Radiocarbon (2026), **68**, pp. 1155–1168
doi:10.1017/RDC.2026.10258

CAMBRIDGE UNIVERSITY PRESS

RESEARCH ARTICLE

Identifying activity pulses in radiocarbon sequences: A Monte Carlo kernel–minima approach and simulation assessment

Christian Mesía-Montenegro

Dirección de Investigación e Innovación, Universidad Privada del Norte, Peru
Email: christian.mesia@upn.edu.pe

Received: 08 December 2025 UTC; **Revised:** 03 April 2026 UTC; **Accepted:** 31 May 2026 UTC

Keywords: changepoint models; kernel density estimation; prehistoric demography; radiocarbon dating; summed probability distributions.

Abstract

Radiocarbon date frequencies are widely used to reconstruct temporal variation in human activity, with peaks and gaps in kernel density estimates (KDEs) often read as discrete episodes. Such readings usually lack explicit segmentation criteria and are sensitive to sampling variability, calibration-curve structure and smoothing choices. This paper proposes and evaluates a Monte Carlo KDE–minima procedure for descriptive segmentation of dated-event intensity. Calibrated posterior distributions for individual dates are sampled to generate an ensemble of KDEs on a fixed calendar grid; local minima are retained as pulse boundaries only when they occur in both the median KDE and its 95% envelope, making the rule explicitly dependent on bandwidth and smoothing scale.

Performance is assessed with two simulation designs. Level A operates in calendar time with three pulses plus background, varying sample size, pulse separation and bandwidth. Pulse width is fixed at 400 yr and separation Δ is defined as the centre-to-centre distance between adjacent pulses. Level B incorporates full radiocarbon calibration under three setups (IntCal20, SHCal20 and Mixed) and compares the envelope rule with a naive single-KDE minima rule and a Poisson changepoint model for binned counts. The bandwidth range $h = 80$–120 yr is treated as a decision-relevant multi-century smoothing regime, not as an optimal selector. High probabilities of recovering the three-pulse structure arise only when episodes are widely separated ($\sim$ 800 yr) and supported by at least $\sim$ 100–200 dates. Calibration further degrades performance. Differences among curves are small relative to sample size, bandwidth and segmentation rule, indicating that curve choice is second-order at this scale. Conditional on correct pulse-count recovery, naive KDE typically places boundaries more tightly, but at the cost of substantially more over-segmentation. The envelope rule is therefore better understood as a low-risk device for suppressing unstable minima than as a precision estimator. Under the calibrated design examined here, the Poisson changepoint model systematically over-segments. The study sets quantitative limits on what KDE-based pulse detection can recover and provides a conservative rule for summarising broad episodes of dated activity rather than estimating population history directly.

1 Introduction

Frequencies of radiocarbon determinations are now a standard tool for describing the tempo of past human and environmental activity. Summed calibrated probability distributions (SPDs) and kernel density estimates (KDEs) are routinely used to visualise peaks and troughs in dated events and to infer temporal phases of occupation, construction or ritual activity (e.g., Timpson et al. 2014; Bevan and Crema 2021; Mendez-Quiros et al. 2023). Yet these inferences typically rely on informal visual inspection, with peaks and gaps identified by eye. Such approaches are vulnerable to sampling variability, taphonomic loss and structure in the calibration curve, and they are further affected by

© The Author(s), 2026. Published by Cambridge University Press on behalf of University of Arizona. This is an Open Access article, distributed under the terms of the Creative Commons Attribution licence (https://creativecommons.org/licenses/by/4.0/), which permits unrestricted re-use, distribution and reproduction, provided the original article is properly cited.

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

choices of temporal aggregation, smoothing and inclusion criteria (Contreras and Meadows 2014; Surovell et al. 2009; Crema 2022).

Recent reviews emphasise that radiocarbon-frequency methods fall into three broad families: (i) reconstructive approaches, which estimate a smoothed intensity function over time (e.g., SPDs, KDEs, mixture models); (ii) hypothesis-testing frameworks, which use Monte Carlo or permutation procedures to test specific models against null expectations; and (iii) explicit model-fitting strategies, including Poisson and Cox-process models in likelihood or Bayesian form (Crema and Bevan 2021; Crema 2022; Heaton et al. 2025; de Navascués et al. 2025). In this view, SPDs and KDEs are reconstructions of an underlying event intensity, not direct measurements of population size (Bronk Ramsey 2017). They do not, by themselves, provide probabilistic statements about the number, timing or magnitude of discrete episodes.

A central practical problem is therefore segmentation: given a smoothed time series of dated activity, under what conditions is it defensible to treat peaks and troughs as distinct episodes? Archaeologists already use an implicit segmentation procedure when they mark off "phases" on an SPD or KDE plot (e.g., Bevan and Crema 2021; Timpson et al. 2014; Mendez-Quiros et al. 2023). However, the criteria behind such decisions are rarely stated, and their performance has not been systematically evaluated with controlled simulations (Contreras and Meadows 2014; Crema 2022).

This paper proposes a simple and deliberately conservative segmentation rule: a Monte Carlo KDE–minima procedure that identifies *activity pulses* as contiguous calendar intervals between persistent minima of an ensemble of KDEs. The method is designed to sit within the reconstructive family of approaches. It formalises a familiar visual practice—reading episodes off a smoothed curve—without introducing the strong parametric assumptions of fully specified changepoint models, while remaining compatible with recent developments in radiocarbon event-count modelling (Crema and Bevan 2021; Heaton et al. 2025; de Navascués et al. 2025).

Two elements are key. First, radiocarbon dates are treated as calibrated posterior distributions rather than point estimates. A Monte Carlo ensemble of KDEs is generated by repeatedly sampling calendar ages from those posteriors (cf. Efron and Tibshirani 1994; Bronk Ramsey 2017). Second, minima are accepted as pulse boundaries only when they appear in both the median KDE and the lower bound of a Monte Carlo envelope, so that shallow troughs likely to reflect sampling noise or calibration-induced shoulders are systematically excluded. The bandwidth and grid resolution explicitly control the scale at which pulses are defined (Silverman 2018; Scott 2015; Sheather 2004). Figure 1 provides simple visual examples of the resulting segmentation for favourable and unfavourable cases.

Substantively, the method is intended for descriptive segmentation of dated-event intensity, not for direct population estimation. It is most relevant when archaeologists are already reading broad phases visually from SPD or KDE curves and need a low-risk formalisation of that exploratory practice. When the research question instead requires posterior estimates of episode number, timing or intensity, explicit model-fitting approaches such as recent Poisson-process methods are preferable (Heaton et al. 2025; de Navascués et al. 2025).

