A Bayesian alternative for Aoristic analyses in archaeology

Enrico R. Crema[1,2]

[1] McDonald Institute for Archaeological Research, University of Cambridge. Downing Street, CB2 3ER, Cambridge, UK.
[2] Department of Archaeology, University of Cambridge. Downing Street, CB2 3DZ, Cambridge, UK.

Abstract
Aoristic analysis is often used to handle chronological uncertainties of datasets where scientific dates (e.g. ^{14}C, OSL, etc.) are not available, and observations are described by association to archaeological periods or phases. While several advances have been made over the last two decades, the basic principle of this approach remains fundamentally the same. Temporal windows of analyses are first divided into regularly sized time-blocks and probability weight is assigned to each of these for every observation. Weights are then aggregated by time-block, and the resulting vector of summed probabilities is interpreted as a curve representing changes in the intensity over time of a particular phenomenon. This paper reviews the basic principles and assumptions of aoristic analyses in archaeology, highlighting several issues with its application and interpretation and advocating for a Bayesian alternative implemented via *baorista*, a new package written in R statistical computing language. The robustness of the proposed solution is evaluated through a series of experiments based on simulated datasets, which showcase key advantages over aoristic analysis. Two specific solutions are considered, a parametric approach where data are fitted to specific growth models and a non-parametric approach that allows a visualisation of the changing frequencies of events that account for sampling error and the peculiarities of archaeological periodisation.

Keywords
Aoristic Analyses, Bayesian Inference, Chronological Uncertainty, Archaeological Periodisation

1. Introduction

Time is undoubtedly one of the most important and ubiquitous yet elusive and problematic concepts in Archaeology. Ontological and epistemological discussions proliferate the archaeological literature (Murray 1999; Lyman and O'Brein 2006; Bailey 2007), and the opportunity of providing long-term perspectives is a trope that is often used to justify its wider disciplinary relevance. It is undeniable, however, that we recurrently need to face the inferential limits imposed by issues such as chronological resolution, time-averaging, and, more broadly, the problem of temporal uncertainty. The widespread use of radiocarbon and other scientific dating techniques offers new opportunities to address many of these concerns. More recently, the use of Bayesian models have dramatically refined our ability to date key events (Buck et al. 1996; Bayliss 2009), whilst the use of a large collection of radiocarbon record has kindled a growing interest in new research areas such as comparative paleodemography (French et al. 2021; Crema 2022).

Whilst absolute chronology remains the golden standard in modern archaeological research, much of the legacy data at our hands are based on traditional forms of relative chronologies, typically inferred from diagnostic properties of the artefacts we recover. Efforts have been made to translate these periodisations into absolute chronology through the use of radiocarbon dating and Bayesian methods (e.g. Ziedler et al. 1998; Crema and Kobayashi 2020; Brunner et al. 2020), but the substantial majority of archaeological data is still typically attributed to a *period* or a *phase* that is chronologically bounded by a (often loosely defined) start and end date. Most diachronic analyses are thus built around these temporal units, which can be severely limiting in the presence of uneven durations or when, depending on the quality of the diagnostic artefacts, a precise attribution to a particular phase or another is not always possible (Bevan et al. 2012).

The current "go-to" solution in analysing archaeological datasets dated by means of relative chronologies is to employ a suite of techniques with a shared origin in *aoristic analysis*. The technique was initially developed in crime science (Ratcliffe and McCullagh 1998), and after a few early applications in the early 2000s (Johnson 2004; Mischka 2004), it experienced mild success within archaeology (Crema 2012; Baxter and Cool 2016; Yubero-Gómez et al. 2016; Verhagen et al. 2016; Orton et al. 2017; Palmisano et al. 2017; Palmisano et al. 2019; Stoddart et al. 2019; Hinz et al. 2019; Brozio et al. 2019; Knitter et al. 2019; Romandini et al. 2020; Kleijne et al. 2020; Romanowska et al. 2021; Pollard 2021; Roalkvam 2022; Taelman 2022; Levy et al. 2022; Franconi et al. 2023; Hoebe et al. 2023), partly aided by the development of several dedicated R packages such as *aoristic, datplot, archSeries, and kairos* (Orton et al. 2017; Steinmann and Weissova 2021; Frerebeau 2022; Ratcliffe 2022). It is worth noting, however, that some of the core ideas of aoristic analyses were independently introduced within archaeology before and after Raticliffe's seminal paper (e.g. Carlson 1983; Roberts et al. 2012).

The main appeal of aoristic analyses (and similar methods) is rooted in their intuitive handling of chronological uncertainty and the minimum requirements for their implementation (see section 2 below for details). Whilst, in principle, aoristic approaches can be part of any analyses involving time, its primary application in archaeology has been the visualisation of time-frequency data. In practical terms, aoristic analyses offer a solution for replacing the x-axis of bar plots from relative chronologies (e.g. "Early Bronze Age", "Middle Bronze Age", etc…) to a sequence of time blocks representing equally sized time intervals (e.g. 3700-3601 BC, 3600-3501 BC, etc..), whilst accounting for the: 1) the uneven duration of the periodisations and phases, and 2) the uncertainties in the assignment of each observation to these periodsiations and phases.

In this paper, I argue that 1) the archaeological application of aoristic methods is often unwarranted and can occasionally lead to a misleading interpretation of the data, and 2) an alternative approach based on Bayesian inference can solve many of its limitations. Section 2 will briefly review aoristic analyses and related techniques developed within archaeology; section 3 will highlight the main issues of its application in archaeological analyses, with a particular emphasis on its use for time-series of frequency data. Section 4 will introduce the basic principle of a Bayesian alternative, whilst section 5 will examine the robustness of this approach through the analyses of simulated datasets. Finally, section 6 will summarise and discuss the main findings presented in the paper, focusing on limitations and potential future methodological advances.

2. Aoristic Analysis

2.1 Core Concept

Aoristic analysis was initially developed by Jerry Ratcliffe and Michael McCullagh (1998) to provide a quantitative framework for analysing crime patterns where each *event* does not have a precise time-stamp (e.g. “the car was stolen at 2:23 PM”), but instead is described by a *time-span* bounded between two points in time representing the interval within the event occurred (e.g. “the car was stolen sometime between 1pm and 3pm”). The key concern here is that time-spans of events cannot be straightforwardly handled in temporal analyses, and basic queries (e.g. “is the crime rate higher at 2pm or 3pm?”) become difficult. The most common solution involves using mid-points (e.g. treating “the car was stolen sometime between 1pm and 3pm” as if it were “the car was stolen at 2:00pm”), or removing samples with coarse chronological resolutions. Neither are satisfactory and can lead to biased patterns in the data. Ratcliffe and McCullagh’s core intuition was to: 1) discretise time into a series of regularly sized blocks (e.g. 1:01-2:00 pm, 2:01-3:00pm, etc.); and 2) assign *aoristic weights* to each of these time-blocks based on the duration of the time-span. Thus, if an event had a time-span of 12:00pm to 2:00pm, it would have an aoristic weight of 0.5 for 12:01-1:00pm and 0.5 for 1:01-2:00pm. This solution is effectively equivalent to describing the uncertainty of the timestamp θ_i of the event i as a uniform probability distribution:

$$\theta_i \sim \text{Uniform}(\alpha_i, \beta_i) \quad [1]$$

where α_i and β_i are the start and end points of the time-span of the event i. Once these weights are assigned, it is possible to add them by each time-block and display a “temporal weight histogram”, which provides an “indication of the probable temporal distribution of events” (Ratcliffe 2000). In mathematical terms, equation [1] is just a particular way to express measurement error employing a uniform distribution rather than a more conventional Gaussian distribution.

The shared concern with temporal uncertainty in crime science and archaeology led to pioneer adaptations of aoristic analyses in archaeology during the early 2000s (Johnson 2004; Mischka 2004). A key step for this adaptation was to infer the time-span of archaeological events from an evaluation of the diagnostic properties of an object and its attribution to one or more archaeological periods. Thus, a potsherd might be assigned to a narrower dating (e.g. ‘late Medieval’, 1200-1537 CE) or a broader one (e.g. ‘unidentified period’, 6300 BCE-2000 CE).

Despite these early attempts, archaeological applications of aoristic analyses took off only in the subsequent decade, in part promoted by the increased availability of computer scripts for executing calculations but also by the proposal of methodological advances that addressed some, but not all, of its shortcomings.

2.2 Issues with the Application of Aoristic Analyses in Archaeology

The basic premise of aoristic analyses is an intuitive approach that enables, in its simplest form, a straightforward visualisation of chronologically uncertain time-frequency data as a time-series of summed weights. While this undoubtedly provides the basis for a quick visual assessment of the archaeological record, aoristic analysis is characterised by several statistical issues that can potentially lead to unwarranted interpretations and conclusions.

2.2.1 The summation problem

Several authors have highlighted how summing aoristic weights can be "misleading" (Crema 2012), and "does not accord with [...] established mathematical definition of probability" (Collins‐Elliott 2019), raising concerns that are similar to those made for the summed probability distribution of radiocarbon dates (Blackwell and Buck 2003; Carleton and Groucutt 2020; Crema 2022).

The problem is that the summation of aoristic weights hinders the underlying chronological uncertainty when examining changes in the frequencies of events over time. A simple example can illustrate this problem. Suppose we have two scenarios, each with five events (A,B,C,D, and E) and five time-blocks (t_1,t_2, … t_5). In the first case, we assume all events have the same time-span covering the five time-blocks (i.e. they have the same aoristic weights across all time-blocks; w_A = {t_1=0.2; t_2=0.2; t_3=0.2; t_4=0.2; t_5=0.2}, w_B = {t_1=0.2; t_2=0.2; t_3=0.2; t_4=0.2; t_5=0.2}, etc.). In the second scenario, we assume there is no temporal uncertainty but also that each event is assigned to a different time-block (i.e. w_A = {t_1=1; t_2=0; t_3=0; t_4=0; t_5=0}, w_B = {t_1=0; t_2=1; t_3=0; t_4=0; t_5=0}, w_C = {t_1=0; t_2=0; t_3=1; t_4=0; t_5=0}, etc.). In both cases, the sum of the aoristic weights over the five time-blocks will be the same (i.e. w_{Total} = {t_1=1; t_2=1; t_3=1; t_4=1; t_5=1}), and as such aoristic analysis does not distinguish between the two scenarios. If we treat aoristic weights as probabilities and ask ourselves whether there was a change in the frequency of the events from t_1 to t_2, we would reach different conclusions. In the second scenario, we are confident that there are no changes, as we *know* that there was only one event in each of the time-blocks. In contrast, the answer for the first scenario is more complex. We have, in fact, a total of 3,125 possible permutations of our events across the time-blocks, but the number of events in t_1 and t_2 are the same only in a subset of 873 permutations. This means that the probability of 'no change' in the number of events between t_1 and t_2 is only ca. 0.28. Clearly, examining just the sum of aoristic weights does not adequately capture how the frequency of events might have changed over time.

