent the fraction of area under probability distribution.

Fig. 6. Posterior probability distribution of the estimated duration of human settlement on the Shilipu mound in the lower Yellow River area, modeled under the assumption of a uniform distribution of calendar ages between occupation onset and abandonment. These results indicate that the most probable settlement duration falls within 2210–2495 years (the 68.2 % highest posterior density credible interval), while durations outside 1955−2560 years (the 95.4 % highest posterior density credible interval) are less strongly supported by the model. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

variability, the presented method could offer more realistic estimates when data integrity is in question.

Illustrative studies using real-world data suggest that this method could clearly improve archaeological interpretations of human dispersal and settlement timing, as well as the confidence with which these conclusions are drawn. The ability to estimate endpoints more reliably, while also considering the effect of different statistical models, makes this approach versatile for a wide range of applications, from site-specific studies to regional or cultural analyses. This method could also be applied in geological and ecological contexts, offering a broader utility beyond archaeology by estimating the timing of environmental or paleontological events from a limited number of chronological data.

Declaration of competing interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Shiyong Yu reports financial support was provided by the National Natural Science Foundation of China. If there are other authors, they declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi. org/10.1016/j.quageo.2025.101697.

Data availability

No data was used for the research described in the article.

References

Anatolyev, S., Kosenok, G., 2005. An alternative to maximum likelihood based on spacings. Econom. Theory 21, 472–476, 410.1017/S0266466605050255.

Andrés Christen, J., Pérez E, S., 2009. A new robust statistical model for radiocarbon data. Radiocarbon 51, 1047–1059, 1010.1017/S003382220003410X.

Becerra-Valdivia, L., Higham, T., 2020. The timing and effect of the earliest human arrivals in North America. Nature 584, 93–97. https://doi.org/10.1038/s41586-41020-42491-41586.

Bronk Ramsey, C., 1995. Radiocarbon calibration and analysis of stratigraphy: the OxCal program. Radiocarbon 37, 425–430. https://doi.org/10.1017/S0033822200030903.

Bronk Ramsey, C., 2000. Comment on 'The use of Bayesian statistics for ¹⁴C dates of chronologically ordered samples: a critical analysis'. Radiocarbon 42, 199–202, 110.1017/S0033822200059002.

Buck, C.E., Christen, J.A., James, G.N., 1999. BCal: an on-line Bayesian radiocarbon calibration tool. Internet Archaeol. 7, 13635387, 13635310.13611141/ia.13635387.13635381.

Buck, C.E., Litton, C.D., Smith, A.F., 1992. Calibration of radiocarbon results pertaining to related archaeological events. J. Archaeol. Sci. 19, 497–512, 410.1016/0305-4403(1092)90025-X.

Cheng, R., Amin, N., 1983. Estimating parameters in continuous univariate distributions with a shifted origin. J. R. Stat. Soc. Ser. B 45, 394–403, 310.1111/j.2517-6161.1983.tb01268.x.

Du, T., Zhang, W., Li, B., Liu, L., Li, Y., Ge, Y., Yu, S.-Y., 2024. Sedimentary evidence for the diversion of the yellow river onto the North China Plain 3000–2600 years ago. Palaeogeogr. Palaeoclimatol. Palaeoecol. 634, 111909, 111910.111016/j.palaeo.112023.111909.

Ekström, M., Inference, 2008. Alternatives to maximum likelihood estimation based on spacings and the Kullback–Leibler divergence. J. Stat. Plann. Inference 138, 1778–1791, 1710.1016/j.jspi.2007.1706.1031.

Ge, J., Xing, S., Grün, R., Deng, C., Jiang, Y., Jiang, T., Yang, S., Zhao, K., Gao, X., Yang, H., Guo, Z., Petraglia, M.D., Shao, Q., 2024. New Late Pleistocene age for the Homo sapiens skeleton from Liujiang southern China. Nat. Commun. 15, 3611, 3610.1038/s41467-41024-47787-41463.