The aim is therefore not to promote a universal demographic tool, but to map the conditions under which KDE–minima segmentation offers a defensible summary of a simple underlying process and the conditions under which it fails. Level A isolates smoothing and sampling in stylised calendar-time simulations. Level B adds laboratory error, calibration and method comparison under IntCal20, SHCal20 and a simple 50/50 mixed assignment. The mixed setup is used as a heterogeneity stress test rather than as a literal demographic scenario. A brief empirical illustration applies the procedure to a published radiocarbon corpus for the Kotosh Religious Tradition (KRT) in the Central Andes, where a Bayesian phase model already exists; this empirical application is discussed in detail in Section 3.3.

The results show that KDE–minima segmentation performs adequately only under relatively favourable conditions—well-separated pulses supported by large samples—and that calibration further reduces performance. The ensemble-based rule nonetheless offers a low-risk tactic for summarising broad episodes, with clearly defined limitations. The focus of the paper is thus on placing quantitative bounds on what KDE-based pulse detection can safely recover, rather than on promoting a new universal tool for demographic reconstruction (cf. Crema 2022).

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

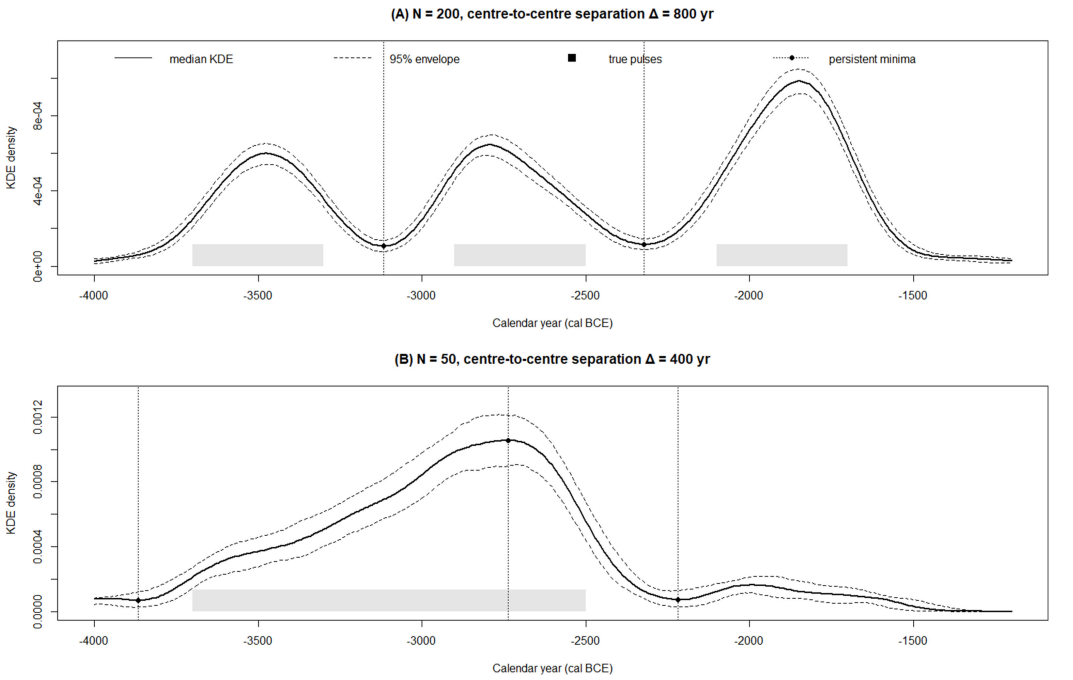


Figure 1. *Illustrative behaviour of the Monte Carlo KDE–minima algorithm under contrasting Level A conditions. (A) High-information case ($N = 200$, centre-to-centre separation $\Delta = 800$ yr) showing recovery of the three-pulse structure. (B) Lower-information case ($N = 50$, $\Delta = 400$ yr), where adjacent 400-year pulses abut and the inferred structure becomes less distinct. Solid line = median KDE; dashed lines = 95% Monte Carlo envelope; grey bars = true pulse intervals; dotted vertical lines and points = persistent minima retained as pulse boundaries. Together, the panels contrast a favourable case with wide separation and a more difficult case with fewer dates and limited separation.*

2 Methods

2.1 Activity pulses and Monte Carlo KDE–minima algorithm

Let $\{p_i(t)\}_{i=1}^{n}$ denote calibrated posterior distributions for n radiocarbon determinations, where t is calendar age. The goal is to partition the time axis into contiguous intervals within which the underlying event intensity can reasonably be treated as approximately homogeneous, and between which there is evidence for a lull in activity. These intervals are termed *activity pulses*. They are explicitly scale-dependent features of a smoothed intensity estimate at a chosen bandwidth h and grid resolution, not direct estimates of a unique demographic trajectory (Silverman 2018; Scott 2015).

The Monte Carlo KDE–minima procedure is:

1. For each replicate $r = 1, \ldots, R$, draw a single calendar age $t_i^{(r)}$ from each posterior $p_i(t)$. Across replicates, these draws approximate integrating over the full posteriors (Efron and Tibshirani 1994; Bronk Ramsey 2017).
2. For each replicate, construct a KDE

$$\hat{f}^{(r)}(t) = \frac{1}{nh} \sum_{i=1}^{n} K\left(\frac{t - t_i^{(r)}}{h}\right)$$

on a fixed calendar grid, where K is the standard normal density and h the bandwidth. The grid spacing is fine enough that minima can be located at a resolution of a few years relative to h (Silverman 2018; Scott 2015).

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

3. At each grid point t_j, compute the pointwise median $\tilde{f}(t_j)$ and the 2.5–97.5% Monte Carlo envelope $[\tilde{f}_{0.025}(t_j), \tilde{f}_{0.975}(t_j)]$ over replicates.
4. Identify local minima of $\tilde{f}(t)$ and of the lower envelope $\tilde{f}_{0.025}(t)$ using the sign pattern of first differences along the grid.
5. Mark a grid index j^* as a *persistent minimum* if it is a local minimum of the median KDE and there exists a minimum of the lower envelope within one grid step of j^*.
6. Let accepted minima occur at calendar positions $b_1 < \ldots < b_{K-1}$. Define K activity pulses as the intervals $(-\infty, b_1], (b_1, b_2], \ldots, (b_{K-1}, +\infty)$, truncated to the calendar range spanned by the data.

The requirement that minima be present in both the median and the lower envelope is intentionally conservative. A minimum visible in $\tilde{f}(t)$ but absent from $\tilde{f}_{0.025}(t)$ is treated as a likely artefact of sampling variability or calibration-induced shoulders. The single-grid-step neighbourhood allows for minor horizontal shifts in minima between the median and envelope. In spirit, the procedure is analogous to multi-scale curve exploration methods such as SiZer, which emphasise persistent structures over those that disappear under small perturbations (Chaudhuri and Marron 1999). The examples in Figure 1 show how the algorithm behaves in an "easy" case with widely separated pulses and a "hard" case with fewer dates and substantially overlapping episodes.