Because the number of permutations becomes quickly intractable with a larger number of events and time-blocks, an analytical solution for deriving probabilities for specific changes in time-frequency is not feasible. Crema (2012) tackles this problem using Monte Carlo simulation, effectively sampling *n* time-frequencies from the universe of all possible permutations and then computing the probability of observing specific scenarios.

2.2.2 Archaeological periodisation

Calculating the time-span of archaeological events generally involves a two-stage process: diagnostic features of each artefact are first associated with one or more archaeological periods, and subsequently, the earliest and latest start and end dates are derived. While the exact procedure might differ case by case, the key point here is that the definition of the time-span is typically based on the association of the focal even to some pre-existing chronological intervals. This step has profound inferential implications: archaeological time-spans are not *random*, and events close in time are more likely to fall into the same archaeological period, resulting in identical time-spans. This is somewhat different from crime events, where events close in time can have different time-spans, and hence observations can be assumed to have independent measurement errors.

Bevan and Crema (Bevan and Crema 2021) have recently shown the implication of non-random measurement error in aoristic analyses by comparing the potential impact of different time-spans assigned to the same simulated dataset. Figure 1 uses a similar simulation approach to showcase how non-random measurement errors derived by archaeological periodisations (fig. 1-a, b, and c) can produce time-series of aoristic sums that can be very different to each other and the ‘true’ underlying time-frequency. In contrast, when time-spans are assigned randomly (fig. -d), aoristic sums can produce sequences far closer to the real time-frequency.

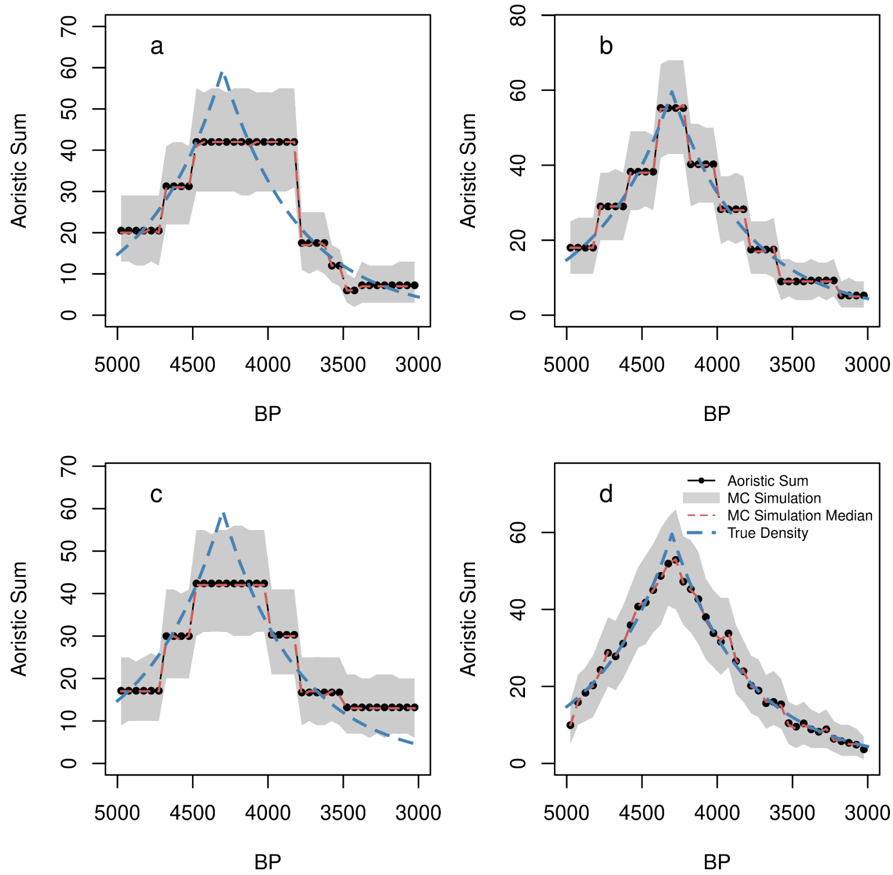

Figure 1. Comparison of aoristic sums and Monte Carlo simulations under different archaeological periodisation (panels a-c) and completely random time-spans (panel d) on the same underlying dataset and statistical population.

Inferential consequences of non-random measurement error in aoristic analysis are clearly dependent on the nature of the underlying archaeological periodisation from which time-spans are derived. Evenly and small-sized phases and periods are less likely to lead to biased outcomes. Conversely, if the underlying archaeological periodisation is coarse and/or has uneven durations, artificial abrupt shifts in the frequency density are likely to be observed at major transitions between phases and periods (e.g. drop in density observed at ca. 3800 BP in figure 1a). The core implication of ignoring the impact of periodisation is the potential of constructing circular arguments. There is an intuitive appeal to assume that social and economic changes co-occur across domains so that, for example, we might expect changes in population size during periods of significant socio-cultural changes, which in turn are reflected by the material record from which we might define our periodisations. This inferential chain remains an untested assumption, and uncritically applying aoristic analyses might lead to a confirmation bias, given that significant changes in aoristic sums will always occur at the transition between archaeological periods and phases.

2.2.3 Descriptive vs Inferential Statistics

The third major limitation of aoristic analyses is that, by its nature, it is a method designed to describe the observed *sample*, and not the underlying *statistical population*. Even when proper treatment of chronological uncertainties is handled via Monte-Carlo simulation and the impact of the underlying archaeological periodisation is negligible, the result of aoristic analyses can only describe the fluctuations in the density of the *sample* events. As a result, any observed changes in the density of the events over time could be a statistical fluke arising from sampling error. To put it more succinctly, at its best aoristic analysis can only be a good descriptive statistic for time-frequencies; it was never designed as a tool to make inferences about the underlying statistical population.

This inferential problem also relates to the granularity of the 'shape' of the temporal density one aims to recover. If one wishes to recover broad trends over multiple millennia, smaller sample sizes and coarser periodisation might well be sufficient. Conversely, if the objective is to recover fluctuations with shorter frequencies and smaller magnitudes, one would require a larger number of events coupled with evenly and regularly sized fine-grained archaeological periodisations. As for any statistical inference, whether the available sample is sufficient in terms of quantity and quality depends on the question that is asked.

2.2.4 Other issues

The three points discussed above are undoubtedly the most common set of limitations and problems associated with aoristic analyses. However, it is worth noting that several other issues have also been raised in the literature.

Crema and Kobayashi (2020) argue that the uncertainty of when an event occurred within its time-span of existence is only one of the three forms of uncertainty associated with archaeological periodisation. They distinguish between *within-phase uncertainty* (the uncertainty on when an event occurred within an archaeologically defined phase), *phase assignment uncertainty* (the uncertainty in assigning an event to a particular archaeological period or phase), and *phase boundary uncertainty* (the uncertainty on the definition of the absolute calendar dates of the start and end of phase), and advocate for an analytical workflow where these uncertainties are estimated from empirical data (e.g. employing Bayesian analyses on radiocarbon dates associated with particular phases) and taken into account together within the Monte-Carlo simulation approach.

Several authors note that a uniform probability distribution (i.e. equation [1] above) is only one possible way to model the uncertainty within the time-span of an event. Baxter and Cool (2016), for example, model this interval using a Beta distribution (technically a Beta-PERT distribution), which provides a flexible range of symmetric and asymmetric unimodal shapes and does include the uniform distribution as a special case. Other authors have similarly used different probability distributions, including Chi-square (Carlson 1983), Gaussian (Bellanger and Husi 2012), Gamma (Steponaitis and Kintigh 1993), and Trapezoid (Crema and Kobayashi 2020). While the exact shape of these distributions differ, they all share an unimodal distribution typically observed in the rise and fall popularity of stylistic traits.

While dimensionless "events" are mathematically straightforward, most archaeological phenomena have some duration in time. If the event is relatively short compared to the chronological resolution of analyses (e.g. the construction of a dwelling), the implications are minimal; events can effectively be treated as points in time. However, when the duration of an event is considerable relative to the time window of analyses (e.g. the occupation of a site), several analytical consequences and challenges arise. Palmisano et al. (Palmisano et al. 2017) approach events with duration by using a modified two-step Monte-Carlo simulation approach. They examined long-term changes in the settlement density of central Italy by first randomly sampling the start date of occupation within the temporal time-span and subsequently simulating the duration of each settlement using a Gaussian distribution informed from the available archaeological record. Collins-Elliot (2019) also employs a Monte-Carlo approach (in this case to examine the abundance of Roman coinage usage) but instead models the duration of coin use via a geometric distribution, with a probability **λ** representing the chances of coin discarded each year.

An under-discussed aspect of the Monte-Carlo simulation approach is the problem of sample interdependence. Consider, for example, a situation where the objective is to assess residential density over time. Now, suppose we have a prehistoric settlement that consists of 20 dwellings, all with an identical time-span between 7,000 and 3,000 BP. How should the dates of the residential units be simulated? If we sample each residential unit independently, our simulations will unlikely contain instances where the 20 dwellings co-exist in time. Indeed, with larger time-spans this will become increasingly unlikely. On the other hand, simulating a single date for 20 residential units comes with a strong assumption of settlement contemporaneity that is not necessarily warranted. Orton and colleagues (2017) discuss this problem in the context of aoristic analyses applied to historic fish consumption in London. They chose to use the context of recovery as an analytical unit, weighting its contribution in the Monte-Carlo approach based on the number of ecofacts recovered.