Guilderson, T.P., Reimer, P.J., Brown, T.A., 2005. The boon and bane of radiocarbon dating. Science 307, 362–364, 310.1126/science.11041.

Herrando-Pérez, S., Saltré, F., 2024. Estimating extinction time using radiocarbon dates. Quat. Geochronol. 79, 101489, 101410.101016/j.quageo.102023.101489.

Holland-Lulewicz, J., Ritchison, B., 2021. How many dates do i need?: using simulations to determine robust age estimations of archaeological contexts. Adv. Archaeol. Pract. 9, 272–287, 210.1017/aap.2021.1010.

Hyndman, R.J., 1996. Computing and graphing highest density regions. Am. Stat. 50, 120–126, 110.1080/00031305.00031996.10474359.

Jing, Z., Rapp, G., Gao, T., 1995. Holocene landscape evolution and its impact on the Neolithic and Bronze Age sites in the Shangqiu area, northern China. Geoarchaeology 10, 481–513, 410.1002/gea.3340100605.

Jing, Z., Rapp, G., Gao, T., 1997. Geoarchaeological aids in the investigation of early Shang civilization on the floodplain of the lower Yellow River, China. World Archaeol. 29, 36–50, 10.1080/00438243.00431997.09980362.

Li, W.-J., Yu, S.-Y., Pan, J., Cao, X., Chen, Y., Wang, Y., 2021. A 2000-year documentary record of levee breaches on the lower Yellow River and their relationship with climate changes and human activities. Holocene 31, 333–345, 310.1177/0959683620972764.

Ranneby, B., 1984. The maximum spacing method. An estimation method related to the maximum likelihood method. Scand. J. Stat. 11, 93–112. https://www.jstor.org/stable/4615946.

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., Palmer, J.G., Pearson, C., van der Plicht, J., Reimer, R.W., Richards, D.A., Scott, E.M., Southon, J.R., Turney, C.S.M., Wacker, L., Adolphi, F., Büntgen, U., Capano, M., Fahrni, S.M., Fogtmann-Schulz, A., Friedrich, R., Köhler, P., Kudsk, S., Miyake, F., Olsen, J., Reinig, F., Sakamoto, M., Sookdeo, A., Talamo, S., 2020. The IntCal20 Northern Hemisphere radiocarbon age calibration curve (0–55 cal kBP). Radiocarbon 62, 725–757, 710.1017/RDC.2020.1041.

Reimer, P.J., Brown, T.A., Reimer, R.W., 2004. Discussion: reporting and calibration of post-bomb 14C data. Radiocarbon 46, 1299–1304, 1210.1017/S0033822200033154.

Saltré, F., Brook, B.W., Rodríguez-Rey, M., Cooper, A., Johnson, C.N., Turney, C.S., Bradshaw, C.J., 2015. Uncertainties in dating constrain model choice for inferring extinction time from fossil records. Quat. Sci. Rev. 112, 128–137, 110.1016/j.quascirev.2015.1001.1022.

Saltré, F., Chadœuf, J., Higham, T., Ochocki, M., Block, S., Bunney, E., Llamas, B., Bradshaw, C.J., 2024. Environmental conditions associated with initial northern expansion of anatomically modern humans. Nat. Commun. 15, 4364, 4310.1038/s41467-41024-48762-41468.

Shang, H., 2010. The Early Modern Human from Tianyuan Cave, China. Texas A&M University Press, College Station.

Shi, C., Zhang, L., Xu, J., Guo, L., 2010. Sediment load and storage in the lower Yellow River during the late Holocene. Geogr. Ann. Ser. A 92, 297–309, 210.1111/j.1468-0459.2010.00396.x.

Signor, P.W., Lipps, J.H., 1982. Sampling bias, gradual extinction patterns and catastrophes in the fossil record. In: Silver, L.T., Schultz, P.H. (Eds.), Geological Implications of Impacts of Large Asteroids and Comets on the Earth. Geological Society of America, pp. 292–296, 210.1130/SPE1190.