In all simulations and empirical analyses, calendar grids are restricted to the window of interest, and pulses abutting the edges of this window must be interpreted cautiously because minima near boundaries are less stable. In practice, analysts should examine whether edge pulses persist when the analysis window is expanded or when h is varied within a plausible range (Silverman 2018; Scott 2015). The bandwidth range used throughout this study, $h = 80$–120 yr, should be read as a decision-relevant multi-century smoothing regime rather than as an attempt to estimate an optimal bandwidth. The substantive target is the recovery of broad, multi-century pulses, not the finest-scale local structure visible in an SPD or KDE curve. Given the 400-year pulse widths used in the simulations, values in this range are large enough to suppress small shoulders and calibration-induced irregularities but still small enough not to force adjacent 400-year pulses to merge a priori. Formal bandwidth selectors exist (Sheather 2004; Silverman 2018; Scott 2015), but for radiocarbon data with irregular calibration structure and posterior uncertainty, sensitivity over a plausible range is more informative than any single plug-in estimate. Bandwidth sensitivity for the KRT application is examined in the empirical illustration (Section 3.3).

2.2 Level A simulations: calendar-time pulses

Level A simulations operate directly in calendar time (BCE), isolating the effects of sample size, episode spacing and bandwidth without calibration. The underlying intensity $\lambda(t)$ is piecewise constant with three pulses and a uniform background. The configuration is summarised in Table 1. In the baseline design, three pulses of width 400 yr are centred at −3500, −2700 and −1900 cal BCE, that is

$$[s_1, e_1]=[-3700, -3300], \quad [s_2, e_2]=[-2900, -2500], \quad [s_3, e_3]=[-2100, -1700],$$

with intensities $\lambda_1 = 1$, $\lambda_2 = 1.5$, $\lambda_3 = 2$ and background intensity $\lambda_{\text{bg}} = 0.1$. The simulation window spans −4000 to −1200 cal BCE. Under these settings, the three pulses account for approximately 92% of expected events over the window. The low background component is included to prevent the simulations from collapsing into an unrealistically clean pulse-only system and to approximate the low-rate dated events that would often exist outside major activity clusters. This stylised, piecewise-constant configuration with equal-width pulses is not intended as a literal model of any particular archaeological sequence; it provides a controlled setting in which the joint effects of spacing, sample size and bandwidth on segmentation can be isolated.

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

Table 1. *Piecewise-constant intensity function for Level A simulations. For each pulse, the table lists the calendar-time start and end (cal BCE) and the event intensity λ_k. A low background intensity $\lambda_{bg} = 0.1$ applies across the intervals outside the pulses. Level B simulations use analogous 400-yr pulses in calendar years BP.*

Pulse ID	t_{start} (cal BCE)	t_{end} (cal BCE)	λ_k
1	−3700	−3300	1.0
2	−2900	−2500	1.5
3	−2100	−1700	2.0

Throughout Level A, pulse width is fixed at 400 yr and separation Δ is defined as the centre-to-centre distance between adjacent pulses. This distinction matters. When $\Delta = 400$ yr, adjacent pulse edges abut; when $\Delta < 400$ yr, adjacent pulses overlap in calendar time; and when $\Delta > 400$ yr, a trough or lull exists between them. Reviewer concern about the recoverability of 400-year pulses at similar separations is therefore addressed here by making the relation between width and spacing explicit in both the design and the figure captions.

Events are generated by sampling segment membership with probability proportional to intensity × duration and drawing times uniformly within segments. This yields true event times $\{t_i^{\text{true}}\}_{i=1}^{N}$. To mimic posterior uncertainty in calendar time without calibration, each event is perturbed by Gaussian error with $\sigma = 50$ yr, and the perturbed times are treated as posterior draws when constructing KDEs. Figure 1 illustrates the resulting KDEs and persistent minima for a well-separated case and a more challenging configuration.

KDEs are evaluated on a fixed grid of 400 equally spaced calendar points between −4000 and −1200 cal BCE. The main design varies the total sample size $N \in \{50, 100, 200, 500\}$, the centre-to-centre separation between pulses $\Delta \in \{200, 400, 600, 800\}$ yr, and the bandwidth $h \in \{80, 100, 120\}$ yr. For each combination of (N, Δ, h), 200 datasets are simulated and analysed with the ensemble KDE–minima rule and with a naive single-KDE minima rule that uses minima of a single KDE constructed from a single draw of calendar ages. The bandwidth values are chosen to interrogate the same multi-century regime across all simulations: they are not intended as optimal selectors, but as a plausible decision range for asking whether 400-year pulses separated by several centuries remain recoverable after smoothing. In other words, the goal is conservative segmentation at the scale of broad episodes, not local feature recovery at the finest temporal resolution. Values substantially larger than 120 yr would increasingly force adjacent episodes to merge, whereas substantially smaller values would make spurious local minima more common. The chosen range therefore brackets a practical compromise between resolution and stability while keeping the design tractable (Bronk Ramsey 2017; Crema and Bevan 2021; Silverman 2018; Scott 2015; Sheather 2004).

2.3 Level B simulations: calibrated dates and alternative segmentation methods

Level B simulations extend Level A by including radiocarbon measurement error and calibration, and by comparing the ensemble rule with two alternatives: the naive single-KDE minima rule and a Poisson changepoint model fitted to binned counts.

The simulation window spans 2000–5000 cal BP. Three pulses of width 400 yr are centred at 4500, 3500 and 2500 cal BP, with equal intensities and a background intensity $\lambda_{bg} = 0.1$, so that pulses contribute approximately 87% of expected events. As in Level A, the low background component prevents the simulations from collapsing into an unrealistically clean pulse-only system and approximates the low-rate dated events expected outside major activity clusters. All pulses have the same height in Level B to focus on calibration effects; the structure is analogous to Table 1 but expressed in calendar years BP. As with Level A, this is a deliberately simplified configuration: real sequences will show unequal pulse heights, variable widths and superimposed long-term trends, so the

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

performance estimates reported here should be read as benchmarks under idealised conditions rather than as universal expectations.

For each true event time t_i^{true}, an expected radiocarbon age μ_i is first obtained from a specified calibration curve by uncalibration. Three calibration setups are considered:

1. Pure IntCal20 (Reimer et al. 2020),
2. Pure SHCal20 (Hogg et al. 2020),
3. A simple "mixed" configuration in which each event is randomly assigned to IntCal20 or SHCal20 with equal probability.