3 Bayesian Solutions

3.1 Similarities between aoristic analyses and summed probability distribution of radiocarbon dates

Many of the problems associated with the application of aoristic analyses and related methods in archaeology resemble those encountered in the application of summed probability distribution of radiocarbon dates (hereafter SPD). Firstly, the problematic interpretation of summing probabilities of calibrated dates is mathematically equivalent to the issues associated with summing aoristic weights. Secondly, the uncertainty associated with each event is both non-random and linked (at least in part) to their absolute dates. In the same way we observe specific signatures from the underlying archaeological periodisations in aoristic analyses, we observe distinctive peaks and troughs in SPDs that result from the slopes and the plateau of the calibration curve. Thirdly, both SPDs and aoristic sums are descriptive rather than inferential statistics.

A fairly extensive number of papers have raised these statistical problems with SPDs (for a recent review see Carleton and Groucutt 2020; Crema 2022), highlighting how biased their visual assessment can be and advocating for more robust analytical solutions. The similarity of these problems to those observed in aoristic analyses has indeed been mentioned and discussed before (Baxter and Cool 2016; Orton et al. 2017; Collins‑Elliott 2019; Bevan and Crema 2021).

While both aoristic (and related) analyses and SPDs have experienced methodological advances, the latter has benefited from a considerably higher number of new approaches in the last decade. In a recent review article, Crema (2022) has distinguished between three broad approaches: 1) Null Hypothesis Significance Testing (NHST); 2) Non-parametric “reconstructive’ methods, and 3) Model-based inference. All three approaches overcome the core statistical issues of SPDs, either by determining whether and when we observe significant deviations from a particular growth model, reconstructing the overall “shape” of the time-frequency distribution whilst displaying confidence envelopes to account for any fluctuations arising from sampling error, or by fitting parameters of particular growth models to identify rates and timing of significant shifts in density.

Given the similarity of the underlying statistical problem, it is worth evaluating whether some of the solutions developed for SPDs can be adapted to instances where typically aoristic analyses are carried out. For example, the NHST approach developed by Shennan and colleagues (Shennan et al. 2013; Timpson et al. 2014) requires a null growth model and an algorithm for ‘back-calibrating’ calendar dates into ^{14}C ages. Details of the implementation are discussed elsewhere (Timpson et al. 2014), but here it is sufficient to know that the core algorithm generates, via Monte-Carlo simulation, *n* sets of calendar dates given a user-defined growth model. These dates are then ‘converted back’ into ^{14}C ages, calibrated again and then aggregated to generate SPDs. The process is repeated for each set, and the ensemble of SPDs is used to generate an envelope to which the observed curves are compared to. This procedure emulates both the impact of sampling error (each set contains the same number of dates as the observed data) and the information loss entailed by the calibration process. Deviations of the observed SPD from the simulation envelope are interpreted as evidence of departures from a particular null model.

A fundamental step for this Monte Carlo test is the ability to transform dates sampled in calendar time (given a particular growth model) into ^{14}C ages through a back-calibration process. To adopt a similar approach for events described with archaeological periodisations, one would thus require an algorithm that similarly assigns each calendar date to one or more periods (and eventually to a time-span). While theoretically possible, a number of issues need to be addressed. For example, archaeological events are often assigned to multiple periods depending on the number and quality of diagnostic elements (i.e. *phase assignment uncertainty*). This uncertainty is not random, and can have specific structures ((e.g. pairs of phases that can be more or less diagnostically separated; cf Bevan et al. 2012; Crema 2015) that cannot be modelled straightforwardly. It is also worth noting that the NHST approach has several other limitations, ranging from the selection of an appropriate null hypothesis to the limited information provided by significance testing in general (Crema 2022).

A more promising direction to undertake is to apply models that estimate the likelihood of a particular growth model whilst accounting for the chronological uncertainty of each observation. Two recently developed approaches (Timpson et al. 2021; Crema et al. 2022) achieve this, and they can be adapted to the analyses of aoristic data.

Timpson and colleagues (2021) calculate the likelihood of a growth model given a set of observations characterised by chronological uncertainty by effectively using probability mass functions instead of density functions. Thus, the intuition here is to treat time as discrete units, assigning to each year (or a larger unit) a probability. A growth model can thus be described by a multinomial distribution, with a vector of probability values assigned to each calendar year. For practical purposes, these growth models can be described by fewer parameters than the number of calendar years within the window of analyses. For example, an exponential growth model can be reduced to three parameters: a start date *a*, an end date *b*, and a growth rate *r* (see equation [1] in Crema and Shoda 2021). Observed events are similarly described by a vector of probabilities using a discrete form of equation [1] or any other statistical distribution.

In the absence of chronological uncertainty, the probability of observing a sample at time *t,* given a particular growth model, is simply given by the model probability at time *t*. Thus, if the growth model is represented by the vector $\{P_{model}(t=1)=0.1$, $P_{model}(t=2)=0.2$, $P_{model}(t=3)=0.3$, $P_{model}(t=4)=0.2$, $P_{model}(t=5)=0.15$, $P_{model}(t=6)=0.05\}$ and the observed sample has t=2, the likelihood would be to 0.2 (i.e. $P_{model}(t=2)$). If the observation is also described by a vector of probabilities, the likelihood becomes the scalar product of the model and observation vectors. For example, if the observed data is described the vector $\{P_{observed}(t=1)=0$, $P_{observed}(t=2)=0$, $P_{observed}(t=3)=0.4$, $P_{observed}(t=4)=0.4$, $P_{observed}(t=5)=0.2$, $P_{observed}(t=6)=0\}$, the likelihood is equal to the sum of the products $P_{model}(t=1) \times P_{observed}(t=1) + P_{model}(t=2) \times P_{observed}(t=2) \ldots P_{model}(t=6) \times P_{observed}(t=6)$, in this case 0.23. The approach clearly does not depend on how $P_{observed}$ is obtained in the first place. Thus, a vector of probability values obtained from calibrated ^{14}C dates can be replaced with a vector of aoristic weights for each calendar year. Furthermore, although Timpson and colleagues employ a maximum likelihood approach for parameters inference and model comparison, the approach can be extended within a Bayesian framework, where the likelihood of a particular model would be equal to the sum of the log of the scalar products of all observation and the log of the priors of the growth model parameters.

Crema and Shoda (2021) build their inferential framework following Timpson and colleagues' intuition of using probability mass functions but undertaking a slightly different approach for calculating the likelihood. The main difference in their approach is to employ a hierarchical measurement error model where the calendar date of each observation becomes also a parameter. In the case of radiocarbon dates, measurement error is modelled as a Gaussian with mean and standard deviation obtained from a combination of the errors and transformations of the calibration curve (see equation [3] in (Crema and Shoda 2021). An aoristic version of this approach would simply replace the Gaussian with a uniform distribution. Because the time-stamp of each observation is a parameter, the hierarchical approach provides a posterior estimate of the time-span of each observation that is conditional to the overall model and the other observations in the dataset. The approach developed by Crema and Shoda is not dissimilar to other Bayesian models of radiocarbon dates, where one can simultaneously infer higher (e.g. the start and the end of stratigraphic phase) as well as lower (e.g. the calendar date of each ^{14}C sample) level parameters. While an aoristic version of this approach can be implemented straightforwardly, its practical application is limited to cases where the measurement error of each event can be described by a parameterised probability distribution (uniform in this case), as such, its application will not be explored in this paper.

3.2. Parametric and Non-parametric Bayesian alternatives to aoristic analysis

This paper presents and explores the robustness of *parametric* and *non-parametric* Bayesian approaches to aoristic data. In both cases, the objective is to describe the time-frequency of some archaeological events where a vector of probabilities describes each observation over the time-blocks constituting the temporal window of analyses. In the case of the *parametric* approach, the general shape of the time-frequency is assumed a priori (e.g. exponential growth, logistic growth, etc.) and the objective is to recover the parameters of a particular growth model. In the *non-parametric* approach, there are no assumptions on the underlying shape of the time-frequency data other than some degree of temporal autocorrelation (see below). At its core, both approaches define a multinomial model with a sequence of z levels, each corresponding to the time-blocks within the temporal window of analyses. Once we infer the z probability values of a discrete probability distribution, we can compute the likelihood using the approach described in section 3.1

In the case of a parametric model, the actual number of parameters can be drastically reduced to fewer values representing some growth curve limited by a start and end date. Thus, for example, exponential growth can be defined by two parameters: the growth rate r and the number of blocks z. More specifically, the probability p_i assigned to the i-th time-block is simply given by:

$$p_i = (1+r)^i / \sum_{j=1}^{z} (1+r)^j \quad [2]$$

Here, the number of blocks z is user-defined and is determined by the chronological resolution of the analyses (i.e. shorter time-blocks will lead to a larger z). It follows that from an inferential standpoint, we just need to estimate the growth rate r. For a given value of this parameter, we can compute the vector of probabilities $p_1, p_2, \ldots p_z$, and as long as we can characterise each of our observations as a vector of probabilities with the same length (z) we can calculate the likelihood and estimate r. The denominator in equation [2] normalises the growth model into probabilities, and as such, any equation that can characterise a population size changing over z discrete intervals can employ a similar solution (see example in Crema and Shoda 2021; Kim et al. 2021; Timpson et al. 2021).

When a strong assumption on the shape of the time-frequency is not available or when the objective is not the calculation of a general exponential growth rate, a non-parametric approach might be more suitable. The term 'non-parametric' is a misnomer, as the objective in this case is to estimate the vector of probabilities $p_1, p_2, \ldots p_z$ directly. The simplest way to achieve this is to estimate directly these parameters using a multinomial distribution with a symmetric Dirichlet distribution as a prior. Such an approach would, however, consider a wide range of possible shapes in the time-frequency distribution of the events with no a priori assumptions to aid the inferential process. An alternative is to restrict the range of possible shapes whilst keeping sufficient flexibility by adding some weak assumptions on temporal autocorrelation. One such approach is to use an intrinsic Gaussian conditional autoregressive model (ICAR; Besag 1974; Rue and Held 2004) to generate a vector of temporally autocorrelated values $g_1, g_2, \ldots g_z$ and use the softmax function to convert the vector into the probabilities $p_1, p_2, \ldots p_z$. This approach would effectively estimate the probability for given time-slice p_i, conditioning it on the values of the abutting slices p_{i-1} slices p_{i+1}. As a result, the vector $p_1, p_2, \ldots p_z$ will exhibit temporal autocorrelation but still sufficiently flexible to allow for large shifts between p_i and p_{i+1} when the observed evidence is sufficiently strong.