Solow, A.R., 1993. Inferring extinction from sighting data. Ecology 74, 962–964, 910.2307/1940821.

Solow, A.R., 1997. Estimating settlement time. Radiocarbon 39, 351–354, 310.1017/S0033822200053327.

Steier, P., Rom, W., 2000. The use of Bayesian statistics for 14C dates of chronologically ordered samples: A criticalaanalysis. Radiocarbon 42, 183–198, 110.1017/S0033822200058999.

Sun, B., 2013. The Longshan culture of Shandong. In: Underhill, A.P. (Ed.), A Companion to Chinese Archaeology. Blackwell Publishing Ltd, pp. 435–458, 410.1002/9781118325698.ch9781118325621.

Telford, R.J., Heegaard, E., Birks, H.J., 2004. The intercept is a poor estimate of a calibrated radiocarbon age. Holocene 14, 296–298, 210.1191/0959683604hl0959683707fa.

Xing, S., Guan, Y., O'Hara, M., Cai, H., Wang, X., Gao, X., 2017. Late Pleistocene hominin teeth from Laoya cave, southern China. Anthropol. Sci. 125, 129–140, 110.1537/ase.170802.

Yu, S.-Y., 2022. MatCalib: a Matlab software package for Bayesian modeling of radiocarbon ages subject to temporal order constraints. AIMS Geosci 8, 16–32. https://doi.org/10.3934/geosci.2022002.

Yu, S.-Y., 2024. Integrating age modeling into a hierarchical Bayesian framework for inferring the pattern and rate of past sea-level changes from uncertainty-prone proxy data. Quat. Geochronol. 85, 101617, 101610.101016/j.quageo.102024.101617.

Yu, S.-Y., Chen, X., Fang, Z., Liu, X., Li, M., Guo, J., 2021. Towards a precise timing of groundwater use in the lower Yellow River area during the late Bronze age: bayesian inference from the radiocarbon ages of ancient water wells at the Liang'ercun site, north China. Quat. Geochronol. 66, 101214, 101210.101016/j.quageo.102021.101214.

Yu, S.-Y., Hou, Z., Chen, X., Wang, Y., Song, Y., Gao, M., Pan, J., Sun, M., Fang, H., Han, J., Kidder, T.R., Chen, F.-H., 2020. Extreme flooding of the lower Yellow River near the Northgrippian-Meghalayan boundary: evidence from the Shilipu archaeological site in southwestern Shandong Province, China. Geomorphology 350, 106878, 106810.101016/j.geomorph.102019.106878.

Yu, S.-Y., Li, W.-J., Zhou, L., Yu, X., Zhang, Q., Shen, Z., 2023. Human disturbances dominated the unprecedentedly high frequency of Yellow River flood over the last millennium. Sci. Adv. 9, eadf8576, 8510.1126/sciadv.adf8576.

Zeeden, C., Dietze, M., Kreutzer, S., 2018. Discriminating luminescence age uncertainty composition for a robust Bayesian modelling. Quat. Geochronol. 43, 30–39, 0.1016/j.quageo.2017.1010.1001.

Zheng, Y., Yu, S.-Y., Fan, T., Oppenheimer, C., Yu, X., Liu, Z., Xian, F., Liu, Z., Li, J., Li, J., 2021. Prolonged cooling interrupted the Bronze Age cultures in northeastern China 3500 years ago. Palaeogeogr. Palaeoclimatol. Palaeoecol. 574, 110461, 110410.111016/j.palaeo.112021.110461.

Zhou, Y., Jiang, Y., Liang, G., Li, Y., Forestier, H., Li, H., Chen, P., Wang, L., Liang, T., He, C., 2019. A technological perspective on the lithic industry of the Bailiandong Cave (36–7 ka) in Guangxi: an effort to redefine the cobble-tool industry in South China. C. R. Palevol 18, 1095–1121, 010.1016/j.crpv.2019.1009.1001.