These three setups are included not because large curve-specific differences are expected a priori, but to test whether the broad conclusions of the method are robust to reasonable hemispheric calibration structure. The mixed configuration is therefore a simple heterogeneity stress test rather than a literal demographic scenario.

In all cases, a laboratory measurement A_i is then simulated from a normal distribution with mean μ_i and standard deviation $\sigma_A = 30$ ^{14}C yr. These simulated radiocarbon ages and their associated errors are recalibrated using the same curve assignment that generated μ_i, yielding posterior discrete distributions for each date. The KDE ensemble is then constructed as in Section 2.1, using $R = 300$ posterior draws and a fixed grid of 400 calendar points between 2000 and 5000 cal BP. The simulation pipeline is therefore: true calendar ages → expected radiocarbon ages under a chosen curve → laboratory error → recalibration → Monte Carlo KDE ensemble.

The Level B design varies $N \in \{50, 100, 200, 500\}$, $h \in \{80, 100, 120\}$ yr and calibration setup, and performs 200 replicates per combination, per segmentation method (ensemble rule, naive KDE, Poisson changepoint). The Poisson model operates on counts in 50-yr bins and assumes a piecewise-constant intensity; the number and locations of changepoints are estimated with the PELT multiple-changepoint algorithm in the changepoint package (cpt.meanvar, test.stat = "Poisson") using a Schwarz information criterion (SIC, i.e. BIC) penalty on the log-likelihood (Heaton et al. 2025; de Navascués et al. 2025).

2.4 Performance metrics

For both simulation levels, performance is evaluated by comparing the inferred number of pulses $\hat{K}$ and their boundaries with the true configuration:

- **Pulse-count accuracy.** For each replicate, $\hat{K}$ is compared to the true $K_{\text{true}} = 3$. Results are summarised by the probabilities of correct recovery (p_{correct}), under-segmentation (p_{under}, $\hat{K} < K_{\text{true}}$) and over-segmentation (p_{over}, $\hat{K} > K_{\text{true}}$).
- **Boundary error.** Boundary accuracy is summarised in two ways. First, for all replicates, an unconditional mean absolute boundary error (yr) is computed using nearest-boundary matching. This metric remains useful descriptively but becomes difficult to compare across methods when they differ strongly in their tendency to over-segment. Second, for replicates with $\hat{K} = K_{\text{true}}$, a conditional mean absolute boundary error is computed by ordered one-to-one pairing of sorted estimated and true boundaries. This conditional metric provides the cleaner comparison of temporal localisation once the correct pulse count has been recovered. For replicates with $\hat{K} \neq K_{\text{true}}$, supplementary tables also report the unconditional error for incorrect-K cases.

These metrics separate recovery of the overall pulse structure from temporal localisation once that structure has been obtained. This dual perspective mirrors recent simulation-based evaluations of radiocarbon-frequency methods and avoids treating boundary precision as commensurate across methods with different segmentation tendencies (Contreras and Meadows 2014; Bevan and Crema 2021; Crema 2022). A compact synthesis across information regimes is provided in Table 2, while detailed simulation values appear in Supplementary Tables S1–S3.

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

Table 2. *Condensed qualitative summary of envelope KDE–minima performance across simulation regimes. Statements are approximate syntheses derived from Supplementary Tables S1–S3 for bandwidths 80–120 yr.*

Scenario	Conditions	Envelope rule performance
Low information	$N \lesssim 50$, any separation	$p_{\text{correct}} < 0.5$; frequent under- and over-segmentation, with mean boundary errors often exceeding $\sim$ 300 yr.
Medium information (calendar time)	Level A; $N = 100$–200, centre-to-centre separations approaching $\sim$ 800 yr	Recovery improves only in the widest-separation settings; p_{correct} commonly rises into the $\sim$ 0.5–0.8 range there, with mean boundary errors on the order of $\sim$ 50–135 yr under idealised calendar-time simulations.
Medium information (calibrated)	Level B; $N = 100$–200, three pulses plus background	Calibration reduces recovery and broadens boundaries; for the envelope rule, p_{correct} is typically in the $\sim$ 0.25–0.55 range, with unconditional mean boundary errors commonly around $\sim$ 200–300 yr. Naive KDE can achieve higher p_{correct} and, conditional on correct recovery, often tighter boundary placement in some settings, but with more over-segmentation; the Poisson changepoint model systematically over-segments under this simulation design.
High information	Level B; $N500$	The envelope rule becomes more stable, with p_{correct} commonly in the $\sim$ 0.4–0.7 range and unconditional mean boundary errors commonly in the $\sim$ 130–240 yr range. Naive KDE often achieves higher p_{correct} and tighter conditional boundary placement, but still at the cost of more frequent spurious pulses.

3 Results

3.1 Level A: calendar-time pulses

Level A simulations confirm that segmentation performance is governed primarily by episode spacing, sample size and bandwidth. When pulses are strongly overlapping ($\Delta = 200$ or 400 yr, with fixed pulse width = 400 yr), all methods struggle: p_{correct} remains low and both under- and over-segmentation are common, even at $N = 500$. Under these conditions, the smooth intensity itself contains limited information about distinct episodes.

For well-separated pulses ($\Delta \approx 800$ yr), performance improves markedly. With $N200$ and moderate bandwidths ($h = 80$–120 yr), the ensemble rule recovers the correct three-pulse structure in a majority of replicates, and boundary errors are typically on the order of one to two bandwidths (Figure 1). However, the method remains conservative: at small and moderate sample sizes ($N = 50$–100), under-segmentation is more common than over-segmentation, and pulses tend to be temporally diffuse.

Figure 2 summarises p_{correct} for the envelope-based rule as a function of N, centre-to-centre separation Δ, and h. High values of p_{correct} are confined to cells with both large N and wide separation