4. Experiment Design, Implementation, and the *baorista* R package

A dedicated R package called *baorista* has been developed to implement the two approaches described in the previous section. At its core, *baorista* provides wrapper and utility functions for fitting Bayesian models using the NIMBLE probabilistic programming language (de Valpine et al. 2017; de Valpine et al. 2020). Users are required to structure their datasets either by defining the time-span of each event (from which aoristic weights are computed assuming a uniform probability distribution) or by providing a matrix containing the probability of each event at each of the time-slices within the window of analyses. The package automatically estimates, via MCMC, parameters of growth models or the vector of probabilities $p_1, p_2, \ldots p_z$ in the case of the non-parametric ICAR model.

In this paper, *baorista* is used on a series of simulated datasets to assess the robustness of the proposed Bayesian solutions. While a comprehensive assessment of these techniques is not viable, four sets of experiments were carried out to answer the following questions:

1. Does a Bayesian approach provide a more accurate and precise estimate of exponential growth rates when compared to regression analyses on aoristic sums? (*experiment #1*)
2. What are the implications of selecting different time-block sizes (i.e. chronological resolution)? (*experiment #2*)
3. What are the inferential limits of coarse archaeological periodisations? (*experiment #3*)

4. How effectively can the non-parametric model recuperate the shape of the time-frequency data under different sample sizes and periodisations? (*experiment #4*)

While details of each experiment differ, in all instances, simulated archaeological samples were generated by first sampling calendar dates from a known probability distribution and subsequently assigning each date into an artificially created periodisation. While this procedure does not account for 'phase assignment' and 'phase boundary uncertainties' (sensu Crema and Kobayashi 2020), it emulates the problems associated with summation, archaeological periodisation, and sampling error.

In experiment #1, 60 datasets with three different sets of sample sizes (n=100, 250, and 500) were generated and analysed. In all cases, the underlying probability distribution had a growth rate of 0.002 with dates sampled between 4999 and 3002 BP. Each replicate was assigned to a randomly generated archaeological periodisation obtained using a breaking stick algorithm based on the Dirichlet distribution. The algorithm consists of first randomly selecting the number of periods between 3 and 10 and subsequently sampling probabilities from a Dirichlet distribution with $\alpha = 0.5$. The so-obtained probabilities were then used to split the duration between 5000 and 3001 BP into periods with durations proportional to the probabilities. Finally each sample falling within a particular period was assigned a time-span equivalent to the start and end date of that period. Each replicate was analysed in two ways. First, an aoristic sum was generated using 10-year resolution time-blocks, and a linear regression was fitted to the log-transformed summed probabilities to estimate the exponential growth rate along with a 95% confidence interval. The same data was then analysed using the Bayesian parametric approach described in the previous section, recording in each case the 95% highest probability density interval (HPDI) of r. We used default MCMC settings implemented in *baorista* (four chains with 100,000 iterations, half discarded for burn-in and with thinning parameter set to 10). The fitting process did not generate any convergence warnings. The prior of the growth rate was modelled using an exponential with a rate of 1.

Experiment #2 loosely followed a similar structure to experiment #1, but with a larger number of replicates (1000), each with a randomly assigned 'true' growth rate (randomly sampled from between -0.002 and 0.002), sample size (between 100 and 500), and periodisation (using the same approach as in experiment 1). For each replicate, we considered two different time-block sizes, one with a coarse setting of 100 years and one with a finer resolution of 10 years. We recorded for each replicate whether the 95% confidence interval (for the regression over aoristic sums) or the 95% HPDI of r included the 'true' growth rate or not. MCMC and prior settings were the default of *baorista* and the same used in experiment #1.

Experiment #3 explored the impact of periodisation in the inferential process, more specifically identifying whether the Bayesian approach introduced here is capable of correctly assessing instances of indeterminacy. The experiment had two stages. First, we consider a range of parameter combinations (growth rate r and inflexion point m) of a logistic growth model ranging between 800 and 301 BC and compute the cumulative probability mass over the intervals 800-501 BC and 500-301 BC, representing the time-span of two hypothetical archaeological periodisations. Because all dates falling within each period are assigned to the same time-span, parameter combinations yielding identical cumulative probability mass over the two intervals are indeterminable. We thus select one of the so-obtained combinations of parameters (r = 0.01154545 and m = 302 BP) and sample 500 calendar dates from the model. Each date was then associated with a time-span based on the abovementioned two periods. The resulting aoristic dataset was then fitted with a logistic growth model using a flat prior for the growth rate ($r \sim Uniform(0.0001,0.03)$) and the inflexion point parameters ($m \sim Uniform(301,800)$). As for experiments #1 and #2, default settings were used for the MCMC.

Finally, experiment #4 assessed the inferential power of the non-parametric approach by determining whether the ICAR model was able to correctly recover the shape of a time-frequency distribution under a combination of different sample sizes (n=50 and n=500) and periodisation (3, 5, and 8 periods, with the duration of each modelled using the Dirichlet distribution with α = 2). The model was fitted over four chains with 4 million iterations, 3 million discarded for burnin and parameters sampled every 100 steps.

All scripts required for replicating the analyses can be found on a dedicated GitHub repository (https://github.com/ercrema/beyond_aoristic), whilst the source code and a quick guide for *baorista* can be found in a separate repository (https://github.com/ercrema/baorista).

5. Results

5.1 Experiment 1

Figure 2 compares the precision and the accuracy of the regression of aoristic analyses and the direct Bayesian approach proposed here. The general expectation of a robust inferential tool is to have stable and high accuracy levels and increased precision (i.e. narrower confidence intervals) with larger samples. There are compelling differences between the two methods in this case. When growth rates are estimated from aoristic sums, the accuracy is generally low, with most the confidence intervals not including the ‘true’ growth rate (here equivalent to 0.002). The precision of the estimate does seem to be weakly associated with sample sizes, but there are significant differences across replicates, indicating how the approach is not robust to sampling error and underlying archaeological periodisations.

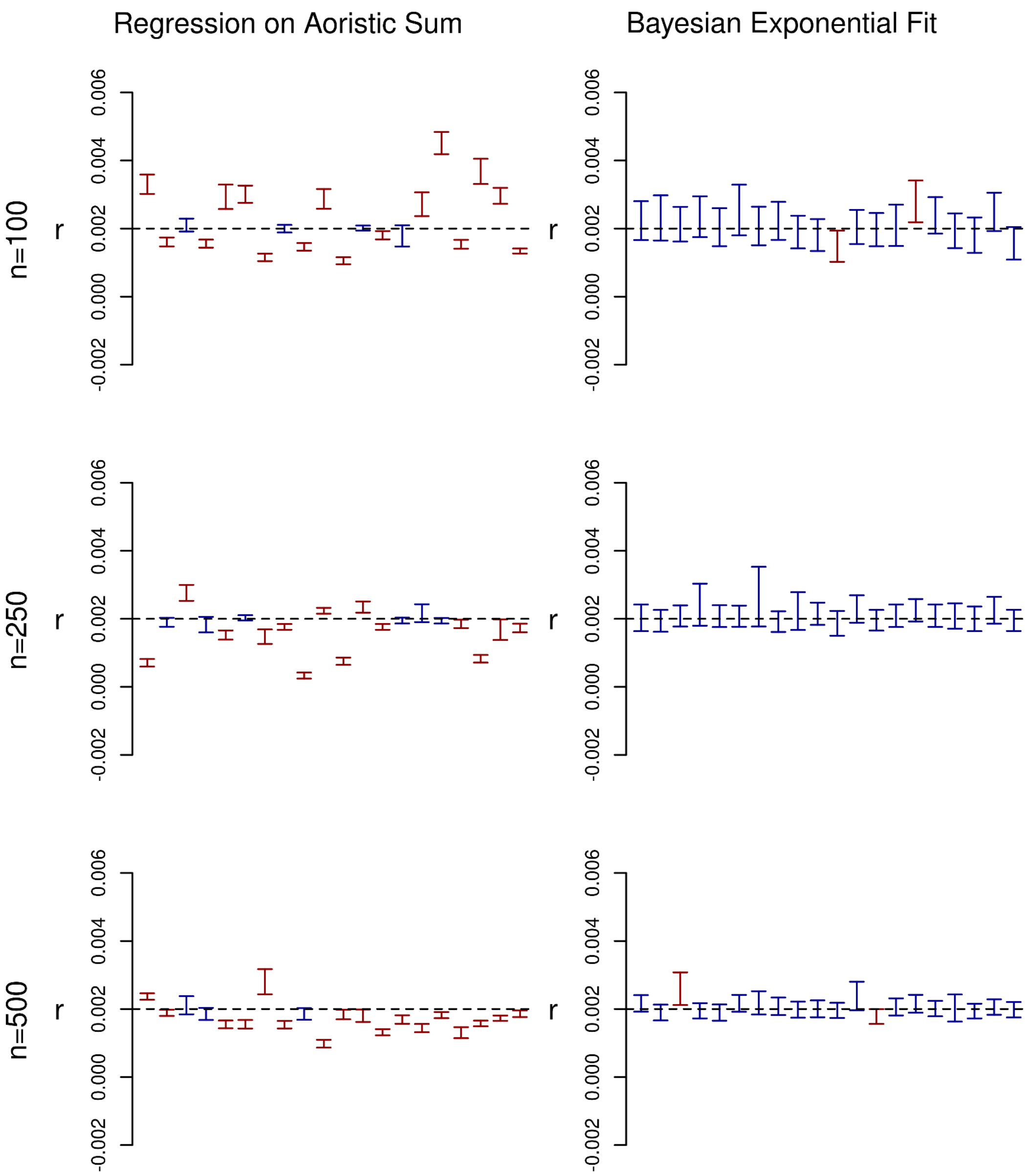

Figure 2. Comparison of growth rate estimates based on regression over aoristic sums (left column; 95% confidence intervals) vs direct Bayesian approach (right column; 95% highest posterior density intervals) on simulated datasets with different sample sizes but same 'true' growth rate (equal to 0.002). Intervals shown in red do not include the true rate within its range.

In contrast, the Bayesian approach shows remarkably higher and consistent levels of accuracy, with only 4 replicates out of 60 sets failing to include the true growth rate. While there are still some differences between replicates, there is a far more consistent association between precision and sample sizes, with larger number of events leading to narrower HPDIs.

5.2 Experiment 2

Results of experiment #2 further highlight the superior accuracy of the Bayesian approach over the regression on aoristic sums. The objective in this case is to compare the two methods with a broader range of settings but also to determine whether the choice of the time-block size has any impact on the inferential process. Figure 3 shows how this is indeed the case, although with opposite outcomes for the two approaches. Estimates based on aoristic sum show generally poor accuracy across different sample sizes (*n*) and 'true' growth rates (*r*), although settings based on coarser time-blocks are more likely to infer *r* correctly.

Once again, the Bayesian approach shows higher levels of accuracy, but this time, the relationship with the choice of time-block size is the opposite. When the resolution is set to 10 years, the model accurately estimated the true growth rate nearly 95% of the time, but this value fell just below 90% when the resolution was decreased to 100-year time-blocks.

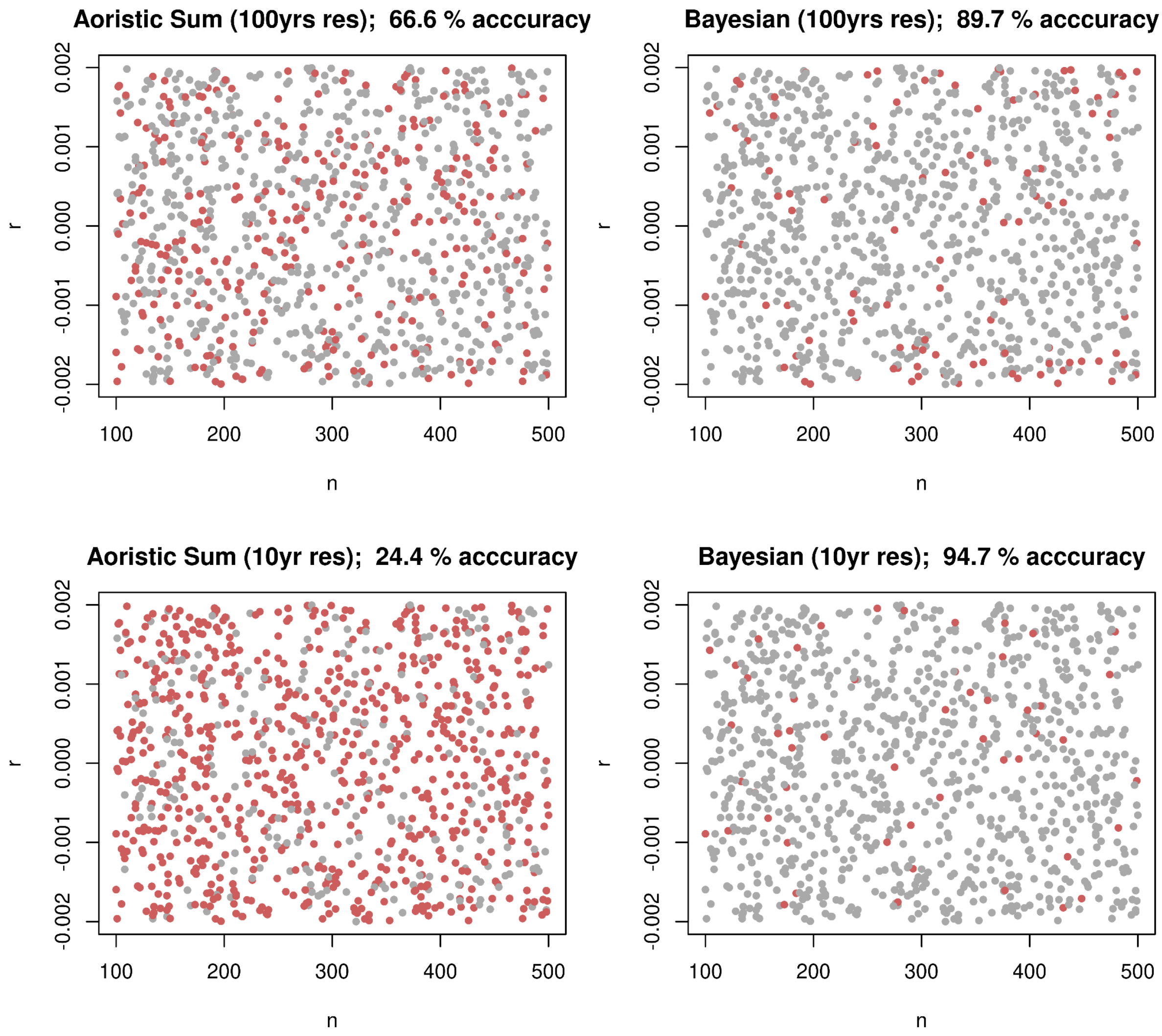

Figure 3 Comparison of growth rate estimates based on regression over aoristic sums (left column) vs direct Bayesian approach (right column) under different sample sizes (*n*), true growth rate (*r*) and time-block resolution (100 and 10 years). Red dots show replicates where the confidence interval or higher prediction interval did not include the true rate.

5.3 Experiment 3

The idiosyncratic and context-dependent nature of archaeological periodisation limits the range of experimental analyses one could pursue. Still, it is possible to examine a particular example that illustrates the limitations imposed by the nature of relative chronology. A simple way to conceptualise this is to consider periodisation as information loss, where different parameter settings of a particular growth model leading to different time-frequency distributions become effectively undistinguishable. Figure 4a, shows a hypothetical example of three logistic growth models with different intrinsic growth rates (*r*) and inflection points (*m*), but having the same cumulative probability mass for the intervals 800-501 BC and 500-301 BC. The three curves are obviously different, but because all dates falling within 800-501 BC are treated in the same way, any shape variations of the curve within this interval are effectively lost when we the chronology is based on these periodisations. Figure 4b is the result of a systematic evaluation of different parameter combinations of *r* and *m*, with instances showing a cumulative probability mass of 0.13 between 800 and 501 BC highlighted in red. While this captures a wide range of shapes, the archaeological record associated with these are expected to yield the same proportion of events for the first and the second of our hypothetical phases.

The objective of experiment #3 is to determine whether the Bayesian approach proposed here is capable of recovering the structure of indeterminacy imposed by the archaeological periodisation. Thus, samples generated from any parameter combinations along the red dots in figure 4b should yield similar results, and the posterior should capture the range of indeterminate parameter combinations. Figure 4c and 4d does indeed show that the proposed Bayesian solution can achieve this. The posterior of time-frequency distribution (figure 4c) shows an envelope capturing a wide range of shapes that includes the extreme cases shown in figure 4a. Similarly, the joint posterior of *r* and *m* (figure 4d) does recuperate the spectrum of indeterminate parameter combinations shown in figure 4b.

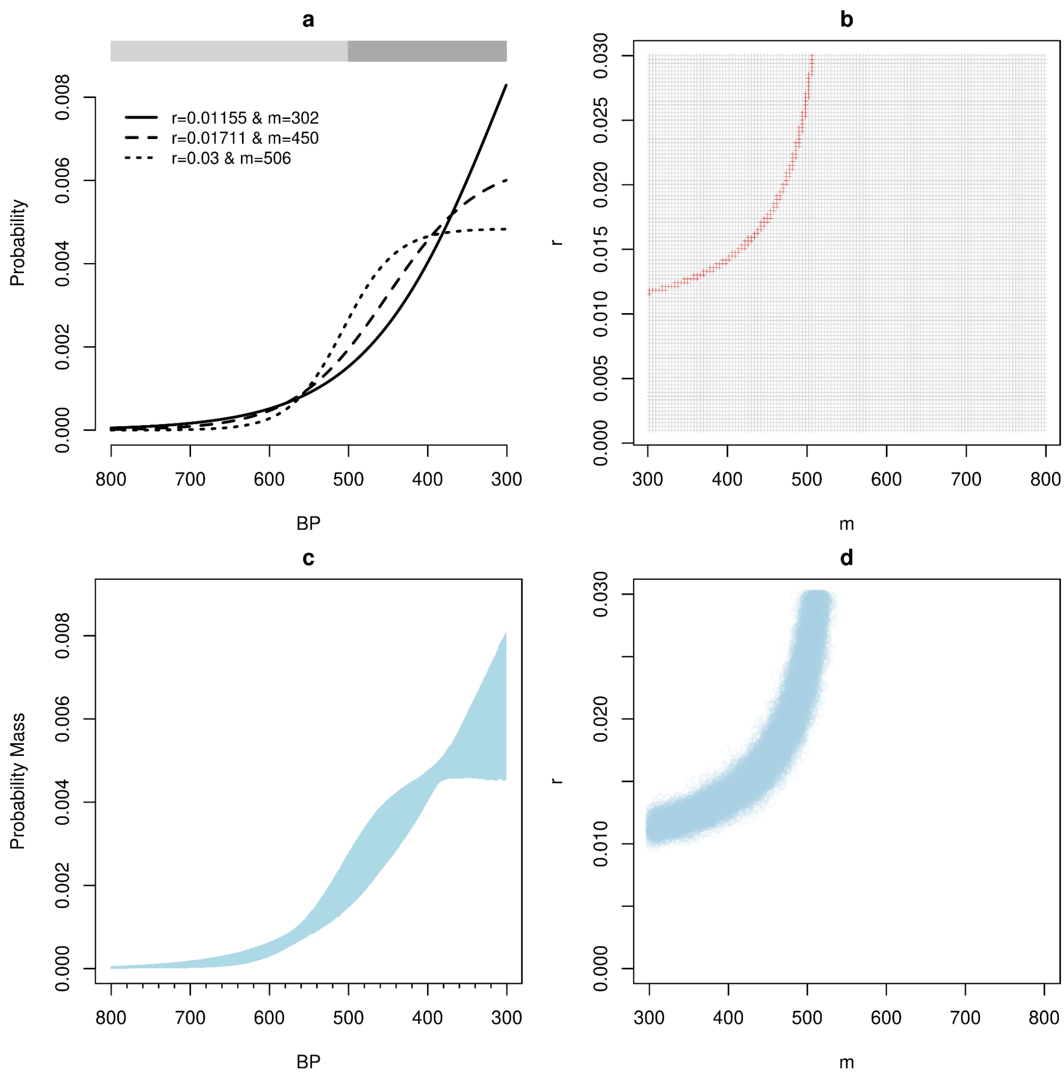

Figure 4. Coarse archaeological periodisation and indeterminacy: a) three parameter combinations of *r* and *m* yielding identical cumulative probabilities for the specific periodisation (phases I and II, shown as different grey bars on the top); b) parameter combinations of *r* and *m* yielding a cumulative probability of 0.13 for phase I and 0.87 for phase II; c) posterior fitted model obtained analysing on a simulated dataset generated with *r*=0.01155 and m=302 (solid line on panel a); d) joint posterior of *r* and *m* of the same analysis.