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

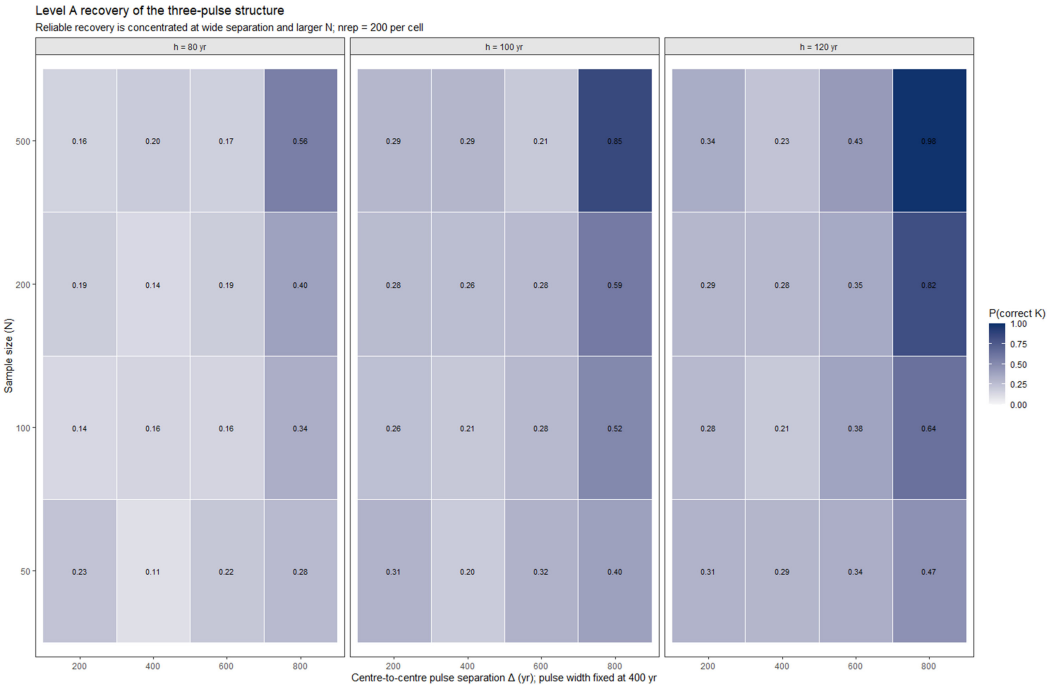


Figure 2. *Level A recovery of the three-pulse structure under the envelope KDE–minima rule. Heatmaps show the probability of correct pulse-count recovery, P(K-hat = K-true), across sample size (N = 50, 100, 200, 500), centre-to-centre pulse separation (Δ = 200, 400, 600, 800 yr), and bandwidth (h = 80, 100, 120 yr). Pulse width is fixed at 400 yr. Reliable recovery is concentrated in the widest-separation settings and improves with larger sample size, showing that recoverability depends jointly on spacing, information content, and smoothing scale. Each cell summarises 200 Monte Carlo replicates.*

(bottom-right corner of each panel), reinforcing the conclusion that KDE–minima segmentation is reliable only under favourable conditions. Table 2 condenses these patterns into four qualitative regimes (low, medium calendar-time, medium calibrated and high information), with approximate ranges for p_{correct} and boundary error.

Compared to the naive single-KDE minima rule, the ensemble rule substantially reduces over-segmentation. The naive rule tends to interpret shallow shoulders and local irregularities as separate pulses, especially at smaller N and smaller h, yielding high p_{over}. The ensemble rule discards many of these minima because they are not stable in the lower envelope, trading a modest loss of sensitivity for greater robustness against spurious episodes. This behaviour is consistent with general insights from multi-scale curve exploration, where persistent features are privileged over local noise (Chaudhuri and Marron 1999). Supplementary Table S1 and Supplementary Figure S02 give full numerical and graphical comparisons for the Level A scenarios. Curve-specific Level B KDE results are reported in Supplementary Table S03 and Supplementary Figure S03.

3.2 Level B: calibrated dates and segmentation methods

Level B simulations show that calibration further degrades segmentation performance. For a given combination of N, h and curve assignment, p_{correct} is generally lower than in Level A, and boundary errors increase. This reflects the additional smoothing and distortion introduced by the calibration curve

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

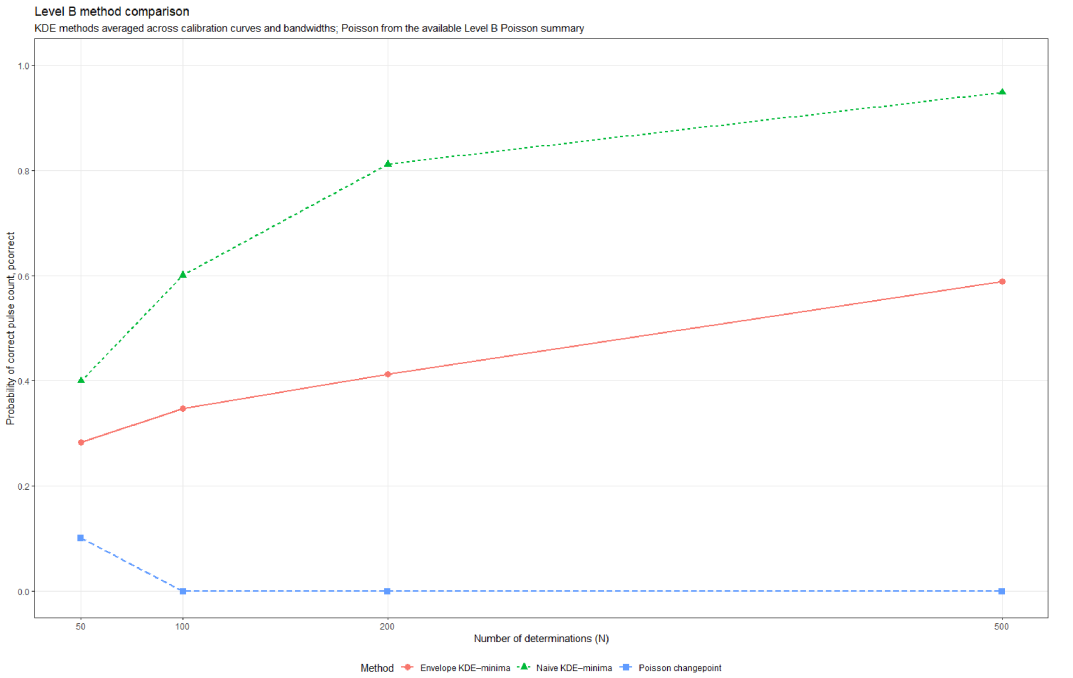


Figure 3. *Collapsed Level B comparison of segmentation methods after averaging KDE results across calibration curves and bandwidths. The plot shows the probability of correct pulse-count recovery for the envelope KDE–minima rule, the naive single-KDE minima rule, and the Poisson changepoint model as a function of sample size. It is intended as a summary of method trade-offs rather than a substitute for the curve- and bandwidth-stratified supplementary results: naive KDE achieves higher recovery but with greater risk of over-segmentation, the envelope rule is more conservative, and the Poisson changepoint model over segments systematically under the present calibrated design.*

and measurement error, as well as the non-linear relationship between true calendar time and radiocarbon age (Reimer et al. 2020; Hogg et al. 2020; Bronk Ramsey 2017).

Differences between calibration setups (IntCal20, SHCal20, mixed) are consistently small relative to the effects of sample size, bandwidth and segmentation rule. This is a negative but informative result, because the three curve setups were included precisely to test whether the broad conclusions are robust to reasonable hemispheric calibration structure. At the multi-century scale considered here, the mixed configuration behaves similarly to the pure curves, suggesting that curve choice is second-order relative to the basic information content of the sample.

Figure 3 compares p_{correct} for the three segmentation methods as a function of N, averaging over bandwidths and calibration setups. It should therefore be read as a collapsed method-comparison synthesis rather than as a substitute for the curve- and bandwidth-stratified supplementary outputs. Curve-specific Level B KDE results are shown in Supplementary Table S3 and the corresponding supplementary figure. Within this synthesis, both KDE-based methods improve with increasing N, but they differ in trade-off structure: the envelope rule is more conservative and suppresses many spurious minima, whereas the naive KDE rule can attain higher p_{correct} at the cost of more frequent over-segmentation. The Poisson changepoint model performs poorly under the present design. In the curve-specific Poisson outputs, p_{correct} is low at $N = 50$ and falls to zero at $N = 100$, 200 and 500 across all three calibration setups, while p_{over} is systematically high. The Poisson model is therefore included here as a contrasting model-fitting benchmark rather than as a superior segmentation solution.

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

Table 3. *Modelled boundaries, medians and 95.4% highest posterior density (HPD) intervals for the three Bayesian episodes of the Kotosh Religious Tradition (KRT) regional sequence. Times are in cal BCE. These values are taken from the OxCal phase model reported in Mesia-Montenegro (2025) and provide the Bayesian frame of reference for the KDE–minima pulses shown in Figure 4.*

Model element	Median (cal BCE)	Start (95.4% HPD)	End (95.4% HPD)
Phase 1			
Boundary Start 1	−3400	−4004	−2991
Boundary End 1	−3000	−3280	−2680
Interval lapse between 1 and 2	295	0	591
Phase 2			
Boundary Start 2	−2700	−2800	−2500
Boundary End 2	−1200	−1260	−1097
Interval lapse between 2 and 3	407	166	619
Phase 3			
Boundary Start 3	−800	−1001	−544
Boundary End 3	−550	−750	−200

Because methods differ sharply in their tendency to over-segment, unconditional boundary error alone is not sufficient for comparing temporal localisation. Supplementary Table S3 therefore also reports boundary error conditional on correct pulse-count recovery. Under this stricter comparison, the naive KDE rule typically yields tighter boundary placement than the envelope rule, whereas the main advantage of the envelope rule lies in its more conservative segmentation and lower susceptibility to unstable minima.

Overall, high probabilities of recovering the true three-pulse structure are achieved only under favourable conditions: widely separated episodes and substantial sample sizes. Even then, boundary estimates remain uncertain at the scale of several tens to a few hundred years. The main practical contrast is therefore not between higher sensitivity and higher precision, but between a conservative envelope rule that limits spurious splitting and a more aggressive naive rule that often localises boundaries more tightly when correct, but at a substantially higher risk of extra pulses.

3.3 Empirical illustration: Kotosh Religious Tradition

To illustrate the method on real data, the ensemble KDE–minima procedure is applied to a published radiocarbon corpus for the Kotosh Religious Tradition (KRT) in the Central Andes. That dataset comprises 170 determinations from 16 Late Archaic and Formative ceremonial sites, modelled in OxCal as a three-episode regional sequence with two nested subtraditions (Mito Tradition and Kotosh Broad Tradition) (Mesia-Montenegro 2025). The present analysis does not introduce new chronological modelling; it uses posterior calendar-age samples from the published OxCal model as input to the KDE–minima algorithm, treating the Bayesian model as a benchmark.

Table 3 summarises the three Bayesian KRT episodes, reporting median boundary ages and 95.4% highest posterior density (HPD) intervals. Posterior samples for all KRT events are pooled and analysed with bandwidths of 80, 100 and 120 yr and a 95.4% envelope. For each bandwidth, the ensemble KDE and persistent minima are computed, and the resulting pulses are compared qualitatively with the three Bayesian regional episodes. Figure 4 summarises this bandwidth sensitivity: the main three-episode structure is robust across bandwidths, but the exact locations of minima and the sharpness of intervening troughs vary. In particular, edge pulses and minor shoulders are more unstable than the central episode, consistent with the simulation results.

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

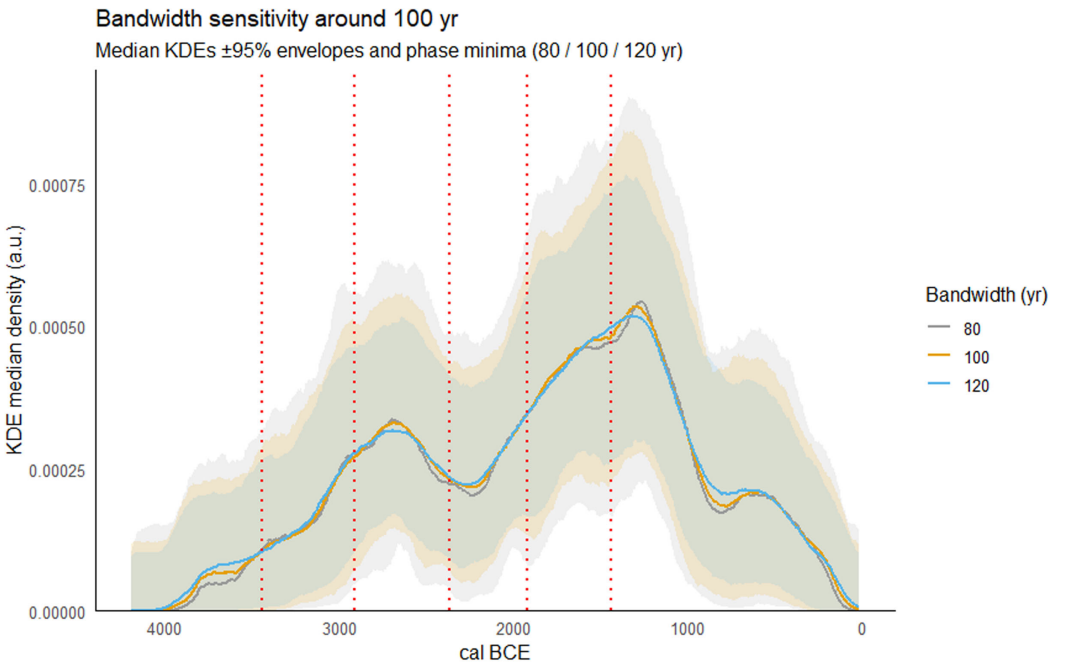


Figure 4. *Bandwidth sensitivity of the Kotosh Religious Tradition illustration. Median KDEs and 95.4% envelopes are shown for bandwidths h = 80, 100, and 120 yr, with vertical dotted lines marking minima identified across the examined bandwidths. The principal multi-episode structure is broadly stable across this range, whereas the exact positions of minima, the sharpness of intervening troughs, and edge or shoulder features remain bandwidth-sensitive. The figure is therefore intended to show robustness of the broad pattern together with uncertainty in finer segmentation details.*