5.4 Experiment 4

Figure 5 shows the six scenarios explored in experiment #4, with the underlying probability distribution of the time-frequency superimposed over the aoristic sum for different sample sizes and periodisations. Unsurprisingly, instances with a larger number of samples and periods (figure 4-f) show closer similarity between the aoristic sum and the shape of the underlying time-frequency distribution. However, in all cases, spurious fluctuations are also visible.

Results of the non-parametric Bayesian model (figure 6) show that in all scenarios, the 95% HPDI includes the actual shape of the time-frequency distribution from which the artificial samples were generated. The precision of the posterior is conditioned by both the number of samples available and the number of archaeological periodisations, although the latter appears to have a more substantial impact, with the two-period scenario (figure 5a-b) showing the widest HPDI ranges.

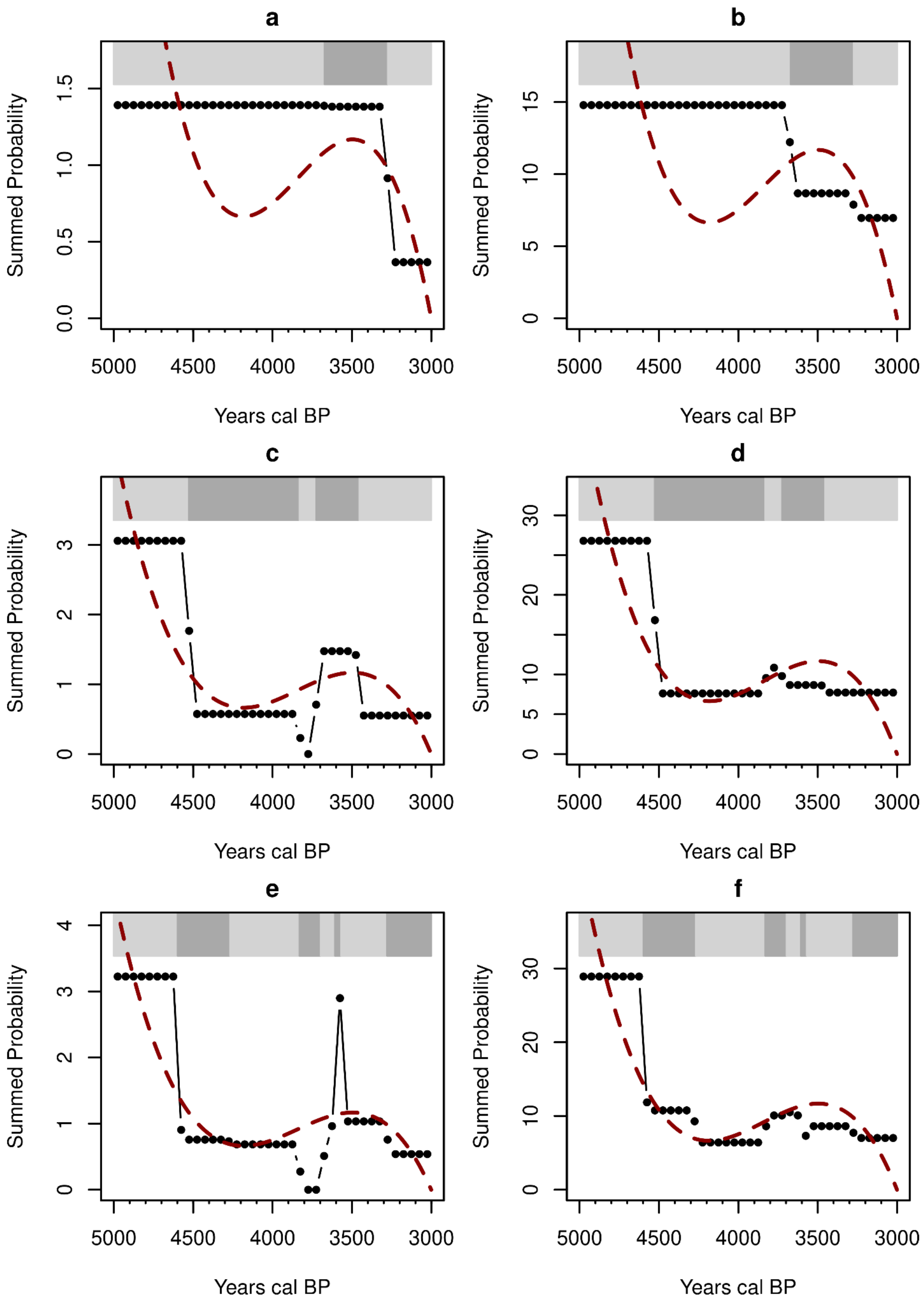

Figure 5. Aoristic sum on simulated data sampled from the same underlying distribution (dashed red line) with different periodisations (top grey bars) and sample sizes (n=50 for *a,c*, and e; n=500 for *b*, *d*, and *f*).

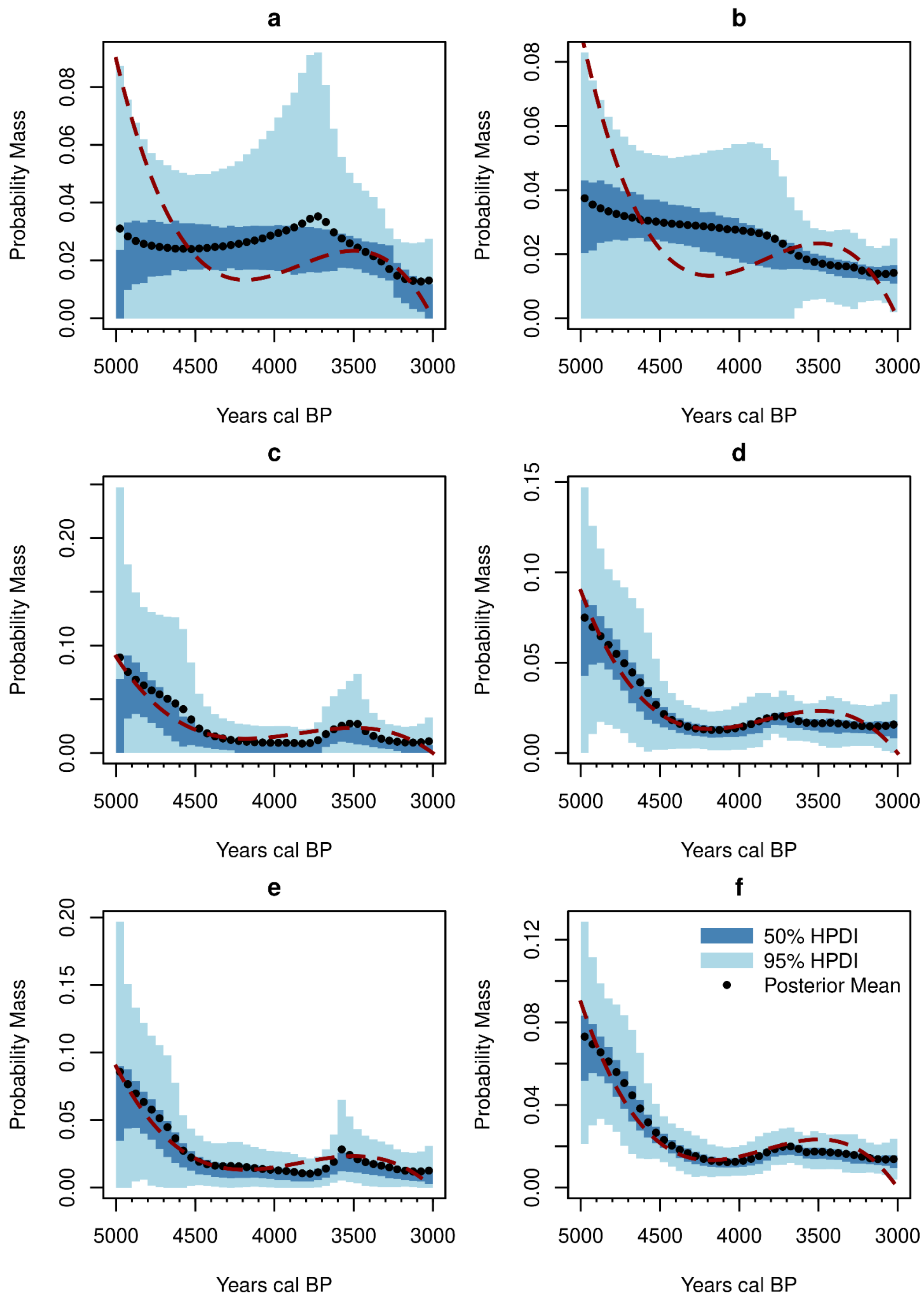

Figure 6. Non-parametric Bayesian estimates of $p_{1,} p_{2,} \dots p_z$ on simulated data sampled from the same underlying distribution (dashed red line) with different periodisations (top grey bars) and sample sizes (n=50 for *a,c*, and e; n=500 for *b*, *d*, and *f*).

6. Discussion

The experiment design devised in this paper is limited to a narrow set of circumstances where the observed data is characterised exclusively by *within-phase uncertainty (Crema and*

Kobayashi 2020). Yet, employing these 'tactical simulations' (Orton 1973) enables us to determine the inferential power of the proposed method and to compare its performance to aoristic analyses. Despite the additional cost in computing performance (see below), the experiments demonstrate that both parametric and non-parametric approaches offer a more robust alternative to aoristic analysis.

Results of experiments 1-3 show that the parametric approach can provide a superior accuracy in recovering the 'true' parameters of a model under a variety of different scenarios. As for SPDs, direct analyses of aoristic sums for statistical inference are unwarranted, and both experiments #1 and #2 demonstrate the implications of pursuing this. Regression estimates over aoristic sums behave inconsistently, with generally a low accuracy and the size of the confidence interval severely impacted by factors external to sample size. In contrast, the parametric Bayesian approach consistently offers higher accuracy, with larger sample sizes yielding higher accuracy as expected.