The KRT example is therefore best read as a demonstration of behaviour on a large, archaeologically structured dataset with a known Bayesian benchmark. It highlights both the utility and the limits of the method: the ensemble rule can summarise broad episodes in a manner broadly consistent with a more complex phase model, but it cannot substitute for explicit modelling when precise boundaries or site-level differences are required (Crema 2022; Mesia-Montenegro 2025). A brief SiZer-based sensitivity check is included in the supplementary material as an additional multi-scale diagnostic rather than as an alternative phase estimate. Under the stricter operational rule adopted in the revised analysis, no stable SiZer episodes were identified for the high-information Level A scenario, reinforcing the scale dependence of exploratory segmentation.

4 Discussion

4.1 Positioning within radiocarbon-frequency methods

The KDE–minima procedure is best understood as a descriptive segmentation of dated-event intensity within the reconstructive family of radiocarbon-frequency approaches. It does not estimate population size directly and is not designed to yield posterior statements about episode number, timing or magnitude. Its practical niche is narrower: to formalise the common archaeological practice of visually reading broad phases from SPD or KDE curves while making the smoothing dependence explicit and limiting spurious minima (Crema and Bevan 2021; Crema 2022). In that sense it complements, rather than competes with, hypothesis-testing approaches and explicit model-fitting frameworks. When the substantive target is posterior inference on rates, changepoints or episode intensity, Poisson- or Cox-process models remain preferable (Heaton et al. 2025; de Navascués et al. 2025).

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

4.2 Practical guidance and limitations

The simulations make clear that KDE–minima pulses should be interpreted cautiously. Figure 5 provides a decision chart that synthesises the main lessons; key points are:

- **Sample size and separation.** Reliable recovery of episode counts and approximate boundaries occurs only when pulses are separated by several bandwidths (here $\approx$ 800 yr, with pulse width fixed at 400 yr) and when each episode is supported by on the order of 100 or more determinations. At smaller N or with closely spaced pulses, under- and over-segmentation are common, consistent with earlier warnings about the limits of demographic inference from sparse radiocarbon series (Contreras and Meadows 2014; Crema 2022). This regime-based view is summarised in Table 2.
- **Bandwidth choice.** The bandwidth h controls the scale at which pulses are visible. Here $h = 80$–120 yr is used as a decision-relevant multi-century range, not as an estimate of an optimal smoothing parameter. The objective is conservative segmentation of broad episodes a few hundred years wide, not recovery of the finest local features in an SPD or KDE curve. Researchers should therefore inspect whether $\hat{K}$ and boundary locations remain stable across a plausible range of h rather than privilege a single value (Silverman 2018; Scott 2015; Sheather 2004; Crema 2022). The KRT example (Figure 4) illustrates both robust features and bandwidth-sensitive details.
- **Calibration effects.** Calibration-curve structure and measurement error further blur episodes and can introduce spurious features. The three Level B curve setups were included to test robustness to reasonable hemispheric calibration structure, not because large differences were expected a priori. Their similarity is therefore informative: at the multi-century scales examined here, curve choice is second-order relative to sample size, bandwidth and segmentation rule. The mixed configuration should be read as a heterogeneity stress test rather than as a literal demographic scenario (Reimer et al. 2020; Hogg et al. 2020; Crema and Bevan 2021).
- **Method trade-offs.** Under the present calibrated design, the naive KDE rule often attains higher pulse-count recovery and, conditional on correct recovery, typically tighter boundary localisation than the envelope rule. Its cost is a markedly higher tendency to over-segment. The envelope rule's main advantage is therefore not sharper timing but lower-risk segmentation through the suppression of unstable minima. The Poisson changepoint model, in contrast, systematically over-segments and does not provide a competitive summary for these simulated data. Figure 3 should therefore be read as a synthesis of these trade-offs rather than as a universal ranking.
- **Uncertainty.** The method as implemented here provides no direct probabilistic measure of uncertainty in K or in boundary locations for empirical data. A natural extension would be to summarise the distribution of $\hat{K}$ and of minima positions across Monte Carlo replicates and across bandwidths, thereby providing informal credibility intervals for pulses (cf. Efron and Tibshirani 1994; Chaudhuri and Marron 1999).

The method is therefore best viewed as a low-risk, exploratory tool for summarising broad episodes in well-sampled sequences, not as a general solution to the phase-identification problem. For many research questions, especially those requiring precise phase boundaries or strong causal inference, explicit Bayesian or likelihood-based event-count models remain preferable (Crema and Bevan 2021; Heaton et al. 2025; de Navascués et al. 2025).

4.3 Future directions

Several extensions remain necessary. The simulations used equal-width pulses and, in Level B, equal heights; heterogeneous pulse shapes, overlapping shoulders and superimposed long-term trends are common in real data and will likely reduce performance further. Non-stationary sampling intensity and taphonomic loss were not explicitly modelled, and integrating plausible sampling processes would sharpen guidance on minimal data requirements and likely biases (Surovell et al. 2009; Bevan and

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

Decision chart for KDE–minima segmentation

Step 1: Sample size (N).

If $N < 50$: *do not attempt segmentation*; use SPDs/KDEs descriptively only.
If $50 \leq N \leq 200$: proceed to Step 2.
If $N > 200$: proceed to Step 2 and, if finer structure is of interest, consider formal Bayesian phase or changepoint models in addition to descriptive KDEs.

Step 2: Expected temporal scale of episodes.

If episodes are decadal or poorly constrained in duration: prefer descriptive SPDs/KDEs or model-based changepoint/phase analysis; KDE–minima is not recommended.
If episodes are multi-century (widths $\sim$300–500 yr) and suspected gaps between them are $\gtrsim$ 600–800 yr: proceed to Step 3.

Step 3: Calibration roughness.

If individual calibrated posteriors are strongly multimodal (wiggly calibration curve, mixed curves, or heterogeneous errors): treat KDE–minima as exploratory and check robustness with simple simulations; Poisson/Bayesian changepoint models may be preferable.
If posteriors are relatively smooth and the dataset resembles the simulation designs (single curve, homogeneous errors): apply the envelope-based KDE–minima rule with $h \approx$ 80–120 yr.

Step 4: Interpretation.

Report pulses as broad clusters of dated activity rather than sharp cultural phases; state the bandwidth and envelope level used, summarise simulation-based limitations, and interpret pulses alongside independent archaeological evidence.

Figure 5. *Heuristic decision chart for KDE–minima segmentation of radiocarbon sequences. The chart summarises the main lessons of the simulation study: segmentation is most defensible for broad multi century episodes, sufficiently large samples, and relatively smooth calibrated posteriors; calibration roughness, small sample sizes, and closely spaced episodes increase the risks of under- and over segmentation. The chart should be read as a practical guide derived from the simulation designs examined here, not as a universal thresholding rule for all archaeological radiocarbon datasets.*

Crema 2021; Crema 2022). Empirical uncertainty summaries for $\hat{K}$ and boundary positions, and closer integration with recent Poisson-process event-count models, remain the most useful next steps (Heaton et al. 2025; de Navascués et al. 2025).

5 Conclusions

The Monte Carlo KDE–minima procedure can recover broad activity pulses only under restricted conditions: episodes must be widely separated relative to their width, sample sizes must be on the order of at least 100–200 dates, and the resulting segmentation must remain stable across a defensible bandwidth range. Under poorer information conditions, especially after calibration, the method readily merges episodes or produces temporally diffuse boundaries. It should therefore be used for descriptive segmentation of dated-event intensity, not as a direct estimator of population history.

Within that restricted role, the envelope rule offers a conservative, low-risk summary because it suppresses unstable minima. Its advantage is not sharper boundary localisation: conditional on correct recovery, naive KDE typically places boundaries more tightly, but it does so alongside a much higher tendency to over-segment. At the multi-century scale examined here, differences among IntCal20, SHCal20 and the mixed calibration setup are small, so curve choice is secondary to sample size, bandwidth and segmentation rule.

A practical decision rule follows. Use KDE–minima segmentation only when the question concerns broad episodes, the dated sample is sufficiently large, and the segmentation is stable across a plausible

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press

bandwidth range. When the target is precise phase boundaries or posterior inference on episode number, timing or intensity, explicit Bayesian or likelihood-based event-count models are preferable. The Kotosh Religious Tradition example shows that, under favourable conditions, the ensemble rule can provide a rapid summary consistent with a published Bayesian benchmark, but at coarser temporal resolution.

Supplementary material. To view supplementary material for this article, please visit https://doi.org/10.1017/RDC.2026.10258

Code and data availability. All R scripts, supplementary tables and figures are archived in an OSF open repository https://doi.org/10.17605/OSF.IO/JZ3PE, enabling reproduction and extension of the analyses. The repository includes R scripts for the Level A and Level B simulation experiments, functions implementing the Monte Carlo KDE–minima and naive segmentation rules, code for computing the performance metrics reported in the tables and figures, and scripts for fitting the Poisson changepoint model and running the SiZer comparison.

Acknowledgments. I would like to thank the reviewers for constructive comments on earlier versions of this work and the Universidad Privada del Norte for supporting this research.

References

Bevan A and Crema ER. (2021) Modifiable reporting unit problems and time series of long-term human activity. *Philosophical Transactions of the Royal Society B: Biological Sciences* **376**(1816), 20190726.

Bronk Ramsey C (2017) Methods for summarizing radiocarbon datasets. *Radiocarbon* **59**(6), 1809–1833.

Chaudhuri P and Marron JS (1999) Sizer for exploration of structures in curves. *Journal of the American Statistical Association* **94**(447), 807–823.

Contreras DA and Meadows J (2014) Summed radiocarbon calibrations as a population proxy: a critical evaluation using a realistic simulation approach. *Journal of Archaeological Science* **52**, 591–608.

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**(4), 1387–1418.

Crema ER. and Bevan A (2021) Inference from large sets of radiocarbon dates: software and methods. *Radiocarbon* **63**(1), 23–39.

de Navascués M, Burgarella C and Jakobsson M (2025) Analysis of the abundance of radiocarbon samples as count data. *Peer Community Journal* **5**, e20.

Efron B and Tibshirani RJ (1994) *An Introduction to the Bootstrap*. Chapman and Hall/CRC, New York.

Heaton TJ, Al-assam S and Bard E (2025) A new approach to radiocarbon summarisation: Rigorous identification of variations/changepoints in the occurrence rate of radiocarbon samples using a poisson process. *Journal of Archaeological Science* **182**, 106237.

Hogg AG, Heaton TJ, Hua Q, Palmer JG, Turney CSM, Southon J, Bayliss A, Blackwell PG, Boswijk G, Bronk Ramsey C et al. (2020) SHCal20 southern hemisphere calibration, 0–55,000 years cal BP. *Radiocarbon* **62**(4), 759–778.

Mendez-Quiros P, Barceló JA, Santana-Sagredo F and Uribe M (2023) Modeling long-term human population dynamics using kernel density analysis of ^{14}C data in the Atacama desert (18º–21ºS). *Radiocarbon* **65**(3), 665–679.

Mesia-Montenegro C (2025) Time, ritual, and bayesian reasoning: Charting the andean kotosh religious tradition. *Archaeological and Anthropological Sciences* **17**, 1–17.

Reimer PJ, Austin WEN, Bard E, Bayliss A, Blackwell PG, Bronk Ramsey C, Butzin M, Cheng H, Edwards RL, Friedrich M. et al. (2020) The IntCal20 northern hemisphere radiocarbon age calibration curve (0–55 cal kBP). *Radiocarbon* **62**(4), 725–757.

Scott DW (2015) *Multivariate Density Estimation: Theory, Practice, and Visualization*, 2 edn. Wiley, Hoboken.

Sheather SJ (2004) Density estimation. *Statistical Science* **19**(4), 588–597.

Silverman BW (2018) *Density Estimation for Statistics and Data Analysis*. Routledge, New York.

Surovell TA, Finley JB, Smith GM, Brantingham PJ and Kelly R (2009) Correcting temporal frequency distributions for taphonomic bias. *Journal of Archaeological Science* **36**(8), 1715–1724.

Timpson A, Colledge S, Crema ER, Edinborough K, Kerig T, Manning K, Thomas MG and Shennan S (2014) Reconstructing regional population fluctuations in the European neolithic using radiocarbon dates: a new case-study using an improved method. *Journal of Archaeological Science* **52**, 549–557.

Cite this article: Mesía-Montenegro C (2026). Identifying activity pulses in radiocarbon sequences: A Monte Carlo kernel–minima approach and simulation assessment. *Radiocarbon* **68**, 1155–1168. https://doi.org/10.1017/RDC.2026.10258

https://doi.org/10.1017/RDC.2026.10258 Published online by Cambridge University Press