While the fundamental methodological framework of the parametric approach presented in experiments 1-3 is the same as the one introduced for radiocarbon dates by Timpson and colleagues (2021), the non-parametric ICAR approach has not been explored previously. Experiment #4 shows that the model does capture the uncertainty of the estimates, showing larger posterior envelopes when sample sizes are and archaeological periodisation are coarse. While the 'true' shape of the frequency distribution was contained within the 95% HPDI, the extent to which the model can provide an accurate and precise recovery is somewhat limited. This is not surprising and reflects the amount of information provided as assumption in the inferential process. In other words, a satisfactory performance can be achieved by better quality data (large sample sizes and fine-grained periodisation) or stronger assumptions in the shape of the time frequency distribution (i.e. using a parametric approach).

Aside from the choice between parametric and non-parametric approaches, the methods introduced here require additional decisions to be made by the analyst. As for aoristic analysis, both the parametric and the non-parametric are based on estimates over user-defined time-blocks. Experiment #2 demonstrates that a finer resolution (i.e. smaller time-blocks) provides substantially higher accuracy. This is most likely due to information loss associated with coarser resolution, where the boundary between periods and phases can be out of sync with the breakpoints of the time-blocks. The general recommendation is thus to use the finest resolution, i.e. an yearly time-block. This choice, however, comes with an added computational cost that can quickly become prohibitive in the case of the non-parametric approach, particularly when dealing with larger sample sizes and/or larger window of analyses. In this case, a compromise can be achieved by using a slightly coarser resolution, bearing in mind the potential loss in the accuracy of the estimated parameters.

The choice of the prior is another key aspect of the inferential process, particularly in the case of the parametric approach. As for any standard Bayesian approaches, priors should be weakly informative, excluding unrealistic scenarios but allowing the model to flexibly learn from the data. The impact of these choices should also be properly explored. In some circumstances, such as the scenario presented in experiment #3, the choice of the prior can lead to major differences. Using non-flat prior for *r* and *m* would not recover the full range of indeterminate outcomes shown in figure 4b, as effectively we are stating that some values of *r* and *m* are more likely before even looking at the data. However, if using a non-flat prior is justified on independent grounds, the Bayesian approach can reduce indeterminacy and greatly aid the inferential process.

It is worth highlighting here some of the limitations of the approach proposed in this paper. Firstly, extreme periodisation and small sample sizes can severely limit the inferential power. Admittedly, this is not a negative aspect of the approach advocated here. A good statistical method should not misguide the analyst and fully capture the uncertainty in the data. Results of experiments #3 and #4 are promising in this regard, but it is difficult to fully assess the extent to which the model can handle appropriately high levels of uncertainty in the input data.

Secondly, parametric approaches are strictly dependent on the selection of the growth model. Fitting a logistic growth model when the true time-frequency distribution is characterised by a rise-and-fall pattern can lead to problematic inference. The only exception to this might be the exponential growth model, whose parameter could still be interpreted the average growth rate within the time window of analyses. In the case of SPDs, posterior predictive checks are possible (Crema and Shoda 2021; Timpson et al. 2021) and can even reveal *when* interesting deviations from the model are observed. Implementing similar approaches with aoristic data is not trivial, with the exception of instances where there is a one-to-one association between a calendar date and an archaeological period. A potential way to tackle this approach is to fully account for the error structure of the data, i.e. estimate the probability of each event assigned to a particular time-span.

Thirdly, the models introduced here do not account for uncertainties in the start and end date of time-spans (i.e. *phase-boundary uncertainty*). When this information is available (e.g. Crema and Kobayashi 2020), one could potentially iteratively sample possible time-spans and generate a custom probability distribution to represent the time-span of each event. While this approach would account for additional levels of uncertainty, it would effectively “collapse” different forms of uncertainty into a single vector of probabilities across all time-blocks, ignoring consequently any interdependencies. Alternative approaches would require additional layers in the hierarchical model where the uncertainty of archaeological periodisation and the inference on the time-frequency distribution is carried out simultaneously. Further studies are required to explore the analytical and computational challenges this and possibly other solutions require.

Finally, it is worth mentioning that other attempts to provide a Bayesian framework to aoristic analysis are currently being developed in other fields. Van Lieshout and Markwits (van Lieshout and Markwitz 2022) approach the problem as a marked point process, whilst Briz-Redón (Briz-Redón 2023) associates a uniform prior to each observation, effectively carrying out a measurement error model similar to the one implemented by Crema and Shoda

(2021) for radiocarbon dates. Both solutions offer similar advantages to the solutions implemented here, although they are not designed to deal with instances where a user-defined vector of probabilities is employed to describe the most likely timing of the event within its time-span.

7. Conclusions

Developing a robust and general inferential tool for events described by archaeological periodisation is a challenging task. In contrast to radiocarbon dating, there are no formal principles that translate the uncertainty associated with the timing of an event into a probability distribution, with much of the legwork left to the subjective guesswork of experts. Start and end dates of archaeological periods are just general chronological reference points, and never designed to be approached as formal analytical units. Yet, the bulk of the archaeological record is described within this chronological framework. Aoristic analysis and related techniques offer a straightforward and easy-to-implement solution that allows archaeologists to explore the rich untapped resource.

I argue that even in an ideal condition, where the only form of uncertainty is determining when an event occurred within a period or phase, aoristic analysis a highly problematic approach for three reasons: 1) the summing of aoristic weights is mathematically unwarranted; 2) the non-random nature of archaeological periodisation can lead to misleading artefacts in aoristic sums; and 3) aoristic analysis is at its best a descriptive rather than an inferential statistic.

The paper introduced two Bayesian approaches and an associated R package that can provide an alternative to aoristic analysis in archaeology. The two approaches are tailored for different sets of objectives, but in both cases, the minimum requirements are the same for the aoristic analysis, with each archaeological event described by a chronological time-span or a vector of probability values for each time-block. While a comprehensive assessment of the robustness of the proposed approach was not possible, a series of experiments have been carried out to determine its accuracy and precision. Under all circumstances, the results showed a superior performance to aoristic analysis. The parametric approach can recuperate key values of interest accurately and consistently while simultaneously capable of formally capturing instances of high indeterminacy dictated by the nature of the underlying periodisation. Similarly, the ICAR-based non-parametric can visualise the extent of the uncertainty dictated by sampling error and periodisation, providing a more sound and cautious basis for the visual inspection of archaeological time-frequency data.

Acknowledgements

This work was supported by a Philip Leverhulme Prize (#PLP-2019–304). For the purpose of open access, the author has applied a Creative Commons Attribution (CC BY) licence to any Author Accepted Manuscript version arising from this submission.

References

Bailey, G., 2007, Time perspectives, palimpsests and the archaeology of time, *Journal of Anthropological Archaeology*, **26**, 198–223.

Baxter, M. J., and Cool, H. E. M., 2016, Reinventing the wheel? Modelling temporal uncertainty with applications to brooch distributions in Roman Britain, *Journal of archaeological science*, **66**, 120–7.

Bayliss, A., 2009, Rolling Out Revolution: Using Radiocarbon Dating in Archaeology, *Radiocarbon*, **51**(1), 123–47.

Bellanger, L., and Husi, P., 2012, Statistical tool for dating and interpreting archaeological contexts using pottery, *Journal of archaeological science*, **39**(4), 777–90.

Besag, J., 1974, Spatial interaction and the statistical analysis of lattice systems, *Journal of the Royal Statistical Society*, **36**(2), 192–225.

Bevan, A., Conolly, J., Hennig, C., Johnston, A., Quercia, A., Spencer, L., and Vroom, J., 2012, Measuring Chronological Uncertainty in Intensive Survey Finds, *Archaeometry*, **55**(2), 318–28.

Bevan, A., and Crema, E. R., 2021, Modifiable reporting unit problems and time series of long-term human activity, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **376**(1816), 20190726.

Blackwell, P. G., and Buck, C. E., 2003, The Late Glacial human reoccupation of north-western Europe: new approaches to space-time modelling, *Antiquity*, **77**(296), 232–40.

Briz-Redón, Á., 2023, A Bayesian aoristic logistic regression to model spatio-temporal crime risk under the presence of interval-censored event times, *arXiv [stat.AP]*.

Brozio, J. P., Müller, J., Furholt, M., Kirleis, W., Dreibrodt, S., Feeser, I., Dörfler, W., Weinelt, M., Raese, H., and Bock, A., 2019, Monuments and economies: What drove their variability in the middle-Holocene Neolithic?, *Holocene*, **29**(10), 1558–71.

Brunner, M., Felten, J. von, Hinz, M., and Hafner, A., 2020, Central European Early Bronze Age chronology revisited: A Bayesian examination of large-scale radiocarbon dating, *PloS one*, **15**(12), e0243719.

Buck, C. E., Cavanagh, W. G., and Litton, C. D., 1996, *Bayesian approach to interpreting archaeological data*, Wiley, Chirchester.

Carleton, W. C., and Groucutt, H. S., 2020, Sum things are not what they seem: Problems with point-wise interpretations and quantitative analyses of proxies based on aggregated radiocarbon dates, *Holocene*, 0959683620981700.

Carlson, D. L., 1983, COMPUTER ANALYSIS OF DATED CERAMICS: ESTIMATING DATES AND OCCUPATIONAL RANGES, *Southeastern Archaeology*, **2**(1), 8–20.

Collins‐Elliott, S. A., 2019, Quantifying artefacts over time: Interval estimation of a Poisson distribution using the Jeffreys prior, *Archaeometry*, **61**(5), 1207–22.

Crema, E. R., 2012, Modelling Temporal Uncertainty in Archaeological Analysis, *Journal of Archaeological Method and Theory*, **19**(3), 440–61.

Crema, E. R., 2015, Time and Probabilistic Reasoning in Settlement Analysis, In *Mathematics and Archaeology* (eds. J. A. Barceló, and I. Bogdanovic), 314–34, CRC Press, Boca Raton.

Crema, E. R., and Kobayashi, K., 2020, A multi-proxy inference of Jōmon population dynamics using bayesian phase models, residential data, and summed probability distribution of 14C dates, *Journal of archaeological science*, **117**, 105136.

Crema, E. R., and Shoda, S., 2021, A Bayesian approach for fitting and comparing demographic growth models of radiocarbon dates: A case study on the Jomon-Yayoi transition in Kyushu (Japan), *PloS one*, **16**(5), e0251695.

Crema, E. R., 2022, Statistical Inference of Prehistoric Demography from Frequency Distributions of Radiocarbon Dates: A Review and a Guide for the Perplexed, *Journal of Archaeological Method and Theory*.

Crema, E. R., Stevens, C. J., and Shoda, S., 2022, Bayesian analyses of direct radiocarbon dates reveal geographic variations in the rate of rice farming dispersal in prehistoric Japan, *Science advances*, **8**(38), eadc9171.

Franconi, T., Brughmans, T., Borisova, E., and Paulsen, L., 2023, From Empire-wide integration to regional localization: A synthetic and quantitative study of heterogeneous amphora data in Roman Germania reveals centuries-long change in regional patterns of production and consumption, *PloS one*, **18**(1), e0279382.

French, J. C., Riris, P., Fernandéz-López de Pablo, J., Lozano, S., and Silva, F., 2021, A manifesto for palaeodemography in the twenty-first century, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **376**(1816), 20190707.

Frerebeau, N., 2022, *kairos: Analysis of Chronological Patterns from Archaeological Count Data*.

Hinz, M., Schirrmacher, J., Kneisel, J., Rinne, C., and Weinelt, M., 2019, The

Chalcolithic–Bronze Age transition in southern Iberia under the influence of the 4.2 kyr event? A correlation of climatological and demographic proxies, *Journal of Neolithic Archaeology*, 1–26.

Hoebe, P. W., Peeters, J. H. M., and Arnoldussen, S., 2023, Parsing prehistoric patterns: Prospects and limitations of a big radiocarbon dataset for understanding the impact of climate on Late Palaeolithic and Mesolithic populations in northwest Europe (16–7.5 ka calBP), *Journal of archaeological science, reports*, **49**(103944), 103944.

Johnson, I., 2004, Aoristic Analysis: seeds of a new approach to mapping archaeological distributions through time, In *[Enter the Past] the E-way into the Four Dimensions of Cultural Heritage: CAA2003. BAR International Series 1227.* (eds. K. F. Ausserer, W. B. Rner, M. Goriany, and L. K.-V. Ckl), 448–52, Archaeopress, Oxford.

Kim, H., Lee, G.-A., and Crema, E. R., 2021, Bayesian analyses question the role of climate in Chulmun demography, *Scientific reports*, **11**(1), 23797.

Kleijne, J., Weinelt, M., and Müller, J., 2020, Late Neolithic and Chalcolithic maritime resilience? The 4.2 ka BP event and its implications for environments and societies in Northwest Europe, *Environmental research letters: ERL [Web site]*, **15**(12), 125003.

Knitter, D., Brozio, J. P., Hamer, W., Duttmann, R., Müller, J., and Nakoinz, O., 2019, Transformations and Site Locations from a Landscape Archaeological Perspective: The Case of Neolithic Wagrien, Schleswig-Holstein, Germany, *Land*, **8**(4), 68.

Levy, E., Finkelstein, I., Martin, M. A. S., and Piasetzky, E., 2022, The Date of Appearance of Philistine Pottery at Megiddo: A Computational Approach, *Bulletin of the American Society of Overseas Research*, **387**, 1–30.

van Lieshout, M. N. M., and Markwitz, R. L., 2022, State estimation for aoristic models, *Scandinavian journal of statistics, theory and applications*.

Lyman, R. L., and O’Brein, M. J., 2006, *Measuring Time with Artifacts: A History of Methods in American Archaeology*, University of Nebraska Press, Lincoln.

Mischka, D., 2004, Aoristische Analyse in der Archäologie, *Archäologische Informationen*, **27**(2), 233–43.

Murray, T., 1999, *Time and Archaeology*, Routledge, London.

Orton, C., 1973, The tactical use of models in archaeology - the SHERD project, In *The explanation of culture change* (ed. C. Renfrew), 137–9, Duckworth, London.

Orton, D., Morris, J., and Pipe, A., 2017, Catch Per Unit Research Effort: Sampling

Intensity, Chronological Uncertainty, and the Onset of Marine Fish Consumption in Historic London, *Open Quaternary*, **3**(1).

Palmisano, A., Bevan, A., and Shennan, S., 2017, Comparing archaeological proxies for long-term population patterns: An example from central Italy, *Journal of archaeological science*, **87**, 59–72.

Palmisano, A., Woodbridge, J., Roberts, N., Bevan, A., Fyfe, R., Shennan, S., Cheddadi, R., Greenberg, R., Kaniewski, D., Langgut, D., Leroy, S. A. G., Litt, T., and Miebach, A., 2019, Holocene landscape dynamics and long-term population trends in the Levant, *Holocene*, 0959683619826642.

Pollard, D., 2021, All equal in the presence of death? A quantitative analysis of the Early Iron Age cemeteries of Knossos, Crete, *Journal of Anthropological Archaeology*, **63**, 101320.

Ratcliffe, J., 2022, *aoristic: Generates Aoristic Probability Distributions*.

Ratcliffe, J. H., and McCullagh, M. J., 1998, Aoristic crime anaysis, *International journal of geographical information science: IJGIS*, **12**(7), 751–64.

Ratcliffe, J. H., 2000, Aoristic analysis: the spatial interpretation of unspecifed temporal events, *Inernational Journal of Geographical Information Science*, **14**(7), 669–79.

Roalkvam, I., 2022, Exploring the composition of lithic assemblages in Mesolithic South-Eastern Norway, *Journal of Archaeological Science: Reports*, **42**, 103371.

Roberts, J. M., Jr, Mills, B. J., Clark, J. J., Haas, W. R., Jr, Huntley, D. L., and Trowbridge, M. A., 2012, A method for chronological apportioning of ceramic assemblages, *Journal of archaeological science*, **39**(5), 1513–20.

Romandini, M., Crezzini, J., Bortolini, E., Boscato, P., Boschin, F., Carrera, L., Nannini, N., Tagliacozzo, A., Terlato, G., Arrighi, S., Badino, F., Figus, C., Lugli, F., Marciani, G., Oxilia, G., Moroni, A., Negrino, F., Peresani, M., Riel-Salvatore, J., Ronchitelli, A., Spinapolice, E. E., and Benazzi, S., 2020, Macromammal and bird assemblages across the late Middle to Upper Palaeolithic transition in Italy: an extended zooarchaeological review, *Quaternary international: the journal of the International Union for Quaternary Research*, Peopling dynamics in the Mediterranean area between 45 and 39 ky ago: state of art and new data, **551**, 188–223.

Romanowska, I., Brughmans, T., Bes, P., Carrignon, S., Egelund, L., Lichtenberger, A., and Raja, R., 2021, A Study of the Centuries-Long Reliance on Local Ceramics in Jerash Through Full Quantification and Simulation, *Journal of Archaeological Method and Theory*.

Rue, H., and Held, L., 2004, *Gaussian Markov random fields: Theory and* *applications*, Chapman & Hall/CRC Monographs on Statistics and Applied Probability, Chapman and Hall, London, England.

Shennan, S., Downey, S. S., Timpson, A., Edinborough, K., Colledge, S., Kerig, T., Manning, K., and Thomas, M. G., 2013, Regional population collapse followed initial agriculture booms in mid-Holocene Europe, *Nature communications*, **4**, ncomms3486.

Steinmann, L., and Weissova, B., 2021, datplot: A New R Package for the Visualization of Date Ranges in Archaeology, *Advances in Archaeological Practice*, **9**(4), 288–98.

Steponaitis, V. P., and Kintigh, K. W., 1993, Estimating site occupation spans from dated artifact types: Some new approaches, In *Archaeology of Eastern North America: Papers in Honor of Stephen Williams* (ed. J. Stoltman), Vol. 25, 349–61, Archaeological Report, Mississippi Department of Archives and History, Jackson.

Stoddart, S., Woodbridge, J., Palmisano, A., Mercuri, A. M., Mensing, S. A., Colombaroli, D., Sadori, L., Magri, D., Di Rita, F., Giardini, M., Mariotti, M., Montanari, C., Bellini, C., Florenzano, A., Torri, P., Bevan, A., Shennan, S., Fyfe, R., and Roberts, N., 2019, Tyrrhenian central Italy: Holocene population and landscape ecology, *Holocene*, 0959683619826696.

Taelman, D., 2022, Marble trade in the Roman Mediterranean: a quantitative and diachronic study, *Journal of Roman Archaeology*, **35**(2), 848–75.

Timpson, A., Colledge, S., Crema, E., Edinborough, K., Kerig, T., Manning, K., Thomas, M. G., 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**, 549–57.

Timpson, A., Barberena, R., Thomas, M. G., Méndez, C., and Manning, K., 2021, Directly modelling population dynamics in the South American Arid Diagonal using 14C dates, *Philosophical transactions of the Royal Society of London. Series B, Biological sciences*, **376**(1816), 20190723.

de Valpine, P., Turek, D., Paciorek, C. J., Anderson-Bergman, C., Lang, D. T., and Bodik, R., 2017, Programming With Models: Writing Statistical Algorithms for General Model Structures With NIMBLE, *Journal of computational and graphical statistics: a joint publication of American Statistical Association, Institute of Mathematical Statistics, Interface Foundation of North America*, **26**(2), 403–13.

de Valpine, P., Adler, C., Turek, D., Michaud, N., Anderson-Bergman, C., Obermeyer, F., Wehrhahn Cortes, C., Rodrìguez, A., Temple Lang, D., and Paganin, S., 2020, *NIMBLE: MCMC, Particle Filtering, and Programmable Hierarchical Modeling*.

Verhagen, P., Vossen, I., Groenhuijzen, M. R., and Joyce, J., 2016, Now you see them, now you don't: Defining and using a flexible chronology of sites for spatial analysis of Roman settlement in the Dutch river area, *Journal of Archaeological Science: Reports*, **10**, 309–21.

Yubero-Gómez, M., Rubio-Campillo, X., and López-Cachero, J., 2016, The study of spatiotemporal patterns integrating temporal uncertainty in late prehistoric settlements in northeastern Spain, *Archaeological and anthropological sciences*, **8**(3), 477–90.

Ziedler, J. A., Buck, C. E., and Litton, C. D., 1998, Integration of Archaeological Phase Information and Radiocarbon Results from the Jama River Valley, Ecuador: A Bayesian Approach, *Latin American Antiquity*, **9**(2), 160–79.